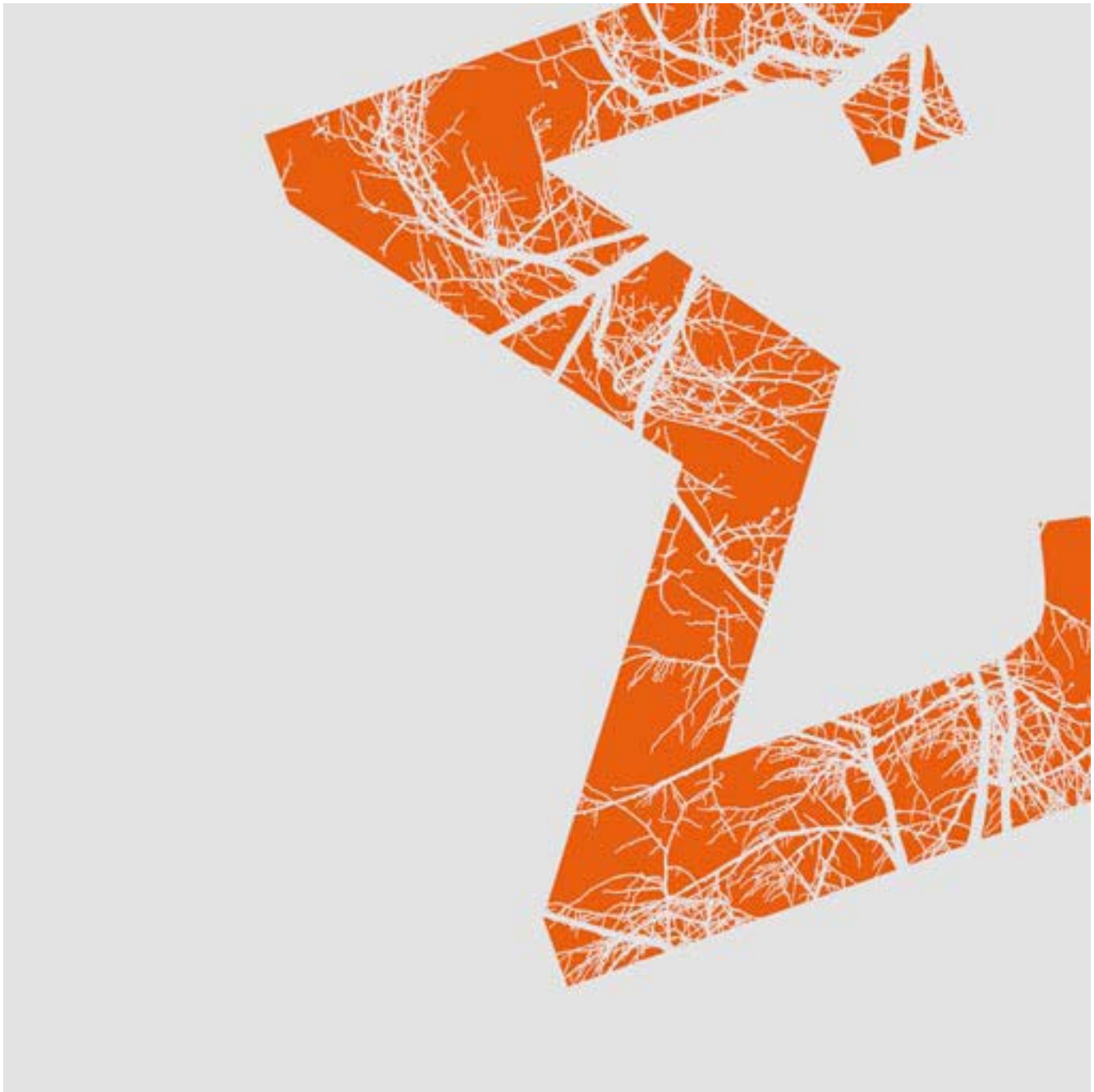
Rapport de méthodes

Enquête sur les chiffres d'affaires du commerce de détail

Plan d'échantillonnage et méthodes d'estimation



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

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

Neuchâtel, 2006

Statistik der Schweiz

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

0 Statistische Grundlagen und Übersichten
1 Bevölkerung
2 Raum und Umwelt
3 Arbeit und Erwerb
4 Volkswirtschaft
5 Preise
6 Industrie und Dienstleistungen
7 Land- und Forstwirtschaft
8 Energie
9 Bau- und Wohnungswesen
10 Tourismus
11 Verkehr und Nachrichtenwesen
12 Geld, Banken, Versicherungen
13 Soziale Sicherheit
14 Gesundheit
15 Bildung und Wissenschaft
16 Kultur, 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 Transports et communications
12 Monnaie, banques, assurances
13 Protection sociale
14 Santé
15 Education et science
16 Culture, 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

Enquête sur les chiffres d'affaires du commerce de détail

Plan d'échantillonnage et méthodes d'estimation

Auteur **Jean-Pierre Renfer**
Office fédéral de la statistique

Editeur **Office fédéral de la statistique**

Neuchâtel, 2006

Préambule

La statistique des chiffres d'affaires du commerce de détail DHU (Detailhandelsumsatzstatistik) est une enquête mensuelle de l'Office fédéral de la statistique (OFS) dont les résultats de décembre 2000 ont été les derniers a avoir été calculés selon l'ancienne statistique, révisée en 1953 pour la dernière fois. A partir du mois de janvier 2001, l'OFS a introduit des méthodes d'enquête modernisées permettant de mesurer l'évolution des chiffres d'affaires de l'ensemble du commerce de détail en recourant à un échantillon aléatoire stratifié selon la taille et l'activité économique de l'entreprise. Lors de la révision de cette statistique, cet échantillon de l'enquête conjoncturelle a été obtenu comme un sous-échantillon aléatoire des entreprises de l'échantillon de l'enquête structurelle lancée en mai 2000.

Jean-Pierre Renfer s'est vu confié les travaux méthodologiques pour établir le plan d'échantillonnage, procéder au tirage puis développer les estimateurs et leur variance en collaboration avec le service des méthodes statistiques (METH) de l'OFS.

Que les personnes ayant oeuvré à la révision de cette statistique conduite initialement par la section PREIS soient ici chaleureusement remerciées pour les nombreuses et intéressantes discussions, en particulier son chef de section Dieter Koch, le responsable du domaine Iba Ndiaye et la cheffe du projet Fabia Ndiaye-Laini.

Ce travail n'aurait pu voir le jour sans le soutien de Beat Hulliger (METH) pour ses conseils toujours avisés tout au long du projet ainsi que la relecture de ce rapport.

Résumé

L'objectif de l'enquête conjoncturelle DHU révisée est de fournir à un rythme mensuel une estimation de l'évolution des chiffres d'affaires du commerce de détail.

Cette étude a été menée dans le but de fournir le plan d'échantillonnage pour la DHU révisée. Dans ce cadre, les éléments essentiels ont pu être développés pour déterminer la meilleure stratégie pour le tirage des échantillons. Un problème important a pu être résolu en procédant à la reconstitution des entreprises à partir des informations obtenues sur leur siège et les établissements qui les constituent. Des solutions satisfaisantes ont également pu être trouvées pour traiter de la problématique des données, notamment en utilisant des modèles de régression robustes basées sur la norme l_1 adaptés au traitement des données aberrantes. Une grande partie de ce rapport a été consacrée aux différentes méthodes d'estimation et au calcul de la précision attendue en fonction du plan d'échantillonnage retenu. La recherche d'estimateurs par le quotient conjointement avec une allocation optimale de Neyman a permis un gain en précision non négligeable.

Mots-clé

allocation optimale; cadre de sondage; données aberrantes; estimateur par le quotient; plan d'échantillonnage; précision; rapport de méthodes; stratification.

Complément d'information:	Jean-Pierre Renfer, tél. 032 713 66 62 Jean-Pierre.Renfer@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:	338-0041
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 2006 La reproduction est autorisée, sauf à des fins commerciales, si la source est mentionnée.
ISBN:	3-303-00341-6

Table des matières

1 Introduction **5**

2 Traitement des données pour établir le plan d'échantillonnage **6**
2.1 Problématique des données 6
2.2 Délimitation de la population et reconstitution des entreprises 8
2.3 Traitement du problème des données aberrantes 12
2.4 Regroupement au niveau NOGA3 16

3 Plan d'échantillonnage **17**
3.1 Construction des strates 17
3.2 Allocation proportionnelle et optimale de Neyman 21
3.3 Comparaison entre un tirage aléatoire simple et stratifié 23

4 Optimisation du plan d'échantillonnage **24**
4.1 Recherche de l'estimateur le plus efficace 24
4.2 Recherche de l'allocation la plus efficace 28
4.3 Analyse de deux strates supplémentaires 29
4.4 Impact des taux de réponse sur la précision des estimateurs 30
4.5 Plan d'échantillonnage définitif : enquête structurelle 31
4.6 Plan d'échantillonnage définitif : enquête conjoncturelle 32

5 Forme des estimateurs et estimation de leur variance **34**
5.1 Notations 34
5.2 Paramètres à estimer 35
5.3 Précision des paramètres estimés 37

A Annexe : prototype du programme d'extrapolation **38**

B Références **67**

1 Introduction

La révision de la statistique des chiffres d'affaires du commerce de détail a pour but principal de fournir des résultats sur le chiffre d'affaires du commerce de détail en introduisant une enquête structurelle annuelle (DHU-S), permettant notamment de combler une information lacunaire jusqu'ici sur l'estimation du volume des chiffres d'affaires, ainsi qu'une enquête conjoncturelle mensuelle (DHU-K). Pour cela, contrairement à la DHU actuelle basée sur le principe du volontariat, le principe d'enquête par échantillonnage aléatoire a été adopté. Pour les deux enquêtes, l'unité de tirage sera l'entreprise. A noter que l'unité d'enquête (unité pour laquelle les informations sont collectées) sera le point de vente pour la DHU-S annuelle alors qu'il s'agira de l'entreprise pour la DHU-K mensuelle.
Un problème important qui doit être traité en amont de tout développement pour le plan d'échantillonnage, est la reconstitution de l'entité entreprise à partir des informations obtenues sur son siège et les établissements qui la constituent, notamment pour déterminer la branche économique dans laquelle elle exerce son activité principale. En effet, pour certaines entreprises possédant des établissements (filiales/succursales), ces derniers ne se trouvent pas tous forcément dans une branche économique de la DHU. Cette problématique doit être traitée de manière appropriée en considérant le nombre d'emplois équivalents à plein temps (en effet les postes à temps partiel et à temps complet peuvent se répartir de manière très inégale selon la branche économique et doivent être traitées sur une base commune) et en utilisant cette variable pour l'attribution de l'activité économique principale à l'entreprise.
Le but de ce document est notamment de fournir le plan d'échantillonnage pour chacune de ces deux enquêtes. Une partie importante sera consacrée à la recherche d'un plan d'échantillonnage efficace, basé sur une stratification judicieuse et permettant de fournir la taille des échantillons, d'indiquer l'allocation la plus efficace, de calculer les estimateurs les mieux adaptés et de retenir ceux dont la précision est la plus élevée.
La source d'information principale pour la mise en œuvre de ce plan d'échantillonnage est le recensement des entreprises et des établissements de 1995 (BZ 95). En effet, à cette occasion de nombreuses informations ont été relevées, notamment les chiffres d'affaires des entreprises ainsi que le nombre d'employés à temps partiel et à plein temps au niveau des établissements.
Les données sur le chiffre d'affaires des entreprises de la BZ95 ont dû être considérées avec circonspection et un traitement statistique approprié leur a été appliqué avant de procéder aux calculs permettant d'établir le plan d'échantillonnage. A noter que les résultats obtenus devraient permettre d'anticiper les problèmes rencontrés et, le cas échéant, fournir des solutions pour les traiter.
Ainsi, la structure de ce document suivra le développement logique de l'élaboration de cette étude en traitant au chapitre 2 de la problématique des données avant de passer à celle de la reconstitution des entreprises, puis au traitement du problème des données aberrantes. Le chapitre 3 est consacré au plan d'échantillonnage et le chapitre 4 à son optimisation. Finalement le chapitre 5 traite de l'estimation des paramètres et de leur précision.

2 Traitement des données pour établir le plan d'échantillonnage

2.1 Problématique des données

Les données étudiées dans ce document proviennent du recensement des entreprises 1995 (BZ 95) et ont été sauvegardées dans le data set SAS BZ95AST. Il s'agit des données au niveau des établissements (et non au niveau des entreprises) dont ceux appartenant à la DHU ont été extraits (selon la branche économique à laquelle ils appartiennent) avec le programme infoast.sas. Les variables d'intérêt sont les suivantes :

- $KANT$: Numéro OFS du canton dans lequel est situé l'établissement ;
- $BURNR$: Numéro OFS du BUR (unique, fourni par le système) ;
- $UNTKZ$: Lien permettant de retrouver les établissements formant une entreprise ;
- $BEART$: Code permettant de distinguer le siège de l'entreprise des filiales ;
- $BETYP$: fournit le type d'entreprise/établissement ;
- $NOGA5$: Code au niveau 5 de la nomenclature des activités économiques ;
- $BETT1$: Nombre d'employés travaillant à temps partiel (entre 50% et 90%) ;
- $BETT2$: Nombre d'employés travaillant à temps partiel (inférieur à 50%) ;
- $BETV$: Nombre d'employés travaillant à plein temps (au moins 90%) ;
- $BETOT = BETT1 + BETT2 + BETV$;
- $UUMS$: Chiffre d'affaires de l'entreprise exprimé en milliers de francs.

La variable $BEART$ est particulièrement importante car sa codification nous permet de distinguer si l'entreprise et l'établissement se confondent ou si une entreprise possède une ou plusieurs filiales/succursales. Dans le premier cas (entreprise=établissement), le code correspondant ($BEART = 4$) nous indique qu'il n'y a qu'un établissement qui correspond à l'entreprise (Einzelbetrieb). Ces entreprises ne posent pas de problème de reconstitution puisqu'elles se fondent en une seule unité de tirage (elles font en effet soit entièrement partie de la population soit n'y appartiennent pas). En revanche, les entreprises possédant une ou plusieurs filiales/succursales doivent être reconstituées en tenant compte de plusieurs variables : tout d'abord, le chiffre d'affaires de l'ensemble de l'entreprise est donné par l'*unique* établissement correspondant au siège de l'entreprise (ayant pour code $BEART = 3$). Les autres établissements d'une entreprise donnée ont pour code ($BEART = 5$). Grâce à la variable $UNTKZ$ on peut ainsi retrouver tous les établissements (siège compris) formant l'entreprise : il s'agit des établissements ayant le même UNTKZ. Les deux problèmes qui se posent pour reconstituer l'entreprise sont les suivants. D'une part, il s'agit de ne tenir compte que des établissements appartenant à la DHU pour imputer le chiffre d'affaires ainsi que le nombre d'emplois, de manière à reconstituer une entreprise DHU. D'autre part, il s'agit de fournir un critère pour déterminer l'activité principale DHU de l'entreprise en fonction de celle de ses établissements pour attribuer un code $NOGA5$ unique à l'entreprise reconstituée.
A noter que dans les branches économiques de la DHU, le nombre d'employés travaillant à temps partiel est important et la variable BETOT (somme de tous les

employés indépendamment de leur taux d'occupation) n'est pas adéquate pour expliquer le chiffre d'affaires de l'entreprise. En effet, il s'agit de calculer les emplois équivalents à plein temps (Vollzeitequivalent : $VZEQ$). Ce calcul a pu être fait à partir du taux d'occupation moyen de la section économique G de la NOGA recouvrant toutes les branches économiques de la DHU (Source : Statistique du volume du travail 1995, données harmonisées au concept du recensement des entreprises) : 59.74% pour les employés travaillant entre 50% et 90% ; pour ceux occupés à moins de 50%, ce taux s'élève à 32.13%. Le nombre d'emplois équivalents à plein temps (dans la suite du texte on parlera d'emplois) a donc été calculé selon l'équation

$$VZEQ = 0.5974 * BETT1 + 0.3213 * BETT2 + BETV.$$

Un autre problème auquel on se trouve confronté est celui des données ayant pour valeur 0. En effet, dans la BZ 95, aucune distinction n'a été faite entre 0 structurels et données manquantes. Si pour le nombre total d'employés, et donc à fortiori pour le nombre d'emplois équivalents à plein temps, ce problème n'existe pratiquement pas (à une ou deux exceptions près), le problème est différent en ce qui concerne le chiffre d'affaires : dans ce cas on trouve environ 20% des entreprises, notamment pour celles totalisant peu d'emplois, ayant un chiffre d'affaires nul dans les données de la BZ 95. Comme il s'agit des chiffres d'affaires du commerce de détail, il est peu probable qu'il s'agisse de 0 structurels. Nous les avons donc considérés ici comme des données manquantes. Ces données n'ont pas subi de traitement particulier, même si les modèles développés dans le cadre de ce document permettent d'imputer les chiffres d'affaires obtenus à partir de ces modèles. La raison principale de ce choix est qu'un tel traitement des données manquantes aurait eu pour conséquence d'obtenir des estimateurs avec une variance plus petite que celle calculée sur les données à disposition. Les conclusions de cette étude auraient ainsi été faussées par le fait qu'une sous-estimation de la variance aurait conduit à une précision par trop optimiste, s'éloignant de la réalité d'une enquête par échantillonnage de ce type. On peut en effet s'attendre à ce que les effets de la non-réponse puissent poser certains problèmes, notamment au niveau des petites entreprises/établissements. Il faut souligner qu'il ne s'agit pas d'un problème directement lié à l'élaboration du plan d'échantillonnage pour lequel les calculs ont pu être faits sur les données nettes et ainsi permettre de calculer les estimateurs en se basant sur l'ensemble de toutes les observations de la population.

Il faut en outre indiquer qu'il est parfaitement envisageable, selon l'ampleur du phénomène, de traiter le problème des données manquantes dans l'enquête proprement dite, en se basant notamment sur les types de modèles développés ici.

Finalement, comme évoqué dans l'introduction, les données sur le chiffre d'affaires des entreprises doivent être considérées avec prudence pour pouvoir les exploiter de manière appropriée. Il s'avère qu'un petit pourcentage des données sur les chiffres d'affaires ne sont pas plausibles et apparaissent comme des données aberrantes, aussi appelées « outliers ». Ce problème semble provenir d'une part de la récolte des données auprès des entreprises qui devaient fournir le chiffre d'affaires en milliers de francs (une erreur d'un facteur mille apparaît lorsque ce chiffre a été fourni en francs) et d'autre part à la saisie des données par lecture optique (une erreur d'un facteur

10 apparaît lorsqu'un 1 a été rajouté au chiffre d'affaires car confondu avec le trait à droite de la case permettant de saisir le chiffre d'affaires). Le taux de contamination est faible, de l'ordre de 2-3% dans la plupart des branches économiques au niveau $NOGA5$, mais a des conséquences désastreuses sur les estimateurs ainsi que sur l'estimation de leur précision. Un traitement statistique présenté dans le paragraphe 2.3 et basé sur une méthode adaptée à cette problématique a permis de résoudre ce problème de manière satisfaisante.
Cette méthode, conjointement avec une méthode classique comme celle des moindres carrés, pourrait également s'avérer particulièrement utile dans le traitement, l'exploitation et l'analyse des données aussi bien de l'enquête structurelle DHU-S que conjoncturelle DHU-K.

2.2 Délimitation de la population et reconstitution des entreprises

Dans un premier temps, tous les établissements ayant un code *NOGA5* de la DHU ont été sélectionnés. On trouve 39387 établissements qui sont confondus avec l'entreprise (*BEART=4*), 4249 sièges d'entreprises (*BEART=3*) et 15566 succursales/filiales (*BEART=5*). Seuls posent problème les 19815 (4249+15566) établissements. Il s'agit de retrouver l'entreprise à laquelle ils appartiennent grâce à la variable *UNTKZ*. On trouve ainsi 5454 *UNTKZ* différents, soit 5454 entreprises dont au moins un établissement fait partie de la DHU. En effet, il y a 1206 entreprises qui n'ont pas leur siège dans la DHU et 741 qui n'ont que leur siège dans la DHU selon l'illustration de la Figure 1.

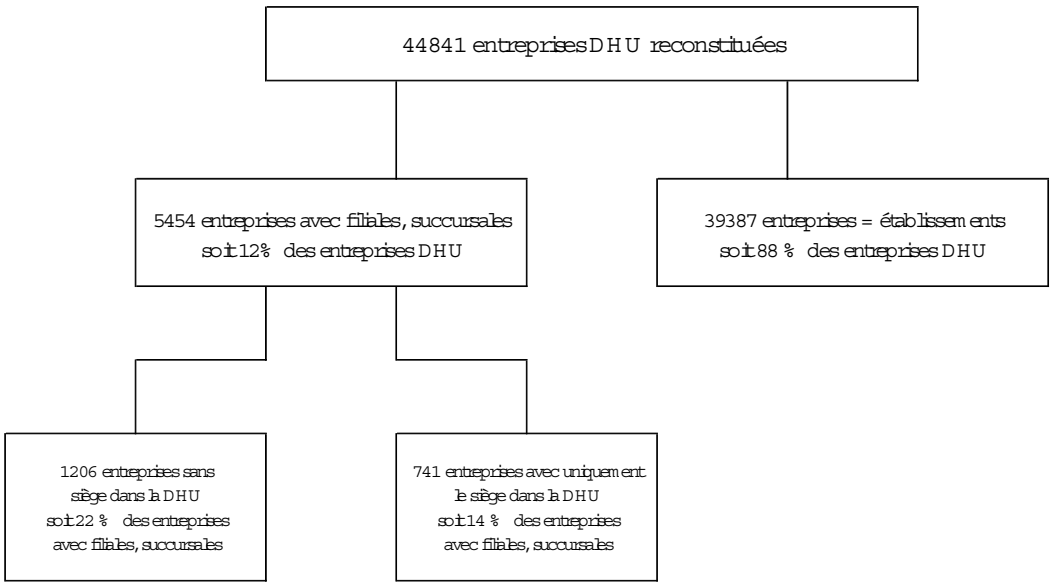


FIG. 1 – Répartition des entreprises de la DHU et problématique des sièges.

