

Statbel

(Direction générale Statistique – Statistics Belgium)

Service Statistiques sociales (SSS)

Collecte de données Enquêtes Citoyens (CEC)

Service méthodologie (SM)

Enquête sur le budget des ménages (EBM) 2018

Note méthodologique

Personne de contact: Kelly Sabbe
Statbel (Direction générale Statistique – Statistics Belgium)
Chaque jour ouvrable entre 9h et 16h.

Table des matières

Liste des tableaux.....	6
Liste des figures.....	8
Liste des abréviations.....	8
1. Introduction générale.....	9
1.1. Qu'est-ce que l'Enquête sur le budget des ménages?.....	9
1.2. Base juridique.....	9
1.3. Quand l'EBM a-t-elle lieu?.....	9
1.4. Nombre de ménages participants.....	10
2. Les sous-échantillons pour EBM 2018.....	12
2.1. Introduction.....	12
2.2. Recrutement à partir des échantillons EFT – Source 1.....	15
2.2.1. Introduction.....	15
2.2.2. Degré 1 : tirage des unités géographiques (ou UPE)......	16
2.2.2.1. Cadre d'échantillonnage des UPE.....	16
2.2.2.2. Le plan de sondage pour le tirage des UPE.....	17
2.2.2.3. Répartition des UPE dans le temps.....	19
2.2.3. Degré 2 : tirage des ménages EFT (UPE).....	20
2.2.3.1. Le cadre d'échantillonnage des ménages EFT.....	20
2.2.3.2. Le plan de sondage pour le tirage des ménages EFT.....	21
2.2.3.3. Répartition des ménages-EFT dans le temps.....	22
2.2.4. Degré 3 : « sélection » des ménages EBM (UTE).....	23
2.2.4.1. Le cadre d'échantillonnage et le plan de sondage pour la sélection des ménages EBM.....	23
2.2.4.2. Répartition des ménages EBM dans le temps.....	23
2.2.5. Probabilités de sélection et poids de sondage.....	24
2.3. Recrutement via l'échantillon EBM 2016 – Source 2.....	26
2.3.1. Introduction.....	26
2.3.2. Sous-échantillons trimestriels d'EBM 2018 de la source 2.....	26
2.3.2.1. Le sous-échantillon d'EBM 2018 T1 de la source 2.....	26
2.3.2.2. Le sous-échantillon d'EBM 2018 T2 de la source 2.....	28
2.3.2.3. Le sous-échantillon d'EBM 2018 T3 de la source 2.....	28
2.3.2.4. Le sous-échantillon d'EBM 2018 T4 de la source 2.....	29

2.3.3.	Répartition des ménages d'EBM 2018 entre les 24 demi-mois	29
2.4.	Tirages complémentaires du RN – Source 3	32
2.4.1.	Introduction	32
2.4.2.	Degré 1 : tirage des unités géographiques (ou UPE)	33
2.4.3.	Degré 2 : tirage des ménages EBM (UPE).....	33
2.4.4.	Probabilités de sélection et poids de sondage.....	36
2.4.4.1.	Pour les ménages de type EFT.....	36
2.4.4.2.	Pour les ménages âgés.....	37
3.	Taux de réponse	39
3.1.	Taux de réponse initial de l'EFT - Source 1	39
3.2.	Taux de réponse initial de l'EBM - Source 2.....	40
3.3.	Taux de réponse initial du tirage complémentaire du RN - Source 3.....	41
3.4.	Echantillon initial total EBM2018 par mois	42
3.5.	Aperçu du travail de terrain et du taux de réponse	43
3.6.	Raisons de la non-réponse	45
4.	Collecte et traitement des données pour le calibrage	46
4.1.	Collecte des données	46
4.2.	Enregistrement des données	47
4.2.1.	Questionnaire ménage.....	47
4.2.2.	Carnet de dépenses.....	48
4.2.2.1.	Fourchettes de prix	48
4.2.2.2.	Nomenclature.....	49
4.3.	Détermination de la personne de référence sur la base du revenu.....	50
5.	Calibrage et estimation	52
5.1.	Introduction	52
5.2.	Cadres de calibrage	52
5.2.1.	Un échantillon EFT 2018 calibré	52
5.2.2.	Population de ménages de type EFT issus du RN.....	54
5.2.3.	Population de ménages âgés issus du RN.....	55
5.3.	Calibrage d'EBM 2018	56
5.3.1.	Calibrage de ménages de type EFT de l'EFT (source 1) et du RN (source 3.1)	56
5.3.2.	Calibrage de ménages de type EFT d'EBM 2016 (source 2).....	57
5.3.3.	Calibrage des ménages âgés (source 3.2).....	58

5.3.4.	Pourcentage de propriétaires de logements dans la population	59
5.4.	Estimation des paramètres.....	61
5.4.1.	Variables d'intérêt, paramètres et poids	61
5.4.2.	Estimation des totaux et des moyennes	62
5.4.2.1.	Exemple 1	64
5.4.2.2.	Exemple 2	65
5.4.3.	Estimation des nombres et proportions de ménages ayant une caractéristique spécifique	66
5.4.3.1.	Exemple 3	66
5.4.3.2.	Exemple 4 (suite de l'exemple 1 au par. 5.4.2)	67
5.4.4.	Estimations des domaines	68
5.4.5.	Moyennes alternatives: par personne et par unité de consommation modifiée	69
5.4.5.1.	Exemple 5 (suite de l'exemple 1 au par. 5.4.2 et de l'exemple 4 au par. 5.4.3)	69
5.5.	Estimation de la variance.....	70
5.5.1.	Introduction	70
5.5.2.	Description du plan de sondage	71
5.5.2.1.	Sources 1 et 3.1	71
5.5.2.2.	Source 2	72
5.5.2.3.	Source 3.2	72
5.5.3.	Calibrage.....	72
5.5.4.	Correction de la non-réponse	73
5.5.5.	Combinaison des sources.....	73
5.5.6.	Aperçu de l'estimation de la variance	74
6.	Traitement des données pour publication et diffusion	76
6.1.	Représentativité du sondage	76
6.1.1.	Répartitions au niveau du ménage	76
6.1.2.	Répartitions au niveau individuel	77
6.2.	Traitement des données	78
6.2.1.	Identification des valeurs aberrantes.....	78
6.2.2.	Imputation des dépenses énergétiques.....	79
6.2.3.	Collecte des informations des carnets.....	79
6.2.4.	Ventilation de factures groupées	80
6.2.5.	Imputer les données manquantes	81
6.3.	Loyer imputé	82

6.3.1. Extrapolation des données.....	82
7. Publication et diffusion	83
7.1. Publication sur le site web	83
7.2. Dépenses courantes / saisonnières.....	83
7.3. Microdonnées disponibles pour la recherche	86
Annexes.....	87
Définitions.....	87
L'échantillonnage à deux degrés en bref – Approche des poids de sondage pour les USE	89
Cadre théorique.....	89
Évolution de la base de sondage entre le degré 1 et le degré 2.....	90
Extension d'un échantillon à deux degrés avec un tirage supplémentaire d'USE	90
Situation particulière: UPE trop petites	91
Références	92
Références aux pages web pertinentes	93

Liste des tableaux

Tableau 1: Liste des abréviations	8
Tableau 2: Répartition des échantillons par région	10
Tableau 3: Aperçu des sources pour les échantillons bimensuels P01 à P24 d'EBM 2018	13
Tableau 4: Degré 1 - tirage des UPE ou GRP, répartition des GRP entre les strates, trimestres et GR - y compris DEU	18
Tableau 5: Répartition uniforme des groupes EFT par trimestre et par GR entre les semaines - DEU inclus	19
Tableau 6: Répartition quasi uniforme des groupes EFT par trimestre et par GR entre les semaines - hors DEU	20
Tableau 7: Degré 2 - tirage des MN: répartition des MN entre strates UPE, trimestres et GR - y compris la DEU	22
Tableau 8: Répartition quasi uniforme des ménages EFT par trimestre et par GR entre les semaines - hors DEU	22
Tableau 9: Fractions approximatives de la répartition des ménages EBM par semaine entre les demi-mois	23
Tableau 10: Aperçu schématique de la création du sous-échantillon d'EBM 2018 via la source 2	27
Tableau 11: Ménages répondants à EBM 2016 invités à participer à EBM 2018 - attribution de la période (« demi-mois ») en 2018	30
Tableau 12: Les 6 échantillons bimestriels complémentaires d'EBM 2018: informations de base	32
Tableau 13: Les 6 échantillons bimestriels complémentaires d'EBM 2018: paramètres d'échantillon et distribution	34
Tableau 14: Les 6 échantillons bimestriels complémentaires d'EBM 2018: Chiffres de la population (limités aux UPE tirées dans les GR1*7)	35
Tableau 15: Poids de sondage dans les échantillons complémentaires EBM 2018 de ménages de type EFT	37
Tableau 16: Poids de sondage dans les échantillons EBM 2018 complémentaires de ménages âgés.	38
Tableau 17: Taux de réponse initial de l'EFT - Source 1	39
Tableau 18: Diminution en raison du changement de modalité d'enquête	39
Tableau 19: Taux de réponse initial de l'EBM - Source 2	40
Tableau 20: Taux de réponse initial du tirage complémentaire du RN - Source 3	41
Tableau 21: Echantillon initial total EBM2018 par mois	42
Tableau 22: Aperçu du travail de terrain et du taux de réponse pour la Belgique et les régions pour EBM 2018	43
Tableau 23: raison du refus de participer à l'EBM	45

Tableau 24: Modalité pour remplir le carnet et le questionnaire des ménages.....	46
Tableau 25: Aperçu des centres de traitement	46
Tableau 26: Validation du revenu pondéré au niveau du ménage collecté dans l'EBM par rapport aux données collectées dans EU-SILC	51
Tableau 27: Répartition du revenu selon les classes.....	51
Tableau 28: Sommes des poids calibrés pour l'échantillon de répondants calibré d'EFT 2018 vague 1	54
Tableau 29: Répartition en pourcentage et en chiffres absolus des classes de revenus du ménage par région - variable INK_5_G	55
Tableau 30: Nombre de ménages âgés selon la région et la taille du MN (RN du 1/07/2018)	55
Tableau 31: Echantillon de répondants EBM 2018 de ménages âgés: nombre de répondants et sommes des poids calibrés selon la province, la taille du ménage et le trimestre	59
Tableau 32: Estimation du nombre de propriétaires de logements (EBM 2018)	60
Tableau 33: Estimations concernant les dépenses pour les 'Articles d'habillement et chaussures' (ECOICOP 03) - EBM 2018	64
Tableau 34: Extraction de la ligne pour 'Articles d'habillement et produits alimentaires' dans le tableau publié HBS_Tab01_BE_2018	65
Tableau 35: Estimations concernant le nombre de voitures que possède chaque ménage - EBM 2018	65
Tableau 36: Estimations concernant le statut de propriété de l'habitation - EBM 2018	67
Tableau 37: Estimations concernant les dépenses effectuées ou non pour des 'Articles d'habillement et chaussures' (ECOICOP 03) - EBM 2018.....	67
Tableau 38: Aperçu de l'estimation de la variance pour les catégories ECOICOP jusqu'au 2e niveau pour la Belgique et les régions	74
Tableau 39: Répartition selon la taille des ménages participants.....	76
Tableau 40: Répartition selon le nombre d'actifs par ménage des ménages participants.....	76
Tableau 41: Répartition selon le sexe des membres des ménages participants.....	77
Tableau 42: Répartition selon l'âge des membres des ménages participants	77
Tableau 43: Différence de montant après l'application de la méthode RFM pour ECOICOP 1 et 2 (non pondéré).....	78
Tableau 44: Nombre de données manquantes par catégorie pour les dépenses énergétiques.....	79
Tableau 45: Nombre de données imputées des carnets pour les dépenses énergétiques	80
Tableau 46: Aperçu des factures groupées pour les dépenses énergétiques	80
Tableau 47: Nombre de données imputées sur la base des modèles de régression pour les dépenses énergétiques.....	81
Tableau 48: Validation du loyer réel et imputé par rapport à l'enquête EU-SILC (pondéré)	82
Tableau 49: Données au niveau du ménage et au niveau individuel	86

Tableau 50: Tailles des UPE dans la population, nombres réalisés de sélections d'UPE et 89

Liste des figures

Figure 1: Flowchart des étapes du traitement des données et du service impliqué 11

Figure 2: Légère hausse des taux de réponse entre les périodes d'EBM 2018 31

Liste des abréviations

Tableau 1: Liste des abréviations

ANT	Anvers
BRU	Région de Bruxelles-Capitale
BWA	Brabant wallon
CAPI	Computer-Assisted Personal Interviewing
CATI	Computer-Assisted Telephone Interviewing
CAWI	Computer-Assisted Web Interviewing
CEC	Collecte de données Enquêtes Citoyens
DEU	Communauté germanophone
EBM	Enquête sur le budget des ménages
ECOICOP	European Classification of Individual Consumption according to Purpose
EFT	Enquête sur les forces de travail
EU-SILC	European Union Statistics on Income and Living Conditions
GR	Groupe(s) de rotation
HAI	Hainaut
HBS	Household Budget Survey
LIE	Liège
LIM	Limbourg
LUX	Luxembourg
MN	Ménage(s)
NAM	Namur
OVL	Flandre orientale
PS	Post Stratification
PPSI	Paper-and-pencil Self-administered Interviewing
RN	Registre national
SM	Service méthodologie
SSS	Service Statistiques sociales
UCM	Unité de consommation modifiée
UPE	Unité(s) primaire(s) d'échantillonnage
USE	Unité(s) secondaire(s) d'échantillonnage
UTE	Unité(s) tertiaire(s) d'échantillonnage
VLA	Région flamande
VLB	Brabant flamand
WAL	Région wallonne
WVL	Flandre occidentale

1. Introduction générale

1.1. Qu'est-ce que l'Enquête sur le budget des ménages?

Statbel réalise l'Enquête sur le budget des ménages (EBM, en anglais HBS: Household Budget Survey) depuis 1853. L'objectif de cette enquête est d'analyser les dépenses des ménages privés et de fournir à ce sujet des données descriptives en lien avec certaines caractéristiques des ménages. Comme cette enquête collecte une grande quantité d'informations uniques, ses résultats sont souvent utilisés afin de définir, suivre et évaluer le comportement d'achat des ménages.

1.2. Base juridique

Il n'y a pas de règlement-cadre européen pour UE-EBM. L'enquête n'est donc pas obligatoire pour les États membres de l'UE. UE-EBM est réalisée sur la base d'un 'gentlemen's agreement'. Cet accord a été obtenu lors de la réunion de 1989 du Comité du programme statistique.

Le *gentlemen's agreement* contient des directives concernant la production d'EBM qui sont conseillées à tous les États membres de l'UE afin que les résultats soient le plus possible comparables au niveau international. Il s'agit par exemple de l'échantillonnage, des définitions des concepts et des variables, des classifications, etc.

L'objectif de l'enquête sur le budget des ménages est de dégager un cadre global pour la production de données statistiques « communautaires » relatives à la consommation des ménages (EU-HBS) sur la base de données transversales (montant, composition,...) sur les postes de leur budget.

Un [arrêté royal de janvier 1996](#) fixe le montant des indemnités aux ménages et un autre [arrêté royal du 17 mai 2006](#) le montant des indemnités accordées aux enquêteurs.

1.3. Quand l'EBM a-t-elle lieu?

L'EBM a été réformée en 2012 pour devenir une enquête bisannuelle afin de produire des chiffres de meilleure qualité, plus adaptés aux demandes de nos principaux utilisateurs et ce, dans des délais plus courts après la période de collecte de données.

L'EBM 2014 a été organisée de la même manière que l'EBM 2012, avec quelques améliorations au niveau du questionnaire ménage (interviews réalisées en CAPI). L'EBM 2016 a été organisée de la même manière qu'en 2014.

Afin d'alléger la charge d'enregistrement des ménages et d'être plus en phase avec les autres pays européens, la période d'enregistrement a été modifiée à partir d'EBM 2018. Les ménages ne doivent plus enregistrer leurs dépenses pendant un mois mais pendant un demi-mois, du 1^{er} au 15^e jour ou du 16^e jour à la fin du mois.

1.4. [Nombre de ménages participants](#)

Depuis 2012, la taille visée de l'échantillon est de 5.000 ménages. La personne de référence du ménage est la personne ayant le revenu net le plus élevé. L'échantillon ne comprend pas de ménages collectifs (couvents, maisons de repos, prisons, etc.). Le Tableau 2 donne un aperçu du nombre de ménages par région pour la période à partir de 2012.

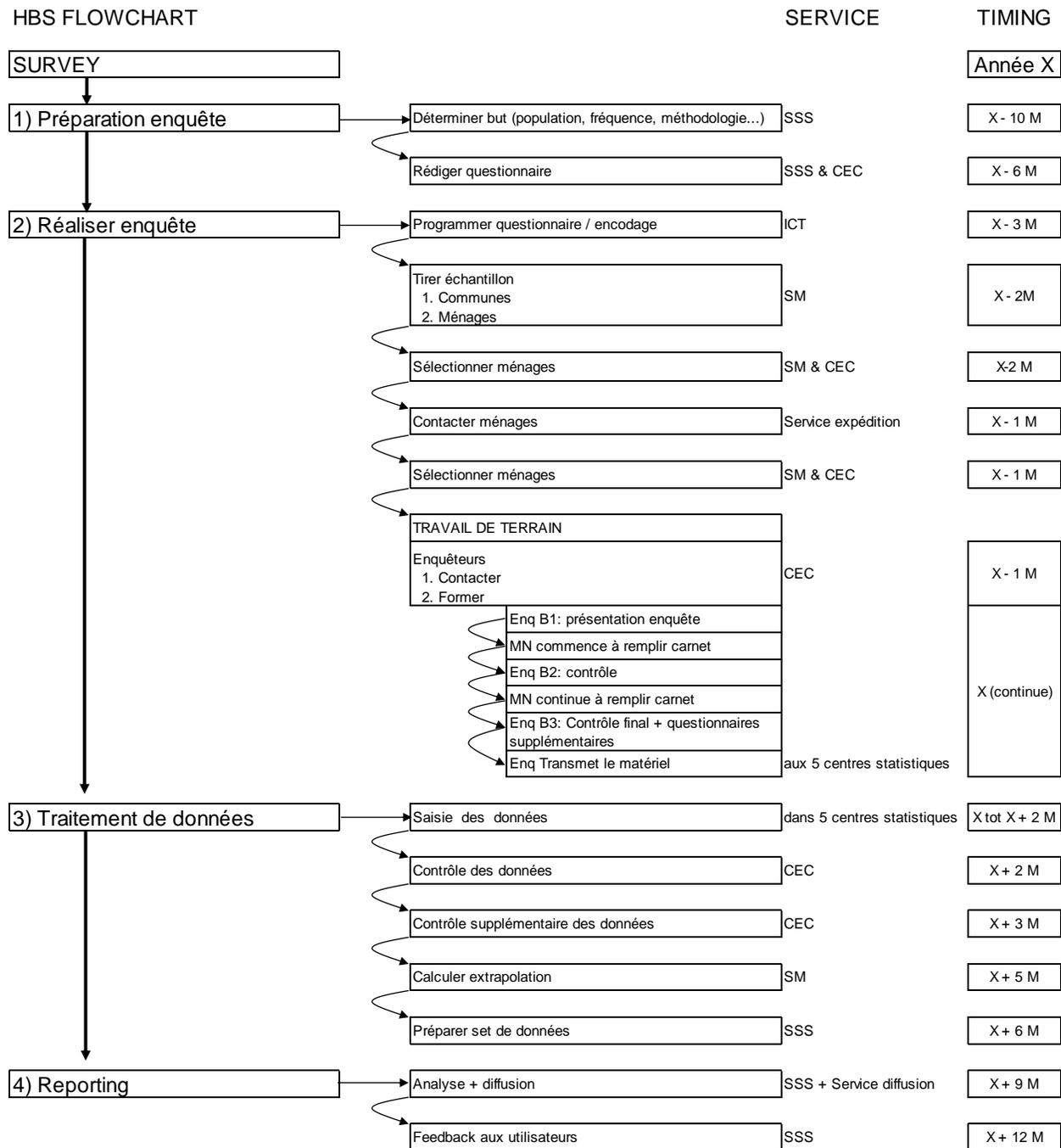
Tableau 2: Répartition des échantillons par région

Nombre de ménages	2012	2014	2016	2018
Région de Bruxelles-Capitale (BRU)	723	633	531	713
Région flamande (VLA)	3.254	2.893	2.175	3.103
Région wallonne (WAL)	2.604	2.609	1.784	2.320
National	6.581	6.135	4.490	6.136

Les FAQ fournissent des informations générales complémentaires sur EBM et son histoireⁱ.

La suite du document donne une description aussi chronologique que possible du tirage de l'échantillon, de la réponse initiale et définitive, de la collecte et du traitement des données pour le calibrage, du calibrage, du traitement des données pour publication et diffusion et de la manière dont se déroulent la publication et la diffusion. La Figure 1 montre quel est le service impliqué lors de chaque étape du traitement.

Figure 1: Flowchart des étapes du traitement des données et du service impliqué



2. Les sous-échantillons pour EBM 2018

2.1. Introduction

L'échantillon de ménages pour EBM 2018 est composé à partir de 3 sources:

- Source 1: échantillons EFT, comparable à la situation pour EBM 2016;
- Source 2: l'échantillon de ménages répondants d'EBM 2016;
- Source 3: un échantillon complémentaire, tiré directement du Registre national (RN), avec
 - Source 3.1: ménages de type EFT
 - Source 3.2: ménages « âgés ».

La source 1 sont les répondants qui ont participé à la dernière vague de l'EFT de 2016 T3 à 2017 T1 et ont indiqué qu'ils voulaient participer à HBS. L'EFT est également toujours basée sur un échantillon de ménages du RN. La source 1 est examinée en détail au par. 2.2.

La source 2 sont les HBS 2016-(répondants) qui ont déjà participé à l'HBS en 2016, et ont précédemment participé à l'enquête LFS (entre 2015 T3 et 2016 T2), qui à leur tour ont également été tirées du RN. Ce point est abordé au par. 2.3.

La source 3 est un échantillon supplémentaire et est examinée au par. 2.4.

De cette manière, tous les ménages d'EBM 2018 proviennent du RN, via des échantillons aléatoires, mais le parcours que ces ménages ont suivi pour arriver dans l'échantillon d'EBM 2018 diffère fortement d'une source à l'autre. Tout cela signifie que la composition de l'échantillon EBM 2018 est complexe et applique un plan de sondage compliqué, ce qui complexifie le traitement des données (calibrage, estimation, estimation de la variance).

Le passage de la tenue d'un carnet pour EBM pendant un mois complet à la tenue d'un carnet pendant un demi-mois complique davantage encore la constitution de l'échantillon EBM 2018¹; voir [Statbel/Delclite \(2020\)](#). Les ménages qui sont recrutés pour EBM 2018 doivent dès lors être attribués à l'un des 24 *échantillons bimensuels*: sept groupes de rotation EFT (source 1), douze échantillons mensuels issus d'EBM 2016 (source 2) et six échantillons complémentaires bimestriels (tirés directement du RN – source 3). Ces groupes doivent donc être répartis sur 24 périodes d'un demi-mois. Nous désignerons ces périodes par P01, P02, ... P24.

Le Tableau 3 présente un bref aperçu des sources d'EBM 2018 et de la réalisation des 24 sous-échantillons EBM bimensuels.

¹ La notion de 'demi-mois' ne doit pas être prise au sens strict: l'année 2018 est scindée en 24 périodes sur la base des mois calendrier. Chaque mois, une première période commence le 1^{er} jour et se termine le 15^e jour du mois, et une deuxième période commence le 16^e jour et se termine le dernier jour du mois.

Tableau 3: Aperçu des sources pour les échantillons bimensuels P01 à P24 d'EBM 2018

Sources				EBM 2018
EFT transversale	EFT panel (*)	EBM 2016	RN	
Source 1: Répondants EFT				
	2016 T3 = union de GR1 [3-] GR2 [4-] GR3 [1(2)2]	}		P01-P04 P05-P08
	2016 T4 = union de GR4 [1(1)2] GR5 [1(2)2] GR6 [2(2)2]			P09-P12 P13-P16 P17-P20
	2017 T1 = GR7 [2(2)2]			P21-P24
Source 2: Répondants EBM 2016 (tirés eux-mêmes des répondants EFT)				
2015 T3	s1-s7		T1	T1 ~ P01-P06
2015 T4	s1-s7		T2	T2 ~ P07-P12
2015 T3 2015 T4 2016 T1	s8-s13 s8-s14 s1-s7		T3	T3 ~ P13-P18
2016 T1 2016 T2	s8-s13 s1-s13		T4	T4 ~ P19-P24
Source 3: Registre national (RN)				
			RN 1	P01-P04
			RN 2	P05-P08
			RN 3	P09-P12
			RN 4	P13-P16
			RN 5	P17-P20
			RN 6	P21-P24
(*) Entre crochets, on mentionne pour chaque GR EFT le scénario qui indique comment sont prévues les interrogations répétées de l'EFT: 'x-' signifie que les ménages du GR sont interrogés pendant x trimestres successifs; 'x(y)z' signifie que les ménages du GR sont interrogés pendant x trimestres, ne le sont plus pendant y trimestres et, enfin, le sont à nouveau pendant z trimestres. A partir du GR7, le scénario est toujours 2(2)2 (et quatre interrogations sont donc prévues).				

Le recrutement à partir des échantillons EFT (source 1) est comparable à l'approche de EBM 2016 (voir [Statbel \(2017a\)](#) ou [Statbel \(2017b\)](#)) mais l'introduction d'un panel pour l'EFT - depuis 2016 T3 en ce qui concerne l'échantillonnage - a eu un impact sur la méthode de travail pour obtenir les échantillons EBM bimensuels. Nous expliquons ça plus en détails au par. 2.2.

