

Statistics Belgium

Working Paper

L'Institut national de statistique propose des informations impartiales à un prix abordable.

Les informations sont diffusées conformément à la loi, notamment pour ce qui concerne leur confidentialité.

Nous classons les statistiques en huit domaines :

Généralités	Économie et finances
Territoire et environnement	Agriculture et activités assimilées
Population	Industrie
Société	Services, commerce et transports

Tous droits de traduction, d'adaptation, de reproduction par tous procédés, y compris la photographie et le microfilm sont soumis à autorisation préalable de l'Institut national de statistique.

Toutefois, la citation de courts extraits, à titre explicatif ou justificatif, dans un article, un compte-rendu ou un livre, est autorisée moyennant indication claire et précise de la source.

Éditeur responsable : Hans D'Hondt



©2003, INSTITUT NATIONAL DE STATISTIQUE *éditeur*

B-1000 Bruxelles – 44 rue de Louvain

Chère lectrice, Cher lecteur,

Ce que vous avez sous les yeux est une publication dans la nouvelle série méthodologique de l'Institut national de Statistique belge, série intitulée "Statistics Belgium Working Papers". Notre propos est de mettre à la disposition de nos utilisateurs et de quiconque s'occupe de statistique, - qu'il appartienne au monde des affaires ou au milieu universitaire-, des études, des rapports et des documents préparatoires. En outre, nous voulons donner la possibilité aux membres de notre staff, aux statisticiens, aux méthodologistes et à d'autres de faire connaître leurs idées et les résultats de leurs travaux.

Ces articles visent à contribuer au développement des connaissances statistiques et à intensifier l'échange des idées.

Cependant, en tant que *working papers*, ces articles ne reflètent le point de vue officiel ni de l'Institut national de Statistique (INS) ni du gouvernement belge. La responsabilité quant au contenu incombe aux seuls auteurs.

L'INS lance cette initiative dans l'espoir de rehausser le niveau des discussions sur les besoins et les méthodes statistiques, et d'assurer une meilleure diffusion des idées et des résultats.

Pour des raisons de commodité, les articles des *Statistics Belgium Working Papers* seront publiés dans la langue du texte original.

Le staff ainsi que tous les cadres de l'INS, et plus particulièrement les auteurs, attendent avec impatience de recevoir tout commentaire ou toute question concernant les articles publiés.

Bruxelles,
Hans D'Hondt
Directeur général,
Institut national de Statistique

Statistics Belgium

Working Paper n° 8

L'enquête sur les Forces de travail :
calibrage et autres développements

Denis Luminet

Actuaire, Bureau Méthodologie et Coordination

Direction générale de la Statistique et de l'Information économique

Sommaire

Résumé	3
Introduction	7
I - Contexte	
I.1 Les comptes nationaux.....	10
I.2 Les comptes de l'emploi.....	10
I.3 Les concepts du BIT	12
I.4 Comparaison enquête/fichiers.....	13
Principaux défauts de l'EFT	13
Principales qualités de l'EFT	14
Différence entre source EFT et source Sécurité sociale.....	14
Spécificités de l'EFT comme enquête auprès des ménages.....	14
I.5 Données de la Sécurité sociale et dossier BCSS.....	15
II – Sondages et extrapolations	
II.1 Sources d'erreurs inhérentes à un sondage	20
II.2 Si l'EFT était un recensement	21
II.3 Plan de sondage	21
II.4 Estimateurs	22
II.5 Extrapolation en l'absence de non-réponse.....	23
II.6 La non-réponse	25
II.7 Extrapolation en présence de non-réponse	27
II.8 Informations auxiliaires supplémentaires	28
II.9 Calage sur marges.....	30
II.10 Distances et fonctions de calibrage.....	31
II.11 Exemples de fonctions de calibrage.....	33
II.12 g-Calib	33
III Mise en oeuvre	
III.1 Options	36
III.2 Quelques corrélations.....	36
III.3 Choix des variables de calibrage	38
III.4 Post-strates et marges de calibrage.....	39
III.5 Calendrier	40
III.6 Totaux de population	41

III.7 Ménages	41
III.8 Fonctions de calibrage.....	42
III.9 Imputation	43
III.10 Processus.....	45
IV Résultats	
IV.1 Résultats globaux.....	48
IV.2 Croisement EFT-BCSS.....	49
IV.3 Variance	53
IV.4 NACEbis.....	53
V Conclusion	
V.1 Les "vrais" résultats de l'EFT.....	56
V.2 Sur le plan belge	56
V.3 Sur le plan européen	56
V.4 Sur le plan de la qualité.....	56
V.5 Bilan et perspectives	57
Annexe	
A Récupération des fichiers BCSS et création de "Bigfile" (d'après J.M. Museux)	60
B Variables supplémentaires	74
Γ Imputation de l'information BCSS (d'après C. Vanderhoeft).....	79
Δ Principaux résultats.....	88
Glossaire des institutions	90
Bibliographie	95

Résumé

RESUME – SAMENVATTING – SUMMARY

En Belgique, l'enquête (par sondage) sur les forces de travail (EFT) est devenue continue en 1999, et les ménages tirés au sort sont interrogés à deux reprises. Ce mémoire montre comment l'utilisation de données auxiliaires de la sécurité sociale peut améliorer la précision de l'extrapolation (méthode de calibrage).

In België is de (steekproef)enquête naar de arbeidskrachten (EAK) doorlopend geworden in 1999, en de getrokken huishoudens worden tweemaal bevraagd. Deze verhandeling laat zien hoe het gebruik van bijkomende gegevens van de sociale zekerheid de precisie van de extrapolatie kan verbeteren (kalibratiemethode).

In Belgium, the labour force (sample) survey (LFS) became continuous in 1999, and the selected households are interrogated twice. This memoir shows how the use of auxiliary data from the social security can improve the accuracy of the extrapolation (calibration method).

Introduction

INTRODUCTION

L'Enquête par sondage sur les Forces de Travail (EFT) est une enquête socio-économique auprès des ménages dont le but principal est de classer la population en âge de travailler (15 ans et plus) en trois groupes exhaustifs et distincts (personnes occupées, chômeurs et inactifs), et de fournir des données sur chacune de ces catégories.

Elle s'inscrit dans le cadre des enquêtes réalisées par les Instituts nationaux de statistique sous la coordination de l'Office statistique des Communautés européennes Eurostat. Il s'agit d'obtenir des informations comparables au niveau européen notamment pour ce qui concerne les taux d'emploi et de chômage (leur somme donnant le taux d'activité) au sens du Bureau international du Travail (BIT), et aussi de recueillir, traiter et diffuser des informations non disponibles par ailleurs.

Organisée chaque printemps entre 1983 et 1998, l'enquête est devenue continue dans notre pays en 1999, de plus, les répondants (de 15 à 64 ans) sont à nouveau interrogés trois mois plus tard (par courrier ou par téléphone)¹.

L'accent a été mis dans le présent rapport sur les **données auxiliaires** de la Sécurité Sociale et leur utilisation pour **calibrer** l'EFT, en améliorant la précision des estimations fournies. Nous nous sommes efforcés de maintenir l'équilibre entre la théorie (mathématique) et la pratique (d'un service public fédéral), entre aspirations scientifiques (investiguer toutes les options) et contraintes politiques (souhait d'une méthode et d'un résultat).

Nous commençons par rappeler le contexte: le cadre institutionnel, les définitions, les objectifs poursuivis et les moyens que nous nous sommes donnés. Suivent des considérations sur l'extrapolation des sondages en général et de l'EFT en particulier. Le troisième chapitre indique comment ont été opérés les choix ou résolues les difficultés pratiques. Les résultats constituent le quatrième chapitre, lequel précède les conclusions. En Annexe figurent les principales syntaxes. Faute de temps, les autres sujets ne seront que brièvement présentés sous forme d'Appendice; nous espérons qu'ils pourront être investigués dès que possible. Soulignons d'ailleurs que notre méthode actuelle de calibrage devra également être évaluée et probablement revue dans les prochaines années.

Si l'EFT était notre centre d'intérêt concret, les méthodes (développées ou à développer) pourraient s'appliquer à d'autres enquêtes (entreprises et ménages). En effet, le rôle des actuaire constituant le Bureau des Méthodes et de Coordination est notamment d'améliorer et d'unifier les techniques statistiques.

¹ Une nouvelle proposition d'Eurostat, visant à poser des questions supplémentaires soit durant certains trimestres, soit à certaines vagues de ménages, pourrait amener l'INS à revoir ce plan de rotation (tirer moins de ménages, les interroger plus souvent).

I Contexte

I CONTEXTE

I.1 **Les comptes nationaux**

La loi du 21 décembre 1994 (Moniteur belge du 23 décembre) a créé un **Institut des Comptes Nationaux** (ICN)¹.

Selon l'article 108, l'ICN a pour mission d'établir, avec le concours des organismes visés à l'article 109, dénommés ci-après autorités associées, mais sous sa propre responsabilité, les statistiques et prévisions économiques suivantes:

- ❑ *les comptes nationaux réels*
- ❑ *les comptes nationaux financiers*
- ❑ *les comptes annuels et trimestriels des administrations publiques*
- ❑ *les comptes nationaux trimestriels*
- ❑ *les produits régionaux bruts*
- ❑ *les statistiques du commerce extérieur, au sein de l'Union européenne et en dehors de celle-ci, ainsi que celles du transit*
- ❑ *les prévisions économiques nécessaires à l'établissement du budget fédéral, appelées aussi budget économique*
- ❑ *les tableaux des entrées et sorties, y compris éventuellement certains comptes satellites.*

Les trois autorités associées de l'ICN sont

le Bureau fédéral du Plan (BFP)

la Banque nationale de Belgique (BNB)

l'Institut national de statistique (INS)

L'article 109 §3 précise: L'ICN confie à la Banque nationale de Belgique l'élaboration des tableaux statistiques visés à l'article 108, a), b), d) et e) de la présente loi. Pour ce faire, la Banque nationale de Belgique se base notamment sur des données collectées par l'Institut national de statistique et établies par l'ICN. «Notamment» indique que la BNB peut utiliser, outre les données transmises par l'INS, d'autres sources, administratives par exemple.

I.2 **Les comptes de l'emploi**

Un des objectifs poursuivis par les statistiques sociales est l'établissement des **comptes de l'emploi**.

Dans l'optique des comptes nationaux, il s'agit de déterminer différents indicateurs dont les trois fondamentaux, selon le **Système Européen de Comptes**² SEC95, sont

le nombre de personnes en emploi (salarié et non-salarié);

les heures travaillées (absolument ou en équivalent temps plein);

les rémunérations des salariés³.

Ces grandeurs peuvent être détaillées selon les caractéristiques démographiques (âge, sexe, domicile du travailleur selon la classification NUTS⁴) ou économiques (localisation de

¹ Cette loi érigeait également l'INS en service de l'Etat à gestion séparée.

² Néerlandais ESR95 (Europees Systeem van Rekeningen); anglais ESA95 (European System of Accounts).

³ Celles-ci en particulier devant s'intégrer dans les comptes nationaux, la cohérence exigeant que la somme des rémunérations versées par les entreprises belges égale la somme des rémunérations perçues par leurs salariés. Nous n'entrons pas dans les détails de la discussion (salaires bruts ou nets, ...)

⁴ NUTS = *Nomenclature des Unités Territoriales Statistiques*. La Belgique se compose de trois NUTS I (les régions), et de 11 NUTS II (les provinces, en y assimilant Bruxelles-Capitale). Dans la suite, il nous arrivera d'écrire –par abus de langage– «province» pour «NUTS II».

l'entreprise/de l'établissement selon la NUTS, secteur d'activité de l'entreprise/de l'établissement selon la NACE⁵, branche SUT⁶).

Parmi les personnes sans emploi, il est utile d'identifier celles qui sont à la recherche d'un emploi et seraient disponibles pour travailler, autrement dit les chômeurs au sens du BIT (voir paragraphe suivant).

Dans son rapport de stage *Werkgelegenheidsrekeningen* (voir aussi [Termote]), A. Termote –après avoir présenté la situation aux Pays-Bas, au Danemark et en Suisse– se penche sur les différentes sources (enquêtes de l'INS, fichiers des institutions de sécurité sociale, comptes annuels et bilan social des entreprises) pouvant contribuer à l'établissement de comptes de l'emploi, vus comme un volet des comptes nationaux et confiés donc à la BNB. Elle examine aussi dans quelle mesure les concepts coïncident avec les définitions requises (du BIT pour l'emploi, du SEC95 plus largement). Il apparaît en particulier que l'EFT serait une bonne source pour l'emploi, moins pertinente pour les heures⁷, et peu fiable pour les salaires⁸.

En ce qui concerne l'emploi, l'ICN a préféré jusqu'à présent les données administratives traitées (corrections pour doubles comptes, statuts particuliers tels celui des marins, évaluation du travail souterrain, etc.) par le Ministère fédéral de l'Emploi et du Travail. Et force est de constater que **les estimations de l'emploi** (national, c'est-à-dire des résidents, quel que soit leur lieu de travail) **diffèrent**: on observe un écart «en faveur» de l'enquête, fortement réduit en 2001 (ce qui s'explique en partie par l'interprétation plus restrictive⁹ et par une baisse mesurée par l'EFT au 4^{ème} trimestre).

TABLEAU I.1 ESTIMATIONS DE L'EMPLOI

EMPLOI NATIONAL	1999	2000	2001	2002
EFT–moyenne de l'année (<i>avant calibrage</i>)	4.006.903	4.092.286	4.051.200	?
ICN – au 30 juin	3.908.600	3.969.400	4.015.100	4.014.100 ¹⁰
Différence EFT – ICN	98.303	122.886	36.100	?

La comparaison est plus floue en ce qui concerne le chômage: si l'EFT estime le chômage BIT, les sources administratives fournissent –comme approximation– le chômage complet indemnisé. Le désir d'harmoniser les sources (et les résultats!) a conduit à la création d'une cellule de coordination (dont A. Versonnen assure la présidence et A. Termote le secrétariat), chargée d'accorder les définitions et de perfectionner l'établissement du volet «emploi» des comptes

⁵ NACE = *Nomenclature générale des Activités économiques dans les Communautés Européennes*. L'EFT collecte l'activité principale de l'entreprise (utilisée actuellement dans les comptes nationaux) et de l'établissement (privilegiée par Eurostat); la qualité de la mesure est néanmoins sujette à caution.

⁶ SUT = *Supply Users Table*.

⁷ Signalons cependant les exploitations réalisées par K. Degroote (Conseil Central de l'Economie) en liaison avec J.M. Museux, puis Y. Colens et A. Termote, avec imputation des non-réponses partielles («ne sait pas») pour les durées hebdomadaires effective –question 20- et habituelle –question 24- du travail.

⁸ L'EFT demande aux salariés -question 82- leur salaire net mensuel pour l'activité principale (ainsi que -question 83- les primes, pécules, bonus et autres rémunérations extraordinaires, sur base annuelle), mais la réponse n'est pas obligatoire (taux légèrement supérieur à 50%). Notre collaborateur scientifique E. Bonsang s'est penché tant sur les non-réponses que sur les réponses, et l'INS espère pouvoir exploiter davantage ces informations dans le futur.

⁹ Principal changement: une personne en interruption de carrière complète (régime général, pour soins palliatifs, ...) ou suspension complète des prestations (dans le nouveau système de crédit-temps) reste comptée dans l'emploi pour autant que la durée de l'interruption n'excède pas 3 mois; cette interprétation Eurostat est en vigueur depuis janvier 2001 (en 1999 et 2000, l'EFT la considérait comme occupée quelle que soit la durée) et son impact peut être évalué à 30.000 personnes.

¹⁰ Prévision. **Remarque (avril 2003) : selon la méthode révisée du Ministère fédéral de l'Emploi et du Travail (désormais SPF Emploi, Travail et Concertation Sociale), adoptée (du moins provisoirement par l'ICN), les chiffres de l'emploi seraient sensiblement supérieurs (de l'ordre de 150 000 unités) à l'estimation antérieure, et donc supérieurs à l'estimation EFT.**

nationaux belges. Cette cellule, rassemblant initialement les trois partenaires de l'ICN, s'est élargie au Ministère fédéral de l'Emploi et du Travail¹¹, et accueille ponctuellement des experts d'autres institutions.

1.3 Les concepts du BIT

L'objectif principal de l'EFT est de classer la population (des 15 ans et plus, résidant en Belgique – les enfants de 0 à 14 ans étant considérés d'office comme inactifs) en 3 catégories:

travailleurs (actifs occupés);

chômeurs (actifs inoccupés);

inactifs.

Rappelons comment ces catégories sont définies par le BIT:

Personne occupée (dite aussi «active occupée» ou «ayant un emploi») signifie: celui/celle qui, durant une certaine semaine, dite «semaine de référence», a travaillé au moins une heure contre rémunération, en tant que salarié(e) ou à son propre compte (ou encore comme aidant familial non rémunéré)¹²; on y assimile la personne ayant un emploi dont elle était temporairement absente (vacances, maladie, maternité, conflit social, intempéries, ...).

Est **chômeur** («actif inoccupé») quiconque n'avait pas d'emploi rémunéré durant la semaine de référence, en recherche activement (au moins une démarche au cours du «mois de référence», à savoir une période de quatre semaines se terminant par la semaine de référence) et serait disponible¹³ pour commencer à travailler dans les deux semaines¹⁴.

Les autres personnes, notamment celles s'occupant exclusivement de leur ménage (homme ou femme au foyer) ou travaillant bénévolement (par exemple, pour une institution humanitaire)¹⁵ seront classées comme **inactives**, catégorie qui regroupe aussi les personnes en incapacité de travail, les étudiants et les pensionnés¹⁶.

Les «forces de travail» (la population active, qu'elle soit occupée ou non) rassemblent les deux premières catégories.

En particulier, la première catégorie correspond à l'emploi (national), une des variables fondamentales des comptes de l'emploi. Elle est souvent subdivisée en emploi salarié (secteur privé et secteur public) et emploi non-salarié¹⁷.

¹¹ Lequel étudie la problématique des comptes de l'emploi dans le cadre d'un projet AGORA.

¹² Quel que soit le statut (travail «déclaré» ou «souterrain»), y compris le travail en ALE (pour la Belgique).

¹³ Concept délicat car virtuel (comment distinguer: «je pourrais commencer à travailler ... à condition qu'on puisse garder mes enfants» de «je ne pourrais pas commencer à travailler ... sauf si on pouvait garder mes enfants»).

¹⁴ Que l'individu soit indemnisé ou non par l'ONEM et enregistré ou non comme demandeur d'emploi par les bureaux régionaux ORBEM, FOREM, VDAB est sans importance.

¹⁵ Ce choix peut être discuté, mais est cohérent avec l'optique des Comptes nationaux, qui ne prennent pas en considération le travail non rémunéré (domestique ou autre) dans le calcul du PIB ou du PNB.

¹⁶ Pour autant que ces personnes n'exercent pas la moindre activité rémunérée.

¹⁷ La distinction entre «ouvrier» et «employé» est une spécificité belge – en voie d'estompement, à l'instar de la séparation entre les secteurs privé et public (songeons aux entreprises publiques autonomes). Même la différence entre salariés et non-salariés est parfois floue, songeons aux «free-lance» ou aux consultants.

Schématiquement:

TABLEAU I.2 STATUTS PROFESSIONNELS

QUESTIONNAIRE INS 7. Parmi les catégories suivantes, laquelle caractérise le mieux le statut professionnel?	CODIFICATION EUROSTAT	SALARIE OU NON
- Secteur privé - ouvrier	Salarié	Salarié
- Secteur privé - employé	Salarié	Salarié
- Secteur public - fonctionnaire statutaire	Salarié	Salarié
- Secteur public - contractuel	Salarié	Salarié
- Indépendant - en nom propre, sans personnel	Indépendant n'employant aucun salarié	Non salarié
- Indépendant - en nom propre, avec personnel	Employeur employant un ou plusieurs salariés	Non salarié
- Indépendant - en société, sans personnel	Indépendant n'employant aucun salarié	Non salarié
- Indépendant - en société, avec personnel	Employeur employant un ou plusieurs salariés	Non salarié
- Aidant familial non rémunéré	Aide familiale	Non salarié

Observons une évolution: alors que naguère le taux de chômage était l'indicateur privilégié, l'EFT (comme les comptes nationaux) se focalise désormais sur l'emploi; ainsi, la Belgique (dont le chômage BIT se situe dans la moyenne européenne) se signale –hélas– par un taux d'emploi des quinquagénaires particulièrement bas.

I.4 Comparaison enquête/fichiers

Ayant constaté que les (traitements de) données administratives (essentiellement ONSS, ONSSAPL et INASTI pour l'emploi; ONEM pour le chômage) donnent des évaluations différentes de celles fournies par l'enquête, on peut se demander si les fichiers ne répondraient pas mieux à notre objectif. Comparons donc **les faiblesses et les forces** de l'enquête par sondage et des fichiers administratifs, en observant que leurs définitions sont parfois différentes, et que l'EFT est une enquête auprès des ménages (grappes d'individus).

Principaux défauts de l'EFT

Imprécision: Les résultats d'une enquête par sondage sont parfois sujets à un biais¹⁸, toujours à un écart type¹⁹, la somme de leurs carrés donnant l'erreur quadratique moyenne²⁰.

¹⁸ Par exemple, une hypothèse (à confirmer ou à infirmer) serait «actifs occupés moins souvent chez eux → plus difficiles à interviewer → sous-représentés parmi les répondants → sous-estimés dans l'extrapolation (il s'agirait moindre probabilité de participation, même si la probabilité d'être tiré au sort dans l'échantillon initial ne dépend pas du statut professionnel). En calibrant sur des données de la sécurité sociale, nous pourrions tenter de réduire ce biais dans l'extrapolation.

¹⁹ Il se pourrait que «par hasard» on ait tiré un peu plus/moins d'actifs occupés dans l'échantillon en 1999 qu'en 1998, ce qui donnera l'impression d'une croissance/décroissance de l'emploi; une augmentation de l'échantillon réduit ces variations aléatoires. Pour situer l'ordre de grandeur, en assimilant l'EFT à un sondage aléatoire simple, on évalue à 6000 l'erreur type sur l'estimation publiée de 375.000 chômeurs BIT en 1999. En calibrant (bis!), nous pourrions espérer stabiliser l'extrapolation

²⁰ Ainsi qu'à des erreurs de mesure (répondant, enquêteur, communication entre eux) et de traitement (codage et introduction des données), que l'on rencontre aussi dans les enquêtes exhaustives ... rappelons que les fichiers administratifs ne sont pas exempts d'erreurs.