Il s'agit donc de traiter le problème de la reconstitution des entreprises possédant des filiales/succursales en fonction du fait qu'une partie des établissements d'une entreprise peut se trouver en dehors du domaine de la DHU. Ce point est particulièrement important pour les entreprises qui ont leur siège en dehors de la DHU puisque seul ce dernier permet de connaître le chiffre d'affaires de l'ensemble de l'entreprise. De plus, même si le siège fait partie de la DHU et qu'ainsi le chiffre d'affaires est connu pour l'entreprise, il s'agit de rechercher les établissements hors DHU de manière à pouvoir redistribuer le chiffre d'affaires en conséquence pour reconstituer une entreprise essentiellement DHU. A noter que pour retrouver tous les établissements dont l'entreprise correspondante est constituée d'au moins un établissement de la DHU, un fichier nommé *construction.xls* a été créé contenant les 5454 UNTKZ différents. En comparant ces codes à ceux de l'ensemble des établissements (sièges ou filiales/succursales au nombre de 102874) on retrouve ainsi tous les établissements à considérer, soit 28785 figurant dans le fichier *astuntpoursplus.txt*.
Pour l'attribution du chiffre d'affaires à l'entreprise DHU, le critère retenu est celui d'une imputation du chiffre d'affaires proportionnelle aux emplois équivalents à plein temps en faisant la somme des emplois des établissements de la DHU (x_1) et celle des emplois des établissements hors de la DHU (x_2) comme indiqué à la Figure 2.

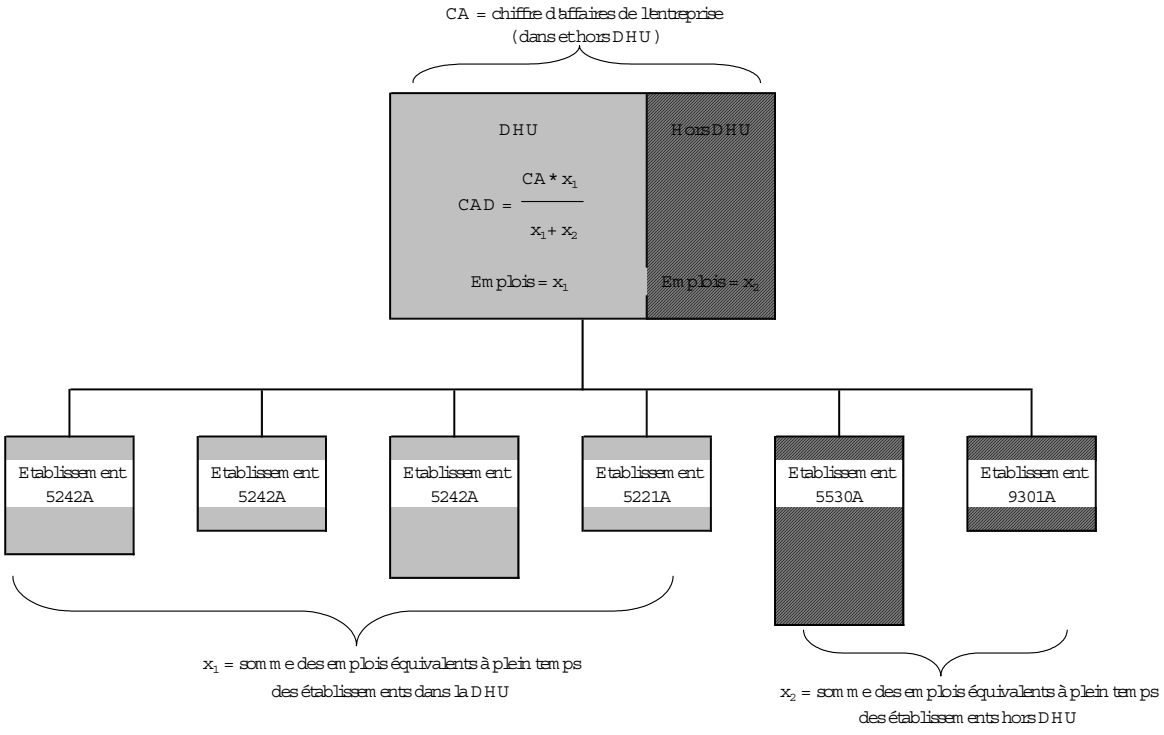


FIG. 2 – Attribution du chiffre d'affaires à l'entreprise DHU.

Comme indiqué sur l'illustration de la Figure 2, le nombre d'emplois de l'entreprise a été calculé en faisant la somme des emplois des établissements se trouvant dans

la DHU. Il reste encore à attribuer le code NOGA5 correspondant à la branche économique dans laquelle l'entreprise reconstituée exerce son activité principale. Ce code a été attribué à partir du code NOGA5 de ses établissements, comme indiqué dans l'exemple illustré à la Figure 3.

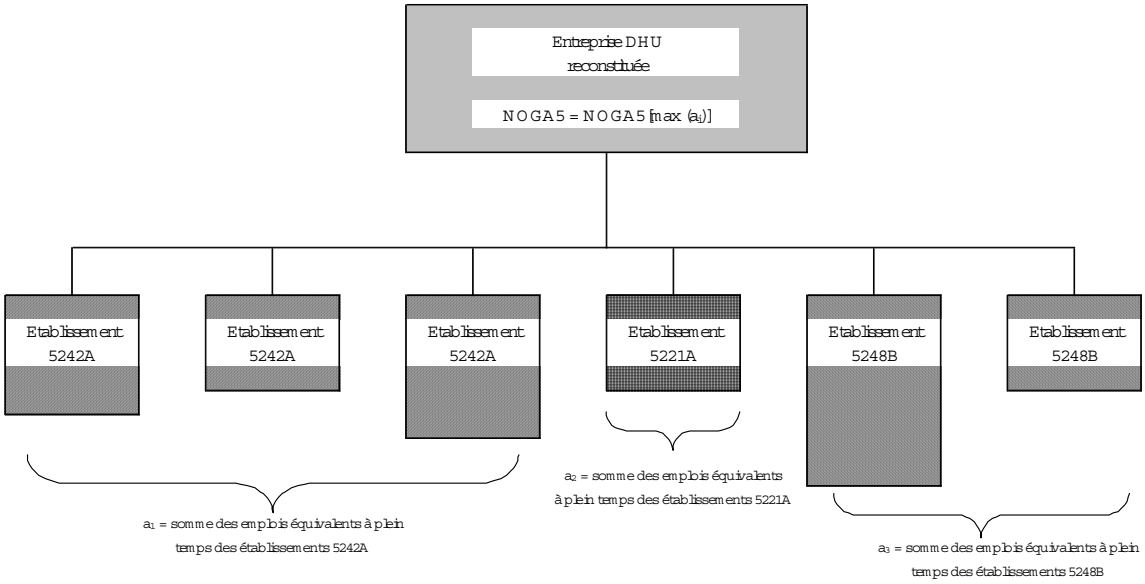


FIG. 3 – Attribution du code NOGA5 à l'entreprise DHU reconstituée.

Le critère qui a été retenu pour déterminer l'activité principale de l'entreprise est le suivant : pour les établissements d'une même branche d'activité économique (même code NOGA5), le nombre total d'emplois a été calculé pour retenir la branche totalisant le plus d'emplois. C'est le code NOGA5 correspondant qui a été attribué à l'entreprise, comme indiqué sur la Figure 3.
A noter qu'un prototype de programme a été développé dans Splus pour la reconstitution des entreprises comme décrit ci-dessus. Après vérification, les codes ont été correctement attribués, sauf dans cinq cas exceptionnels : il s'agit d'entreprises dont tous les établissements de la DHU n'ont pas donné leur nombre d'employés alors que tous les établissements hors de la DHU les ont fournis. Dans ces cinq cas, le code correspondant au nombre maximal d'établissements de la DHU a été attribué après coup à ces entreprises.
La Figure 4 permet de visualiser dans quelle mesure les entreprises reconstituées ont été touchées par le fait qu'une partie de leurs établissements se trouvaient hors de la DHU.
Par exemple, la hauteur de la première barre indique qu'il y a environ 400 entreprises dont le nombre d'emplois dans la DHU ne s'élève pas à plus de 10% ou, en d'autres termes, l'ensemble des établissements hors de la DHU totalisent entre 90% et 100% des emplois de l'entreprise. La dernière barre, soit un peu plus de 3000 entreprises ne sont pas (ou de façon très marginale) touchées par ce problème.

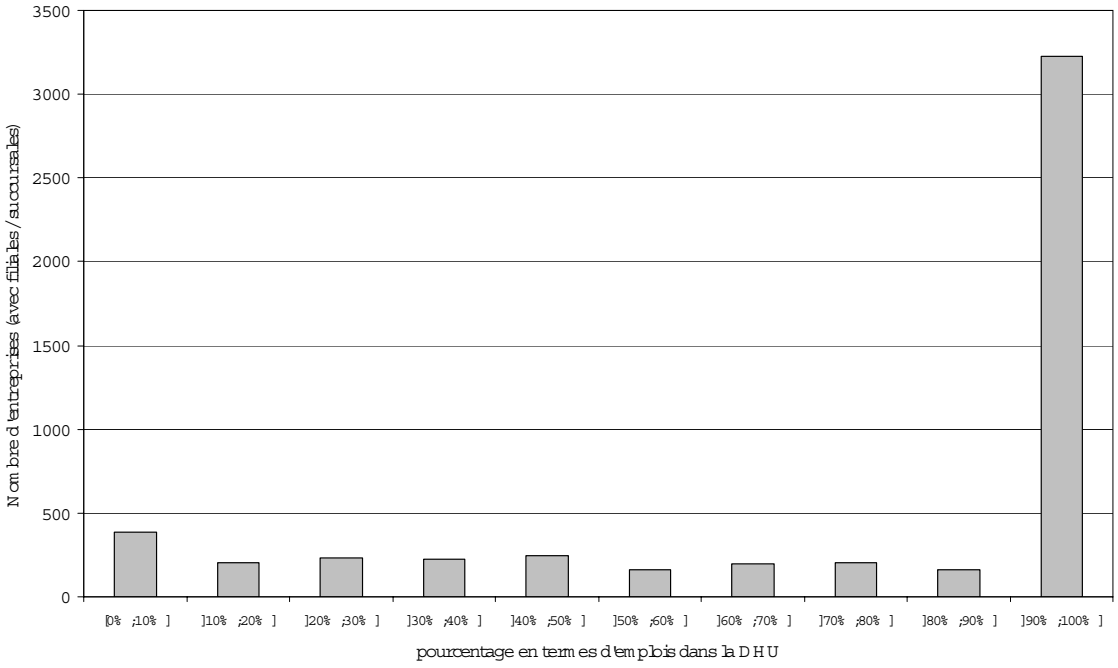


FIG. 4 – Répartition empirique des entreprises touchées par le problème en/hors DHU.

Ainsi, sur les 5454 entreprises reconstituées, près des 2/3 ne sont pas touchées par le problème lié à des établissements hors de la DHU. Par contre, pour les autres, soit environ 1/3 d'entreprises, celles-ci sont touchées par ce phénomène et se répartissent de manière relativement uniforme en termes de pourcentage d'emplois des établissements de la DHU.
Pour terminer ce paragraphe, mentionnons encore que le chiffre d'affaires de chaque entreprise reconstituée a ainsi pu être distribué proportionnellement au nombre d'emplois de ses établissements, comme indiqué sur la Figure 5.
Nous avons ainsi finalement pu obtenir un fichier entreprises avec les informations nécessaires à l'étude de l'ensemble des entreprises de la DHU en regroupant les entreprises reconstituées et celles ne présentant pas ce problème (Einzelbetrieb).
D'autre part, un fichier établissements a pu être constitué à partir des établissements des entreprises reconstituées et de ceux qui se confondent avec l'entreprise (Einzelbetrieb).

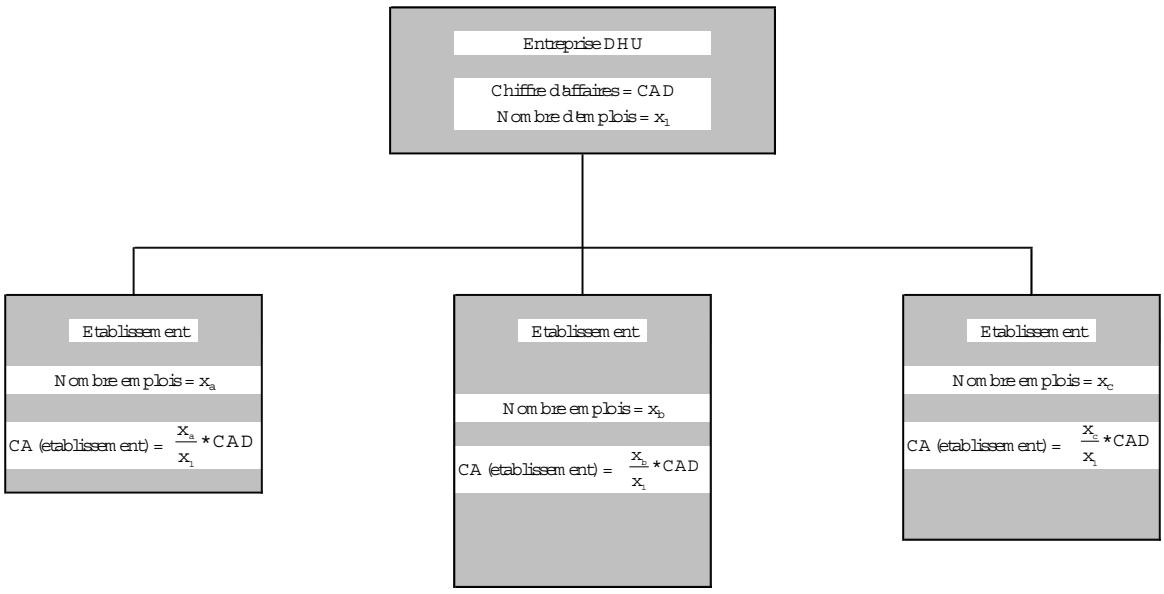


FIG. 5 – Attribution du chiffre d'affaires aux établissements de l'entreprise DHU.

2.3 Traitement du problème des données aberrantes

Comme exposé dans le paragraphe consacré à la problématique des données, il faut traiter de manière appropriée les données contaminées (d'un facteur 1000 ou 10 respectivement). Pour cela, des méthodes ont été élaborées qui permettent de se prémunir contre les effets de ce genre de données. Ces méthodes, dites robustes, sont particulièrement adaptées à la situation rencontrée dans notre ensemble de données sur les chiffres d'affaires du commerce de détail. De manière à illustrer le problème rencontré, nous donnons une représentation graphique sur la Figure 6 de l'une des 70 positions NOGA5 en termes d'emplois équivalents à plein temps en abscisse et du chiffre d'affaires correspondant (exprimé en milliers de francs) en ordonnée. Le résultat obtenu, tout à fait comparable à celui obtenu pour les autres positions NOGA5, est décevant et il ne semble pas y avoir de relation linéaire entre le nombre d'emplois et les chiffres d'affaires. Cela provient essentiellement des unités sur l'axe vertical où l'on voit clairement le problème des données aberrantes (des chiffres d'affaires de l'ordre de 1.5 milliard de francs avec 1 à 5 emplois ne sont pas plausibles dans cette position NOGA5 : il s'agit sans doute de chiffres d'affaires exprimés en francs plutôt qu'en milliers de francs). D'ailleurs la corrélation entre ces deux variables (calculée sans les données manquantes, c'est-à-dire sur les données de base) ne s'élève qu'à 7%.

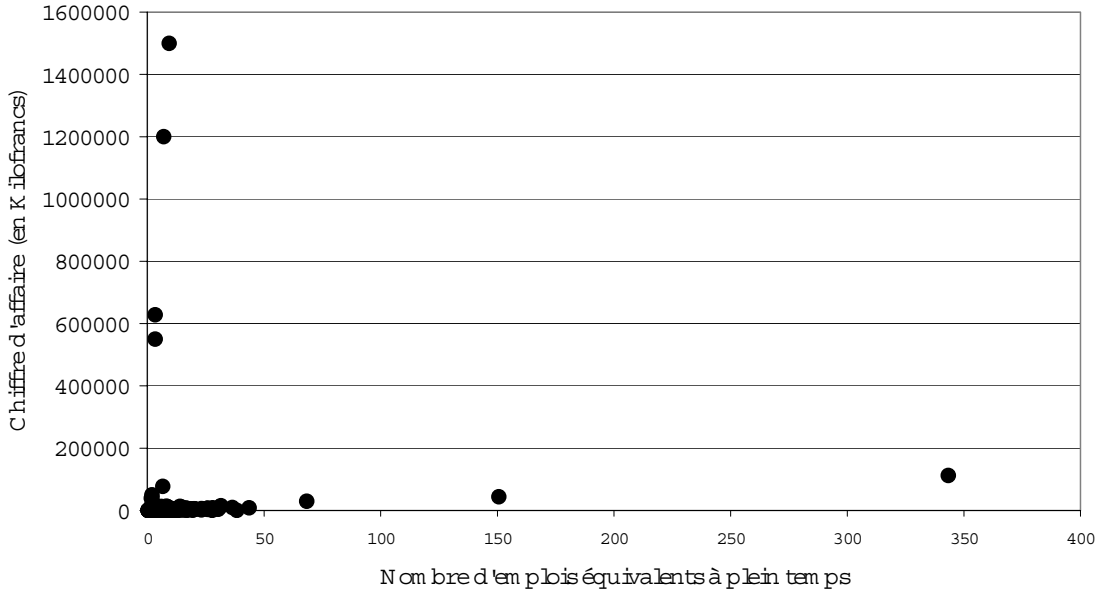


FIG. 6 – Chiffre d'affaires et emploi des données de base.

Des modèles de régression linéaire (droites passant par l'origine) peuvent être calculés pour expliquer les chiffres d'affaires à partir des emplois équivalents à plein temps. La méthode la plus connue est celle des moindres carrés (estimation l_2), qui peut être considérée comme une généralisation de la moyenne dans le cas bivarié. La pente de la droite ainsi obtenue est fortement influencée (à la hausse) par les données aberrantes, et est donc certainement surestimée. A noter que la pente d'une droite reliant l'origine à une observation représente le chiffre d'affaires moyen par emploi pour l'entreprise considérée. Une mesure plus appropriée serait une généralisation de la médiane dans le cas bivarié, puisque moins sensible aux données aberrantes. Il s'avère qu'une méthode robuste consistant à minimiser la somme des résidus en valeur absolue (estimation l_1), revient en fait à calculer la médiane pondérée (par les emplois) et fournit une estimation de la pente pratiquement pas influencée par les données aberrantes. La Figure 7 illustre le comportement des droites l_1 et l_2 en présence de ces données aberrantes (les échelles ont été choisies ici de manière à représenter la partie comprenant l'essentiel des données en agrandissant la partie inférieure gauche du graphique). A noter que la plupart des observations de la partie supérieure gauche de ce graphique ont un chiffre d'affaires se terminant par un 1, et sont ainsi symptomatiques du problème lié à la lecture optique des données (contamination d'un facteur 10 environ des chiffres d'affaires).
A noter que la pente de la droite obtenue par l'estimation l_2 vaut plus du double de celle obtenue par l'estimation l_1 !

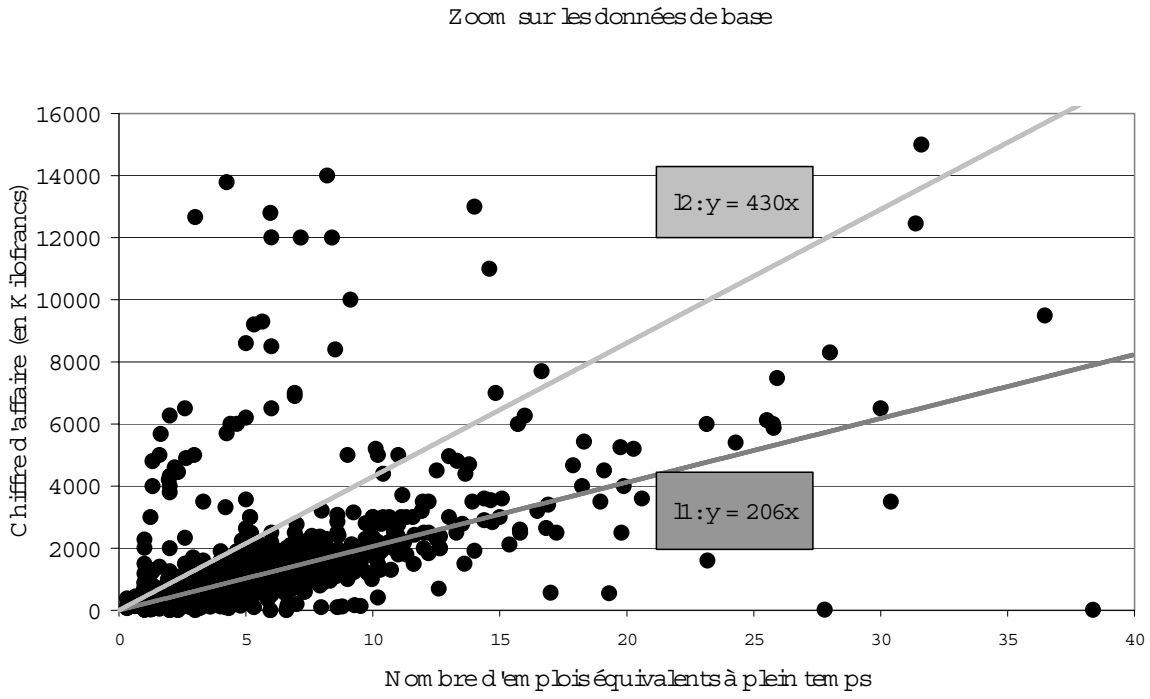


FIG. 7 – Régression linéaire l_1 et l_2 sur les données de base.