Remarquez également que le Tableau 3 indique que l'échantillon EFT pour 2016 T3 se compose à vrai dire de 3 GR (ou sous-échantillons) qui devaient être tirés dans le cadre du lancement du panel EFT. Le regroupement de ces trois GR a toutefois été conçu comme un échantillon trimestriel transversal pour l'estimation des résultats de l'EFT pour 2016 T3. La situation est analogue pour 2016 T4. On indique par ailleurs que, en ce qui concerne EBM 2018, les GR3 et GR 4 sont fusionnés. Il convient de remarquer que ces GR contiennent chacun environ la moitié du nombre de ménages des autres GR.

Le recrutement à partir de deux autres sources est une nouveauté par rapport à EBM 2016. Il complique le traitement des données (comme nous le verrons plus tard) mais s'est avéré nécessaire pour les raisons suivantes. En raison du passage d'une enquête trimestrielle transversale à un panel rotatif trimestriel pour l'EFT depuis 2017 (voir [Termote en Depickere \(2018a\)](#) ou [Termote et Depickere \(2018b\)](#)), les échantillons EFT (qui sont appelés *groupes de rotation* (GR) dans le cadre du panel) sont devenus plus petits. Il a dès lors été décidé d'effectuer le recrutement également à partir d'anciens échantillons EBM (source 2). Il est toutefois apparu que les échantillons EFT et EBM ensemble ne fournissaient toujours pas assez de ménages EBM participants. Un tirage supplémentaire à partir du RN a donc finalement été réalisé (source 3). Ce dernier a aussi permis d'accorder une attention particulière à un certain degré de (sur)représentation des personnes âgées dans l'échantillon EBM 2018.

Le recrutement à partir d'EBM 2016 (source 2) est expliqué au par. 2.3. Le tirage des échantillons complémentaires du RN est expliqué au par. 2.4.

2.2. Recrutement à partir des échantillons EFT – Source 1

2.2.1. Introduction

Les échantillons ou groupes de rotation (GR) EFT trimestriels sont des *échantillons à deux degrés*. Pour EBM, un recrutement complémentaire de ménages est effectué dans plusieurs GR. Ce recrutement est décrit ci-dessous comme le *troisième degré*. Une première partie (soit la partie de la source 1) des échantillons EBM pour 2018 est donc présentée ici comme des *échantillons à trois degrés*.

- Degré 1: tirage des *unités primaires d'échantillonnage* (UPE) à partir d'un cadre d'échantillonnage dérivé du RN, (voir paragraphe 2.2.2.1) et composé d'UPE qui sont des unités géographiques construites;
- Degré 2: tirage des *unités secondaires d'échantillonnage* (USE) d'un cadre d'échantillonnage dérivé du RN composé d'USE, qui sont les ménages *éligibles*, appartenant aux UPE sélectionnées au degré 1;
- Degré 3: 'tirage' des *unités tertiaires d'échantillonnage* (UTE) de l'échantillon de ménages du degré 2 ; les unités tertiaires sont donc aussi des ménages.

Les échantillons d'EFT résultent donc des tirages au degré 1 et au degré 2. Les échantillons d'EBM sont obtenus au degré 3. Cette méthode de travail a permis d'obtenir 6 échantillons EBM bimestriels de la manière suivante (voir Tableau 3):

- EBM 2018 Jan-Fév de EFT GR1 (w3 au 2017 T1 – w1 au 2016 T3).
- EBM 2018 Mar-Avr de EFT GR2 (w4 au 2017 T2 – w1 au 2016 T3).
- EBM 2018 Mai-Jui de EFT GR3&4 (w3 au 2017 T3 – w1 au 2016 T3&4).
- EBM 2018 Juil-Août de EFT GR5 (w3 au 2017 T4 – w1 au 2016 T4).
- EBM 2018 Sep-Oct de EFT GR6 (w4 au 2018 T2 – w1 au 2016 T4).
- EBM 2018 Nov-Déc de EFT GR7 (w4 au 2018 T2 – w1 au 2017 T3).

Précisons qu'un premier échantillon EBM bimestriel composé de ménages qui ont été interrogés pour l'EBM en janvier ou février 2018 résulte du premier groupe de rotation (GR1) qui a été tiré dans la cadre du panel pour l'EFT. Ce groupe de rotation a été tiré peu de temps avant le trimestre 2016 T3 afin d'être interrogé une première fois pour l'EFT au trimestre 2016 T3 – nous parlons alors de la première vague (ou *wave*) et nous la notons w1 – et une 3^e et dernière fois (w3) en 2017 T1. L'interprétation des descriptions ci-dessus pour les cinq autres échantillons EBM bimestriels est analogue. Toutefois, pour certains GR, le nombre d'interrogations n'est pas de 3 mais de 4 (voir le commentaire sur les scénarios au Tableau 3) et, pour le troisième échantillon EBM bimestriel, les deux GR ont été fusionnés. Cette dernière opération était nécessaire car la taille des GR3 et GR 4 (nombre de ménages) ne représente environ que la moitié des autres GR. Il convient de remarquer que le GR3 et le GR4 ont été lancés durant des trimestres différents (respectivement 2016 T3 et T4) mais que, pour les deux, le nombre d'interrogations est de 3 et la dernière interrogation se situe au 2017 T3. Cela correspond au scénario défini pour ces GR et décrit au Tableau 3. Certains éléments résultent d'une phase de lancement complexe pour le panel EFT. Vous trouverez davantage d'informations à ce sujet dans [Termote en Depickere. \(2018a\)](#) ou [Termote et Depickere \(2018b\)](#).

Remarquez que l'interrogation d'EBM 2018 a eu lieu entre deux et quatre trimestres après la dernière interrogation d'EFT. Pour EBM 2016, la période entre l'(unique) interrogation pour EFT et l'interrogation pour EBM était invariablement de 2 trimestres.

Les ménages EFT auxquels il a été effectivement demandé s'ils souhaitaient participer à EBM 2016 sont les ménages qui ont achevé leurs dernières interrogations EFT prévues (la 3^e ou la 4^e). Cette méthode diffère de la méthode appliquée pour EBM 2016 pour laquelle non seulement les ménages qui avaient répondu à EFT avaient été invités à participer à EBM mais l'ensemble des ménages sélectionnés initialement.

2.2.2. Degré 1 : tirage des unités géographiques (ou UPE).

2.2.2.1. Cadre d'échantillonnage des UPE

Le cadre d'échantillonnage des UPE au degré 1 se compose d'une liste d'unités géographiques qui sont des quartiers ("sections statistiques") ou des regroupements de ces quartiers au sein de parties de communes ("lettres statistiques"). Nous discuterons de cela plus en détail dans la suite de ce paragraphe. Chaque UPE contient un nombre de ménages privés (USE) suffisant pour former au moins un *groupe de ménages* (au degré 2). Nous appelons ce nombre la taille de l'UPE (il détermine la probabilité que l'UPE soit tirée). Chaque groupe de ménages est attribué dans son ensemble à un seul enquêteur et doit, pour des raisons organisationnelles, contenir un certain nombre de ménages, qui représente la *taille du groupe*.

Les UPE pour les GR1 à GR6, dont les trois premiers constituent l'échantillon trimestriel EFT pour 2016 T3 et les trois suivants l'échantillon trimestriel pour 2016 T4, ont été tirées sous la forme de six échantillons indépendants d'un même cadre d'échantillonnage. Le cadre d'échantillonnage (des UPE) est dérivé d'une extraction du RN datant du 02/04/2016. Les UPE du GR7 (soit l'échantillon trimestriel EFT complet pour 2017 T1) ont été tirées du cadre d'échantillonnage d'UPE qui a été dérivé d'une extraction du RN datant du 10/09/2016. Remarquez que ces extractions datent de plusieurs mois avant le début des périodes de référence ou des trimestres EFT (resp. 2016 T3, 2016 T4 en 2017 T1) des enquêtes EFT. C'est nécessaire car les UPE sélectionnées doivent être connues à temps pour recruter les enquêteurs.

Pour la création d'un cadre d'échantillonnage d'UPE, on part d'une liste de ménages tirée du RN qui indique, pour chaque ménage, l'adresse de son domicile. Statbel détermine le *secteur statistique* sur base de cette adresse (qui comprend le code REFNIS à cinq chiffres de la commune). Les secteurs statistiques au sein d'une commune sont identifiés par un code à quatre signes (lettres, chiffres, tiret): p.ex. A101, A10-, etc.², ii, iii, iv Nous disons que les secteurs d'une commune, dont les codes commencent par les 2 mêmes signes, font partie de la même *section (statistique)* (dans une commune); le code de cette section se compose dès lors de ces deux premiers signes; par exemple, la section statistique A1 contient les secteurs statistiques A101, A10-, etc. De cette manière, nous pouvons créer une liste de sections – aussi appelées *quartiers* – et déterminer pour chaque section un indicateur de sa taille, à savoir le nombre de ménages privés. On contrôle ensuite quelles sections sont "suffisamment grandes" et quelles sections ne le sont pas. Les sections «suffisamment grandes» sont celles qui contiennent un nombre de ménages privés suffisant pour construire au moins un groupe de MN pour l'EFT. La taille du groupe est une valeur définie à l'avance; pour les enquêtes à partir de 2016, la valeur est de 23 en Flandre et en Wallonie et de 26 en Région de Bruxelles-Capitale. Les sections de taille insuffisante sont soit regroupées pour obtenir un groupe suffisamment grand de sections, soit regroupées avec la(es) plus petite(s) des sections suffisamment grandes. Le regroupement éventuel des sections s'effectue toujours au sein de ce que l'on appelle les *lettres (statistiques)* ou *parties de commune*.³ Les UPE sont les unités qui ressortent de ce processus. Les UPE "suffisamment grandes" sont finalement celles

² Un secteur (statistique) est une zone géographique identifiée au sein d'une commune par un code qui se compose de 4 signes et dont le premier signe est normalement une lettre (A, B, ...).

³ Une commune fusionnée ou lettre (statistique), qui est identifiée au sein d'une commune par le premier signe, à savoir une lettre (A, B, ...), de tous les secteurs statistiques qui appartiennent à la lettre ou à la commune avant fusion. Voir note de bas de page précédente.

qui contiennent un nombre de ménages privés suffisant pour construire au moins un groupe de ménages. La liste finale d'UPE suffisamment grandes constitue le cadre d'échantillonnage d'UPE au degré 1 du plan de sondage.

Des sections insuffisamment grandes, qui subsistent après la méthode de fusion décrite ci-dessus et qui ne peuvent donc pas être fusionnées avec d'autres sections, ne sont pas retenues dans le cadre d'échantillonnage des UPE.

Il arrive que pour certaines adresses, le secteur statistique ne puisse pas être déterminé. C'est, par exemple, le cas si les noms de rue ont changé, si de nouvelles rues ont été créées ou si les numéros de maison ont été modifiés. Statbel n'est pas toujours informé à temps de ce type de changements. Dans ce cas, le code "9999" – pour "missing" – est attribué au secteur statistique. Les ménages pour lesquels ce type de problèmes se posent ne peuvent pas non plus se voir attribuer une section statistique et une lettre statistique (ou commune fusionnée). Il ne peut dès lors pas être tenu compte de ces ménages dans la construction des UPE. Par conséquent, le nombre de ménages privés dans les secteurs, les sections et les UPE peut être sous-estimé. Dans les cas extrêmes, il arrive même qu'il ne soit pas possible de construire des UPE suffisamment grandes dans une commune. La commune sort alors complètement du cadre d'échantillonnage.⁴

Le cadre d'échantillonnage des UPE pour le GR 7 (2017 T1) a été complété par les caractéristiques suivantes des UPE: 1) le nombre de ménages privés, 2) le taux de chômage, 3) le revenu moyen du ménage. *Par région* et pour chacune de ces trois caractéristiques, les UPE sont ensuite classées selon cette caractéristique (de la plus petite à la plus grande). Les UPE sont ensuite réparties en quintiles (environ 20% des UPE classées en première en position sont classées dans le premier quintile, environ 20% des UPE qui suivent sont classées dans le deuxième quintile, etc.). De cette manière, les UPE reprises dans le cadre d'échantillonnage sont de plus caractérisées par: (1*) le quintile du nombre de ménages privés (par région), (2*) le quintile du taux de chômage (par région) et (3*) le quintile du revenu moyen du ménage (par région). Il s'agit de ces caractéristiques transformées (1*), (2*) et (3*) – qui sont toutefois moins détaillées que les caractéristiques initiales (1), (2) et (3) – qui ont été utilisées dans le plan de sondage pour le tirage des UPE.

2.2.2.2. Le plan de sondage pour le tirage des UPE

L'échantillon des UPE pour un GR EFT donné est un *échantillon systématique stratifié* avec un point de départ aléatoire dans chaque strate et des probabilités de sélection proportionnelles à la taille des UPE.

- [1] La stratification des UPE se base sur la subdivision du royaume selon la classification NUTS 2. La Région de Bruxelles-Capitale et les provinces, hormis celle de Liège, constituent chacune une strate. Deux strates ont été définies dans la province de Liège. Elles correspondent respectivement aux communes francophones et aux neuf communes germanophones. On a donc travaillé avec 12 strates d'UPE.
- [2] Au sein de chaque strate, les UPE sont classées comme suit dans le cadre d'échantillonnage.
 - Pour les trimestres 2016 T3 et 2016 T4, à savoir pour les GR1 à GR6, les UPE sont classées selon la taille, qui comme nous l'avons dit, est mesurée comme le nombre de ménages privés dans l'UPE.

⁴ Le cas échéant, on peut malgré tout récupérer pour la sélection les ménages ayant 9999 comme code de secteur, en traitant le *secteur fictif 9999* et donc la *section fictive 99* comme une UPE (particulière).

- Pour le trimestre 2017 T1, les UPE sont classées en alternance selon les trois caractéristiques des UPE, (1*) quintile de taille, (2*) quintile du taux de chômage et (3*) quintile du revenu moyen du ménage tel que défini au paragraphe 2.2.2.1. Le tri en alternance (en anglais *serpentine sorting*) selon (1*), (2*) et (3*) signifie qu'au sein de chaque strate d'échantillonnage, on trie d'abord selon la caractéristique (1*), ensuite, au sein des classes successives de (1*), selon la caractéristique (2*) mais en alternance à la hausse et à la baisse, et enfin, au sein des combinaisons successives de (1*) et (2*), selon la caractéristique (3*), également en alternance à la hausse et à la baisse. Cette modification du plan de sondage (pour l'EFT) depuis 2017 T1 a été effectuée afin de favoriser la représentativité et ce, pour plusieurs caractéristiques et non plus uniquement pour la caractéristique de la taille des UPE.

[3] Ensuite, le nombre de sélections d'UPE est défini *par trimestre* pour chaque strate. Cette opération est effectuée en fonction du nombre de ménages répondants (c'est-à-dire qui participent effectivement) souhaité, du taux de réponse attendu (dans l'Eft) et de la taille souhaitée des groupes de ménages. Le nombre de sélections d'UPE est le nombre de groupes de ménages à constituer pour la strate, qui sont chacun attribués à un enquêteur. (étant entendu que chaque enquêteur peut éventuellement prendre en charge plusieurs groupes de ménages).

Le Tableau 4 fournit les informations suivantes pour les trimestres EFT 2016 T3, 2016 T4 et 2017 T1 pour le tirage des UPE au degré 1: les strates et les nombres de sélections d'UPE par strate, par trimestre et par GR.

Tableau 4: Degré 1 - tirage des UPE ou GRP, répartition des GRP entre les strates, trimestres et GR - y compris DEU

Strate UPE		Trimestres EFT et GR								
		2016 T3				2016 T4				2017 T1
		GR1	GR2	GR3	Tot.	GR4	GR5	GR6	Tot.	GR7
1	ANT	28	31	19	78	10	28	31	69	31
2	BRU	39	39	26	104	13	39	39	91	39
3	WVL	24	27	17	68	8	24	27	59	27
4	OVL	26	29	17	72	10	26	29	65	29
5	HAI	26	28	17	71	9	26	28	63	28
6	LIE	22	25	14	61	8	22	25	55	25
7	LIM	20	22	14	56	8	20	22	50	22
8	LUX	14	16	10	40	6	14	16	36	16
9	NAM	15	17	10	42	6	15	17	38	17
10	VLB	22	25	15	62	8	22	25	55	25
11	BWA	14	16	10	40	5	14	16	35	16
12	DEU	10	11	0	21	0	10	11	21	11
Total		260	286	169	715	91	260	286	637	286

[4] Enfin, *par GR*, on tire un échantillon systématique stratifié avec un point de départ aléatoire dans chaque strate et des probabilités de sélection proportionnelles aux tailles des UPE. Les caractéristiques des tirages de ce type sont *par GR*:

- Les plus grandes UPE sont certainement tirées, à savoir si l'UPE est plus grande que la longueur du degré dans le tirage systématique (la probabilité de sélection est alors de 1).
- Les grandes UPE peuvent en principe être sélectionnées plusieurs fois mais à cause des nombres relativement faibles d'UPE à tirer (Tableau 4), ce n'est pas le cas en pratique. Les

cardinalités (par GR) des UPE tirées sont donc toujours de 1. Remarquez que le GR 3 et le GR4, qui sont fusionnés pour les échantillons EBM bimestriels, ont bien des UPE communes.

- Les petites UPE peuvent ne pas être tirées.
- Mais si de petites UPE, qui se suivent dans le triage, dépassent ensemble la longueur du degré, la sélection de l'une ou plusieurs d'entre elles est garantie.
- Nous remarquons que:
 - le tri (au sein de chaque strate) selon la taille des UPE (à savoir le nombre de ménages privés) pour les GR1 à GR6 favorise la représentativité de l'échantillon, si nous pouvons supposer que la corrélation entre la taille des UPE et certaines caractéristiques des MN qui résident dans une UPE est suffisante;
 - le tri en alternance (au sein de chaque strate) pour le GR7, comme expliqué au point [2] ci-dessus, favorise la représentativité de l'échantillon selon les trois caractéristiques qui sont utilisées dans le tri en alternance.

2.2.2.3. Répartition des UPE dans le temps

Par trimestre (ou période de référence EFT), les sélections d'UPE ou de groupes EFT (voir Tableau 5) sont répartis uniformément entre les 13 semaines de référence du trimestre. Ainsi, les 715 groupes EFT du trimestre 2016 T3 sont répartis uniformément entre 13 semaines, ce qui donne dès lors 55 groupes EFT pour chacune des 13 semaines. Pour 2016 T4, la répartition uniforme des 637 groupes EFT donne 49 groupes par semaine. Pour 2017 T1, la répartition uniforme des 286 groupes EFT donne 22 groupes par semaine. Le Tableau 6 montre qu'une répartition uniforme entre les semaines est réalisée pour chaque GR.

Tableau 5: Répartition uniforme des groupes EFT par trimestre et par GR entre les semaines - DEU inclus

Année	Trim	GR	Semaine du trimestre													Tot.	
			1	2	3	4	5	6	7	8	9	10	11	12	13		
2016	3	1	20	20	20	20	20	20	20	20	20	20	20	20	20	20	260
		2	22	22	22	22	22	22	22	22	22	22	22	22	22	22	286
		3	13	13	13	13	13	13	13	13	13	13	13	13	13	13	169
		Tot.	55	55	55	55	55	55	55	55	55	55	55	55	55	55	715
	4	4	7	7	7	7	7	7	7	7	7	7	7	7	7	7	91
		5	20	20	20	20	20	20	20	20	20	20	20	20	20	20	260
		Tot.	49	49	49	49	49	49	49	49	49	49	49	49	49	49	637
2017	1	7	22	22	22	22	22	22	22	22	22	22	22	22	22	286	
Total			126	126	126	126	126	126	126	126	126	126	126	126	126	1638	

La répartition uniforme parfaite est aussi possible pour la strate BRU (Région de Bruxelles-Capitale), par trimestre et même par GR parce que les nombres à répartir sont tous des multiples de 13 comme on le voit au Tableau 5. Pour les autres strates, une répartition uniforme parfaite par trimestre et/ou par GR n'est que rarement possible. Dans ces cas, on se rapproche le plus possible d'une répartition uniforme, de telle manière que l'on se rapproche davantage de la répartition uniforme pour les niveaux supérieurs de l'agrégation que pour les niveaux inférieurs.

Le Tableau 6 est analogue au précédent mais les groupes EFT de la Communauté germanophone (DEU) ne sont pas inclus car les ménages de cette communauté n'entrent pas en ligne de compte pour EBM. En raison du nombre relativement faible de groupes EFT dans la DEU, une répartition uniforme entre les semaines n'est pas possible au sein de cette strate d'échantillonnage, ce qui se répercute dans la répartition des groupes des 11 autres strates, et ce par GR (sauf le GR3 et le GR4 car ils ne contiennent pas de groupes de la DEU) comme par trimestre et globalement.

Tableau 6: Répartition quasi uniforme des groupes EFT par trimestre et par GR entre les semaines - hors DEU

Année	Trim	GR	Semaine du trimestre													Tot.
			1	2	3	4	5	6	7	8	9	10	11	12	13	
2016	3	1	19	19	19	19	19	19	19	19	19	19	20	20	20	250
		2	22	22	21	21	21	21	21	21	21	21	21	21	21	275
		3	13	13	13	13	13	13	13	13	13	13	13	13	13	169
		Tot.	54	54	53	54	54	54								
	4	4	7	7	7	7	7	7	7	7	7	7	7	7	7	91
		5	19	19	19	19	19	19	19	19	19	19	20	20	20	250
		6	22	22	21	21	21	21	21	21	21	21	21	21	21	275
		Tot.	48	48	47	48	48	48								
2017	1	7	21	22	21	21	21	21	21	21	21	22	21	21	21	275
Total			123	124	121	122	123	123	123	1585						

Ensuite, chaque sélection d'UPE, et donc chaque groupe EFT de ménages (encore à sélectionner), est définitivement liée à une semaine de référence.

2.2.3. Degré 2 : tirage des ménages EFT (UPE)

2.2.3.1. Le cadre d'échantillonnage des ménages EFT

Les échantillons de ménages sont tirés trimestre après trimestre et un cadre d'échantillonnage de ménages est à nouveau construit chaque trimestre, chaque fois sur la base d'une extraction aussi récente que possible du RN:

- Pour l'EFT 2016 T3 (et EBM 2018 Jan-Juin) : extraction datant du 07/05/2016
- Pour l'EFT 2016 T4 (et EBM 2018 Mai-Oct) : extraction datant du 13/08/2016
- Pour l'EFT 2017 T3 (et EBM 2018 Nov-Déc) : extraction datant du 19/11/2016

Le cadre d'échantillonnage de ménages (USE) pour un trimestre donné, dérivé de l'extraction correspondante du RN dans laquelle on peut trouver tous les ménages belges, se compose de ménages qui répondent aux critères suivants:

- [1] Seuls les ménages qui appartiennent aux UPE sélectionnées au degré 1 sont retenus.

- [2] Nous nous limitons aux ménages privés. Les ménages collectifs comme les maisons de repos, les prisons,... sont donc exclus selon les instructions de l'EFT d'Eurostat..
- [3] Les ménages, qui ont été sélectionnés dans un échantillon EFT pour l'un des 10 trimestres précédents, n'entrent plus en ligne de compte pour une sélection.
- [4] Au moins un membre du ménage a atteint l'âge de 15 ans minimum et et 76 ans maximum. L'âge des membres du ménage est calculé le dimanche de la semaine de référence (c'est-à-dire le dernier jour de la semaine de référence), déterminée selon le paragraphe 2.2.2.3 pour le groupe EFT auquel le ménage appartient.

2.2.3.2. Le plan de sondage pour le tirage des ménages EFT

L'échantillon des ménages de l'EFT (USE) pour un trimestre donné, issu du cadre d'échantillonnage d'USE construit pour ce trimestre, est un *échantillon aléatoire simple stratifié*:

- [1] Les strates sont les *différentes* UPE sélectionnées pour le trimestre au degré 1.
- [2] Le nombre de ménages sélectionnés dans chaque strate ou chaque UPE est le produit de la taille du groupe pour l'UPE (26 pour les UPE de la Région de Bruxelles-Capitale, 23 pour les UPE de Flandre et de Wallonie) et du nombre de fois que l'UPE est sélectionnée (c'est-à-dire la « cardinalité ») pour le trimestre donné.
 - a. Pour le trimestre 2016 T3, on a réuni les UPE de 3 GR (GR1 à GR3). Les cardinalités des UPE par GR s'élèvent en effet à 1 mais certaines UPE (plus grandes) sont tirées dans deux ou trois GR, ce qui fait que, dans le regroupement, les cardinalités peuvent grimper jusque 3.
 - b. Pour le trimestre 2016 T4, on a réuni les UPE de 3 GR (GR4 à GR6). Les cardinalités des UPE par GR s'élèvent en effet à 1 mais certaines UPE (plus grandes) sont tirées dans deux ou trois GR, ce qui fait que, dans le regroupement, les cardinalités peuvent grimper jusque 3.
 - c. Pour le trimestre 2017 T1, des UPE sont tirées pour un seul GR (GR7). Elles ont toutes une cardinalité de 1.

- [3] Les ménages font l'objet d'un tirage simple aléatoire au sein de chaque strate ou de chaque UPE.

Si une UPE est sélectionnée plusieurs fois (pour le trimestre donné), alors les ménages tirés seront répartis de manière aléatoire et uniforme entre les sélections d'UPE, c'est-à-dire les GR (voir point [2] ci-dessus). Un nombre identique de ménages sera de ce fait attribué à chaque sélection d'UPE. On appelle cela un groupe de ménages (EFT) et le nombre de ménages dans le groupe est la taille du groupe.

Le Tableau 7 montre les nombres de ménages sélectionnés selon la strate d'échantillonnage des UPE, le trimestre et le GR. Les nombres de ménages sélectionnés sont les tailles des groupes multipliées par les nombres de sélections d'UPE ou les groupes EFT du Tableau 6.