Fardeau de réponse L'interview (durant de 5 à 10 minutes par personne) représente une charge; certains ménages nous disent «Mais vous avez déjà tout ça dans vos ordinateurs!»²¹

Principales qualités de l'EFT

Comparabilité Les définitions internationales du BIT, souvent précisées par Eurostat, sont appliquées strictement, facilitant les comparaisons internationales.

Actualité Les résultats doivent être transmis à Eurostat 12 semaines après la fin du trimestre, alors que certains fichiers administratifs ne sont définitivement stabilisés qu'après un an ou deux.

Complétude L'interview permet de déterminer l'emploi (national) des salariés et non-salariés, y compris –en théorie– le travail occasionnel non déclaré, tout en détectant les cumuls d'activité (sans devoir apparier des fichiers)

Différences entre source EFT et sources Sécurité Sociale

Concepts différents Pour certaines situations marginales, l'interprétation «administrative» peut différer de l'interprétation «enquête»²².

Mesures différentes Pour un même concept (NACE, durée du travail,...) les fichiers de la sécurité sociale (alimentés par l'employeur) et le travailleur (participant à l'enquête) peuvent répondre différemment.

Moments différents Tandis que l'EFT est continue (52 semaines de référence par année), les institutions de sécurité sociale ne fournissent que des statistiques trimestrielles – voire moins fréquentes.

Champs différents Alors que l'EFT mesure l'emploi national (actifs occupés qui habitent le Royaume, quel que soit leur lieu de travail), les fichiers de la sécurité sociale indiqueront l'emploi intérieur (actifs occupés en Belgique, y habitant ou non), ce qui impose des corrections.

Spécificités de l'EFT comme enquête auprès des ménages

Résultats par ménage Utile intrinsèquement (le taux de «jobless households» figure parmi les indicateurs d'exclusion sociale), et pour le benchmarking d'autres enquêtes de l'INS (Tourisme, Budget des Ménages); mais SEC95 s'intéresse surtout aux individus.

Possibilité de proxy La faculté de répondre pour un autre membre du ménage améliore le taux de réponse (... aux dépens de la qualité?).

Biais de participation Le taux de réponse dépend de la composition du ménage (voir II.6)²³.

Une idée naturelle consiste –plutôt que de les juxtaposer²⁴– à **intégrer sources administratives et enquête**. Un cours TES (Strasbourg, 1999) a fait découvrir à F. Montaigne et J.M. Museux les techniques de calibrage et l'information auxiliaire²⁵, permettant d'améliorer grâce à des données externes l'extrapolation des enquêtes (par sondage, et comportant donc un caractère aléatoire

²¹On considère (en notre ère de simplification administrative) qu'il vaut mieux –à qualité égale– recourir aux sources existantes qu'à une enquête (et si une enquête s'impose, un sondage est préférable à un recensement). Néanmoins, certaines données subjectives de l'EFT (p. ex. recherche d'un autre emploi, et dans ce cas, motif) ne peuvent être recueillies que par interview.

²² Exemples:

-Les parents d'accueil sont «indemnisés»: s'agit-il d'une activité rémunérée ou non? Au moment où ces lignes sont écrites, une clarification du statut est annoncée, qui peut inciter davantage les répondants à estimer avoir un emploi.

-Une personne prestant des services en ALE une semaine sur deux peut-elle être considérée comme temporairement absente de son emploi les autres semaines?

²³ Par exemple, si la famille compte un enfant en bas-âge, la probabilité qu'un des parents au moins (pouvant le cas échéant répondre au nom du conjoint) soit présent est plus élevée que pour les couples sans enfants.

²⁴ Même si la pluralité des sources a ses partisans...

²⁵ Nous préférons «calage» ou «calibrage» à l'anglicisme «calibration».

pouvant être aggravé par le phénomène de non-réponse²⁶). Pensant que l'EFT serait un bon champ d'application, le second cité a exposé l'idée à la cellule de coordination et introduit le dossier à la BCSS. Si nous conservons²⁷ les réponses à l'EFT, l'information de la sécurité sociale interviendra dans le calcul des poids.

Ce projet n'aurait pas abouti sans **outil informatique** adapté. Trois membres du Bureau des Méthodes et de Coordination (J.M. Museux lui-même, C. Vanderhoeft et E. Waeytens) avaient perçu l'utilité d'un programme SPSS (le logiciel statistique répandu à l'INS) [voir VWM] analogue à CALMAR (SAS) dû à l'INSEE. C'est à C. Vanderhoeft que revient le mérite d'avoir mis au point g-Calib (voir entre autres [Vanderhoeft] et ses liens). Le logiciel a été rendu plus convivial depuis lors, avec l'aide de L. Préal et P. Celis.

1.5 Données de la Sécurité Sociale et dossier BCSS

L'accès aux données personnelles ne va pas de soi (Article 22 de la Constitution: Chacun a droit au respect de sa vie privée et familiale, sauf dans les cas et conditions fixés par la loi).

L'INS a dû motiver sa demande, et se restreindre au strict nécessaire, à savoir:

les données individuelles pour les membres de l'échantillon, qui transitent par la Banque Carrefour de la Sécurité Sociale;

les données globales (totaux pour la population), transmises directement par les différentes institutions.

Pour le calibrage, il est essentiel que mêmes définitions soient utilisées pour les données individuelles et globales, et que ces données soient déchargées au même moment. En revanche, certaines analyses (liaisons entre variables d'enquête et variables de la sécurité sociale) ne nécessitent que les informations individuelles, d'autres (comparaisons de totaux) sont possibles avec les seules informations globales.

Six institutions²⁸ (AdP-INASTI-ONEM-ONP-ONSS-ONSSAPL) ont été sollicitées, l'INS leur demandant une vingtaine de variables. En septembre 2000, le Comité de Surveillance près la BCSS a donné son feu vert.

Les premières données reçues portaient sur l'échantillon 1999, et donnaient la situation au 31 décembre de ladite année. Depuis, le calendrier (convenu entre l'INS, la BCSS et les six institutions) est le suivant:

TABLEAU I.3 CALENDRIER DES VARIABLES DE LA SECURITE SOCIALE

Institution	Fréquence	Données de...	Livrées après...
ONSS	2 fois/an	deuxième et quatrième trimestres	7 mois
ONSSAPL	2 fois/an	deuxième et quatrième trimestres	7 mois
INASTI	1 fois/an	quatrième trimestre	6 mois
ONEM	2 fois/an	juin et décembre	5 mois
ONP	1 fois/an	quatrième trimestre	6 mois
AdP	1 fois/an	quatrième trimestre	6 mois

Le flux des données individuelles (Schéma S002) comporte un aller-retour: l'INS transmet, via la BCSS, les numéros nationaux²⁹ des personnes (ménages effectifs et ménages de réserve) interrogeables; ces institutions nous renvoient (via la BCSS) les informations individuelles.

²⁶ Répondre impliquant: *être tiré, et être visité, et être présent, et accepter*, voir chapitre II.

²⁷ Ce principe connaît une exception: l'activité NACE de l'entreprise (voir IV.4).

²⁸ Le glossaire des institutions figure en fin de rapport.

²⁹ Le NISS (numéro d'identification à la sécurité sociale) concorde avec le numéro national pour les résidents du Royaume.

Ensuite, une syntaxe ODBC due à J.M. Museux permet d'adjoindre ces informations de la sécurité sociale aux données de l'enquête, créant un fichier SPSS baptisé «bigfile», qui (après corrections et imputations éventuelles) sera introduit dans g-Calib.

La fusion des fichiers dans «bigfile» s'opère sur la base des numéros de groupe (2392 par an), de ménage dans le groupe, et de personne au sein du ménage. Un phénomène (ennuyeux sur le plan des analyses, même si les répercussions sur les résultats globaux sont faibles) est apparu, celui des permutations. En effet, dans le fichier d'enquête, la personne de référence «socioéconomique» du ménage est la personne active occupée la plus âgée ou, à défaut, celle qui apporte les principales ressources. Cette définition ne correspond pas nécessairement à la définition «administrative» retenue dans les registres communaux et donc dans le registre national qui en est dérivé et sur la base duquel le formulaire de ménage a été imprimé, à savoir: «la personne de référence est le membre du ménage qui est habituellement en contact avec l'administration pour les affaires concernant le ménage». A l'issue des entretiens, l'enquêteur vérifie si la «personne de référence socioéconomique» coïncide avec la «personne de référence du registre national» et adapte si nécessaire le formulaire de ménage; dans ce cas, la fusion attribuera à un individu les données BCSS d'un autre membre du ménage (voir IV.2). Bien que le problème soit limité (moins de 1% des individus), il serait évidemment désirable de le résoudre.

Les totaux que reçoit l'INS sont en principe calculés par commune (de résidence)³⁰, croisée avec le sexe et la catégorie d'âge (ce qui, dans les petites municipalités, nous met à la limite du secret...). Pour calibrer, nous les agrégeons par NUTS II.

Quelques problèmes pratiques ont du être rencontrés, sur le plan... :

de la simultanéité: selon les institutions, la livraison des totaux précède ou suit celle des informations individuelles; ce qui signifie simplement que nous devons attendre (un écart temporel entre le déchargement des bases de données individuelles et le calcul des totaux serait beaucoup plus ennuyeux);

du format: une des institutions nous avait d'abord envoyé des totaux (difficiles à lire vu la taille énorme du fichier) à un niveau de détail dépassant ce qui était demandé, par la suite, une version simplifiée a été établie, où l'agrégation se fait par arrondissement, de manière donc moins fine que convenue mais suffisante puisque nous calibrons par NUTS II;

des définitions: les totaux d'une autre institution semblaient anormalement élevés, et il est apparu que certains individus y étaient comptés plusieurs fois, alors qu'au niveau individuel, les valeurs possibles n'étaient que 0 ou 1. Grâce entre autres aux précisions fournies par J.P. Georges, ces totaux ont été corrigés (en général et sauf erreur manifeste, par exemple une brusque rupture d'un trimestre à l'autre, les totaux sont supposés corrects!).

Voici la liste des variables, sur laquelle nous reviendrons au Chapitre III.

Les variables ONSS et ONSSAPL sont parallèles, avec une exception: l'ONSS nous communique la NACE codée à 4 chiffres, l'ONSSAPL se limite à la division (code de 2 chiffres... qui est toujours 75).

³⁰ Par conséquent, «nos» totaux de l'ONSS, de l'ONSSAPL et (dans une moindre mesure) de l'INASTI sont inférieurs à ceux généralement diffusés par ces institutions, qui incluent les frontaliers entrants. Rien de grave, tant la différence peut être expliquée, et que le même concept est utilisé pour les totaux et les membres de l'échantillon.

TABLEAU I.4 LISTE DES VARIABLES DE LA SECURITE SOCIALE

ORGANISME	VARIABLE	VALEURS POSSIBLES et REMARQUES
ONSS (LATG)	NISS = numéro national	0: inconnu 1: connu
ONSS (LATG)	Personne inscrite en fin de trimestre?	0: non 1: oui
ONSS (LATG)	Code travailleur chez l'employeur (principal) Si plusieurs employeurs, le principal est celui redevable du salaire le plus élevé	1: employé 2: employé (Agent Contractuel Subventionné) 3: ouvrier 4: ouvrier (Agent Contractuel Subventionné) 5: statutaire 6: autre
ONSS (LATG)	Code NACE de l'employeur (principal)	NACE 4 chiffres
ONSS (LATG)	Volume de travail (total des employeurs)	Somme sur le trimestre de tous les ETP (tels que calculés par l'ONSS)
ONSS (LATG)	Salaire (total des employeurs)	Somme sur le trimestre de tous les salaires bruts (primes exclues)
ONSSAPL	NISS = numéro national	0: inconnu 1: connu
ONSSAPL	Personne inscrite en fin de trimestre?	0: non 1: oui
ONSSAPL	Code travailleur chez l'employeur (principal) Si plusieurs employeurs, le principal est celui redevable du salaire le plus élevé	1: employé 2: employé (Agent Contractuel Subventionné) 3: ouvrier 4: ouvrier (Agent Contractuel Subventionné) 5: statutaire 6: autre (pompiers volontaires, aumôniers, ...).
ONSSAPL	Code NACE de l'employeur	NACE 2 chiffres (Toujours 75)
ONSSAPL	Volume de travail (total des employeurs)	Somme sur le trimestre de tous les ETP
ONSSAPL	Salaire (total des employeurs)	Somme sur le trimestre de tous les salaires bruts (primes exclues)
INASTI	NISS = numéro national	0: inconnu 1: connu
INASTI	Qualité	1: indépendant 2: aidant
INASTI	Catégorie de cotisations	1: activité principale 2: activité secondaire 3: actif après pension
ONP	NISS = numéro national	0: inconnu 1 connu
ONP	Pensionné à un moment du trimestre?	0: non 1: oui
ADP	NISS = numéro national	0: inconnu 1: connu
ADP	Pensionné à un moment du trimestre?	0: non 1: oui
ONEM	NISS = numéro national	0: inconnu 1: connu
ONEM	Dernière situation durant le mois (de référence) Il s'agit uniquement des personnes indemnisées par l'ONEM, à l'exclusion p.ex. des demandeurs d'emploi libres inscrits au FOREM/ORBEM/VDAB	0: inconnu (aucune allocation pour le mois visé) 1: sans emploi demandeur d'emploi 2: sans emploi non-demandeur d'emploi 3: travailleur qui cumule avec une allocation 4: prépensionné à temps plein 5: prépensionné à temps partiel 6: interruption complète d'un emploi à temps plein 7: interruption complète d'un emploi à temps partiel 8: interruption de carrière - prestations réduites

II Sondages et extrapolations

II SONDAGES ET EXTRAPOLATIONS

II.1 Sources d'erreurs inhérentes à un sondage

Dans les enquêtes par sondage, on est confronté à quatre causes principales d'erreur [Tillé p. 254]:

- ❑ Couverture
- ❑ Mesure
- ❑ Echantillonnage
- ❑ Non-réponse

La **couverture** de l'EFT est satisfaisante: la base de sondage (le registre national, restreint aux ménages privés, mais sans restriction quant à la région linguistique¹) reprend la grosse majorité de la population-cible, à savoir les (membres des) ménages (privés) résidant en Belgique². Observons que cette population change constamment (naissances et immigration, décès et émigration). Quant aux «sans domicile fixe», soit ils ne figurent pas (plus) dans la base de sondage (défaut de couverture), soit ils y sont encore mentionnés mais contribueront à la non-réponse.

Des efforts sont faits pour améliorer la qualité de la **mesure**³: questionnaire dans les trois langues officielles et version anglaise officieuse, formation écrite et orale des enquêteurs, questions et modalités de réponses aussi claires que possible, adaptation aux (modifications des) interprétations Eurostat, personnel qualifié pour la codification. Certains points noirs subsistent, notamment la question 6 relative à l'interruption de carrière/crédit-temps.

Bien que nous essayions d'améliorer les techniques d'extrapolation, notre enquête est et restera sujette à des fluctuations d'**échantillonnage**.

La complexité de l'extrapolation est accrue par le phénomène de **non-réponse**, dont le mécanisme, en fonction de facteurs géographiques (province, caractère urbain ou rural), démographiques (âge et sexe, mais aussi nationalité ou état-civil, sans oublier la composition du ménage) et autres (situation d'emploi... que l'enquête vise justement à déterminer!) reste obscur. L'INS s'efforce (notamment en prévenant les ménages et en garantissant l'observation du secret statistique) de comprimer le taux de refus.

Nous supposons dans la suite la couverture parfaite et la mesure correcte, et nous concentrerons sur l'échantillonnage et la non-réponse.

Histoire de rendre la discussion plus concrète, le statut de chômage BIT (ou non) sera pris comme caractère d'intérêt.

¹ Certaines enquêtes de l'Institut ne couvrent pas la région de langue allemande.

² *Le champ principal de l'enquête est constitué par la population des ménages privés résidant sur le territoire économique de chaque État membre. Si possible, ce champ principal constitué par la population des ménages privés est complété par la population des ménages collectifs* [Règlement 577/98, art.2 §2]. En Belgique, seuls les ménages privés sont susceptibles d'être interrogés, mais on extrapole à la population totale, en admettant que la post-stratification redresse l'exclusion des membres des ménages collectifs (surtout des résidences pour personnes âgées), étant donné que leur situation d'(in)activité correspond à celles des individus (de même âge) faisant partie de ménages privés.

³ Et à moyen terme, développement du questionnaire électronique (CAPI = *Computer Aided Personal Interview*).

II.2 Si l'EFT était un recensement...

Si l'enquête était exhaustive (type recensement) et exempte de non-réponse, les informations recueillies par les enquêteurs suffiraient, sans la moindre incertitude de sondage (variance nulle), à estimer (déterminer!) le nombre de chômeurs BIT⁴ dans la population:

$$\hat{y}_0 = \sum_{k \in U} y_k$$

où \hat{y}_0 représente l'estimation (le calcul!) du **nombre** ou du **total** Y de chômeurs, l'indicatrice y_k vaut 1 ou 0 selon que l'individu k est chômeur ou non, et la sommation est étendue à l'ensemble de la population (univers = base de sondage = échantillon = répondants): nous pouvons numéroter (arbitrairement) de 1 à N (de l'ordre de 10.200.000, en incluant les enfants de 0 à 14 ans) les personnes (qui participent toutes à l'enquête), formant notre univers U , censé coïncider avec la base de sondage. La seule information essentielle réside dans les réponses (supposées exactes) y_k . Pour estimer (déterminer!) la **proportion** P de chômeurs dans la population, il suffit de diviser le total des chômeurs par la population; afin de mesurer celle-ci, nous disposons de deux possibilités (en principe équivalentes)

(source intrinsèque) le nombre de répondants ($\sum_{k \in U} 1$)

(source auxiliaire) la population indiquée par le registre national (N)

Nous venons ainsi de faire connaissance avec la notion de **donnée auxiliaire**, issue d'une source autre que l'enquête: ici, la population totale (que le registre national ventile par âge, sexe, nationalité, état civil et autres caractéristiques, comme la composition du ménage; en revanche, le statut de chômage BIT n'y figure pas).

II.3 Plan de sondage

Pour d'évidentes raisons pratiques, une enquête continue sur les forces de travail n'est réalisable que sous la forme d'un sondage.

Le plan de sondage (pour les spécialistes, une mesure de probabilité sur l'ensemble des parties non vides de l'univers) de l'enquête sur les forces de travail est relativement complexe.

La stratification se fait par NUTS II; le nombre de ménages à tirer trimestriellement dans chaque province est déterminé (les provinces les moins peuplées sont proportionnellement plus sondées).

Au sein des NUTS II, on tire des sections (parties de communes) comme unités primaires, proportionnellement au nombre de ménages (répétitions possibles) et on leur attribue une semaine de référence parmi les 13 du trimestre.

Les unités secondaires sont les ménages; dans chaque section sélectionnée, on tire 20 ménages effectifs ainsi que 8 ménages de réserve, formant un groupe et confiés à un même enquêteur.

Ces ménages peuvent enfin être considérés comme grappes d'individus, ceux-ci formant les unités tertiaires.

Il s'agit d'un plan de sondage probabiliste (chaque individu a une probabilité non nulle d'être tiré), à plusieurs degrés et à probabilités inégales⁵; cependant, dans une province donnée, chaque ménage (et donc chaque individu) a la même probabilité d'être tiré⁶ – ce qui ne signifie pas la même

⁴ Idées similaires pour toute variable nominale (p. ex. salarié ou non), et analogues pour les variables numériques (durée du travail).

⁵ On peut comparer avec un atlas: rien n'empêche que les pays soient représentés à différentes échelles – pour autant que l'échelle soit toujours indiquée, et chaque pays y figure.

⁶ Au sein de la NUTS II, le plan est donc autopondéré (comparer [Tillé p. 179]).

probabilité de répondre⁷. La non-réponse constitue une complication sérieuse, que le tirage de ménages de réserve essaye de résoudre... en posant de nouveaux problèmes⁸.

En prenant toujours l'individu comme unité de base, nous numérotions les personnes **susceptibles de participer** à l'enquête, qui forment notre univers U . Le plan de sondage, suivi du tirage, aboutit à la sélection de l'échantillon trimestriel s (n -uple non ordonné, sans remise⁹) [n voisin de 38.000, y compris les membres des ménages de réserve]. C'est alors que peut commencer le travail de terrain («enquête» proprement dite).

L'ensemble r des répondants est contenu dans s .

II.4 Estimateurs

Après le travail de terrain, le problème statistique fondamental consiste, en fonction d'une information partielle (les y_k , pour $k \in r$), à estimer le total (plutôt, sa moyenne trimestrielle, car l'enquête est continue) des chômeurs dans l'univers $Y = \sum_{k \in U} y_k$. Puisque la population totale est

connue grâce au registre national, chaque estimation du total équivaut à une estimation de la proportion.

Tous les estimateurs (variables aléatoires, car dépendant des répondants... et éventuellement d'autres critères) décrits ci-après seront linéaires¹⁰, autrement dit de la forme:

$$\hat{y} = \sum_{k \in r} w_k y_k$$

où les w_k sont des nombres réels (que le praticien préfère positifs), appelés poids ou coefficients d'extrapolation; w_k peut dépendre non seulement de k lui-même mais aussi du plan de sondage (avant tirage) et/ou de l'échantillon obtenu (après tirage), ainsi que du plan de réponse¹¹ et/ou des répondants. Bien que non obligatoire théoriquement, il semble naturel d'utiliser le même système de coefficients pour toutes les estimations (pas seulement du chômage, mais aussi des autres caractères).

Chaque choix d'une famille de poids ($w_k : k \in r$) définit un estimateur, et nous cherchons «le meilleur» selon différents critères:

précision: on cherche à minimiser l'erreur quadratique moyenne, somme des carrés du biais¹² et de l'écart type¹³, en d'autres termes, on souhaite que les estimations avoisinent la vraie valeur;

⁷ Article 3 de la loi du 4 juillet 1962 relative à la statistique publique: [...] *les personnes appelées à répondre sont désignées [...] suivant une méthode impliquant, pour toutes les personnes comprises dans une même catégorie, la même probabilité d'être astreintes à déclarer.*

Dans notre cas, cette probabilité dépend de la NUTS II; le sur-échantillonnage des provinces les moins peuplées répond à un critère de précision [Règlement 577/98, art.3 §1] *Pour un groupe de chômeurs représentant 5 % de la population d'âge actif, l'erreur-type relative pour l'estimation des moyennes annuelles [...] ne doit pas excéder 8 % au niveau de la NUTS II de la sous-population en cause.*

Remarquons qu'aucune référence n'est faite à un éventuel biais.

