

Rapport de méthodes

Estimation de la couverture du recensement de la population de l'an 2000

0

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

Échantillon pour l'estimation
de la sous-couverture
(P-sample) et qualité
du cadre de sondage
des bâtiments

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

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

0	Statistische Grundlagen und Übersichten	0	Bases statistiques et produits généraux
1	Bevölkerung	1	Population
2	Raum und Umwelt	2	Espace et environnement
3	Arbeit und Erwerb	3	Vie active et rémunération du travail
4	Volkswirtschaft	4	Economie nationale
5	Preise	5	Prix
6	Industrie und Dienstleistungen	6	Industrie et services
7	Land- und Forstwirtschaft	7	Agriculture et sylviculture
8	Energie	8	Energie
9	Bau- und Wohnungswesen	9	Construction et logement
10	Tourismus	10	Tourisme
11	Verkehr und Nachrichtenwesen	11	Transports et communications
12	Geld, Banken, Versicherungen	12	Monnaie, banques, assurances
13	Soziale Sicherheit	13	Protection sociale
14	Gesundheit	14	Santé
15	Bildung und Wissenschaft	15	Education et science
16	Kultur, Medien, Zeitverwendung	16	Culture, médias, emploi du temps
17	Politik	17	Politique
18	Öffentliche Verwaltung und Finanzen	18	Administration et finances publiques
19	Rechtspflege	19	Droit et justice
20	Einkommen und Lebensqualität der Bevölkerung	20	Revenus et qualité de vie de la population
21	Nachhaltige Entwicklung und regionale Disparitäten	21	Développement durable et disparités régionales

Estimation de la couverture du recensement de la population de l'an 2000

Échantillon pour l'estimation de la sous-couverture
(P-sample) et qualité du cadre de sondage des bâtiments

Auteurs

Anne Renaud et Jann Potterat

Office fédéral de la statistique

Editeur

Office fédéral de la statistique

Préambule

Le projet «estimation de la couverture du recensement» fait partie des projets visant à évaluer la qualité du recensement de la population de l'an 2000 (RFP2000). Ce rapport présente la construction et la pondération de l'échantillon permettant d'estimer la composante de sous-couverture du recensement. Il décrit également une analyse de la qualité de la base de sondage utilisée pour le tirage de cet échantillon. Le travail a été réalisé par Anne Renaud et Jann Potterat du service de méthodes statistiques (METH).

Les résultats décrits dans ce rapport ont été obtenus grâce à l'apport de plusieurs personnes. Un grand merci à Philippe Eichenberger de METH pour les discussions méthodologiques et la relecture du rapport. Merci également à Rachel Fritschi, Stéphane Rey et Françoise Rossé du projet RFP2000 pour le travail effectué sur les données de l'enquête de couverture et celles du recensement.

Résumé

Les défauts de couverture du recensement de la population de l'an 2000 (RFP2000) sont estimés selon le système dual, c'est-à-dire sur la base de deux échantillons de personnes. Ces deux échantillons permettent respectivement d'estimer les composantes de sous-couverture et de sur-couverture.

Le P-sample de 49'883 personnes est l'échantillon permettant d'estimer la composante de sous-couverture du recensement. Il découle d'un plan d'échantillonnage à plusieurs niveaux. La pondération du P-sample prend en compte l'information sur le plan d'échantillonnage, la non-réponse des ménages et l'incertitude sur l'appartenance ou pas à la population d'intérêt.

La qualité de la base de sondage des bâtiments utilisée pour la construction du P-sample est également analysée afin d'avoir une information sur le biais potentiel des estimations de couverture.

Mots-clé

rapport de méthodes; recensement de la population; VZ2000; RFP2000; enquête de couverture; EC2000; sous-couverture; plan d'échantillonnage; post-enumeration; coverage; P-sample

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:	338-0023
Prix:	gratuit
Série:	Statistique de la Suisse
Domaine:	0 Bases statistiques et produits généraux
Langue du texte original:	Français
Graphisme/Layout:	OFS
Copyright:	OFS, Neuchâtel 2004 La reproduction est autorisée, sauf à des fins commerciales, si la source est mentionnée.
ISBN:	3-303-00276-2

Table des matières

1	Introduction	5
2	Rappel	5
2.1	Population d'intérêt	5
2.2	Plan d'échantillonnage de l'EC2000	5
3	Construction du P-sample et des listes de référence pour la pondération	6
3.1	Liste des bâtiments de l'EC2000	6
3.2	Liste des ménages	8
3.3	Intégration des données sur les interviews et P-sample provisoire	8
3.4	Intégration des informations complémentaires et P-sample final	9
4	Pondération du P-sample	10
4.1	Ajustement pour la non-réponse au niveau des ménages	10
4.2	Poids finaux des ménages et personnes	16
4.3	Autres approches pour la correction de la non-réponse	19
5	Dernières modifications du P-sample	19
6	Qualité du cadre de sondage des bâtiments	20
6.1	Première allure	21
6.2	Comparaison entre le REAB et le RBL2000	21
6.3	Analyse au niveau des bâtiments de l'EC2000	25
6.4	Couverture des bâtiments	26
7	Conclusion	27

1 Introduction

L'estimation de la couverture du recensement fédéral de la population de l'an 2000 (RFP2000) repose sur la méthode duale (dual system method).

L'estimateur dual du total de la population est basé sur deux échantillons de personnes : le E-sample (échantillon d'*Enumeration*) et le P-sample (échantillon de *Post-enumeration*).

Le E-sample est un échantillon dans les données du recensement. Il permet d'estimer la surcouverture du recensement ; voir Renaud (2003) pour le plan et le tirage.

Le P-sample est un échantillon indépendant du recensement. Il correspond à un sous-ensemble des personnes interrogées lors de l'enquête de couverture (EC2000) ; voir Renaud et Eichenberger (2002) et Renaud (2002). Il permet d'estimer la composante de sous-couverture du recensement.

Ce rapport documente la construction du P-sample, depuis l'échantillonnage en plusieurs phases jusqu'à l'intégration des informations recueillies lors de la phase d'appariement avec le recensement. Il présente également la méthode de pondération et une analyse de la qualité de la base de sondage de bâtiments utilisée pour la construction de cet échantillon.

Le présent rapport fait suite aux rapports de méthodes Renaud et Eichenberger (2002) et Renaud (2003) déjà publiés ; rapports dans lesquels la démarche générale de l'estimation de la couverture est présentée.

2 Rappel

2.1 Population d'intérêt

La population d'intérêt pour les estimations est la population résidente en Suisse, selon le domicile économique, vivant dans un ménage privé et dans l'état au jour du recensement (référence exacte : minuit dans la nuit du 4 au 5 décembre 2002).

Le P-sample est l'ensemble des personnes interrogées lors de l'enquête de couverture (EC2000) et faisant partie de la population d'intérêt.

2.2 Plan d'échantillonnage de l'EC2000

Le plan d'échantillonnage de l'EC2000 est un plan à plusieurs niveaux avec deux bases d'échantillonnage indépendantes.

On a tiré 285 parmi les 3699 numéros postaux (NPA) à 6 chiffres de la Suisse sans le Tessin (plan NORD) et 20 parmi les 250 communes/secteurs du Tessin (plan TICINO). L'échantillon au premier niveau est donc composé de 305 PSUs (Primary Sampling Units) dans 28 strates (variable : `stradap`).

Deux PSUs du NORD n'ont pas participé à l'EC2000 : Founex (`stradap`=9, NPA=129700) et Mayens-de-Riddes (`stradap`=13, NPA=191800). Un ensemble de tournées de facteurs puis

de bâtiments ont été tirés dans les 303 PSUs restants. La liste des ménages dans les bâtiments de l'échantillon a été construite avec l'aide des facteurs. Un sous-échantillonnage de bâtiments a ensuite permis d'atteindre l'objectif de env. 27'000 ménages. Ce sous-échantillonnage correspond à une réduction de 17'992 à 15'877 bâtiments.

Des interviews ont été organisées pour l'ensemble des env. 27'000 ménages. Les informations de base du recensement, complétées par quelques questions nécessaires aux estimations, ont été recueillies pour tous les membres des ménages contactés.

3 Construction du P-sample et des listes de référence pour la pondération

La construction du P-sample et des listes de référence nécessaires à la pondération est effectuée comme suit :

1. mise en ordre de la liste des bâtiments de l'EC2000 (base de la pondération du P-sample) ;
2. ajustement de la liste des ménages dans les bâtiments (cadre de sondage au niveau des ménages) ;
3. intégration des données provenant des interviews et définition du P-sample provisoire ;
4. intégration des informations provenant de la procédure d'appariement et définition du P-sample final.

3.1 Liste des bâtiments de l'EC2000

La liste des bâtiments de l'EC2000 forme la base pour la sélection et la pondération des ménages et personnes atteintes durant l'EC2000. Il s'agit d'un total de 15'877 bâtiments, dont 14'697 pour le plan NORD et 1'180 pour le plan TICINO.

Chaque bâtiment k a une probabilité de sélection π_k définie par le plan d'échantillonnage à plusieurs niveaux de l'EC2000 ; voir Renaud et Eichenberger (2002). L'inverse de cette probabilité est le poids d'échantillonnage $w_k = 1/\pi_k$ du bâtiment.