De manière à pouvoir fournir une estimation fiable de la moyenne et de la variance des chiffres d'affaires dans les différentes positions NOGA, il est indispensable de traiter ces données aberrantes qui faussent complètement les calculs. Il s'agit en effet de trouver un critère permettant de ne garder pour les calculs que les données fiables et d'éliminer celles considérées comme aberrantes. Les données fiables étant bien modélisées par la droite l_1, celle-ci intervient de manière naturelle dans un critère lié à sa pente : les observations pour lesquelles le chiffre d'affaires moyen par emploi dépasse 4 fois la pente de la droite l_1 seront considérées comme aberrantes et donc éliminées des calculs ultérieurs pour établir le plan d'échantillonnage. A noter que le facteur 4 de la pente de la droite l_1 a été retenu après plusieurs tentatives sur les 70 positions NOGA5 de considérer d'autres facteurs. Il est apparu qu'un facteur inférieur à 4 conduisait à considérer trop d'observations comme aberrantes (critère trop sévère) alors qu'un facteur supérieur à 4 ne permettait pas de discriminer toutes les données aberrantes, notamment celles contaminées par un facteur 10 (facteur trop conservateur). La Figure 8 représente les données nettes, obtenues par le traitement décrit ci-dessus, ainsi que les modèles l_1 et l_2 recalculés sur ces données. On constate que l'estimation de la pente de la droite l_1 est restée pratiquement la même (insensibilité aux données aberrantes) alors que celle de l'estimation l_2 s'est clairement modifiée à la baisse, fournissant ainsi une estimation plus réaliste qu'avec les données contaminées.

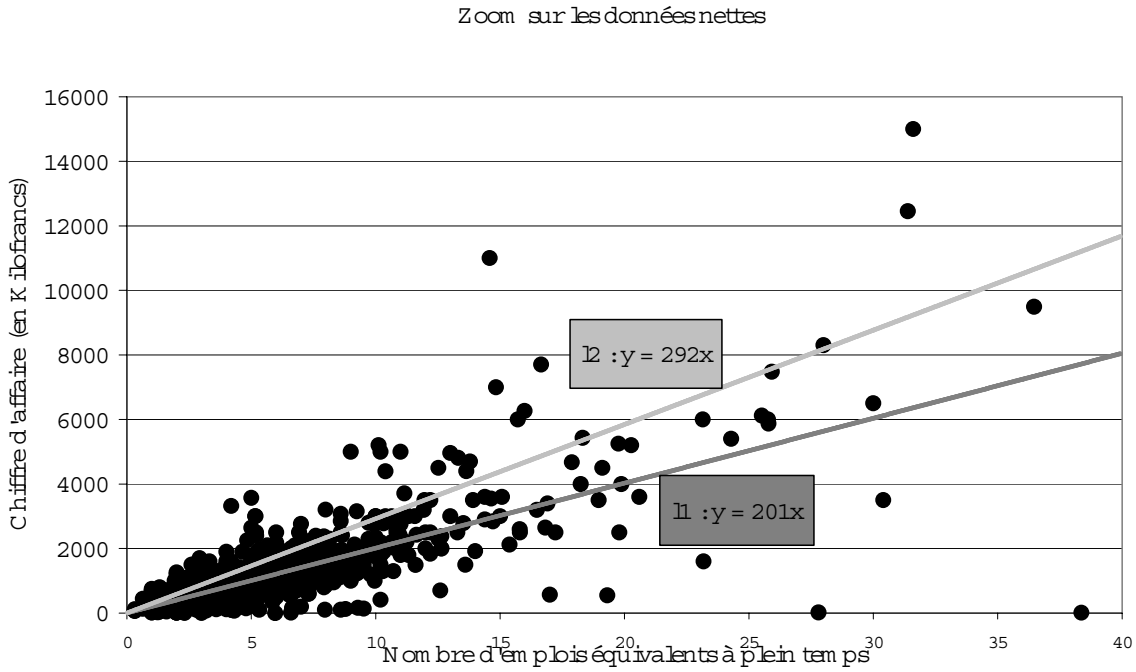


FIG. 8 – Régression linéaire l_1 et l_2 sur les données nettes.

Grâce au logiciel Splus qui contient une fonction permettant de calculer les estimateurs l_1, un programme a été développé pour le traitement des données aberrantes dans chacune des 70 positions NOGA5. Tous les résultats statistiques relatifs à ce traitement ont été sauvegardés et permettent d'évaluer les différences obtenues avant et après élimination des données aberrantes. Dans l'exemple illustré dans ce paragraphe, le pourcentage de données aberrantes se monte à 4%. En éliminant ces données aberrantes et en recalculant la corrélation entre le chiffre d'affaires et le nombre d'emplois, celle-ci passe de 7% à 97%, montrant ainsi qu'un faible taux de contamination peut non seulement avoir un effet catastrophique sur l'estimation de la moyenne des chiffres d'affaires mais également sur le coefficient de corrélation entre chiffres d'affaires et emplois. La Figure 9 met en évidence ce phénomène en fournissant les coefficients de corrélation entre ces deux variables avant et après élimination des données aberrantes dans les différentes positions NOGA5, préalablement triées dans l'ordre croissant des coefficients de corrélation obtenus avant traitement des données aberrantes.

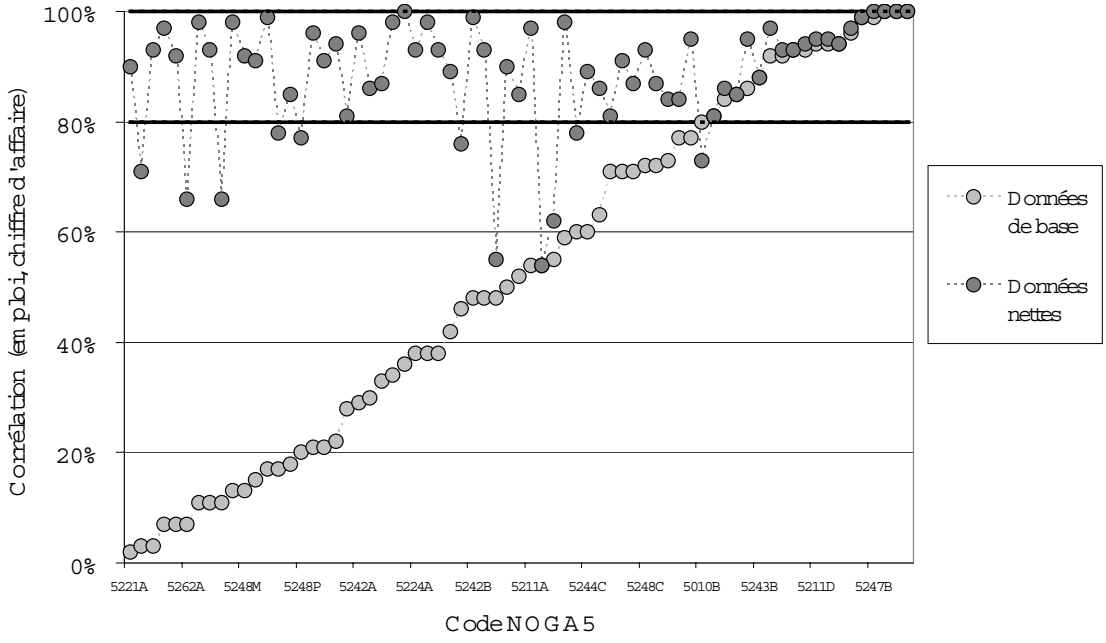


FIG. 9 – Gain en corrélation après traitement des données ("outliers" éliminés).

2.4 Regroupement au niveau NOGA3

Le but de cette étude étant de développer un plan d'échantillonnage aléatoire stratifié, il est naturel de retenir la branche d'activité économique pour la formation de strates dans lesquelles les chiffres d'affaires sont relativement homogènes. Pour cela, il n'est pas concevable de rester au dernier niveau de la NOGA. Il s'agit donc de regrouper les entreprises selon une position hiérarchiquement supérieure au niveau NOGA5. Plusieurs regroupements ont été envisagés, avant de porter notre choix sur la position NOGA3. En effet, cette dernière présente l'avantage de réduire le nombre d'activités économiques différentes à 11, tout en conservant l'homogénéité relative des chiffres d'affaires en procédant à un regroupement à ce niveau. A noter que ce regroupement s'est fait après traitement des données aberrantes au niveau NOGA5. Pour illustrer cette situation, la Figure 10 permet de visualiser un tel regroupement pour l'une des positions au niveau NOGA3. Puisqu'il s'agit de données nettes, on constate que le modèle fourni par la droite l_2 traduit bien la relation linéaire existant entre les chiffres d'affaires et les emplois.

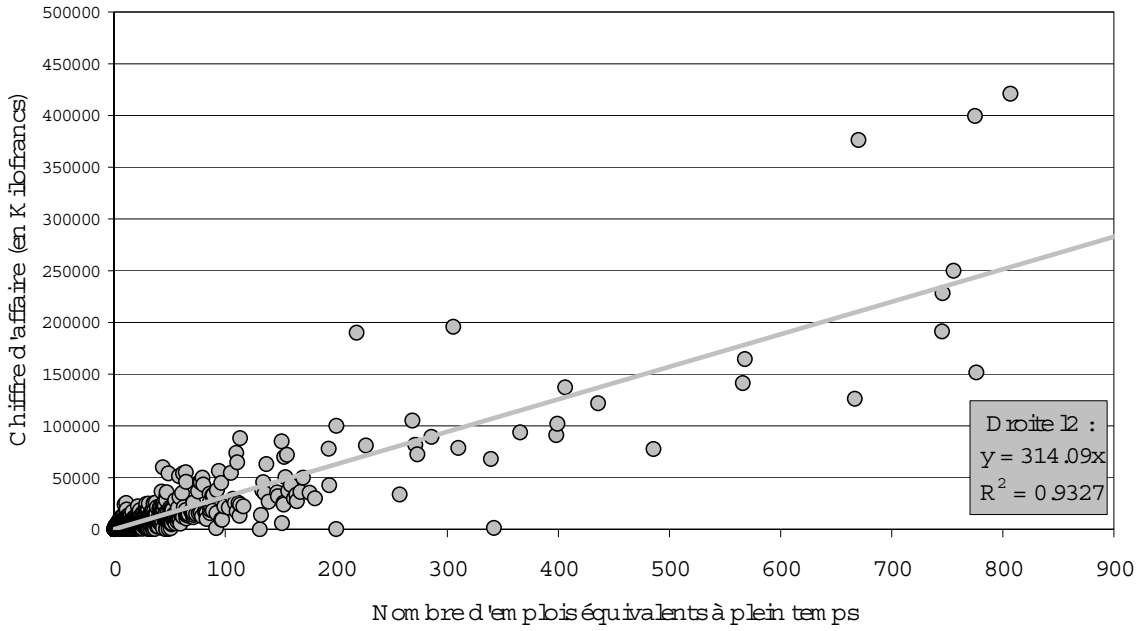


FIG. 10 – Régression linéaire par les moindres carrés sur les données nettes.

3 Plan d'échantillonnage

3.1 Construction des strates

De manière à pouvoir établir un plan d'échantillonnage aussi adapté que possible à la réalité des entreprises en Suisse, il est prévu de catégoriser les entreprises selon leur taille en trois grands groupes : les petites, les moyennes et les grandes entreprises. Le but est de procéder à un échantillonnage des petites et moyennes entreprises alors que les grandes seront recensées. Pour cela, du fait de la très bonne corrélation entre le nombre d'emplois et le chiffre d'affaires, il est raisonnable de faire la distinction entre ces trois groupes d'entreprises en fonction du nombre d'emplois de ces dernières. Il reste à déterminer jusqu'à combien d'emplois une entreprise peut encore être considérée comme une petite entreprise et à partir de combien d'emplois doit-elle être considérée comme une grande entreprise. Ces deux coupures suffisent à déterminer les trois catégories (les moyennes entreprises sont celles qui se trouvent comprises entre ces deux bornes). Cette problématique connaît des solutions dans la littérature, notamment la technique cumulée $\sqrt{f}$, brièvement rappelée ici. Soit x le nombre d'emplois et $f_x(x)$ sa fonction de densité. On définit $a_0 = A_L$ et $a_H = A_U$, où H est le nombre de strates à former. Il s'agit donc de trouver $(H-1)$ bornes telles que $A_L = a_0 < a_1 < a_2 < \cdots < a_{H-1} < a_H = A_U$. Cette technique permet de construire H strates de manière à ce que pour $h = 1, \ldots, H$, l'on ait :

$$\int_{a_{h-1}}^{a_h} \sqrt{f_x(x)}dx = \frac{1}{H}\int_{A_L}^{A_U} \sqrt{f_x(x)}dx$$

La forme exacte de la fonction de densité continue n'est en général pas connue et il faut se contenter d'une approximation, comme celle basée sur un histogramme comprenant J classes de longueur égale. La règle revient ainsi à calculer $\sqrt{f_j}$, c'est-à-dire la racine de la fréquence des valeurs de x dans chaque classe de façon à être aussi proche que possible de

$$\frac{1}{H}\sum_{j=1}^{J} \sqrt{f_j}.$$

Une analyse préliminaire de la répartition du nombre d'emplois dans les entreprises au niveau $NOGA3$ a été menée. Les résultats obtenus sont très semblables à ceux présentés sur l'exemple de la Figure 11. On constate que la forme de la courbe de Lorenz obtenue est pratiquement la même dans les 11 positions, avec un changement de courbure très marqué au niveau d'environ 15 emplois, fournissant ainsi de manière empirique une première borne pour catégoriser les petites entreprises. Pour déterminer la seconde borne, nous pouvons appliquer la technique cumulée $\sqrt{f}$ aux entreprises ayant plus de 15 emplois.

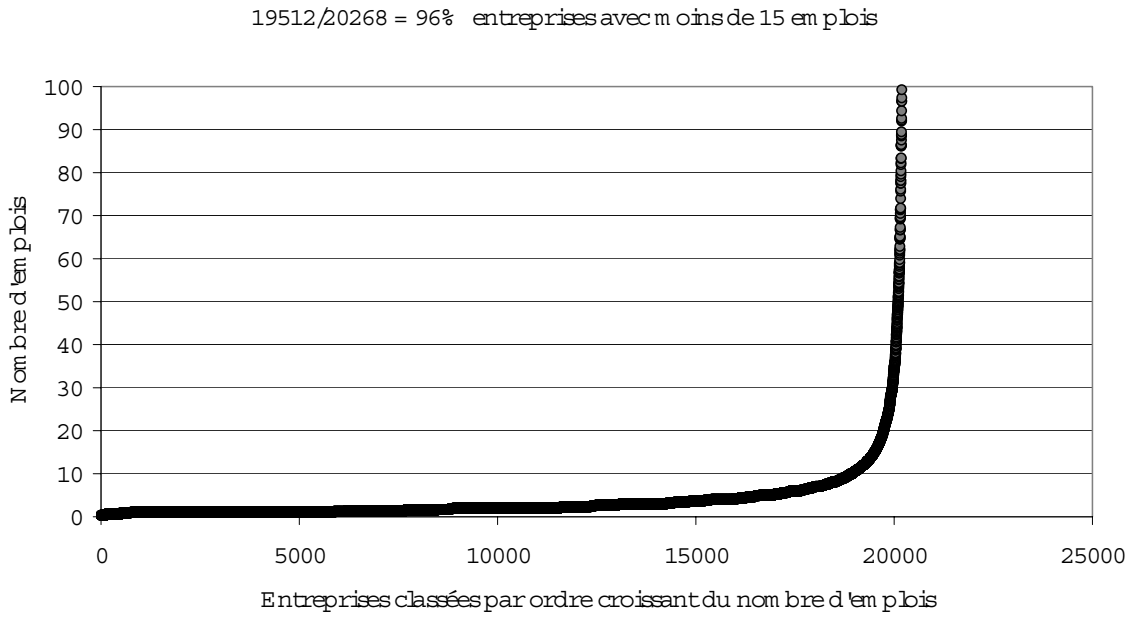


FIG. 11 – Courbe de Lorentz de la distribution du nombre d'emplois.

Pour illustrer la méthode de calcul, nous indiquons comment obtenir les résultats pour l'une des positions $NOGA3$ en considérant 10 intervalles de longueur 20 (le

dernier intervalle contenant toutes les entreprises avec plus de 195 emplois). La fréquence correspondante aux entreprises comprenant entre 15 et 35 emplois est de 496. En calculant sa racine, on obtient $\sqrt{f_1} = \sqrt{496} = 22.27$, alors que pour le second intervalle (nombre d'emplois compris entre 35 et 55) on trouve $\sqrt{f_2} = \sqrt{115} = 10.72$ et ainsi de suite. Le cumul de ces valeurs vaut 64.66. Il s'agit donc de trouver à partir de quel intervalle, le cumul partiel de ces valeurs est le plus proche de 64.66/2=32.33. On constate qu'en faisant la somme des deux premières valeurs obtenues (22.27+10.72=32.99), cette condition est satisfaite. Par conséquent, la borne recherchée se trouve dans le second intervalle, soit celui comprenant entre 35 et 55 emplois. Les résultats obtenus dans les autres positions $NOGA3$ sont semblables, à l'exception de la position 521 comprenant, notamment les hypermarchés. Les coupures obtenues (en considérant l'extrémité droite de l'intervalle retenu) sont fournies dans le Tableau 1.

Activité économique $NOGA3$	501	503	504	505	521	522	523	524	525	526	527
coupure	**30**	**35**	**35**	**30**	**275**	**35**	**35**	**55**	**45**	**55**	**45**

Tableau 1 : Coupures pour déterminer la limite des grandes entreprises.

On constate ainsi que les résultats sont homogènes (à l'exception de la position 521 citée ci-dessus) et qu'il est donc raisonnable de considérer la seconde coupure à 45 emplois. Finalement, la démarche adoptée dans ce paragraphe nous conduit à fournir la catégorisation suivante des entreprises : seront considérées comme *petites entreprises*, celles comprenant jusqu'à 15 emplois, comme *moyennes entreprises* celles comptant entre 15 et 45 emplois et finalement comme *grandes entreprises* celles comprenant plus de 45 emplois. A noter qu'il est possible de calculer globalement le facteur permettant de passer du nombre d'emplois à celui du nombre d'employés, en divisant par 0.8311 le nombre d'emplois (résultat obtenu à partir d'un modèle de régression linéaire passant par l'origine sur l'ensemble des observations en considérant d'une part la somme de tous les employés et d'autre part le nombre d'emplois équivalents à plein temps correspondant, obtenu par la méthode décrite dans le paragraphe 2.1).

Grâce à la détermination des tailles limites des entreprises du paragraphe précédent et à l'activité économique au niveau $NOGA3$ dans laquelle l'entreprise exerce son activité principale, il est désormais possible de construire les strates résultant de ces informations. Comme il y a 11 positions $NOGA3$ auxquelles peut appartenir l'entreprise et que celle-ci peut être catégorisée selon sa taille (en termes d'emplois) en petite, moyenne ou grande entreprise, il y aura au total 33 strates.

Pour chacune d'elle, il s'agit d'obtenir les informations nécessaires à l'établissement du plan d'échantillonnage. Pour cela, nous fournissons dans le Tableau 2 les tailles de la population dans chaque strate, désignées par N_h : il s'agit du nombre d'entreprises recensées en 1995. Dans ce même tableau, nous indiquons les tailles nettes N_h^* qui

correspondent aux observations qui ont effectivement pu être utilisées pour les calculs (la différence correspond en grande partie aux données manquantes et dans une moindre mesure aux données aberrantes).

Activité économique $NOGA3$	Petites entreprises (≤ 15 *emplois*)		Moyennes entreprises		Grandes entreprises (> 45 *emplois*)		Total	
	N_h	N_h^*	N_h	N_h^*	N_h	N_h^*	N_h	N_h^*
501	503	347	13	11	4	3	520	361
503	304	242	9	8	4	4	317	254
504	1131	874	7	5	3	1	1141	880
505	684	480	12	10	2	2	698	492
521	3203	2528	68	60	87	75	3358	2663
522	6958	5424	352	310	50	45	7360	5779
523	2639	2222	89	77	21	17	2749	2316
524	24780	19512	648	564	203	192	25631	20268
525	1030	718	10	8	5	5	1045	731
526	726	521	43	37	31	29	800	587
527	1203	865	13	9	6	5	1222	879
Total	43156	33733	1264	1099	416	378	44841	35210

Tableau 2 : Tailles brutes et nettes des strates.