⁸ *La substitution de ménages de réserve aux non-réponses est généralement la source de biais importants. Cette substitution doit être évitée* [Avignon]. La Belgique a pourtant maintenu les réserves, pour des raisons pratiques. Dans l'extrapolation (post-stratification ou calibrage), nous ne distinguerons pas ménages effectifs et ménages de réserve.

⁹ La probabilité qu'un individu (p. ex. après déménagement) soit interrogé plusieurs fois au cours d'une année donnée, a fortiori le même trimestre, peut être négligée.

¹⁰ Selon la tendance moderne; [Tillé p. 40] parle de «linéaires et homogènes».

¹¹ Audacieuse expression de [Tillé p. 256].

¹² On se restreint parfois aux estimateurs non biaisés (qui ne tendent pas à sur- ni à sous-estimer), dont l'erreur quadratique moyenne n'est autre que la variance.

cohérence: les «estimations» de totaux déjà connus grâce à une autre source (revoici les données auxiliaires!) devraient correspondre, par exemple, dans un croisement «chômage BIT x sexe». Exposons d'abord une situation idéalisée (supposant la réponse totale: $r = s$)¹⁴, avant de nous pencher sur la dure réalité ($r < s$) [le nombre trimestriel de répondants est environ 26.000].

II.5 Extrapolation en l'absence de non-réponse

Supposons (provisoirement et pédagogiquement) que chaque individu tiré participe effectivement ($r = s$). On pourrait faire l'hypothèse que la proportion de chômeurs BIT dans la population peut être estimée par cette proportion dans l'échantillon s

$$\sum_{k \in U} \frac{y_k}{N} = \sum_{k \in r} \frac{y_k}{n} = \bar{y}$$

et donc prendre comme **estimateur «naïf»** pour le total

$$\hat{y}_1 = \frac{N}{n} \sum_{k \in r} y_k$$

mais cet estimateur est critiquable. Ainsi, on pourrait (soupçonner l'INS de) manipuler à la baisse l'estimation du chômage en modifiant le plan de sondage afin de tirer plus (ou moins) de personnes dans certaines provinces (ou tranches d'âge), supposées moins (ou plus) exposées au chômage.

* * *

Pallions cet inconvénient. Le plan de sondage assigne à chaque individu une certaine probabilité de sélection, notée π_k pour l'individu k ($\pi_k = P(k \in s) = P(k \in r) > 0 \forall k$).

Nous définissons (plus exactement, ils ont défini [HT]) **l'estimateur de Horvitz-Thompson**

$$\hat{y}_2 = \sum_{k \in r} \frac{y_k}{\pi_k}.$$

Les ingrédients sont

les réponses des personnes interrogées (y_k)

les probabilités de sélection, dépendant du plan de sondage (π_k)

l'échantillon s (le seul élément aléatoire!).

Par exemple, dans la province de Luxembourg (101.247 ménages au 1 janvier 2002), le plan de sondage stipule que 1120 ménages (800 effectifs + 320 réserves) sont tirés par trimestre, ce qui donne, pour chaque ménage –et ipso facto chaque individu– une probabilité de sélection au premier trimestre de $1120/101.247 = 0,011 = \pi_k$.

L'inverse de π_k , souvent noté d_k , est appelé **poids de sondage** de l'individu k . Il dépend (fortement) de la province et (faiblement) du trimestre, mais ni des caractéristiques individuelles (âge, sexe,...), ni de la composition (nombre de membres, âge moyen...) du ménage.

Sur le plan de la précision, la principale vertu de cet estimateur est son absence de biais pour notre variable d'intérêt (... et les autres variables), c'est-à-dire que son espérance coïncide avec la vraie valeur. Sa variance peut en principe être estimée (sur la base des probabilités jointes¹⁵ π_{jk}); les formules sont cependant assez compliquées.

¹³ Un calcul –ou plus souvent, une estimation– de l'écart type, assorti(e) de l'hypothèse (acceptable pour de grands échantillons, eu égard aux versions appropriées du théorème central-limite et au faible taux de sondage estompant la différence entre sondage sans ou avec remise) de normalité approximative pour la distribution, permettra de construire des intervalles de confiance.

¹⁴ Et absurde, puisque l'enquêteur ne contacte les ménages de réserve qu'à défaut de ménages effectifs.

¹⁵ Dans le cas d'un sondage par grappes (ici par ménages) le tirage de l'individu j équivaut à celui de l'individu k s'ils appartiennent à la même grappe.

L'estimateur de Horvitz-Thompson présente pourtant des inconvénients¹⁶. en ce qui concerne la cohérence. Posons $z_k = 1$ pour chaque individu de l'échantillon; dans ce cas la variable aléatoire (car dépendant de l'échantillon)

$$\sum_{k \in s} \frac{z_k}{\pi_k} = \sum_{k \in s} \frac{1}{\pi_k}$$

sera un estimateur sans biais de la population totale... laquelle est connue. Et un statisticien ou un actuaire doit-il estimer ce qu'il connaît déjà –même si en moyenne son estimation coïncidera avec la vraie valeur¹⁷?

Une solution serait de stratifier (davantage ou à plusieurs degrés) l'échantillon, en modifiant le plan de sondage afin d'imposer (outre un nombre fixe de ménages) une condition sur leur nombre de membres, mais elle est difficile à réaliser en pratique. Et si nous désirons que les marges démographiques connues des tableaux croisés tels «statut BIT x sexe x âge x province» concordent avec les données (du registre national) dont nous disposons, équilibrer en plus le tirage sur le sexe et/ou l'âge deviendrait une gageure (toujours sous notre hypothèse idéale de réponse totale!).

* * *

C'est pourquoi (sans modifier le plan de sondage décrit au §3) nous allons donner la priorité à la cohérence. Prenons comme post-strates les croisements pour lesquels la cohérence démographique nous importe (province x âge quinquennal¹⁸ x sexe, soit $11 \times 16 \times 2 = 352$ par trimestre dans notre cas) et définissons **l'estimateur post-stratifié**

$$\hat{y}_3 = \sum_{k \in s} \frac{y_k}{p_k}$$

où la sommation porte toujours sur les répondants $k \in s$ ($s =$ échantillon), et p_k est le rapport entre le nombre de répondants de la post-strate à laquelle k appartient et la population de celle-ci. On peut considérer p_k comme une version empirique de π_k : au lieu de la probabilité de tirage (déterminée a priori par le plan de sondage, et fonction de la province), on observe la proportion (la fréquence relative) de sondés (dépendant aussi de la classe d'âge et du sexe).

Une expression alternative de l'estimateur post-stratifié est

$$\hat{y}_3 = \sum_{h \in H} N_h \bar{y}_h$$

où H est l'ensemble des (352) post-strates, N_h représente la population de la post-strate h et \bar{y}_h la

proportion de chômeurs $\bar{y}_h = \frac{\sum_{k \in h} y_k}{n_h}$. Le lecteur sagace aura deviné que n_h égale l'effectif de

l'échantillon appartenant à h , qu'il est donc important que $n_h \neq 0$, et que le choix de post-strates trop fines (arrondissements au lieu de la province et/ou années au lieu de l'âge quinquennal) augmenterait le risque de n'avoir aucun répondant dans une catégorie¹⁹. De même, on pourrait faire intervenir la semaine d'enquête comme variable de post-stratification, pour redresser le fait que l'échantillon (personnes sélectionnées) varie légèrement au cours du trimestre. Si les effets sur le

¹⁶ L'exemple folklorique des *éléphants de Basu* en est un autre: un estimateur sans biais peut être entaché d'une grande variance [Tillé p. 44].

¹⁷ Deux remarques:

Hájek [Tillé p. 38] propose une solution... mais qui soulève d'autres difficultés;

la loi des grands nombres (généralement utilisée pour un sondage avec remise, nous travaillons sans remise, ce qui renforce la conclusion) nous assure qu'avec une grande probabilité, la population estimée sera proche de la population réelle, mais «probablement proche» ne nous suffit pas.

¹⁸ Avec regroupement des individus de 75 ans et plus.

¹⁹ Ou bien un seul répondant qui se verrait attribuer un poids élevé, faisant grimper la variance.

statut BIT semblent assez faibles, les résultats portant sur la durée effective (au cours de la semaine de référence) du travail seraient peut-être affinés.

Autrement dit, nous découpons la population en catégories supposées homogènes, utilisons notre estimateur naïf pour estimer dans chacune le nombre de chômeurs, avant de sommer sur ces catégories. Le procédé garantit la cohérence avec les informations démographiques, la précision étant tributaire de la non-vacuité des post-strates [Tillé p. 191], laquelle est «presque» garantie par notre choix de la province et de l'âge quinquennal.

II.6 La non-réponse

Redevenons réaliste: certains individus tirés au sort ne participent pas à l'enquête (réserve non utilisée²⁰, décès ou départ du Royaume²¹, refus, absence de longue durée, barrière linguistique, défaillance de l'enquêteur, ...)

Deux précisions relatives à l'EFT:

l'enquête n'est considérée comme positive que si les renseignements ont été recueillis (éventuellement par proxy) pour tous les membres du ménage âgés de 15 ans et plus; similairement, un formulaire individuel n'est accepté que si on a répondu à tous les items indiqués par le cheminement²² (à l'exception éventuelle des questions 82, 83 et 84); l'enquêteur peut être tenté de recourir à *l'imputation sauvage*, selon l'expression de J. Langers (Statec)).

Est-il possible de définir un profil des (non-)répondants? Les pourcentages qui suivent portent sur le fichier complet (répondants et non-répondants, membres des ménages effectifs et de réserve) de 2001, non pondéré.

Le taux de réponse global atteint 68,1% (n'oublions pas que tous les ménages, de réserve comme effectifs, figurent au dénominateur).

Le *TABLEAU II.1* donne les extrêmes, selon différentes caractéristiques (la première saisonnière, la deuxième géographique, les quatre suivantes démographiques individuelles, la sixième démographique collective, les deux dernières socio-économiques transmises par la BCSS et déjà abordées au Chapitre I)

²⁰ Il peut paraître bizarre de considérer une réserve non utilisée comme un cas de non-réponse...mais cela nous permet de considérer tout répondant potentiel les répondants comme partie de l'échantillon initial. D'ailleurs, les numéros nationaux de tous les membres des ménages (effectifs et réserves) sélectionnés sont transmis à la BCSS (au début du trimestre, donc avant de savoir qui répondra).

²¹ Signifiant que l'individu a quitté le champ de l'enquête.

²² Certaines questions admettent «ne sait pas» comme modalité de réponse.

TABLEAU II.1 TAUX DE REPONSE (MIN/MAX) SELON CERTAINES CARACTERISTIQUES (ENSEMBLE)

TAUX DE RÉPONSE PAR...	% MINIMUM	% MAXIMUM
Trimestre	67,7 (Troisième)	68,5 (Premier)
NUTS II	62,3 (Bruxelles)	70,3 (Limbourg)
Sexe	66,9 (Hommes)	69,2 (Femmes)
Age quinquennal	58,1 (30-34 ans)	85,2 (70-74 ans)
Etat civil	62,7 (Célibataires)	79,6 (Veufs)
Groupe de nationalités ²³	59,0 (Etrangers non UE)	68,9 (Belges)
Nombre de membres du ménage ²⁴	62,6 (6 personnes)	75,3 (2 personnes)
Présent ONSS fin de trimestre?	64,1 (oui)	69,3 (non)
Statut ONEM	60,0 (sans emploi demandeur)	79,3 (prépension pl. temps)

Bien entendu, ces caractéristiques ne sont pas indépendantes: ainsi, les veufs (et surtout les veuves) sont nombreux (nombreuses) dans la catégorie des 70-74 ans, âge auquel la présence à l'ONSS ou à l'ONEM est rarissime.

Pour rendre la comparaison plus pertinente, nous avons recalculé ces pourcentages (*TABLEAU II.2*) en nous restreignant aux âges entre 20 et 59 ans, et en excluant les réserves non utilisées (nous limitant donc aux ménages effectifs et aux réserves que l'enquêteur a au moins cherché à contacter), ce qui élève le taux global à 79%. Ce tableau confirme certaines conclusions du précédent (taux de réponse par province²⁵, par sexe ou par nationalité²⁶) en modifie d'autres (la classe d'âge la moins répondante devient celle des 25-29), et inverse (par l'exclusion des plus jeunes et des plus vieux) celle touchant le statut ONSS.

Ici comme ailleurs, la prudence s'impose (liaison entre situation individuelle «célibataire» et collective «ménage d'une personne»; Bruxelles cumulant caractère urbain et proportion importante d'étrangers; réponse ou non de tout le ménage) à celui qui aimerait expliquer ces phénomènes.

TABLEAU II.2 TAUX DE REPONSE (MIN/MAX) SELON CERTAINES CARACTERISTIQUES (20-59 ANS, HORS RÉSERVES NON UTILISÉES)

TAUX DE RÉPONSE DES 20-59 ANS, HORS RÉSERVES NON UTILISÉES, PAR	% MINIMUM	% MAXIMUM
Trimestre	78,3 (Deuxième)	79,8 (Quatrième)
NUTS II	65,3 (Bruxelles)	90,4 (Limbourg)
Sexe	77,9 (Hommes)	80,0 (Femmes)
Age quinquennal	72,5 (25-29 ans)	84,4 (55-59 ans)
Etat civil	73,8 (Célibataires)	82,0 (Mariés)
Groupe de nationalités	65,9 (Etrangers non UE)	80,6 (Belges)
Nombre de membres du ménage	62,3 (1 personne)	83,1 (4 personnes)
Présent ONSS fin de trimestre?	77,5 (non)	80,3 (oui)
Statut ONEM	72,3 (sans emploi demandeur)	88,3 (prépension pl. temps)

²³ Trois groupes: Belges – Autres Union Européenne – Hors Union Européenne.

²⁴ Limité aux ménages de 8 personnes au plus. Signalons en passant une corrélation positive (... et suspecte) entre nombre de membres et fréquence des réserves non utilisées.

²⁵ L'écart entre Bruxelles et Limbourg, comblé dans le tableau 1 par les réserves, est manifeste dans le tableau 2.

²⁶ Barrière linguistique?

II.7 Extrapolation en présence de non-réponse

Sur un plan général, les trois réactions principales face à la non-réponse sont: insister (visites répétées, courriers comminatoires), attitude de l'INS vis à vis des entreprises mais pas des ménages;

imputer, autrement dit «deviner» quelle aurait été la réponse (en fonction d'un modèle, ce qui est politiquement délicat pour un Institut de statistique officiel);

repondérer, c'est-à-dire supposer que le comportement des répondants (d'une catégorie déterminée) représente celui des non-répondants (de la même catégorie), ce que nous allons essayer de préciser²⁷.

Considérons au sein de s le sous-ensemble r des répondants. Grande est la tentation d'ajouter un degré au sondage (une fois s déterminé, «tirer au sort» –pas nécessairement à probabilités conditionnelles égales!– le sous-ensemble r par un «plan de réponse») et de poser

$$P(k \in r) = P(k \in s) \cdot P(k \in r | k \in s)$$

c'est-à-dire exprimer la probabilité de réponse ρ_k comme produit de la probabilité de sélection $\pi_k = P(k \in s)$ et de la probabilité conditionnelle de répondre, à condition d'être sélectionné, notée $\theta_k = P(k \in r | k \in s)$ et d'introduire un **estimateur d'Horvitz-Thompson modifié**

$$\hat{y}_4 = \sum_{k \in r} \frac{y_k}{\rho_k} = \sum_{k \in r} \frac{y_k}{\pi_k \theta_k}.$$

Cette formule fait intervenir

les réponses des personnes interrogées (y_k)

les probabilités de sélection dépendant du plan de sondage (π_k)

les probabilités de réponse des individus (sélectionnés) (θ_k)

l'ensemble r des répondants (ceci veut dire que nous n'imputons pas).

Hélas, les θ_k –en admettant qu'il s'agisse de probabilités– ne sont pas indiqués par le plan de sondage, et difficiles à calculer²⁸, car la non-réponse peut être de trois types différents [Tillé p. 256], par ordre de gravité croissante

uniforme (ne dépendant ni de la variable d'intérêt, ni des variables auxiliaires, par exemple une défaillance de l'enquêteur), comme si le plan de sondage nous disait «présélectionnons d'abord 1000 ménages, puis tirons-en 900 au hasard»;

aléatoire (liée à des variables auxiliaires connues, telles l'âge, mais indépendante de la variable d'intérêt), ce que la post-stratification compensera;

confondue (liée à la variable d'intérêt, ici le statut BIT), risquant de biaiser les résultats.

* * *

Que fait un statisticien ou un actuairer face à un paramètre inconnu comme θ_k ? estimer!

Si 26.000 des 38.000 personnes tirées ont répondu, on peut considérer (en première approximation uniforme) que $P(k \in r | k \in s) = \theta_k = 0,68$ pour tout k . Puisque les caractéristiques démographiques peuvent influencer sur les taux de réponse θ_k , et que π_k dépend de la province, nous calculerons une estimation r_k de $\rho_k = \pi_k \theta_k$ par classe d'âge, par sexe, et par province. Par contre, nous renonçons (pour l'instant) à redresser les effets de la non-réponse confondue.

Par exemple, vu que

selon le registre national, 2352 individus mâles de 40 à 44 ans résidaient dans la province de Luxembourg au 1^{er} janvier 2002,

²⁷ La frontière entre imputation et repondération n'est pas étanche; songeons à une catégorie représentée dans s par un répondant et un non-répondant: imputer à celui-ci la réponse de celui-là, ou doubler le poids du répondant, reviennent au même.

²⁸ Dans certains cas, θ_k peut être estimé directement; un exemple classique est la méthode de Politz-Simmons, qui consiste à évaluer la probabilité de réponse en fonction de la fréquence des présences (nombre de soirs sur une semaine) de l'individu ou du ménage à son domicile.

d'après le fichier d'enquête, 64 d'entre eux (ont été tirés et) ont effectivement répondu au cours du premier trimestre 2002,

on affectera chaque réponse (de ces 64 individus) d'un coefficient r_k^{-1} (r_k = estimation de ρ_k = probabilité de réponse d'un Luxembourgeois jeune quadragénaire²⁹) égal à $2352/64 = 36,755$.

Nous aboutissons à l'**estimateur post-stratifié en présence de non-réponse**, utilisé dans l'extrapolation de l'enquête depuis 1983³⁰.

$$\hat{y}_5 = \sum_{k \in r} \frac{y_k}{r_k}$$

L'expression alternative de l'estimateur post-stratifié devient

$$\hat{y}_5 = \sum_{h \in H} N_h \bar{y}_h$$

avec $\bar{y}_h = \frac{\sum_{k \in h} y_k}{n_h}$, où à présent n_h désigne l'effectif des répondants appartenant à la strate h .

Notons que cette expression fait intervenir r_k (estimant $\rho_k = \pi_k \theta_k$) mais π_k ni θ_k n'y figurent (explicitement); autrement dit, on peut «oublier» le plan de sondage³¹. Par contre, il est indispensable de disposer de totaux de population ventilés par âge x sexe x province³². En effet, nous avons besoin

des réponses (supposées sincères) des personnes interrogées (y_k)

du nombre de répondants dans chaque post-strate (n_h)

de la population de chaque post-strate (N_h)

L'estimateur \hat{y}_5 reconstitue exactement (et pas seulement en moyenne) la population (croisée ou non) par âge, sexe et lieu de résidence. En ce qui concerne la précision, il corrige la partie de la non-réponse (uniforme ou aléatoire) indépendante de la variable d'intérêt, mais pas la liaison (éventuelle) avec celle-ci.

II.8 Informations auxiliaires supplémentaires

La question³³: «**L'échantillon** de l'EFT est-il **représentatif**?» devrait être reformulée:

«**L'extrapolation** de l'EFT est-elle **représentative**?»; s'il n'est nullement indispensable³⁴ que les

²⁹ La loi des grands nombres ne s'applique plus ici: la constance de π_k garantit une représentation approximativement proportionnelle des hommes et femmes, des jeunes et des vieux, dans s ; mais pas dans r car θ_k dépend de l'âge et du sexe. Par exemple, dans la province de Luxembourg, la proportion masculine s'élevait à 49,3% dans la population au 1^{er} janvier 2001 (248576 personnes), et à 49,0% dans l'échantillon de l'année 2001 (10571 personnes), pour tomber à 47,9% parmi les répondants (7220 personnes).

³⁰ En conformité avec [Règlement 577/98, art.3 §5]: *Les coefficients de pondération sont calculés en tenant compte notamment des probabilités de sélection et de données exogènes sur la distribution par sexe, par classes d'âge (classes de cinq ans) et par région (niveau de la NUTS II)*. «Notamment» veut dire que d'autres variables peuvent intervenir. Alors que les consignes d'Eurostat relatives à l'ordre et la formulation des questions sont très strictes, une grande diversité règne entre les Quinze en ce qui concerne les méthodes d'extrapolation, voir V.3.

³¹ Ceci est particulièrement utile lorsque des ménages de réserve peuvent être interrogés à la place de ménages effectifs non-répondants. Dans ce cas, le calcul des θ_k devient particulièrement compliqué, la probabilité qu'un ménage de réserve soit contacté à défaut d'un ménage effectif dépendant du comportement de celui-ci, et de son statut de première, seconde, ... réserve.

³² Ni la nationalité regroupée (raisons politiques et risque de petites catégories), ni l'état civil (risque de catégories vides lorsque croisée avec l'âge) n'interviennent.

³³ Nous nous intéressons ici à la cohérence et à un volet de la précision (le biais éventuel et les façons de le réduire); la manière naturelle de minimiser la variance consiste à augmenter l'échantillon.

différentes catégories soient représentées de manière exactement proportionnelle parmi les répondants, il est souhaitable qu'après extrapolation, on retrouve leurs effectifs si possible exacts (cohérence de ce que nous connaissons: âge, sexe, ...), à défaut approchés (précision de ce que nous estimons: chômage BIT). Reste à savoir ce que nous entendons par catégorie.

Bien entendu, poser

«La somme des poids des répondants en chômage BIT
égale
le total du chômage BIT»

...serait (certes désirable, pour corriger aussi la non-réponse confondue, mais) téméraire; si nous connaissions le nombre de travailleurs-de chômeurs-d'inactifs parmi la population, l'enquête aurait été superflue!