Le poids des bâtiments est adapté pour tenir compte de la non-réponse des 2 PSUs du plan NORD. On reprend l'ajustement pondéré décrit dans Renaud (2003) :

$$f_{ih}^{(2)} = f_h^{(2)} = \frac{\sum_{i \in s_h} w_{ih}}{\sum_{i \in \tau_h} w_{ih}} \quad (1)$$

avec s_h l'échantillon de taille m_h tiré parmi les M_h PSUs de la strate h , $\tau_h \subset s_h$ le sous-échantillon de taille r_h formé des PSUs ayant participé à l'enquête, et $w_{ih} = 1/\pi_{ih}$ l'inverse de la probabilité d'inclusion π_{ih} du PSU i de la strate h dans s_h .

Le poids final $w_k^{(2)}$ du bâtiment k est donc :

$$w_k^{(2)} = f_{ih}^{(2)} w_k = \begin{cases} 1.03186 w_k & \text{dans les 14 PSUs de la strata 9} \\ 1.24418 w_k & \text{dans les 3 PSUs de la strata 13} \\ w_k & \text{dans les 286 PSUs restants} \end{cases} \quad (2)$$

Parmi les 14'697 bâtiments du NORD, 12'710 (86.5%) ont été tirés dans la liste du Registre des adresses des bâtiments (REAB) utilisée pour la préimpression des questionnaires du recensement (**Objet** $\in \{1, 4\}$; voir Table 1). Les 1987 (13.5%) restants ont été tirés parmi les bâtiments non recensés du REAB (**Objet** $\in \{2, 8, 9\}$). Les bâtiments du NORD ont un identificateur univoque dans les données de l'EC2000 et dans les données du recensement (**Ref_code**).

TAB. 1 – Variable **Objet** dans le REAB.

Objet	Description
1	avec utilisation à des fins d'habitation
2	sans utilisation à des fins d'habitation
4	habitation mobile/provisoire
7	bâtiment fictif (1 par commune, "Kirchspitze")
8	inconnu
9	démoli

Les 1'180 bâtiments du Tessin ont été tirés dans une liste de bâtiments fournie par l'Ufficio di statistica del Cantone Ticino (Ustat). Cette liste, provenant d'une statistique fiscale exhaustive (Stima), ne contient pas d'identificateur permettant de faire un lien univoque avec les données du recensement. Pour les besoins du projet, nous créons des identificateurs **Ref_code** fictifs, supérieurs aux valeurs du NORD (numérotation des cas + 10^8).

Le code **Codbat1** est déterminé pour chaque bâtiment de l'EC2000 afin d'indiquer s'il est habité ou pas ; voir Table 2. On dit qu'un bâtiment est habité si on a trouvé un ou des ménages lors de l'élaboration des listes de ménages. Les différents types de bâtiments vides sont déterminés sur la base des formulaires bâtiments remplis par les facteurs (travail effectué dans la section BHS de l'OFS) et l'intégration de différents cas particuliers ; voir Section 3.2.

TAB. 2 – Codification **Codbat1** des bâtiments de l'EC2000. Nombre de cas n et proportion p .

Codbat1	Description	n	p
1	habité	12'939	81.5%
2	vide mais clairement habitable	538	3.4%
3	vide et clairement inhabitable	1791	11.3%
4	vide doublons	37	0.2%
5	vide avec un statut peu clair	269	1.7%
6	inexistant (pas trouvé)	303	1.9%
	Total	15'877	100%

Il est intéressant de noter que certains bâtiments considérés comme inhabitables (**Objet** $\notin \{1, 4\}$) sont en fait habitables selon les informations relevées par les facteurs (572 sur 1987, *i.e.* 29%). Le choix de l'élargissement de la base de sondage aux bâtiments notés "sans utilisation à des fins d'habitation", "inconnu" ou "démoli" était donc adéquat. Il faut noter que ces différences peuvent être dues à la qualité de la liste de bâtiments mais également à des erreurs de localisation

de la part des facteurs ; notamment lorsque l'identification se faisait à l'aide de cartes (bâtiments sans adresse postale). Les questions relatives à la qualité de la liste de base des bâtiments sont abordées dans la Section 6.

3.2 Liste des ménages

La liste des ménages habitant les bâtiments de l'EC2000 forme le cadre de sondage au niveau des ménages. On fait l'hypothèse que la liste construite avec l'aide des facteurs est complète. Ces derniers étaient en effet chargés d'élaborer des listes exhaustives de ménages par bâtiment à partir des informations partielles fournies sur la base de l'annuaire de téléphone électronique ; voir la procédure d'enquête de l'EC2000 dans Renaud et Eichenberger (2002).

La qualité de la liste est importante pour les estimations de couverture. Il est probable que la difficulté de recenser une personne est corrélée avec la difficulté de lister son ménage dans la liste de ménages. On a donc potentiellement tendance à sous-estimer la sous-couverture de la population. Des efforts particuliers ont donc été mis dans l'élaboration de cette liste afin d'éviter des défauts systématiques.

La liste exhaustive des ménages reprend la liste des 27'331 ménages préparée pour les interviews (rem : 27'330 après la perte d'un élément). Elle est adaptée pour tenir compte de différents problèmes et cas particuliers :

1. Les doubles détectés avant les interviews sont exclus de la liste (9 cas).
2. Les ménages oubliés durant la saisie et retrouvés ensuite sont ajoutés à la liste (59 cas).
3. Les ménages de Brusino-Arsizio listés après le début des interviews sont ajoutés (6 cas).
4. Les ménages de 17 bâtiments pour lesquels les facteurs avaient indiqué des changements d'adresse sur le formulaire du bâtiment n'avaient pas été saisis. 4 bâtiments sont clairement existants mais sans ménages, 3 bâtiments sont des parties de bâtiments multiples et considérés comme pas réels, et 10 bâtiments sont clairement habités par 12 ménages à saisir. Ces 12 ménages sont donc ajoutés à la liste.

La liste servant de base pour l'échantillonnage comprend donc 27'398 ménages sur un total de 27'407 listés. On notera que 77 ménages font partie de la liste de référence mais n'avaient aucune chance d'être contactés parce qu'ils font partie des cas 2-4 définis ci-dessus.

3.3 Intégration des données sur les interviews et P-sample provisoire

Parmi les 27'321 ménages traités par l'institut, 24'914 ont été contactés¹ (91%).

Un ensemble de 21'350 (78%) ont répondu à l'interview, sont des ménages privés et ont au moins un membre faisant partie de la population d'intérêt (contacté au domicile économique,

¹On considère qu'un ménage a été contacté s'il y a eu un contact, même bref, par téléphone ou en personne, ou si une visite sur le terrain a permis de déterminer que le ménage cherché est en fait une entreprise ou un logement de vacances.

résidant au jour du recensement et pas un double) ; voir Section 4 pour plus d'information sur la non-réponse.

Il est important de noter que les personnes contactées à leur domicile civil non économique ou dans un ménage collectif ne font pas partie de la population.

La liste des ménages est formée sur la base des listes élaborée par les facteurs du 18 janvier au 9 février 2001 pour le NORD, resp. du 28 février au 16 mars 2001 pour le Tessin. De son côté, le P-sample est défini sur la base des informations relevées durant les interviews du 17 avril au 29 mai 2001. On espère cependant que ce décalage, qui peut être lié à des modifications des types de domicile et de ménages, n'aura pas une grande influence sur les résultats.

Les données des 147 interviews partielles n'ont pas été fournies par l'institut de sondage. Seul le numéro de la dernière question saisie était disponible. Les réponses partielles sont par conséquent considérées comme des non-réponses complètes, avec information parfois sur le fait d'être ou pas un vrai ménage. On considère par exemple avoir atteint un vrai ménage si l'information a été saisie pour 2 personnes et que l'interview a été interrompue au milieu de la troisième. Après contrôle dans la section BHS de l'OFS, un ensemble de 109 parmi les 147 interviews partielles peuvent être considérées comme des vrais ménages.

Un ensemble de 355 ménages sont des doubles (1.3% du total des ménages traités). Il s'agit de 280 ménages ayant répondu 2 fois entièrement à l'interview (par exemple ménages uniques atteints par deux noms différents) et de 75 ménages pas interrogés car ils avaient indiqué qu'ils avaient déjà été contactés pour la même enquête. Ces derniers ont été retenus parmi les 147 ménages ayant indiqué qu'ils avaient déjà été interviewés (cas retravaillés dans la section BHS de l'OFS ; 72 refusés).

Les 21'350 ménages ayant répondu aux interviews contiennent 50'070 personnes faisant partie de la population d'intérêt. Ces dernières forment le P-sample provisoire utilisé pour la recherche des appariements avec le recensement et comme base pour le plan d'échantillonnage du E-sample ; voir Renaud (2003).

3.4 Intégration des informations complémentaires et P-sample final

L'appariement entre le P-sample provisoire et le recensement a permis de découvrir de nouveaux éléments hors population dans le P-sample provisoire. Un ensemble de 184 nouveaux éléments hors population ont été trouvés dans le P-sample provisoire, notamment par le biais d'un contrôle manuel des personnes appariées avec une personne unique dans le recensement (principalement des doublons, travail effectué dans la section BHS de l'OFS).

Le P-sample final passe donc de 50'070 à 49'886 personnes. Cette correction au niveau des personnes diminue le nombre de ménages de 21'350 à 21'284.

Le nombre de personnes dans le P-sample final par variante, langue et taille de commune montre que les objectifs définis dans le plan d'échantillonnage sont presque entièrement atteints, voir Table 3. Exemple : au moins 10'000 personnes dans la variante TRANSIT, dans les moyennes et grandes communes. Le total de 10'000 n'est pas tout à fait atteint pour la variante CLASSIC.