Il faut encore connaître dans chaque strate la moyenne et l'écart-type sur les emplois ($\overline{x}_h$, respectivement S_{x_h}) qui figure dans le tableau 3 ainsi que sur les chiffres d'affaires ($\overline{y}_h$, respectivement S_{y_h}), comme fourni dans le tableau 4.

Activité économique $NOGA3$	Petites entreprises (≤ 15 *emplois*)		Moyennes entreprises		Grandes entreprises (> 45 *emplois*)	
	$\overline{x}_h$	S_{x_h}	$\overline{x}_h$	S_{x_h}	$\overline{x}_h$	S_{x_h}
501	2.44	2.57	27.1	7.73	55.3	5.96
503	2.99	2.63	26.9	8.14	69.9	23.0
504	2.61	2.04	25.6	11.0	157	.
505	3.29	2.54	28.6	8.61	147	60
521	2.66	2.23	23.8	7.14	930	950
522	4.12	3.10	23.2	7.37	90	92
523	4.83	2.87	21.8	6.62	102	75
524	2.78	2.51	24.1	7.65	184	351
525	1.60	1.43	28.9	10.0	78	33
526	2.74	2.90	24.9	7.96	152	178
527	1.62	1.47	28.6	8.61	188	141

Tableau 3 : Moyenne et écart-type des emplois (x) par strate.

Activité économique $NOGA3$	Petites entreprises (≤ 15 *emplois*)		Moyennes entreprises		Grandes entreprises (> 45 *emplois*)	
	$\overline{y}_h$	S_{y_h}	$\overline{y}_h$	S_{y_h}	$\overline{y}_h$	S_{y_h}
501	1182	1963	12123	7812	11797	15038
503	843	961	8550	5162	20494	9336
504	605	731	9051	9995	105386	.
505	2583	3014	14897	8788	55214	77932
521	824	1158	8276	5016	323530	327018
522	672	660	3441	4630	17804	32035
523	1139	907	5046	2785	31241	27351
524	621	933	6151	5237	55717	116227
525	146	241	2557	3190	6904	12381
526	776	1556	9847	9156	57635	60056
527	185	278	5094	3961	38860	25692

Tableau 4 : Moyenne et écart-type des chiffres d'affaires (y) par strate.

3.2 Allocation proportionnelle et optimale de Neyman

Avant de nous concentrer sur les estimateurs à utiliser et le calcul de leur précision, nous mentionnons dans ce paragraphe les types d'allocation considérés dans

ce document. Dans le cas d'un tirage aléatoire stratifié il s'agit de déterminer, pour une taille d'échantillon n donnée, la taille de l'échantillon dans chaque strate h, c'est-à-dire n_h. L'allocation la plus simple est celle consistant à allouer la taille de l'échantillon dans la strate h proportionnellement à la taille de la population N_h de cette strate. Une autre stratégie pour déterminer la taille de l'échantillon dans chaque strate, plus connue sous le nom d'allocation optimale de Neyman, revient à prendre en compte le coût des observations à récolter. Si le coût unitaire est constant dans chaque strate, ce qui est relativement réaliste dans notre cas, le taux de sondage dans chaque strate devra être proportionnel à l'écart-type de sa population. En effet, dans notre cas, plus l'écart-type des chiffres d'affaires dans une strate est élevé, plus il faudra effectuer un nombre important d'observations pour estimer avec précision le total du chiffre d'affaires pour cette strate. La variabilité des chiffres d'affaires pour les grandes entreprises étant nettement plus importante que pour les moyennes et petites entreprises, ceci explique la pratique de recenser les grandes entreprises et de procéder à un sondage pour les moyennes et petites entreprises. Nous avons ainsi considéré l'allocation proportionnelle et l'allocation optimale en calculant la taille de chaque strate de la manière suivante :

$$\text{Allocation proportionnelle} : n_h = n\frac{N_h}{N};$$

$$\text{Allocation optimale de Neyman} : n_h = n\frac{N_h S_{y_h}}{\sum_h N_h S_{y_h}}.$$

Les calculs ont été effectués en prenant une taille d'échantillon de n=1800 pour les deux types d'allocation et apparaissent dans le tableau 5. Les tailles de chaque strate devant être des entiers, les résultats ont été arrondis à l'entier le plus proche. En ce qui concerne l'allocation optimale, comme les grandes entreprises sont recensées, il se peut que la contrainte ne soit pas saturée ou au contraire que la taille allouée aux grandes entreprises dépasse le nombre de grandes entreprises à recenser. Dans les deux cas, une nouvelle répartition a été déterminée à partir des petites entreprises dans la même position NOGA 3, sans toucher aux tailles des moyennes entreprises. Cette problématique d'une éventuelle réallocation peut être traitée différemment et fera l'objet d'un traitement particulier dans le cadre de l'optimisation du plan d'échantillonnage. Il est important de souligner ici les différences considérables enregistrées entre les deux types d'allocation, et ceci dans toutes les strates. A noter que dans les strates correspondant aux grandes entreprises, et dans une moindre mesure aux moyennes, les tailles de la population sont faibles en regard de celles des petites entreprises, ce qui explique que l'allocation proportionnelle alloue presque toutes les ressources aux petites entreprises.

Activité économique $NOGA3$	Petites entreprises (≤ 15 *emplois*)		Moyennes entreprises		Grandes entreprises (> 45 *emplois*)	
	n_{hp}	n_{ho}	n_{hp}	n_{ho}	n_{hp}	n_{ho}
501	20	14	1	2	0	4
503	12	2	0	1	0	4
504	45	11	0	1	0	3
505	27	36	0	3	0	2
521	129	415	3	68	3	87
522	279	59	14	29	2	50
523	106	31	4	4	1	21
524	995	405	26	270	8	203
525	41	1	0	1	0	5
526	29	20	2	9	1	31
527	48	3	1	1	0	6

Tableau 5 : Allocation proportionnelle (n_{hp}) et optimale (n_{ho}) pour $n = 1800$.

3.3 Comparaison entre un tirage aléatoire simple et stratifié

Ce paragraphe est consacré à l'étude comparative de l'efficacité des différentes méthodes de tirage aléatoire. Il est en effet important de connaître le gain en précision que l'on peut attendre d'un tirage aléatoire stratifié, comme nous proposons de le faire, par rapport à un tirage aléatoire simple. Pour cela, nous avons l'information nécessaire au calcul de l'estimateur du total du chiffre d'affaires en Suisse, que ce soit pour un tirage aléatoire simple ou stratifié. Les formules de calcul pour les estimateurs ainsi que leur variance sont données ci-dessous pour les trois méthodes considérées :

$$\text{Tirage aléa. simple : } \widehat{Y} = N\overline{y},\ V(\widehat{Y}) = N^2\frac{(1-f)}{n}S^2,\ f = \frac{n}{N} \text{ et } S^2 = \frac{\sum_{i=1}^{N}(y_i - \overline{y})^2}{N-1}.$$

$$\text{Tirage aléa. stratifié, alloc. prop. : } \widehat{Y}_{ST} = \sum_{h=1}^{H} N_h\overline{y}_h,\ V(\widehat{Y}_{ST}) = \frac{N(N-n)}{n}\sum_{h=1}^{H}\frac{N_h}{N}S^2_{y_h}.$$

$$\text{Tirage aléa. stratifié, alloc. optimale. : } \widehat{Y}_{ST} = \sum_{h=1}^{H} N_h\overline{y}_h,\ V(\widehat{Y}_{ST}) = \sum_{h=1}^{H}\frac{N_h(N_h-n_h)}{n_h}S^2_{y_h}.$$

Les calculs ont été effectués sur les données nettes et les estimateurs du total obtenus sont respectivement $\widehat{Y} = 86.9$ et $\widehat{Y}_{ST} = 80.8$ **Milliards de francs**.

La précision de l'estimation est fournie par le coefficient de variation cv exprimé en %, où cv est obtenu en divisant l'écart-type de l'estimateur par lui-même. Les résultats obtenus en termes de coefficients de variation (en %) sont donnés dans le Tableau 6.

Taille de l'échantillon	Tirage aléatoire simple	Tirage aléatoire stratifié, allocation proportionnelle	Tirage aléatoire stratifié, allocation optimale de Neyman
n	cv (en %)	cv (en %)	cv (en %)
500	54	41	5.1
1000	38	29	3.6
1500	31	23	2.2
1800	28	21	**1.9**
2000	27	20	1.7
2500	23	18	1.4
3000	21	16	1.3
5000	16	12	**0.9**
10000	11	8	0.6

Tableau 6 : Coefficients de variation du total des chiffres d'affaires en fonction de n.

Ces résultats sont éloquents et montrent le gain en précision d'un tirage aléatoire stratifié, basé sur une allocation optimale par rapport à une allocation proportionnelle et à plus forte raison sur un tirage aléatoire simple. A noter qu'une taille de 1800 pour l'enquête conjoncturelle semble être raisonnable avec un coefficient de variation inférieur à 2% de même qu'une taille de 5000 pour l'enquête structurelle avec un coefficient de variation inférieur à 1%.

4 Optimisation du plan d'échantillonnage

4.1 Recherche de l'estimateur le plus efficace

La précision des estimateurs obtenue à ce stade de l'étude n'est probablement pas optimale. C'est pourquoi il est nécessaire d'envisager d'autres possibilités pour déterminer s'il est possible d'obtenir des estimateurs encore plus précis. La première tentative a été de former des strates différentes. Pour cela, nous avons dans un premier temps modifié les positions $NOGA3$. Celles-ci ont été divisées de manière sensée, notamment au niveau 524 qui contient près de la moitié des données. Ainsi, outre le regroupement des positions 501, 503 et 504, nous avons considéré quatre positions au lieu d'une seule pour la position 524, à savoir un regroupement respectif des positions (5241, 5242, 5243), (5244, 5245, 5246), (5247) et (5248). Les résultats

obtenus dans ce cas ont été décevants dans la mesure où la précision des estimateurs évoqués dans le paragraphe précédent est pratiquement la même que pour les positions $NOGA3$ usuelles. La seconde tentative, qui est aussi liée à la modification des strates a consisté à modifier les tailles limites des entreprises pour leur catégorisation en petites, moyennes et grandes entreprises. Tout en gardant les positions $NOGA$ évoquées ci-dessus, nous avons étudié l'impact d'un changement au niveau des petites entreprises (moins de 10 emplois), des moyennes entreprises (entre 10 et 35 emplois) et finalement des grandes entreprises (plus de 35 emplois). Là encore, les résultats ont été décevants, n'apportant aucune amélioration au niveau de la précision des estimateurs.
Ces tentatives n'ayant pas été fructueuses, il reste comme voie d'étude la recherche d'un estimateur plus efficace que celui de la moyenne stratifiée présentée au paragraphe précédent, en particulier les estimateurs par le quotient. Ainsi, trois nouvelles tentatives ont été menées sur les strates originales en considérant les estimateurs suivants ainsi que leur variance :

$$\text{Estimateur par le quotient : } \widehat{Y}_R = \frac{\overline{y}}{\overline{x}}X,\ V(\widehat{Y}_R) = N^2\frac{(1-f)}{n}\left[\frac{\sum_{i=1}^{N}(y_i - Rx_i)^2}{N-1}\right].$$

$$\text{Estimateur par le quotient séparé : } \widehat{Y}_{RS} = \sum_{h=1}^{H}\frac{\overline{y}_h}{\overline{x}_h}X_h = \sum_{h=1}^{H}R_hX_h, \text{ avec}$$

$$V(\widehat{Y}_{R_s}) = \sum_{h=1}^{H}\frac{N_h(N_h-n_h)}{n_h}S_{d_h}^2, \text{ où } S_{d_h}^2 = \frac{1}{N_h-1}\sum_{i=1}^{N_h}d_{h_i}^2 \text{ et } d_{h_i} = y_{h_i} - R_hx_{h_i}.$$

$$\text{Estimateur par le quotient combiné : } \widehat{Y}_{RC} = \frac{\widehat{Y}_{ST}}{\widehat{X}_{ST}}X,\ V(\widehat{Y}_{RC}) = \sum_{h=1}^{H}\frac{N_h(N_h-n_h)}{n_h}S_{U_h}^2,$$

$$\text{où } S_{U_h}^2 = \frac{1}{N_h-1}\sum_{i=1}^{N_h}(U_{h_i} - \overline{U}_h)^2 = \frac{1}{N_h-1}\sum_{i=1}^{N_h}\left[(y_{h_i} - \overline{y}_h) - R(x_{h_i} - \overline{x}_h)\right]^2.$$

Les calculs ont été faits sur les données nettes à partir des informations qui figurent dans les Tableaux 7 et 8 respectivement.

Activité économique $NOGA3$	Petites entreprises (≤ 15 *emplois*)		Moyennes entreprises		Grandes entreprises (> 45 *emplois*)	
	X_h	R_h	X_h	R_h	X_h	R_h
501	1134	484	367	447	322	213
503	876	282	231	318	279	293
504	2842	232	187	354	282	671
505	2137	785	341	521	294	376
521	8279	310	1583	348	77855	348
522	27757	163	8150	149	4393	198
523	12515	236	1965	232	2105	306
524	66254	223	15557	256	37240	303
525	1655	91	278	88	388	89
526	1851	283	1047	395	4559	379
527	1863	114	373	178	1249	207

Tableau 7 : Valeurs X_h et R_h nécessaires au calcul des estimateurs quotients.

Activité économique $NOGA3$	Petites entreprises (≤ 15 *emplois*)		Moyennes entreprises		Grandes entreprises (> 45 *emplois*)	
	S_{d_h}	S_{U_h}	S_{d_h}	S_{U_h}	S_{d_h}	S_{U_h}
501	1344	1519	6707	6598	15594	15775
503	642	642	3229	3435	4239	4416
504	477	469	10950	10625	.	.
505	2149	2569	8253	8209	55336	61342
521	781	802	4112	4161	82388	97108
522	557	723	4351	4324	21930	21140
523	578	585	2039	2001	21300	20883
524	693	693	4524	4499	32100	34985
525	186	302	3590	4888	9585	4480
526	1102	1108	8511	8590	33654	30067
527	190	274	3264	3145	21521	27869

Tableau 8 : Valeurs S_{d_h} et S_{U_h} nécessaires aux variances des estimateurs quotients.

L'estimateur par le quotient s'obtient à partir de $R = \frac{1937.06}{7.017} = 276$ en multipliant cette valeur par le nombre total d'emplois $X = 286208$. Ainsi $\widehat{Y}_R = 79.0$ Mia. Sa variance a été calculée dans le cas d'un tirage aléatoire simple pour comparer la précision obtenue en considérant un estimateur par le quotient plutôt qu'un estimateur du total obtenu à partir de la moyenne des chiffres d'affaires, comme celui calculé

dans le paragraphe précédent.
Les deux autres estimateurs par le quotient séparé et par le quotient combiné respectivement s'obtiennent à partir des informations des Tableau 7 et 8. On trouve $\widehat{Y}_{RS} = 78.39$ Mia. et $\widehat{Y}_{RC} = 78.54$ Mia. respectivement. La variance de ces estimateurs a été calculée dans le cadre d'un tirage aléatoire stratifié en appliquant l'allocation optimale de Neyman telle que décrite dans le paragraphe précédent en remplaçant les valeurs de S_y par S_d et S_U respectivement.
Les résultats obtenus quant à la précision de ces estimateurs sont donnés sous forme de coefficients de variation (en %) dans le Tableau 9.

Taille de l'échantillon	Tirage aléatoire simple, estimateur par le quotient	Tirage aléatoire stratifié, allocation optimale, estimateur par le quotient séparé	Tirage aléatoire stratifié, allocation optimale, estimateur par le quotient combiné
n	cv (en %)	cv (en %)	cv (en %)
500	16	5.0	5.6
1000	11	1.9	2.2
1500	9	1.4	1.5
1800	8	1.2	**1.3**
2000	8	1.0	1.2
2500	7	0.97	1.0
3000	6	0.87	0.92
5000	5	0.61	**0.66**
10000	3	0.38	0.41

Tableau 9 : Coefficients de variation des estimateurs quotients en fonction de n.

En comparant la précision des estimateurs par le quotient à ceux du paragraphe précédent, il apparaît que ces derniers sont plus efficaces. En effet, l'estimateur par le quotient dans le cas d'un tirage aléatoire simple est nettement plus précis que celui basé sur la moyenne des chiffres d'affaire. Il en va de même dans le cas d'un tirage aléatoire stratifié, avec allocation optimale de Neyman, comme l'indiquent les deux dernières colonnes du Tableau 9. Il est à noter que dans la pratique, il peut être délicat de calculer l'estimateur par le quotient séparé (dans certaines strates le quotient peut ne pas être disponible en raison du problème de la non-réponse, notamment dans les strates avec peu d'observations). La différence entre les estimateurs par le quotient séparé et combiné est minime et il est préférable dans ce cas d'opter pour l'estimateur par le quotient combiné. A noter cette fois une précision tout à fait acceptable pour une taille de 1800 (enquête conjoncturelle) avec un coefficient de variation de 1.3%. De même, avec une taille de 5000 (enquête structurelle), le coefficient de variation n'est que de 0.66%.
Finalement le gain en précision est relativement important en considérant ce type d'estimateur et il semble judicieux de retenir l'estimateur par le quotient combiné

dans le cadre d'un tirage aléatoire stratifié en procédant à une allocation optimale de Neyman. Cette allocation peut encore être raffinée comme le montre le prochain paragraphe.

4.2 Recherche de l'allocation la plus efficace

La manière dont a été traitée l'allocation optimale de Neyman jusqu'ici n'est peut-être pas la plus adéquate. On peut en effet procéder à une itération du processus d'allocation en saturant les contraintes relatives aux grandes entreprises devant être recensées et en réitérant le processus d'allocation optimale aux moyennes et petites entreprises. La différence en termes de précision n'est pas flagrante mais permet tout de même d'obtenir des coefficients de variation légèrement inférieurs à ceux du Tableau 9, soit $cv = 1.2\%$ **pour** $n = 1800$ et $cv = 0.62\%$ **pour** $n = 5000$. L'allocation obtenue dans ce cas est fournie dans le Tableau 10 pour les deux tailles de 1800 et 5000 respectivement.

Activité économique *NOGA*3	Petites entreprises (≤ 15 *emplois*)		Moyennes entreprises		Grandes entreprises (> 45 *emplois*)	
	$n = 1800$	$n = 5000$	$n = 1800$	$n = 5000$	$n = 1800$	$n = 5000$
501	29	96	3	11	4	4
503	7	24	1	4	4	4
504	20	66	3	9	3	3
505	66	219	4	12	2	2
521	97	321	11	35	87	87
522	190	629	57	190	50	50
523	58	193	7	22	21	21
524	648	2148	110	365	203	203
525	12	39	2	6	5	5
526	30	101	14	46	31	31
527	12	41	2	5	6	6

Tableau 10 : Allocation optimale de Neyman pour $n = 1800$ et $n = 5000$.

A noter que dans l'enquête structurelle, bien que l'unité de tirage soit l'entreprise, l'unité d'enquête est l'établissement. Nous disposons de l'information pour déterminer la répartition dans les strates de la population, aussi bien au niveau des entreprises, des entreprises qui ne forment qu'un avec les établissements (Einzelbetrieb) que des établissements. A partir de là, on peut estimer en moyenne le nombre d'établissements qui participeront à l'enquête structurelle lors d'un tirage d'entreprises. Les calculs permettant une telle estimation sont basés sur l'hypothèse que tous les établissements d'une entreprise de l'échantillon participent à l'enquête structurelle (tirage en grappes). L'allocation optimale de Neyman donnée dans le

Tableau 10 permet de connaître les taux de sondage des entreprises dans chaque strate. Il est donc possible d'appliquer ces taux de sondage d'une part aux entreprises=établissements (Einzelbetrieb) ainsi qu'aux établissements liés respectivement aux petites, moyennes et grandes entreprises (de la même $NOGA3$). Pour le cas d'un échantillon d'entreprises de taille $n = 5000$, on trouve ainsi 14690 établissements en moyenne. La précision de l'estimateur par le quotient combiné peut être recalculée en faisant l'hypothèse que la variance des chiffres d'affaire dans les strates se comporte de manière analogue pour les entreprises que pour les établissements. On peut dans ce cas fournir une approximation de la précision en recalculant la variance de l'estimateur par le quotient combiné sur la base des tailles obtenues pour les établissements. On trouve dans ce cas un coefficient de variation $cv = 0.44\%$. Il faut cependant préciser que cette approximation est peut-être légèrement optimiste du fait qu'elle ne tient pas compte de l'effet de grappe (conduisant en général à une estimation un peu moins précise).
Les informations obtenues sur la précision aussi bien au niveau des entreprises que des établissements pour l'enquête structurelle nous fournissent cependant des bornes sur la précision de l'estimateur considéré : en termes de coefficient de variation, il se situe dans la fourchette $\mathbf{0.44\%} \leq cv \leq \mathbf{0.62\%}$.
L'estimation du total des chiffres d'affaires sera donc précise à 1% près environ ($cv \approx 0.5\%$), un gain en précision notable en regard des premiers résultats obtenus avant de procéder à la recherche de l'estimateur le plus efficace.