Tableau 7: Degré 2 - tirage des MN: répartition des MN entre strates UPE, trimestres et GR – y compris la DEU

Strate UPE		Taille du groupe	Trimestres EFT et GR								
			2016 T3				2016 T4				2017 T1
			GR1	GR2	GR3	Tot.	GR4	GR5	GR6	Tot.	GR7
1	ANT	23	644	713	437	1794	230	644	713	1587	713
2	BRU	26	1014	1014	676	2704	338	1014	1014	2366	1014
3	WVL	23	552	621	391	1564	184	552	621	1357	621
4	OVL	23	598	667	391	1656	230	598	667	1495	667
5	HAI	23	598	644	391	1633	207	598	644	1449	644
6	LIE	23	506	575	322	1403	184	506	575	1265	575
7	LIM	23	460	506	322	1288	184	460	506	1150	506
8	LUX	23	322	368	230	920	138	322	368	828	368
9	NAM	23	345	391	230	966	138	345	391	874	391
10	VLB	23	506	575	345	1426	184	506	575	1265	575
11	BWA	23	322	368	230	920	115	322	368	805	368
12	DEU	23	230	253	0	483	0	230	253	483	253
Total		-	6097	6695	3965	16757	2132	6097	6695	14924	6695

2.2.3.3. Répartition des ménages-EFT dans le temps

Au par. 2.2.2.3, nous avons discuté la répartition des groupes EFT entre les semaines de référence. On en déduit facilement la répartition des ménages EFT entre les semaines de référence. Le Tableau 8 suivant montre cette répartition, sans tenir compte des ménages de la DEU, étant donné qu'ils n'entrent pas en ligne de compte pour l'EFT.

Tableau 8: Répartition quasi uniforme des ménages EFT par trimestre et par GR entre les semaines - hors DEU

Année	Trim	GR	Semaine du trimestre													Tot.	
			1	2	3	4	5	6	7	8	9	10	11	12	13		
2016	3	1	446	446	446	446	446	446	446	446	446	446	446	469	469	469	5867
		2	515	515	492	492	492	492	492	492	492	492	492	492	492	492	6442
		3	305	305	305	305	305	305	305	305	305	305	305	305	305	305	3965
		Tot.	1266	1266	1243	1266	1266	1266	16274								
	4	4	164	164	164	164	164	164	164	164	164	164	164	164	164	164	2132
		5	446	446	446	446	446	446	446	446	446	446	446	469	469	469	5867
		Tot.	1125	1125	1102	1125	1125	1125	14441								
2017	1	7	436	459	436	436	436	436	436	436	436	459	436	436	436	5714	
Total			2827	2850	2781	2804	2827	2827	2827	36429							

2.2.4. Degré 3 : « sélection » des ménages EBM (UTE)

2.2.4.1. Le cadre d'échantillonnage et le plan de sondage pour la sélection des ménages EBM

Au par. 2.2.1, nous avons déjà indiqué que les 6 échantillons EBM bimestriels de ménages sont dérivés des échantillons EFT pour les GR1 à GR7. La correspondance est simple: cinq GR (GR1, GR2, GR5, GR6 et GR7) donnent chacun un échantillon EBM 2018 bimestriel (respectivement jan-fév, mar-avr, juil-août, sep-oct, nov-déc.) alors que les GR 3 et GR4 donnent l'échantillon pour mai-juin.

Les ménages qui appartiennent aux échantillons EFT forment à proprement parler le cadre d'échantillonnage de la « sélection » de ménages pour EBM 2018 au degré 3. Il s'agit ici d'une *autosélection* de ménages, dans le sens où seuls les ménages qui ont entièrement terminé la dernière interrogation prévue (la 3^e ou la 4^e) ont été invités à participer à l'EBM. La probabilité qu'un ménage qui a été sélectionné pour EFT (GR 1 à GR 7) appartienne à l'échantillon EBM 2018 correspond donc à sa probabilité de réponse à l'EFT.

2.2.4.2. Répartition des ménages EBM dans le temps

Il nous reste encore seulement à expliquer comment les ménages sont répartis entre chacun des échantillons bimestriels entre chaque période de 4 demi-mois (p.ex. GR1 → jan-fév → P01-P04).

Les ménages de chaque GR sont répartis de manière pratiquement équitable entre les 13 semaines du trimestre durant lequel le GR est lancé. Le Tableau 9 montre comment les ménages sont attribués – de manière systématique – au quatre demi-mois (ou 4 « périodes ») durant les deux mois qui correspondent au GR. L'objectif est de parvenir à une répartition aussi uniforme que possible des ménages sur l'ensemble de l'année 2018.

Tableau 9: Fractions approximatives de la répartition des ménages EBM par semaine entre les demi-mois

Demi-Mois	Semaine du trimestre														
	1	2	3	4	5	6	7	8	9	10	11	12	13		
1	1	1	1	0,25											3,25
2				0,75	1	1	0,5								3,25
3							0,5	1	1	0,75					3,25
4										0,25	1	1	1		3,25
	1	1	1	1	1	1	1	1	1	1	1	1	1	1	13

Répartir équitablement les ménages en 13 semaines sur quatre demi-mois peut être réalisé de manière approximative en attribuant à chaque demi-mois les ménages en 3,25 semaines.⁵ Cela peut se faire de manière systématique en attribuant 100% des ménages des 3 premières semaines et 25% des ménages de la semaine 4 au premier demi-mois. Les 75% restants des ménages de la semaine 4, nous les attribuons au deuxième demi-mois.⁶ Au deuxième demi-mois, nous attribuons ensuite aussi tous les ménages (100%) des semaines 5 et 6 et 50% des ménages de la semaine 7. Nous poursuivons de cette manière – systématique –

⁵ Nous entendons par « ménages d'une semaine (déterminée) » les ménages dont la semaine de référence EFT est cette semaine (déterminée).

⁶ Les pourcentages et fractions de ce paragraphe et au Tableau 9 sont approximatifs car les *nombres* par semaine (p.ex. 19) doivent être répartis en *nombres* sur deux demi-mois (p.ex. 5 et 14, ce qui se rapproche de la répartition 25-75%).

l'attribution des ménages de la semaine 7 (50%) et des semaines 8 à 13 aux troisième et quatrième demi-mois.

Remarquez que tous les ménages d'un même groupe EFT (de 23 ou 26 ménages) sont attribués à la même semaine de référence EFT. Attribuer tous les ménages d'une semaine au même demi-mois équivaut donc à attribuer tous les groupes EFT d'une semaine donnée au même demi-mois. Quand les ménages d'une semaine donnée sont ventilés sur 2 demi-mois, cette opération s'effectue en fait au niveau des groupes si bien que tous les ménages d'un même groupe EFT sont attribués au même demi-mois. Une répartition équitable approximative des ménages EBM sur les 24 demi-mois ou périodes est importante afin de garantir que chaque période d'observation est suffisamment représentée dans les données collectées, afin de limiter ainsi au maximum un éventuel biais en raison de fluctuations saisonnières du comportement d'achat.

Enfin, il convient de souligner que la répartition des ménages au fil du temps discutée ici ne joue aucun rôle dans le calcul des probabilités d'inclusion et des poids de sondage, qui jouent pourtant un rôle essentiel dans le calibrage. La période d'observation ne sera pas non plus reprise comme variable dans le calibrage (afin d'atteindre éventuellement un équilibre entre les périodes dans l'échantillon calibré, via des variables de contraste).

2.2.5. Probabilités de sélection et poids de sondage

L'échantillonnage à chacun des trois degrés destiné à permettre de procéder au tirage (ou recrutement) de ménages pour EBM via les tirages pour EFT est probabiliste. En d'autres termes, à chaque degré, chaque unité statistique du cadre d'échantillonnage d'unités éligibles a une probabilité strictement positive d'être sélectionnée, que l'on appelle la probabilité de sélection ou d'inclusion (en anglais, *selection probability*, *inclusion probability*). L'inverse d'une probabilité de sélection est un poids de sondage (en anglais *sampling weight*, *design weight*). Il est important de pouvoir calculer ces poids de sondage (ou du moins de bien pouvoir les estimer) pour les unités répondantes. Ce calcul est généralement suivi par une correction de la non-réponse et/ou d'une correction par calibrage car elles jouent un rôle important dans le calcul des estimations des caractéristiques des unités.

Dans ce paragraphe, nous abordons très brièvement les probabilités de sélection et les poids de sondage, en nous limitant aux éléments essentiels dans le cadre de l'estimation pour EFT et EBM.

La sélection au degré 1 (par. 2.2.2) d'unités géographiques (UPE) est surtout importante pour l'organisation de l'EFT: l'objectif est de créer des groupes de ménages EFT. Chaque groupe est attribué dans son ensemble à un seul enquêteur et dès lors les ménages au sein de chaque groupe ne doivent pas avoir des domiciles trop éloignés les uns des autres. Comme nous ne voulons toutefois pas calculer d'estimations des caractéristiques d'UPE, les probabilités d'inclusion et les poids de sondage des UPE ne sont pas importants en soi.

La sélection au degré 2 (par. 2.2.3) des ménages EFT (USE) s'effectue au sein de chaque UPE sélectionnée au degré 1 par SRS. Cela entraîne des probabilités d'inclusion *conditionnelles* très simples: chaque ménage éligible au sein d'une UPE h a une probabilité d'inclusion n_h/N_h dans laquelle N_h est le nombre de ménages éligibles et n_h le nombre de ménages sélectionnés dans l'UPE h . Il en résulte les poids de sondage conditionnels N_h/n_h , qui pourraient être utilisés pour calculer des estimations des caractéristiques des ménages au sein de l'UPE h (après correction de la non-réponse et/ou calibrage éventuels). Mais les estimations de ce type au sein des UPE ne sont pas non plus importantes en soi, notamment parce que les échantillons au sein des UPE sont très réduits et les estimations au sein des UPE ont dès lors une très faible précision.

Les estimations sur la base de l'EFT combinent toujours des informations des observations dans un nombre suffisamment grand d'UPE. C'est la raison pour laquelle nous considérons dans ce texte les sélections d'UPE du degré 1 et de ménages EFT au degré 2 comme indissociables et nous déterminons pour chaque ménage éligible une probabilité de sélection non conditionnelle. On peut démontrer – voir l'annexe L'échantillonnage à deux degrés en bref – Approche des poids de sondage pour les USE pour une introduction sommaire et simplifiée au calcul de probabilités de sélection et de poids de sondage pour un échantillonnage à deux degrés – qu'étant donné le plan de sondage combiné au degré 1 et au degré 2, la probabilité de sélection d'un MENAGE est (approximativement) égale à mg/N dans laquelle m est le nombre de groupes de ménages, g la taille du groupe et N le nombre total de ménages éligibles. En raison de la stratification au niveau des UPE, telle que discutée au par.2.2.2.2, les paramètres m , g et N et donc aussi la probabilité de sélection, varient selon la strate d'échantillonnage des UPE (et le GR). Les probabilités de sélection inverses $N/(mg)$ sont les poids de sondage qui doivent être utilisés, généralement après correction de la non-réponse et/ou après calibrage, pour calculer des estimations de caractéristiques des ménages.

Les publications basées sur l'EFT, ne contiennent toutefois pas d'estimations au niveau des ménages. Les estimations de paramètres concernant le marché du travail doivent être produites et publiées au niveau individuel, par exemple, le taux de chômage qui est une proportion du nombre estimé d'individus. Etant donné que parmi les ménages EFT sélectionnés, tous les membres du ménage de l'échantillon sont repris, les probabilités de sélection ont encore la forme mg/N avec comme poids de sondage correspondants $N/(mg)$. Les estimations au niveau individuel ne sont toutefois pas pertinentes pour EBM si bien que nous pouvons entièrement laisser de côté l'aspect des estimations au niveau individuel ainsi que les probabilités de sélection et les poids de sondage, dans le cadre d'EBM.

La sélection des ménages EBM 2018 au degré 3 a été décrite au par. 2.2.4 comme une autosélection de ménages EFT, dont la probabilité de sélection (conditionnelle) est la probabilité de réponse à EFT. Nous entendons par cette dernière probabilité, la probabilité qu'un ménage qui a été sélectionné pour EFT participe à EFT pendant toutes les vagues prévues (3 ou 4 selon le GR).

La probabilité de sélection (non conditionnelle) pour un ménage EBM 2018 est finalement le produit de la probabilité de sélection (approximative) pour EFT (dont la forme est mg/N) et de la probabilité de réponse à EFT. Contrairement à la probabilité de sélection pour EFT, qui découle du plan de sondage à deux degrés d'EFT, la probabilité de réponse à EFT doit être évaluée sur la base des observations.

La méthode GRH - qui est l'abréviation de groupe de réponse homogène (en anglais *respons homogeneity group*; voir [Särndal et al \(1992\)](#) pp.575-582) - est une méthode simple d'estimation des probabilités de réponse à EFT. Cette méthode consiste à diviser l'échantillon en sous-ensembles (groupes) de ménages qui sont supposés avoir tous la même probabilité de réponse. Par sous-ensemble, nous calculons alors une estimation de la probabilité de réponse sous la forme du rapport entre le nombre de ménages répondants et le nombre de ménages sélectionnés.

Les autres méthodes sont basées sur des modèles de régression (logistiques). Une méthode de pondération à deux degrés est appliquée depuis le passage d'EFT d'une enquête transversale à un panel et pour la production d'estimations pour les indicateurs EFT depuis 2017 T1. Cette méthode consiste, au premier degré, à estimer les probabilités de réponse au niveau des ménages et à les utiliser pour corriger les poids de sondage, et, au deuxième degré, à corriger par calibrage les poids de sondage corrigés de la non-réponse selon les distributions de la population. L'estimation des probabilités de réponse au premier degré s'effectue au moyen d'un modèle de *mixed effects logistic regression* (voir [Vanderhoeft et al. 2021](#)). Nous pouvons utiliser ces estimations de probabilités de réponse pour obtenir une estimation des probabilités de sélection à trois degrés des ménages qui ont répondu à EBM 2018.

2.3. [Recrutement via l'échantillon EBM 2016 – Source 2](#)

2.3.1. [Introduction](#)

L'échantillon (de répondants) d'EBM 2016 est une deuxième source pour obtenir les répondants d'EBM 2018: les répondants à EBM 2016 ont été invités à participer à nouveau à EBM 2018.

L'échantillon d'EBM 2016 est issu de quatre échantillons trimestriels d'EFT. Cela a été décrit en détails dans *Enquête sur le budget des ménages 2016 - Note méthodologique* ([Statbel \(2017a\)](#) ou [Statbel \(2017b\)](#)) au par. 3. Au Tableau 10 ci-après, nous présentons de manière schématique la création du sous-échantillon d'EBM 2018 via cette deuxième source. Ce schéma doit permettre de justifier le calcul des poids de sondage (approximatifs) et d'aborder la répartition de la source 2 entre les quatre sous-échantillons trimestriels d'EBM 2018.

Le Tableau 3 montre que les quatre sous-échantillons trimestriels d'EBM 2016 ne sont pas issus « en droite ligne » de quatre échantillons trimestriels d'EFT ; le contexte a été décrit au par. 3.6 de *Enquête sur le budget des ménages 2016 - Note méthodologique* ([Statbel \(2017a\)](#) ou [Statbel \(2017b\)](#)) au par. 3. comme un « recrutement postal complémentaire ». Le Tableau 3 indique également que les quatre sous-échantillons trimestriels d'EBM 2018 de la source 2 ne sont pas non plus issus « en droite ligne » de quatre sous-échantillons d'EBM 2016. Cette situation compliquée est détaillée au Tableau 10 ci-dessous, qui indique également comment peuvent être calculés les poids initiaux (approximatifs) pour le calibrage des répondants d'EBM 2018 issus de la source 2. C'est le sujet du par. suivant 2.3.2.

2.3.2. [Sous-échantillons trimestriels d'EBM 2018 de la source 2](#)

2.3.2.1. [Le sous-échantillon d'EBM 2018 T1 de la source 2](#)

Ce sous-échantillon est le résultat d'un échantillonnage à 4 degrés comportant deux volets (issu du RN) tiré comme suit:

Un *premier échantillonnage à 2 degrés* tiré du RN a donné l'échantillon de ménages d'EFT 2015 T3. Selon l'argumentation de l'annexe L'échantillonnage à deux degrés en bref – Approche des poids de sondage pour les USE, les poids de sondage correspondants peuvent être estimés de manière approximative au moyen de la formule $N_1/n_{rel.1}$ dans laquelle $n_{rel.1}$ est le nombre de ménages effectivement sélectionnés dans l'échantillon d'EFT 2015 T3 et N_1 le nombre total de ménages éligibles dans la population belge (au moment où le tirage des ménages est effectué au degré 2).

Ensuite, au troisième degré, les ménages qui ont été attribués aux 7 premières semaines de référence de l'EFT du trimestre 2015 T3 sont retenus de l'échantillon d'EFT 2015 T1 afin de constituer un sous-échantillon d'EBM 2016 T1. Chacun des ménages de l'échantillon d'EFT 2015 T3 a donc une probabilité de sélection (approximative conditionnelle) de 7/13 de se retrouver dans un sous-échantillon d'EBM 2016 T1. En combinaison avec les degrés 1 et 2, cela donne un poids de sondage (approximatif) $(N_1/n_{rel.1})/(7/13)$ après les 3 premiers degrés.

Tableau 10: Aperçu schématique de la création du sous-échantillon d'EBM 2018 via la source 2

					Commentaires
	RN (1)	RN (2)	RN (3)	RN (4)	Plusieurs extractions du RN utilisées
Degrés 1 & 2	$n_{rel,1}/N_1$	$n_{rel,2}/N_2$	$n_{rel,3}/N_3$	$n_{rel,4}/N_4$	Probabilités de élection n_{rel}/N pour ménages EFT <ul style="list-style-type: none"> • n_{rel} et N varient selon le trimestre et la strate d'UPE • pour N : fusionner les strates LIE et DEU • pour n_{rel} : conserver uniquement la strate LIE
	EFT 2015 T3	EFT 2015 T4	EFT 2016 T1	EFT 2016 T2	Echantillons trimestriels EFT (EFT continue)
Degré 3	$\sim 7/13$ $\sim 6/13$	$\sim 7/14$ $\sim 7/14$	$\sim 7/13$ $\sim 6/13$	$13/13$	Probabilités de sélection conditionnelles exprimées en « semaines »
	EBM 2016 T1	EBM 2016 T2	EBM 2016 T3 (A) (B) (C)	EBM 2016 T4 (A) (B)	} Voir par. 3.6 « Recrutement postal complémentaire » dans Statbel (2017a) ou Statbel (2017b)
Degré 4	p_{11}	p_{12}	p_{21}	p_{22}	Probabilités d'(auto)sélection conditionnelles ~ probabilités de réponse <ul style="list-style-type: none"> • varient éventuellement selon le trimestre • peuvent être « modélisées »
	EBM 2018 T1	EBM 2018 T2	EBM 2018 T3	EBM 2018 T4	Echantillons trimestriels de répondants EBM 2018

Etant donné que seuls les répondants de l'échantillon d'EBM 2016 T1 sont invités à participer à EBM 2018, nous pouvons affirmer qu'au 4^e degré, les ménages de l'échantillon d'EBM 2016 T1 ont une probabilité d'(auto)-sélection (conditionnelle) qui est la probabilité de réponse individuelle p_{11} de ces ménages. En combinaison avec les degrés 1, 2 et 3, cela donne un poids de sondage approximatif $\binom{N_1/n_{rel,1}}{(7/13)}/p_{11}$ après les 4 degrés de la sélection. Il convient de remarquer que la probabilité d'(auto)-sélection p_{11} varie selon le ménage; un modèle de réponse (par exemple, un modèle de régression logistique) pourrait être utilisé pour estimer ces probabilités.

Un *deuxième échantillonnage à 2 degrés* tiré du RN a donné l'échantillon de ménages d'EFT 2015 T4. Ensuite, au troisième degré, les ménages qui ont été attribués aux 7 premières semaines de référence de l'EFT du trimestre 2015 T4 (qui comporte 14 semaines) sont retenus de l'échantillon d'EFT 2015 T4 afin de constituer un échantillon pour EBM 2016 T2 (voir Tableau 10). Enfin, au 4^e degré, les répondants de l'échantillon d'EBM 2016 T2 sont invités à participer à EBM 2018. Ces ménages reçoivent un poids de sondage (approximatif) qui peut être exprimé sous la forme $\binom{N_2/n_{rel,2}}{(7/14)}/p_{12}$.

En bref:

- RN (1) → EFT 2015 T3 → EBM 2016 T1 → EBM 2018 T1
 - poids initial à calibrer $\binom{N_1/n_{rel,1}}{(7/13)}/p_{11}$
- RN (2) → EFT 2015 T4 → EBM 2016 T2 → EBM 2018 T1
 - poids initial à calibrer $\binom{N_2/n_{rel,2}}{(7/14)}/p_{12}$

2.3.2.2. Le sous-échantillon d'EBM 2018 T2 de la source 2

Par analogie avec le sous-échantillon d'EBM 2018 T1 de la source 2, le sous-échantillon d'EBM 2018 T2 de la source 2 a été constitué via deux canaux:

- RN (2) → EFT 2015 T4 → EBM 2016 T2 → EBM 2018 T2
 - poids initial à calibrer $\binom{N_2/n_{rel,2}}{(7/14)}/p_{21}$
- RN (3) → EFT 2016 T1 → EBM 2016 T3 → EBM 2018 T2
 - poids initial à calibrer $\binom{N_3/n_{rel,3}}{(7/13)}/p_{22}$

2.3.2.3. Le sous-échantillon d'EBM 2018 T3 de la source 2

La composition du sous-échantillon d'EBM 2018 T3 de la source 2 est analogue mais plus complexe:

- RN (1) → EFT 2015 T3 → EBM 2016 T3 → EBM 2018 T3
 - poids initial à calibrer $\binom{N_1/n_{rel,1}}{(6/13)}/p_{31}$
- RN (2) → EFT 2015 T4 → EBM 2016 T3 → EBM 2018 T3
 - poids initial à calibrer $\binom{N_2/n_{rel,2}}{(7/14)}/p_{31}$
- RN (3) → EFT 2016 T1 → EBM 2016 T1 → EBM 2018 T3

- poids initial à calibrer $(N_3/n_{rel,3})/(7/13)/p_{31}$
- RN (4) → EFT 2016 T2 → EBM 2016 T4 → EBM 2018 T3
 - poids initial à calibrer $(N_4/n_{rel,4})/(13/13)/p_{32}$

2.3.2.4. Le sous-échantillon d'EBM 2018 T4 de la source 2

La composition du sous-échantillon d'EBM 2018 est:

- RN (3) → EFT 2016 T1 → EBM 2016 T4 → EBM 2018 T4
 - poids initial à calibrer $(N_3/n_{rel,3})/(6/13)/p_{31}$
- RN (4) → EFT 2016 T2 → EBM 2016 T4 → EBM 2018 T4
 - poids initial à calibrer $(N_4/n_{rel,4})/(13/13)/p_{41}$

2.3.3. Répartition des ménages d'EBM 2018 entre les 24 demi-mois

La lettre d'invitation aux ménages mentionne pendant quelle période ou « demi-mois » il doivent remplir les carnets d'EBM 2018. Afin de déterminer ces demi-mois, les ménages répondants d'EBM 2016 ont été répartis équitablement entre les 24 demi-mois (périodes) en 2018 de la manière suivante:

- Les 4.475⁷ ménages répondants d'EBM 2016 sont classés selon le mois durant lequel les carnets ont été remplis en 2016 et selon le numéro de groupe durant le mois.
- Etant donné que $4.475 / 24 = 186,458$, 186 ou 187 ménages ont dès lors pu être attribués à chaque demi-mois.
- Toutefois, pour des raisons pratiques, les ménages qui se trouvaient dans le même groupe en 2016 ont été attribués au même demi-mois.

Cela a entraîné une redistribution des ménages répondants d'EBM 2016 entre les demi-mois en 2018. Cette redistribution est résumée au Tableau 11.

⁷ La différence entre 4475 et 4490 dans le panel de l'ancienne EBM: il y avait 4490 ménages dans l'échantillon 2016. Finalement, seuls 4475 ménages ont reçu une lettre car nous savions qu'il y avait un problème pour 15 ménages: 7 étaient décédés et 8 avaient déménagé à une adresse inconnue.