Il faut noter que les valeurs diffèrent selon le choix de la commune de référence². La commune politique n'est en effet pas relevée dans l'EC2000 ; voir discussion dans Renaud (2003).

TAB. 3 – Nombre de personnes dans le P-sample, selon la variante, la langue et la taille de la commune majoritaire dans le PSU (lien par le PSU).

Variante	Langue				Taille commune			Total
	D	F	I	R	Pet.	Moy.	Grand	
CLASSIC	6031	3091	369	124	4367	5248		9615
SEMI-CLASSIC	2347	2067	67	359	909	3931		4840
TRANSIT	23787	9239		135	135	18099	14927	33161
FUTURE	727					281	446	727
TESSIN			1543		313	542	688	1543
Total	32892	14397	1979	618	5724	28101	16061	49886

4 Pondération du P-sample

La pondération des 49'886 personnes du P-sample tient compte de la probabilité d'inclusion de la personne dans l'EC2000, d'une correction pour la non-réponse des deux PSUs manquants et d'une correction pour la non-réponse au niveau des ménages. Toutes les personnes sont interviewées dans les ménages de l'échantillon et il n'y a pas de non-réponse au niveau des personnes. Le poids de la personne du P-sample est donc identique au poids de son ménage.

Le poids w_p de la personne p dans le ménage q du bâtiment k dans le PSU i de la strate h a la forme :

$$w_p = w_k f_{ih}^{(2)} g_q = w_k^{(2)} g_q \quad (3)$$

avec w_k le poids d'échantillonnage du bâtiment, $f_{ih}^{(2)}$ l'ajustement pour la non-réponse des PSUs et g_q l'ajustement pour la non-réponse au niveau des ménages. On note $w_k^{(2)} = w_k f_{ih}^{(2)}$ le poids final du bâtiment, voir Section 3.1.

Il n'y a pas de correction supplémentaire des poids car aucune information auxiliaire fiable et indépendante du recensement n'est disponible pour le nombre de personnes, ménages ou bâtiments en Suisse.

4.1 Ajustement pour la non-réponse au niveau des ménages

Les critères pour la création de la liste des ménages dans les bâtiments de l'échantillon étaient peu restrictifs dans la procédure d'enquête de l'EC2000. La liste comporte donc un nombre non négligeable de faux ménages tels que des commerces ou des bureaux.

²Les valeurs obtenues pour la commune majoritaire dans le PSU d'échantillonnage, diffèrent peu de celles obtenues sur la base de la commune majoritaire liée à l'adresse au jour du recensement. Des différences sont cependant sensibles si la commune du bâtiment tiré dans le REAB pour l'EC2000 est utilisée au lieu de la commune majoritaire.

L'effet "population ou pas" doit être pris en compte dans la correction de la non-réponse car on ne peut pas partir de l'hypothèse que les ménages sans information sont tous de faux ou de vrais ménages.

L'ajustement g_q pour la non-réponse au niveau des ménages est effectuée dans des cellules considérées comme homogènes du point de vue des probabilités de réponse. Les cellules sont construites sur la base d'informations sur le plan d'échantillonnage (strates, PSUs) et sur la méthode de relevé (CATI ou CAPI³)

Dans chaque cellule, on détermine un facteur d'ajustement des poids. Cet ajustement prend en compte le fait que certains ménages non-répondants ne font en fait pas partie de la population d'intérêt (entreprise, ménages collectifs, etc.). Il combine donc une probabilité de faire partie de la population avec une probabilité de répondre.

L'ajustement de la non-réponse est effectuée en 4 phases : (1) définition de variables auxiliaires, (2) traitement des ménages absents des listes de ménages à interroger, (3) création des cellules d'ajustement, et (4) calcul de la valeur de l'ajustement dans chaque cellule.

4.1.1 Définition de variables auxiliaires

Des variables auxiliaires sont construites afin d'identifier les ménages non-répondants, leur appartenance ou pas à la population d'intérêt et la méthode de relevé utilisée lors de l'EC2000 (CATI ou CAPI) ; voir Table 4.

TAB. 4 – Variables auxiliaires au niveau du ménage : `feh_lco` pour l'existence des données de relevé, `inner_h` pour l'appartenance à la population et `art` pour la méthode de relevé.

Variable	Description
<code>feh_lco</code>	1 : informations manquantes (non-contact ou contact suivi d'une non-réponse) 0 : informations existantes (contact et réponse)
<code>inner_h</code>	1 : dans la population 0 : pas dans la population '.' (missing) : statut inconnu
<code>art</code>	1 : CATI 0 : CAPI '.' (missing) : pas d'enquête

On rappellera qu'un ménage fait partie de la population d'intérêt si le ménage est privé et qu'au moins un de ces membres fait partie de la population résidente selon le domicile économique, au jour du recensement.

Nous avons des informations complètes pour 80% des ménages listés ; voir Table 5. Parmi ceux-ci, un ensemble de 650 ne font pas partie de la population d'intérêt car il s'agit de doubles, de ménages hors domicile économique ou de ménages collectifs.

³CATI : Computer Assisted Telephone Interview, CAPI : Computer Assisted Personal Interview.

TAB. 5 – Distribution des variables `fehlco`, `innerh` et `art`. Nombres absolus n et proportions p [%].

	<code>fehlco</code>		<code>innerh</code>		<code>art</code>	
	n	p	n	p	n	p
.			4'291	15.7%	77	0.3%
0	21'934	80.0%	1'832	6.7%	3'240	11.8%
1	5'473	20.0%	21'284	77.7%	24'090	87.9%
Total	27'407	100%	27'407	100%	27'407	100%

On note que 11.8% des ménages étaient contactés par CAPI et que le taux de réponse varie fortement selon le type de relevé (82% pour CATI et 68% pour CAPI).

Parmi le nombre non négligeable de 4'291 ménages sans information sur l'appartenance ou pas à la population (`fehlco`=.), le nombre de non-contacts l'emporte sur le nombre de contacts avec non-réponse ; voir Table 6. La différence est nette pour le CAPI (`art`=0) ; moins pour le CATI (`art`=1).

TAB. 6 – Cas sans information sur l'appartenance à la population (`fehlco`=.). Non-contact, resp. contact avec non-réponse, par type de relevé.

	<code>art</code>			
	.	0	1	Total
non-contact	77	549	1'961	2'587
contact mais non-réponse	0	78	1'626	1'704
Total	77	627	3'587	4'291

Pour les besoins des ajustements, on définit les variables indicatrices c_j , $j = 1, \dots, 4$. Ces variables prennent les valeurs 0 ou 1 selon différentes combinaisons de `fehlco` et `innerh` ; voir Table 7. La variable c_1 prend par exemple la valeur 1 pour le ménage q s'il a donné des informations (`fehlco`=0) et qu'il fait partie de la population (`innerh`=1). On définit également le nombre de ménage $n_{c_j} = \sum_q c_j$ et la somme des poids des ménages $w_{c_j} = \sum_q c_j w_q$ pour $j = 1, \dots, 4$.

TAB. 7 – Définition des variables c_j , et nombre de ménages n_{c_j} , $j = 1, \dots, 4$.

Variable	<code>innerh</code>	<code>fehlco</code>	n_{c_j}
c_1	1	0	21'284
c_2	0	0	650
c_3	0	1	1'182
c_4	.	1	4'291
Total			27'407

Les variables `fehlco` et `innerh`, et donc également les C_j , sont construites sur la base du code `code_inter` de l'EC2000 ; voir Table 8 :

- `fehlco`=0 pour `code_inter` $\in \{6, 7, 28, 52, 53, 99\}$ et 1 sinon ;
- `innerh`=1 pour `code_inter` $\in \{99\}$, 0 pour `code_inter` $\in \{4, 5, 6, 7, 26, 27, 28, 52, 53\}$ et une partie du code 99, et missing sinon.

On notera que les ménages avec interviews partielles (`code_inter`= 50 ou 51) sont tous considérés comme peu clairs du point de vue de l'appartenance ou pas à la population. L'information sur l'appartenance probable (ou pas) de ces ménages à la population n'est pas utilisée.

4.1.2 Types de relevé `art` manquants

Un groupe de 77 ménages n'avaient aucune chance d'être contactés car ils ont été listés après le début des interviews ; voir Section 3.2 (`art`=.).

Le type de relevé `art` étant important dans le cadre de la correction de la non-réponse, les 77 cas sans information sur le type de relevé sont mis par défaut à CATI (`art`=1). Ces cas particuliers seront donc traités avec la masse des CATI afin d'éviter un apport d'instabilité parmi le petit nombre de CAPI.