4.3 Analyse de deux strates supplémentaires

Ce paragraphe contient une analyse de deux strates supplémentaires pour les entreprises dont l'activité principale sont respectivement les codes $NOGA5020a$ (Entretien et réparation de véhicules automobiles) et $8512a$ (pratique médicale).
Pour mener à bien cette analyse complémentaire, il a fallu reprendre les données de la BZ 95 et les différents programmes de traitement des données développés deux ans auparavant lors de l'établissement initial du plan d'échantillonnage. Les tableaux 11 et 12 fournissent les résultats obtenus avec S-plus avant et après traitement des outliers par la méthode l_1 pour les codes $NOGA5020a$ et $8512a$ respectivement.

$NOGA5$ $5020a$	Nombre d'obs.	Res. Stand. Error	Multiple R-Square	coeff. l_2	std. err. l_2	coeff. l_1
	n	RSE	R^2	β_{l_2}	se	β_{l_1}
avec outliers	8155	53949	11.5%	891	27.4	350
sans outliers	8024	3080	82.5%	412	2.1	335

Tableau 11 : Résultats obtenus avant et après traitement des outliers ($NOGA5020a$).

$NOGA5$ 8512a	Nombre d'obs.	Res. Stand. Error	Multiple R-Square	coeff. l_2	std. err. l_2	coeff. l_1
	n	RSE	R^2	β_{l_2}	se	β_{l_1}
avec outliers	6314	17196	0.3%	260	57.3	171
sans outliers	6222	300	79.8%	157	1.0	170

Tableau 12 : Résultats obtenus avant et après traitement des outliers ($NOGA8512a$).

Les résultats des tableaux 11 et 12 montrent de manière éloquente combien un très faible pourcentage d'outliers (2% pour la $NOGA5020a$ et 1% pour la $NOGA8512a$) peut avoir des conséquences désastreuses sur l'estimation obtenue par la méthode des moindres carrés (l_2). On voit ici l'intérêt d'une méthode robuste comme l'estimation l_1 pour traiter du problème des outliers.
Il est désormais possible d'estimer les contributions de ces deux nouvelles strates au chiffre d'affaires total calculé au paragraphe 4.1. L'estimateur par le quotient séparé (qui avait fournit une estimation de 78.39 Mia. de francs sans les nouvelles strates) fournit pour le code $NOGA502$ un chiffre d'affaire de 20.64 Mia. francs (11.08 Mia. pour les petites entreprises, 4.95 Mia. pour les moyennes et 4.61 Mia. pour les grandes). Pour ce qui est du code $NOGA851$, le chiffre d'affaires se monte à 4.101 Mia. francs (3.93 Mia. pour les petites entreprises, 0.077 Mia. pour les moyennes et 0.094 Mia. pour les grandes).
Au total, en considérant ces strates, le chiffre d'affaires total s'élève à 99.03 Mia. avec le code NOGA 502 et à 103.13 Mia. si on rajoute la strate 851.

4.4 Impact des taux de réponse sur la précision des estimateurs

Lors de l'enquête test, il est apparu que les taux de réponse pour les grandes entreprises (qui doivent être recensées) étaient mauvais, se situant en moyenne à environ 40%, et présentaient de ce fait un impact considérable sur la précision attendue. Pour fournir une indication chiffrée de l'impact de ce taux de réponse parmi les grandes entreprises sur la précision du volume du chiffre d'affaires, nous avons simulé différents taux de réponse et calculé la précision correspondante. Pour cela, nous avons considéré un taux de réponse identique dans les différentes strates. Les résultats sont éloquents et apparaissent dans la Figure 12. Nous y avons également inclus le coefficient de variation qu'il serait possible d'obtenir en prenant toutes les petites et moyennes entreprises (environ 55000 entreprises et en gardant un taux de réponse de 40% parmi les grandes entreprises). Ce calcul apparaît comme le dernier point sur le graphique et correspond à la meilleure précision qu'il est possible d'obtenir avec un taux de réponse de seulement 40% parmi les grandes entreprises.

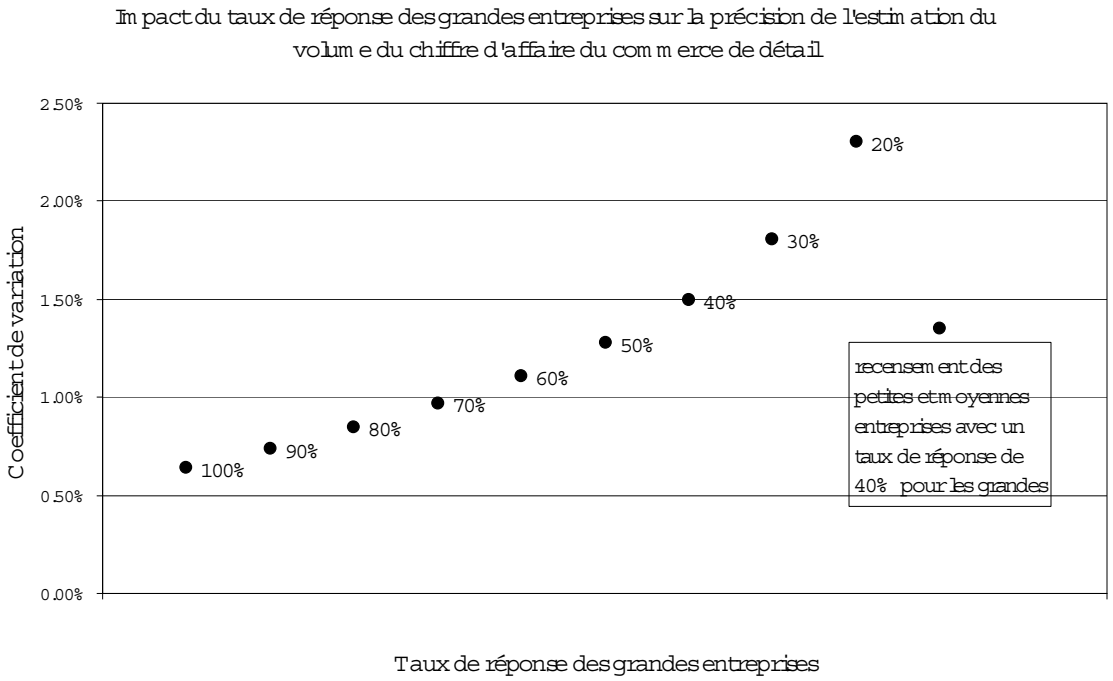


FIG. 12 – Coefficient de variation de l'estimation du volume total des chiffres d'affaires en fonction du taux de réponse des grandes entreprises.

Par rapport au plan d'échantillonnage initial pour lequel l'estimation du coefficient de variation se situait à 0.5% (0.66% au niveau entreprises) il se situe cette fois au même niveau (0.63% au niveau entreprises) en admettant un recensement des grandes entreprises. En tenant compte de la non-réponse des grandes entreprises lors de l'enquête test (non obligatoire) l'estimation ne pourra se faire qu'avec un coefficient de variation d'environ 1.5% (au niveau entreprises). **L'impact est considérable puisque l'on est deux fois moins précis avec un taux de réponse de 40% parmi les grandes entreprises.**
Il convient dès lors de prendre particulièrement soin de la non-réponse parmi les 416 grandes entreprises recensées.
De manière à obtenir une précision raisonnable, ces taux de réponse ne devraient pas passer sous la barre des 2/3 (67%), conduisant ainsi au niveau des entreprises à un coefficient de variation de l'ordre de 1% et au niveau des établissements à environ 0.75% selon les arguments avancés dans le paragraphe du plan d'échantillonnage initial.

4.5 Plan d'échantillonnage définitif : enquête structurelle

Une allocation optimale de Neyman a été faite en tenant compte uniquement de la strate 502. Les résultats apparaissent dans le Tableau 13 avec les taux de réponse (t_{rep}) qui permettent d'estimer les tailles brutes. A noter que grâce aux taux de

réponse relativement bons dans les petites et moyennes entreprises, il est possible de tirer au total dans l'échantillon brut seulement 13941 entreprises (au lieu des 15000 prévues). Avec les taux de réponse obtenus lors de l'enquête test, il faut compter avec 5805 entreprises dans l'échantillon net. **Il apparaît ici que les moyennes entreprises peuvent être recensées**. D'autre part, pour les petites entreprises, nous avons adapté les tailles de manière à ce que les poids initiaux ne dépassent pas 15.

Activité économique $NOGA3$	Petites entreprises (≤ 15 *emplois*)			Moyennes entreprises			Grandes entreprises (> 45 *emplois*)		
	n_h^{net}	t_{rep}	n_h^{brut}	n_h^{net}	t_{rep}	n_h^{brut}	n_h^{net}	t_{rep}	n_h^{brut}
501	96	42%	229	4	0%	4	0	100%	0
502	785	40%	1963	198	39%	422	83	35%	83
503	60	35%	171	5	40%	12	2	0%	2
504	115	48%	240	5	17%	5	3	0%	3
505	219	45%	487	8	33%	8	4	50%	4
521	321	41%	783	35	41%	63	66	50%	66
522	629	38%	1655	190	43%	259	24	35%	24
523	193	53%	364	44	55%	80	20	37%	20
524	2252	42%	5362	273	49%	556	179	48%	179
525	114	40%	285	6	83%	7	4	25%	4
526	101	37%	273	42	58%	42	25	30%	25
527	102	42%	243	5	11%	12	6	33%	6
TOTAL	4987	42%	12055	815	44%	1470	416	41%	416

Tableau 13 : Enquête structurelle : allocation optimale et tailles brutes.

4.6 Plan d'échantillonnage définitif : enquête conjoncturelle

En ce qui concerne la détermination des tailles pour l'enquête conjoncturelle comme sous-échantillon de l'enquête structurelle, il apparaît comme important de recenser les moyennes et grandes entreprises également. Ceci permet en effet d'obtenir un très bon taux de recouvrement. Pour le calcul des tailles dans les strates, la marge de manoeuvre est étroite puisque les moyennes et grandes entreprises totalisent 1886 entreprises et qu'il n'est pas envisageable de tirer beaucoup plus de 4000 entreprises (échantillon brut). La taille totale de l'échantillon net étant fixée à 1800. Par rapport au plan d'échantillonnage initial (pour lequel nous avions calculé un coefficient de variation de l'ordre de 1.3% en recensant les grandes entreprises et en faisant une allocation optimale parmi les petites et moyennes), nous avons adapté les tailles à l'aide des **deux critères** suivants : d'une part, la **taille nette minimale par strate doit être de 30** et d'autre part les **poids initiaux ne doivent pas dépasser 100**

(en appliquant ces deux critères et en procédant à un recensement des moyennes entreprises, le coefficient de variation attendu vaut 1.45% s'écartant légèrement de celle obtenue avec une allocation optimale). Du fait que le taux de recouvrement est plus important et que les critères adoptés permettent d'éviter certains problèmes au moment de l'extrapolation (trop peu d'observations dans les strates, poids initiaux très élevés), il est raisonnable de retenir cette solution.
A noter que toutes les remarques faites à propos de la précision par rapport à la non-réponse des grandes entreprises restent valables ici !
Le Tableau 14 fournit les tailles de la population des entreprises de la DHU reconstituées (à partir du BUR réactualisé en mars 2000), c'est-à-dire du cadre de sondage (N_h). Il contient également les tailles brutes (n_h^{brut}) et nettes par strate (n_h^{net}).

Activité économique $NOGA3$	Petites entreprises (≤ 15 *emplois*)			Moyennes entreprises			Grandes entreprises (> 45 *emplois*)		
	N_h	n_h^{brut}	n_h^{net}	N_h	n_h^{brut}	n_h^{net}	N_h	n_h^{brut}	n_h^{net}
501	785	71	30	4	4	4	0	0	0
502	11854	375	150	422	422	165	83	83	30
503	902	86	30	12	12	5	2	2	2
504	1719	63	30	5	5	1	3	3	3
505	1013	67	30	8	8	3	4	4	2
521	4014	100	41	63	63	26	66	66	33
522	8966	237	90	259	259	112	24	24	9
523	2859	57	30	80	80	44	20	20	8
524	33782	1024	430	556	556	273	179	179	86
525	1697	75	30	7	7	6	4	4	1
526	960	81	30	42	42	25	25	25	8
527	1525	71	30	12	12	2	6	6	2
TOTAL	70076	2307	951	1470	1470	666	416	416	184

Tableau 14 : Enquête conjoncturelle : tailles de la population, brutes et nettes.

5 Forme des estimateurs et estimation de leur variance

5.1 Notations

Introduisons les notations suivantes :

C : Population d'entreprises DHU reconstituées du cadre de sondage ;

G : Ensemble des 11 plus grandes entreprises DHU de C (traitées séparément) ;

$U = C \setminus G$: Population d'entreprises DHU reconstituées, de taille $\mid U \mid = N$;

U_h : Strate (NOGA3 $\times$ taille entreprise), de taille $\mid U_h \mid = N_h$, $h = 1, \ldots, H$;

X_h : Emplois équivalents plein temps (variable auxiliaire connue sur la population) ;

S_h : Echantillon aléatoire simple pour la DHU-S, de taille $\mid S_h \mid = n_{s_h}$;

s_h : Sous-échantillon aléatoire simple ($s_h \subseteq S_h$) pour la DHU-K, de taille $\mid s_h \mid = n_h$;

$r_{h,t} \subseteq s_h$: sous-échantillon des répondants ($y_{k,t} \geq 0$), de taille $\mid r_{h,t} \mid = m_{h,t}$;

t : vague sous revue (mois) pour lequel les entreprises doivent fournir leurs chiffres d'affaires ; mai 2000 $\equiv$ 1,..., **juin 2001** $\equiv$ 14,..., juillet 2002 $\equiv$ 27,... ;

j : groupe de marchandise (**au total 14 groupes dont un "non ventilable"**) ;

k : indice désignant l'entreprise k ;

Ainsi, au niveau de chaque entreprise k, on a :

$$y_{k,t} = \sum_{j=1}^{14} y_{k,j,t} \text{ : Chiffre d'affaires total pour le mois sous revue}$$

$$y_{k,t-1} = \sum_{j=1}^{14} y_{k,j,t-1} \text{ : Chiffre d'affaires total pour le mois précédent}$$

$$y_{k,t-12} = \sum_{j=1}^{14} y_{k,j,t-12} \text{ : Idem pour le même mois de l'année précédente}$$

5.2 Paramètres à estimer

Le but de l'enquête conjoncturelle est d'estimer les variations des volumes des chiffres d'affaires du commerce de détail, y compris au niveau des strates ainsi que des groupes de marchandises. Au niveau du total, il s'agit dans un premier temps d'estimer le volume total pour le mois sous revue ainsi que pour le même mois de l'année précédente.

$$\widehat{Y}_t = \sum_{j=1}^{14} \widehat{Y}_{j,t} \text{ et } \widehat{Y}_{t-12} = \sum_{j=1}^{14} \widehat{Y}_{j,t-12};$$

Les variations par rapport au même mois de l'année précédente s'obtiennent au niveau global par :

$$\widehat{V}_{t,t-12} = \frac{\widehat{Y}_t - \widehat{Y}_{t-12}}{\widehat{Y}_{t-12}} \times 100\%;$$

Et au niveau des groupes de marchandises par :

$$\widehat{V}_{j,t,t-12} = \frac{\widehat{Y}_{j,t} - \widehat{Y}_{j,t-12}}{\widehat{Y}_{j,t-12}} \times 100\%;$$

Estimateur d'un total :

Le volume total :

$$Y_t = \sum_{k \in U} y_{k,t}$$

est estimé par :

$$\begin{aligned}
\widehat{Y}_t &= \sum_{h=1}^{H} \frac{1}{\pi_h} \cdot \frac{1}{\tau_h} \cdot \frac{1}{\rho_{h,t}} \cdot c_{h,t} \cdot y_{h,t} \\
&= \sum_{h=1}^{H} w_{h,t} \cdot c_{h,t} \cdot y_{h,t}, \text{ avec} \\
&= \sum_{h=1}^{H} \sum_{k=1}^{m_{h,t}} w_{h,t} \cdot c_{h,t} \cdot y_{k,t} \text{ , où}
\end{aligned}$$

$\frac{1}{\pi_h} \cdot \frac{1}{\tau_h} = \frac{N_h}{ns_h} \cdot \frac{n_{s_h}}{n_h} = \frac{N_h}{n_h}$: poids de sondage,

$\frac{1}{\rho_{h,t}} = \frac{n_h}{m_{h,t}}$: poids dû à la non-réponse,

$c_{h,t}$: réalise une calibration des poids $w_{h,t}$ sur la variable auxiliaire X_h (ou X).

Trois estimateurs du total ont été envisagés :

– Estimateurs par expansion ($c_{h,t} \equiv 1$), donné par :

$$\begin{aligned} \widehat{Y}_{ST,t} &= \sum_{h=1}^{H} w_{h,t} \cdot y_{h,t}; \\ \widehat{X}_{ST,t} &= \sum_{h=1}^{H} w_{h,t} \cdot x_{h,t}; \end{aligned}$$

– Estimateur par le quotient séparé ($c_{h,t} \equiv \frac{X_h}{w_{h,t} \cdot x_{h,t}}$), donné par :

$$\widehat{Y}_{RS,t} = \sum_{h=1}^{H} \frac{y_{h,t}}{x_{h,t}} \cdot X_h;$$

– Estimateur par le quotient combiné ($c_{h,t} \equiv \frac{X}{\sum_h w_{h,t} \cdot x_{h,t}}$), donné par :

$$\widehat{Y}_{RC,t} = \frac{\widehat{Y}_{ST,t}}{\widehat{X}_{ST,t}} \cdot X.$$

On obtient ainsi les trois estimateurs finaux du total en rajoutant $Y_{G,t}$ aux estimateurs définis ci-dessus, où $Y_{G,t} = \sum_{k \in G} y_{k,t}$ des 11 très grandes entreprises :

$$\begin{aligned} \widehat{Y}_{FST,t} &= \widehat{Y}_{ST,t} + \sum_{k \in G} y_{k,t} \\ \widehat{Y}_{FRS,t} &= \widehat{Y}_{RS,t} + \sum_{k \in G} y_{k,t} \\ \widehat{Y}_{FRC,t} &= \widehat{Y}_{RC,t} + \sum_{k \in G} y_{k,t} \end{aligned}$$

5.3 Précision des paramètres estimés

Les estimations de variance suivantes ont été implémentées :

$$\widehat{VAR}(\widehat{Y}_{ST,t}) = \sum_{h=1}^{H} \frac{N_h(N_h - m_{h,t})}{m_{h,t}} \cdot S^2_{y_{h,t}}$$

et

$$\begin{aligned} \widehat{VAR}(\widehat{Y}_{RS,t}) &= \sum_{h=1}^{H} \frac{N_h(N_h - m_{h,t})}{m_{h,t}} \cdot S^2_{d_{h,t}} \text{ où} \\ S^2_{d_{h,t}} &= \frac{1}{N_h - 1} \sum_k d^2_{h_k} \text{ avec} \\ d_{h_k} &= y_{h_k} - R_h x_{h_k} \text{ et } R_h = \frac{y_{h,t}}{x_{h,t}} \end{aligned}$$

Enfin, les coefficients de variation ont été calculés respectivement par :

$$cv(\widehat{Y}_{ST,t}) = \frac{\sqrt{\widehat{VAR}(\widehat{Y}_{ST,t})}}{\widehat{Y}_{FST,t}} \times 100\%$$

et

$$cv(\widehat{Y}_{RS,t}) = \frac{\sqrt{\widehat{VAR}(\widehat{Y}_{RS,t})}}{\widehat{Y}_{FRS,t}} \times 100\%.$$