En fait, l'équation (nous vérifions âge, sexe et domicile des répondants, tandis que le registre national nous fournit le second membre)

«La somme des poids des répondants hommes Luxembourgeois de 40 à 44 ans
égale
le total des hommes Luxembourgeois de 40 à 44 ans»

nous permet de déterminer (totalement, car nous attribuons le même poids à chaque membre de la post-strate) les coefficients de l'estimateur \hat{y}_5 (post-stratification en présence de non-réponse). Ensuite, une fois les coefficients déterminés,

«La somme des poids des répondants en chômage BIT
estime
le total du chômage BIT»

est une estimation. C'est ainsi (aux détails près) que l'EFT belge a été extrapolée depuis 1983.

Nous disions que \hat{y}_5 a pour inconvénient de négliger l'effet que la variable d'intérêt elle-même peut avoir sur le taux de réponse; cependant, poser

«La somme des poids des répondants en chômage BIT
égale
le total du chômage BIT»

...serait (certes désirable, mais) téméraire; si nous connaissions le nombre de travailleurs-de chômeurs-d'inactifs parmi la population, l'enquête aurait été superflue!

Mais voici l'idée de J.M. Museux: en admettant que (sans coïncider) «demandeur d'emploi ONEM» et «chômeur BIT» soient liés (corrélés), une équation supplémentaire

«La somme des poids des répondants demandeurs d'emploi ONEM
égale
le total des demandeurs d'emploi ONEM»

imposerait sans doute de modifier les poids, mais permettrait d'affiner l'estimation du chômage BIT. En d'autres termes, nous utiliserons une variable auxiliaire liée à la variable d'intérêt pour tenter de redresser la non-réponse confondue³⁵ et de stabiliser les fluctuations aléatoires³⁶.

Pour écrire (et résoudre) cette équation, deux conditions doivent être remplies savoir quels répondants sont demandeurs d'emploi ONEM³⁷ disposer du total des demandeurs d'emploi ONEM³⁸

³⁴ Et difficile pratiquement, en raison du sondage à plusieurs degrés et de la non-réponse.

³⁵ Hypothèse: les répondants demandeurs d'emploi ONEM **toujours** moins nombreux (taux de refus systématiquement plus élevé); réaction: augmenter leur poids.

³⁶ Hypothèse: moins de demandeurs d'emploi ONEM dans l'échantillon d'un certain **trimestre** (par hasard); réaction: augmenter leur poids (ils seront moins à se partager le gâteau).

³⁷ Information qui transite –vu son caractère personnel– par la BCSS; savoir quels non-répondants sont dans le cas est utile pour les analyses, mais pas pour l'extrapolation: nous allons uniquement modifier les poids des répondants.

³⁸ Information globale, transmise directement à l'INS.

Population

(Registre national pour Age & ONEM pour Statut)

	Age1	Age2	Age3	Total
Statut1	N_{11}	N_{12}	N_{13}	N_{1+}
Statut2	N_{21}	N_{22}	N_{23}	N_{2+}
Total	N_{+1}	N_{+2}	N_{+3}	$N = N_{++}$

Répondants

(extrait du Registre national et vérifié par enquêteur pour Age & ONEM via BCSS pour Statut)

	Age1	Age2	Age3	Total
Statut1	n_{11}	n_{12}	n_{13}	n_{1+}
Statut2	n_{21}	n_{22}	n_{23}	n_{2+}
Total	n_{+1}	n_{+2}	n_{+3}	$n = n_{++}$

Poids

(6 inconnues à déterminer, poids égaux pour tous les répondants d'une même cellule)

	Age1	Age2	Age3	Total
Statut1	w_{11}	w_{12}	w_{13}	
Statut2	w_{21}	w_{22}	w_{23}	
Total				

Population reconstituée

	Age1	Age2	Age3	Total
Statut1	$w_{11} n_{11}$	$w_{12} n_{12}$	$w_{13} n_{13}$	
Statut2	$w_{21} n_{21}$	$w_{22} n_{22}$	$w_{23} n_{22}$	
Total				

Comparons les deux procédés:

<p>Post-stratification croisée: Reconstituer les cellules (et donc les marges) $n_{11}w_{11} = N_{11}$, etc. 6 équations affines pour les 6 inconnues w_{ij} Très simple à résoudre; <i>unique solution</i>: $w_{ij} = N_{ij}/n_{ij}$, facile calculer «à la main» ou avec une calculatrice de poche – pour autant que chaque $n_{ij} \neq 0$.</p>	<p>Calage sur marges: Reconstituer les marges (pas nécessairement les cellules) $n_{11}w_{11} + n_{12}w_{12} + n_{13}w_{13} = N_{1+}$, etc. 5 équations affines (4 indépendantes, car $N_{1+} + N_{2+} = N_{+1} + N_{+2} + N_{+3} = N_{++} \dots$) pour les 6 inconnues w_{ij} Système indéterminé; <i>une des solutions</i> est de prendre $w_{ij} = N_{ij}/n_{ij}$ pour autant que tous les $n_{ij} \neq 0$; une solution peut exister même si un ou plusieurs $n_{ij} = 0$ (tant qu'aucune ligne ni aucune colonne des n_{ij} n'est identiquement nulle)</p>
--	--

II.10 Distances et fonctions de calibrage

Comme les équations de calage sur marge ont généralement plusieurs solutions, nous allons privilégier celle qui modifie le moins les poids de sondage.

En d'autres termes, nous cherchons des **poids calibrés** w_k

vérifiant les contraintes (telles que la somme des poids des répondants hommes de 40 à 44 ans égale la population totale des hommes de 40 à 44 ans et la somme des poids des répondants demandeurs d'emploi ONEM égale la population totale des demandeurs d'emploi ONEM)

aussi **proches** que possible **des poids de sondage** $d_k = \pi_k^{-1}$ définis au §5.

Il s'agit d'un problème d'optimisation sous contraintes.

Dans le cas de l'EFT, la situation se simplifie, car tous les individus d'une strate (province donnée pour un trimestre donné) ont la même probabilité de tirage, π_k et son inverse d_k seront indépendants de k . Que les poids calibrés w_k soient proches des poids de sondage d_k impliquera ici que les w_k sont proches les uns des autres, bonne nouvelle pour la variance. Savoir si les groupes comptent 28 ménages (chacun susceptible d'être interrogé) ou 20 ménages (maximum interrogeable) est sans importance. En fait, g-Calib calcule d'abord un coefficient global (*scale parameter*) et des poids échelonnés (*scaled weights*), puis ajuste les poids individuels.

Reste à définir ce que nous entendons nous par «proches». Pour des raisons techniques, nous allons (dans le respect des contraintes) minimiser

$$\sum_{k \in r} d_k G\left(\frac{w_k}{d_k}\right)$$

où la fonction G , définie sur un voisinage convexe de 1, vérifie les conditions:

G est positive;

G est strictement convexe;

G est deux fois continûment dérivable (à l'intérieur de son domaine);

$G(1) = 0$;

$G'(1) = 0$;

$G''(1) = 1$.

Les conditions a) et d) impliquent que chaque terme (k fixe) de la somme à minimiser est lui-même minimal lorsque $w_k = d_k$, ce qui traduit notre volonté de modifier le moins possible les poids de sondage. Rappelons que les d_k sont strictement positifs (heureusement, car ils apparaissent en dénominateurs!), tandis que les w_k peuvent (dans certains cas) être négatifs.

Le théorème des fonctions implicites affirme que la fonction réciproque F de G' , définie et continûment dérivable au voisinage de 0, satisfait $F(0) = 1$ et $F'(0) = 1$. F est appelée **fonction de calibrage**.

Sous certaines hypothèses, on montre que le problème d'optimisation sous contraintes admet une et une seule solution [VWM].

En supposant, pour ne pas infliger de calcul matriciel au non-spécialiste, une seule contrainte

$$\sum_{k \in r} c_k w_k = w$$

(par exemple, c_k valant 1 si le répondant k est un homme âgé de 40 à 44 ans, 0 sinon), où w représente le total de calage, la solution du problème de minimisation sous contrainte (extremum lié) fait intervenir la fonction (lagrangienne)

$$L(w_k, \lambda) = \sum_{k \in r} d_k G\left(\frac{w_k}{d_k}\right) - \lambda \left(\sum_{k \in r} c_k w_k - w \right).$$

En annulant les dérivées partielles, on a

$$(*) \sum_{k \in r} c_k w_k = w \quad \left(\text{puisque } \frac{\partial L}{\partial \lambda} = 0 \right)$$

$$G'\left(\frac{w_j}{d_j}\right) = \lambda c_j \quad \left(\text{car } \frac{\partial L}{\partial w_j} = 0 \right)$$

et donc

$$(**) \quad w_j = d_j F(\lambda c_j)$$

En introduisant (**) dans (*), on obtient (ici, une équation à une inconnue)

$$\sum_{k \in r} c_k d_k F(\lambda c_k) = w$$

nous permettant de calculer λ , et, grâce à (**), les w_k .

II.11 Exemples de fonctions de calibrage

Nous avons de la latitude dans le choix du critère de proximité, donc de G et de F .

Une manière de procéder (qui évoquera le test de χ^2 chez les plus avertis de nos lecteurs) est de minimiser

$$\sum_{k \in r} \frac{(d_k - w_k)^2}{2d_k} \quad (\text{soit } G(x) = \frac{(x - 1)^2}{2})$$

Cette méthode⁴¹ est appelée **linéaire**, car la résolution du problème d'extremum lié à l'aide des multiplicateurs de Lagrange fait intervenir des fonctions affines⁴²: la réciproque $(G')^{-1}$ de la fonction dérivée de G , en d'autres termes la fonction de calibrage, est définie par

$$F(u) = 1+u, \text{ puisque } G'(x) = x-1.$$

Des variantes consistent à imposer une borne inférieure (souvent 0) et/ou une borne supérieure aux quotients $g_k = w_k/d_k$.

Parmi les alternatives, nous pouvons considérer comme fonction à minimiser

$$\sum_{k \in r} w_k \log \frac{w_k}{d_k} + d_k - w_k \quad (\text{soit } G(x) = x \log x - x + 1, x > 0)$$

appelée méthode⁴³ **exponentielle**, qui postule la positivité des w_k . Leur détermination nécessite la résolution (par exemple, par la méthode de Newton-Raphson) d'un système d'équations non linéaires: la fonction de calibrage est ici

$$F(u) = \exp u, \text{ car } G'(x) = \log x.$$

On consultera [Tillé p. 217, notations légèrement différentes], [VWM] ou [Vanderhoeft] pour les détails relatifs aux fonctions de calibrage.

II.12 g-Calib

Une résolution du problème de minimisation sous contraintes a été proposée par des membres de l'Institut National de la Statistique et des Etudes Economiques (CALMAR, sous le logiciel SAS). Parmi les variantes, citons GES (Statistique Canada) et Bascula (Centraal Bureau voor de Statistiek). Nous ferons appel à g-Calib, réalisé à l'INS dans un environnement SPSS, et dont la souplesse de certaines caractéristiques (notamment la faculté de croiser ou non certaines variables, et d'intégrer éventuellement la dimension du ménage) représentent une avancée.

⁴¹ Similaire à la méthode d'estimation par la régression [Tillé p. 221].

⁴² Linéaires, dans l'ancienne terminologie.

⁴³ Aussi nommée méthode d'estimation par ratissage [Tillé p. 222].

III Mise en oeuvre

III MISE EN OEUVRE

III.1 Options

Si le chapitre précédent indiquait la ligne générale suivie (tenir compte non seulement des variables démographiques, mais aussi de celles de la sécurité sociale), celui-ci expose les choix faits (souvent pour des raisons plus pratiques que théoriques) en ce qui concerne

- ❑ les variables à utiliser;
- ❑ l'utilisation croisée ou marginale de ces variables;
- ❑ les calendriers d'input (à quel moment recueille-t-on l'information sociale) et d'output (à quelle fréquence et dans quel délai l'INS doit-il livrer les résultats calibrés), ces deux étant évidemment liés;
- ❑ les totaux de population;
- ❑ l'intégration éventuelle de la dimension «ménage»;
- ❑ la méthode de calibrage (linéaire ou exponentiel);
- ❑ le sort à réserver aux répondants pour lesquels l'information sociale manque.

Sans prétendre que la méthode présentée ici soit idéale ni définitive, elle nous semble acceptable. Quelques alternatives seront discutées. Nous aurions aimé multiplier les simulations – mais (outre le temps limité dont nous disposons) il faut rester prudent: une variante peut avoir un impact (minime ou sensible) sur une province ou un trimestre, sans qu'on puisse tirer la même conclusion partout et toujours.

III.2 Quelques corrélations

Comme nous le disions au premier chapitre, l'INS a demandé (et obtenu) l'accès à une vingtaine de variables, provenant (directement pour les totaux, via la BCSS pour les données individuelles) de six institutions. Pour chacune des institutions, la première information consiste à savoir si l'individu est connu (a un dossier) ou non. Parmi les autres variables, certaines sont destinées à améliorer l'extrapolation, d'autres servent plutôt aux analyses et aux contrôles.

Pour le calibrage linéaire¹ (équivalent à l'estimation par la régression généralisée), le gain de précision (la diminution de la variance) est lié au coefficient de détermination du modèle; en particulier

une variable de calibrage apporte d'autant plus de précision qu'elle est **corrélée avec la variable d'intérêt** (par exemple, prendre «présent en fin de trimestre à l'ONSS» ou «connu à l'ONSS» devrait nous permettre d'améliorer la précision de l'estimation de l'emploi);

deux variables de calibrage fortement liées entre elles (par exemple, «présent en fin de trimestre à l'ONSS» et «connu à l'ONSS») ne nous apporteront guère plus qu'une seule de ces deux variables. Penchons-nous d'abord sur l'emploi BIT; quelles informations pourraient-elles améliorer significativement la précision de nos estimations?

Nous avons calculé quelques corrélations (entre couples de variables binaires) basées sur le fichier 1999 (enquête + données BCSS au quatrième trimestre, en se limitant aux individus pour lesquels celle-ci était connue, donc sans imputation, et sans pondérer ... ce qui induirait un cercle vicieux, car nous cherchons précisément à déterminer les variables qui nous permettent de mieux calculer les coefficients). Notons que cette analyse est indépendante des totaux (et de leurs retards ou erreurs éventuels). En revanche, le phénomène de permutation (voir I.5) a pour effet de perturber (légèrement, et à la baisse) ces corrélations.

Comme genre et âge entrent de toutes façons en ligne de compte, nous avons pris trois classes d'âge, et donné les coefficients de corrélation pour les hommes et femmes.

¹ Nous verrons au §8 que les résultats produits par la méthode exponentielle (ratissage) sont très voisins.

C'est naturellement avec les variables de l'ONSS que le lien est le plus fort (vu que la majorité des travailleurs sont salariés, et la plupart de ceux-ci sont inscrits à l'ONSS, plutôt qu'à l'ONSSAPL, sans oublier ALE, frontaliers sortants et travail occasionnel). La primauté² donnée à «présent en fin de trimestre ONSS» sur «connu ONSS» est motivée par la corrélation plus élevée (précision!) et le fait que les statistiques de la BNB se basent sur «présent en fin de trimestre ONSS» (cohérence!). La corrélation (entre «présent fin de trimestre ONSS» et «occupé BIT») –comme on pouvait s'y attendre– devient un peu meilleure si on se limite au quatrième trimestre 0,663 pour les 20-24 ans, 0,535 pour les 40-44 et 0,517 chez les 60-64 ans), car le décalage temporel est limité. Mutatis mutandis pour l'ONSSAPL.

Pour l'INASTI, une telle nuance (entre «présent en cours de trimestre» et «présent à la fin du trimestre») n'existe pas³.

Notons encore que le gain total de précision n'est pas égal (mais supérieur, vu la corrélation négative entre chaque couple parmi «présent fin de trimestre ONSS», «présent fin de trimestre ONSSAPL», «connu INASTI») à la somme des carrés des 3 coefficients de corrélation. Par exemple pour les 40-44 ans, cette somme vaut $0,26 + 0,02 + 0,03 = 0,31$... mais si nous définissons la variable «emploi sécu» –valant 1 si au moins une des trois «présent fin de trimestre ONSS», «présent fin de trimestre ONSSAPL», «connu INASTI» égale 1, et 0 sinon– nous trouvons une corrélation de 0,753 (carré 0,57) entre «emploi sécu» et «occupé BIT».

Nous renonçons aux variables «ONP» et «AdP», car la corrélation est faible (en fait, l'âge est déjà un bon indicateur du statut de non-activité) et l'intérêt pour la cohérence limité: les pensionnés (très majoritairement sans emploi et n'en cherchant plus) ne forment qu'une partie des inactifs.

TABLEAU III.1 CORRELATION AVEC INDICATRICE «OCCUPE BIT» (ANNEE 1999)

Age	connu ONSS	présent fin ONSS	connu APL	présent fin APL	connu INASTI	connu ONP	connu AdP	connu ONEM
Utilisé?		OUI		OUI	OUI			Voir infra
20-24	0,537	0,550	0,096	0,089	0,145	(vide)	0,020	-0,052
square	0,288	0,302	0,009	0,008	0,021	***	0,000	0,003
H	0,541	0,557	0,086	0,077	0,155	***	0,018	0,019
F	0,528	0,536	0,115	0,110	0,124	***	0,022	0,079
40-44	0,488	0,513	0,132	0,132	0,186	-0,096	-0,061	-0,236
square	0,238	0,263	0,017	0,017	0,034	0,009	0,004	0,056
H	0,343	0,403	0,075	0,076	0,143	-0,046	-0,111	-0,336
F	0,545	0,558	0,195	0,193	0,176	-0,098	-0,034	-0,175
60-64	0,212	0,486	0,212	0,214	0,472	-0,147	-0,076	-0,128
square	0,045	0,236	0,045	0,046	0,223	0,022	0,006	0,016
H	0,101	0,494	0,172	0,168	0,507	-0,136	-0,115	-0,288
F	0,366	0,432	0,284	0,296	0,327	-0,100	-0,033	0,032

Quant aux catégories de l'ONEM, leur interprétation est plus délicate. Ainsi, on peut conjecturer que «sans emploi demandeur d'emploi» est lié à «chômeur BIT», «sans emploi non demandeur d'emploi»⁴ laisse présumer «inactif BIT», alors que «travailleur avec allocations»⁵ correspondra souvent à «actif occupé BIT».

²Comme expliqué supra, la combinaison «connu ONSS» ET «présent fin de trimestre ONSS» n'apporte pas beaucoup plus que «présent fin de trimestre ONSS», puisque la corrélation entre ces deux variables vaut 0,915 (un peu plus pour les quadragénaires dont la situation est généralement stable, un peu moins pour les tranches d'âge extrêmes qui entrent dans/sortent plus souvent de l'emploi).

³ La spécification «catégorie» (avec trois modalités: activité principale - activité secondaire - actif après pension) pourrait présenter de l'intérêt, surtout dans l'optique des doubles comptes.

⁴ Principalement des chômeurs complets âgés, et aussi des chômeurs complets dispensés temporairement de l'inscription comme demandeur d'emploi pour raison d'études ou de difficultés familiales.

⁵ Notamment les travailleurs à temps partiel avec allocation de garantie de revenus.

Les tableaux croisés ci-dessous le confirment (vu les faibles effectifs, nous avons combiné les sexes). En se focalisant sur les indicatrices «chômeur BIT» et «sans emploi demandeur d'emploi», on trouve des coefficients de corrélation de 0,416 (20-24 ans), 0,560 (40-44 ans) et 0,167 (60-64 ans), suffisamment élevés pour qu'on intègre cette variable.

Nous proposons de conserver «Sans emploi non demandeur d'emploi» et «pré-pensionné» (corrélation avec «inactif BIT» >0,2 pour les 60-64 ans, inférieure pour les plus jeunes), ainsi que «emploi+allocation» (corrélation avec «occupé BIT» proche de 0,1). Même si le gain de précision absolu n'est pas énorme, n'oublions pas qu'une diminution de l'écart type de 1000 unités –mettons– serait relativement plus importante pour 300 000 chômeurs que pour 4 000 000 de travailleurs.

TABLEAUX III.2 STATUT BIT x STATUT ONEM, PAR CLASSE D'AGE (ANNEE 1999)

	ss emploi	ss emploi			
20-24 ans Ensemble	demandeur	non dem	emploi+alloc	prépension	
INAC	42,4	21,0	62,8	8,0	***
TRAV	45,9	24,1	18,6	82,8	***
CHOM	11,7	54,9	18,6	9,2	***
personnes	5928	519	43	163	0

	ss emploi	ss emploi			
40-44 ans Ensemble	demandeur	non dem	emploi+alloc	prépension	
INAC	16,7	27,9	44,4	2,4	***
TRAV	77,3	17,2		93,8	***
CHOM	5,9	54,9	55,6	3,8	***
personnes	8090	548	9	289	0

	ss emploi	ss emploi			
60-64 ans Ensemble	demandeur	non dem	emploi+alloc	prépension	
INAC	86,9	76,2	95,5	25,0	99,0
TRAV	12,9	9,5	1,7	75,0	1,0
CHOM	0,3	14,3	2,9		
personnes	5979	21	242	8	669

III.3 *Choix des variables de calibrage*

Le *TABLEAU III.3* présente les **informations de la sécurité sociale** (actuellement) **utilisées** (en plus des données démographiques) **pour calibrer** (années 1999, 2001 et 2001), en fonction de l'objectif principal, c'est-à-dire estimer les 3 catégories BIT (en distinguant éventuellement salariés et non-salariés). Il s'agit de sept variables muettes ou indicatrices. Bien entendu, calibrer par exemple selon la NACE ou le volume de travail (ce dernier constituant une variable authentiquement quantitative) aurait intéressé ceux qui étudient particulièrement la répartition selon les activités ou la durée du travail.

TABLEAU III.3 VARIABLES UTILISEES POUR CALIBRER

ORGANISME	VARIABLE	VALEURS POSSIBLES
ONSS (LATG)	Personne inscrite en fin de trimestre?	0 : non 1 : oui
ONSSAPL	Personne inscrite en fin de trimestre?	0 : non 1 : oui
INASTI	NISS = numéro national	0 : inconnu 1 : connu
ONEM	Dernière situation durant le mois (de référence) ⁶	1 : demandeur d'emploi inoccupé 2: non-demandeur d'emploi inoccupé 3 : travailleur qui cumule avec une allocation 4: prépensionné à temps plein

Ne croyons pas que la montagne (21 variables demandées - *TABLEAU I.4*) a accouché d'une souris (4 variables de calibrage - *TABLEAU III.3*), car d'autres variables...