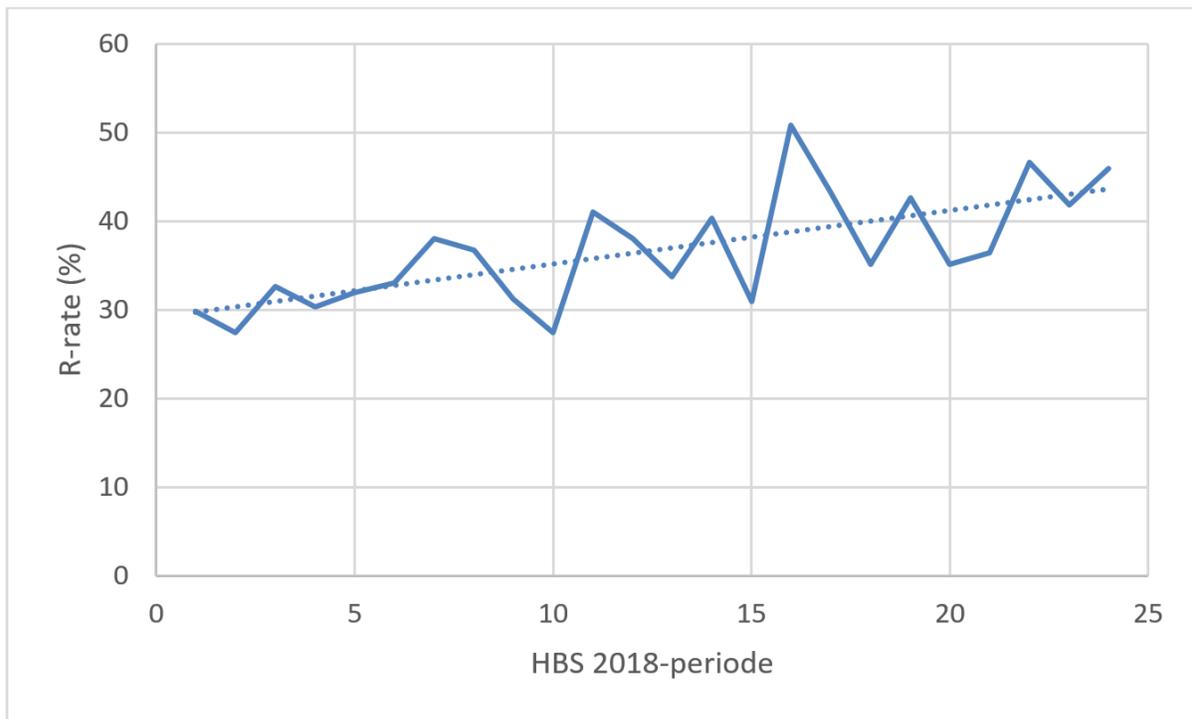
Tableau 11: Ménages répondants à EBM 2016 invités à participer à EBM 2018 - attribution de la période (« demi-mois ») en 2018

Période EBM 2018	Mois EBM 2016												Tot.	Rép.	R-rate (%)	
	1	2	3	4	5	6	7	8	9	10	11	12				
01	181	181	54	29,8
02	108	71	179	49	27,4
03	.	184	184	60	32,6
04	.	70	118	188	57	30,3
05	.	.	161	27	188	60	31,9
06	.	.	.	188	188	62	33,0
07	.	.	.	71	116	187	71	38,0
08	122	66	188	69	36,7
09	186	186	58	31,2
10	7	179	186	51	27,4
11	188	188	77	41,0
12	32	152	184	70	38,0
13	187	187	63	33,7
14	153	33	186	75	40,3
15	188	188	58	30,9
16	185	185	94	50,8
17	68	117	.	.	.	185	80	43,2
18	185	.	.	.	185	65	35,1
19	172	16	.	.	188	80	42,6
20	185	.	.	185	65	35,1
21	187	.	.	187	68	36,4
22	137	52	189	88	46,6
23	189	189	79	41,8
24	194	194	89	45,9
Total	289	325	279	286	238	259	399	492	474	474	525	435	4475	1642	36,7	
Rép.	83	96	92	95	93	83	140	185	197	183	206	189	1642			
R-rate (%)	28,7	29,5	33,0	33,2	39,1	32,0	35,1	37,6	41,6	38,6	39,2	43,4	36,7			

Nous constatons que les nombres de ménages invités pour EBM 2018 par période ou demi-mois (colonne ayant pour titre "Tot.") sont relativement constants.

Le Tableau 11 montre que les nombres finaux (« Rép. ») et les pourcentages (« R-rate (%) ») de ménages répondants à EBM 2018 – de la source 2 – selon le mois d'EBM 2016 et la période d'EBM 2018. Il convient de remarquer la répartition relativement uniforme des ménages invités entre les périodes d'EBM 2018, ainsi que la légère croissance des taux de réponse correspondants:

Figure 2: Légère hausse des taux de réponse entre les périodes d'EBM 2018



Le Tableau 11 donne une explication pour les flèches du degré 4 au Tableau 10. Nous constatons que, par exemple, tous les ménages des trois premiers mois d'EBM 2016 (T1) sont attribués aux périodes d'EBM 2018 01-05 (flèche verticale la plus à gauche) et qu'une partie des ménages du quatrième mois d'EBM 2016 (T2) est attribuée aux périodes d'EBM 2018 05-06 (flèche oblique la plus à gauche); les périodes d'EBM 2018 01-06 constituent le trimestre 2018 T1. Idem pour les autres flèches.

2.4. Tirages complémentaires du RN – Source 3

2.4.1. Introduction

Tout comme le recrutement des groupes de rotation de l'EFT (source 1 – voir Tableau 3, colonnes « panel EFT » et « EBM 2018 », ainsi que le par. 2.2.1) a donné en premier lieu 6 échantillons bimestriels d'EBM 2018, qui ont ensuite été répartis entre 4 échantillons bimensuels d'EBM 2018, un échantillon complémentaire a été tiré directement du RN tous les deux mois. Il était ensuite réparti entre 4 échantillons bimensuels complémentaires d'EBM 2018 Tableau 3 (voir , colonnes « RN » et « EBM 2018 »).

Les 6 échantillons bimestriels complémentaires (voir colonne [1] au Tableau 12 ont été ajustés au recrutement issu d'EFT. Ceci est expliqué au Tableau 12.

Tableau 12: Les 6 échantillons bimestriels complémentaires d'EBM 2018: informations de base

Sous-échantillon complémentaire EBM		GR EFT sous-jacents	Nombre d'UPE (hors DEU)	Nombre d'UPE vides	Nombre de MN éligibles
[1]		[2]	[3]	[4]	[5]
Jan-Fév	(P01-P04)	GR1	250	0	588.557
Mar-Avr	(P05-P08)	GR2	275	0	637.719
Mai-Juin	(P09-P12)	GR3 & GR4	243	1	505.916
Jui-Août	(P13-P16)	GR5	250	1	577.660
Sep-Oct	(P17-P20)	GR6	275	0	617.442
Nov-Déc	(P21-P24)	GR7	275	1	613.205

- La colonne [2] indique quel GR EFT a servi de point de départ pour chaque échantillon bimestriel EBM. Il convient de remarquer que pour l'échantillon EBM Mai-Juin, deux GR ont été fusionnés car ils étaient environ deux fois plus petits que les autres GR.
- La colonne [3] indique combien d'UPE différentes, à l'exclusion des UPE des communes germanophones, sont tirées pour former les GR pour l'EFT. Exemple (a) pour jan-fév: pour le GR1 260 UPE, dont 10 dans les communes germanophones, ont été tirées au premier degré du tirage pour l'EFT. Exemple (b) pour mai-juin: les deux GR plus petits, GR3 et GR4, contiennent respectivement 169 et 91 UPE, dont 17 UPE communes. Le nombre d'UPE différentes est dès lors plus faible ici, à savoir $243 = (169 + 91) - 17 = 260 - 17$; Les GR3 et GR4 ne contiennent pas d'UPE dans les communes germanophones.
- La colonne [4] indique combien d'UPE, mentionnées à la colonne [3], ne contiennent pas les ménages privés éligibles ou sont « vides ». (Le concept d'« éligible » dans ce contexte de tirages complémentaires est précisé au par. 2.4.3). Ce phénomène est la conséquence de l'utilisation de deux versions différentes du RN: une version pour la construction et le tirage des UPE pour l'EFT et une version ultérieure pour le tirage des ménages dans l'échantillon complémentaires pour EBM 2018. Une UPE est « vide » si on ne peut pas déterminer le secteur statistique (ce qui est effectué par Statbel) des adresses de cette UPE, ce qui est par exemple le cas lorsque les noms de rue, les numéros de maison,... ont changé; voir par. 2.2.2.1. Ainsi, par exemple, le tirage des UPE pour le GR1 est basé sur une extraction du RN du 02/04/2016, alors que le tirage des ménages pour l'échantillon complémentaire d'EBM 2018 jan-fév utilise une extraction du 09/09/2017.

- La colonne [5] donne le nombre de ménages éligibles dans les UPE dont le nombre est la différence entre les nombres des colonnes [3] en [4]. Nous spécifierons plus tard (voir par. 2.4.3) quels ménages sont « éligibles ». Exemple pour mai-juin: dans $243 - 1 = 242$ UPE non germanophones et non vides, 505.916 ménages sont éligibles. Pour chaque tirage, une nouvelle extraction du RN mise à jour a été utilisée: le Tableau 3 l'indique pour de RN 1 à RN 6.

Les tirages bimestriels complémentaires de ménages du RN, qui ont été discutés dans ce paragraphe, peuvent tous être considérés comme un échantillonnage à deux degrés avec la sélection des UPE au degré 1, suivie par la sélection au degré 2 des ménages parmi les UPE sélectionnées au degré 1.

2.4.2. Degré 1 : tirage des unités géographiques (ou UPE)

Le tirage des UPE n'est pas un nouveau tirage: les UPE avaient déjà été tirées pour l'EFT (GR1 à GR7) comme expliqué au par. 2.2.2.

2.4.3. Degré 2 : tirage des ménages EBM (UPE)

L'objectif des tirages complémentaires directs du RN est quadruple:

- Au total, nous voulions réaliser environ 2.000 ménages répondants supplémentaires.
- Nous voulions répartir ce nombre équitablement entre les 6 échantillons bimestriels.
- Nous souhaitons qu'environ 80% de ces ménages soient du type EFT et 20% du type non EFT. On entend par ménage de type EFT, un ménage dont au moins un membre appartient à la classe d'âge des 15-76, à savoir un ménage qui serait aussi éligible pour le tirage d'EFT.
- Nous voulions atteindre cette proportion 80:20 pour chaque échantillon bimestriel.

De plus, nous voulions répartir équitablement les 6 échantillons entre les UPE concernées. Pour atteindre ces objectifs, un tirage aléatoire stratifié a été réalisé. Les strates d'échantillonnage sont les cellules au croisement de deux variables:

- Une variable qui détermine pour chaque ménage à quelle UPE il appartient.
- Une variable qui détermine pour chaque ménage s'il est de type EFT.

En ce qui concerne la deuxième variable de stratification, nous remarquons que si un ménage n'est pas de type EFT, à savoir qu'il ne contient aucun membre dans la tranche d'âge 15-76 ans et donc contient uniquement des membres de 77 ans au moins (une personne « âgée »), uniquement des membres de 14 ans maximum (un « enfant »), ou uniquement au moins un enfant et au moins une personne âgée. Nous avons constaté que (dans les UPE sélectionnées) les ménages qui n'étaient pas du type EFT contenaient généralement uniquement des personnes de 77 ans ou plus (nous les appelons ci-après les « ménages âgés »). *C'est la raison pour laquelle il a été décidé de limiter les ménages éligibles de chaque échantillon bimestriel aux ménages de type EFT, d'une part, et aux ménages âgés, d'autre part.* La deuxième variable de stratification mentionnée ci-dessus distingue donc les ménages de type EFT et les ménages âgés (et élimine donc les autres MN du cadre d'échantillonnage).

Nous remarquons que, via les sources 1 (échantillons EFT; voir par. 2.2) et 2 (échantillon EBM 2016; voir par. 2.3), seuls des ménages de type EFT pouvaient se retrouver dans l'échantillon EBM 2018. La source 3 (RN) et le plan de sondage (déjà en partie) décrit ci-dessus ont aussi permis la sélection de ménages âgés. Il

s'agissait d'un souhait des utilisateurs d'EBM en raison de leurs préoccupations quant à la représentativité de l'échantillon d'EBM.

Remarquez toutefois que les ménages de type EFT contiennent au moins une personne dans la tranche d'âge 15-76 ans au moment où ils sont (peuvent être) tirés pour l'EFT, ce qui tire vers le bas le nombre de personnes âgées dans les échantillons EFT. Lors du recrutement ultérieur pour EBM, et plus encore lors l'interrogation ultérieure (beaucoup plus tard), la composition (notamment ne ce qui concerne l'âge) a plus ou moins changé. Le 'vieillissement' des échantillons des sources 1 et 2 peut donc résulter en une augmentation du nombre de personnes âgées dans les échantillons EBM dérivés de ces sources.

Etant donné le nombre de ménages souhaité (2.000) et la stratification décrite ci-dessus, la taille initiale de l'échantillon et l'allocation, c'est-à-dire la répartition de l'échantillon entre les strates, ont été calculées par échantillon bimestriel. À cette fin, les probabilités de réponses attendues (au niveau des ménages) devaient encore être déterminées. Le Tableau 13 montre, aux colonnes [2], [3a] et [3b], les paramètres qui varient selon les échantillons EBM bimestriels et le type de ménage:

- La colonne [2] montre le nombre de catégories de la première variable de stratification: il est dérivé des colonnes [3] et [4] du Tableau 12. Pour tous les échantillons bimestriels, sauf le troisième (mai-juin), ce nombre correspond à la valeur de la colonne [3] du Tableau 12 moins la valeur de la colonne [4] du Tableau 12, soit le nombre d'UPE différentes non vides, ce qui correspond aussi au nombre de groupes EFT (GRP) de ménages dans les UPE non vides (en non germanophones). Pour mai-juin, ce calcul (colonne [3] moins colonne[4]) donne la valeur 242. Pour les 17 UPE présentes tant dans le GR3 EFT que dans le GR4 EFT, nous ne voulions pas tirer 1 mais 2 groupes de ménages supplémentaires, comme c'est également le cas dans les tirages pour EFT. C'est la raison pour laquelle, nous avons compté sur $260 = 169 + 91$ GRP EFT dans les calculs des tailles des échantillons.
- La ventilation des colonnes [3a] et [3b], ainsi que des colonnes [4a] et [4b] ainsi que [5a] et [5b] correspond aux deux catégories pour la deuxième variable de stratification (ménages de type EFT versus ménages âgés). La plupart des probabilités de réponse attendues aux colonnes [3a] et [3b] s'élèvent à 5%, ce qui est basé sur les faibles probabilités de réponse observées dans le passé pour EBM, lorsque l'enquête n'était pas basée sur l'EFT et le recrutement se réalisait par courrier. Pour les ménages âgés et les trois derniers échantillons bimestriels, la probabilité de réponse a été réduite à 3,5%, parce qu'il a été constaté dans les échantillons bimestriels précédents qu'une prévision de réponse de 5% pour les ménages âgés était trop élevée; l'échantillon de jan-fév a montré que 3,5% était une valeur plus réaliste.

Tableau 13: Les 6 échantillons bimestriels complémentaires d'EBM 2018: paramètres d'échantillon et distribution

Échantillon complémentaire EBM	Nombre GRP	Probabilité de réponse (Niveau du MN)		Distribution par GRP		Distribution totale (réalisée)	
		Type EFT	Âgé	Type EFT	Âgé	Type EFT	Âgé
[1]	[2]	[3a]	[3b]	[4a]	[4b]	[5a]	[5b]
Jan-Fév	250	5%	5%	21	5	5.249	1.247
Mar-Avr	275	5%	5%	19	5	5.225	1.375
Mai-Juin	260	5%	5%	21	5	5.410	1.280
Jui-Août	249	5%	3,5%	21	8	5.229	1.977
Sep-Oct	275	5%	3,5%	19	7	5.225	1.922
Nov-Déc	274	5%	3,5%	19	7	5.206	1.900
Total	1 583	-	-	-	-	31.544	9.701

L'allocation par GRP des colonnes [4a] et [4b] a été calculée de la manière suivante. Pour les ménages de type EFT de la colonne [4a] (dans laquelle “#” signifie “nombre” et “rép.” “répondants”):

$$\begin{aligned} & \left[\frac{\text{totaal gewenst \# responderende huishoudens}}{\text{\# tweemaandelijke steekproeven}} \times (\text{aandeel LFS-type}) / (\text{\# GRP}'n) \right] \times \frac{1}{\text{responskans}} \\ &= \left[\frac{\text{totaal gewenst \# resp. LFS-type huishoudens per 2 maanden}}{\text{\# GRP}'n} \right] \times \frac{1}{\text{responskans}} \\ &= [\text{gewenst \# resp. LFS-type huishoudens per 2 maanden en per GRP}] \times \frac{1}{\text{responskans}} \end{aligned}$$

La division par la probabilité de réponse est une augmentation d'un nombre (souhaité) de ménages répondants à un nombre initial de ménages. Cela donne pour jan-fév après avoir arrondi:

$$\left[\frac{2.000}{6} \times 80\% / 250 \right] \times \frac{1}{5\%} = 21,33 \cong 21$$

Pour les ménages âgés de la colonne [4b], une formule analogue est d'application:

$$\left[\frac{\text{totaal gewenst \# responderende huishoudens}}{\text{\# tweemaandelijke steekproeven}} \times (\text{aandeel oudere}) / (\text{\# GRP}'n) \right] \times \frac{1}{\text{responskans}}$$

De sorte que pour p.ex. Mai-juin:

$$\left[\frac{2.000}{6} \times 20\% / 260 \right] \times \frac{1}{5\%} = 5,13 \cong 5$$

et pour Nov-Dec:

$$\left[\frac{2.000}{6} \times 20\% / 274 \right] \times \frac{1}{3,5\%} = 6,95 \cong 7$$

La définition des strates d'échantillonnages – sur la base du Tableau 13, nous concluons qu'il y a $1.583 \times 2 = 3.166$ strates – et du nombre calculé de ménages à sélectionner par strate, comme indiqué dans les colonnes [4a] et [4b], permettent de tirer l'échantillon. Le tirage au sein de chaque strate est aléatoire (en anglais, *simple random sampling*).

En principe, nous pouvons calculer les nombres des colonnes [5a] et [5b] à partir des colonnes [2], [4a] et [4b], par exemple pour mars-avr.: $275 \times 19 = 5.225$ et $275 \times 5 = 1.375$. Certains des nombres des colonnes [5a] et [5b] ne sont pas exactement des produits de ce type. Pour sep-oct, $275 \times 7 = 1.925 \neq 1.922$ s'applique, car 3 des 275 UPE ne contiennent que 6 et non au moins 7 ménages éligibles. C'est la raison pour laquelle nous avons repris aux colonnes [5a] et [5b] non pas les tailles d'échantillon calculées mais les tailles d'échantillon réalisées. Le Tableau 14 suivant clôture ce paragraphe.

Tableau 14: Les 6 échantillons bimestriels complémentaires d'EBM 2018: Chiffres de la population (limités aux UPE tirées dans les GR1*7)

Sous-échantillon complémentaire EBM	Nombre de MN éligibles		
	Type EFT	Âgé	Total
[1]	[2a]	[2b]	[3]
Jan-Fév	536.166	52.391	588.557
Mar-Avr	580.268	57.451	637.719
Mai-Juin	459.663	46.253	505.916
Jui-Août	527.242	50.418	577.660
Sep-Oct	561.585	55.857	617.442
Nov-Déc	557.897	55.308	613.205

Les chiffres de la population de la colonne [3] sont repris de la colonne [5] dans le Tableau 13. Les colonnes [2a] et [2b] montrent la ventilation des chiffres de la colonne [3] selon le type de ménage. Nous constatons que les chiffres de la colonne [2b] représentent environ 9 % des totaux de la colonne [3].⁸ Si nous comparons avec le ratio 80:20 du plan de sondage, nous pouvons conclure que les ménages plus âgés sont surreprésentés dans l'échantillon des ménages répondants. L'objectif de cette surreprésentation était de réaliser un nombre suffisant de ménages répondants âgés afin de permettre l'analyse (sommaire) des schémas des dépenses de cette sous-population.

2.4.4. Probabilités de sélection et poids de sondage

2.4.4.1. Pour les ménages de type EFT

Par tirage bimestriel et par strate d'échantillonnage d'UPE, le tirage complémentaire de ménages *de type EFT* pour EBM 2018 est la combinaison de:

- un tirage antérieur des UPE pour l'EFT (voir par.2.4.2) qui ne tient compte que des ménages de type EFT;
- le tirage de ménages dans la strate des ménages âgés (voir par. 2.4.3) qui est effectué formellement de la même manière qu'un tirage - au degré 2 - de ménages pour l'EFT (voir par.2.2.3).

Si les hypothèses telles que formulées en annexe 'L'échantillonnage à deux degrés en bref – Approche des poids de sondage pour les USE' sont valables, alors en principe, les probabilités de sélection sont à nouveau ici de la forme mg/N et les poids de sondage de la forme $N/(mg)$. Nous trouvons les valeurs de g à la colonne [4a] du Tableau 13; les valeurs de m se trouvent au Tableau 4 (à l'exclusion des valeurs de la DEU). Les sommes des valeurs de m entre les strates d'échantillonnage d'UPE par tirage bimestriel sont les valeurs de la colonne [2] du Tableau 13.

Cependant, nous nous écartons de la formule $N/(mg)$ de deux façons :

- Comme, lors du tirage des ménages, certains groupes ne peuvent pas être réalisés complètement (parce que les UPE correspondantes contiennent moins de g ménages de type EFT), nous remplaçons le dénominateur mg dans la formule des poids de sondage par le nombre de ménages réellement réalisé que nous notons sous la forme n_{rel} . Les sommes des strates d'échantillonnage des UPE de ce genre de nombres effectifs par tirage bimestriel se trouvent à la colonne [5a] du Tableau 13. Nous constatons que l'impact de cette adaptation est négligeable. C'est pourquoi la formule $N/(mg)$ est remplacée par N/n_{rel} .
- Davantage encore que lors d'un tirage à deux degrés pour l'EFT, le tirage des ménages est effectué plus tard que le tirage des UPE. Les extractions du RN qui sont utilisées lors des deux tirages successifs peuvent alors nettement différer, dans ce sens que les valeurs N – au total et par UPE – peuvent avoir nettement augmenté compte tenu de la croissance progressive de la population. Dans l'hypothèse (complémentaire et de simplification) que la population croît de manière égale entre les UPE, nous substituerons les valeurs N de l'extraction qui est utilisée lors du tirage des ménages dans la formule des poids de sondage. Remarquez que si nous utilisons les valeurs N de

⁸ Ce pourcentage s'applique également si l'on considère non seulement les ménages des UPE tirées pour les GR1 à GR7, mais aussi l'ensemble de la population belge.

la première extraction, les poids de sondage seraient réduits d'un certain facteur (plus ou moins constant), ce qui n'aurait aucune incidence sur le calibrage.

Le Tableau 15 montre les poids de sondage N/n_{rel} (approximatifs) pour les ménages de type EFT.

Tableau 15: Poids de sondage dans les échantillons complémentaires EBM 2018 de ménages de type EFT

Strate UPE		Echantillons bimestriels					
		Jan-Fév	Mar-Avr	Mai-Juin	Jui-Août	Sep-Oct	Nov-Déc
		GR1	GR2	GR3-4	GR5	GR6	GR7
1	ANT	1.198,64	1.199,50	1.162,88	1.206,08	1.209,53	1.209,51
2	BRU	615,34	682,92	620,38	621,12	688,42	686,27
3	WVL	913,47	898,98	880,38	917,73	904,45	904,86
4	OVL	1.055,87	1.049,28	1.022,55	1.063,10	1.056,83	1.057,53
5	HAI	975,74	1.003,19	979,25	979,64	1.007,36	1.007,70
6	LIE	968,78	941,12	969,63	970,01	945,43	944,99
7	LIM	777,02	782,04	709,12	780,82	786,09	786,57
8	LUX	368,38	357,04	380,33	399,26	359,26	383,77
9	NAM	612,92	599,29	577,56	616,71	603,00	603,18
10	VLB	907,39	887,33	874,20	914,59	891,97	890,72
11	BWA	504,44	489,56	473,33	507,70	491,57	491,54

2.4.4.2. Pour les ménages âgés

Par tirage bimestriel et par strate d'échantillonnage d'UPE, le tirage complémentaire de ménages âgés pour EBM 2018 est la combinaison de:

- un tirage antérieur des UPE pour l'EFT (voir par.2.4.2) qui ne tient compte que des ménages de type EFT;
- un tirage de ménages dans la strate des ménages âgés (voir par. 2.4.3) qui est effectué formellement de la même manière qu'un tirage - au degré 2 - de ménages pour l'EFT (voir par.2.2.3).

Le calcul des poids de sondage est, dans ce cas, compliqué à deux niveaux: (1) le tirage des ménages est effectué beaucoup de temps après le tirage des UPE; (2) l'extraction du RN qui est utilisée lors du tirage des UPE se limite aux ménages de type EFT alors que l'extraction lors du tirage des ménages se limite aux ménages âgés. Si l'on utilisait la même extraction de ménages âgés pour les deux tirages, alors une formule de la forme $N/(mg)$ peut encore être justifiée (moyennant les hypothèses telles que décrites en annexe L'échantillonnage à deux degrés en bref – Approche des poids de sondage pour les USE), avec les valeurs m de la colonne [4b] du Tableau 13 et les valeurs N de l'extraction des ménages âgés.

Comme lors du tirage de ménages de type EFT (par. 2.4.4.1), nous remplaçons aussi les dénominateurs mg dans la formule $N/(mg)$ pour les poids de sondage par les nombres réalisés n_{rel} de ménages âgés sélectionnés. La formule devient donc N/n_{rel} . Dans cette formule, nous conservons, les valeurs N de l'extraction de ménages âgés, ce qui est adéquat pour le calibrage des ménages âgés sélectionnés selon la population correspondante de ménages âgés.

Le rapport entre le nombre de ménages âgés et le nombre de ménages de type EFT est assez constant entre les UPE et stable au fil du temps (voir le dernier alinéa du par. 2.4.3), ce qui est une justification complémentaire de ce choix des valeurs N dans le calcul des poids de sondage, en dépit du tirage d'UPE sur

la base d'une extraction de ménages de type EFT et du tirage ultérieur de ménages âgés sur la base d'une extraction plus récente de ménages âgés.

Le Tableau 16 montre les poids de sondage (approximatifs) N/n_{rel} pour les ménages âgés.

Tableau 16: Poids de sondage dans les échantillons EBM 2018 complémentaires de ménages âgés.

Strate UPE		Echantillons bimestriels					
		Jan-Fév	Mar-Avr	Mai-Juin	Jui-Août	Sep-Oct	Nov-Déc
		GR1	GR2	GR3-4	GR5	GR6	GR7
1	ANT	552,20	499,41	533,48	348,58	356,23	356,16
2	BRU	187,96	187,58	186,83	116,22	132,69	134,25
3	WVL	526,53	468,54	510,11	328,80	335,81	336,08
4	OVL	503,85	451,63	483,73	313,13	321,45	320,85
5	HAI	381,34	353,45	379,54	240,35	251,42	249,45
6	LIE	390,61	343,79	389,07	242,01	245,57	243,74
7	LIM	333,22	304,45	305,61	209,96	219,58	224,70
8	LUX	155,26	131,88	172,66	108,65	93,83	106,37
9	NAM	240,56	212,27	225,10	149,12	150,78	157,04
10	VLB	430,22	378,70	411,20	267,66	270,15	270,32
11	BWA	206,36	178,61	190,36	126,79	127,35	127,94

3. Taux de réponse

3.1. Taux de réponse initial de l'EFT - Source 1

Depuis l'édition précédente d'EBM 2016, l'EFT a changé fondamentalement à la suite du passage à un panel alors que l'EFT était auparavant une enquête ponctuelle, les ménages sont maintenant interrogés à plusieurs reprises depuis le lancement du panel (mi-2016). Il a été décidé de poser la question de la participation à l'EBM lors de la dernière vague. De cette façon, poser la question de la participation à l'EBM ne peut avoir aucun impact sur la participation à l'EFT elle-même. L'inconvénient de cette méthode est que cette dernière vague est réalisée via une enquête CAWI ou CATI et non plus face à face par un enquêteur (CAPI). Ce n'est que dans le cas d'une enquête CATI qu'un enquêteur peut poser la question de recrutement et éventuellement donner des informations complémentaires aux personnes.