TAB. 8 – Classification des ménages par `code_inter` et c_j . Les codes `code_inter` entre 1 et 19 sont réservés aux CATI, et ceux entre 20 et 31 sont réservés aux CAPI. Les codes `code_inter` entre 50 et 71 ont été ajoutés après les interviews.

code_ inter	Description	n_{c_j}				Total
		1	2	3	4	
1	refus du ménage	0	0	0	747	747
2	pas d'info au téléphone	0	0	0	288	288
3	méfiance face au mandataire	0	0	0	39	39
4	adresse commerciale	0	0	645	0	645
5	logement/maison de vacances	0	0	133	0	133
6	ménage collectif	0	49	0	0	49
7	ménage a déjà répondu - doublons effectifs	0	75	0	0	75
8	autre raison de non-réponse	0	0	0	218	218
9	problème de langue	0	0	0	26	26
10	maladie/infirmité	0	0	0	160	160
11	répondeur automatique	0	0	0	3	3
12	voice-Mail-System	0	0	0	1	1
13	modem/Fax	0	0	0	1	1
14	fausse connection sous ce numéro	0	0	0	61	61
15	pas de connection sous ce numéro	0	0	0	11	11
16	numéro de téléphone non valable	0	0	0	2	2
17	autres problèmes techniques	0	0	0	141	141
18	numéro de téléphone a changé	0	0	0	549	549
19	ménage pas atteint	0	0	0	1'129	1'129
20	refus du ménage	0	0	0	77	77
21	les personnes ont déménagé vers un lieu inconnu	0	0	0	101	101
22	ménage pas atteint	0	0	0	226	226
23	ménage (personne) inconnue	0	0	0	75	75
24	ménage n'existe plus (décès)	0	0	0	32	32
25	ménage pas trouvé (adresse incomplète)	0	0	0	9	9
26	adresse commerciale	0	0	62	0	62
27	logement/maison de vacances	0	0	342	0	342
28	ménage collectif	0	23	0	0	23
29	autres problèmes de langue	0	0	0	1	1
30	autres	0	0	0	98	98
31	formulaire manque	0	0	0	8	8
32	effacé pour raison de mauvaise qualité	0	0	0	1	1
50	Interview partielle auprès d'un ménage	0	0	0	109	109
51	Interview partielle, probablement hors population	0	0	0	38	38
52	Ménages à double détectés avant interviews (IHA et OFS)	0	9	0	0	9
53	Ménages à double détectés après les tests de plausibilité	0	280	0	0	280
54	Ménages saisis après interviews (Brusino-Arsizio)	0	0	0	6	6
55	Ménages oubliés lors de la saisie	0	0	0	59	59
56	Ménages dans des bâtiments avec changement d'adresse	0	0	0	12	12
71	Doublons annoncés par les ménages, mais non vérifiés	0	0	0	63	63
99	Interview complète (au niveau du ménage)	21'284	214	0	0	21'498
Total		21'284	650	1'182	4'291	27'407

4.1.3 Cellules pour l'ajustement de la non-réponse

L'ajustement de non-réponse est effectuée dans des cellules h^* considérées comme homogènes du point de vue de la probabilité de réponse. On a donc $g_q = g_{h^*}$, si le ménage q est dans la cellule h^* .

Dans notre cas, les cellules sont basées sur la strate h du premier niveau d'échantillonnage de l'EC2000 (**stradap**). Un partage plus fin est effectué dans le cas des plus grandes **stradap**. Ce dernier fait intervenir le type de relevé (**art**) et, pour les plus grandes, la taille des PSUs.

Les **stradap** sont classifiées en trois types selon leur taille en terme de ménages répondants et faisant partie de la population. Les tailles pondérées $w_{c_1,h}$ et non pondérées $n_{c_1,h}$ sont utilisées pour définir les types ; voir Table 9. Les taux de réponse ne sont pas considérés dans le partage.

TAB. 9 – Définition des types de **stradap** et liste par type.

Type	Critère	stradap
grande	$w_{c_1,h} > 100'000$ ou $n_{c_1,h} > 2000$	1,2,9,15,16,17,18,19
petite	$n_{c_1,h} < 40$	14,20,102
moyenne	sinon	3,5,6,7,8,10,11,12,13,21,41,42, 101,103,104,105,106

Les cellules de correction de la non-réponse h^* sont définies par les combinaisons suivantes :

$$\text{cellules} = \begin{cases} (\text{stradap} \times \text{art} \times \text{psugrup}) & \text{si grande stradap} \\ (\text{stradap}) & \text{si petite stradap} \\ (\text{stradap} \times \text{art}) & \text{si moyenne stradap} \end{cases}$$

avec **psugrup** défini au niveau des PSUs comme suit :

$$\text{psugrup} = \begin{cases} 1 & \text{si le PSU a beaucoup de ménages} \\ 0 & \text{si le PSU a peu de ménages} \end{cases}$$

La limite entre PSUs avec peu de ménages et PSUs avec beaucoup de ménages dépend de la strate. Elle est déterminée après un tri des PSUs selon leur nombre de ménages et un partage en deux de façon à obtenir un nombre de ménages similaire dans les deux parties.

4.1.4 Calcul de l'ajustement

L'ajustement pour la non-réponse est effectuée en deux étapes dans chaque cellule h^* .

La première étape consiste à estimer, dans chaque cellule, la proportion p_{c_1,h^*} de ménages faisant partie de la population. Cette estimation est basée sur les ménages pour lesquels cette information est connue (c_1 , c_2 et c_3) :

$$\hat{p}_{c_1,h^*} = \frac{n_{c_1,h^*}}{\left(\sum_j n_{c_j,h^*}\right) - n_{c_4,h^*}} \quad (4)$$

avec n_{c_j, h^*} le nombre de ménages du type j se trouvant dans la cellule h^* , $j = 1, \dots, 4$.

La deuxième étape reprend la valeur estimée $\hat{p}_{c_1, h^*}$ et estime le total pondéré $\hat{p}_{c_1, h^*} w_{c_4, h^*}$ de ménages du groupe $c_4 = 1$ faisant partie de la population.

Le facteur de correction g_{h^*} est alors donné par :

$$g_{h^*} = \frac{\hat{p}_{c_1, h^*} w_{c_4, h^*} + w_{c_1, h^*}}{w_{c_1, h^*}} \quad (5)$$

Nous notons que la proportion $\hat{p}_{c_1, h^*}$ est basée sur les tailles non pondérées alors que g_{h^*} se base sur les valeurs pondérées. L'utilisation de la pondération dans les deux valeurs ou dans aucune de ces dernières serait également possible. Le choix ne semble pas avoir une grande influence sur les résultats.

Les cellules et les détails numériques de la correction de la non-réponse se trouvent dans la Table 10.

4.2 Poids finaux des ménages et personnes

Le poids w_p de la personne p du ménage q du P-sample se trouvant dans la cellule h^* , dans le bâtiment k du PSU i de la *stradap* h , est déterminé par :

$$w_p = \begin{cases} w_k f_{ih}^{(2)} g_{h^*} & \text{si la personne fait partie du P-sample} \\ 0 & \text{sinon} \end{cases} \quad (6)$$

avec w_k le poids d'échantillonnage du bâtiment k , $f_{ih}^{(2)}$ l'ajustement pour la non-réponse au niveau des PSUs et g_{h^*} la correction pour la non-réponse au niveau des ménages.

La distribution des poids w_p par *stradap*, voir Table 11, est similaire à celle du P-sample provisoire ; voir Table 2 de Renaud (2003).

Avec la nouvelle pondération des personnes et des ménages, le total estimé des ménages dans la population en Suisse d'intérêt $\sum_q w_q = 3'217'476$ est supérieur à la valeur correspondante du recensement (3'181'569 ménages). De même, le total estimé des personnes $\sum_p w_p = 7'318'915$ est supérieur à la population d'intérêt de 7'121'626 personnes dans le recensement. Ces valeurs sont cependant peu fiables car leur coefficient de variation est de l'ordre de 4.3% pour les ménages et 3.5% pour les personnes (calculs grossiers). Le P-sample n'est pas construit de façon à avoir une estimation fiable d'un total mais dans le but d'être utilisé dans l'estimation duale. Le total estimé sur la base le P-sample ne peut donc pas prouver l'existence d'une sous-couverture globale de la population.

TAB. 10 – Construction des cellules et des poids finaux des ménages. Voir la notation dans le texte. Nombre de cas $n = \sum_j n_{c_j}$ et somme des poids $w = \sum_q w_q$.