A Annexe : prototype du programme d'extrapolation

```
/*                                                                    */
/* Programme SAS permettant de calculer le modèle de pondération pour */
/* l'enquête conjoncturelle comme sous-échantillon de l'enquête structurelle */
/* lancée en mai 2000 et d'extrapoler les résultats de la vague en cours (mois */
/* sous revue) au niveau des volumes des groupes de marchandises      */
/*                                                                    */
/* Auteur : Jean-Pierre Renfer, OFS.                                  */
/* Date de création : 20 février 2001                                 */
/* Dernière modification : 20 mars 2001 à 18h25                       */
/*                                                                    */
options linesize=137 pagesize=41 date number pageno=1;
/*                                                                    */
/* Importation des trois fichiers input qui contiennent respectivement */
/* 1. input_tailles_vzeq_du24-02-01.xls : les informations sur le cadre de */
/* sondage et celles sur les échantillons bruts;                      */
/* 2. untwelle.xls : les informations au niveau entreprise de la vague en cours */
/* 3. warenwelle.xls : les informations pour les groupes de marchandise de la */
/* vague en cours;                                                    */
/*                                                                    */
/*                                                                    */
/* Importation des informations sur : cadre de sondage, échantillons bruts */
/* structurel et conjoncturel;                                        */
/*                                                                    */
PROC IMPORT OUT= SASUSER.TIRAGE
DATAFILE= "T :\21 DHU, Projekte\DHU 2000\Production\
input_tailles_vzeq_du24-02-01.xls"
DBMS=EXCEL5 REPLACE;
SHEET="Feuil1";
RANGE="<ENTIRE SHEET>";
GETNAMES=YES;
RUN;
/*                                                                    */
/* Importation des chiffres d'affaires des entreprises répondantes    */
/*                                                                    */
PROC IMPORT OUT= SASUSER.UNTWELLE
DATAFILE= "T :\21 DHU, Projekte\DHU 2000\Production\untwelle.xls"
DBMS=EXCEL2000 REPLACE;
GETNAMES=YES;
RUN;
/*                                                                    */
/* Importation des informations sur : chiffres d'affaires au niveau groupes de */
```

```
*/ marchandises des entreprises répondantes;                                    */
/*                                                                              */
PROC IMPORT OUT= SASUSER.WARENWELLE
DATAFILE= "T :\21 DHU, Projekte\DHU 2000\Production\warenwelle.xls"
DBMS=EXCEL2000 REPLACE;
GETNAMES=YES;
RUN;
/*                                                                              */
/* Regroupement des strates pour lesquelles les entreprises ont été recensées   */
/* (moyennes et grandes entreprises traitées ensemble);                         */
/* La variable UTAILLE vaut 1 pour les petites entreprises et 2 pour les        */
/* moyennes et grandes entreprises;                                             */
/* La probabilité d'inclusion vaut 1 pour les moyennes et les grandes entreprises */
/* Le taux de réponse est calculé pour l'ensemble des moyennes et grandes       */
/* entreprises par activité économique NOGA3;                                   */
/*                                                                              */
data INFOCADRE;
set SASUSER.TIRAGE;
if UTAILLE=3 then UTAILLE=2;
if UNOGA3=501 AND UTAILLE=2 then UTAILLE=1;
run;
proc sort data=INFOCADRE;
by UNOGA3 UTAILLE;
run;
/*                                                                              */
/* Le data set INFOCADRE a été trié par strate (après regroupement)             */
/* Les variables de ce data set sont les suivantes :                            */
/* NCADRE : nombre d'entreprises dans le cadre de sondage par strate            */
/* VZEQCADR : Emplois équivalent plein temps correspondant dans le cadre        */
/* NSBRUT : Nombre d'observations par strate de l'échantillon brut structurel   */
/* VZEQSBRU : Emplois équivalent plein temps correspondant de l'échantillon     */
/* brut structurel                                                              */
/* NKBRUT : Nombre d'entreprises par strate de l'échantillon brut               */
/* conjoncturel;                                                                */
/* VZEQKBRU : Emplois équivalent plein temps correspondant dans l'échantillon*/
/* brut conjoncturel                                                            */
/*                                                                              */
proc univariate noprint data=INFOCADRE;
var NCADRE VZEQCADR NSBRUT VZEQSBRU NKBRUT VZEQKBRU;
output out=INFOTAILLEVZEQ
sum=Nbcadre AVZEQcad NbSbrut AVZEQSbr NbKbrut AVZEQKbr;
by UNOGA3 UTAILLE;
run;
/*                                                                              */
/* Le data set INFOTAILLEVZEQ contient les informations de INFOCADRE            */
```

```
/* regroupées selon les nouvelles strates                                    */
/* Il y a 24 strates (12 activités économiques et 2 classes de taille : petites et  */
/* (moyennes+grandes))                                                       */
/* Les variables de ce data set sont les suivantes :                         */
/* Nbcadre : nombre d'entreprises dans le cadre de sondage par strate        */
/* AVZEQcad : Emplois équivalent plein temps correspondant dans le cadre     */
/* NbSbrut : Nombre d'observations par strate de l'échantillon brut structurel  */
/* AVZEQSbr : Emplois équivalent plein temps correspondant dans l'échantillon */
/* brut structurel                                                           */
/* NbKbrut : Nombre d'observations par strate de l'échantillon brut          */
/* conjoncturel ;                                                            */
/* AVZEQKbr : Emplois équivalent plein temps correspondant dans l'échantillon */
/* brut conjoncturel                                                         */
/*                                                                           */
/*                                                                           */
/* Data set donnees_base à partir du fichier input untwelle                  */
/*                                                                           */
data donnees_base ;
set SASUSER.UNTWELLE ;
NOGA3=substr(noga_stufe5_orig,1,3) ;
if NOGA3='501' then UNOGA3=501 ;
if NOGA3='502' then UNOGA3=502 ;
if NOGA3='503' then UNOGA3=503 ;
if NOGA3='504' then UNOGA3=504 ;
if NOGA3='505' then UNOGA3=505 ;
if NOGA3='521' then UNOGA3=521 ;
if NOGA3='522' then UNOGA3=522 ;
if NOGA3='523' then UNOGA3=523 ;
if NOGA3='524' then UNOGA3=524 ;
if NOGA3='525' then UNOGA3=525 ;
if NOGA3='526' then UNOGA3=526 ;
if NOGA3='527' then UNOGA3=527 ;
UTAILLE=groessenklasse_id ;
run ;
/*                                                                           */
/* Indication de l'année et de la vague traitée                              */
/*                                                                           */
proc univariate noprint data=donnees_base ;
var JAHR MONAT ;
output out=ANNEEVAGUE mean=ANNEE MOIS ;
run ;
proc print data=ANNEEVAGUE ;
title 'ATTENTION : Résultats pour la vague et l"année :' ;
run ;
/*                                                                           */
```

```
/* On traite à part les 11 plus grandes entreprises (plus de Chf. 50 Millions      */
/* mensuellement, exception faite de l'entreprise 3364 car non répondante         */
/* sur plusieurs mois. Ces 11 très grandes entreprises doivent impérativement     */
/* faire partie des données relevées pour le mois sous revue, les identifiants    */
/* (variable unternehmen_id) sont les suivants : 3348, 3269, 3322, 3368, 3373,    */
/* 3384, 3245, 3328, 3701, 3262 et 3334 ;                                         */
/*                                                                                */
data tgrandes ;
set donnees_base ;
where unternehmen_id=3348 OR unternehmen_id=3269 OR
unternehmen_id=3322 OR unternehmen_id=3368 OR unternehmen_id=3373 OR
unternehmen_id=3384 OR unternehmen_id=3245 OR unternehmen_id=3328 OR
unternehmen_id=3701 OR unternehmen_id=3262 OR unternehmen_id=3334 ;
run ;
/*                                                                                */
/* Le data set tgrandes contient les chiffres d'affaires de ces 11 très grandes   */
/* entreprises au niveau du total du chiffre d'affaires pour le mois sous revue et
*/
/* pour le même mois de l'année précédente ;                                      */
/* Le data set restgrandes contient respectivement pour la variable               */
/* UMSATZ_TOTAL et UMSATZ_TOT_VORJAHR, le nombre de                               */
/* répondants et la somme des chiffres d'affaires de ces 11 très grandes          */
/* entreprises (tg_umsatz_total et tg_umsatz_tot_vorjahr respectivement).         */
/*                                                                                */
proc univariate noprint data=tgrandes ;
var UMSATZ_TOTAL UMSATZ_TOT_VORJAHR ;
output out=restgrandes N=nbumstot nbumstotvj
sum=tg_umsatz_total tg_umsatz_tot_vorjahr ;
run ;
/*                                                                                */
/* Le data set res11entreprises contient les résultats de ces 11 très grandes     */
/* entreprises au niveau du nombre de répondants sur chacune des deux             */
/* variables volume total du chiffre d'affaires du mois sous revue et du même     */
/* mois de l'année précédente ainsi que les variations correspondantes            */
/* exprimées en pourcent : ces résultats sont ensuite imprimés avec le titre      */
/* correspondant.                                                                 */
/*                                                                                */
data res11entreprises ;
set restgrandes ;
aidemerge=1 ;
tg_variation_pourcent=
((tg_umsatz_total-tg_umsatz_tot_vorjahr)/tg_umsatz_tot_vorjahr)*100 ;
run ;
proc print data=res11entreprises ;
var nbumstot nbumstotvj tg_umsatz_total
```

```
tg_umsatz_tot_vorjahr tg_variation_pourcent;
title 'Très grandes entreprises (impérativement 11!!!) :
Volumes et variation de leurs chiffres d"affaire';
run;
/*                                                                    */
/* Partie B : Calcul au niveau des strates des chiffres d'affaires des 11 très  */
/* grandes entreprises : la comparaison ne se fera pas directement avec la  */
/* partie extrapoléee puisque seules trois activités économiques sont représentées */
/*                                                                    */
data tgrandespourstra;
set tgrandes;
NOGA3=substr(noga_stufe5_orig,1,3);
if NOGA3='501' then UNOGA3=501;
if NOGA3='502' then UNOGA3=502;
if NOGA3='503' then UNOGA3=503;
if NOGA3='504' then UNOGA3=504;
if NOGA3='505' then UNOGA3=505;
if NOGA3='521' then UNOGA3=521;
if NOGA3='522' then UNOGA3=522;
if NOGA3='523' then UNOGA3=523;
if NOGA3='524' then UNOGA3=524;
if NOGA3='525' then UNOGA3=525;
if NOGA3='526' then UNOGA3=526;
if NOGA3='527' then UNOGA3=527;
UTAILLE=groessenklasse_id;
run;
proc sort data=tgrandespourstra;
by UNOGA3;
run;
proc univariate noprint data=tgrandespourstra;
var UMSATZ_TOTAL UMSATZ_TOT_VORJAHR;
output out=restgrandesstra N=nbumstotstra nbumstotvjstra
sum=tg_umsatz_total_stra tg_umsatz_tot_vorjahr_stra;
by UNOGA3 UTAILLE;
run;
data restgrandestra;
set restgrandesstra;
aidemerge=1;
run;
data res11tgstrate;
set restgrandestra;
tg_variation_pourcent_stra=
((tg_umsatz_total_stra-tg_umsatz_tot_vorjahr_stra)/
tg_umsatz_tot_vorjahr_stra)*100;
run;
```

```
title 'Très grandes entreprises :
Volumes et variation de leurs chiffres d"affaire par strate';
proc print data=res11tgstrate;
var UNOGA3 UTAILLE nbumstotstra nbumstotvjstra tg_umsatz_total_stra
tg_umsatz_tot_vorjahr_stra tg_variation_pourcent_stra;
run;
/*                                                                  */
/* Le data set donnees_utilisees contient les entreprises répondantes sauf les 11  */
/* très grandes. On procède au regroupement des strates recensées, comme  */
/* pour la population. On ne garde comme répondants que les entreprises ayant  */
/* fourni un chiffre d'affaires non négatif que ce soit pour la variable  */
/* UMSATZ_TOTAL ou UMSATZ_TOT_VORJAHR.                              */
/*                                                                  */
data donnees_utilisees;
set donnees_base;
if unternehmen_id=3348 then delete;
if unternehmen_id=3269 then delete;
if unternehmen_id=3322 then delete;
if unternehmen_id=3368 then delete;
if unternehmen_id=3373 then delete;
if unternehmen_id=3384 then delete;
if unternehmen_id=3245 then delete;
if unternehmen_id=3328 then delete;
if unternehmen_id=3701 then delete;
if unternehmen_id=3262 then delete;
if unternehmen_id=3334 then delete;
if (UMSATZ_TOTAL<0 OR UMSATZ_TOTAL=.) AND
(UMSATZ_TOT_VORJAHR<0 OR UMSATZ_TOT_VORJAHR=.) then delete;
if UTAILLE=3 then UTAILLE=2;
if UNOGA3=501 AND UTAILLE=2 then UTAILLE=1;
run;
/*                                                                  */
/* On traite le reste des entreprises répondantes sauf les 11 très grandes  */
/* pour obtenir les statistiques sur la partie échantillon uniquement  */
/* (celle qui sera extrapolée).                                     */
/*                                                                  */
/*                                                                  */
/* Le data set statuntwelle contient les résultats de la partie échantillon sans les  */
/* 11 très grandes avec respectivement pour la variable UMSATZ_TOTAL et  */
/* UMSATZ_TOT_VORJAHR le nombre d'entreprises répondantes ainsi que  */
/* la somme des chiffres d'affaires pour le mois sous revue et le même mois de  */
/* l'année précédente (pmg_umsatz_total respectivement  */
/* pmg_umsatz_tot_vorjahr).                                         */
/*                                                                  */
proc univariate noprint data=donnees_utilisees;
```

```
var UMSATZ_TOTAL UMSATZ_TOT_VORJAHR;
output out=statuntwelle N=rep_umsatz rep_umsatz_vorjahr
sum=pmg_umsatz_total pmg_umsatz_tot_vorjahr;
run;
data respmgentreprises;
set statuntwelle;
aidemerge=1;
pmg_variation_pourcent=
((pmg_umsatz_total-pmg_umsatz_tot_vorjahr)/pmg_umsatz_tot_vorjahr)*100;
run;
proc print data=respmgentreprises;
var rep_umsatz rep_umsatz_vorjahr
pmg_umsatz_total pmg_umsatz_tot_vorjahr
pmg_variation_pourcent;
title 'Reste de l"échantillon (sans les 11 très grandes) :
Volumes et variation de leurs chiffres d"affaire';
run;
/*                                                                    */
/* Le data set resechantillon contient les résultats sur tout l'échantillon y  */
/* compris les 11 très grandes avec la somme des chiffres d'affaires pour le   */
/* mois sous revue et le même mois de l'année précédente (ech_umsatz_total     */
/*respectivement ech_umsatz_tot_vorjahr).                                      */
/*                                                                    */
data resechantillon;
merge res11entreprises respmgentreprises;
by aidemerge;
run;
data resechentreprises;
set resechantillon;
nbrep_umsatz=nbumstot+rep_umsatz;
nbrep_umsatz_vorjahr=nbumstotvj+rep_umsatz_vorjahr;
ech_umsatz_total=tg_umsatz_total+pmg_umsatz_total;
ech_umsatz_tot_vorjahr=tg_umsatz_tot_vorjahr+pmg_umsatz_tot_vorjahr;
ech_variation_pourcent=
((ech_umsatz_total-ech_umsatz_tot_vorjahr)/ech_umsatz_tot_vorjahr)*100;
run;
proc print data=resechentreprises;
var nbrep_umsatz nbrep_umsatz_vorjahr ech_umsatz_total
ech_umsatz_tot_vorjahr ech_variation_pourcent;
title 'Ensemble de l"échantillon (y compris les 11 très grandes) :
Volumes et variation de leurs chiffres d"affaire';
run;
/*                                                                    */
/* Résultats extrapolés pour la vague en cours                        */
/*                                                                    */
```

```
/*                                                                    */
/* Calculs au niveau du chiffre d'affaires total de l'entreprise      */
/*                                                                    */
/*                                                                    */
/* On traite séparément les données sur la variable UMSATZ_TOTAL>=0   */
/* et celles sur UMSATZ_TOT_VORJAHR>=0                                */
/*                                                                    */
/*                                                                    */
/* Traitement des données sur la variable UMSATZ_TOTAL>=0             */
/*                                                                    */
data donnees;
set donnees_utilisees;
where UMSATZ_TOTAL>=0;
run;
proc sort data=donnees;
by UNOGA3 UTAILLE;
run;
/*                                                                    */
/* Calcul par strate du nombre de répondants (stra_rep_umsatz), du total */
/* du chiffre d'affaires des répondants (stra_pmg_umsatz_total) et du nombre */
/* correspondant d'emplois équivalents plein temps du BUR             */
/* (stra_pmg_bur_vzeq).                                               */
/*                                                                    */
proc univariate noprint data=donnees;
var UMSATZ_TOTAL BUR_VOLLZEITSTELLEN;
output out=input_donnees N=stra_rep_umsatz stra_rep_bur_vzeq
sum=stra_pmg_umsatz_total stra_pmg_bur_vzeq;
by UNOGA3 UTAILLE;
run;
/*                                                                    */
/* Obtention par strate des données permettant de calculer la pondération : */
/* tailles par strate du cadre de sondage, des échantillons bruts structurel et */
/* conjoncturel, des répondants ainsi que l'emploi équivalent plein temps */
/* correspondant.                                                     */
/*                                                                    */
data pour_poids;
merge INFOTAILLEVZEQ input_donnees;
by UNOGA3 UTAILLE;
run;
/*                                                                    */
/* Calcul effectif du modèle de pondération avec probabilités d'inclusion dans */
/* l'échantillon brut structurel, dans l'échantillon brut conjoncturel et taux de */
/* réponse. Distinction entre le poids (inverse de la probabilité d'inclusion) de */
/* tirage et celui de la non-réponse (le poids définitif étant obtenu par le */
/* produit du poids de tirage et celui dû à la non-réponse). Un estimateur par */
```

```
/* le quotient séparé est calculé ici (est_quotient_sep). L'estimateur par      */
/* expansion est calculé également (umsatz_pondere).                            */
/*                                                                              */
data resultats;
set pour_poids;
PROBINCS=NBSBRUT/NBCADRE;
PROBINCKS=NBKBRUT/NBSBRUT;
PROBINCK=PROBINCS*PROBINCKS;
POIDSTIRAGE=1/PROBINCK;
POIDSNRUMSATZ=NBKBRUT/stra_rep_umsatz;
poids_umsatz=POIDSTIRAGE*POIDSNRUMSATZ;
umsatz_pondere=poids_umsatz*stra_pmg_umsatz_total;
vzeq_pondere=poids_umsatz*stra_pmg_bur_vzeq;
stra_quotient=(stra_pmg_umsatz_total/stra_pmg_bur_vzeq);
est_quotient_sep=stra_quotient*AVZEQCAD;
run;
/* */
/* Calcul de la variance de l'estimateur par le quotient séparé. Pour cela il faut */
/* reprendre les données de base pour calculer les di2.                         */
/*                                                                              */
data pourvar;
merge resultats donnees;
by UNOGA3 UTAILLE;
if stra_rep_umsatz=. then delete;
di2=(UMSATZ_TOTAL-stra_quotient*BUR_VOLLZEITSTELLEN)*
(UMSATZ_TOTAL-stra_quotient*BUR_VOLLZEITSTELLEN);
run;
/*                                                                              */
/* Calcul par strate de la somme des résidus au carré et de la variance de la   */
/* variable UMSATZ_TOTAL.                                                       */
/*                                                                              */
proc univariate noprint data=pourvar;
var di2 UMSATZ_TOTAL;
output out=sommedi2 sum=sumdi2 sum_umsatz_total var=varnimpdi2 varum-
satz;
by UNOGA3 UTAILLE;
run;
/* */
/* Calcul (sans tenir compte des 11 très grandes entreprises) des coefficients de */
/* variation par strate pour l'estimateur par expansion et par le quotient séparé. */
/*                                                                              */
data resvariance;
merge sommedi2 resultats;
by UNOGA3 UTAILLE;
varYST=((Nbcadre*(Nbcadre-stra_rep_umsatz))/
```

```
(stra_rep_umsatz))*varumsatz;
varYRS=
((Nbcadre*(Nbcadre-stra_rep_umsatz))/
(stra_rep_umsatz*(stra_rep_umsatz-1)))*sumdi2;
cvYST=(sqrt(varYST)/umsatz_pondere)*100;
cvYRS=(sqrt(varYRS)/est_quotient_sep)*100;
aidemerge=1;
run;
title 'Résultats par strate pour la pondération avec la variable UMSATZ_TOTAL';
proc print data=resvariance;
var UNOGA3 UTAILLE NBCADRE NBSBRUT NBKBRUT
stra_rep_umsatz POIDSTIRAGE POIDSNRUMSATZ poids_umsatz;
run;
title 'Résultats par strate pour l"extrapolation avec la variable UMSATZ_TOTAL';
proc print data=resvariance;
var UNOGA3 UTAILLE stra_pmg_umsatz_total umsatz_pondere
stra_pmg_bur_vzeq vzeq_pondere stra_quotient est_quotient_sep cvYRS;
run;
/*                                                                  */
/* Calcul pour l'extrapolation avec l'estimateur par le quotient combiné  */
/* (solution retenue). Ici X correspond à la somme totale des équivalents */
/* plein temps du BUR (sans les 11 très grandes entreprises), Yst   */
/* correspond à l'estimateur par expansion pour le chiffre d'affaires total */
/* extrapolé, Xst correspond à la somme pondérée des équivalents plein */
/* temps pour l'ensemble de l'échantillon et YRs est l'estimateur par le */
/* quotient séparé du total des chiffres d'affaires.                */
/*                                                                  */
proc univariate noprint data=resvariance;
var AVZEQCAD umsatz_pondere vzeq_pondere est_quotient_sep
varYST varYRS;
output out=resech sum=X Yst Xst YRs varianceYST varianceYRS;
run;
/*                                                                  */
/* Partie B : Calcul de l'estimateur par le quotient combiné proprement dit */
/* dans chaque strate pour établir une comparaison des résultats par strate. */
/*                                                                  */
data resechpmerge;
set resech;
aidemerge=1;
run;
data resultechstra;
merge resvariance resechpmerge;
by aidemerge;
YRcstra=(umsatz_pondere/Xst)*X;
aidemerge=1;
```

```
run;
/*                                                                    */
/* Calcul de l'estimateur par le quotient combiné proprement dit. Calcul de  */
/* l'écart-type du total pour l'estimateur par expansion et pour l'estimateur */
/* par le quotient séparé.                                            */
/*                                                                    */
data resultech;
set resech;
     YRc=(Yst/Xst)*X;
     stdYST=sqrt(varianceYST);
     stdYRS=sqrt(varianceYRS);
     aidemerge=1;
     run;
/*                                                                    */
/* Traitement des données sur la variable UMSATZ_TOT_VORJAHR>=0       */
/*                                                                    */
data donnees_vorjahr;
set donnees_utilisees;
where UMSATZ_TOT_VORJAHR>=0;
run;
proc sort data=donnees_vorjahr;
by UNOGA3 UTAILLE;
run;
/* */
/* Calcul par strate du nombre de répondants (stra_rep_umsatz_vorjahr),  */
/* du total du chiffre d'affaires des répondants                         */
/* (stra_pmg_umsatz_tot_vorjahr) et du nombre correspondant d'emplois    */
/* équivalents plein temps du BUR                                        */
/* (stra_pmg_bur_vzeq_vorjahr)                                           */
/*                                                                       */
proc univariate noprint data=donnees_vorjahr;
var UMSATZ_TOT_VORJAHR BUR_VOLLZEITSTELLEN;
output out=input_donnees_vorjahr N=stra_rep_umsatz_vorjahr
stra_rep_bur_vzeq_vorjahr sum=stra_pmg_umsatz_tot_vorjahr
stra_pmg_bur_vzeq_vorjahr;
by UNOGA3 UTAILLE;
run;
/* */
/* Obtention par strate des données permettant de calculer la pondération : */
/* tailles par strate du cadre de sondage, des échantillons bruts structurel */
/* et conjoncturel, des répondants ainsi que l'emploi équivalent plein temps */
/* correspondant.                                                        */
/*                                                                       */
data pour_poids_vorjahr;
merge INFOTAILLEVZEQ input_donnees_vorjahr;
```

```
by UNOGA3 UTAILLE;
run;
/*                                                                    */
/* Calcul effectif du modèle de pondération avec probabilités d'inclusion dans   */
/* l'échantillon brut structurel, dans l'échantillon brut conjoncturel et taux de  */
/* réponse. Distinction entre le poids (inverse de la probabilité d'inclusion) de  */
/* tirage et celui de la non-réponse (le poids définitif étant obtenu par le      */
/* produit du poids de tirage et celui dû à la non-réponse). Un estimateur par    */
/* le quotient séparé est calculé ici (est_quotient_sep_vorjahr). L'estimateur    */
/* par expansion est calculé également (umsatz_vorjahr_pondere).                  */
/*                                                                    */
data resultats_vorjahr;
set pour_poids_vorjahr;
     PROBINCS=NBSBRUT/NBCADRE;
     PROBINCKS=NBKBRUT/NBSBRUT;
     PROBINCK=PROBINCS*PROBINCKS;
     POIDSTIRAGE=1/PROBINCK;
     POIDSNRUMSATZ_VORJAHR=NBKBRUT/stra_rep_umsatz_vorjahr;
     poids_umsatz_vorjahr=POIDSTIRAGE*POIDSNRUMSATZ_VORJAHR;
     umsatz_vorjahr_pondere=
     poids_umsatz_vorjahr*stra_pmg_umsatz_tot_vorjahr;
     vzeq_vorjahr_pondere=
     poids_umsatz_vorjahr*stra_pmg_bur_vzeq_vorjahr;
     stra_quotient_vorjahr=
     (stra_pmg_umsatz_tot_vorjahr/stra_pmg_bur_vzeq_vorjahr);
     est_quotient_sep_vorjahr=stra_quotient_vorjahr*AVZEQCAD;
run;
/*                                                                    */
/* Calcul de la variance de l'estimateur par le quotient séparé       */
/* (variable UMSATZ_TOT_VORJAHR). Pour cela il faut reprendre les     */
/* données de base pour calculer les di2.                             */
/*                                                                    */
data pourvar_vorjahr;
merge resultats_vorjahr donnees_vorjahr;
by UNOGA3 UTAILLE;
if stra_rep_umsatz_vorjahr=. then delete;
di2_vorjahr=
(UMSATZ_TOT_VORJAHR-stra_quotient_vorjahr*
BUR_VOLLZEITSTELLEN)*
(UMSATZ_TOT_VORJAHR-stra_quotient_vorjahr*
BUR_VOLLZEITSTELLEN);
run;
/*                                                                    */
/* Calcul par strate de la somme des résidus au carré                 */
/* et de la variance de la variable UMSATZ_TOT_VORJAHR.               */
```

```
/*                                                                    */
proc univariate noprint data=pourvar_vorjahr;
var di2_vorjahr UMSATZ_TOT_VORJAHR;
output out=sommedi2_vorjahr sum=sumdi2_vorjahr sum_umsatz_tot_vorjahr
var=varnimpdi2_vorjahr varumsatz_vorjahr;
by UNOGA3 UTAILLE;
run;
/*                                                                    */
/* Calcul (sans tenir compte des 11 très grandes entreprises) des coefficients de  */
/* variation par strate pour l'estimateur par expansion et par le quotient séparé  */
/*                                                                    */
data resvariance_vorjahr;
merge sommedi2_vorjahr resultats_vorjahr;
by UNOGA3 UTAILLE;
varYST_vorjahr=((Nbcadre*(Nbcadre-stra_rep_umsatz_vorjahr))/
(stra_rep_umsatz_vorjahr))*varumsatz_vorjahr;
varYRS_vorjahr=((Nbcadre*(Nbcadre-stra_rep_umsatz_vorjahr))/
(stra_rep_umsatz_vorjahr*(stra_rep_umsatz_vorjahr-1)))*sumdi2_vorjahr;
cvYST_vorjahr=(sqrt(varYST_vorjahr)/umsatz_vorjahr_pondere)*100;
cvYRS_vorjahr=(sqrt(varYRS_vorjahr)/est_quotient_sep_vorjahr)*100;
aidemerge=1;
run;
title 'Résultats par strate pour la pondération
avec la variable UMSATZ_TOT_VORJAHR';
proc print data=resvariance_vorjahr;
var UNOGA3 UTAILLE NBCADRE NBSBRUT NBKBRUT
stra_rep_umsatz_vorjahr POIDSTIRAGE
POIDSNRUMSATZ_VORJAHR poids_umsatz_vorjahr;
run;
title 'Résultats par strate pour l"extrapolation
avec la variable UMSATZ_TOT_VORJAHR';
proc print data=resvariance_vorjahr;
var UNOGA3 UTAILLE stra_pmg_umsatz_tot_vorjahr umsatz_vorjahr_pondere
stra_pmg_bur_vzeq_vorjahr vzeq_vorjahr_pondere
stra_quotient_vorjahr est_quotient_sep_vorjahr cvYRS_vorjahr;
run;
/*                                                                    */
/* Calcul pour l'extrapolation avec l'estimateur par le quotient combiné          */
/* (solution retenue). Ici X=X_vorjahr correspond à la somme totale des           */
/* équivalents plein temps du BUR (sans les 11 très grandes entreprises),         */
/* Yst_vorjahr correspond à l'estimateur par expansion pour le chiffre            */
/* d'affaire du total extrapolé du même mois de l'année précédente,               */
/* Xst_vorjahr correspond à la somme pondérée des équivalents plein               */
/* temps pour l'ensemble de l'échantillon répondant pour la variable              */
/* UMSATZ_TOT_VORJAHR et YRs_vorjahr est l'estimateur par le                      */
```

```
/* quotient séparé du total des chiffres d'affaires pour cette variable.      */
/*                                                                            */
proc univariate noprint data=resvariance_vorjahr;
var AVZEQCAD umsatz_vorjahr_pondere vzeq_vorjahr_pondere
est_quotient_sep_vorjahr varYST_vorjahr varYRS_vorjahr;
output out=resech_vorjahr sum=X_vorjahr Yst_vorjahr Xst_vorjahr
YRs_vorjahr varianceYST_vorjahr varianceYRS_vorjahr;
run;
/*                                                                            */
/* Partie B : Calcul pour la variable portant sur le chiffre d'affaires du même */
/* mois de l'année précédente de l'estimateur par le quotient combiné         */
/* proprement dit dans chaque strate pour comparer les résultats par strate.  */
/*                                                                            */
data resechpmerge_vorjahr;
set resech_vorjahr;
aidemerge=1;
run;
data resultechstra_vorjahr;
merge resvariance_vorjahr resechpmerge_vorjahr;
by aidemerge;
YRcstra_vorjahr=(umsatz_vorjahr_pondere/Xst_vorjahr)*X_vorjahr;
aidemerge=1;
run;
/*                                                                            */