On s'attendait à ce qu'avec le passage de CAPI à CAWI comme méthode d'interrogation, la volonté des personnes de participer l'EBM diminue. Les chiffres le confirment.

Tableau 17: Taux de réponse initial de l'EFT - Source 1

Enquête EFT	Enquête EBM	Nombre de MN EFT positifs W4	Nombre de MN EFT positifs	% (Cf. 2016: 33%)
EFT 2017 T1	EBM 2018 JAN	1610	431	27
	EBM 2018 FÉV	1651	479	29
EFT 2017 T2	EBM 2018 MAR	1745	467	27
	EBM 2018 AVR	1657	489	30
EFT 2017 T3	EBM 2018 MAI	1552	389	25
	EBM 2018 JUIN	1553	439	28
EFT 2017 T4	EBM 2018 JUIL	1547	432	27
	EBM 2018 AOÛT	1679	472	28
EFT 2018 T1	EBM 2018 SEP	1744	424	24
	EBM 2018 OCT	1546	402	26
EFT 2018 T2	EBM 2018 NOV	1747	421	24
	EBM 2018 DÉC	1692	416	24
	Total annuel	19723	5261	27

La volonté générale de participer à l'enquête EBM diminue de 33 % pour EBM 2016 à 27 % pour EBM 2018. Cette diminution est quasi entièrement due au changement de mode d'enquête. C'est ce qui ressort du Tableau 18 suivant.

Tableau 18: Diminution en raison du changement de modalité d'enquête

Mode d'enquête question de recrutement	Volonté de participer à l'EBM (%)
CAWI (sans enquêteur)	25
CATI (avec enquêteur)	33

3.2. Taux de réponse initial de l'EBM - Source 2

Les ménages d'EBM 2016 ont été répartis sur les 24 périodes de référence d'EBM2018. La répartition de 2016 n'a en effet pas pu être reprise parce qu'en 2016, il y avait eu un biais dans la répartition par trimestre en raison du taux de réponse décevant au cours des deux premiers trimestres et de l'échantillon plus large au cours des deux derniers trimestres.

Chaque trimestre, un courrier a été adressé aux ménages des six périodes de référence afin de les inviter à participer à l'EBM. Il leur était demandé de renvoyer un formulaire sur lequel ils pouvaient indiquer s'ils étaient intéressés ou non par une participation. S'ils étaient intéressés, ils pouvaient indiquer leurs coordonnées.

Tableau 19: Taux de réponse initial de l'EBM - Source 2

Enquête EBM	Envoi				Résultats		
	Date 1 ^{er} envoi	Nombre de lettres	Date rappel	Nombre de rappels	Nombre négatifs	Nombre de MN positifs	% oui / total
Trimestre 1	17/10/2018	1123	09/11/2017	678	307	383	34%
Trimestre 2	24/01/2018	1119	15/02/2018	674	235	438	39%
Trimestre 3	24/04/2018	1116	15/05/2018	658	196	489	43%
Trimestre 4	16/7/2018	1132	16/08/2018	661	166	409	36%
TOTAL		4490				1719	38%

3.3. Taux de réponse initial du tirage complémentaire du RN - Source 3

Le SM fournissait un échantillon tous les deux mois. Le CEC était ensuite chargé d'attribuer la période de référence aux ménages et de calculer un numéro de ménage. Chaque mois, un courrier était adressé aux ménages des deux périodes de référence afin de les inviter à participer à l'EBM. Il leur était demandé de renvoyer un formulaire sur lequel ils pouvaient indiquer s'ils étaient intéressés ou non par une participation. S'ils étaient intéressés, ils pouvaient indiquer leurs coordonnées.

- Nombre de ménages par groupe EFT et par UPE: 21 de la population EFT + 5 de la population âgée
- A partir de l'échantillon de juillet : nombre par groupe: 21 de la population EFT + 8 de la population âgée

Note : ces chiffres sont des objectifs : chaque tirage s'est avéré avoir des UPE dans lesquelles il n'y avait pas assez de ménages pour atteindre ces chiffres => le nombre total de lettres varie donc chaque mois.

Tableau 20: Taux de réponse initial du tirage complémentaire du RN - Source 3

Mois EBM	Envoi		Résultats		
	Nombre total de lettres	Nombre de rappels	Nombre de réactions négatives	Nombre de réactions positives	% oui / total
Janvier	3222	2870	512	215	6.67%
Février	3274	3018	471	219	6.68%
Mars	3336	2949	597	227	6.80%
Avril	3264	2910	510	216	6.61%
Mai	3348	3037	565	212	6.33%
Juin	3340	2974	427	224	6.70%
Juillet	3557	3285	588	182	5.11%
Août	3558	3285	413	137	3.85%
Septembre	3612	2933	514	193	5.34%
Octobre	3532	3147	463	230	6.51%
Novembre	3577	3181	481	249	6.96%
Décembre	3528	3138	561	235	6.66%
TOTAL	41148	36727	6102	2539	6.17%

3.4. Echantillon initial total EBM2018 par mois

Chaque mois, le CEC établit un échantillon, composé des ménages positifs de l'EFT, des nouveaux ménages du RN et des ménages du panel de l'enquête EBM 2016 qui ont accepté de participer. Cet échantillon couvre alors les deux périodes de référence de ce mois.

Tableau 21: Echantillon initial total EBM2018 par mois

Enquête EBM	Nombre de MN EFT positifs	Nombre de MN EBM positifs 2016	Nombre de nouveaux MN positifs (RN)	Total
EBM 2018 JAN	431	109	215	755
EBM 2018 FÉV	479	131	219	829
EBM 2018 MAR	468	132	227	827
EBM 2018 AVR	489	150	216	855
EBM 2018 MAI	389	121	212	722
EBM 2018 JUIN	442	167	224	833
EBM 2018 JUIL	432	155	182	769
EBM 2018 AOÛT	472	166	137	775
EBM 2018 SEP	423	157	193	773
EBM 2018 OCT	402	150	230	782
EBM 2018 NOV	421	174	249	844
EBM 2018 DÉC	416	183	235	834
Total	5264	1795	2539	9598

Lorsque l'échantillon mensuel est prêt, le CEC crée un dossier pour chaque enquêteur reprenant ses listes d'adresses et le nombre de carnets nécessaire.

3.5. Aperçu du travail de terrain et du taux de réponse

Tableau 22: Aperçu du travail de terrain et du taux de réponse pour la Belgique et les régions pour EBM 2018

Partie travail de terrain / taux de réponse	D'un point de vue géographique	EFT	EBM 2016	Nouveau MN de type EFT	Nouveau MN âgés	Total
Recrutement						
Potentiels	Belgique	19.723	4.490	31.544	9.701	65.458
	Flandre	9.717	2.175	15.324	4.729	31.945
	Wallonie	7.050	1.784	11.540	3.532	23.906
	Bruxelles	2.956	531	4.680	1.440	9.607
Recrutés	Belgique	5.264	1.795	2.220	319	9.598
	Flandre	2.216	929	1.110	150	4.405
	Wallonie	2.257	672	835	120	3.884
	Bruxelles	791	194	275	49	1.309
Taux de réponse recrutement = recrutés/potentiels	Belgique	26,69%	39,98%	7,04%	3,29%	14,66%
	Flandre	22,81%	42,71%	7,24%	3,17%	13,79%
	Wallonie	32,01%	37,67%	7,24%	3,40%	16,25%
	Bruxelles	26,76%	36,53%	5,88%	3,40%	13,63%
Travail de terrain (1)						
Non effectué	Belgique	5	1	6	0	12
	Flandre	0	1	3	0	4
	Wallonie	5	0	3	0	8
	Bruxelles	0	0	0	0	0
<p>Potentiels: nombre de ménages qui ont été invités à participer à l'EBM 2018.</p> <ul style="list-style-type: none"> • Pour la partie de l'échantillon qui provient de l'EFT, il s'agit de tous les ménages qui ont effectivement participé à l'EFT. • Pour la partie de l'EBM 2016, il s'agit de tous les ménages qui ont participé à l'EBM en 2016. • Pour la partie du nouvel échantillon (tant du type EFT que les ménages âgés), il s'agit de tous les ménages qui ont été tirés du RN. <p>Recrutés: nombre de ménages qui ont confirmé qu'ils sont intéressés à participer à l'EBM 2018. Ils ont réagi 'positivement' à notre invitation.</p> <p>Non effectués: nombre de ménages qui n'ont pas été contactés, par manque d'enquêteur par exemple.</p>						

Partie travail de terrain / taux de réponse	D'un point de vue géographique	EFT	EBM 2016	Nouveau MN de type EFT	Nouveau MN âgés	Total
Travail de terrain (2)						
Refus de participer	Belgique	2654	128	423	77	3282
	Flandre	952	75	178	36	1241
	Wallonie	1256	32	184	24	1496
	Bruxelles	446	21	61	17	545
Aucune information	Belgique	49	9	20	3	81
	Flandre	18	5	12	1	36
	Wallonie	16	4	3	1	24
	Bruxelles	15	0	5	1	21
Refusés	Belgique	50	15	19	3	87
	Flandre	10	4	8	1	23
	Wallonie	22	7	6	1	36
	Bruxelles	18	4	5	1	28
Entièrement finalisés	Belgique	2 506	1 642	1 752	236	6 136
	Flandre	1 236	844	909	112	3101
	Wallonie	958	629	639	94	2320
	Bruxelles	312	169	204	30	715
Taux de réponse travail de terrain =Entièrement finalisés/Rrecrutés	Belgique	47,61%	91,48%	78,92%	73,98%	63,93%
	Flandre	55,78%	90,85%	81,89%	74,67%	70,40%
	Wallonie	42,45%	93,60%	76,53%	78,33%	59,73%
	Bruxelles	39,44%	87,11%	74,18%	61,22%	54,62%
Recrutement + travail de terrain						
Réponse globale =Entièrement finalisés/potentiels	Belgique	12,71%	36,57%	5,55%	2,43%	9,37%
	Flandre	12,72%	38,80%	5,93%	2,37%	9,71%
	Wallonie	13,59%	35,26%	5,54%	2,66%	9,70%
	Bruxelles	10,55%	31,83%	4,36%	2,08%	7,44%
<p>Refus de participer = nombre de ménages qui n'ont finalement pas participé.</p> <p>Aucune info: nombre de ménages pour lesquels l'enquêteur n'a pas transmis toutes les données.</p> <p>Refusés: nombre de ménages qui ont rempli un carnet et un questionnaire ménage, mais qui n'ont pas été acceptés comme répondant, parce que les contrôles démontrent que la qualité des données n'est pas satisfaisante.</p> <p>Entièrement finalisés: nombre de ménages pour lesquels la qualité des données est suffisante pour être acceptés comme répondant.</p>						

3.6. Raisons de la non-réponse

Dans la lettre d'invitation qui est envoyée aux ménages, ceux-ci ont également la possibilité de donner la raison pour laquelle ils ne veulent pas participer. Le Tableau 23 donne un aperçu de ces raisons. Si l'on analyse ces raisons, les principales raisons invoquées par les ménages âgés sont qu'ils se considèrent trop âgés pour participer ou qu'ils ont un problème de santé qui les empêche de participer ou qu'ils considèrent que leur propre situation n'est pas pertinente (par exemple parce qu'ils ne font pas les courses eux-mêmes et n'ont donc pas une vue d'ensemble de leurs dépenses). Les ménages issus de l'échantillon de type EFT et du panel EBM indiquent comme raison principale de ne pas participer le fait qu'ils n'ont pas le temps ou qu'ils ne sont pas disponibles pendant la période proposée.

Tableau 23: raison du refus de participer à l'EBM

	Panel EBM (%)	Nouveau MN de type EFT (%)	Nouveau MN âgés (%)
Raison du refus (du recrutement et du travail de terrain)	Pourcentage		
Pas le temps, trop occupé, pas disponible pendant la période	57,96	45,08	8,90
Déjà (trop) participé	10,02	1,24	0,34
Je suis trop âgé / malade	9,69	9,79	68,78
Raison familiale (un proche est malade / décédé, je viens de déménager, de me séparer, etc.)	7,39	1,52	0,94
Autre	5,42	12,65	1,71
Pas intéressé	4,76	14,77	5,56
Atteinte à la vie privé / trop privé	2,12	6,21	2,65
Doute quant à l'utilité de l'enquête	0,99	1,15	0,43
Ma situation n'est pas intéressante (isolé, indépendant, faibles revenus, cultive mes propres légumes, ne travaille pas, etc.)	0,99	5,06	3,42
En maison de repos	0,33	1,01	6,67
Problèmes linguistiques	0,33	1,29	0,51
Négatif envers les autorités (impôts, ils savent déjà tout, etc.)	0,00	0,23	0,09
Total	100,00	100,00	100,00
Nombre total de refus reçus	609	2174	1169

À la fin de la dernière vague, il est seulement demandé aux participants au panel EFT s'ils souhaitent ou non participer à l'EBM. Si la réponse est négative, aucune autre question n'est posée sur la raison.

4. Collecte et traitement des données pour le calibrage

4.1. Collecte des données

Le travail de terrain est mené par des enquêteurs qui ont contact à trois reprises avec les ménages après la remise d'une lettre d'introduction qui a reçu une réponse positive des ménages.

1. Lors de la première visite, l'enquêteur se rend au domicile du ménage pour remettre tous les documents et fournir des explications.
2. Le deuxième contact se déroule soit de manière physique, si les ménages remplissent leur carnet sur papier (PPSI) ou s'ils en manifestent le besoin, soit par téléphone si les ménages remplissent leur carnet en ligne. Il existe en effet depuis 2012 une application numérique (CAWI) pour le carnet des ménages ^Y.
3. La troisième visite a lieu après la période de référence pendant laquelle les ménages ont consigné leurs dépenses. L'enquêteur réalise alors une interview (CAPI depuis 2014). Dans des circonstances exceptionnelles, ce troisième contact peut également se dérouler par téléphone (CATI).

Tableau 24: Modalité pour remplir le carnet et le questionnaire des ménages

Manière de participer	2012 (%)	2014 (%)	2016 (%)	2018 (%)
Carnet				
CAWI	20	30	55	58
PPSI	80	70	45	42
Questionnaire ménage				
CAPI	0	100	100	99.5
CATI	0	0	0	0.5
Sur papier	100	0	0	0

Après la période de référence, les enquêteurs envoient les carnets des ménages collectés par la poste aux centres de traitement. Il existe 5 centres régionaux de traitement où sont traitées les enquêtes de certaines provinces ou régions.

Tableau 25: Aperçu des centres de traitement

Centre	Région	Province
Brussel	RU	BRU (ville)
Bruxelles	BRU	BWA VLB
Gand	VLA	OVL LIM WVL
Anvers	VLA	ANT
Liège	WA	LIE LUX
Charleroi	WA	HAI NAM

Les collaborateurs des centres de traitement régionaux enregistrent les carnets reçus et procèdent à un contrôle de qualité visuel. Si les carnets sont de qualité suffisante, ils vérifient s'ils ont reçu le questionnaire ménage. Si ces deux éléments sont en ordre, il est procédé au paiement du ménage. Le paiement des enquêteurs est également effectué à ce moment. Les enquêteurs sont payés par visite et peuvent donc également être (partiellement) payés pour des enquêtes qui ont finalement été interrompues.

Un [arrêté royal du 17 janvier 1996](#) fixe le montant de l'indemnité aux ménages. Les ménages reçoivent pour leur participation un montant de 45 euros, ainsi qu'éventuellement un bonus de 5 euros s'ils ont complété le carnet entièrement en ligne. Un [arrêté royal du 17 mai 2006](#) fixe le montant de l'indemnité aux enquêteurs. Les enquêteurs reçoivent un montant brut de 15.85 euros par visite. Pour une enquête entièrement terminée, 4 visites sont rémunérées. Comme la dernière visite est assez longue parce qu'il faut remplir le questionnaire ménage et vérifier le carnet, cette visite sera facturée deux fois. L'enquêteur reçoit également une prime de 2,81 euros par questionnaire ménage complété pour avoir posé les questions supplémentaires sur l'énergie ajoutées depuis 2016.

4.2. Enregistrement des données

Tous les membres du ménage doivent consigner leurs dépenses dans un carnet des ménages^{vi} pendant une période de référence donnée. Cette période de référence est définie à l'avance. En 2012, 2014 et 2016, cette période de référence était d'un mois calendrier. Depuis 2018, la période de référence a été ramenée à 15 jours, soit du 1^{er} au 15 du mois, soit du 16 à la fin du mois. A la fin de la période de référence, la personne de référence du ménage est interrogée au moyen d'un questionnaire ménage.^{vii}

4.2.1. *Questionnaire ménage*

Le questionnaire ménage aborde les sujets suivants ^{viii}:

1. Composition du ménage
2. Questions sur l'habitation (résidence principale et seconde résidence)
(depuis 2016: y compris le module énergie)
3. Dépenses périodiques (énergie et logement)
4. Équipement du logement
5. Achat de biens durables
6. Assurances
7. Revenus
8. Avantages en nature
9. Relations entre les membres du ménage
10. Absences

Les dépenses qui sont demandées dans le questionnaire ménage sont les dépenses périodiques et couvrent les dépenses effectuées au cours des trois mois qui précèdent la période de référence et le mois de la période de référence, sauf indication contraire. Selon les feedbacks des enquêteurs, le questionnaire ménage prend en moyenne 40 à 50 minutes à remplir.

4.2.2. Carnet de dépenses

Pendant la période de référence, les ménages enregistrent toutes leurs dépenses dans le carnet. Cela peut se faire dans l'application en ligne ou sur un carnet papier. Les encodeurs des centres de traitement régionaux introduisent les carnets papier dans l'application en ligne.

Le carnet consiste en un enregistrement quotidien de toutes les dépenses ordinaires et un enregistrement (limité) des produits issus de la propre culture ou de la propre entreprise du ménage. Une description détaillée de chaque produit est demandée dans les différentes colonnes:

1. Description du produit, la plus détaillée possible;
2. Les montants dépensés au moment de l'achat;
3. Le solde qui doit encore être payé ultérieurement;
4. La quantité;
5. L'unité (litre, kilogramme, ou pièces);
6. Le lieu de l'achat;
7. Le pays d'achat;
8. Les dépenses mixtes pour indépendants.

L'application en ligne présente l'avantage que le produit acheté peut être placé immédiatement dans la catégorie correspondante de la nomenclature ECOICOP. Cela se fait au moyen d'une liste de termes établie au préalable dans laquelle chaque article est lié à un code de la nomenclature. Un autre avantage de l'application en ligne est qu'un certain nombre de contrôles de qualité peuvent être effectués immédiatement. La première étape consiste à vérifier si une quantité a été encodée. Ensuite, il y a aussi un contrôle des prix des produits. Il est vérifié si le prix indiqué est plausible par rapport à la quantité indiquée. Ce contrôle est effectué sur la base de fourchettes de prix déterminées à l'avance.

4.2.2.1. Fourchettes de prix

Des fourchettes de prix sont utilisées dans le programme d'enregistrement en ligne des dépenses. Chaque catégorie de la nomenclature ECOICOP se voit attribuer un prix minimum et un prix maximum par unité achetée. Si les ménages enregistrent une dépense dont le prix d'achat se situe en dehors des fourchettes de prix, ils reçoivent un avertissement. L'avertissement consiste en un triangle rouge avec un point d'exclamation qui apparaît à côté de la ligne de la dépense.

Avant le début de chaque édition de l'EBM, les fourchettes de prix sont déterminées en utilisant les enregistrements de l'édition précédente. Pour chaque catégorie de la ECOICOP, les centiles des dépenses par unité sont calculés. La limite inférieure est C5, la limite supérieure C95. Les centiles sont ensuite arrondis à 1 eurocent. Un contrôle manuel des fourchettes de prix obtenues est ensuite effectué. Afin d'effectuer ce contrôle, la qualité des anciennes fourchettes de prix de l'édition précédente est également vérifiée. À cette fin, on examine combien de fois un produit a été enregistré lors de l'édition précédente en dehors des fourchettes de prix (et n'a pas été corrigé), à la fois en nombres absolus et en termes relatifs (pourcentage).

Le contrôle manuel s'effectue dans un tableau Excel. Toutes les fourchettes de prix sont vérifiées et peuvent être ajustées, mais certaines catégories sont marquées de manière à être certainement contrôlées. Le marquage indique que:

- Plus de 15 % des enregistrements dans cette catégorie de la ECOICOP se situaient en dehors des anciennes fourchettes de prix.

- Plus de 10 % des enregistrements dans cette catégorie de l'ECOICOP se situaient en dehors des anciennes fourchettes de prix, ce qui représentait plus de 250 enregistrements en chiffres absolus.
- Très peu de dépenses d'un certain produit ont été enregistrées et le calcul n'est donc pas très fiable (nombre de dépenses inférieur à 70).
- Il y a une grande différence entre les anciennes fourchettes de prix et les nouvelles (la différence est supérieure à 50 %).
- Le prix minimum est inférieur à 5 cents par unité.

La personne responsable a la possibilité d'ajuster les fourchettes de prix dans la liste Excel. Pour les nouvelles catégories de l'ECOICOP, les prix seront trouvés sur Internet. Ensuite, la personne responsable détermine une unité standard ainsi qu'un prix minimum et un prix maximum.

4.2.2.2. Nomenclature

Les dépenses sont agrégées au niveau des ménages et extrapolées aux dépenses annuelles selon la nomenclature ECOICOP^{ix}. Cette nomenclature contient 12 catégories ECOICOP de dépenses:

01. Produits alimentaires et boissons non alcoolisées
02. Boissons alcoolisées, tabac et stupéfiants
03. Articles d'habillement et chaussures
04. Logement, eau, gaz, électricité et autres combustibles
05. Meubles, articles de ménage et d'entretien courant du logement
06. Santé
07. Transports
08. Communication
09. Culture et loisirs
10. Formation
11. Restaurants et hôtels
12. Biens et services divers

Chacune de ces catégories est subdivisée en sous-catégories. Le niveau européen va jusqu'à 5 positions. Il y a un 6^e niveau qui est une subdivision nationale supplémentaire qui fournit davantage de détails sur les catégories. Les résultats nationaux sont diffusés jusqu'au sixième niveau. Les résultats régionaux sont diffusés jusqu'au quatrième niveau. Il n'y a pas de diffusion d'un niveau plus détaillé dans les régions car les résultats obtenus ne sont pas suffisamment précis en raison de la taille limitée des échantillons par région.

4.3. Détermination de la personne de référence sur la base du revenu

Dans l'EBM, la personne de référence est la personne ayant le revenu net le plus élevé. Cette personne doit d'abord être déterminée avant de pouvoir effectuer le calibrage.

Jusqu'en 2010, les revenus étaient inscrits dans le carnet de manière exhaustive, tout comme les dépenses. Depuis 2012, les revenus sont collectés de manière plus globale dans un questionnaire afin de limiter la charge de travail des répondants. Cette méthode de travail offre la possibilité de comparer les dépenses des ménages en fonction de leur niveau de vie (p.ex. quartiles de revenu) mais les tableaux détaillés sur la composition des revenus ne seront plus basés sur l'EBM.

Pour toutes les études sur les revenus, nous recommandons d'utiliser l'enquête EU-SILC (enquête communautaire sur les revenus et conditions de vie), qui est également organisée par Statbel^x.

Pour le calcul du revenu disponible, tous les revenus du ménage sont additionnés, aussi bien les revenus communs que les revenus individuels des membres du ménage. Il s'agit plus spécifiquement des revenus mensuels nets suivants au niveau individuel:

- R1: salaire net
- R2: revenu net indépendant
- R3: revenu net des investissements (y compris les intérêts, les dividendes...)
- R4 : revenu net d'immeubles loués (loyers nets reçus)
- R5: pension de vieillesse
- R6: prépension (employeur + chômage)
- R7: allocations de chômage
- R8 : pension de veuf et d'orphelin
- R9: allocations familiales
- R10: allocations maladie/invalidité
- R11: revenu de l'activité secondaire
- R12: autre revenu
- R13: avantages en nature
- R14: remboursement mutualité
- R15: remboursement assurance
- R16: remboursement impôts

Les revenus sont vérifiés en détails par catégorie (identification des valeurs aberrantes, vérification des ménages qui n'ont pas rempli correctement les informations : questionnaire ménage vide, revenus trop faibles ou non conformes aux informations sur le statut socioprofessionnel, secteur d'activité ...).

Ensuite, un revenu est calculé par personne en prenant la somme de toutes les catégories. Le revenu du ménage est calculé en prenant la somme de tous les revenus des individus du ménage.

La somme de chaque type de revenu R1 à R16 est comparée au revenu déclaré par le ménage, en partant du principe que le revenu auto-déclaré par le ménage est plus précis parce que les ménages le connaissent mieux.

Un premier contrôle de qualité est effectué en validant le revenu moyen d'EBM par rapport au revenu moyen déclaré par les ménages dans l'enquête EU-SILC. Le Tableau 26 donne un aperçu de la comparaison des valeurs statistiques de base pour le revenu au niveau du ménage dans l'EBM avec le revenu au niveau du ménage trouvé dans l'enquête EU-SILC pour la même année de référence.