stradap	art	psugrup	n_{c_1}	w_{c_1}	n_{c_2} $+n_{c_3}$	w_{c_2} $+w_{c_3}$	n_{c_4}	w_{c_4}	n	$\bar{p}_{c_1, h^*}$ in %	g_{h^*}	w
1	0	0	79	1'522.7	72	1'866.8	8	179.4	159	52.3	1.0617	1'616.6
		1	59	1'079.8	7	128.1	6	109.8	72	89.4	1.0909	1'177.9
	1	0	367	6'940.8	20	392.4	102	2'006.1	489	94.8	1.2741	8'843.2
		1	358	6'551.9	15	274.5	61	1'116.4	434	96.0	1.1635	7'623.4
2	0	0	74	1'640.3	36	873.4	5	107.8	115	67.3	1.0443	1'712.9
		1	44	948.9	5	107.8	10	215.7	59	89.8	1.2040	1'142.5
	1	0	458	9'978.2	21	497.4	90	2'025.8	569	95.6	1.1941	11'915.2
		1	421	9'079.2	17	366.6	66	1'423.3	504	96.1	1.1507	10'447.3
3	0	.	79	1'844.4	34	793.8	20	466.9	133	69.9	1.1770	2'170.9
	1	.	446	10'412.8	16	373.6	90	2'101.2	552	96.5	1.1948	12'441.3
5	0	.	38	1'304.8	22	811.8	11	399.5	71	63.3	1.1939	1'557.8
	1	.	143	5'059.6	10	365.5	31	1'105.0	184	93.5	1.2041	6'092.4
6	0	.	9	309.4	6	206.2	1	34.4	16	60.0	1.0666	330.0
	1	.	131	4'502.8	8	275.0	23	790.6	162	94.2	1.1655	5'247.9
7	0	.	17	709.3	4	163.0	4	165.9	25	81.0	1.1893	843.6
	1	.	90	3'757.6	7	290.2	24	996.6	121	92.8	1.2461	4'682.3
8	0	.	25	734.9	13	382.1	12	352.7	50	65.8	1.3157	966.9
	1	.	84	2'469.2	11	323.3	17	499.7	112	88.4	1.1789	2'911.0
9	0	0	51	1'584.1	13	412.5	15	461.6	79	79.7	1.2322	1'951.9
		1	33	1'011.8	6	184.6	8	244.3	47	84.6	1.2043	1'218.5
	1	0	489	15'189.3	21	663.9	72	2'220.6	582	95.9	1.1402	17'318.4
		1	412	12'662.1	32	987.9	76	2'347.0	520	92.8	1.1720	14'840.0
10	0	.	19	756.3	7	278.6	4	159.2	30	73.1	1.1539	872.7
	1	.	89	3'542.7	3	119.4	12	477.7	104	96.7	1.1304	4'004.8
11	0	.	41	1'408.1	12	484.7	12	568.0	65	77.4	1.3121	1'847.5
	1	.	413	13'581.0	31	1'063.4	62	2'097.8	506	93.0	1.1437	15'532.3
12	0	.	18	562.7	3	93.8	1	31.3	22	85.7	1.0476	589.5
	1	.	101	3'157.2	5	156.3	23	719.0	129	95.3	1.2170	3'842.2
13	0	.	12	613.4	13	664.5	1	51.1	26	48.0	1.0399	637.9
	1	.	102	5'213.7	4	204.5	27	1'380.1	133	96.2	1.2547	6'541.8
14	.	.	29	538.1	6	110.0	21	388.7	56	82.9	1.5984	860.1
15	0	0	144	22'942.6	57	9'030.9	27	4'121.9	228	71.6	1.1287	25'895.6
		1	173	27'380.9	16	2'470.9	35	5'119.1	224	91.5	1.1711	32'066.6
	1	0	2'068	327'778.8	123	19'533.9	391	62'293.7	2'582	94.4	1.1794	386'575.4
		1	1'988	311'759.7	87	13'584.2	312	48'660.3	2'387	95.8	1.1495	358'379.7
16	0	0	138	22'031.6	28	4'466.0	53	8'471.7	219	83.1	1.3197	29'074.3
		1	199	31'756.7	21	3'349.9	81	12'921.3	301	90.5	1.3680	43'444.6
	1	0	1'959	312'684.3	147	23'455.8	360	57'467.0	2'466	93.0	1.1710	366'140.0
		1	1'621	258'575.8	112	17'865.6	339	54'081.7	2'072	93.5	1.1956	309'162.3
17	0	0	89	15'017.3	30	5'045.6	18	2'986.1	137	74.8	1.1487	17'250.6
		1	77	12'852.6	16	2'701.8	35	5'924.3	128	82.8	1.3816	17'757.7
	1	0	1'332	227'573.2	78	12'888.6	196	34'189.0	1'606	94.5	1.1419	259'870.9
		1	1'283	216'041.2	74	12'419.3	229	38'795.2	1'586	94.5	1.1698	252'720.8
18	0	0	141	23'555.1	38	6'365.2	61	10'198.0	240	78.8	1.3410	31'588.2
		1	169	28'257.4	20	3'345.2	107	17'906.2	296	89.4	1.5666	44'268.8
	1	0	1'594	266'543.0	165	27'618.5	328	54'898.8	2'087	90.6	1.1866	316'292.2
		1	1'320	220'856.5	81	13'547.7	270	45'166.1	1'671	94.2	1.1927	263'411.3
19	0	0	43	11'849.0	7	1'928.9	22	6'062.3	72	86.0	1.4400	17'062.6
		1	17	5'132.3	12	3'720.0	4	1'102.2	33	58.6	1.1259	5'778.4
	1	0	203	55'938.3	13	3'582.3	48	13'226.8	264	94.0	1.2222	68'369.1
		1	115	32'515.9	15	4'581.2	69	21'252.4	199	88.5	1.5782	51'316.1
20	.	.	24	3'421.5	3	404.5	11	1'483.3	38	88.9	1.3854	4'740.0
21	0	.	12	857.5	4	285.9	8	571.7	24	75.0	1.4999	1'286.2
	1	.	299	21'365.0	24	1'715.0	47	3'358.5	370	92.6	1.1455	24'474.0
41	0	.	60	1'078.9	30	675.8	21	370.6	111	66.7	1.2289	1'325.9
	1	.	630	11'228.6	31	572.9	99	1'806.4	760	95.3	1.1533	12'950.3
42	0	.	11	609.9	.	.	2	110.9	13	100.0	1.1818	720.8
	1	.	160	8'871.1	16	887.1	23	1'275.2	199	90.9	1.1307	10'030.4
101	0	.	19	1'712.4	18	2'222.0	5	405.1	42	51.4	1.1215	1'920.5
	1	.	71	5'901.9	7	588.2	33	2'683.8	111	91.0	1.4139	8'344.9
102	.	.	35	3'832.7	14	2'024.7	20	2'223.2	69	71.4	1.4143	5'420.7
103	0	.	4	382.2	10	1'288.9	.	.	14	28.6	1.0000	382.2
	1	.	46	4'290.0	4	374.0	18	1'594.6	68	92.0	1.3420	5'757.0
104	0	.	39	7'039.6	10	1'843.2	8	1'408.7	57	79.6	1.1593	8'160.8
	1	.	203	37'097.4	27	4'780.6	50	8'975.2	280	88.3	1.2135	45'019.0
105	0	.	13	3'367.6	4	1'100.0	2	444.0	19	76.5	1.1008	3'707.1
	1	.	56	14'308.0	1	328.0	7	1'794.8	64	98.2	1.1232	16'071.3
106	0	.	58	4'673.8	5	412.2	7	562.4	70	92.1	1.1108	5'191.6
	1	.	140	11'394.9	4	324.9	30	2'440.6	174	97.2	1.2082	13'767.7
Total			21'284	2'701'142.3	1'832	225'245.1	4'291	561'596.1	27'407	92.1	1.1912	3'217'476.2

TAB. 11 – Statistique des poids finaux w_p des personnes du P-sample. Strate **stradap** 1-42 pour le plan NORD et 101-106 pour le plan TICINO. Nombre de cas n , somme sum , moyenne $mean$, coefficient de variation cv , minimum min , médiane P_{50} et maximum max .

stradap	n	sum	$mean$	cv	min	P_{50}	max
1	2'209	49'303.6	22.3	9.02	19.4	21.3	36.9
2	2'599	65'792.1	25.3	4.58	22.5	25.8	30.6
3	1'412	39'312.8	27.8	0.50	27.5	27.9	27.9
5	471	19'906.5	42.3	7.09	40.6	40.9	48.6
6	407	16'212.9	39.8	2.12	36.7	40.1	40.1
7	275	14'229.1	51.7	2.14	49.2	51.8	53.2
8	283	10'068.9	35.6	4.76	34.7	34.7	38.7
9	2'404	86'251.7	35.9	7.18	5.2	36.1	46.4
10	296	13'361.2	45.1	0.74	45.0	45.0	45.9
11	1'162	44'189.2	38.0	8.34	35.8	37.2	80.3
12	330	12'384.4	37.5	4.18	32.8	38.0	38.0
13	300	18'955.1	63.2	4.90	53.2	64.1	64.1
14	65	1'931.2	29.7	2.22	29.3	29.3	30.8
15	10'733	1'975'071.2	184.0	12.18	89.6	185.5	275.8
16	7'893	1'502'512.6	190.4	3.81	183.0	187.6	218.7
17	6'881	1'350'279.4	196.2	14.81	51.4	195.4	372.0
18	6'648	1'346'055.0	202.5	6.53	194.5	199.3	263.9
19	945	354'863.0	375.5	13.33	310.3	336.8	489.2
20	61	12'004.7	196.8	7.61	186.8	186.8	218.9
21	727	60'013.4	82.6	5.02	81.9	81.9	107.2
41	1'823	37'761.5	20.7	5.73	18.3	20.5	26.3
42	419	26'332.3	62.9	1.03	62.7	62.7	65.5
101	211	24'739.4	117.3	28.92	76.8	100.8	213.7
102	96	14'956.2	155.8	18.46	145.0	149.1	312.8
103	119	14'550.7	122.3	16.93	83.0	116.9	167.7
104	551	121'963.5	221.4	7.16	200.8	210.2	247.5
105	164	47'174.2	287.7	15.92	244.4	281.7	368.5
106	402	38'739.8	96.4	4.75	88.0	95.7	100.6
Total	49'886	7'318'915.5	146.7	56.03	5.2	186.6	489.2

4.3 Autres approches pour la correction de la non-réponse

Différentes corrections de la non-réponse ont été envisagées avant de choisir la méthode décrite dans la Section 4.1.

Le traitement des ménages sans information sur l'appartenance ou pas à la population (`innerh=.`) a été testé selon deux approches : (1) imputation d'une valeur 0-1, ou (2) utilisation d'une probabilité d'appartenance à la population, sans fixer la valeur à 0 ou 1. Les deux approches peuvent être appliquées dans des cellules bien choisies (ex. `stradap×art` ou `h*` décrites plus haut). Dans les deux cas, on estime une probabilité d'appartenance à la population. La première approche fixe une valeur 0 ou 1, par exemple selon processus de Poisson, et la deuxième utilise uniquement la probabilité (ex. $\hat{p}_{c_1, h^*}$ décrite plus haut).