- interviennent dans le calcul de la NACEbis (voir IV.4); c'est le cas du «code NACE» de l'ONSS, mais aussi des «volumes de travail» ONSS et ONSSAPL (afin de déterminer l'activité principale d'un salarié inscrit dans ces deux institutions);
- n'ont pas été prises en compte, lorsque le risque qu'une catégorie n'ait aucun représentant dans l'échantillon (rendant impossible le respect d'un total de calibrage petit mais non nul) était réel, par exemple les «prépensionnés à temps partiel» de l'ONEM
- ont permis de calculer des corrélations, lesquelles s'avéraient parfois trop faibles pour que les variables apportent un gain de précision significatif; notamment les variables «pensionné» (ONP et ADP); comme dans le cas de la régression multiple, on cherchera un compromis entre les objectifs de précision et de parcimonie⁷.
- indiquent que la qualité de la mesure reste médiocre pour certaines variables d'enquête (ainsi, la confrontation entre la question 6 de l'EFT et les codes 6 et 7 de «Dernière situation durant le mois (de référence)» ONEM.

En conclusion (provisoire, et sans démonstration mathématique que ce choix soit optimal), 7 variables binaires de la sécurité sociale (l'utilité des trois dernières restant incertaine) interviendront dans le calcul des poids, dont voici les abréviations :

- «ONSS» présent fin de trimestre ONSS
- «APL» présent fin de trimestre ONSSAPL
- «INASTI» connu INASTI
- «ONEM1» sans emploi demandeur d'emploi avec allocation ONEM
- «ONEM2» sans emploi non demandeur d'emploi avec allocation ONEM
- «ONEM3» en emploi avec allocation ONEM
- «ONEM4» prépensionné à plein temps avec allocation ONEM

III.4 *Post-strates et marges de calibrage*

Une fois les variables de calibrages (de la sécurité sociale) choisies, il faut savoir comment les organiser. Il est naturel de travailler par trimestre (même si nous ne produisons jusqu'à présent des résultats calibrés que tous les ans, ceux-ci doivent être ventilés par trimestre) croisé avec la NUTS II (puisque les probabilités de sélection dépendent de la province, mais non de l'âge ni du sexe).

⁶ Convertie en quatre variables muettes «ONEM1», «ONEM2», «ONEM3», «ONEM4».

⁷ Signalons à ce sujet l'article [NS] aimablement communiqué par C. Vanderhoeft, que D. Luminet regrette de n'avoir pas encore pu étudier en profondeur.

Reste à savoir quelles données seront croisées, et quelles utilisées marginalement (voir II.9). Pour différentes raisons –dont la continuité avec la post-stratification traditionnelle n'est pas la moindre), nous gardons le croisement âge quinquennal x sexe. Dans chaque province trimestrielle (ou trimestre provincial), on a opté pour les $32+7 = 39$ contraintes (dans la notation préconisée par C. Vanderhoeft)

(AGE*SEXE) + ONSS + ONSSAPL + INASTI + ONEM1 + ONEM2 + ONEM3 + ONEM4
autrement dit, on exige notamment que

- la somme des poids des répondants hommes de 40-44 ans coïncide avec le total de la population masculine de cette tranche d'âge (sans faire intervenir les informations BCSS)
- la somme des poids des répondants demandeurs d'emploi inoccupés coïncide avec le total des demandeurs d'emploi inoccupés (sans faire intervenir l'âge ni le sexe)⁸.

Dans l'optique des doubles comptes et de leur estimation, il aurait été désirable de combiner ONSS et ONSSAPL, ce qui est malheureusement impossible: si nous disposons de l'information croisée pour les membres de l'échantillon (répondants ou non), les totaux calculés séparément par chaque institution ne permettent pas de savoir exactement combien de personnes figurent dans les deux.

III.5 Calendrier

A l'instar des variables de calibrage, le calendrier ci-dessous doit être considéré comme provisoire, et pourrait être revu selon les besoins des utilisateurs ... et surtout les possibilités de l'INS.

Tout en notant que

alors que l'ONSS, l'ONSSAPL et l'ONEM nous communiquent leurs données deux fois par an (situation fin juin et fin décembre), l'INASTI ne peut fournir que la situation fin décembre, le nombre de mois nécessaire à la livraison des données individuelles et globales (rappelons que les deux sont indispensables si l'on veut calibrer) varie⁹ entre 5 (ONEM) et 7 (ONSS & ONSSAPL)¹⁰, il a fallu concilier **cohérence** (idéalement, des données portant sur le même moment, notamment pour mieux estimer les doubles comptes, et contemporaines des données démographiques), **précision** (idéalement, des données portant sur une période proche du trimestre d'enquête, et donc mieux corrélées avec la situation selon l'EFT) ... et **réalisme** (de toutes façons, il faut que les données soient là).

Dans la mesure où des résultats pour l'année N sont souhaités en juin N+1, et vu que l'ONEM arrive à fournir ses totaux dans les 5 mois, l'INASTI a besoin de 6 mois, alors qu'ONSSAPL et ONSS sont plus lents, l'information la plus pertinente disponible serait

TABLEAU III.4 CALENDRIER «OPTIMAL» (POUR LA PRECISION)

Trimestre	1	2	3	4
ONSS	Décembre N-1	Juin N	Juin N	Juin N
ONSSAPL	Décembre N-1	Juin N	Juin N	Juin N
INASTI	Décembre N-1	Décembre N-1	Décembre N-1	Décembre N-1
ONEM	Décembre N-1	Juin N	Juin N	Décembre N

⁸ Comme observé au chapitre précédent, croiser avec l'âge serait périlleux; avec le sexe était/sera envisageable.

⁹ ...officiellement; un peu plus en réalité.

¹⁰ Pour certaines institutions, les données personnelles précèdent les totaux, pour d'autres c'est l'inverse.

En fait, nous avons travaillé en 1999 avec (pas le choix!)

TABLEAU III.5 CALENDRIER DE MISE EN ROUTE

Trimestre	1	2	3	4
ONSS	Décembre 99	Décembre 99	Décembre 99	Décembre 99
ONSSAPL	Décembre 99	Décembre 99	Décembre 99	Décembre 99
INASTI	Décembre 99	Décembre 99	Décembre 99	Décembre 99
ONEM	Décembre 99	Décembre 99	Décembre 99	Décembre 99

et en 2000 et 2001 avec (essayant dans la mesure du possible d'assurer la simultanéité des informations, seul l'INASTI étant déphasé-priorité à la cohérence sur la précision!)

TABLEAU III.6 CALENDRIER DE CROISIÈRE

Trimestre	1	2	3	4
ONSS	Juin N	Juin N	Juin N	Juin N
ONSSAPL	Juin N	Juin N	Juin N	Juin N
INASTI	Décembre N-1	Décembre N-1	Décembre N-1	Décembre N-1
ONEM	Juin N	Juin N	Juin N	Juin N

Observons encore que J.M. Museux pensait, plus ambitieusement, à une double approche («conjoncturelle» pour produire des estimations trimestrielles rapides, «structurelle» comme consolidation annuelle). En effet, pour certains trimestres d'enquête, nous recevons d'abord la situation de sécurité sociale avant la semaine de référence, ultérieurement cette situation après (mais à un moment plus proche de) ladite semaine. Nous y reviendrons au Chapitre V.

III.6 Totaux de population

En marge du calendrier des données de la sécurité sociale se pose (pour calibrer comme pour post-stratifier) la question des totaux démographiques. Au niveau individuel, la stabilité de la province est assurée¹¹, l'évolution de l'âge prévisible¹², et les changements de sexe rarissimes. En revanche, la population totale doit-elle être calculée au 1^{er} janvier (pour les 4 trimestres) comme le veut la tradition (datant déjà des enquêtes printanières organisées jusqu'en 1998) ou au 30 juin (pour assurer une meilleure comparabilité avec les chiffres administratifs)? Les estimations calibrées pour 1999 et 2000 (voir Chapitre IV) se basent sur la population au début d'année; pour 2001 nous avons pris celle de la mi-année. Les différences sont faibles (l'évolution démographique est lente, même si la régularisation de personnes qui ne figuraient pas au registre national tend à l'accélérer). Le débat sera poursuivi au Chapitre V.

III.7 Ménages

La prise en compte éventuelle (possible avec g-Calib) du ménage dans le calcul des coefficients a alimenté des discussions passionnantes (et souvent passionnées).

Les options, ou «niveaux» (terminologie de C. Vanderhoeft) étaient:

niveau 1: le poids dépend seulement des caractéristiques de l'individu;

niveau 2: le poids dépend des moyennes des caractéristiques des individus dans le ménage, et est donc identique pour tous les membres d'un même ménage;

niveau 3: le poids dépend seulement des caractéristiques du ménage;

¹¹ En cas de déménagement, un enquêteur peut interroger le ménage initialement tiré pour autant que celui-ci réside encore dans la même commune (donc la même province).

¹² C'est la classe d'âge à la fin de la semaine de référence qui intervient dans l'extrapolation.

niveau 4 le poids dépend des caractéristiques des individus et du ménage.

Jusqu'à présent, nous avons calibré sans tenir compte de la dimension ménage (Niveau 1), vu que la plupart des résultats de l'EFT concernent les individus; en particulier, les comptes nationaux se focalisent sur les personnes. La simplicité (les alternatives nécessitent d'identifier ménages et individus) est une autre raison.

Si le niveau 3 (perte des informations individuelles) apparaît trop réducteur, les arguments en faveur des niveaux 2 et 4 sont:

la structure du ménage a un impact sur le taux de réponse (TABLEAUX II.1 & II.2)

certaines exploitations de l'EFT s'intéressent en particulier aux ménages (p.ex. ceux où personne ne travaille, *jobless households*);

l'EFT sert de base à l'extrapolation d'autres enquêtes de l'INS (Budget des Ménages, ...);

le problème des permutations (de données au sein d'un même ménage), voir I.5) pourrait trouver une solution.

Pour alimenter la discussion, extrayons encore d'[Avignon] :

Une pondération basée sur les individus peut amener certaines discordances lorsque les résultats de l'enquête sont analysés au niveau des ménages, les individus au sein d'un même ménage n'ayant en général pas le même poids statistique. Eurostat souhaiterait¹³ que tous les individus d'un même ménage reçoivent le même poids. Une première justification de cette demande est que la plupart des pays tirent au hasard des ménages et non des individus pour constituer leur échantillon. Une seconde résulte de la tendance de plus en plus affirmée d'analyser le comportement sur le marché de l'emploi relativement à la situation du ménage et non de l'individu.

Les pays nordiques, qui eux sélectionnent des individus et non des ménages, sont réticents vis-à-vis de cette approche pour des raisons essentiellement culturelles. L'Irlande et le Royaume-Uni sont dubitatifs pour des raisons méthodologiques, tandis que la France et les Pays-Bas appliquent un coefficient unique par ménage et ont exposé leur méthodologie.

Il est conseillé aux Etats membres de tester les méthodes de pondération unique par ménage et d'analyser les éventuelles distorsions ou améliorations vis-à-vis de leur enquête telle qu'actuellement traitée.

III.8 Fonctions de calibrage

Rappelons que (toujours par province x trimestre) notre plan de sondage est autopondéré, signifiant que chaque individu a la même probabilité de **tirage**. Que les classes d'âge croisées avec les sexes, et les différentes catégories (inscrit ou non à l'ONSS, à ONSSAPL, à INASTI; inscrit en telle qualité à l'ONEM) ne soient pas représentées proportionnellement parmi les **répondants** s'explique d'une (petite: loi des grands nombres!) part, par les fluctuations aléatoires dans l'échantillon; de l'autre, par le comportement de réponse qui dépend (peu du sexe, mais surtout) de l'âge et du statut socio-professionnel administratif (ceux-ci étant liés), comme nous avons vu au Chapitre II (TABLEAUX II.1 & II.2).

Abstraction faite de la province et du trimestre, la différence sans doute la plus spectaculaire dans le TABLEAU II.1 (58,1% de réponse chez les 30-34 ans, 85,2 % chez les 70-74 ans) a pour effet que les poids de ceux-ci seront en moyenne les deux-tiers des poids de ceux-là¹⁴. Comme ces taux de réponse ne diffèrent pas énormément, les méthodes de calibrage linéaire et exponentielle (voir II.11) produiront des résultats voisins¹⁵. Dans les deux cas (comme dans notre bonne vieille post-

¹³ Ce souhait figurait dans une version préliminaire, mais a disparu de la rédaction finale du [Règlement 577/98].

¹⁴ Si (l'EFT n'était pas obligatoire, et que) les taux de réponse tombaient de—mettons— 50%, soit 8,1% contre 35,2%, la distorsion pondérale augmenterait, comme l'écart entre les «coefficients linéaires» et les «coefficients exponentiels».

¹⁵ Mathématiquement, cela se justifie ainsi: les poids calibrés étant proches des poids échelonnés (voir II.10), la différence entre $F_{lin}(u) = 1 + u$ et $F_{exp}(u) = exp u$, de l'ordre de u^2 , sera très petite si u est petit.

stratification), il s'agit de reconstituer les populations par âge et par sexe; interviennent alors les autres contraintes, qui nous forcent à modifier les poids suivant la situation sociale.

Voici un quadruple **exemple** (l'information démographique étant complétée, ou non, par celle de la sécurité sociale; avec calibrage linéaire ou exponentiel) numérique (Province de Luxembourg, quatrième trimestre 2001) montre

au mathématicien, la moyenne (non pondérée) et l'écart type des poids au statisticien du travail, les estimations (pondérées) qui en résultent

On vérifie que

- la moyenne des poids, à l'instar de la population totale extrapolée, est toujours la même quelle que soit la méthode (et on peut en tirer le nombre de répondants);
- l'écart type des poids devient un peu supérieur lorsqu'on adjoint l'information de la sécurité sociale

— si l'on utilise la seule information démographique, les calibrages linéaire et exponentiel coïncident (entre eux, et) avec la post-stratification;

— avec l'information de la sécurité sociale, les calibrages linéaire et exponentiel ne coïncident pas, mais produisent des résultats (tant les coefficients¹⁶ que les estimations résultantes) très proches.

TABLEAU III.7. COMPARAISON DES CALIBRAGES

Données utilisées	démographiques	démographiques	démo + sécu	démo + sécu
Méthode	Lin	Exp	Lin	Exp
POIDS (non pondéré)				
Moyenne	137,8711	137,8711	137,8711	137,8711
Ecart type	34,77629	34,77629	35,93186	35,98453
STATUT (pondéré)				
Salarié	79369	79369	79007	79005
Nonsalarié	16185	16185	15660	15643
Chômeur	7153	7153	7656	7693
Inactif	146564	146564	146949	146930
Total	249271	249271	249271	249271

D'après ce tableau (et d'autres calculs analogues), la différence entre les deux (principales) méthodes est minime. Nous avons préféré la linéaire, car son exécution est un peu plus rapide. Revenons sur le problème des permutations (dans quelques ménages, l'information BCSS, l'information BCSS a été permutée entre les individus, ce qui modifie 7 des 39 contraintes): comme les coefficients de variation (rapports de l'écart type à la moyenne) des poids sont assez faibles (inférieurs au quart, quelle que soit la méthode), et que l'âge et le sexe continuent à intervenir dans le calcul, l'effet global sur les estimations sera (heureusement) minime.

III.9 Imputation

Pour une petite fraction¹⁷ de l'échantillon, l'information individuelle de la BCSS n'est pas disponible. Il s'agit d'individus non repris dans l'échantillon initial (ni parmi les ménages effectifs, ni parmi les ménages de réserve), mais qui

α soit se sont intégrés (naissance ou «migration») à un ménage tiré au sort (et sont tenus de répondre);

¹⁶ En particulier, nous n'avons (ni ici ni ailleurs) jamais rencontrés de poids négatifs, théoriquement possibles en méthode linéaire.

¹⁷ Autour de 2500 (dont un millier de nouveaux-nés) sur les 100 000 personnes interrogées chaque année, proportionnellement plus à Bruxelles.

β soit font partie d'un ménage nouvel occupant d'un logement tiré au nom d'un autre ménage (et peuvent répondre si l'enquêteur est suffisamment persuasif).

Leur N° national n'ayant pas été transmis (au moment du tirage initial) à la BCSS, nous ignorons dans quels fichiers (et si oui, avec quelles données) ils figurent.

Différentes options se présent(ai)ent:

ne pas tenir compte de ces répondants¹⁸, ce qui présente plusieurs inconvénients (base de données rétrécie; dans le cas α , composition du ménage inexacte; dans le cas β , suppression d'un ménage dont les caractéristiques étaient vraisemblablement similaires à celles du ménage initial).

les garder avec leur poids initial (suggestion J.M. Museux) échelonné (*scaled weight*); s'il est assez facile de "geler" leur poids, il faudrait aussi modifier les totaux, ce qui n'est pas aisé.

envoyer une requête complémentaire à la BCSS (à envisager... si le calendrier le permet)

imputation *hot deck* (par un plus proche voisin).

Nous avons choisi l'**imputation** (Syntaxe SPSS de C. Vanderhoeft). L'idée consiste à attribuer à chaque répondant dont l'information BCSS manque, les données de la sécurité sociale (uniquement les quatre/sept variables utilisées pour le calibrage - imputer une NACE ou un salaire serait trop scabreux) d'un autre individu «voisin» pour lequel ces informations sont disponibles. Par voisin, nous entendons: même trimestre, même province, même âge quinquennal, même statut (salarié/non-salarié/chômeur/inactif), sans condition sur le sexe (qui augmenterait le risque de laisser un «receveur» sans «donneur»). S'il y a plusieurs donneurs potentiels, on tire au sort. Ce dernier point est peut-être philosophiquement délicat¹⁹. Nous avons opté pour *seed*²⁰ = 1999000 (en 1999), 2000000 (en 2000) et 2001000 (en 2001); l'impact sur le résultat de l'imputation est faible²¹. A titre d'exemple, considérons, pour le premier trimestre 1999, la province avec le moins de répondants (Brabant Wallon).

Quelques remarques (valant généralement pour autres provinces – trimestres - années):

beaucoup de jeunes enfants (en fait 12 des 16 receveurs 0-4 ans n'avaient pas encore soufflé leur première bougie au moment de l'enquête) parmi les individus à imputer;

chaque cellule vide côté donneurs l'est également côté receveurs;

la proportion de chômeurs est supérieure parmi les receveurs (surtout parmi les receveurs de 15 ans et plus: $3/27 = 11\%$, contre $55/1442 = 4\%$ chez les donneurs), soulignant le danger qu'il y aurait à les laisser tomber (première option);

la proportion de non-salariés inférieure parmi les receveurs (parmi les receveurs de 15 ans et plus: $1/27$, contre $151/1442$ chez les donneurs), soulignant l'utilité de séparer salariés (relevant généralement de l'ONSS ou de l'ONSSAPL) et non salariés (le plus souvent inscrits à l'INASTI) avant d'imputer.

¹⁸ Et modifier nos instructions aux enquêteurs. On constate à nouveau que calibrage "théorique" et travail de terrain "pratique" sont indissociables.

¹⁹ Après le travail de terrain, les résultats dépendront encore (faiblement) d'un élément arbitraire ou aléatoire.

²⁰ *Seed* = valeur initiale

²¹ Une mini-simulation donne un écart-type inférieur à 4 sur le nombre d'individus imputés comme présents en fin de trimestre à l'ONSS, l'effet (extrapolé, trimestriel, pour le Royaume) sur le nombre de salariés ne dépassant pas 250.

TABLEAU III.8 DONNEURS ET RECEVEURS POUR L'IMPUTATION

Age	sala	nonsa	inac	cho	Donneurs	sala	nonsa	inac	cho	Receveurs
00-04			100		100			16		16
05-09			111		111			1		1
10-14			103		103					
15-19	5		111	2	118			1		1
20-24	29	3	59	8	99	3				3
25-29	64	5	13	14	96	3	1			4
30-34	83	15	10	4	112	4			1	5
35-39	91	31	16	8	146	1				1
40-44	103	27	14	8	152					
45-49	79	24	15	8	126	3		1		4
50-54	67	20	41	2	130				2	2
55-59	24	7	57	1	89	3		1		4
60-64	12	11	71		94			3		3
65-69	2	6	93		101					
70-74			81		81					
=>75	1	2	95		98					
TOTAL	560	151	990	55	1756	17	1	23	3	44

III.10 *Processus*

Pour clore ce chapitre, voici une représentation schématique de la production des résultats calibrés, les principales syntaxes spécifiques seront données en annexe

TABLEAU III.9 TRAITEMENT ET PRODUCTION POUR LE CALIBRAGE

Fichier «enquête» (Y. Colens)	Fichiers «BCSS» (D. Roctus)
Fusion des fichiers: «bigfile» (SYNTAXE α)	
Variables complémentaires (Enquête - BCSS - Combinaison) (SYNTAXE β)	
Imputation des personnes au n° national inconnu (SYNTAXE γ)	
Totaux population	Totaux sécurité sociale
Calibrage proprement dit (calcul des poids) (g-CALIB)	
Calcul et interprétation des estimations (SYNTAXE δ)	

IV Résultats

IV RESULTATS

IV.1 Résultats globaux

Voici les résultats pour (les trimestres de) 1999, 2000 et 2001. La variable d'intérêt est le statut BIT, en distinguant, parmi les travailleurs, salariés et non-salariés. Bien entendu, tous les résultats (par âge, sexe, région; autres variables...) peuvent être (re)calculés avec les poids calibrés.

TABLEAU IV.1 COMPARAISON CALIBRAGE/POST-STRATIFICATION

	CALIBRE				POST-STRATIFIE			
1999	99T1	99T2	99T3	99T4	99T1	99T2	99T3	99T4
Salariés	3248901	3242999	3310763	3290628	3339228	3298776	3387956	3359692
Non sala	666684	692474	668287	675834	650604	688636	648564	654156
Chômeurs	408878	388736	390949	345658	398408	377504	376108	347952
Inactifs	5889289	5889542	5843753	5901632	5825512	5848836	5801124	5851952
Total	10213752	10213751	10213752	10213752	10213752	10213752	10213752	10213752
2000	00T1	00T2	00T3	00T4	00T1	00T2	00T3	00T4
Salariés	3312085	3343629	3404094	3410151	3406504	3458996	3496140	3422044
Non sala	664323	675019	653963	651863	638072	661036	637164	649188
Chômeurs	323023	305307	334383	308369	319396	290492	320760	303080
Inactifs	5939654	5915129	5846646	5868703	5875112	5828560	5785020	5864772
Total	10239085	10239084	10239086	10239086	10239084	10239084	10239084	10239084
2001	01T1	01T2	00T3	00T4	01T1	01T2	00T3	00T4
Salariés	3391705	3379453	3415301	3381744	3435508	3424000	3472232	3432664
Non sala	645301	652399	634292	626612	623928	612424	610588	593456
Chômeurs	275762	260679	291232	331531	273152	265420	282332	326576
Inactifs	5963173	5983409	5935114	5936053	5930832	5961572	5898284	5910724
Total	10275941	10275940	10275939	10275940	10263420	10263416	10263436	10263420

Le contraste sera plus frappant si nous calculons la différence (calibrage – post-stratification).