Tableau 26: Validation du revenu pondéré au niveau du ménage collecté dans l'EBM par rapport aux données collectées dans EU-SILC

Weighted Basic Statistical Measures (Location & Variability)	HBS 2018	EU-SILC 2018
Mean	41073.55	39238.46
Std Deviation	964186	829930
Median	32640.00	32536.38
Variance	9.29655 ^{E11}	6.88785 ^{E11}
Mode	14400.00	14400.00
Range	814885	976140
Interquartile Range	31286	32875

Un autre contrôle de qualité est effectué en répartissant les revenus en catégories ou en classes. Si aucun revenu ne peut être trouvé pour un ménage à partir des revenus déclarés entre R1 et R16 et qu'aucune donnée n'est disponible dans l'EFT, les classes sont utilisées pour imputer les valeurs médianes des valeurs manquantes. Les valeurs des classes obtenues sont à nouveau comparées avec les valeurs trouvées dans l'enquête EU-SILC. Le Tableau 27 donne un aperçu des classes de revenus pour EBM et EU-SILC pour la même année de référence.

Tableau 27: Répartition du revenu selon les classes

Income class	Income HBS 2018		Income EU-SILC 2018	
	Percent	Cumulative Percent	Percent	Cumulative Percent
[0-250[0.29	0.29	0.69	0.69
[250-500[0.15	0.44	0.31	1.00
[500-1000[2.13	2.57	3.87	4.87
[1000-1500[13.79	16.36	15.81	20.68
[1500-2000[16.55	32.91	13.49	34.16
[2000-2500[12.26	45.17	10.97	45.13
[2500-3000[10.15	55.32	10.13	55.26
[3000-5000[26.83	82.15	26.54	81.8
[5000+	17.85	100.00	18.20	100

5. Calibrage et estimation

5.1. Introduction

Pour le calibrage de l'échantillon de répondants EBM 2018, nous nous basons sur le modèle développé pour EBM 2016, qui constituait à divers égards en un renouvellement des modèles de calibrage ou de pondération des enquêtes EBM 2014 et antérieures. La raison est la recherche d'une continuité dans la méthodologie, afin d'éviter les ruptures dans les statistiques publiées dues aux modifications méthodologiques.

Néanmoins, lors du calibrage de l'enquête EBM 2018, nous devons tenir compte d'une complexité accrue, due en grande partie à la modification du recrutement : les répondants sont désormais issus de trois "sources" différentes (EFT, EBM et RN), comme nous l'avons vu au paragraphe 2. Cela rend à son tour le calibrage plus complexe, dans le sens où différents sous-échantillons d'EBM 2018 sont calibrés indépendamment les uns des autres et que les modèles respectifs sont adaptés aux sous-échantillons.

Une partie importante du travail, en raison de la complexité accrue, est effectuée plus tôt, lors de la préparation du calibrage. Au par. 2, nous avons déjà abordé – dans le cadre du recrutement à partir de trois sources – le calcul des poids de sondage (c'est-à-dire les poids initiaux dans le calibrage). Il convient de noter que le calcul de ces pondérations initiales comprend l'estimation de toutes sortes de probabilités de réponse et d'auto-sélection. Ces probabilités représentent les étapes successives du processus de recrutement – variant selon la source – des ménages invités à participer à l'EBM 2018. L'ensemble des ménages invités à participer à l'EBM 2018 peut être considérée comme l'échantillon EBM 2018 initial. Le poids initial correspondant à chacun de ces ménages peut être considéré comme un poids de sondage (bien que *poids de recrutement* serait un terme plus approprié, pour insister sur la complexité de la procédure de recrutement).

Chaque ménage recruté et invité à participer à l'EBM 2018 a une probabilité de réponse intrinsèque, qui sera abordée à la section 0. Les poids initiaux (ou poids de sondage ou poids de recrutement) mentionnés au paragraphe précédent ne sont pas corrigés sur la base de ces probabilités de réponse intrinsèques. Les poids initiaux sont donc directement corrigés par le calibrage, ce qui suppose de manière implicite que le calibrage corrige suffisamment la non-réponse à partir du recrutement et de l'invitation à participer à l'EBM 2018.

Au par. 5.2 ci-après, nous aborderons les cadres de calibrage à partir desquels sont calculés les totaux de calibrage ou benchmarks. Ensuite, au par. 5.3, nous aborderons les modèles de calibrage qui seront finalement utilisés pour les sous-échantillons d'EBM 2018. Le choix de ces modèles sera également discuté. Pour les détails techniques et l'interprétation, nous nous référons à *Enquête sur le budget des ménages 2016 - Note méthodologique* ([Statbel \(2017a\)](#) ou [Statbel \(2017b\)](#)), où les modèles renouvelés sont discutés en détail.

5.2. Cadres de calibrage

5.2.1. Un échantillon EFT 2018 calibré

Dans *Enquête sur le budget des ménages 2016 - Note méthodologique* ([Statbel \(2017a\)](#) ou [Statbel \(2017b\)](#)), il a été longuement discuté de la manière dont l'échantillon de répondants de l'EFT 2016 a été calibré selon la population, pour estimer les totaux de calibrage appropriés pour le calibrage de l'échantillon de répondants d'EBM 2016. Le modèle utilisé pour l'EFT 2016 était une nouveauté par rapport aux anciens modèles de

calibrage de l'EFT : nous sommes passés d'un calibrage au niveau individuel à un calibrage simultané au niveau individuel et du ménage, car les poids de calibrage fournis par cette méthode sont davantage étayés pour les ménages. Le modèle de calibrage de l'EFT 2018 est formellement le même (voir par. 5.1.3 de *Enquête sur le budget des ménages 2016 - Note méthodologique* ([Statbel \(2017a\)](#) ou [Statbel \(2017b\)](#))):

{HH; STRAT12; SEX × AGE // HHSIZE; Logit}

Ce modèle (avec notamment la légende entre parenthèses) a été abordé en détail au par. 5.1.3 dans [Statbel \(2017a\)](#) ou [Statbel \(2017b\)](#). Quelques différences entre EFT 2016 et EFT 2018 méritent toutefois l'attention:

- L'EFT 2016 a été calibrée sur une population de 11.141.994 individus dans 4.852.486 ménages privés ; l'EFT 2018 a été calibrée sur une population de 10.663.202 individus dans 4.465.140 ménages privés *de type EFT*. Cette opération est nécessaire parce que les sous-échantillons d'EBM 2018 issus de différentes sources seront calibrés séparément et parce que l'échantillon calibré de l'EFT 2018 ne fournira que les totaux de calibrage pour les calibrages des sous-échantillons d'EBM 2018 contenant des ménages de type EFT, c'est-à-dire les sources 1, 2 et 3,1.
- À partir de 2017, l'EFT est passée à une enquête par panel (qui a démarré au T3 de 2016). Par conséquent, les échantillons trimestriels - que nous appelons les *groupes de rotation* (GR) dans le cadre du panel - sont devenus beaucoup plus petits : les GR qui ont été utilisés sur le terrain pour l'EFT en 2018 (GR n° 11, 12, 13 et 14) contiennent chacun 6.695 ménages. Lors d'une première interrogation ou première « vague » - au cours du trimestre durant lequel les GR ont été lancés - nous avons recueilli les données de 18.369 ménages répondants (environ 68,5% de $6.695 \times 4 = 26.780$) contenant 43.637 individus. En 2016, nous disposions des données de 41.014 ménages répondants auxquels appartenait 98.568 individus répondants.
- Les poids de sondage des 18.369 ménages répondants de l'EFT 2018 (vague 1) ont été corrigés sur la base d'une probabilité de réponse estimée ; cela n'a pas été fait pour les 41.014 ménages répondants de l'EFT 2016. Par conséquent, en 2018, les poids de sondage corrigés de la non-réponse sont ceux qui ont été corrigés à nouveau lors du calibrage. En 2016, les poids de sondage ont été corrigés directement par le calibrage. La correction explicite de la non-réponse a été introduite pour l'EFT en 2017 T1 (comme première étape du calibrage des poids de sondage), c'est-à-dire lors du premier calibrage trimestriel pour l'estimation des indicateurs du marché du travail basés sur des données du panel. Voir [Vanderhoeft et al \(2021\)](#) pour une discussion du *mixed effects logistic regression model* pour l'estimation des probabilités de réponse ; les prédicteurs utilisés dans le modèle de régression sont *HHtype, Origin, Province* et *Urbanisation*.

Les limites inférieure et supérieure des *g-weights* - c'est-à-dire les facteurs de correction déterminés par le calibrage - lors de l'application de la méthode logit sont, tout comme en 2016, respectivement 0 et 10.

Les totaux de calibrage du modèle de calibrage de l'EFT 2018 (vague 1) – comme par exemple ceux de la ligne “Total” dans le Tableau 28 ci-après – sont calculés sur la base d'une extraction du RN du 1/07/2018.

Le Tableau 28 donne les sommes des poids calibrés par région et selon la variable de “poststratification” PS, qui joueront un rôle dans le calibrage de l'échantillon d'EBM 2018 (voir par. 5.3.1 et 5.3.2).

Tableau 28: Sommes des poids calibrés pour l'échantillon de répondants calibré d'EFT 2018 vague 1

PS : Post-strate (Au sein de la Région)	Région			Total
	BRU	VLA	WAL	
Salarié, 1P	98.206,36	320.055,53	162.196,68	580.458,57
Salarié, 2P, <=1act	34.042,98	186.973,47	111.555,60	332.572,05
Salarié, 2P, 2act	27.197,41	234.959,70	87.389,28	349.546,39
Salarié, 3+P, <=1act	53.488,91	173.456,74	131.208,52	358.154,17
Salarié, 3P, 2+act	21.707,30	214.621,16	99.625,90	335.954,35
Salarié, 4+P, 2+act	37.353,16	359.926,42	167.913,37	565.192,96
Indépendant	61.753,69	274.951,16	125.881,57	462.586,42
Non actif, <=59, 1P	55.057,37	99.651,87	108.361,09	263.070,33
Non actif, <=59, 2+P	40.155,20	69.650,51	99.727,22	209.532,92
Non actif, 60-69, 1P	29.909,89	123.372,73	100.314,71	253.597,33
Non actif, 60-69, 2+P	15.161,70	173.280,86	94.403,94	282.846,50
Non actif, 70+, 1P	21.565,96	102.804,16	69.024,51	193.394,63
Non actif, 70+, 2+P	15.221,07	179.990,69	83.021,63	278.233,39
Total	510.821,00	2.513.695,00	1.440.624,00	4.465.140,00

Les ventilations estimées de la population de l'échantillon EFT 2018 calibré seront calculées et seront utilisées comme benchmarks (totaux de calibrage) dans le calibrage d'EBM 2018: voir par. 5.3.1. Il s'agit ici des variables qui sont utilisées dans la construction de la variable de "post stratification" PS, à savoir le statut socio-professionnel et l'âge de la personne de référence du ménage, et le nombre de membres du ménage qui appartiennent ou non) la classe d'âge active. Le modèle de calibrage d'EFT 2018 est supposé réduire suffisamment le biais dans l'échantillon de répondants EFT, pour que le calibrage de l'EFT n'introduise pas de biais dans le calibrage de l'EBM.

5.2.2. Population de ménages de type EFT issus du RN

L'extraction du RN datant du 1/07/2018, qui a été complétée par le couplage au registre fiscal reprenant le revenu net imposable de 2016 (pour les individus et les ménages) a permis la construction d'un deuxième cadre de calibrage, dans le but de calculer des totaux de calibrage afin que certaines variables de revenu puissent être incluses dans les modèles de calibrage pour EBM 2018.

Ce deuxième cadre de calibrage est limité aux ménages de type EFT (et donc privés) dans lesquels les variables géographiques (région et province) et le revenu (net imposable) du ménage sont transformés en variables de classe de revenu INK_5, INK_5_G et INK_5_P (respectivement au niveau global, régional et provincial). La construction de ces variables est similaire au calibrage de l'EBM 2016 et est expliquée en détail au par. 5.2.3 de *Enquête sur le budget des ménages 2016 - Note méthodologique* ([Statbel \(2017a\)](#) ou [Statbel \(2017b\)](#)). La seule différence est que nous nous limitons maintenant aux ménages de type EFT, alors que pour EBM 2016, tous les ménages privés avaient été retenus.

Le "_5" dans le nom des variables des classes de revenu fait référence aux quintiles de population sous-jacents K_1, K_2, K_3, K_4 , qui sont calculés pour l'ensemble de la population ainsi que par région et par province, et qui sont également nécessaires pour calculer les variables INK_5, INK_5_G et INK_5_P pour les MN répondants d'EBM 2018 (voir par. 5.2.3 de *Enquête sur le budget des ménages 2016 - Note méthodologique* ([Statbel \(2017a\)](#) ou [Statbel \(2017b\)](#))).

Le Tableau 29 présente la distribution relative (en pourcentage) de INK_5_G par région, telle qu'elle est dérivée de la population des ménages de type LFS dont il est question ici. La distribution absolue (par région)

est obtenue en multipliant les totaux du Tableau 28 par les pourcentages. De cette façon, il y a une cohérence immédiate entre les totaux de calibrage dans le Tableau 28 et le Tableau 29.

Tableau 29: Répartition en pourcentage et en chiffres absolus des classes de revenus du ménage par région – variable INK_5_G

INK_5_G	BRU		VLA		WAL	
	Abs.	%	Abs.	%	Abs.	%
1	98.333,04	19,25	86.722,48	3,45	77.793,70	5,40
2	82.497,59	16,15	485.394,50	19,31	272.566,06	18,92
3	82.497,59	16,15	485.394,50	19,31	272.566,06	18,92
4	82.497,59	16,15	485.394,50	19,31	272.566,06	18,92
5	82.497,59	16,15	485.394,50	19,31	272.566,06	18,92
6	82.497,59	16,15	485.394,50	19,31	272.566,06	18,92
Total	510.821,00	100,00	2.513.695,00	100,00	1.440.624,00	100,00

Les ventilations de INK_5 et INK_5_P ne sont pas présentées parce que ces variables n'ont finalement pas été retenues dans les modèles de calibrage d'EBM 2018.

5.2.3. Population de ménages âgés issus du RN

Le (petit) sous-échantillon d'EBM 2018 de ménages âgés (voir par. 2.4) sera calibré séparément selon une sous-population correspondante de ménages âgés selon le RN. A partir d'une extraction du RN du 1/07/2018 5.2.1- la même que celle utilisée pour le calibrage d'EFT 2018 (par. 5.2.1) – nous obtenons une population de 456.272 ménages privés âgés, répartis de la manière suivante selon la région et la taille du ménage (personne seule vs. ménage comptant plusieurs personnes).

Tableau 30: Nombre de ménages âgés selon la région et la taille du MN (RN du 1/07/2018)

	Seul	Pas seul	Total	%
BRU	28.907	7.164	36.071	7,91%
VLA	194.567	91.743	286.310	62,75%
WAL	100.656	33.235	133.891	29,34%
Total	324.130	132.142	456.272	100,00%
%	71,04%	28,96%	100,00%	

Les nombre au Tableau 30 sont les totaux de calibrage du sous-échantillon EBM 2018 de ménages âgés (source 3.2); voir par. 5.3.3.

5.3. Calibrage d'EBM 2018

5.3.1. Calibrage de ménages de type EFT de l'EFT (source 1) et du RN (source 3.1)

Le sous-échantillon EBM 2018 de ménages répondants qui sont issus des sources 1 et 3.1 se rapproche le plus de l'échantillon de répondants d'EBM 2016; nous remarquons que ce sous-échantillon contient au total 4.258 ménages répondants dont 2.506 de la source 1 et 1.752 de la source 3.1. Cela laisse supposer que le modèle de calibrage d'EBM 2016 est adapté pour ce sous-échantillon EBM 2018. Dès lors, nous avons testé, comme pour EBM 2016, diverses variantes du modèle

$$\{HH; REGION \times PS + PROV \times (C12 + C13 + C14) + REGION \times INK_5_G; TrLin\}$$

et nous avons constaté que ce modèle est, en effet, encore toujours adéquat.

Pour des explications détaillées sur ce modèle (et les variantes), nous vous renvoyons au par. 5.2 de *Enquête sur le budget des ménages 2016 - Note méthodologique* ([Statbel \(2017a\)](#) ou [Statbel \(2017b\)](#)). Pour rappel :

- Le premier terme $REGION \times PS$, constitue une post-stratification classique par région et type de ménage (selon le statut d'occupation de la personne de référence et des autres membres du ménage, la taille du ménage et l'âge de la personne de référence : la variable PS est exactement telle que définie pour EBM 2016, avec 13 catégories (cf. Tableau 28));
- Le deuxième terme, $PROV \times (C12 + C13 + C14)$, apporte un équilibre entre les trimestres (via les variables de contraste C12, C13 en C14) au niveau provincial. Les alternatives $(C12 + C13 + C14)$ et $REGION \times (C12 + C13 + C14)$ ont été testées mais ne se sont pas révélées significativement différentes (meilleures).

Le troisième terme, $REGION \times INK_5_G$, permet d'aligner le ratio propriétaires/locataires sur celui estimé à partir du SILC : les modèles avec revenus dans le calibrage corrigent bien la surestimation des propriétaires par les modèles sans revenus. L'alternative plus détaillée $PROV \times INK_5_P$ n'a pas pu être appliquée pour la raison purement technique que, pour certaines provinces, la catégorie missing pour INK_5_P n'est pas représentée dans l'échantillon des répondants.

Ce modèle de calibrage a vu le jour progressivement dans le cadre du calibrage de l'EBM 2016, à partir d'un modèle comportant exclusivement le premier terme $REGION \times PS$. Après une extension au terme $REGION \times INK_5_G$, on peut se demander s'il n'y a aucun lien entre les variables INK_5_G et PS et si, par conséquent, on peut rechercher une simplification du modèle (p.ex. construire PS sans utiliser le statut professionnel, qui peut être une variable liée au revenu). Nous n'avons pas abordé cette problématique plus en détail, pour les raisons suivantes: 1) le statut professionnel ne joue qu'un rôle limité dans la construction de PS; 2) nous attendons tout de même une importance variation des revenus dans chaque catégorie de statut professionnel; 3) nous estimons que la présence d'un effet spécifique via différents termes dans les modèles de calibrage n'a aucune conséquence négative sur les poids calibrés (tant que la variabilité des poids ne devient pas trop grande par l'inclusion de nouvelles variables). Dans ce dernier point, les modèles de calibrage se distinguent des modèles de régression: dans les modèles de régression, une variable deviendra moins significative si elle est déjà vérifiée pour d'autres variables corrélées avec cette variable ; mais finalement, également dans les modèles de régression, l'estimation de la variable dépendante ne changera pas de manière significative si une variable corrélée avec d'autres variables indépendantes est en plus incluse dans le modèle.

Les totaux de calibrage (benchmarks) ont été obtenus comme suit :

- Pour le premier terme, REGION × PS, ils sont calculés à partir d'un échantillon EFT 2018 calibré; voir 5.2.1 et Tableau 28;
- Pour le deuxième terme, PROV × (C12 + C13 + C14), ils sont tous égaux à 0, car ce terme concerne les variables de contraste: voir par. 5.2.2 dans [Statbel \(2017a\)](#) ou [Statbel \(2017b\)](#) pour une description détaillée de la construction des variables de contraste C12, C13 et C14, et pour le pourquoi du 0 des totaux de calibrage correspondants;
- Pour le troisième terme, REGION × INK_5_G, ils sont calculés à partir d'une population de ménages de type EFT du RN (complétée par les revenus nets imposables) : voir par. 5.2.2 et Tableau 29.

Le modèle ci-dessus a également été appliqué aux échantillons des sources 1 et 3.1 séparément ; les nombres de ménages répondants dans ces échantillons sont respectivement de 2.506 et 1.752. Les résultats étaient analogues. Au final, le calibrage a été effectué sur l'ensemble de l'échantillon de 4.258 ménages, ce qui reflète également un rapport réaliste entre les sources : aussi bien non pondérée que pondérée avec les poids initiaux, la partie issue de la source 1 représente un petit 60 % du total des sources 1 et 3.1.

Enfin, nous notons que l'échantillon supplémentaire de la source 3.1 peut être considéré comme une extension de l'échantillon de la source 1 : voir annexe Extension d'un échantillon à deux degrés avec un tirage supplémentaire d'USE. Cela fournit une autre justification à la fusion des échantillons des sources 1 et 3.1.

5.3.2. Calibrage de ménages de type EFT d'EBM 2016 (source 2)

La fusion de l'échantillon des 1.642 ménages répondants de la source 2 avec l'échantillon de la source 1 et/ou de la source 3.1 est beaucoup moins évidente que la fusion des échantillons des sources 1 et 3.1. La première raison est que les UPE dans les différents sous-échantillons ont été tirées indépendamment les unes des autres ; une deuxième raison est que les ménages répondants recrutés à partir de la source 2 ont suivi un parcours complètement différent, comme expliqué au paragraphe 2.3. Le calibrage du sous-échantillon EBM 2018 de la source 2 (EBM 2016) a donc été effectué séparément.

Formellement, le modèle de calibrage appliqué aux ménages répondants de la source 2 est le même que pour le calibrage du sous-échantillon des sources 1 et 3.1 prises ensemble (voir paragraphe précédent 5.3.1) ; c'est-à-dire

$$\{HH; REGION \times PS + PROV \times (C12 + C13 + C14) + REGION \times INK_5_G; TrLin\}$$

Les totaux de calibrage ont été calculés exactement de la même manière qu'au par. 5.3.1. Le calcul des poids initiaux pour le calibrage tient compte du cheminement complexe : sélection aléatoire dans les échantillons de l'EFT, à partir de là, recrutement pour l'EBM 2016, puis recrutement pour l'EBM 2018 à partir des ménages répondants à l'EBM 2016 ; voir Tableau 10 au par. 2.3. À cette fin, les probabilités d'(auto)-sélection p_{ij} du Tableau 10 ont simplement été "modélisées" en fonction de la province dans laquelle vit le ménage, le trimestre au cours duquel le ménage a participé à l'EBM en 2016 et le trimestre au cours duquel il est invité à participer à l'EBM en 2018. Le "modèle" est simplement un croisement des trois variables mentionnées, et le calcul de la valeur p_{ij} observée par cellule dans ce croisement ; un modèle plus ingénieux (mais pas nécessairement meilleur) n'a pas été envisagé au moment où le calibrage pouvait être effectué, car l'on ne disposait pas d'autres caractéristiques des ménages.

Comme pour le calibrage des échantillons des sources 1 et 3.1 (prises ensemble), des variantes du modèle de calibrage final mentionné ci-dessus ont été testées pour le calibrage de l'échantillon de la source 2 (cf. par. 5.3.1) ; là encore, ces modèles alternatifs n'ont pas apporté d'amélioration significative.

5.3.3. Calibrage des ménages âgés (source 3.2)

L'échantillon de répondants d'EBM 2018 de ménages privés âgés ne contient que 236 ménages. C'est la raison pour laquelle le modèle de calibrage est également plus limité que les modèles discutés ci-dessus pour le calibrage des sous-échantillons des ménages de type EFT :

$$\{HH; REGION \times PS^* + PROV \times (C12 + C13 + C14); TrLin\}$$

La variable PS^* a peu à voir avec la variable (de post-stratification) PS (qui compte 13 classes) utilisée dans les paragraphes précédents :

- $PS^* = 1$: Ménage d'une personne (single)
- $PS^* = 2$: Ménage de 2 ou plusieurs personnes (pas single)

En forçant un peu, PS^* peut être considéré comme une forte simplification de PS ; d'où le maintien de la notation "PS". Cette simplification est évidemment liée au fait que:

- tous les membres des ménages « âgés » ont au moins 77 ans;
- la majorité des membres sont inactifs;
- la majorité de ces ménages ne comptent que 1 ou 2 membres.

Le revenu du ménage a bien été pris en considération pour l'intégrer dans le calcul via le calibrage. Un problème technique empêche toutefois l'utilisation de cette variable. Dans la population, il y a des ménages âgés qui ne sont pas inclus au registre fiscal (c'est-à-dire sans revenu), mais dans le petit échantillon de 236 ménages âgés, tous les ménages âgés ont un revenu familial (net imposable) qui est inclus au registre fiscal.

Le Tableau 31 ci-dessous montre la répartition de l'échantillon de répondants non pondéré et calibré selon les variables de calibrage. Notez que les sommes de poids en italique et en gras sont des totaux de calibrage correspondant au terme $REGION \times PS^*$ dans le modèle de calibrage (voir aussi Tableau 30; les sommes uniquement en italiques sont des totaux de calibrage implicites en raison du deuxième terme supplémentaire $PROV \times (C12 + C13 + C14)$).

Tableau 31: Echantillon de répondants EBM 2018 de ménages âgés: nombre de répondants et sommes des poids calibrés selon la province, la taille du ménage et le trimestre

Région	Province	Nombres de répondants							Sommes des poids calibrés				
		Trimestre EBM				Taille du ménage (PS*)		Total	par EBM trimestre	Taille du ménage (PS*)		Total	
		T1	T2	T3	T4	Seul	Non Seul			Seul	Pas seul		
BRU	BRU	5	11	6	8	18	12	30	9.017,75	28.907,00	7.164,00	36.071,00	
VLA	ANT	6	1	9	8	11	13	24	12.498,53	40.193,82	9.800,32	49.994,14	
	WVL	8	8	4	5	11	14	25	17.235,15	40.460,16	28.480,43	68.940,59	
	OVL	6	4	8	7	12	13	25	19.430,42	49.769,71	27.951,98	77.721,70	
	LIM	3	6	4	4	8	9	17	8.797,66	21.225,98	13.964,64	35.190,62	
	VLB	6	7	3	5	15	6	21	13.615,74	42.917,32	11.545,63	54.462,95	
	Total	29	26	28	29	57	55	112	71.577,50	194.567,00	91.743,00	286.310,00	
WAL	HAI	7	2	4	11	16	8	24	10.034,07	31.015,68	9.120,59	40.136,27	
	LIE	4	4	6	7	14	7	21	11.956,27	36.979,19	10.845,89	47.825,08	
	LUX	1	3	3	4	9	2	11	2.115,56	7.690,98	771,27	8.462,25	
	NAM	2	5	6	8	11	10	21	5.762,83	14.722,42	8.328,90	23.051,32	
	BWA	8	3	4	2	10	7	17	3.604,02	10.247,73	4.168,36	14.416,08	
	Total	22	17	23	32	60	34	94	33.472,75	100.656,00	33.235,00	133.891,00	
Total		56	54	57	69	135	101	236	114.068,00	324.130,00	132.142,00	456.272,00	

5.3.4. Pourcentage de propriétaires de logements dans la population

L'ajout du terme REGION x INK_5_G était déjà, dans le cadre du calibrage de l'EBM 2016, une manière d'aligner le pourcentage de propriétaires de logements estimé sur la base d'EBM avec l'estimation de SILC, qui s'élève à 68,8 % pour SILC 2018. Cela a rendu l'échantillon calibré d'EBM 2016 plus représentatif. Le Tableau 32 montre que nous y parvenons également dans une large mesure pour EBM 2018: le pourcentage global estimé des propriétaires de logements s'élève à 69,35 %. Cette estimation s'obtient de la façon suivante (voir également Exemple 3 au par. 5.4.3).