On notera qu'il est possible de traiter l'appartenance à la population et la non-réponse selon des cellules identiques ou différentes. Les cellules différentes offrent plus de flexibilité car elles permettent de travailler avec des modèles différents pour les questions d'appartenance à la population et de non-réponse. Cependant, ce partage a un sens uniquement si on a des informations sur les processus en jeu.

Dans le cadre de l'EC2000, il n'y a pas d'information auxiliaire sur les processus de l'appartenance à la population et de la non-réponse. Par conséquent, la version choisie suit l'approche de la probabilité pour l'appartenance à la population, avec des cellules identiques pour les deux composantes ; ce qui simplifie la procédure.

L'utilisation des nombres absolus $n_{c_1, h}$ et $n_{c_4, h}$ au lieu des nombres pondérés $w_{c_1, h}$ et $w_{c_4, h}$ ont également été testés dans le calcul de la correction ; tout comme différentes définitions de cellules. On n'observe cependant pas de grandes différences entre les résultats des variantes testées, mais pas retenues, et les résultats de la méthode retenue.

5 Dernières modifications du P-sample

Le P-sample de 49'886 personnes comporte 3 éléments hors population. Ces personnes, qui ont leur domicile à l'étranger au jour du recensement, ont été détectées lors des analyses des résultats des appariements.

Dans les trois cas, la personne avait déménagé entre le jour du recensement et l'enquête de couverture, et le numéro postal indiqué au jour du recensement était manquant. Les informations sur le lieux au jour du recensement (France, resp. Allemagne) permettent d'exclure ces éléments.

Le P-sample final contient donc 49'883 personnes.

Cette modification ne nécessite pas d'adaptation de la pondération décrite dans la Section 4.

6 Qualité du cadre de sondage des bâtiments

Des imperfections dans le cadre de sondage des bâtiments servant de base à l'EC2000 et au P-sample, notamment une sous-couverture, peuvent induire un biais dans les estimations de la couverture de la population.

Le cadre de sondage de l'EC2000 est la liste des bâtiments du Registre des adresses de bâtiments (REAB) pour le plan NORD et celle de la Dichiarazione per la stima ufficiale della sostanza immobiliare (Stima) pour le plan du Tessin. Le REAB est dans l'état de novembre 2000 et la Stima est dans celui reçu de l'Ustat en décembre 2000 ; voir Renaud et Eichenberger (2002). L'identificateur des bâtiments est le `Ref_code` pour le NORD et un `Ref_code` fictif pour le Tessin ; voir Section 3.1.

Deux autres listes de bâtiments sont actuellement disponibles : la liste du Recensement fédéral des bâtiments et des logements de l'an 2000 (RBL2000) et le Registre des bâtiments et logements (RegBL).

Le RBL2000 comporte les bâtiments habités ou habitables dans l'état du 5 décembre 2000. Il contient également des bâtiments fictifs (au maximum 1 par commune) qui regroupent les ménages et les personnes qui n'ont pu être assignés aux bâtiments recensés. L'identificateur est le `Ref_code` correspondant au REAB du NORD.

Le RegBL contient les bâtiments habités, habitables ou non habitables (ex. entreprises, locaux administratifs). Il se base sur le RBL2000, avec l'ajout de données telles que celles du recensement des entreprises. Le jeu de données dans l'état d'août 2003 a été fourni pour les analyses décrites dans cette section. Il est donc difficilement comparable à un état au 5 décembre 2000. Les années de construction ou de destruction donnent cependant des informations partielles sur les changements de la liste. L'identificateur est l'EGID ; équivalent du `Ref_code` du REAB et du RBL2000.

Dans cette section, la comparaison entre le REAB et le RBL2000 donne une information sur la couverture de la population d'intérêt par le cadre de sondage de l'EC2000. Le RegBL est utilisé pour analyser les différences entre le REAB et le RBL2000. Une analyse au niveau des bâtiments de l'EC2000, avec information sur les habitants, permet d'aborder les problèmes de localisation et de couverture.

Seule la partie du NORD est considérée dans nos comparaisons. Il n'y a en effet pas d'identificateurs communs entre la Stima et les deux autres listes, et aucun appariement ne peut être fait de façon satisfaisante avec les variables communes à disposition. Le résumé des tailles et de l'état des données est donné dans la Table 12.

TAB. 12 – Taille et état des données du REAB, du RBL2000 et du RegBL, avec le total des bâtiments N_{tot} et le total sans le Tessin et le Liechtenstein N .

Liste	N_{tot}	N	État
REAB	1'609'909	1'609'909	novembre 2000
RBL2000	1'474'807	1'374'572	5 décembre 2000
RegBL	1'569'647	1'469'630	août 2003

6.1 Première allure

Un ensemble de 1'366'231 bâtiments, sur le total de 1'626'995 bâtiments existants dans tout ou partie des trois listes, se trouve dans les trois listes (84%); voir Table 13. Notons que les bâtiments fictifs, identifiés par `fictif=1`, existent uniquement dans les données du REAB et du RBL2000. Ils ne sont plus considérés dans la suite de l'analyse.

TAB. 13 – Nombre de bâtiments N dans les trois listes (sans le Tessin et le FL), avec '1' si dans la liste et '0' sinon. Les bâtiments fictifs sont identifiés par `fictif=1`.

Type	RegBL	RBL2000	REAB	fictif	N
1	0	0	1	0	153'506
2	0	0	1	1	125
3	0	1	0	0	2
4	0	1	0	1	2
5	0	1	1	0	1'142
6	0	1	1	1	2'588
7	1	0	0	0	12'475
8	1	0	1	0	86'317
9	1	1	0	0	4'607
10	1	1	1	0	1'366'231
Total					1'626'995

Les bâtiments réels se trouvant dans le RBL2000 et pas dans le RegBL sont des bâtiments détruits en 2001 ou plus tard (types 3 et 5). Ceux dans le RegBL et pas dans les autres listes (type 7) sont des bâtiments non habitables ou construits dès 2001 (au moins 645 cas).

6.2 Comparaison entre le REAB et le RBL2000

Trois groupes de bâtiments sont analysés pour déterminer la qualité de la couverture du REAB comme base de sondages des bâtiments de l'EC2000 : (1) bâtiments dans le RBL2000 mais pas dans le REAB (sous-couverture du REAB), (2) bâtiments dans le REAB mais pas dans le RBL2000 (sur-couverture du REAB) et (3) bâtiments dans le REAB et le RBL2000 (couverture commune).

6.2.1 Bâtiments dans le RBL2000 mais pas dans le REAB

La sous-couverture du REAB est de premier intérêt dans l'analyse de la qualité de cette base de sondage. Une personne habitant un bâtiment du RBL2000 ne se trouvant pas dans le REAB n'avait en effet en principe aucune chance d'être atteinte dans l'EC2000.

Parmi les 4'609 bâtiments dans le RBL2000 mais pas le REAB, *i.e.* 0.3% du REAB (types 3 et 9 de la Table 13), on observe que :

- 45% des bâtiments ont été construit entre 1996 et 2000 alors que globalement seuls 6.5% des bâtiments du RBL2000 sont dans ce cas ;

- comme dans la liste globale, env. 81% des bâtiments sont uniquement à usage d’habitation ($G_Gart = 2$, voir Table 15);
- les bâtiments contiennent 8’579 logements (variable G_Azwet), dont 95% ont moins de 6 logements, alors que 95% des bâtiments ont moins de 9 logements dans le RBL2000 ;
- les bâtiments contiennent 18’809 personnes à leur domicile économique (variable G_Apwg). On note que 75% de ces bâtiments ont 0 ou 1 personne (entre 0 et 5 personnes dans le RBL2000) et 95% ont entre 0 et 13 personnes (entre 0 et 17 dans le RBL2000).

En résumé, les 4’609 bâtiments sont principalement récents, avec peu de logements, et habités par peu de personnes. Seule une faible proportion de 0.26% de la population résidante de 7’280’010 personnes n’avait pas la possibilité d’être atteint par l’EC2000. On verra de plus qu’une partie de ces personnes peut tout de même être atteinte dans le cas d’erreurs de localisation ; voir la Section 6.3.

6.2.2 Bâtiments dans le REAB mais pas dans le RBL2000

La sur-couverture du REAB est également un élément intéressant dans l’analyse de la qualité du cadre de sondage. Elle peut en effet être liée à une sous-couverture des bâtiments du recensement.

Sur les 239’823 bâtiments dans le REAB mais pas le RBL2000, *i.e.* 15% du REAB (types 1 et 8 de la Table 13), 23’736 sont notés comme habités (env. 10%) et seuls 86’317 (36%) sont retrouvés dans le RegBL ; voir Table 14. Il est donc possible que quelques bâtiments habités au jour du recensement aient disparu du RBL2000. De plus, la liste des bâtiments a été beaucoup modifiée entre le REAB et le RegBL.

TAB. 14 – Nombre de bâtiments des types 1 et 8, par **Objet** (REAB) et avec information sur la présence ou pas dans le RegBL.

Objet	Description	RegBL		Total
		0	1	
1	avec utilisation à des fins d’habitation	10’586	13’150	23’736
2	sans utilisation à des fins d’habitation	118’841	52’462	171’303
4	habitation mobile/provisoire	456	164	620
8	inconnu	16’488	12’761	29’249
9	démoli	7’135	7’780	14’915
Total		153’506	86’317	239’823

Nous n’avons pas d’information sur le nombre d’habitants dans les bâtiments du REAB. Seules les valeurs relevées pour un échantillon de bâtiments durant l’EC2000 peuvent servir dans une comparaison REAB-RBL2000 au niveau de la population ; voir Section 6.3.