/* Calcul de l'estimateur par le quotient combiné proprement dit. Calcul de   */
/* l'écart-type du total pour l'estimateur par expansion et pour l'estimateur */
/* par le quotient séparé pour la variable UMSATZ_TOT_VORJAHR                 */
/*                                                                            */
data resultech_vorjahr;
set resech_vorjahr;
     YRc_vorjahr=(Yst_vorjahr/Xst_vorjahr)*X_vorjahr;
     stdYST_vorjahr=sqrt(varianceYST_vorjahr);
stdYRS_vorjahr=sqrt(varianceYRS_vorjahr);
aidemerge=1;
     run;
/*                                                                            */
/* Résultats finaux sur la base des trois méthodes d'estimation : estimateur par */
/* expansion, par le quotient séparé et par le quotient combiné du volume total */
/* de la partie extrapolée (sans les 11 très grandes entreprises : les résultats */
/* sont regroupés por la variable UMSATZ_TOTAL,                               */
/* UMSATZ_TOT_VORJAHR ainsi que pour les 11 très grandes entreprises.         */
/* On calcule ensuite les résultats finaux en additionnant les résultats      */
/* extrapolés aux résultats des 11 très grandes entreprises. Les coefficients de */
/* variation au niveau du total sont calculés en divisant l'écart-type de     */
/* l'estimateur par la somme de l'estimateur et du volume du chiffre d'affaire */
```

```
/* des 11 très grandes entreprises.                                    */
/*                                                                     */
/*                                                                     */
/* Partie B : résultats finaux par strate                              */
/*                                                                     */
data resultfinauxstra;
     merge resultechstra resultechstra_vorjahr restgrandestra;
by UNOGA3 UTAILLE;
if YRcstra=. then YRcstra=0;
if YRcstra_vorjahr=. then YRcstra_vorjahr=0;
if tg_umsatz_total_stra=. then tg_umsatz_total_stra=0;
if tg_umsatz_tot_vorjahr_stra=. then tg_umsatz_tot_vorjahr_stra=0;
Variationechstra=.;
Variationpopstra=.;
Variation11tgstra=.;
Variationtotstra=.;
YFRCstra=YRcstra+tg_umsatz_total_stra;
YFRCstra_vorjahr=YRcstra_vorjahr+tg_umsatz_tot_vorjahr_stra;
if stra_pmg_umsatz_tot_vorjahr<>0 then
Variationechstra=((stra_pmg_umsatz_total-stra_pmg_umsatz_tot_vorjahr)/
stra_pmg_umsatz_tot_vorjahr)*100;
if YRcstra_vorjahr<>0 then Variationpopstra=((YRcstra-YRcstra_vorjahr)/
YRcstra_vorjahr)*100;
if tg_umsatz_total_stra<>0 then
Variation11tg=((tg_umsatz_total_stra-tg_umsatz_tot_vorjahr_stra)/
tg_umsatz_tot_vorjahr_stra)*100;
if YFRCstra_vorjahr<>0 then
Variationtotstra=((YFRCstra-YFRCstra_vorjahr)/YFRCstra_vorjahr)*100;
run;
title 'Résultats par strate avec l"estimateur
par le quotient combiné pour l"échantillon';
proc print data=resultfinauxstra;
var UNOGA3 UTAILLE stra_rep_umsatz stra_pmg_umsatz_total
stra_rep_umsatz_vorjahr stra_pmg_umsatz_tot_vorjahr Variationechstra
tg_umsatz_total_stra tg_umsatz_tot_vorjahr_stra Variation11tg;
run;
title 'Résultats par strate avec l"estimateur
par le quotient combiné pour l"extrapolation';
proc print data=resultfinauxstra;
var UNOGA3 UTAILLE YRcstra YRcstra_vorjahr
Variationpopstra YFRCstra YFRCstra_vorjahr Variationtotstra;
run;
title 'Résumé : Résultats par strate des variations';
proc print data=resultfinauxstra;
var UNOGA3 UTAILLE stra_rep_umsatz stra_rep_umsatz_vorjahr
```

```
Variationechstra Variationpopstra Variation11tg Variationtotstra;
run;
/*                                                                    */
/* Partie B : résultats finaux par activité économique                */
/*                                                                    */
proc sort data=resultfinauxstra;
by UNOGA3;
run;
proc univariate noprint data=resultfinauxstra;
var stra_rep_umsatz stra_rep_umsatz_vorjahr stra_pmg_umsatz_total
stra_pmg_umsatz_tot_vorjahr YRcstra YRcstra_vorjahr tg_umsatz_total_stra
tg_umsatz_tot_vorjahr_stra YFRCstra YFRCstra_vorjahr;
output out=resNOGA3 sum=sum_rep_umsatz_NOGA3
sum_rep_umsatz_NOGA3_vorjahr sumechNOGA3 sumechNOGA3_vorjahr
sumNOGA3pop sumNOGA3pop_vorjahr sumNOGA311tg
sumNOGA311tg_vorjahr sumtotNOGA3 sumtotNOGA3_vorjahr;
by UNOGA3;
run;
data resultfinauxNOGA3;
set resNOGA3;
VariationechNOGA3=.;
VariationpopNOGA3=.;
Variation11tgNOGA3=.;
VariationtotNOGA3=.;
if sumechNOGA3_vorjahr<>0 then VariationechNOGA3=
((sumechNOGA3-sumechNOGA3_vorjahr)/sumechNOGA3_vorjahr)*100;
if sumNOGA3pop_vorjahr<>0 then VariationpopNOGA3=
((sumNOGA3pop-sumNOGA3pop_vorjahr)/sumNOGA3pop_vorjahr)*100;
if sumNOGA311tg_vorjahr<>0 then Variation11tgNOGA3=
((sumNOGA311tg-sumNOGA311tg_vorjahr)/sumNOGA311tg_vorjahr)*100;
if sumtotNOGA3_vorjahr<>0 then VariationtotNOGA3=
((sumtotNOGA3-sumtotNOGA3_vorjahr)/sumtotNOGA3_vorjahr)*100;
run;
title 'Résultats par activité économique avec l"estimateur par
le quotient combiné pour l"échantillon';
proc print data=resultfinauxNOGA3;
var UNOGA3 sum_rep_umsatz_NOGA3 sumechNOGA3
sum_rep_umsatz_NOGA3_vorjahr sumechNOGA3_vorjahr
VariationechNOGA3 sumNOGA311tg sumNOGA311tg_vorjahr
Variation11tgNOGA3;
run;
title 'Résultats par activité économique avec l"estimateur par
le quotient combiné pour l"extrapolation';
proc print data=resultfinauxNOGA3;
var UNOGA3 sumNOGA3pop sumNOGA3pop_vorjahr VariationpopNOGA3
```

```
sumtotNOGA3 sumtotNOGA3_vorjahr VariationtotNOGA3;
run;
title 'Résumé : Résultats par activité économique des variations';
proc print data=resultfinauxNOGA3;
var UNOGA3 sum_rep_umsatz_NOGA3
sum_rep_umsatz_NOGA3_vorjahr
VariationechNOGA3 VariationpopNOGA3
Variation11tgNOGA3 VariationtotNOGA3;
run;
/*                                                                    */
/* Partie B : résultats finaux par classe de taille                   */
/*                                                                    */
proc sort data=resultfinauxstra;
by UTAILLE;
run;
proc univariate noprint data=resultfinauxstra;
var stra_rep_umsatz stra_rep_umsatz_vorjahr stra_pmg_umsatz_total
stra_pmg_umsatz_tot_vorjahr YRcstra YRcstra_vorjahr tg_umsatz_total_stra
tg_umsatz_tot_vorjahr_stra YFRCstra YFRCstra_vorjahr;
output out=resTAILLE sum=sum_rep_umsatz_TAILLE
sum_rep_umsatz_TAILLE_vorjahr sumechTAILLE sumechTAILLE_vorjahr
sumTAILLEpop sumTAILLEpop_vorjahr sumTAILLE11tg
sumTAILLE11tg_vorjahr sumtotTAILLE sumtotTAILLE_vorjahr;
by UTAILLE;
run;
data resultfinauxTAILLE;
set resTAILLE;
VariationechTAILLE=.;
VariationpopTAILLE=.;
Variation11tgTAILLE=.;
VariationtotTAILLE=.;
if sumechTAILLE_vorjahr<>0 then VariationechTAILLE=
((sumechTAILLE-sumechTAILLE_vorjahr)/sumechTAILLE_vorjahr)*100;
if sumTAILLEpop_vorjahr<>0 then VariationpopTAILLE=
((sumTAILLEpop-sumTAILLEpop_vorjahr)/sumTAILLEpop_vorjahr)*100;
if sumTAILLE11tg_vorjahr<>0 then Variation11tgTAILLE=
((sumTAILLE11tg-sumTAILLE11tg_vorjahr)/sumTAILLE11tg_vorjahr)*100;
if sumtotTAILLE_vorjahr<>0 then VariationtotTAILLE=
((sumtotTAILLE-sumtotTAILLE_vorjahr)/sumtotTAILLE_vorjahr)*100;
run;
title 'Résultats par classe de taille avec l"estimateur par
le quotient combiné pour l"échantillon';
proc print data=resultfinauxTAILLE;
var UTAILLE sum_rep_umsatz_TAILLE sumechTAILLE
sum_rep_umsatz_TAILLE_vorjahr sumechTAILLE_vorjahr
```

```
VariationechTAILLE sumTAILLE11tg sumTAILLE11tg_vorjahr
Variation11tgTAILLE;
run;
title 'Résultats par classe de taille avec l"estimateur par
le quotient combiné pour l"extrapolation';
proc print data=resultfinauxTAILLE;
var UTAILLE sumTAILLEpop sumTAILLEpop_vorjahr
VariationpopTAILLE sumtotTAILLE sumtotTAILLE_vorjahr
VariationtotTAILLE;
run;
title 'Résumé : Résultats par classe de taille des variations';
proc print data=resultfinauxTAILLE;
var UTAILLE sum_rep_umsatz_TAILLE sum_rep_umsatz_TAILLE_vorjahr
VariationechTAILLE VariationpopTAILLE Variation11tgTAILLE
VariationtotTAILLE;
run;
data resultfinaux;
     merge resultech resultech_vorjahr res11entreprises;
by aidemerge;
YFST=Yst+tg_umsatz_total;
YFRS=YRs+tg_umsatz_total;
YFRC=YRc+tg_umsatz_total;
YFST_vorjahr=Yst_vorjahr+tg_umsatz_tot_vorjahr;
YFRS_vorjahr=YRs_vorjahr+tg_umsatz_tot_vorjahr;
YFRC_vorjahr=YRc_vorjahr+tg_umsatz_tot_vorjahr;
VariationYFST=((YFST-YFST_vorjahr)/YFST_vorjahr)*100;
VariationYFRS=((YFRS-YFRS_vorjahr)/YFRS_vorjahr)*100;
VariationYFRC=((YFRC-YFRC_vorjahr)/YFRC_vorjahr)*100;
cvYFST=(stdYST/YFST)*100;
cvYFRS=(stdYRS/YFRS)*100;
run;
/*                                                                    */
/* Les résultats pour l'estimation du volume total sont présentées pour les  */
/* trois méthodes d'estimation.                                       */
/*                                                                    */
/*                                                                    */
/* Estimation par expansion                                           */
/*                                                                    */
title 'Résultats avec l"estimateur par expansion pour le volume total';
proc print data=resultfinaux;
var YFST YFST_vorjahr VariationYFST cvYFST;
run;
/*                                                                    */
/* Estimation par le quotient séparé                                  */
/*                                                                    */
```

```
title 'Résultats avec l"estimateur par le quotient séparé pour le volume total';
proc print data=resultfinaux;
var YFRS YFRS_vorjahr VariationYFRS cvYFRS;
run;
/*                                                                    */
/* Estimation par le quotient combiné                                 */
/*                                                                    */
title 'Résultats avec l"estimateur par le quotient combiné pour le volume total';
proc print data=resultfinaux;
var YFRC YFRC_vorjahr VariationYFRC;
run;
/*                                                                    */
/* Données sur les groupes de marchandises                            */
/*                                                                    */
/*                                                                    */
/* Format pour l'interprétation des résultats dans les groupes de marchandise  */
/*                                                                    */
/*                                                                    */
/* Séparation entre véhicules à moteurs (13) et carburants (12).      */
/*                                                                    */
proc format;
value codfn
-2 = "-2 : indéterminé"
-1 = "-1 : inconnu"
1 = "1 : Alimentation"
2 = "2 : Boissons"
3 = "3 : Tabac, articles pour fumeurs"
4 = "4 : Vêtements, chaussures"
5 = "5 : Effets personnels"
6 = "6 : Santé, soins, beauté"
7 = "7 : Aménagement du logement"
8 = "8 : Cuisine, ménage"
9 = "9 : Electronique de divertissement et de bureau"
10 = "10 : Do-it-yourself, loisirs"
11 = "11 : Culture"
12 = "12 : Carburants, combustibles"
13 = "13 : Véhicules à moteur (501-504)";
run;
/*                                                                    */
/* On reprend dans le data set waren le fichier d'input warenwelle    */
/*                                                                    */
data waren;
set SASUSER.warenwelle;
NOGA3=substr(NOGA_STUFE5_ORIG,1,3);
if NOGA3='501' then UNOGA3=501;
```

```
if NOGA3='502' then UNOGA3=502;
if NOGA3='503' then UNOGA3=503;
if NOGA3='504' then UNOGA3=504;
if NOGA3='505' then UNOGA3=505;
if NOGA3='521' then UNOGA3=521;
if NOGA3='522' then UNOGA3=522;
if NOGA3='523' then UNOGA3=523;
if NOGA3='524' then UNOGA3=524;
if NOGA3='525' then UNOGA3=525;
if NOGA3='526' then UNOGA3=526;
if NOGA3='527' then UNOGA3=527;
UTAILLE=GROESSENKLASSE_ID;
if UNOGA3=501 AND warengruppe_id=12 then warengruppe_id=13;
if UNOGA3=502 AND warengruppe_id=12 then warengruppe_id=13;
if UNOGA3=503 AND warengruppe_id=12 then warengruppe_id=13;
if UNOGA3=504 AND warengruppe_id=12 then warengruppe_id=13;
run;
/*                                                                    */
/* On traite à part les 11 très grandes entreprises au niveau des groupes de  */
/* marchandise. On code sous -2 (indéterminé) les warengruppe_id=-1   */
/* puisqu'il n'y a qu'une seule entreprise qui a des chiffres d'affaires codés en -1. */
/*                                                                    */
data tgrandeswaren;
set waren;
where unternehmen_id=3348 OR unternehmen_id=3269 OR
unternehmen_id=3322 OR unternehmen_id=3368 OR
unternehmen_id=3373 OR unternehmen_id=3384 OR
unternehmen_id=3245 OR unternehmen_id=3328 OR
unternehmen_id=3701 OR unternehmen_id=3262 OR
unternehmen_id=3334;
if warengruppe_id=-1 then warengruppe_id=-2;
run;
proc sort data=tgrandeswaren;
by warengruppe_id;
run;
/*                                                                    */
/* On calcule la somme des chiffres d'affaires de ces 11 très grandes entreprises */
/* par groupe de marchandise pour la variable UMSATZ (chiffre d'affaires par */
/* groupe de marchandise (sum_tg_umsatzwaren)) ainsi que pour la variable */
/* UMSATZ_VORJAHR (pour le même mois de l'année précédente            */
/* (sum_tg_umsatzwaren_vorjahr)).                                     */
/*                                                                    */
proc univariate noprint data=tgrandeswaren;
var UMSATZ UMSATZ_VORJAHR;
output out=restgrandeswaren N=nb_rep_11tg_umsatz
```

```
nb_rep_11tg_umsatz_vorjahr sum=sum_tg_umsatzwaren
sum_tg_umsatzwaren_vorjahr;
by warengruppe_id;
run;
/*                                                                    */
/* On traite le reste des entreprises répondantes sauf les 11 très grandes pour  */
/* obtenir les statistiques sur la partie échantillon uniquement (celle qui sera  */
/* extrapolée).                                                       */
/*                                                                    */
/*                                                                    */
/* Le data set waren_utilisees contient les données sans les 11 très grandes  */
/* pour UMSATZ>=0. Attention : il faut systématiquement regrouper les  */
/* strates des moyennes et grandes entreprises.                       */
/*                                                                    */
data waren_utilisees;
set waren;
if unternehmen_id=3348 then delete;
if unternehmen_id=3269 then delete;
if unternehmen_id=3322 then delete;
if unternehmen_id=3368 then delete;
if unternehmen_id=3373 then delete;
if unternehmen_id=3384 then delete;
if unternehmen_id=3245 then delete;
if unternehmen_id=3328 then delete;
if unternehmen_id=3701 then delete;
if unternehmen_id=3262 then delete;
if unternehmen_id=3334 then delete;
where UMSATZ>=0;
if UTAILLE=3 then UTAILLE=2;
if UNOGA3=501 AND UTAILLE=2 then UTAILLE=1;
run;
proc sort data=waren_utilisees;
by UNOGA3 UTAILLE;
run;
/*                                                                    */
/* On reprend les poids qui ont été calculés dans le data set resultats.  */
/* On attribue le poids de chaque strate à la variable UMSATZ         */
/* (umsatz_waren_pondere).                                            */
/*                                                                    */
data warenpoids;
merge resultats waren_utilisees;
by UNOGA3 UTAILLE;
umsatz_waren_pondere=poids_umsatz*umsatz;
if umsatz_waren_pondere=. then delete;
run;
```

```
proc sort data=warenpoids;
by warengruppe_id;
run;
/*                                                                    */
/* On calcule la somme pondérée de la variable UMSATZ                 */
/* (sum_umsatz_waren_pondere) par groupe de marchandises : ceci permet de */
/* contrôler la consistance des estimateurs implémentés lorsque les données sont */
/* consistantes : il s'agit d'un contrôle à faire avant de publier les résultats. */
/*                                                                    */
proc univariate noprint data=warenpoids;
var umsatz_waren_pondere umsatz;
output out=reswarenextrapole N=nb_rep_umsatz_pondere nb_rep_umsatz
sum=sum_umsatz_waren_pondere sumech_umsatz_waren;
by warengruppe_id;
run;
/*                                                                    */
/* Préparation du data set sur le volume total extrapolé à partir des chiffres */
/* d'affaire des groupes de marchandises : ce data set devra être associé à celui */
/* sur les 11 très grandes entreprises.                               */
/*                                                                    */
data warenpourresultats;
     set reswarenextrapole;
     aidemerge=1;
     run;
/*                                                                    */
/* On reprend les informations du data set resultech sur Xst et X.    */
/*                                                                    */

     data warenresultats;
     merge resultech warenpourresultats;
     keep warengruppe_id nb_rep_umsatz sum_umsatz_waren_pondere
     sumech_umsatz_waren Xst X;
     by aidemerge;
     run;
/*                                                                    */
/* Calcul effectif de l'estimateur par le quotient combiné pour chaque groupe de */
/* marchandises.                                                      */
/*                                                                    */

data warenresultatspmg;
set warenresultats;
     YRcwaren=(sum_umsatz_waren_pondere/Xst)*X;
run;
/*                                                                    */
/* On associe dans le data set resfinauxwaren les résultats extrapolés en */
```

```
/* utilisant l'estimateur par le quotient et les résulats des 11 très grandes      */
/* entreprises ne comptant que pour elles-mêmes.                                   */
/*                                                                                 */
data resfinauxwaren;
merge warenresultatspmg restgrandeswaren;
by warengruppe_id;
run;
/*                                                                                 */
/* Test de consistance en calculant l'estimation du volume total selon les deux    */
/* méthodes d'estimation par expansion (YFstwaren) et par le quotient combiné      */
/* (YFRcwaren) en se basant sur les chiffres d'affaires par groupe de              */
/* marchandises.                                                                   */
/*                                                                                 */
data pourresaggregeswaren;
set resfinauxwaren;
     YFstwaren=sum_umsatz_waren_pondere+sum_tg_umsatzwaren;
YFRcwaren=YRcwaren+sum_tg_umsatzwaren;
     run;
title 'Résultats par groupe de marchandises pour l"extrapolation avec
la variable UMSATZ';
proc print data=pourresaggregeswaren;
var warengruppe_id nb_rep_umsatz sumech_umsatz_waren YRcwaren
nb_rep_11tg_umsatz sum_tg_umsatzwaren YFRcwaren;
format warengruppe_id codfn.;
run;
     proc univariate noprint data=pourresaggregeswaren;
     var YFstwaren YFRcwaren;
     output out=resaggregeswaren sum=sum_YFstwaren sum_YFRcwaren;
     run;
title 'Résultats pour le volume total avec la variable UMSATZ
(par expansion et par le quotient combiné) ';
     proc print data=resaggregeswaren;
     var sum_YFstwaren sum_YFRcwaren;
     run;
/*                                                                                 */
/* Le data set waren_utilisees contient les données sans les 11 très grandes pour */
/* UMSATZ_VORJAHR>=0. Attention : il faut systématiquement regrouper               */
/* les strates des moyennes et grandes entreprises.                                */
/*                                                                                 */
data waren_utilisees_vorjahr;
set waren;
if unternehmen_id=3348 then delete;
if unternehmen_id=3269 then delete;
if unternehmen_id=3322 then delete;
if unternehmen_id=3368 then delete;
```

```
if unternehmen_id=3373 then delete;
if unternehmen_id=3384 then delete;
if unternehmen_id=3245 then delete;
if unternehmen_id=3328 then delete;
if unternehmen_id=3701 then delete;
if unternehmen_id=3262 then delete;
if unternehmen_id=3334 then delete;
where UMSATZ_VORJAHR>=0;
if UTAILLE=3 then UTAILLE=2;
if UNOGA3=501 AND UTAILLE=2 then UTAILLE=1;
run;
proc sort data=waren_utilisees_vorjahr;
by UNOGA3 UTAILLE;
run;
/*                                                                    */
/* On reprend les poids qui ont été calculés dans le data set resultats.On  */
/* attribue le poids de chaque strate à la variable UMSATZ_VORJAHR    */
/* (umsatz_waren_pondere).                                            */
/*                                                                    */
data warenpoids_vorjahr;
merge resultats_vorjahr waren_utilisees_vorjahr;
by UNOGA3 UTAILLE;
umsatz_waren_pondere_vorjahr=poids_umsatz_vorjahr*umsatz_vorjahr;
if umsatz_waren_pondere_vorjahr=. then delete;
run;
proc sort data=warenpoids_vorjahr;
by warengruppe_id;
run;
/*                                                                    */
/* On calcule la somme pondérée de la variable UMSATZ_VORJAHR         */
/* (sum_umsatz_waren_pondere_vorjahr) par groupe de marchandises : ceci */
/* permet de contrôler la consistance des estimateurs implémentés lorsque les */
/* données sont consistantes : il s'agit d'un contrôle à faire également sur la */
/*variable UMSATZ_VORJAHR avant de publier les résultats.             */
/*                                                                    */

proc univariate noprint data=warenpoids_vorjahr;
var umsatz_waren_pondere_vorjahr umsatz_vorjahr;
output out=reswarenextrapole_vorjahr
N=nb_rep_umsatz_pondere_vorjahr nb_rep_umsatz_vorjahr
sum=sum_umsatz_waren_pondere_vorjahr sumech_umsatz_waren_vorjahr;
by warengruppe_id;
run;

/*                                                                    */
```

```
/* Préparation du data set sur le volume total du même mois de l'année      */
/* précédente extrapolé à partir des chiffres d'affaires des groupes de     */
/* marchandises : ce data set devra être associé à celui sur les 11 très    */
/* grandes entreprises.                                                     */
/*                                                                          */
     data warenpourresultats_vorjahr;
     set reswarenextrapole_vorjahr;
     aidemerge=1;
     run;
/*                                                                          */
/* On reprend les informations du data set resultech_vorjahr sur Xst et X.  */
/*                                                                          */

     data warenresultats_vorjahr;
     merge resultech_vorjahr warenpourresultats_vorjahr;
     keep warengruppe_id nb_rep_umsatz_vorjahr
     sum_umsatz_waren_pondere_vorjahr
     sumech_umsatz_waren_vorjahr Xst_vorjahr X_vorjahr;
     by aidemerge;
     run;
/*                                                                          */
/* Calcul effectif de l'estimateur par le quotient combiné pour chaque groupe */
/* de marchandises avec la variable sum_umsatz_pondere_vorjahr et les       */
/* informations Xst_vorjahr et X=X_vorjahr.                                 */
/*                                                                          */
data warenresultatspmg_vorjahr;
set warenresultats_vorjahr;
    YRcwaren_vorjahr=
     (sum_umsatz_waren_pondere_vorjahr/Xst_vorjahr)*X_vorjahr;
run;
/*                                                                          */
/* On associe dans le data set resfinauxwaren_vorjahr les résultats         */
/* extrapolés pour le même mois de l'année précédente en utilisant          */
/* l'estimateur par le quotient et les résulats des 11 très grandes         */
/* entreprises ne comptant que pour elles-mêmes.                            */
/*                                                                          */
data resfinauxwaren_vorjahr;
merge warenresultatspmg_vorjahr restgrandeswaren;
by warengruppe_id;
run;
/*                                                                          */
/* Test de consistance en calculant l'estimation du volume total pour       */
/* le même mois de l'année précédente selon les deux méthodes               */
/* d'estimation par expansion (YFstwaren_vorjahr) et par le quotient        */
/* combiné (YFRcwaren_vorjahr) en se basant sur les chiffres d'affaires     */
```

```
/* par groupe de marchandises.                                          */
/*                                                                      */
data pourresaggregeswaren_vorjahr;
set resfinauxwaren_vorjahr;
      YFstwaren_vorjahr=sum_umsatz_waren_pondere_vorjahr+
      sum_tg_umsatzwaren_vorjahr;
      YFRcwaren_vorjahr=YRcwaren_vorjahr+
       sum_tg_umsatzwaren_vorjahr;
      run;
title 'Résultats par groupe de marchandises pour l"extrapolation
avec la variable UMSATZ_VORJAHR';
proc print data=pourresaggregeswaren_vorjahr;
var warengruppe_id nb_rep_umsatz_vorjahr sumech_umsatz_waren_vorjahr
YRcwaren_vorjahr nb_rep_11tg_umsatz_vorjahr sum_tg_umsatzwaren_vorjahr
YFRcwaren_vorjahr;
format warengruppe_id codfn.;
run;
      proc univariate noprint data=pourresaggregeswaren_vorjahr;
      var YFstwaren_vorjahr YFRcwaren_vorjahr;
      output out=resaggregeswaren_vorjahr sum=sum_YFstwaren_vorjahr
      sum_YFRcwaren_vorjahr;
      run;
title 'Résultats pour le volume total avec la variable UMSATZ_VORJAHR
(par expansion et par le quotient combiné) ';
      proc print data=resaggregeswaren_vorjahr;
      var sum_YFstwaren_vorjahr sum_YFRcwaren_vorjahr;
      run;
/*                                                                      */
/* Association des data set permettant de calculer les variations du mois */
/* sous revue par rapport au même mois de l'année précédente par groupe */
/* de marchandises.                                                     */
/*                                                                      */
data zproduction;
merge pourresaggregeswaren pourresaggregeswaren_vorjahr;
by warengruppe_id;
run;
/*                                                                      */
/* Calcul par groupe de marchandises avec l'estimateur par le quotient  */
/* combiné pour la partie extrapolée YRcpopwaren et YRcpopwaren_vorjahr */
/* respectivement, calcul pour les 11 très grandes entreprises au niveau */
/* groupe de marchandises Y11tgwaren et Y11tgwaren_vorjahr              */
/* respectivement et finalement les résultats finaux (somme de la partie */
/* extrapolée et des 11 très grandes entreprises) au niveau groupe de   */
/* marchandises YFRcwaren et YFRcwaren_vorjahr. Les variations          */
/* respectives se trouvent respectivement dans variapop varia11tg et    */
```

```
/* variaFRc : variaFRc est la variation effectivement utilisée pour la          */
/* publication.                                                                  */
/*                                                                               */

data zfinalproduction;
set zproduction;
YRcechwaren=sumech_umsatz_waren/1000000;
YRcpopwaren=YRcwaren/1000000;
Y11tgwaren=sum_tg_umsatzwaren/1000000;
YFRcwaren=YFRcwaren/1000000;
YRcechwaren_vorjahr=sumech_umsatz_waren_vorjahr/1000000;
YRcpopwaren_vorjahr=YRcwaren_vorjahr/1000000;
Y11tgwaren_vorjahr=sum_tg_umsatzwaren_vorjahr/1000000;
YFRcwaren_vorjahr=YFRcwaren_vorjahr/1000000;
variaech=((YRcechwaren-YRcechwaren_vorjahr)/
YRcechwaren_vorjahr)*100;
variapop=((YRcpopwaren-YRcpopwaren_vorjahr)/
YRcpopwaren_vorjahr)*100;
varia11tg=((Y11tgwaren-Y11tgwaren_vorjahr)/Y11tgwaren_vorjahr)*100;
variaFRc=((YFRcwaren-YFRcwaren_vorjahr)/YFRcwaren_vorjahr)*100;
run;
title 'Résultats finaux pour l"échantillon par groupe de marchandises
pour la vague en cours (en millions de francs)';
proc print data=zfinalproduction;
var warengruppe_id nb_rep_umsatz YRcechwaren nb_rep_umsatz_vorjahr
YRcechwaren_vorjahr variaech Y11tgwaren Y11tgwaren_vorjahr varia11tg;
format warengruppe_id codfn.;
run;
title 'Résultats finaux pour l"extrapolation par groupe de marchandises pour
la vague en cours (en millions de francs)';
proc print data=zfinalproduction;
var warengruppe_id nb_rep_umsatz YRcpopwaren nb_rep_umsatz_vorjahr
YRcpopwaren_vorjahr variapop YFRcwaren YFRcwaren_vorjahr variaFRc;
format warengruppe_id codfn.;
run;
title 'Résumé : Résultats finaux par groupe de marchandises pour la vague en cours';
proc print data=zfinalproduction;
var warengruppe_id nb_rep_umsatz nb_rep_umsatz_vorjahr
variaech variapop varia11tg variaFRc;
format warengruppe_id codfn.;
run;

proc univariate noprint data=zfinalproduction;
var YRcpopwaren Y11tgwaren YFRcwaren YRcpopwaren_vorjahr
Y11tgwaren_vorjahr YFRcwaren_vorjahr;
```

```
output out=totalfinalproduction sum=SYRcpopwaren SY11tgwaren
SYFRcwaren SYRcpopwaren_vorjahr SY11tgwaren_vorjahr SYFRcwaren_vorjahr;
run;
/*                                                                    */
/* Calcul au niveau du volume total avec l'estimateur par le quotient combiné  */
/* pour la partie extrapolée TOTYRcpopwaren et TOTYRcpopwaren_vorjahr  */
/* respectivement, calcul pour les 11 très grandes entreprises au niveau du total */
/* TOTY11tgwaren et TOTY11tgwaren_vorjahr respectivement et finalement  */
/* les résultats finaux (somme de la partie extrapolée et des 11 très grandes  */
/* entreprises) au niveau du volume total TOTYFRcwaren et  */
/* TOTYFRcwaren_vorjahr. Les variations respectives se trouvent resp.  */
/* dans TOTVARIAPOP TOTVARIA11tg et TOTVARIAFRc :  */
/* TOTVARIAFRc est la variation utilisée pour la publication.  */
/*                                                                    */
data ztotalfinalproduction;
set totalfinalproduction;
TOTYRcpopwaren=SYRcpopwaren;
TOTY11tgwaren=SY11tgwaren;
TOTYFRcwaren=SYFRcwaren;
TOTYRcpopwaren_vorjahr=SYRcpopwaren_vorjahr;
TOTY11tgwaren_vorjahr=SY11tgwaren_vorjahr;
TOTYFRcwaren_vorjahr=SYFRcwaren_vorjahr;
TOTVARIApop=((TOTYRcpopwaren-TOTYRcpopwaren_vorjahr)/
TOTYRcpopwaren_vorjahr)*100;
TOTVARIA11tg=((TOTY11tgwaren-TOTY11tgwaren_vorjahr)/
TOTY11tgwaren_vorjahr)*100;
TOTVARIAFRc=((TOTYFRcwaren-TOTYFRcwaren_vorjahr)/
TOTYFRcwaren_vorjahr)*100;
run;
title 'Résultats finaux aggrégés pour la vague en cours exprimés en millions de
francs';
proc print data=ztotalfinalproduction;
var TOTYRcpopwaren
TOTY11tgwaren
TOTYFRcwaren
TOTYRcpopwaren_vorjahr
TOTY11tgwaren_vorjahr
TOTYFRcwaren_vorjahr
TOTVARIApop
TOTVARIA11tg
TOTVARIAFRc;
run;
/*                                                                    */
/* Sauvegarde dans des fichiers Excel                                 */
/*                                                                    */
```

```
data outprodwaren;
set zfinalproduction;
keep warengruppe_id YFRcwaren YFRcwaren_vorjahr variaFRc;
run;
PROC DBLOAD DBMS=EXCEL DATA=WORK.OUTPRODWAREN;
PATH='T :\21 DHU, Projekte\DHU 2000\production\outproductionwaren';
PUTNAMES YES;
LIMIT=0;
LOAD;
RUN;
```