Les sources 1+3.1 et 2 sont toutes deux considérées comme représentant la population de 4.465.140 ménages de type EFT. Par conséquent, les pourcentages de propriétaires de logements de 66,65 % et 73,68 % respectivement sont des estimations du même paramètre. Pour obtenir une moyenne, nous utilisons les nombres respectifs de répondants comme poids:

$$\frac{66,65\% \times 4258 + 73,68\% \times 1642}{4258 + 1642} = 66,65\% \times 0,7217 + 73,68\% \times 0,2783 = 68,61\%$$

Pour obtenir une estimation pour toute la population de 4.921.412 ménages privés, nous travaillons avec les nombres absolus estimés de propriétaires de logements:

$$\frac{68,61\% \times 4.465.140 + 76,62\% \times 456.272}{4.921.412} = \frac{3.063.532,554 + 349.595,6064}{4.921.412} = \frac{3.413.128,1604}{4.921.412} = 69,35\%$$

Cette méthode est appliquée pour l'estimation de tous les paramètres similaires (nombres et proportions/pourcentages).

Tableau 32: Estimation du nombre de propriétaires de logements (EBM 2018)

Source	Nombre MN répondants	Population de MN (total de calibrage)	Estimation des Propriétaires de logements	
			Nombre	%
MN de type EFT des sources 1 (EFT) + 3.1 (RN)	4.258 (=2.506 + 1.752)	4.465.140	2.976.031,8	66,65%
MN de type EFT de la source 2 (EBM 2016)	1.642	4.465.140	3.290.049,2	73,68%
MN âgés de la source 3.2 (RN)	236	456.272	349.607,6	76,62%
Global	6.136	4.921.414	3.413.030,5	69,35%
Global sans la source 3.2	5.900	4.465.140	3.063.422,9	68,61%

Le Tableau 32 montre que l'effet des répondants de la source 3.2 (ménages privés âgés) est relativement faible. À cet égard, il convient de noter que l'échantillon de 236 ménages âgés répondants se compose principalement de ménages ayant un revenu supérieur (qui sont plus susceptibles d'être propriétaires).

Nous constatons que l'estimation du pourcentage de propriétaires de logements s'élève à 73,68 % si nous nous limitons aux ménages répondants de la source 2 (EBM 2016). Nous supposons qu'il s'agit d'un effet cumulé de l'autosélection: dans EBM 2016, le nombre de revenus plus élevés (et, par conséquent, le nombre de propriétaires de logements) était déjà surreprésenté, ce qui a entraîné une correction via le terme REGION x INK_5_G dans le modèle de calibrage. Et cette réponse sélective se prolonge d'EBM 2016 à EBM 2018.

À cet égard, il convient de noter que les sources 1+3.1 et 2 doivent également contenir des ménages 'âgés': initialement, c-à-d lors du tirage des échantillons EFT, tous ces ménages étaient de type EFT, mais depuis ce ou ces tirages initiaux, beaucoup de temps a passé et certains de ces ménages ont 'vieilli' si bien qu'ils pourraient être considérés comme des ménages 'âgés' lors de la collecte de données pour EBM 2018.

5.4. Estimation des paramètres

5.4.1. Variables d'intérêt, paramètres et poids

L'objectif (principal) de l'EBM est l'étude des dépenses (pour un très large éventail de catégories de produits ou de catégories ECOICOP) effectuées par la population belge de ménages privés. Les paramètres de telles variables, comme les dépenses annuelles totales effectuées par la population belge et les dépenses annuelles moyennes par ménage de la population belge, doivent être estimés sur la base de l'EBM. Outre les dépenses annuelles, nous nous intéressons également aux variables telles que le fait d'effectuer ou non une certaine dépense (au moins une fois dans l'année), avec comme paramètres correspondants le nombre et la proportion de ménages belges qui effectuent cette dépense. Enfin, nous observons également des variables connexes et estimons leurs paramètres. La quantification de chaque variable d'intérêt possible permet une représentation simple, mais malgré tout mathématiquement correcte, des variables d'intérêts, de leurs paramètres et de leurs estimations.

Supposons que la population belge totale contient N ménages privés, et que le nombre total de ménages répondants est égal à n . Nous savons de ce qui précède que $N = 4.921.414$ et $n = 6.136$. Nous utilisons l'indice i pour identifier les ménages dans la population. Pour indiquer que i varie sur l'ensemble de la population, nous écrivons: $i = 1, \dots, N$. Mais nous pouvons également utiliser l'indice i pour identifier les ménages dans l'échantillon de répondants, et $i = 1, \dots, n$ indique alors que i varie sur l'échantillon de répondants.

Par ailleurs, nous notons x_i pour la valeur (toujours sur base annuelle!) d'une variable numérique x pour un ménage i (dans la population ou dans l'échantillon de répondants). Les paramètres que nous voulons estimer pour la variable x sont la population totale $X = \sum_{i=1}^N x_i$ et la moyenne (par ménage) $\bar{X} = \frac{\sum_{i=1}^N x_i}{N} = X/N$.

En outre, en raison du calibrage nécessaire par source (ou de la fusion de deux sources) soit selon la sous-population des ménages privés belges de type EFT, soit selon la sous-population des ménages privés belges âgés, il est utile d'introduire les symboles et grandeurs suivants :

- $N_1 = 4.465.140$ = Le nombre de ménages privés belges de type EFT dans la population;
- $N_2 = 456.272$ = Le nombre de ménages privés belges âgés dans la population;
- $n_{11} = 4.258$ = Le nombre de ménages répondants de type EFT des sources 1+3.1;
- $n_{12} = 1.642$ = Le nombre de ménages répondants de type EFT de la source 2;
- $n_1 = n_{11} + n_{12} = 5.900$ = Le nombre de ménages répondants de type EFT (des sources 1+3.1 ou 2);
- $n_2 = 236$ = Le nombre de ménages privés belges âgés répondants (de la source 3.2).

Ensuite, naturellement:

- $n = n_1 + n_2 = n_{11} + n_{12} + n_2 = 6.136$
- $N = N_1 + N_2 = 4.921.414$

Les paramètres de la variable x peuvent être définis par source et type de ménage:

- $X_1 = \sum_{i=1}^{N_1} x_i$ et $\bar{X}_1 = \frac{\sum_{i=1}^{N_1} x_i}{N_1} = X_1/N_1$, pour le total et la moyenne dans la sous-population de ménages privés belges de type EFT;

- $X_2 = \sum_{i=1}^{N_2} x_i$ et $\bar{X}_2 = \frac{\sum_{i=1}^{N_2} x_i}{N_2} = X_2/N_2$, pour le total et la moyenne dans la sous-population des ménages privés belges âgés.

Les estimations des totaux X , X_1 et X_2 et des moyennes \bar{X} , \bar{X}_1 et \bar{X}_2 utilisent uniquement les observations x_i ($i = 1, \dots, n$) dans l'échantillon de répondants. Par ailleurs, afin de calculer de bonnes estimations, nous avons besoin de poids qui extrapolent les observations selon la population. Nous pouvons considérer deux poids pour chaque ménage répondant:

- Le *poids calibré initial* w_i pour le ménage $i = 1, \dots, n$, qui résulte des calibrages par source (1+3.1, ou 2, ou 3.2) selon la sous-population de ménages correspondante. Ce sont les poids calibrés qui ont été abordés aux par. 5.3.1, 5.3.2 et 5.3.3. Ces poids permettent de calculer séparément les estimations basées sur les trois sources 1+3.1, 2 et 3.2.
- Un *poids calibré ajusté* \tilde{w}_i pour le ménage $i = 1, \dots, n$, qui est présenté au par. 5.4.2 suivant, est introduit. Avec ces pondérations, nous pourrions calculer des estimations sur la base des sources 1+3.1 et 2 ensemble, ou sur les trois sources 1+3.1, 2 et 3.2 ensemble.

5.4.2. Estimation des totaux et des moyennes

Comme déjà indiqué au par. 5.4.1, les variables de dépenses sont à peu près les variables d'intérêt les plus importantes dans l'EBM. De nombreuses dépenses sont notées dans le carnet (ménage): pour chaque ménage répondant $i = 1, \dots, n$, nous trouvons dans le carnet la dépense x_i^0 , p.ex. pendant la période de référence (d'un demi-mois) à laquelle le ménage i est affecté, pour tout produit ou catégorie ECOICOP P , par exemple. Notez que si le ménage i n'a pas enregistré de dépense pour P dans le carnet, $x_i^0 = 0$ sera automatiquement inscrit. Ces dépenses x_i^0 observées sont ramenées sur un an en les multipliant par 24, c-à-d $x_i = x_i^0 \times 24$. (Une justification de cette transformation est donnée au par. 7.2) Avec les poids calibrés w_i ($i = 1, \dots, n$), nous obtenons ensuite diverses estimations pour les totaux et les moyennes de la variable de dépense x , soit la dépense annuelle par ménage pour P .

- $\hat{X}_{11} = \sum_{i=1}^{n_{11}} w_i x_i$ = la dépense annuelle totale estimée pour P dans la population de ménages de type EFT, sur la base des ménages répondants de la source 1+3.1. La dépense annuelle moyenne pour P par ménage dans la population de ménages de type EFT est ensuite estimée comme $\hat{\bar{X}}_{11} = (\sum_{i=1}^{n_{11}} w_i x_i) / N_1 = \hat{X}_{11} / N_1$, sur la base des ménages répondants de la source 1+3.1.
- $\hat{X}_{12} = \sum_{i=1}^{n_{12}} w_i x_i$ = la dépense annuelle totale estimée pour P dans la population de ménages de type EFT, sur la base des ménages répondants de la source 2. La dépense annuelle moyenne de P par ménage dans la population de ménages de type EFT est estimée comme $\hat{\bar{X}}_{12} = (\sum_{i=1}^{n_{12}} w_i x_i) / N_1 = \hat{X}_{12} / N_1$, sur la base des ménages répondants de la source 2.
- $\hat{X}_2 = \sum_{i=1}^{n_2} w_i x_i$ = la dépense annuelle totale estimée pour P dans la population de ménages âgés, sur la base des ménages répondants de la source 3.2. La dépense annuelle moyenne pour P par ménage dans la population de ménages âgés est ensuite estimée comme $\hat{\bar{X}}_2 = (\sum_{i=1}^{n_2} w_i x_i) / N_2 = \hat{X}_2 / N_2$, sur la base des ménages répondants de la source 3.2.

Notez également que $N_1 = \hat{N}_1 = \sum_{i=1}^{n_{11}} w_i = \sum_{i=1}^{n_{12}} w_i$ et $N_2 = \hat{N}_2 = \sum_{i=1}^{n_2} w_i$, et donc aussi $N = N_1 + N_2 = \hat{N}_1 + \hat{N}_2$. Il en ressort que les chiffres de la population N_1 , N_2 et N peuvent être retrouvés sous la forme d'estimations, à savoir des sommes des poids calibrés w_i ; le fait que ces estimations soient exactement égales aux chiffres de la population découle du fait que les chiffres de la population sont des totaux de calibrage ou benchmarks.

Bien entendu, le but ne peut être d'utiliser ou de publier deux estimations, \hat{X}_{11} et \hat{X}_{12} , pour la dépense annuelle totale X_1 pour P dans la population de ménages de type EFT. C'est pourquoi nous combinons ces estimations en une seule, comme suit:

$$\hat{X}_1 = \alpha \hat{X}_{11} + (1 - \alpha) \hat{X}_{12}$$

avec $0 < \alpha < 1$. Un choix naturel pour les facteurs de pondération α et $1 - \alpha$ sont les parts relatives dans l'échantillon total des ménages répondants de type EFT des sous-échantillons des sources 1+3.1 et 2, c'est-à-dire

$$\alpha = \frac{n_{11}}{n_{11} + n_{12}} = \frac{4.258}{4.258 + 1.642} = 0,7217$$

$$1 - \alpha = \frac{n_{12}}{n_{11} + n_{12}} = \frac{1.642}{4.258 + 1.642} = 0,2783$$

Les dépenses annuelles totales moyennes estimées pour P dans la population de ménages de type EFT est alors:

$$\hat{\bar{X}}_1 = \frac{\hat{X}_1}{N_1} = \frac{\alpha \hat{X}_{11} + (1 - \alpha) \hat{X}_{12}}{N_1} = \alpha \frac{\hat{X}_{11}}{N_1} + (1 - \alpha) \frac{\hat{X}_{12}}{N_1} = \alpha \hat{\bar{X}}_{11} + (1 - \alpha) \hat{\bar{X}}_{12}$$

ce qui montre que l'estimation $\hat{\bar{X}}_1$ est également une combinaison de deux estimations $\hat{\bar{X}}_{11}$ et $\hat{\bar{X}}_{12}$ pour le même paramètre \bar{X}_1 .

Nous pouvons réécrire l'expression de \hat{X}_1 comme suit:

$$\begin{aligned} \hat{X}_1 &= \alpha \hat{X}_{11} + (1 - \alpha) \hat{X}_{12} \\ &= \alpha \sum_{i=1}^{n_{11}} w_i x_i + (1 - \alpha) \sum_{i=1}^{n_{12}} w_i x_i \\ &= \sum_{i=1}^{n_{11}} \alpha w_i x_i + \sum_{i=1}^{n_{12}} (1 - \alpha) w_i x_i \\ &= \sum_{i=1}^{n_1} \tilde{w}_i x_i \end{aligned}$$

où \tilde{w}_i est défini comme suit:

$$\begin{aligned} \tilde{w}_i &= \alpha w_i && \text{voor elk responderend huishouden } i \text{ uit bron 1+3.1} \\ &= (1 - \alpha) w_i && \text{voor elk responderend huishouden } i \text{ uit bron 2} \end{aligned}$$

Nous appelons ce \tilde{w}_i le *poids calibré ajusté* pour le ménage répondant de type EFT $i = 1, \dots, n_1 = n_{11} + n_{12}$. Un poids calibré ajusté peut également être défini pour chaque ménage âgé répondant de la source 3.2, simplement comme:

$$\tilde{w}_i = w_i \quad (i = 1, \dots, n_2)$$

L'estimation des dépenses annuelles totales pour P pour tous les ménages privés belges est alors:

$$\begin{aligned} \hat{X} &= \sum_{i=1}^n \tilde{w}_i x_i \\ &= \sum_{i=1}^{n_1} \tilde{w}_i x_i + \sum_{i=1}^{n_2} \tilde{w}_i x_i = \sum_{i=1}^{n_1} \tilde{w}_i x_i + \sum_{i=1}^{n_2} w_i x_i = \hat{X}_1 + \hat{X}_2 \\ &= \alpha \hat{X}_{11} + (1 - \alpha) \hat{X}_{12} + \hat{X}_2 \end{aligned}$$

La dernière expression de \hat{X} permet de voir comment les estimations \hat{X}_{11} , \hat{X}_{12} et \hat{X}_2 sont combinées sur la base des sous-échantillons indépendants issus de trois sources distinctes en l'estimateur global pour X . La première expression montre comment cette estimation peut être écrite comme un total pondéré sur toutes les observations (transformées sur base annuelle) x_i , en utilisant les poids calibrés ajustés \tilde{w}_i .

5.4.2.1. Exemple 1

Imaginons que nous sommes intéressés par les dépenses totales et moyennes par ménage effectuées pour des 'articles d'habillement et chaussures' (catégorie ECOICOP 03) dans la population belge (de ménages privés). Chaque ménage répondant note pendant la période de référence d'un demi-mois les diverses dépenses effectuées pour des articles d'habillement et des chaussures dans le carnet, et ces dépenses sont additionnées pour obtenir x_i^0 . Si un ménage n'a pas noté de dépenses pour des articles d'habillement, alors $x_i^0 = 0$. Nous transformons ensuite ces dépenses x_i^0 observées directement en une dépense x_i sur base annuelle: $x_i = x_i^0 \times 24$. Diverses estimations peuvent être établies, comme le montre le Tableau 33. Notez que dans ce tableau, nous appelons une grandeur mathématique telle que \hat{X} un 'estimateur', alors que la valeur numérique d'une telle grandeur est appelée une 'estimation'.

Tableau 33: Estimations concernant les dépenses pour les 'Articles d'habillement et chaussures' (ECOICOP 03) – EBM 2018

Paramètre	Estimateur	Estimation (en €)	Description
X_1	\hat{X}_{11}	8.014.715.172,45	Dépense annuelle totale estimée pour 'Articles d'habillement et chaussures' dans la population de ménages de type EFT, sur la base de la source 1+3.1
	\hat{X}_{12}	7.196.622.416,33	Dépense annuelle totale estimée pour 'Articles d'habillement et chaussures' dans la population de ménages de type EFT, sur la base de la source 2
	\hat{X}_1	7.787.039.958,42	Dépense annuelle totale estimée pour 'Articles d'habillement et chaussures' dans la population de ménages de type EFT, sur la base des sources 1+3.1 et 2 ensemble
X_2	\hat{X}_2	376.191.612,35	Dépense annuelle totale estimée pour 'Articles d'habillement et chaussures' dans la population de ménages âgés, sur la base de la source 3.2
X	\hat{X}	8.163.231.570,78	Dépense annuelle totale estimée pour 'Articles d'habillement et chaussures' dans l'ensemble de la population de ménages
N	\hat{N}	4.921.414,00	Nombre de ménages dans l'ensemble de la population
\bar{X}	$\hat{\bar{X}}$	1.658,72	Dépense annuelle moyenne estimée par ménage pour 'Articles d'habillement et chaussures' dans l'ensemble de la population

On vérifie facilement les liens suivants (à l'exception des erreurs d'arrondi):

- $0,7217 \times 8.014.715.172,45 + 0,2783 \times 7.196.622.416,33 = 7.787.039.958,42$
- $7.787.039.958,42 + 376.191.612,35 = 8.163.231.570,78$
- $8.163.231.570,78 / 4.921.414 = 1.658,72$

Dans les tableaux Excel publiés, nous trouvons les résultats de la catégorie 'Articles d'habillement et produits alimentaires' tels que montrés dans le Tableau 34. Le résultat grisé de la colonne (4) correspond précisément au résultat obtenu ci-dessus pour $\hat{\bar{X}}$. Les résultats des colonnes (3), (5) et (6) seront abordés ci-après dans les exemples 4 et 5.

Tableau 34: Extraction de la ligne pour 'Articles d'habillement et produits alimentaires' dans le tableau publié HBS_Tab01_BE_2018^{xi}

Catégorie de produits		Nombre de ménages (échantillon)	Nombre de ménages (population)	Dépenses moyennes pour l'ensemble des ménages (par an en euro)	Dépenses moyennes par personne et par an (€)	Dépenses moyennes par unité de consommation et par an (€)
(1)		(2)	(3)	(4)	(5)	(6)
03	ARTICLES D'HABILLEMENT ET CHAUSSURES	3815	2948118,15	1658,72	738,32	1079,27

5.4.2.2. Exemple 2

Un autre type de variables (que les dépenses notées dans le carnet) sont les variables dont la valeur est notée dans le questionnaire (ménage). Un exemple est le nombre x de voitures que possède chaque ménage dans la population étudiée. Les valeurs x_i pour les ménages répondants sont alors trouvées directement dans le questionnaire (et ne sont naturellement pas la conséquence d'une transformation sur base annuelle d'une observation sous-jacente!). Les estimations de cette variable sont présentées au Tableau 35.

Tableau 35: Estimations concernant le nombre de voitures que possède chaque ménage – EBM 2018

Paramètre	Estimateur	Estimation	Description
X_1	\hat{X}_{11}	5.031.917,7	Nombre total estimé de voitures dans la population de ménages de type EFT, sur la base de la source 1+3.1
	\hat{X}_{12}	5.005.602,1	Nombre total estimé de voitures dans la population de ménages de type EFT, sur la base de la source 2
	\hat{X}_1	5.024.594,0	Nombre total estimé de voitures dans la population de ménages de type EFT, sur la base des sources 1+3.1 et 2 ensemble
X_2	\hat{X}_2	350.046,3	Nombre total estimé de voitures dans la population de ménages âgés, sur la base de la source 3.2
X	\hat{X}	5.374.640,4	Nombre total estimé de voitures dans l'ensemble de la population de ménages
N	\hat{N}	4.921.414,0	Nombre de ménages dans l'ensemble de la population
\bar{X}	$\hat{\bar{X}}$	1,0921	Nombre moyen estimé de voitures par ménage dans la population totale

Le lecteur peut facilement vérifier les liens sous-jacents entre les estimations. Notez que nous avons arrondi les nombres estimés (sauf le chiffre de la population N) à la 1^{ère} décimale, alors que la moyenne est arrondie à la 4^e décimale. L'objectif est d'indiquer que les estimations pour les nombres (c'est-à-dire les variables entières) sont aussi généralement des nombres décimaux. Quelques autres estimations pourront également intéresser le lecteur, à savoir $\hat{X}_1 = \frac{5.024.594,0}{4.465.140} = 1,12$ et $\hat{X}_2 = \frac{350.046,3}{456.272} = 0,77$; les ménages âgés possèdent donc en moyenne moins de voitures que les ménages de type EFT.

5.4.3. Estimation des nombres et proportions de ménages ayant une caractéristique spécifique

Le cadre formel élaboré au par. 5.4.2 pour les variables numériques générales (et illustré pour deux variables spécifiques) et pour lequel les définitions et la notation nécessaires ont été introduites au par. 5.4.1 est également applicable pour estimer le nombre et les proportions de ménages qui présentent une caractéristique ou un trait particulier E . La possession d'une caractéristique donnée E peut en effet être quantifiée comme suit:

$$\begin{aligned} x_i &= 1 && \text{als het huishouden } i \text{ het kenmerk } E \text{ bezit} \\ &= 0 && \text{als het huishouden } i \text{ het kenmerk } E \text{ niet bezit} \end{aligned}$$

Cette variable x est appelée variable 0/1 ou dummy. Il s'agit d'une variable numérique (qui quantifie la possession d'une caractéristique E), pour laquelle toutes les formules des paragraphes précédents sont valables. L'interprétation des paramètres tels que X et \bar{X} , et leurs estimations respectives \hat{X} et $\hat{\bar{X}}$, doivent naturellement être adaptées:

- X est le nombre de ménages avec la caractéristique E dans l'ensemble de la population; \hat{X} est le nombre estimé de ménages avec la caractéristique E dans l'ensemble de la population;
- \bar{X} est la proportion de ménages avec la caractéristique E dans l'ensemble de la population; $\hat{\bar{X}}$ est la proportion estimée de ménages avec la caractéristique E dans l'ensemble de la population.

Naturellement, on peut également définir ces paramètres pour la sous-population de ménages de types EFT et pour la sous-population de ménages âgés.

5.4.3.1. Exemple 3

Le questionnaire ménage demande le statut de propriété en ce qui concerne l'habitation qui abrite le ménage (c-à-d la résidence principale); nous distinguons les propriétaires (y compris les co-propriétaires et les usufruitiers) des non-propriétaires (locataires et occupants à titre gratuit), et nous quantifions cette caractéristique comme suit:

$$\begin{aligned} x_i &= 1 && \text{als huishouden } i \text{ eigenaar is van de woning die het betreft} \\ &= 0 && \text{als huishouden } i \text{ niet de eigenaar is van de woning die het betreft} \end{aligned}$$

Notez qu'une estimation telle que \hat{X} est alors une somme des poids calibrés ajustés:

$$\hat{X} = \sum_{i=1}^n \tilde{w}_i x_i = \sum_{i \text{ is eigenaar}} \tilde{w}_i \times 1 + \sum_{i \text{ is niet-eigenaar}} \tilde{w}_i \times 0 = \sum_{i \text{ is eigenaar}} \tilde{w}_i$$

Tableau 36: Estimations concernant le statut de propriété de l'habitation – EBM 2018

Paramètre	Estimateur	Estimation	Description
X_1	\hat{X}_{11}	2.976.031,8	Nombre estimé de ménages de type EFT qui sont propriétaires de l'habitation concernée, sur la base de la source 1+3.1
	\hat{X}_{12}	3.290.049,2	Nombre estimé de ménages de type EFT qui sont propriétaires de l'habitation concernée, sur la base de la source 2
	\hat{X}_1	3.063.422,9	Nombre estimé de ménages de type EFT qui sont propriétaires de l'habitation concernée, sur la base des sources 1+3.1 et 2 ensemble
X_2	\hat{X}_2	349.607,6	Nombre estimé de ménages âgés qui sont propriétaires de l'habitation concernée, sur la base de la source 3.2
X	\hat{X}	3.413.030,5	Nombre estimé de ménages qui sont propriétaires de l'habitation concernée dans l'ensemble de la population de ménages
N	\hat{N}	4.921.414,0	Nombre de ménages dans l'ensemble de la population
\bar{X}	$\hat{\bar{X}}$	69,35%	Proportion estimée de ménages qui sont propriétaires de l'habitation concernée dans l'ensemble de la population de ménages

Les liens entre les diverses estimations du Tableau 36 peuvent être facilement vérifiés.