6.2.3 Bâtiments dans le REAB et le RBL2000

Dans le cas de bâtiments présents dans le REAB et le RBL2000, *i.e.* 85% du REAB (types 5 et 10 de la Table 13), il est intéressant de comparer les caractéristiques des bâtiments. Si le bâtiment se trouve également dans le RegBL, les caractéristiques peuvent être comparées entre les 3 listes.

Les caractéristiques principales sont classifiées selon différentes variables. Pour le REAB, on prendra la variable **Objet** décrite dans la Table 14. Pour le RBL2000 et le RegBL, on prendra respectivement les variables **G_Gart** et **GKAT** décrites dans la Table 15.

TAB. 15 – Définition de la catégorie du bâtiment dans le RBL2000 (**G_Gart**) et le RegBL (**GKAT**).

G_Gart	Description
1	affecté principalement à d'autres usages que l'habitation
2	exclusivement à usage d'habitation
3	principalement à usage d'habitation
4	habitation de fortune mobile ou provisoire
5	fictif
GKAT	Description
1010	habitation de fortune
1021	maison individuelle
1025	maison à plusieurs logements
1030	autre bât. d'habitation
1040	accessoirement à l'habitation
1060	sans usage d'habitation
1080	construction particulière

La Table 16 présente le nombre de bâtiments dans les différentes combinaisons de **Objet**, **G_Gart** et **GKAT**. Dans 98.9% des cas, les catégories des trois listes sont consistantes (voir les valeurs en gras dans la table). Certaines inconsistances sont cependant observées. C'est le cas par exemple de bâtiments considérés comme démolis dans le REAB (**Objet**=9) mais qui deviennent habitables dans le RBL2000 et le RegBL.

TAB. 16 – Nombre de bâtiments dans le REAB et le RBL2000 (types 5 et 10), selon les catégories **Objet**, **G_Gart** et **GKAT**. Les bâtiments avec **GKAT='.'** sont des bâtiments absents du RegBL. Les valeurs en gras sont consistantes ; voir le texte.

Objet	GKAT	G_Gart				Total
		1	2	3	4	
1	.	538	5'285	1'300	1	7'124
	1010	.	.	.	738	738
	1021	352	749'809	622	1	750'784
	1025	321	332'316	989	.	333'626
	1030	336	1'994	191'613	1	193'944
	1040	78'909	865	364	2	80'140
	1060	59	9	5	1	74
	1080	12	.	.	.	12
2	.	3	21	4	.	28
	1010	.	.	.	4	4
	1021	3	75	2	.	80
	1025	.	23	1	.	24
	1030	1	11	25	.	37
	1040	174	8	.	.	182
	1060	3	3	.	.	6
	1080	.	4	.	.	4
4	1010	7	65	15	152	239
	1021	.	113	.	.	113
	1025	.	7	.	.	7
	1030	.	.	20	.	20
	1040	82	.	.	.	82
	1060	1	5	.	.	6
8	1021	.	20	1	.	21
	1025	2	15	1	.	18
	1030	.	2	6	.	8
	1040	16	2	.	.	18
	1060	.	4	.	.	4
9	1021	2	7	1	.	10
	1025	1	7	1	.	9
	1030	.	3	1	.	4
	1040	3	3	1	.	7
	1060	.	4	.	.	4
Total		80'825	1'090'676	194'972	900	1'367'373

6.3 Analyse au niveau des bâtiments de l'EC2000

Une analyse au niveau des 14'697 bâtiments du plan NORD fait ressortir un ensemble de 438 bâtiments dans lesquels on a trouvé des personnes lors de l'EC2000 mais qui ne se trouvent pas dans les données du RBL2000 ; voir Table 17.

TAB. 17 – Information sur les bâtiments du plan NORD de l'EC2000.

Bâtiments dans RBL2000	avec personnes du P-sample		Total
	non	oui	
non	1'754	438	2'192
oui	2'325	10'180	12'505
Total	4'079	10'618	14'697

Parmi les 438 bâtiments habités du P-sample et pas retrouvés dans le RBL2000, 205 sont trouvés dans le RegBL. Ces 205 bâtiments ne sont pas récents (202 construits avant 1996 et 3 entre 1996 et 2000). Ils se répartissent en 137 (67%) habitables ($GKAT \in \{1010, 1021, 1025, 1030, 1040\}$), 45 (22%) non habitables ($GKAT=1060$) et 23 (11%) avec statut peu clair ($GKAT$ manquant).

Les 438 bâtiments comportent un ensemble de 1'828 personnes dans 721 ménages (dans le P-sample). Seules 45 (2.5%) n'ont pas été appariées à un enregistrement du recensement ($MatchG=20$). Les 1'783 restantes ($MatchG=10$) ont donc été assignées à des bâtiments différents dans le P-sample et dans le recensement.

Les erreurs de localisation sont possibles dans le cadre du recensement et de l'EC2000, spécialement si les adresses des bâtiments ne sont pas claires. On peut espérer que le travail effectué par les facteurs sur le terrain donne des résultats plus fiables que ceux effectués dans le cadre du recensement mais une étude complémentaire serait nécessaire pour aboutir à un résultat fiable à ce sujet.

Si on se restreint aux personnes n'ayant annoncé aucun déménagement entre le jour du recensement (5 décembre 2000) et le jour de l'EC2000 (avril-mai 2001), *i.e.* $mover_code=1$, le nombre de personnes du P-sample dans des bâtiments hors RBL2000, resp. dans le RBL2000, passe à 1'784, resp. 45'007 ; voir Table 18. Un ensemble de 551 personnes parmi celles dans des bâtiments du RBL2000 n'ont pas été appariées (1.2%). 1'739 personnes sur les 1'784 hors bâtiments du RBL2000 sont appariées et sont donc des cas d'erreur dans la localisation.

Parmi les bâtiments hors RBL2000 et qui contiennent des personnes qui n'ont pas déménagé, on note que 110 personnes ont été appariées avec des personnes dans 25 bâtiments fictifs ; voir Table 19. Il est donc probable que tout ou partie de ces bâtiments notés comme habités dans l'EC2000 et pas retrouvés dans le RBL2000 soient en fait des bâtiments manquants dans le RBL2000.

Deux personnes sont appariées avec des personnes dans un bâtiment absent du RBL2000 ($G_Gart = '.'$). Il s'agit probablement d'appariements non valables car certains enregistrements dans le recensement ont été retirés après la procédure d'appariement (75 cas).

TAB. 18 – Nombre de personnes n absolu et n_{pond} pondéré selon le statut d'appariement (MatchG) et l'appartenance ou pas du bâtiment au RBL2000 (sans le Tessin, seulement mover_code=1).

	Bât. pas dans RBL2000		Bât. dans RBL2000		Total	
	n	n_{pond}	n	n_{pond}	n	n_{pond}
MatchG=10	1'739	259'990.8	44'456	6'466'770.9	46'195	6'726'761.7
MatchG=20	45	7'244.3	551	85'253.1	596	92'497.4
Total	1'784	267'235.1	45'007	6'552'024.0	46'791	6'819'259.1

TAB. 19 – Nombre de personnes n et de bâtiments m , par catégories G_Gart. Restriction aux personnes dans des bâtiments hors RBL2000 et dans lesquels on a des personnes n'ayant pas déménagé du P-sample.

G_Gart	Description	n	m
.		2	1
1	princ. autres usages que habit.	77	30
2	excl. habitation	1'061	240
3	princ. habitation	464	128
4	habit. fortune/mobile	25	2
5	fictif	110	25
Total		1'739	426

6.4 Couverture des bâtiments

La comparaison des listes de bâtiments permet d'affirmer que la couverture de la base de sondage de l'EC2000 est bonne. Les biais liés à la très faible sous-couverture sont donc négligeables.

Il est cependant difficile de donner un ordre de grandeur sur la sous-couverture au niveau des personnes. Les erreurs de localisation et la sous-couverture sont en effet difficilement séparables.

Quelques bâtiments habités selon l'EC2000 ne se trouvent pas pas dans le recensement. Il est cependant possible que la comparaison soit entachées d'incertitudes car certains identificateurs ont été modifiés dans le cadre des travaux du recensement. Ces modifications, d'une ampleur inconnue, ont par exemple eu lieu lors de partages ou regroupements de bâtiments (ex. un bâtiment dans le REAB transformé en 2 bâtiments dans le RFB2000, *i.e.* nouveau EGID pour même bâtiment selon le REAB).

L'analyse présentée dans cette section est assez grossière. Elle montre cependant qu'il y a un potentiel pour une analyse plus détaillée ; également pour le Tessin.

7 Conclusion

Le P-sample final, échantillon indépendant du recensement, comporte 49'886 personnes. Il est construit sur la base des résultats de l'enquête de couverture (EC2000), avec l'intégration de diverses informations complémentaires recueillies après l'enquête, et satisfait les critères de tailles par sous-groupes.

La pondération du P-sample est basée sur le plan d'échantillonnage de l'EC2000 et intègre une correction pour la non-réponse. Elle tient compte également de l'incertitude sur l'appartenance ou pas des non-répondants à la population. Cette incertitude est un point important dans le cadre de l'EC2000 et diverses méthodes ont donc été testées avant de choisir une méthode adéquate. La variabilité des poids finaux est relativement élevée.