On observera que :

les totaux de population, égaux pour 1999 et 2000 (à l'ajustement statistique près), différent pour 2001 (voir III.6);

la tendance est constante: par rapport aux résultats post-stratifiés, **l'EFT calibrée donne une estimation inférieure de l'emploi salarié, et supérieure de l'emploi non-salarié¹**; le chômage est généralement (10 trimestres sur 12) légèrement réévalué. Ce mouvement semble général (les deux sexes et les trois régions y sont sujets), mais l'ampleur des différences varie selon l'année et le trimestre.

¹ Rappelons que la distinction entre emploi (principal) salarié et non-salarié se fonde sur la déclaration de l'intéressé (ou du membre de son ménage répondant proxy), voir TABLEAU I.2.

TABLEAU IV.2 DIFFERENCE CALIBRAGE/POST-STRATIFICATION

1999	99T1	99T2	99T3	99T4
Salariés	-90327	-55777	-77193	-69064
Non sala	16080	3838	19723	21678
Chômeurs	10470	11232	14841	-2294
Inactifs	63777	40706	42629	49680
Total	0	-1	0	0
2000	00T1	00T2	00T3	00T4
Salariés	-94419	-115367	-92046	-11893
Non sala	26251	13983	16799	2675
Chômeurs	3627	14815	13623	5289
Inactifs	64542	86569	61626	3931
Total	1	0	2	2
2001	01T1	01T2	00T3	00T4
Salariés	-43803	-44547	-56931	-50920
Non sala	21373	39975	23704	33156
Chômeurs	2610	-4741	8900	4955
Inactifs	32341	21837	36830	25329
Total	12521	12524	12503	12520

Nous pouvons aussi compléter le TABLEAU I.1, en nous limitant aux trois années écoulées.

TABLEAU IV.3 ESTIMATIONS DE L'EMPLOI (BIS)

Emploi national	1999	2000	2001
EFT-moyenne de l'année (<i>post-stratifié</i>)	4.006.903	4.092.286	4.051.200
EFT-moyenne de l'année (<i>calibré</i>)	3.949.143	4.028.782	4.031.702
ICN – au 30 juin	3.908.600	3.969.400	4.015.100

Certains trouveront (ont trouvé!) ce tableau *trop beau pour être vrai*, au sens que les résultats calibrés se situent entre les résultats «traditionnels» et l'estimation administrative. Effectivement, dans la mesure où le calibrage adjoint aux données démographiques de post-stratification des informations de la sécurité sociale, il n'y a pas lieu de s'étonner.

Rappelons encore qu'en 2001, une définition plus stricte de l'emploi EFT (quelle que soit l'extrapolation) a été utilisée, ce qui explique la moindre différence avec l'ICN.

IV.2 Croisement EFT-BCSS

En plus de l'extrapolation globale, voici un peu de statistique descriptive.

Le lien avec notre objectif de cohérence (...ou non) apparaîtra plus clairement à la lecture du TABLEAU IV.4 dont la structure a été inspirée par E. Fagnoul². Nous croisons les **4 lignes** (salariés/non-salariés/chômeurs/inactifs, selon enquête, quatrième trimestre 1999) avec les **8 colonnes**³ (combinaison de «présent fin de trimestre ONSS», «présent fin de trimestre ONSSAPL» et «connu INASTI» au 31 décembre 1999); ainsi «onss» signifie présent dans cette institution seulement, etc.).

² J. Ouziel suggérait d'y adjoindre, outre les trois institutions gérant les travailleurs, l'ONEM. Nous y avons renoncé faute de temps.

³ Après imputation si nécessaire (voir III.9)

Notons que la suggestion de K. Hendrickx (prendre, pour ONSS et ONSSAPL, «présent en cours de trimestre» plutôt que «présent à la fin du trimestre»), qui aurait réduit le nombre de «rien» mais augmenté les doubles comptages, n'a pas été retenue, étant donné que les totaux de calibrage portent sur la présence en fin de trimestre.

TABLEAU IV.4 INFORMATION ADMINISTRATIVE ET ENQUÊTE (1999T4)

	rien	onss	apl	inasti	Onss & apl	Onss & ina	Apl & ina	trois	TOTAL
sala	249734	2638512	241403	16375*	21455	112944	9951	254	3290628
nonsa	111019	8726*	150*	544947		9902	1089		675833
cho	285186	46020	1495	10215	677	2064			345657
inac	5699012	114815	7003	79072		1145	585		5901632
TOTAL	6344951	2808073	250051	650609	22132	126055	11625	254	10213750

On vérifie les totaux de calibrage...

ONSS : $2808073+22132+126055+254 = 2956514$ (domiciliés en Belgique!)

ONSSAPL : $250051+22132+11625+254 = 284062$ (domiciliés en Belgique!)

INASTI : $650609+126055+11625+254 = 788543$ (domiciliés en Belgique!)

Attention : «2808073» (présents à l'ONSS seulement) est une estimation, pas un total; de même «22132» pour les cumulards ONSS+ONSSAPL n'est qu'une estimation (relativement plus risquée).

Si la situation de 9.409.060 individus (soit 92% de la population) est «normale», c'est-à-dire emploi salarié: présent à l'ONSS ou (exclusif) à l'ONSSAPL

emploi non-salarié: connu (seulement) de l'INASTI

non-emploi (chômage ou inactivité): rien,

il subsiste des «anomalies», notées

* : emploi salarié (ou non salarié) selon EFT, situation administrative non salarié (ou salarié).

Souligné : situation (emploi ou non) «anormale»

Italique : doubles (triples) comptes administratifs.

Nous expliquons –dans la mesure du possible– ces discordances (*trop vilain pour être vrai!?*) par

Le phénomène de permutation

Le décalage temporel

Les cumuls d'activité

Le travail atypique

Le travail frontalier

a. **Le phénomène de permutation** peut être vérifié en scindant le tableau IV.4 (version non pondérée) selon le numéro de la personne dans le ménage après rectification éventuelle par l'enquêteur.

Supposons une famille de 3 personnes, numérotées père=1, mère=2, enfant=3, dans le registre national. Si l'enquête révèle que les parents ne travaillent plus, et que l'enfant a une activité (indépendante, et déclarée à l'INASTI, mettons), ils seront renumérotés enfant=1, père=2, mère=3; avec pour conséquence qu'au moment de la fusion, l'enfant=1 se verra attribuer l'information BCSS de son père, celui-ci de son épouse, et cette dernière de son enfant. En résumé

TABLEAU IV.5 CHANGEMENT DE NUMEROTATION AU SEIN D'UN MENAGE

PERSONNE	Statut BIT	N° registre	N° revu EFT	Info correcte	Info utilisée
Père	Inactif	1	2	rien	rien
Mère	Inactif	2	3	rien	INASTI
Enfant	Occupé	3	1	INASTI	rien

A la lumière de cet exemple, considérons les tableaux (non pondérés, et portant sur toute l'année 1999):

TABLEAU IV.6 INFORMATION ADMINISTRATIVE ET ENQUÊTE (NON PONDERE), AN 1999, NUMÉRO (REVU EFT) DE LA PERSONNE = 1

	rien	onss	apl	inasti	Onss & apl	Onss & ina	Apl & ina	trois	TOTAL
sala	1698	14138	1208	174	102	822	79	11	18232
nonsa	196	57	2	3740		117	19	1	4132
cho	19483	383	20	598		13	1		20498
inac	1459	250	17	53	3	7			1789
TOTAL	22836	14828	1247	4565	105	959	99	12	44651

Plus de personnes occupées sans être répertoriées à la sécurité sociale (1894) que l'inverse (1345),

TABLEAU IV.7 INFORMATION ADMINISTRATIVE ET ENQUÊTE (NON PONDERE) AN 1999, NUMÉRO (REVU EFT) DE LA PERSONNE = 3

	rien	onss	apl	inasti	Onss & apl	Onss & ina	Apl & ina	trois	TOTAL
sala	373	2826	162	35	18	68	7		3489
nonsa	29	4	1	343		16	3		396
cho	11485	517	21	36	6	6			12071
inac	336	227	9	13	1	4			590
TOTAL	12223	3574	193	427	25	94	10		16546

Moins de personnes occupées sans être répertoriées à la sécurité sociale (402) que l'inverse (840).

b. **Le décalage temporel** et la détérioration des données dans le temps peuvent être appréciés en comparant le croisement du quatrième trimestre 1999 (*TABLEAU IV.4*) avec celui du premier trimestre 1999 (rappelons que les données de la sécurité sociale étaient déchargées au 31 décembre)⁴

⁴ Pour 1999, le décalage absolu moyen entre la semaine d'enquête et les données administratives était de 6 mois; depuis 2000, il est réduit à 3 mois pour l'ONSS et l'ONSSAPL, et reste 6 mois pour l'INASTI.

TABLEAU IV.8 INFORMATION ADMINISTRATIVE ET ENQUÊTE (1999T1)

	rien	onss	apl	inasti	Onss & apl	Onss & ina	Apl & ina	trois	TOTAL
sala	334463	2520563	241916	32050		1264	30319	1014	3248902
nonsa	94170	14021	469	530896		400	30830	246	666685
cho	285167	103939	9367	9656			934		408878
inac	5621090	179295	8511	78463	792	1138			5889289
TOTAL	6334890	2817818	260263	651065	11520	12591 7	11021	1260	10213754

On observe plus de discordances entre situation EFT début 1999 et registres fin 1999, que pour les enquêtes du quatrième trimestre.

c. **Les cumuls d'activité.** On peut conjecturer qu'une personne relevant de deux (a fortiori, les trois) institutions de sécurité sociale déclarera souvent plusieurs activités à l'EFT (le raisonnement est moins convaincant en sens inverse: on peut exercer deux emplois relevant tous les deux de l'ONSS). C'est pourquoi nous allons restreindre le TABLEAU IV.4 aux répondants exerçant (selon l'enquête) une (ou plusieurs) activité(s) accessoire(s). Il apparaît que, des 126055 personnes (estimées) connues de l'ONSS et de l'INASTI, mais pas de l'APL, moins de la moitié avoue deux emplois (ou davantage) à l'enquêteur. En revanche, 3411 cumulards EFT passent incognito à travers les mailles de la sécurité sociale, et aucun des 254 individus (= un seul, extrapolé) connu des 3 institutions n'avoue(nt) d'activité secondaire. Quant aux 48584 cumulards ONSS, rappelons qu'ils peuvent très bien travailler pour plusieurs employeurs (privé/Etat) relevant dudit office.

TABLEAU IV.9 INFORMATION ADMINISTRATIVE ET ENQUÊTE (1999T4) PLUSIEURS ACTIVITES SELON EFT

	aucune	onss	apl	inasti	Onss & apl	Onss & ina	Apl & ina	trois	TOTAL
sala	2602	48584	6857	2425	5001	57857	5098		128424
nonsa	809			19051		4002	142		24004
TOTAL	3411	48584	6857	21476	5001	61859	5240		152428

d. **Le travail atypique.** Signalons que (pour 1999T4), des 249734 salariés inconnus de la sécurité sociale, 66526 occupent un emploi temporaire, dont 29595 en ALE et 1496 un boulot occasionnel.

e. **Le travail frontalier**⁵. Une personne (résidant en Belgique, mais) travaillant à l'étranger a «une bonne excuse» d'être inconnu de la BCSS. Limitons le TABLEAU IV.4 aux répondants exerçant (selon l'enquête) leur activité (principale) hors du Royaume. En effet, la majorité (67540 sur 79323) est inconnue de la Sécurité Sociale (belge)⁶.

⁵ Important pour comparer les emplois intérieur et national

⁶ Sauf de l'INAMI.

**TABLEAU IV.10 INFORMATION ADMINISTRATIVE ET ENQUÊTE (1999T4)-
ACTIVITE (PRINCIPALE) A L'ETRANGER SELON EFT**

	aucune	onss	apl	inasti	Onss & apl	Onss & ina	Apl & ina	trois	TOTAL
sala	62992	7066	604	1557		469			72688
nonsa	4548			2087					6635
TOTAL	67540	7066	604	3644		469			79323

Pour conclure, des 249734 personnes déclarant une activité salariée à l'EFT, mais non renseignés comme présents par l'ONSS, l'ONSSAPL ou même l'INASTI, on peut en expliquer 62292 par le travail frontalier et (en négligeant l'intersection) $29595+1496= 31091$ par les activités occasionnelles ou non assujetties à la sécurité sociale (ALE, marins, ...). Le reste doit être mis sur le compte des permutations, du décalage dans le temps, et des erreurs de mesure. Peut-être faut-il observer que le but principal était et reste l'amélioration de la pondération (globale) formant la première section de ce chapitre, nonobstant l'intérêt des croisements ci-dessus.

IV.3 Variance

Nous nous sommes bornés jusqu'à présent à fournir des estimations, sans les assortir de propos sur la précision (biais, écart-type) ni d'intervalles de confiance (dont certains raffolent, d'autres préférant au contraire les estimations ponctuelles).

Des résultats cités à la première section, il ressort qu'un **biais** a été corrigé. Nous constatons que le poids moyen des individus connus de l'ONSS et de l'APL est en général inférieur à la moyenne (de la classe d'âge), alors que celui des administrés de l'INASTI est supérieur (reflétant une réticence des indépendants à participer à l'enquête? De même, il semble que les personnes inscrites à l'ONEM participent moins volontiers. Il serait utile d'étudier le fichier des non-répondants, et voir dans quelle mesure il s'agit de refus, de déménagement, d'absence de longue durée, etc. Bref, l'actuaire ou le statisticien (aidé du sociologue) a encore du travail pour commenter les résultats. Pour l'**écart type** (sur le nombre de personnes en emploi ou au chômage), les publications⁷ se contentent jusqu'à présent de formules à la \sqrt{npq} , parfois affinées par NUTS I ou NUTS II.

Des analyses plus fines pourraient utiliser:

des logiciels ad hoc (POULPE par exemple),

des méthodes de *bootstrap* (étudiant la stabilité des résultats par rapport à des sous-ensembles de l'échantillon.),

des propriétés des estimateurs par la régression (généralisée).

IV.4 NACEbis

Un croisement (limité aux personnes en emploi selon l'EFT et présentes à l'ONSS) entre la réponse à l'enquête et l'information BCSS a montré que les sections (1 lettre) NACE de l'entreprise ne correspondaient que dans 80% des cas environ (les permutations, le décalage temporel et les cumuls d'activité ne l'expliquant que très partiellement). Dans la mesure où *on* accorde plus de foi à la déclaration de l'employeur à la sécurité sociale⁸ qu'aux réponses du travailleur à l'enquêteur (codifiées ultérieurement à l'INS), on peut créer dans bigfile (pour les actifs occupés) une variable

⁷ Des résultats post-stratifiés, donc apparemment biaisés.

⁸ Une troisième piste serait le répertoire des entreprises DBRIS (DataBase des Redevables de l'Information Statistique), qui modifie parfois la NACE déclarée par une entreprise à l'ONSS en fonction d'autres critères (inscription à la TVA), corrigeant p.ex. le *biais Maribel*.

nouvelle NACEbis (à 2 chiffres) qui donne la priorité à l'information BCSS lorsque celle-ci est disponible. La construction de NACEbis (voir Annexe β) est indépendante du calcul des poids.

En fait, on peut distinguer neuf situations, selon que la personne est connue ou non des trois institutions (les fréquences –non pondérées– se rapportent à 1999); il s'agit, pour ONSS et l'ONSSAPL, de la présence en fin de trimestre, tandis que l'INASTI mentionne simplement l'inscription. La différence avec les tableaux à huit colonnes de la section 2 est que nous n'imputons pas lorsque les données BCSS manquent. Pour près des deux tiers des travailleurs, la NACE EFT sera «gommée» au profit de la NACE ONSS.

TABLEAU IV. 11 CALCUL DE NACEbis

	Fréq. total	Fréq. occupés	Construction de NACEbis
Rien	61,2	10,6	EFT
Onss	26,0	63,2	ONSS
Apl	2,4	6,1	APL (donc 75)
Inasti	6,1	14,0	EFT
onss+apl	0,2	0,5	ONSS ou APL (selon heures)
onss+ina	1,2	3,2	ONSS ou EFT (selon EFT q7)
apl+ina	0,1	0,4	APL ou EFT (selon EFT q7)
trois	0,0	0,0	ONSS APL ou EFT (q7 puis heures)
manquant	2,6	2,1	EFT

Remarques:

l'INASTI essaie actuellement d'harmoniser ses «codes professions» avec la NACE;

si une personne est connue de l'ONSS et de l'APL, le nombre d'heures (FTE) déterminera s'il faut lui attribuer le code ONSS ou le code APL (=75);

si un individu est connu de l'ONSS ou de l'APL, ainsi que de l'INASTI, c'est la réponse à la question 7 (salarié/non, voit TABLEAU I.2) qui est déterminante.

pour un certain nombre de personnes, la NACE EFT «entreprise» (question 11b) manque, nous prenons alors la NACE «unité locale» (question 11a)

V Conclusion

V CONCLUSIONS

V.1 *Les «vrais» résultats de l'EFT*

La **coexistence** de «résultats EFT post-stratifiés» (envoyés trimestriellement à Eurostat) et de «résultats EFT calibrés» (fournis annuellement à l'ICN, avec ventilation trimestrielle) risque de créer des confusions (quelles données seront reprises dans les communiqués de presse que nous aimerions trimestriels? dans la publication annuelle?)¹. Une solution (maximaliste, et peu compatible avec les moyens humains actuels de l'Institut, même si, grâce à l'expérience acquise, les processus peuvent être automatisés) serait d'accélérer le calendrier (voir III.5), afin de calibrer chaque trimestre dans les délais prescrits par les services de la Commission. La «correction à la louche» (augmenter les estimations post-stratifiées de l'emploi non-salarié et du chômage BIT d'un pourcentage fixe, et diminuer les estimations de l'emploi salarié) serait trop hasardeuse (redresserait le biais observé ces trois dernières, sans réduire la variance).

V.2 *Sur le plan belge*

Que décidera l'ICN? La tendance serait d'accepter EFT calibrée pour l'estimation des trois catégories BIT, mais conserver les données administratives pour le détail, notamment la répartition NACE.

Autre point politiquement sensible: après calibrage, le taux d'emploi belge –indicateur important au niveau européen – diminue...

V.3 *Sur le plan européen*

Si Eurostat déploie des efforts louables pour harmoniser les questionnaires (ordre et formulation des questions) des enquêtes sur les forces de travail, l'harmonisation des techniques d'extrapolations reste un objectif lointain (et difficile, eu égard aux traditions et possibilités de chaque Etat membre: participation volontaire ou obligatoire, échantillon d'individus ou de ménages, existence ou non d'un registre de population, structures diverses des institutions de sécurité sociale, connaissances plus ou moins développées des méthodes de calibrage,...). A notre connaissance, parmi les Quinze, seuls Danemark, Finlande et Suède² utilisent des informations de la sécurité sociale (en fait, uniquement le chômage enregistré).

En ce qui concerne les comptes de l'emploi (le volet emploi des comptes nationaux), il faudrait aussi progresser vers l'harmonisation des sources (enquête? sécurité sociale? enquête avec sécurité sociale comme information externe?).

V.4 *Sur le plan de la qualité*

Le calibrage représente-t-il une plus-value sur le plan de la qualité? Passons rapidement en revue les critères:

Pertinence

Inchangée objectivement, la pertinence (*relevance*) de l'EFT (calibrée) serait augmentée si l'ICN l'adoptait comme source de base pour les comptes de l'emploi.

¹ Même conflit possible à Eurostat, entre les divisions «Population et conditions sociales» et «Economie et Finances», qui recevraient les résultats belges respectivement post-stratifiés (de l'INS) et calibrés (via l'ICN)

² Laquelle fait déjà intervenir le registre de l'emploi dans son plan de sondage.

Précision

La précision (*accuracy*) est probablement améliorée (c'était un des buts poursuivis), mais dans quelle mesure ? Il semble empiriquement qu'un biais ait été réduit, mais la variance reste à étudier (voir IV.3)

Cohérence

La cohérence (*coherence*) entre EFT et données administratives est renforcée (c'était un des buts poursuivis), à la nuance que «nos» totaux de calibrage ne correspondent pas à ceux traités par nos collègues de la Cellule (cas des frontaliers), et –plus grave– le prix à payer pourrait être une coexistence entre plusieurs extrapolations de la même enquête (voir V.1).

Actualité

L'actualité (*timeliness*) dépendra du calendrier (voir V.1).

Complétude

La complétude (*completeness*) est inchangée, puisque nous modifions les techniques d'extrapolation mais conservons les variables (à l'exception de la NACEbis, voir IV.4).

Comparabilité

La comparabilité (*comparability*) dans le temps est mise en péril, parce que les résultats calibrés de 1999-2000-2001 indiquent que la post-stratification tendait à surestimer l'emploi (voir IV.1) d'environ 1%; d'aucuns désireront rétropoler les EFT des vingt dernières années. Quant à la comparabilité dans l'espace (de l'Union Européenne), elle reste difficile à réaliser (voir V.3).

Accessibilité

Au refrain (V.1) en ce qui concerne l'accessibilité (*accessibility*): il faudrait décider si le calibrage devient la méthode officielle et unique d'extrapolation pour l'EFT, ou si les résultats calibrés seront communiqués uniquement à l'ICN alors que le grand public (et Eurostat) utiliseront les résultats post-stratifiés.

V.5 *Bilan et perspectives*

Bouteille à moitié pleine (résultats globaux dans la ligne des prévisions) ou à moitié vide (choix discutables³, améliorations souhaitables⁴)? Au lecteur d'en juger! Le chemin, tracé par différents collègues, s'est montré plus cahoteux que prévu; allons-nous le désaffecter ou l'asphalter?