B Références

Cochran, W. G. (1977). *Sampling Techniques.* John Wiley & Sons, New York.

Gourieroux, C. (1981). *Théorie des sondages.* Edition Economica, Paris.

Särndal, C.E., Swensson, B. and Wretman, J. (1992). *Model Assisted Survey Sampling.* Springer, New York.

Venables, W.N. and Ripley, B.D. (1994). *Modern Applied Statistics with S-Plus.* Springer, 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

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). Erhebung über Forschung und Entwicklung in der schweizerischen Privatwirtschaft 2000. Bereinigung der Stichprobe, Ersatz fehlender Werte und Schätzverfahren. Bestellnummer: 338-0018

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Publikationsprogramm BFS

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

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

Programme des publications de l'OFS

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

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

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

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

L'objectif de l'enquête conjoncturelle DHU révisée est de fournir à un rythme mensuel une estimation de l'évolution des chiffres d'affaires du commerce de détail.

Cette étude a été menée dans le but de fournir le plan d'échantillonnage pour la DHU révisée. Dans ce cadre, les éléments essentiels ont pu être développés pour déterminer la meilleure stratégie pour le tirage des échantillons. Un problème important a pu être résolu en procédant à la reconstitution des entreprises à partir des informations obtenues sur leur siège et les établissements qui les constituent. Des solutions satisfaisantes ont également pu être trouvées pour traiter de la problématique des données, notamment en utilisant des modèles de régression robustes basées sur la norme l_1 adaptés au traitement des données aberrantes. Une grande partie de ce rapport a été consacrée aux différentes méthodes d'estimation et au calcul de la précision attendue en fonction du plan d'échantillonnage retenu. La recherche d'estimateurs par le quotient conjointement avec une allocation optimale de Neyman a permis un gain en précision non négligeable.

N° de commande: 338-0041
Commandes: 032 713 60 60
Fax: 032 713 60 61
E-Mail: order@bfs.admin.ch
Prix: gratuit
ISBN 3-303-00341-6