L'analyse de la couverture du cadre de sondage des bâtiments utilisé pour l'EC2000 montre que le risque de biais, dû à une sous-couverture des bâtiments, est faible pour les estimations de la couverture de la population. Certains problèmes de localisation sont apparus car une partie des personnes sont assignées à un bâtiment alors qu'ils habitent un autre bâtiment. La provenance du problème est cependant difficile à détecter (EC2000 ou recensement ?). Les résultats montrent également qu'un petit nombre de bâtiments parmi ceux tirés pour l'EC2000 et dans lesquels on a trouvé des personnes semblent manquer dans le recensement.

Annexe : Programmes SAS

Les travaux décrits dans le présent rapport ont été effectués dans SAS v8.2 sur les réseaux PC et Unix de l'OFS.

Les programmes SAS sont les suivants :

- construction de la liste des bâtiments (et intégration des cas particuliers au niveau ménages) :
... \ECresults\ProgSAS\listebat2.sas (listebat.out)
- reprise des informations sur les matches :
... \Estimations\sysVZ\repriseMatch\base.sas
- intégration des corrections sur les codes de match pour le P-sample :
... \Estimations\sysVZ\repriseMatch\adapt.sas
- data set final du P-sample :
... \Estimations\sysVZ\repriseMatch\creaPsampFin.sas
- pondération du P-sample :
... \NeuesProgMai03Var3.sas
- qualité du cadre d'échantillonnage des bâtiments :
... \GWR-tests.sas

References

- Renaud, A. (2002). Methodology of the Swiss Census 2000 Coverage Survey. In *ASA, 2001 Proceedings of the Section on Survey Research*. American Statistical Association.
- 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). Rapport de méthodes 338-0019, OFS.
- Renaud, A. et P. Eichenberger (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. Rapport de méthodes 338-0009, OFS.

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

- Renaud, A., Potterat, J. (2004). Estimation de la couverture du recensement de la population de l'an 2000. Echantillon pour l'estimation de la sous-couverture (P-sample) et qualité du cadre de sondage des bâtiments. Numéro de commande: 338-0023
- Graf, M. (2004). Fusion de données. Etude de faisabilité. Numéro de commande: 338-0022
- Potterat, J. (2003). Mietpreis-Strukturerhebung 2003. Entwicklung des Stichprobenplans und Ziehung der Stichprobe. Bestellnummer: 338-0021
- Potterat, J. (2003). Landwirtschaftliche Betriebszählung 2003. Stichprobenplan der Zusatzerhebung. Bestellnummer: 338-0020
- Renaud, A. (2003). Estimation de la couverture du recensement de la population de l'an 2000. Echantillon pour l'estimation de la sur-couverture (E-sample). Numéro de commande: 338-0019
- Hulliger, B. (2003). Erhebung über Forschung und Entwicklung in der schweizerischen Privatwirtschaft 2000. Bereinigung der Stichprobe, Ersatz fehlender Werte und Schätzverfahren. Bestellnummer: 338-0018
- Renfer, J.-P. (2003). Enquête 2000 sur la recherche et le développement dans l'économie privée en Suisse. Plan d'échantillonnage. Numéro de commande: 338-0017
- Potterat, J. (2003). Kosten und Nutzen der Berufsbildung aus Sicht der Betriebe. Schätzverfahren. Bestellnummer: 338-0016
- Graf, M., Matei, A. (2003). Stratégie de choix des modèles de désaisonnalisation. Application aux séries de l'emploi total. Numéro de commande : 338-0015
- Potterat, J., Salamin, P.A. (2002). Betriebszählung 2001. Methoden für die Datenbereinigung. Bestellnummer: 338-0014
- Renaud, A. (2002). Programme international pour le suivi des acquis des élèves (PISA). Plans d'échantillonnage pour PISA 2000 en Suisse. Numéro de commande: 338-0013
- Renfer, J.-P. (2002). Enquête 2001 sur les coûts et l'utilité de la formation des apprentis du point de vue des établissements. Plan d'échantillonnage. Numéro de commande: 338-0012
- Potterat, J., Salamin, P.A. (2002). Betriebszählung 2001. Stichprobenplan und Schätzverfahren für die provisorischen Ergebnisse. Bestellnummer: 338-0011
- Graf, M. (2002). Enquête suisse sur la structure des salaires 2000. Plan d'échantillonnage, pondération et méthode d'estimation pour le secteur privé. Numéro de commande: 338-0010
- Renaud, A., Eichenberger P. (2002). Estimation de la couverture du recensement de la population de l'an 2000. Procédure d'enquête et plan d'échantillonnage de l'enquête de couverture. Numéro de commande: 338-0009
- Kilchmann, D., Hulliger, B. (2002). Stichprobenplan für die Obstbaumzählung 2001. Bestellnummer: 338-0008
- Graf, M. (2002). Passage du concept établissement au concept entreprise. Numéro de commande: 338-0007

- Salamin, P.A. (2001). La technique de la double enquête pour la statistique du transport routier de marchandise. Numéro de commande: 338-0006
- Peters, R., Renfer, J.-P. et Hulliger, B. (2001). Statistique de la valeur ajoutée 1997-1998. Procédure d'extrapolation des données. Numéro de commande: 338-0005
- Potterat, J., Hulliger, B. (2001). Schätzung der Sägereiproduktion mit der Sägerei-Erhebung PAUL. Bestellnummer: 338-0004
- Graf, M. (2001). Désaisonnalisation. Aspects méthodologiques et application à la statistique de l'emploi. Numéro de commande: 338-0003
- Hüsler, J., Müller, S. (2001). Schlussbericht Betriebszählung 1995 (BZ 95), Mehrfach imputierte Umsatzzahlen. Bestellnummer: 338-0002
- Renaud, A. (2001). Statistique suisse des bénéficiaires de l'aide sociale. Plan d'échantillonnage des communes. Numéro de commande: 338-0001
- Hulliger, B., Eichenberger, P. (2000). Stichprobenregister für Haushalterhebungen: Umstellung auf Telefonnummern ohne Namen und Adressen, Abläufe für Erstellung und Stichprobenziehung. Bestellnummer: 338-0000
- de Rossi, F.-X. (1998). Méthodes statistiques pour le compte routier suisse.
- Hulliger, B., Kassab, M. (1998). Evaluation of Estimation Methods for the Survey on Environment Protection Expenditures of Swiss Communes.
- Salamin, P.A. (1998). Etablissement d'une clef de passage pondérée entre l'ancienne (NGAE 85) et la nouvelle nomenclature (NOGA 95) générale des activités économiques.
- Peters, R. (1998). Extrapolation des données de l'enquête de structure sur les loyers.
- Bender, A., Hulliger, B. (1997). Enquête suisse sur la population active: rapport de pondération pour 1996.
- Salamin, P.A. (1997). Evaluation de la Statistique de l'emploi.
- Peters, R. (1997). Etablissement du plan d'échantillonnage pour l'enquête 1996 sur la recherche et le développement dans l'économie privée en Suisse.
- Peters, R. (1997). Enquête 1996 sur la structure des salaires en Suisse: établissement du plan d'échantillonnage.
- Peters, R. (1996). Pondération des données de l'enquête sur la famille en Suisse.
- Comment, T., Hulliger, B., Ries, A. (1996). Gewichtungsverfahren für die Schweizerische Arbeitskräfteerhebung (1991-1995).
- Hulliger, B. (1996). Haushalterhebung Familie 1994: Stichprobenplan, Stichprobenziehung und Reservestichproben.
- Peters, R., Hulliger, B. (1996). Schätzverfahren für die Lohnstruktur-Erhebung 1994 / Procédure d'estimation pour l'enquête de 1994 sur la structure des salaires.
- Peters, R. (1996). Schéma de pondération des indices PAUL.

- Hulliger, B., Peters, R. (1996). Enquête sur le comportement de la population suisse en matière de transport en 1994: plan d'échantillonnage et pondération.
- Hulliger, B. (1996). Gütertransportstatistik 1993: Schätzverfahren mit Kompensation der Antwortausfälle.
- Salamin, P.A. (1995). Estimation des flux pour le module II des comptes globaux du marché de travail.
- Peters, R. (1995). Enquête de structure sur les loyers: établissement d'un plan d'échantillonnage stratifié.
- Hulliger, B. (1995). Konjunktuelle 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:

<i>Diffusionsmittel</i>	<i>Kontakt N° à composer</i>	<i>Moyen de diffusion</i>
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 [>>News>>Neuerscheinungen](http://www.statistik.admin.ch).

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 [>>Actualités>>Nouvelles publications](http://www.statistique.admin.ch).

Methodenberichte des Dienstes Statistische Methoden Rapports de méthodes du Service de méthodes statistiques Methodological reports of the Statistical Methods Unit

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

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

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

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

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

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

Les défauts de couverture du recensement de la population de l'an 2000 (RFP2000) sont estimés selon le système dual, c'est-à-dire sur la base de deux échantillons de personnes. Ces deux échantillons permettent respectivement d'estimer les composantes de sous-couverture et de sur-couverture.

Le P-sample de 49'883 personnes est l'échantillon permettant d'estimer la composante de sous-couverture du recensement. Il découle d'un plan d'échantillonnage à plusieurs niveaux. La pondération du P-sample prend en compte l'information sur le plan d'échantillonnage, la non-réponse des ménages et l'incertitude sur l'appartenance ou pas à la population d'intérêt.

La qualité de la base de sondage des bâtiments utilisée pour la construction du P-sample est également analysée afin d'avoir une information sur le biais potentiel des estimations de couverture.