³ Plus (ou moins) de variables de calibrage (III.3)?

⁴ En particulier, les permutations dans la numérotation des membres du ménage, dont l'impact est restreint au point de vue «macro» (IV.1), mais ennuyeux pour les analyses «micro» (IV.2) et sur le plan des principes. Comme pour les personnes sans données BCSS (III.9), il y a plusieurs solutions les identifier et les appairer correctement (possible via la date de naissance?) les éliminer (risque de biais, et discourtois vis-à-vis de ces répondants); dire aux enquêteurs de ne plus modifier la numérotation (à partir de 2003); calibrer par ménage (voir III.7), résolvant le problème «macro» mais laissant le «micro» ouvert.

Annexes

ANNEXE

Nous reprenons ici les principales **syntaxes** SPSS employées spécifiquement¹ pour le calibrage de l'EFT (voir III.10 pour leur agencement). Certaines variantes sont possibles selon l'année (modification du répertoire –inévitables, changement du nom des variables– à éviter). Pour des raisons diverses, les dénominations françaises et néerlandaises des institutions alternent avec les commentaires² en anglais.

A– Récupération des fichiers BCSS et création de “Bigfile” (d’après J.M. Museux)

Cette syntaxe ODBC va chercher les données individuelles, les traite (recodage, variables supplémentaires, ajout de labels), puis les fusionne avec le fichier d'enquête, créant «bigfile».

* récupération des fichiers administratifs

preserve.

SET MXMEMORY=500000.

set workspace=200000.

****1) rsz

GET DATA /TYPE=ODBC /CONNECT=

'DSN=dbr15ksz'

/SQL =

'SELECT '

'T5.year_sta AS v1, '

'T5.groupe AS v2, '

'T5.household AS v3, '

'T5.pers AS v4, '

'T5.period_send AS v5, '

'T5.period_pref AS v6, '

'T5.stat_mom AS v7, '

¹ A l'exclusion donc de g-Calib

² Signalons que les lignes précédées d'un astérisque ne sont pas exécutées.

```
'T5.stat_end AS v8, '
'T5.code_empl AS v9, '
'T5.nace AS v10, '
'T5.work_volume AS v11, '
'T5.salary AS v12 '
'FROM dbr15ksz:srb71.trsz T5 '
' WHERE T5.year_sta =2000'.
```

```
recode v1 to v5 v7 to v9 v11 (convert) into
year_sta groupe househo pers per_send.
recode v7 to v9 v11 (convert) into s_rsz s_e_rsz empl_rsz fte_rsz.
```

```
string per_ref (a6).
compute per_ref=v6.
```

```
string nace_1d (a1) nace_4d (a4).
compute nace_1d = substr(v10,1,1).
compute nace_4d = substr(v10,2,4).
recode nace_4d (convert) into nace_rsz.
format nace_rsz (f4.0).
```

```
compute fte_rsz=fte_rsz/100.
format fte_rsz (f10.2).
compute sal_rsz=v12.
format sal_rsz (f10.2).
```

VARIABLE LABELS

```
year_sta "year_sta"
groupe "groupe"
househo "household"
pers "pers"
per_send "period_send"
per_ref "period_pref"
s_rsz "status (gekend/niet gekend) rsz"
```

```
s_e_rsz "persoon ingeschreven op het einde van het trimester (RSZ)"  
empl_rsz "code_emploi rsz"  
nace_rsz "nace rsz 4 chiffres"  
nace_1d "nace 17 divisions"  
fte_rsz "work_volume in FTE"  
sal_rsz "salary in Euro".
```

```
val lab s_rsz  
1 "gekend/connu"  
0 "niet gekend/inconnu".
```

```
val lab s_e_rsz  
0 "neen/non"  
1 "ja/oui".
```

```
val lab empl_rsz  
1 "bediende/employé"  
2 "bediende (GESCO)/employé (ACS)"  
3 "arbeider/ouvrier"  
4 "arbeider (GESCO)/ouvrier (ACS)"  
5 "statutaire ambtenaar/fonctionnaire statutaire"  
6 "andere/autre".
```

```
execute.
```

```
sort cases by year_sta groupe househo pers.
```

```
SAVE OUTFILE='C:\Documents and settings\denis.luminet\My Documents\My Documents\Calib2000\rsz_2000.sav'  
/COMPRESSED.
```

```
**** 2) rszppo  
GET DATA /TYPE=ODBC /CONNECT=  
'DSN=dbr15ksz'
```

```

/SQL =
'SELECT '
'T5.year_sta AS v1, '
'T5.groupe AS v2, '
'T5.household AS v3, '
'T5.pers AS v4, '
'T5.period_send AS v5, '
'T5.period_pref AS v6, '
'T5.stat_mom AS v7, '
'T5.stat_end AS v8, '
'T5.code_empl AS v9, '
'T5.nace AS v10, '
'T5.work_volume AS v11, '
'T5.salary AS v12 '
'FROM dbr15ksz:srb71.trszppo T5 '
' WHERE T5.year_sta =2000'

```

recode v1 to v5 v7 to v11 (convert) into

```

year_sta
groupe
househo
pers
per_send
s_ppo
s_e_ppo
empl_ppo
nace_ppo
fte_ppo.

```

```

string per_ref (a6).
compute per_ref=v6.

```

```

string nppo_1d (a1).
if (nace_ppo=75) nppo_1d = 'L'.

```

```
compute sal_ppo=v12.  
format sal_ppo (f10.2).
```

```
VARIABLE LABELS
```

```
year_sta "year_sta"  
groupe "groupe"  
househo "household"  
pers "pers"  
per_send "period_send"  
per_ref "periode de référence"  
s_ppo "status (gekend/niet gekend) rszppo"  
s_e_ppo "persoon ingeschreven op het einde van het trimester"  
empl_ppo "code_emploi"  
nace_ppo "nace ppo 2 chiffres"  
nppo_1d "nace 17 divisions"  
fte_ppo "work_volume"  
sal_ppo "salary in Euro".
```

```
val lab s_ppo  
1 "gekend/connu"  
0 "niet gekend/inconnu".
```

```
val lab s_e_ppo  
0 "neen/non"  
1 "ja/oui".
```

```
val lab empl_ppo  
1 "bediende/employé"  
2 "bediende (GESCO)/employé (ACS)"  
3 "arbeider/ouvrier"  
4 "arbeider (GESCO)/ouvrier (ACS)"  
5 "statutaire ambtenaar/fonctionnaire statutaire"  
6 "andere/autre".
```

execute.

sort cases by year_sta groupe househo pers.

```
SAVE OUTFILE='C:\Documents and settings\denis.luminet\My Documents\My Documents\Calib2000\rszppo_2000.sav'
/COMPRESSED.
```

**** 3) RSVZ

```
GET DATA /TYPE=ODBC /CONNECT=
'DSN=dbr15ksz'
/SQL =
'SELECT '
'T4.year_sta AS v1, '
'T4.groupe AS v2, '
'T4.household AS v3, '
'T4.pers AS v4, '
'T4.period_send AS v5, '
'T4.period_pref AS v6, '
'T4.stat_mom AS v7, '
'T4.nace AS v8, '
'T4.stat_indep AS v9, '
'T4.cat AS v10 '
'FROM dbr15ksz:srb71.trsvz T4 '
' WHERE T4.year_sta =2000'
```

recode v1 to v5 v7 to v10 (convert) into

```
year_sta
groupe
househo
pers
per_send
```

```
s_rsvz  
nac_rsvz  
ind_rsvz  
cat_rsvz.
```

```
string per_ref (a6).  
compute per_ref=v6.
```

VARIABLE LABELS

```
year_sta "year_sta"  
groupe "groupe"  
househo "household"  
pers "pers"  
per_send "period_send"  
per_ref "period_pref"  
s_rsvz "status (gekend/niet gekend) rsvz"  
nac_rsvz "nace rsvz"  
ind_rsvz "type indépendant"  
cat_rsvz "categorie".
```

```
val lab s_rsvz  
1 "gekend/connu"  
0 "niet gekend/inconnu".
```

```
val lab ind_rsvz  
1 "zelfstandige/indépendant"  
2 "helper/aidant".
```

```
val lab cat  
1 "hoofdberoep/profession principale"  
2 "bijberoep/deuxième profession "  
3 "Actief na pensioenleeftijd/actif après pension".
```

```
execute.
```

sort cases by year_sta groupe househo pers.

```
SAVE OUTFILE='C:\Documents and settings\denis.luminet\My Documents\My Documents\Calib2000\rsvz_2000.sav'
/drop= v1 to v10
/COMPRESSED.
```

** 4) Adp

```
GET DATA /TYPE=ODBC /CONNECT=
'DSN=dbr15ksz'
/SQL =
'SELECT '
'T0.year_sta AS v1, '
'T0.groupe AS v2, '
'T0.household AS v3, '
'T0.pers AS v4, '
'T0.period_send AS v5, '
'T0.period_pref AS v6, '
'T0.pension AS v7 '
'FROM dbr15ksz:srb71.tadp T0 '
' WHERE T0.year_sta = 2000'
```

```
recode v1 to v5 v7 (convert) into
year_sta
groupe
househo
pers
per_send
s_adp
```

```
string per_ref (a6).
compute per_ref=v6.
```

VARIABLE LABELS

```
year_sta "year_sta"  
groupe "groupe"  
househo "household"  
pers "pers"  
per_send "period_send"  
per_ref "period_pref"  
s_adp "status (gekend/niet gekend) adp".
```

```
val lab s_adp  
0 "nee/non pension"  
1 "ja/oui pension".
```

```
execute.
```

```
sort cases by year_sta groupe househo pers.
```

```
SAVE OUTFILE='C:\Documents and settings\denis.luminet\My Documents\My Documents\Calib2000\adp_2000.sav'
```

```
/drop= v1 to v7  
/COMPRESSED.
```

```
** 5) Rvp
```

```
GET DATA /TYPE=ODBC /CONNECT=  
'DSN=dbr15ksz'  
/SQL =  
'SELECT '  
'T0.year_sta AS v1, '  
'T0.groupe AS v2, '  
'T0.household AS v3, '  
'T0.pers AS v4, '  
'T0.period_send AS v5, '  
'T0.period_pref AS v6, '
```

```
'T0.pension AS v7 '  
'FROM dbr15ksz:srb71.trvp T0 '  
' WHERE T0.year_sta = 2000'
```

```
recode v1 to v5 v7 (convert) into  
year_sta  
groupe  
househo  
pers  
per_send  
s_rvp
```

```
string per_ref (a6).  
compute per_ref=v6.
```

```
VARIABLE LABELS  
year_sta "year_sta"  
groupe "groupe"  
househo "household"  
pers "pers"  
per_send "period_send"  
per_ref "period_pref"  
s_rvp "status (gekend/niet gekend) rvp".
```

```
val lab s_rvp  
0 "nee/non pension"  
1 "ja/oui pension".
```

```
execute.
```

```
sort cases by year_sta groupe househo pers.
```

```
SAVE OUTFILE='C:\Documents and settings\denis.luminet\My Documents\My Documents\Calib2000\rvp_2000.sav'  
/drop= v1 to v7
```

/COMPRESSED.

```

*** rva
GET DATA /TYPE=ODBC /CONNECT=
'DSN=dbr15ksz'
/SQL =
'SELECT '
'T7.year_sta AS v1, '
'T7.groupe AS v2, '
'T7.household AS v3, '
'T7.pers AS v4, '
'T7.period_send AS v5, '
'T7.period_pref AS v6, '
'T7.stat_mom AS v7, '
'T7.sit AS v8 '
'FROM dbr15ksz:srb71.trva T7 '
' WHERE T7.year_sta = 2000 '
.
recode v1 to v5 v7 to v8 (convert) into
year_sta
groupe
househo
pers
per_send
s_rva
sit_rva.

string per_ref (a6).
compute per_ref=v6.

VARIABLE LABELS
year_sta "year_sta"
groupe "groupe"

```

```
househo "household"  
pers "pers"  
per_send "period_send"  
per_ref "period_pref"  
s_rva "status (gekend/niet gekend) rva"  
sit_rva "laatste werkloosheidscode/dernier code chômage".
```

```
val lab s_rva  
1 "gekend/connu"  
0 "niet gekend/inconnu".
```

```
val lab sit_rva  
0 "niet gekend/inconnu"  
1 "niet-werkende werkzoekende/non-actif qui cherche du travail"  
2 "niet-werkende niet-werkzoekende/non-actif qui ne cherche pas de travail "  
3 "werknemer die cumuleert met een uitkering/employé qui cumule avec une indemnité"  
4 "voltijds bruggepensioneerd /prépensionné à temps plein"  
5 "deeltijds bruggepensioneerd /prépensionné à temps partiel"  
6 "loopbaanonderbreker (volledig/voltijdse baan)/interruption de carrière complet/temps plein)"  
7 "loopbaanonderbreker (volledig/deeltijdse baan)/interruption de carrière (complète/temps partiel)"  
8 "loopbaanonderbreker (verminderde prestaties)/interruption de carrière (prestations réduites)".
```

```
execute.
```

```
sort cases by year_sta groupe househo pers.
```

```
SAVE OUTFILE='C:\Documents and settings\denis.luminet\My Documents\My Documents\Calib2000\rva_2000.sav'  
/COMPRESSED.
```

```
**** MATCH FILES
```

```
new file.  
MATCH FILES
```

```
/FILE='C:\Documents and settings\denis.luminet\My Documents\My Documents\Data EFT\Datasav\Calib00.sav'  
/TABLE='C:\Documents and settings\denis.luminet\My Documents\My Documents\Calib2000\rsz_2000.sav'  
/BY year_sta groupe househo pers  
/MAP.  
execute.
```

MATCH FILES

```
/FILE=*  
/TABLE='C:\Documents and settings\denis.luminet\My Documents\My Documents\Calib2000\rszppo_2000.sav'  
/BY year_sta groupe househo pers  
/MAP.  
execute.
```

MATCH FILES

```
/FILE=*  
/TABLE='C:\Documents and settings\denis.luminet\My Documents\My Documents\Calib2000\rsvz_2000.sav'  
/BY year_sta groupe househo pers  
/MAP.  
execute.
```

MATCH FILES

```
/FILE=*  
/TABLE='C:\Documents and settings\denis.luminet\My Documents\My Documents\Calib2000\adp_2000.sav'  
/BY year_sta groupe househo pers  
/MAP.  
execute.
```

MATCH FILES

```
/FILE=*  
/TABLE='C:\Documents and settings\denis.luminet\My Documents\My Documents\Calib2000\rvp_2000.sav'  
/BY year_sta groupe househo pers  
/MAP.  
execute.
```

MATCH FILES

```
/FILE=*
```

```
/TABLE='C:\Documents and settings\denis.luminet\My Documents\My Documents\Calib2000\rva_2000.sav'
```

```
/BY year_sta groupe househo pers
```

```
/MAP.
```

```
execute.
```

FREQUENCIES

```
VARIABLES=s_rsvz s_rsz s_ppo s_rvp s_rva s_adp
```

```
/ORDER= ANALYSIS .
```

```
SAVE OUTFILE='C:\Documents and settings\denis.luminet\My Documents\My Documents\Data EFT\Datasav\Calib00\eft_bigfile_00.sav'
```

```
/COMPRESSED.
```

```
new file.
```

MATCH FILES

```
/FILE='C:\Documents and settings\denis.luminet\My Documents\My Documents\Calib2000\rszva_00.sav'
```

```
/TABLE='C:\Documents and settings\denis.luminet\My Documents\My Documents\Calib2000\rszppo_2000.sav'
```

```
/BY year_sta groupe househo pers
```

```
/MAP.
```

```
execute.
```

```
restore.
```

B– Variables supplémentaires

Cette syntaxe calcule (ou renomme) des variables ne figurant pas dans le fichier “standard” (SANO: séparation des actifs occupés en salariés/non salariés, FRON: séparation des actifs occupés selon le lieu de travail en Belgique/à l'étranger, CUMUL: séparation s des actifs occupés selon le nombre d'activités) (après préparation), détermine la variable ADM (statut vis-à-vis de l'ONSS, ONSSAPL et INAST, voir IV.2) permettant enfin de déterminer la NACEbis (voir IV.4)

*Variables supplémentaires

* age_5 à la place de agecat

```
RECODE
  agec
  (1 thru 16=Copy) INTO age_5 .
EXECUTE .
```

* Enfants de 0 à 14 ans: statbit missing

```
RECODE
  statbit
  (1=2) (2=1) (3=3) (SYSMIS=4) INTO BIT4 .
EXECUTE .
val lab BIT4
1 "occ"
2 "chom"
3 "inac adulte"
4 "inac enfant".
```

* Occupés: séparation salariés/non

```
IF (q7<5) SANO = 0 .
EXECUTE .
IF (q7>4) SANO = 1 .
EXECUTE .
```

```
IF (BIT4=2) SANO =2.
EXECUTE.
IF (BIT4>2) SANO =3.
EXECUTE.
val lab SANO
0 "sala"
1 "nonsal"
2 "chom"
3 "inac".
```

* Occupés: séparation emploi (principal) étranger/Belgique

```
IF (q10<9999) FRON = 0 .
EXECUTE .
IF (q10>9999) FRON = 1 .
EXECUTE .
IF (BIT4=2) FRON=2.
EXECUTE.
IF (BIT4>2) FRON=3.
EXECUTE.
val lab FRON
0 "occ etr"
1 "occ bel"
2 "chom"
3 "inac".
```

Occupés: nombre d'emplois EFT

```
IF (q26=1) CUMUL = 0 .
EXECUTE .
IF (q26=2) CUMUL = 1 .
EXECUTE .
IF (BIT4=2) CUMUL=2.
```

```
EXECUTE.
IF (BIT4>2) CUMUL=3.
EXECUTE.
val lab CUMUL
0 "plus empl"
1 "un empl"
2 "chom"
3 "inac".
```

*VARIABLES SUPPLEMENTAIRES (Fichier BCSS)

*connu ou non des trois institutions

```
compute adm=9.
execute .
```

code 9 pour les personnes dont le numéro n'a pas été transmis

```
IF (s_e_rsz=0 & s_e_ppo=0 & s_rsvz=0) adm = 0 .
EXECUTE .
IF (s_e_rsz=1 & s_e_ppo=0 & s_rsvz=0) adm = 1 .
EXECUTE .
IF (s_e_rsz=0 & s_e_ppo=1 & s_rsvz=0) adm = 2 .
EXECUTE .
IF (s_e_rsz=0 & s_e_ppo=0 & s_rsvz=1) adm = 3 .
EXECUTE .
IF (s_e_rsz=1 & s_e_ppo=1 & s_rsvz=0) adm = 4 .
EXECUTE .
IF (s_e_rsz=1 & s_e_ppo=0 & s_rsvz=1) adm = 5 .
EXECUTE .
IF (s_e_rsz=0 & s_e_ppo=1 & s_rsvz=1) adm = 6 .
EXECUTE .
IF (s_e_rsz=1 & s_e_ppo=1 & s_rsvz=1) adm = 7 .
```

EXECUTE .

val lab adm

0 'none'

1 'rsz only'

2 'ppo only'

3 'rsvz only'

4 'rsz&ppo'

5 'rsz&rsvz'

6 'ppo&rsvz'

7 'all'

9 'miss'.

*VARIABLE MIXTE (NACEbis)

COMPUTE Nacebis = q11a .

EXECUTE .

IF q11b<555 Nacebis = q11b .

EXECUTE .

q11b = NACE entreprise selon enquête (2 chiffres) - parfois missing

*q11a = NACE unite locale (2chiffres)

*nace_rsz = NACE entreprise selon ONSS (4 chiffres, à recoder en 2) *

confirmation: NACE APL = 75, en tous cas pour ce que nous recevons via BCSS

* prendre unite locale ssi entreprise manque

* si connu seulement ONSS, on prend NACE ONSS

IF (adm=1) Nacebis = TRUNC(nace_rsz/100) .

EXECUTE .

* si connu seulement APL, on prend leur NACE, càd 75

```
IF (adm=2) Nacebis = 75 .  
EXECUTE .
```

```
*si ONSS et APL, priorité aux heures  
IF (adm=4 & 100*fte_rsz>=fte_ppo) Nacebis = TRUNC(nace_rsz/100) .  
EXECUTE .  
IF (adm=4 & 100*fte_rsz<fte_ppo) Nacebis = 75 .  
EXECUTE .
```

```
*si ONSS et INASTI, priorité au statut (selon déclaration)  
IF (adm=5 & q7<5) Nacebis = TRUNC(nace_rsz/100) .  
EXECUTE .
```

```
*si APL et INASTI, priorité au statut (selon déclaration)  
IF (adm=6 & q7<5) Nacebis = 75 .  
EXECUTE .
```

```
*si ONSS APL ET INASTI, combinaison des critères  
IF (adm=7 & q7<5 & 100*fte_rsz>=fte_ppo) Nacebis = TRUNC(nace_rsz/100) .  
EXECUTE .  
IF (adm=7 & q7<5 & 100*fte_rsz<fte_ppo) Nacebis = 75 .  
EXECUTE .
```

FREQUENCIES

```
VARIABLES=Nacebis  
/ORDER= ANALYSIS .
```

CROSSTABS

```
/TABLES=Nacebis BY adm  
/FORMAT= AVALUE TABLES  
/CELLS= COUNT .
```

F – Imputation de l'information BCSS (d'après C. Vanderhoeft)

Cette syntaxe impute l'information BCSS lorsque celle-ci est manquante (voir III.9) dans bigfile, et élimine du fichier les variables démographiques (p.ex. état civil), d'enquête (p.ex. travail de nuit) et de sécurité sociale (p.ex. statut ONSS) non essentielles.

```
*****
* Imputefull.sps  VERSION 2000                                     *
*****
* This syntax:                                                *
* - reads data from @BASIC                                     *
* - ENABLED: selects data in strata @FIRST to @LAST          *
* - initialises INIWEI                                       *
* - defines PROVTERM (alternative stratum variable)          *
* - performs RANDOM HOT DECK imputation                     *
* - sets imputation cells by @ESTRAT & @COMMON              *
* - completes variables in @NEW                              *
* - could be run separately; some macros then need to be activated *
*****
```

***** Parameters used also in the calibration job.

* MUST BE SET ON.

```
define @WORKDIR()      /* Workdir used in g-Calib job */
'C:\Documents and Settings\denis.luminet\My Documents\My Documents\Calib2000\'.
!end define
```

```
define @IDENT()       /* Identification variable used in g-Calib job */
pers_id.
!end define.
```

```
define @EWEIGHT()     /* Initial weights used in g-Calib job */
iniwei.
!end define.
```

```
define @ESTRAT()      /* Initial stratification used in g-Calib job */
prov.
!end define.
```

```
define @XDATA()      /* Completed data */
'Completed.sav'.
!end define.
```

```
***** Parameters specific to imputation module.
*   Should never be set off.
```

```
define @SEED()       /* Initialise seed: RANDOM, */
2000000.             /* set 2,000,000 for year 2000 or 1...2,000,000,000; default = 150,000,000 */
!end define.
```

```
define @COMMON()     /* More variable(s) in DONOR & INCOMPLETE file DL uses age, salnon (no sex)*/
age_5 salnon term . /* ... defining matching cells */
!end define.
```

```
define @OTHER()      /* Variable(s) in DONOR & INCOMPLETE file */
sex iniwei prov ilo known from cumul q6 q7 q11a q9a q10 q11b nace_1d q15a q16 q20 q78 q81 nacebis coef2000 part_dir salnon bit4.
!end define.         /* ... e.g. used after g-Calib is finished */
```

```
define @NEW()        /* Variable(s) in DONOR file only */
s_e_rsz sit_rva s_e_ppo s_rsvz. /* ... to be imputed in INCOMPLETE file */
!end define.
```

```
define @BASIC()      /* Basic survey data file, possibly with MVs */
'bigfile_00.sav'.
!end define.
```

```
define @DDICT()      /* Data dictionary completion */
variable label pers_id "ID for sampled individuals"
```

```

/age_5 "5-year age group"
/prov "Province (NUTS 2) of residence"
/sex "Gender"
/s_e_rsz "Registered at RSZ at end of survey quarter"
/sit_rva "Last known un employment status at RVA"
/s_e_ppo "Registered at RSZPPO at end of survey quarter"
/s_rsvz "Registered at RSVZ ..."
/term "Survey quarter (2000)"
/ilo "Status according to ILO definitions"
/salnon "employee or selfemployed"
/known "NN sent to Crossroads Bank"
/fron "Workplace"
/cumul "Nomber jobs LFS"
/nacebis "NACE LFS other"
/bit4 "inactive: child or adult"
/q6 "career break"
/q7 "employment status"
/q9a "ISCO"
/q11a "NACE local"
/q11b "NACE company"
/q15a "type temporary"
/nace_1d "NACE ONSS 1lettre"
/q16 "full or part time"
/q20 "hours reference week"
/q78 "diploma"
/q81 "self defined status"
/iniwei "Constant weight (=1)"
/Coef2000 "PS weight"
/part_dir "proxy or not"
/provterm "Strata by province and quarter".

```

!end define.

```

define @ICELLS(!POS !ENCLOSE("","")) /* Check imputation cells */
TEMPORARY.

```

```
NUMERIC T0.  
LEAVE T0.  
VARIABLE LABEL T0 'Table Total'.  
VALUE LABELS T0 0 ' '.  
TABLES /FORMAT BLANK MISSING('.')  
/TABLES (@ESTRAT > age_5 + T0) BY (sex > (ilo) + T0) > (STATISTICS)  
/STATISTICS count((F5.0))  
/TITLE !1.  
!end define.
```

```
***** Exploring the extended survey data file before imputation.  
* Some corrections and completions are made.
```

```
matrix.  
print /title='Exploring & preparing survey data in ' + @BASIC.  
end matrix.
```

```
get file=@WORKDIR + @BASIC.  
* select if(@ESTRAT >= @FIRST & @ESTRAT <= @LAST).  
weight off.
```

```
compute newstr = 100 * prov + term.  
formats newstr (F4.0).  
autorecode newstr /into provterm.  
AUTORECODE  
  VARIABLES=newstr /INTO termprov  
/PRINT.
```

```
*compute iniwei = 1.  
* if (ilo=2) s_nonact=1.
```

```
@DDICT.  
execute.
```

```
display dictionary /variables=@COMMON @NEW @OTHER.
weight off.
sort cases by @ESTRAT.
split file layered by @ESTRAT.
frequencies /variables=@COMMON @NEW known.
split file off.
```

***** Separating complete from incomplete data.

```
matrix.
print /title='Imputing now ...'.
end matrix.
```

```
temporary.
select if (known=1).
save outfile=@WORKDIR + 'Donor.sav'
  /keep=@IDENT @EWEIGHT @ESTRAT @COMMON @NEW @OTHER /compressed.
temporary.
select if (known=0).
save outfile=@WORKDIR + 'Incomplete.sav'
  /keep=@IDENT @EWEIGHT @ESTRAT @COMMON @OTHER /compressed.
* (Incomplete variables are erased completely: see below for
* the 'method implemented here').
```

*** Invariable program part *****

```
*** Method implemented here:
*** for any incomplete case, all values of incomplete variables are
*** overwritten by values from selected donor case.
*** Advantage:
*** this excludes combinations for incomplete variables that do not
*** exist in the donor file.
```

***** To start a new session with given (new) seed.

```
set seed=@SEED.
```

***** Imputation.

* Prepare donor file: random sorting in matching cells.

```
GET FILE=@WORKDIR + 'Donor.sav'.
@ICELLS "Occupation of imputation cells in DONOR.sav".
sort cases by @ESTRAT @COMMON.
show seed.
compute rndrank=UNIFORM(1).
formats rndrank (F10.8).
sort cases by @ESTRAT @COMMON rndrank.
save outfile=@WORKDIR + 'Donor-X.sav' /keep=@ESTRAT @COMMON @NEW.
```

* Impute.

```
GET FILE=@WORKDIR + 'Incomplete.sav'.
@ICELLS "Occupation of imputation cells in INCOMPLETE.sav".
sort cases by @ESTRAT @COMMON.
compute id=$CASENUM.
MATCH FILES /FILE=* /FILE=@WORKDIR + 'Donor-X.sav' /BY @ESTRAT @COMMON.
select if (~missing(id)).
save outfile=@WORKDIR + 'Imputed.sav'
  /keep=@IDENT @EWEIGHT @ESTRAT @COMMON @NEW @OTHER /compressed.
```

* Result for imputed cases.

```
matrix.
print /title='Exploring completed data in Imputed.sav'.
end matrix.
```

```
sort cases by @ESTRAT.
split file layered by @ESTRAT.
frequencies /variables=@NEW.
split file off.
```

* Append Imputed file to Donor file.

```
GET FILE=@WORKDIR + 'Donor.sav'.
ADD FILES /FILE=* /FILE=@WORKDIR + 'Imputed.sav'.
save outfile=@WORKDIR + @XDATA /compressed.
```

* Results for completed survey data.

```
matrix.
print /title='Exploring completed data in ' + @XDATA.
end matrix.
```

```
@ICELLS "Occupation of imputation cells in COMPLETED file".
sort cases by @ESTRAT.
split file layered by @ESTRAT.
frequencies /variables=@NEW.
```

*** End of invariable program part *****

***** Case-specific command(s) showing results of imputation.

```
TABLES
/FORMAT BLANK MISSING('.')
/OBSERVATION s_e_rsz /FTOTAL $t 'Group Total'
/TABLES (age_5 + $t) BY (sex > s_e_rsz > (STATISTICS) + $t)
/STATISTICS count((F5.0)) sum( ) valid().
TABLES
/FORMAT BLANK MISSING('.')
/OBSERVATION sit_rva /FTOTAL $t 'Group Total'
```

```
/TABLES (age_5 + $t) BY (sex > sit_rva > (STATISTICS) + $t)
/STATISTICS count((F5.0)) sum( ) valid().
```

TABLES

```
/FORMAT BLANK MISSING('.')
/OBSERVATION s_e_ppo /FTOTAL $t 'Group Total'
/TABLES (age_5 + $t) BY (sex > s_e_ppo > (STATISTICS)+ $t)
/STATISTICS count((F5.0)) sum( ) valid().
```

TABLES

```
/FORMAT BLANK MISSING('.')
/OBSERVATION s_rsvz /FTOTAL $t 'Group Total'
/TABLES (age_5 + $t) BY (sex > s_rsvz > (STATISTICS)+ $t)
/STATISTICS count((F5.0)) sum( ) valid().
```

split file off.

FREQUENCIES

```
VARIABLES=s_e_rsz s_e_ppo s_rsvz sit_rva
/ORDER= ANALYSIS .
```

*appendice D. Luminet, admp prolongeant adm et permettant la construction des tableaux de la section IV.2

*connu ou non des trois institutions, après imputation

COMPUTE admp=9.

EXECUTE.

*code 9 pour les personnes sans donneur

```
IF (s_e_rsz=0 & s_e_ppo=0 & s_rsvz=0) admp = 0 .
```

EXECUTE .

```
IF (s_e_rsz=1 & s_e_ppo=0 & s_rsvz=0) admp = 1 .
```

EXECUTE .

```
IF (s_e_rsz=0 & s_e_ppo=1 & s_rsvz=0) admp = 2 .
```

EXECUTE .

```
IF (s_e_rsz=0 & s_e_ppo=0 & s_rsvz=1) admp = 3 .
```

```
EXECUTE .  
IF (s_e_rsz=1 & s_e_ppo=1 & s_rsvz=0) admp = 4 .  
EXECUTE .  
IF (s_e_rsz=1 & s_e_ppo=0 & s_rsvz=1) admp = 5 .  
EXECUTE .  
IF (s_e_rsz=0 & s_e_ppo=1 & s_rsvz=1) admp = 6 .  
EXECUTE .  
IF (s_e_rsz=1 & s_e_ppo=1 & s_rsvz=1) admp = 7 .  
EXECUTE .
```

```
val lab admp  
0 'nonep'  
1 'rsz onlyp'  
2 'ppo only'  
3 'rsvz onlyp'  
4 'rsz&ppop'  
5 'rsz&rsvzp'  
6 'ppo&rsvzp'  
7 'allp'  
9 'missp'.
```

A – Principaux résultats

Cette ultime syntaxe (venant après g-Calib et pouvant d'ailleurs y être intégrée) prend les coefficients calibrés et permet de “tout” calculer. Nous présentons ici la version allégée, avec comme seul résultat d'intérêt la fréquence de “SANO” (voir β).

```
GET FILE= @WORKDIR+ @OUTWEI .
sort cases by @IDENT .
SAVE OUTFILE = @WORKDIR+ @OUTWEI .
GET FILE=@WORKDIR + 'E-DesMat.sav' .
sort cases by @IDENT .
match files file=@WORKDIR + @OUTWEI /table=* /by @IDENT .
WEIGHT
BY calwei .

FREQUENCIES
  VARIABLES=sano
  /ORDER= ANALYSIS .
```

Glossaire des institutions

GLOSSAIRE DES INSTITUTIONS

Les organismes internationaux figurent en **gras**.

Français (site web)	Néerlandais	Commentaire
AdP Administration des Pensions http://ap.fgov.be/apinter/f/contact.htm	AdP Administratie der Pensioenen	Gère, au sein du Ministère des Finances, les pensions de l'Etat, des Régions, des Communautés et des Organismes d'intérêt public
BCSS Banque Carrefour de la Sécurité Sociale http://ksz-bcss.fgov.be	KSZ Kruispuntpank van de Sociale Zekerheid	Chargée notamment de la coordination entre les institutions de sécurité sociale et le Registre national
BFP Bureau Fédéral du Plan http://www.plan.be	FPB Federaal Planbureau	Un des partenaires de l'ICN, spécialisé dans les prévisions économiques
BIT Bureau International du Travail http://www.ilo.org	IAB Internationaal Arbeidsbureau	En anglais ILO (International Labour Office). Administration (siégeant à Genève) de l'OIT (Organisation internationale du Travail), dépend de l'ONU.
BNB Banque Nationale de Belgique http://www.bnb.be/sg/index.htm	NBB Nationale Bank van België	Un des partenaires de l'ICN. Collecte, analyse et diffuse des informations économiques et financières.
CCE Conseil Central de l'Economie http://www.ccecrb.fgov.be/	CRB Centrale Raad voor het Bedrijfsleven	Parfois invité par la Cellule de Coordination
EUROSTAT Office Statistique des Communautés Européennes http://www.europa.eu.int/comm/eurostat/	EUROSTAT Statistische Dienst van de Europese Gemeenschappen	Service de la Commission Européenne, basé à Luxembourg. Coordonne les programmes statistiques des Quinze (+ candidats + associés)
ICN Institut des Comptes Nationaux http://inr-icn.fgov.be	INR Instituut voor de Nationale Rekeningen	Décrit en I.1
INAMI Institut National d'Assurance Maladie- Invalidité http://inami.fgov.be	RIZIV Rijksinstituut voor Ziekte- en Invaliditeitsverzekering	Indemnise (en liaison avec les mutualités) les soins de santé, les congés de maternité, etc.
INASTI Institut National d'Assurances Sociales pour Travailleurs Indépendants http://www.inasti.be/	RSVZ Rijksdienst voor de Sociale Verzekering der Zelfstandigen	Comme son nom l'indique...

INS Institut National de Statistique http://www.statbel.fgov.be/	NIS Nationaal Instituut voor de Statistiek	Un des partenaires de l'ICN.
META Ministère fédéral de l'Emploi et du Travail http://meta.fgov.be	META Federaal Ministerie van Tewerkstelling en Arbeid	Associé à la Cellule de Coordination. Futur SPF Emploi, Travail et Concertation Sociale
ONEM Office National de l'Emploi http://www.onem.fgov.be/	RVA Rijksdienst Voor Arbeidsvoorziening	Chargé entre autres de l'indemnisation du chômage (le placement et la formation étant du ressort des offices régionaux et communautaires)
ONP Office National des Pensions http://www.onprvp.fgov.be/onprvp	RVP Rijksdienst voor Pensioenen	Gère et paye les pensions des (ex-)travailleurs du secteur privé
ONSS Office National de Sécurité Sociale http://www.onssrszls.fgov.be/	RSZ Rijksdienst voor Sociale Zekerheid	Gère la sécurité sociale des salariés du secteur privé, de l'Etat, des Régions et des Communautés.
ONSSAPL ¹ Office national de sécurité sociale des administrations provinciales et locales http://www.onssapl.fgov.be/	RSZPPO Rijksdienst voor Sociale Zekerheid van de Provinciale en Plaatselijke Overheidsdiensten	Gère la sécurité sociale des salariés des provinces et des communes (y compris CPAS)

Le site <http://socialsecurity.fgov.be/broch-fr.htm> donne un bon aperçu de la sécurité sociale en Belgique.

¹ Parfois abrégé en APL dans ce rapport.

Bibliographie

BIBLIOGRAPHIE

[Avignon]

D. Perreux et E. Waeytens

Compte-rendu du séminaire Eurostat «Pondération de l'enquête sur les forces de travail» – Avignon (juin 1996)

INS, 1996

[HT]

D.G. Horvitz et D.J. Thompson

A generalization of sampling without replacement from a finite universe

Journal of the American Statistical Association, 27, 663-685 (1952)

[NS]

P.L.D. Nascimento Silva et C.J. Skinner

Variable Selection for Finite Regression in Finite Populations

Survey Methodology, June 1997, Vol.23, No 1.

[Règlement 577/98]

Communautés Européennes

Règlement (CE) n° 577/98 du Conseil du 9 mars 1998 relatif à l'organisation d'une enquête par sondage sur les forces de travail dans la Communauté

Journal officiel n° L 077 du 14/03/1998 p. 0003 - 0007

[Termote]

A. Termote

Eindrapport van de Coördinatieceel INR: Methodologie ESA Employment

ICN-INR, 2000

[Tillé]

Y. Tillé

Théorie des sondages –Echantillonnage et estimations en population finie

Dunod, Paris, 2001

[Vanderhoeft]

C. Vanderhoeft

Generalised Calibration

http://www.statbel.fgov.be/studies/cal_en.asp

(dernière modification : janvier 2003)

[VWM]

C. Vanderhoeft, E. Waeytens et J.M. Museux

Generalized calibration with SPPS 9.0 for Windows base

In *Enquêtes, Modèles Et Applications*, édité par J.J. Droesbeke et L. Lebart,

Dunod, Paris, 2001

Informations

L'Institut national de statistique

Où trouver l'information de l'INS ?

Dans cinq grandes villes du pays, l'INS met à la disposition du public :

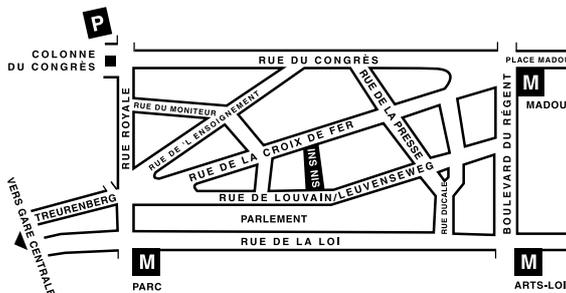
- ◇ Des annuaires et des publications spécialisées ainsi qu'une sélection de disquettes et de cédéroms.
- ◇ Une salle de lecture où il est possible de consulter les publications de l'INS, d'autres ministères ou d'institutions belges et internationales.

Toutes nos bibliothèques sont accessibles les jours ouvrables de 9h à 12h et de 13h à 16h.

Bruxelles

Rue de Louvain 44, 1000 Bruxelles
 tél. 02/548.63.65 – 02/548.63.66 fax 02/548.63.67
 e-mail : info@statbel.mineco.fgov.be

Train (B) : Gare Centrale ou Congrès
 Métro (M) : ligne 1A et 1B, station Parc ou Arts-Loi
 ligne 2, station Arts-Loi ou Madou
 Tram-Bus : 92, 93, 94 arrêt Parc
 65, 66 arrêt Presse
 29, 63 arrêt Congrès
 Parking (P) : Cité administrative (payant)



Anvers

Italiëlei 124 - bus 85, 2000 Antwerpen
 tél. 03/229.07.00 fax 03/233.28.30
 e-mail : info.antwerpen@statbel.mineco.fgov.be

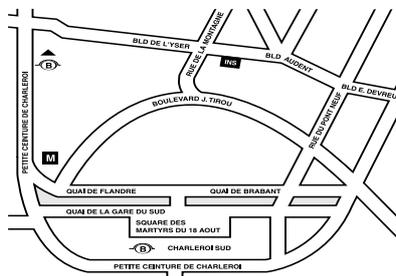
Train (B) : Centraal Station
 Métro (M) : arrêt Opera
 Tram-Bus : accès facile (Fr. Rooseveltplaats)



Charleroi

Tour Biarent, Bd Audent 14/5, 6000 Charleroi
 tél. 071/27.44.14 fax 071/27.44.19
 e-mail : info.charleroi@statbel.mineco.fgov.be

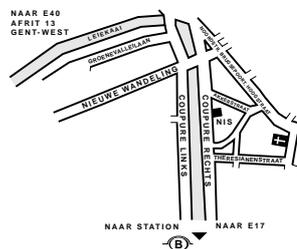
Train (B) : Charleroi Sud, 20 min depuis la gare (Place Buisset, Rue du Collège, place Charles II, bd Tirou, rue de la Montagne)
 Bus : arrêt Tirou
 Autoroute : petite ceinture de Charleroi - sortie Gare du Sud
 Parking (P) : payant face à l'INS



Gand

Coupure rechts 620, 9000 Gent
 tél. 09/267.27.00 fax 09/267.27.29
 e-mail : info.gent@statbel.mineco.fgov.be

Train (B) : Gent St. Pieters
 Tram-Bus : 40, 43 arrêt Theresianenstraat
 Autoroute : accès aisé par autoroute E40 (sortie N° 13 - Gent - West/Drongen)
 Parking (P) : au long de la "Coupure Rechts"



Liège

Bd de la Sauvenière 73-75, 4000 Liège
 tél. 04/223.84.11 fax 04/222.49.94
 e-mail : info.liege@statbel.mineco.fgov.be

Train (B) : Gare des Guillemins ou Gare du Palais
 Tram-Bus : (Guillemins) 1 et 4 arrêt Sauvenière
 Parking (P) : Neujean (à 20 m - même trottoir)
 Mercure (en face)



L'INS diffuse de nombreux produits qui donnent une image chiffrée de la réalité socio-économique belge.

Ces produits, repris dans notre catalogue, sont disponibles auprès de nos centres régionaux ou auprès de notre service de Documentation - vente de Bruxelles. Notre catalogue vous sera envoyé sur simple demande. (voir adresses ci-contre).

Vous trouverez également un extrait de nos données, ainsi que la liste de nos publications sur notre site Internet : <http://statbel.fgov.be>

Société : autres publications

Causes de décès

Cette publication annuelle vous informe sur le nombre de décès qui ont eu lieu durant l'année de référence, par sexe et par groupe d'âges, selon la cause initiale de la mort qu'elle soit naturelle ou le fait d'une cause extérieure. Ces données sont classées suivant les codes de la « Classification internationale des maladies » de l'Organisation mondiale de la santé. Vous y trouverez également une information sur la mortalité infantile, sur les décès des personnes en visite dans notre pays ainsi que sur les morts violentes, notamment sur les suicides et sur les morts considérées juridiquement comme suspectes.

Enquête sur le budget des ménages

L'Enquête sur le budget des ménages mesure la structure des dépenses des ménages belges. Chaque mois, un peu plus de 300 ménages ont noté leurs revenus et dépenses, ce qui porte à 3 745 le nombre de ménages suivis pendant l'année 2000. Ces données fournies par l'enquête sont extrapolées aux quatre millions de ménages environ que compte notre pays. Ces statistiques sont disponibles par région et pour le Royaume et vous informent sur le détail des dépenses par produit ou service utilisé, en fonction de différents critères comme l'âge du chef de ménage, le statut professionnel, ... Ces données sont également disponibles sur fichier informatique.

Quelques autres publications

Publications générales

Annuaire de statistiques régionales
Bulletin de statistique – *Mensuelle*
Communiqué hebdomadaire

Territoire et environnement

Statistique de l'occupation du sol (**disquette**)
Aperçu Environnement - *Annuelle*

Population

Ménages et noyaux familiaux - *Annuelle*
Population totale au 1er janvier - *Annuelle*
Tables de mortalité - *Annuelle*

Économie et finances

Vente de biens immobiliers – *Annuelle*

Agriculture

Recensement agricole et horticole au 15 mai –
Annuelle

Industrie

Production industrielle et construction - *Mensuelle*

Commerce, services et transports

Statistiques mensuelles du transport - *Mensuelle*
Commerce intérieur - *Annuelle*



Achevé d'imprimer
par l'imprimerie de l'INS
B-1000 Bruxelles

Décembre 2003