5.4.3.2. Exemple 4 (suite de l'exemple 1 au par. 5.4.2)

Il peut aussi être intéressant d'estimer combien de ménages privés belges effectuent des dépenses pour les 'Articles d'habillement et chaussures' (catégorie ECOICOP 03). Avec x_i^0 comme dans l'exemple 1, nous définissons:

$$\begin{aligned}
 x_i &= 1 && \text{als huishouden } i \text{ uitgaven aan "Kleding en schoenen" rapporteert, m.a.w. als } x_i^0 > 0 \\
 &= 0 && \text{als huishouden } i \text{ geen uitgaven aan "Kleding en schoenen" rapporteert, m.a.w. als } x_i^0 = 0
 \end{aligned}$$

Les estimations de paramètres basées sur cette variable 0/1 sont rassemblées dans le Tableau 37.

Tableau 37: Estimations concernant les dépenses effectuées ou non pour des 'Articles d'habillement et chaussures' (ECOICOP 03) – EBM 2018

Paramètre	Estimateur	Estimation	Description
X_1	\hat{X}_{11}	2.779.350,1	Nombre estimé de ménages de type EFT qui effectuent des dépenses pour des 'Articles d'habillement et chaussures', sur la base de la source 1+3.1
	\hat{X}_{12}	2.706.667,5	Nombre estimé de ménages de type EFT qui effectuent des dépenses pour des 'Articles d'habillement et chaussures', sur la base de la source 2
	\hat{X}_1	2.759.122,6	Nombre estimé de ménages de type EFT qui effectuent des dépenses pour des 'Articles d'habillement et chaussures', sur la base des sources 1+3.1 et 2 ensemble
X_2	\hat{X}_2	188.995,6	Nombre estimé de ménages âgés qui effectuent des dépenses pour des 'Articles d'habillement et chaussures', sur la base de la source 3.2

X	\hat{X}	2.948.118,1	Nombre estimé de ménages qui effectuent des dépenses pour des 'Articles d'habillement et chaussures' dans l'ensemble de la population de ménages
N	\hat{N}	4.921.414,0	Nombre de ménages dans l'ensemble de la population
\bar{X}	$\hat{\bar{X}}$	59,90%	Proportion estimée de ménages qui effectuent des dépenses pour des 'Articles d'habillement et chaussures' dans l'ensemble de la population de ménages

Le résultat de la colonne (3) du Tableau 34 correspond précisément au résultat obtenu ci-dessus pour \hat{X} .

5.4.4. Estimations des domaines

Une étape supplémentaire mène aux paramètres des *domaines* ou *sous-populations* des ménages. Sont des exemples de domaines:

- ménages de la Région de Bruxelles-Capitale; ...
- ménages avec enfants, isolés, ...
- familles monoparentales en Région flamande; ...

Pour chacun de ces domaines, nous voulons éventuellement calculer une estimation, par exemple:

- du nombre et de la proportion de propriétaires d'habitation; ou de manière générale, du nombre et du pourcentage de ménages ayant la caractéristique E ;
- Des dépenses totales et moyennes par ménage pour la catégorie de produits P ; ou de manière générale: du total et de la moyenne d'une variable numérique x .

Nous pouvons noter les estimations des totaux (en particulier des nombres) et des moyennes (en particulier des proportions) d'une variable x pour un domaine D comme \hat{X}_D et $\hat{\bar{X}}_D$, respectivement. En utilisant la notation $i \in D$, pour indiquer chaque ménage répondant aléatoire dans le domaine D , alors:

$$\hat{X}_D = \sum_{i \in D} \tilde{w}_i x_i$$

$$\hat{\bar{X}}_D = \frac{\hat{X}_D}{\hat{N}_D} = \frac{\sum_{i \in D} \tilde{w}_i x_i}{\sum_{i \in D} \tilde{w}_i}$$

Notez que le dénominateur \hat{N}_D sera la plupart du temps une estimation du nombre de ménages dans le domaine D . Cela joue un rôle important dans l'estimation de la variance du ratio $\hat{\bar{X}}_D$.

Enfin, nous pouvons encore remarquer que la sous-population de ménages de type EFT et la sous-population de ménages âgés peuvent être traités comme des domaines. Dans ce cas, les dénominateurs \hat{N}_D sont les chiffres connus de la population, à savoir les nombres N_1 et N_2 , respectivement, tels que définis et utilisés dans les paragraphes précédents.

5.4.5. Moyennes alternatives: par personne et par unité de consommation modifiée

Les moyennes dont nous avons parlé aux paragraphes 5.4.2, 5.4.3 et 5.4.4 précédents sont toutes par ménage. Les dénominateurs des formules pour \hat{X} sont soit un nombre connu (du RN) des ménages privés dans la population (N), soit un nombre de ménages estimé sur la base de l'EBM 2018 (\hat{N}).

De tels dénominateurs peuvent facilement être remplacés par:

- Un nombre connu ou estimé de personnes (N_I ou \hat{N}_I) qui appartiennent aux ménages privés belges;
- Un nombre connu ou estimé d'UCM (C ou \hat{C}).

Ces grandeurs peuvent être obtenues comme suit:

- $N_I = \sum_{i=1}^N a_i$, si a_i est le nombre de membres du ménage $i = 1, \dots, N$ dans la population (selon le RN);
- $\hat{N}_I = \sum_{i=1}^n \tilde{w}_i a_i$, si a_i est le nombre de membres de ménages répondants $i = 1, \dots, n$ dans l'EBM 2018;
- $C = \sum_{i=1}^N c_i$, si c_i est le nombre d'UCM du ménage $i = 1, \dots, N$ dans la population (selon le RN);
- $\hat{C} = \sum_{i=1}^n \tilde{w}_i c_i$, si c_i est l'UCM du ménage répondant $i = 1, \dots, n$ dans l'EBM 2018.

Notez que $\hat{N} \leq \hat{C} \leq \hat{N}_I$ (toujours), de sorte que toujours:

$$\frac{\hat{X}}{\hat{N}} \geq \frac{\hat{X}}{\hat{C}} \geq \frac{\hat{X}}{\hat{N}_I}$$

où les estimations dans les dénominateurs éventuels peuvent être remplacés par les chiffres exacts de la population (N , C et N_I , respectivement).

5.4.5.1. Exemple 5 (suite de l'exemple 1 au par. 5.4.2 et de l'exemple 4 au par. 5.4.3)

Pour l'EBM 2018:

- $\hat{N}_I = 11.056.446,09$
- $\hat{C} = 7.563.676,67$

Par conséquent, avec $\hat{X} = 8.163.231.570,78$ du Tableau 34.

- Résultat dans la colonne (5) du Tableau 34:

$$\frac{\hat{X}}{\hat{N}_I} = \frac{8.163.231.570,78}{11.056.446,09} = 738,32$$

- Résultat dans la colonne (6) du Tableau 34:

$$\frac{\hat{X}}{\hat{C}} = \frac{8.163.231.570,78}{7.563.676,67} = 1079,27$$

5.5. Estimation de la variance

5.5.1. *Introduction*

L'estimation de la variance doit prendre en compte au mieux toutes les étapes de l'enquête :

- Le plan de sondage, et le tirage de l'échantillon. La théorie des sondages définit pour chaque étape du plan de sondage des formules pour estimer la variance de l'estimateur d'un total. Quand un plan de sondage est à plusieurs degrés, les formules se combinent pour tenir compte de chaque degré du plan de sondage.
- La non-réponse et la manière dont elle est corrigée. La non-réponse est une étape de l'enquête non contrôlée par le statisticien. Dans le cadre de l'estimation de la variance, la non-réponse est modélisée par une seconde phase du plan de sondage, la première phase étant le plan de sondage déterminé pour l'enquête. La probabilité de réponse est par définition inconnue. On doit donc l'estimer. Dans l'enquête HBS, la non-réponse n'est pas corrigée explicitement. On va donc estimer la probabilité de répondre par le rapport entre le poids final (poids après calage) et le poids de sondage. Le processus de non-réponse est alors modélisé comme un tirage de Poisson, la probabilité de tirage étant estimée par la probabilité de réponse définie ci-dessus.
- Le calage. Le calage est l'étape qui consiste à modifier les poids de sondage (ou les poids corrigés pour la non-réponse) de sorte que les totaux de certaines variables connues sur l'univers entier soient estimés exactement sur base de l'échantillon et des poids adaptés, et ce tout en conservant les poids les plus proches possible des poids initiaux. Pour l'estimation de la variance de l'estimateur d'un total par calage, l'estimateur de la variance s'obtient en substituant la variable d'intérêt par le résidu de la régression linéaire de cette variable sur les variables de calage.
- L'estimateur utilisé. Les formules habituelles d'estimation de la variance sont définies pour des estimations de totaux sur l'univers. Quand le paramètre estimé est un ratio de totaux, on substitue à la variable d'intérêt une variable que l'on dit « linéarisée ». L'estimation du total de la variable linéarisée a asymptotiquement la même variance que le ratio auquel on s'intéresse.

Le calcul de la variance est réalisé par le logiciel POULPE. POULPE est un ensemble de macro programmées sous SAS par l'INSEE. Il permet de calculer la variance pour différents plans de sondage, et de prendre en compte la non-réponse et le calage. POULPE permet d'estimer la variance pour des totaux, mais offre aussi une solution pour les variables calculées à partir de totaux (par exemple les moyennes, mais aussi la différence de deux totaux...). POULPE permet donc d'estimer la variance en prenant en compte toutes les étapes du traitement de l'enquête.

Le plan de sondage de l'enquête HBS est complexe. La combinaison de diverses sources pour l'échantillon a conduit à un estimateur complexe (calage des différentes sources, suivi de la combinaison des différentes sources pour obtenir un estimateur final). Cette complexité a conduit à devoir simplifier la modélisation du processus pour rendre l'estimation possible.

5.5.2. Description du plan de sondage

L'échantillon de HBS est constitué de quatre sources (voir les détails dans la partie 2) de sélection des ménages. Le plan de sondage, pour chacune de ces sources, est le suivant :

- Premier degré : sélection (éventuellement avec répétition) selon un tirage systématique proportionnel à la taille (SYSPPSWR, *systematic probability proportional to size with replacement*) d'unités primaires (UP) (les unités primaires sont des unités géographiques). La taille d'une unité primaire est le nombre de ménages éligibles pour l'enquête LFS dans l'unité primaire, selon le Registre national.
- Second degré : sélection par un sondage aléatoire simple (SRS) d'un nombre fixe de ménages dans les unités sélectionnées au premier degré.

En raison du tirage systématique au premier degré, il n'existe pas d'estimateur non biaisé de la variance pour ce type de plan de sondage. Ceci découle du fait que certaines paires d'unités primaires voisines (les unités primaires de taille inférieure au pas de tirage) n'ont aucune chance d'être tirées en même temps. Les probabilités d'inclusion doubles sont donc nulles pour certaines unités primaires, et donc on ne peut pas définir un estimateur non biaisé de la variance.

POULPE propose une solution approximative, pour les tirages systématiques à probabilité égale, sans répétition. Cette solution n'est pas applicable ici, en raison du tirage avec répétition, et des probabilités de tirage inégales.

Dans le cadre de l'estimation de la variance de l'enquête LFS, nous avons pu montrer par simulation que dans le cadre de ce plan de sondage (SYSPPSWR suivi d'un SRS de taille fixe par UP sélectionnée) la variance pouvait être estimée de manière satisfaisante en considérant que chaque groupe est le résultat d'un sondage aléatoire simple de taille égale à la taille du groupe, dans une strate de taille égale à la taille du pas de tirage du tirage systématique des unités primaires. Ceci fournit une solution pour estimer la variance due au plan de sondage. Il suffit pour chaque groupe de déterminer la taille de l'échantillon (connu) et la taille du pas, utilisé comme taille de la strate, ces deux éléments étant suffisants pour estimer la variance d'un total dans le cas d'un SRS stratifié.

5.5.2.1. Sources 1 et 3.1

Les ménages sélectionnés pour les sources 1 et 3.1 sont issus des mêmes unités primaires sélectionnées pour l'enquête LFS. Les ménages de la source 1 sont les ménages sélectionnés pour LFS et qui ont accepté de participer à HBS. Les ménages de la source 3.1 sont les ménages sélectionnés dans les mêmes UP. Ces ménages ont été sollicités par courrier pour participer à LFS. Comme ces ménages sont issus des mêmes UP, ils ont été traités ensemble au niveau du calage (après détermination d'un poids avant calage, intégrant en partie le processus de non-réponse lié à l'enquête LFS ; voir 5.3.1).

Pour les groupes de la source combinée, on a donc comme taille de la strate la taille du pas de sélection des UP, et comme taille de l'échantillon le nombre de ménage tirés à l'origine pour l'enquête LFS plus le nombre de ménages tirés pour l'échantillon supplémentaire.

5.5.2.2. Source 2

Les ménages sélectionnés pour la source 2 sont d'anciens ménages ayant répondu en 2016 à l'enquête HBS. Ces ménages ont donc été tirés à l'origine pour l'enquête LFS, ont ensuite participé à l'enquête HBS 2016, et ont accepté de répondre à nouveau à HBS 2018.

Comme pour la source 1, la modélisation du plan de sondage devrait être une strate par groupe tiré à l'origine pour LFS, la taille de la strate égale au pas de tirage utilisé à l'origine, et la taille de l'échantillon égale à la taille du groupe tiré à l'origine pour LFS.

Cependant, cette stratégie ne peut être appliquée. En effet, l'estimateur de la variance dans le cadre d'un sondage aléatoire simple stratifié est donnée par

$$\hat{V}_{ST}(\hat{t}) = \sum_{h=1}^H \hat{V}(\hat{t}_h) = \sum_{h=1}^H N_h^2 \left(1 - \frac{n_h}{N_h}\right) s_h^2$$

où

$$s_h^2 = \sum_{i=1}^{n_h} \frac{(y_i - \bar{y}_h)^2}{n_h - 1}$$

En cas de non-réponse, s_h^2 est calculé sur les répondants. Il faut donc au minimum deux répondants par groupe pour pouvoir réaliser l'estimation.

Or les 1642 répondants de la source 2 se répartissent entre 1075 groupes différents ayant au moins un répondant (sur 2496 groupes sélectionnés à l'origine pour LFS), parmi lesquels 663 groupes ayant seulement 1 répondant. Pour permettre que tous les répondants soient pris en compte dans l'estimation de la variance, on a procédé au regroupement des groupes. Les strates sont donc redéfinies en regroupant tous les groupes d'un trimestre par province. Les 44 strates ainsi définies (Prov12*TRIM) comprennent de 17 à 59 répondants.

5.5.2.3. Source 3.2

La source 3.2 couvre les ménages composés exclusivement de personnes âgées de plus de 75 ans. Les unités primaires utilisées sont les mêmes que pour les sources 1 et 3.1. Le nombre de ménages sélectionnés par groupe varie de 5 à 8 selon la période. On pourrait a priori utiliser la même modélisation du plan de sondage que pour la source 1. Cependant, le taux de réponse dans cet échantillon est extrêmement faible (seulement 236 ménages ont répondu à l'enquête dans ce groupe). Comme pour la source 2, on a été contraint de fortement simplifier la modélisation du plan de sondage. Dans le cadre de l'estimation de la variance, on considère que l'échantillon est le résultat d'un sondage aléatoire simple dans l'univers des ménages 75+.

5.5.3. Calibrage

Pour l'estimation de la variance de l'estimateur d'un total par calage, l'estimateur de la variance s'obtient en substituant à la variable d'intérêt le résidu de la régression linéaire de cette variable sur les variables de calage. Cette fonctionnalité est intégrée dans POULPE.

Le calage est réalisé indépendamment sur les trois regroupements mentionnés ci-dessus. Les variables de calage sont éventuellement différentes pour les trois groupements utilisés pour le calage. Or POULPE utilise

l'entièreté de l'échantillon répondant comme un seul échantillon. L'on doit donc étendre les variables de calage en leur donnant une valeur 0 en dehors de la source sur laquelle elles ont en premier lieu été définies.

5.5.4. Correction de la non-réponse

Dans le processus d'estimation de la variance, la non-réponse est modélisée comme une seconde phase de tirage, selon un processus de Poisson (on suppose que chaque ménage répond indépendamment des autres ménages, selon une probabilité p_i . Ici aussi, l'hypothèse n'est certainement pas complètement vérifiée, la non-réponse pouvant par exemple dépendre de l'enquêteur, et donc ne pas être totalement indépendante pour des ménages d'un même groupe). Les probabilités de réponse des ménages sont bien sûr inconnues, mais sont estimées, dans le cadre de l'estimation de la variance, pour les ménages répondants par $\tilde{p}_i = d_i/w_i$ où d_i est le poids de sondage et où w_i est le poids de calage.

Cette seconde phase de sondage, nécessaire pour modéliser la non-réponse, est intégrée dans les fonctionnalités de POULPE.

5.5.5. Combinaison des sources

Les trois sources ci-dessus sont chacune calées sur le total de la population correspondante. Pour obtenir une estimation sur la population totale, on combine les résultats de la manière suivante (voir 5.4) :

$$\begin{aligned}\hat{T} &= \frac{\hat{T}_1 \times 4258 + \hat{T}_2 \times 1642}{4258 + 1642} + \hat{T}_3 \\ &= \alpha \hat{T}_1 + (1 - \alpha) \hat{T}_2 + \hat{T}_3\end{aligned}$$

La « mise à l'échelle » des totaux 1 et 2 se justifie parce que chaque source a été traitée indépendamment, et calée sur le total de l'univers des ménages type LFS. Le total 3 est traité indépendamment et calé sur le total de l'univers 75+. Donc

$$\hat{V}(\hat{T}) = \hat{V}(\alpha \hat{T}_1 + (1 - \alpha) \hat{T}_2 + \hat{T}_3) = \hat{V}(\hat{T}'_1 + \hat{T}'_2 + \hat{T}_3)$$

où, p.ex.

$$\hat{T}'_1 = \alpha \hat{T}_1 = \alpha \sum_{i \in r_1} w_i y_i = \sum_{i \in r_1} \alpha d_i \tilde{p}_i y_i$$

et où r_1 est l'échantillon des ménages répondants venant des sources 1 et 3.1.

Pour estimer la variance d'un total sur des trois sources combinées, on a considéré chaque source comme une strate séparée, avec les poids corrigés des facteurs α et $(1 - \alpha)$.

5.5.6. Aperçu de l'estimation de la variance

Tableau 38: Aperçu de l'estimation de la variance pour les catégories ECOICOP jusqu'au 2e niveau pour la Belgique et les régions

COICOP	D'un point de vue géographique	Variance	Consommation	Ecart type	Coefficients de variation	Intervalle de confiance	
						Minimum	Maximum
00	Belgique	103726,41	35764,08	322,07	0,01	35132,83	36395,33
	Flandre	201322,91	36894,99	448,69	0,01	36015,56	37774,43
	Wallonie	294316,24	34589,27	542,51	0,02	33525,95	35652,59
	Bruxelles	683145,31	33356,25	826,53	0,02	31736,26	34976,24
01	Belgique	3108,91	5018,20	55,76	0,01	4908,92	5127,49
	Flandre	5756,62	5127,46	75,87	0,01	4978,75	5276,17
	Wallonie	7813,28	4907,16	88,39	0,02	4733,91	5080,41
	Bruxelles	36112,48	4778,51	190,03	0,04	4406,05	5150,98
02	Belgique	1231,86	726,04	35,10	0,05	657,24	794,83
	Flandre	882,99	641,91	29,72	0,05	583,67	700,15
	Wallonie	5983,62	853,43	77,35	0,09	701,82	1005,04
	Bruxelles	27008,04	789,98	164,34	0,21	467,87	1112,09
03	Belgique	2761,52	1658,72	52,55	0,03	1555,72	1761,71
	Flandre	6529,22	1878,85	80,80	0,04	1720,47	2037,22
	Wallonie	5084,77	1357,63	71,31	0,05	1217,87	1497,39
	Bruxelles	10346,62	1398,52	101,72	0,07	1199,16	1597,89
04	Belgique	2802,57	10837,40	52,94	0,00	10733,64	10941,17
	Flandre	5258,96	10924,75	72,52	0,01	10782,61	11066,88
	Wallonie	7327,41	10518,29	85,60	0,01	10350,51	10686,06
	Bruxelles	28377,61	11308,97	168,46	0,01	10978,80	11639,15
05	Belgique	7327,12	1877,66	85,60	0,05	1709,89	2045,44
	Flandre	8152,79	1955,83	90,29	0,05	1778,86	2132,81
	Wallonie	40958,02	1825,92	202,38	0,11	1429,25	2222,59
	Bruxelles	40041,10	1626,43	200,10	0,12	1234,23	2018,63
06	Belgique	1975,47	1636,46	44,45	0,03	1549,34	1723,57
	Flandre	3766,03	1640,02	61,37	0,04	1519,74	1760,30
	Wallonie	5242,58	1684,78	72,41	0,04	1542,87	1826,70
	Bruxelles	17805,01	1479,10	133,44	0,09	1217,57	1740,63
07	Belgique	36257,01	4065,45	190,41	0,05	3692,24	4438,66
	Flandre	79066,52	4103,82	281,19	0,07	3552,70	4654,95
	Wallonie	95836,27	4505,90	309,57	0,07	3899,13	5112,66
	Bruxelles	69138,64	2600,95	262,94	0,10	2085,59	3116,32
08	Belgique	84,40	1104,36	9,19	0,01	1086,36	1122,37
	Flandre	184,84	1129,39	13,60	0,01	1102,74	1156,04
	Wallonie	155,98	1074,30	12,49	0,01	1049,83	1098,78
	Bruxelles	697,06	1062,76	26,40	0,02	1011,02	1114,51
09	Belgique	4187,00	2579,81	64,71	0,03	2452,99	2706,64
	Flandre	9944,01	2786,18	99,72	0,04	2590,73	2981,63
	Wallonie	6634,31	2307,75	81,45	0,04	2148,11	2467,39
	Bruxelles	23441,78	2306,50	153,11	0,07	2006,41	2606,59

COICOP	D'un point de vue géographique	Variance	Consommation	Ecart type	Coefficients de variation	Intervalle de confiance	
						Minimum	Maximum
10	Belgique	1041,20	256,44	32,27	0,13	193,19	319,68
	Flandre	1222,90	232,26	34,97	0,15	163,72	300,80
	Wallonie	867,76	136,95	29,46	0,22	79,22	194,69
	Bruxelles	45094,31	724,24	212,35	0,29	308,03	1140,46
11	Belgique	2665,08	2351,01	51,62	0,02	2249,83	2452,20
	Flandre	4844,71	2590,77	69,60	0,03	2454,34	2727,19
	Wallonie	7112,78	1926,04	84,34	0,04	1760,74	2091,34
	Bruxelles	29879,35	2347,02	172,86	0,07	2008,22	2685,82
12	Belgique	5070,61	3652,52	71,21	0,02	3512,95	3792,09
	Flandre	13304,22	3883,77	115,34	0,03	3657,69	4109,84
	Wallonie	4718,18	3491,12	68,69	0,02	3356,49	3625,75
	Bruxelles	22821,25	2933,25	151,07	0,05	2637,15	3229,34

Dans le Tableau 38 sont présentés la variance, le coefficient de variation et l'intervalle de confiance pour la consommation générale (ECOICOP 0) et les 12 catégories principales ECOICOP pour la Belgique et les régions.

Pour la Belgique, ces variables par catégorie ECOICOP sont également calculées par groupe d'âge (<30 ans, 30-44 ans, 45-59 ans ou 60+), statut professionnel (ouvrier dans l'industrie et le secteur des services, employé dans l'industrie et le secteur des services, indépendant, chômeur ou 'autre' inactif) et type de ménage (isolé, deux adultes, trois adultes ou plus, parents de familles monoparentales avec enfants à charge, deux adultes avec enfants dépendants ou trois adultes ou plus avec enfants dépendants). Ces données peuvent être obtenues sur demande.

6. Traitement des données pour publication et diffusion

6.1. Représentativité du sondage

6.1.1. Répartitions au niveau du ménage

La composition de l'échantillon est contrôlée afin de s'assurer que les données sont représentatives. À cette fin, la distribution au niveau des ménages pour certaines variables est comparée à la distribution effective de cette variable dans la population sur la base des sources administratives disponibles, telles que le Registre national.

Tableau 39: Répartition selon la taille des ménages participants

Taille du ménage	EBM 2018				Ratio
	Non-pondérée		Pondérée		
	Fréquence	%	Fréquence	%	
1	1.817	29,61	1.732.699	35,21	1.19
2	2.221	36,20	1.565.843	31,82	0.87
3	885	14,42	687.264	13,96	0.97
4	844	13,75	650.830	13,22	0.96
5	288	4,69	207.115	4,21	0.90
6	61	0,99	59.189	1,20	1.21
7	11	0,18	13.916	0,28	1.56
8	7	0,11	3.860	0,08	0.73
9	1	0,02	285	0,01	0.5
9+	1	0,02	28	0,01	0.5
Totaal	6.136	100,00	4.921.414	100,00	1

Remarques

1. Dans l'échantillon des ménages répondants, la taille maximale du ménage n'est pas du même ordre de grandeur que la taille maximale du ménage dans la population.
2. La distribution relative pondérée et non pondérée de la taille du ménage diffèrent, uniquement en raison d'un calibrage limité selon la taille de ménage dans l'EBM.

Tableau 40: Répartition selon le nombre d'actifs par ménage des ménages participants

Nombre d'actifs par ménage	EBM 2018				Ratio
	Non pondérée		Pondérée		
	Fréquence	%	Fréquence	%	
0	2.404	39,18	1.862.431	37,84	0.97
1	1.865	30,39	1.555.951	31,62	1.04
2	1.753	28,57	1411.467	28,68	1.00
3	98	1,60	83.926	1,71	1.07
4	15	0,24	7.122	0,14	0.58
5	1	0,02	517	0,01	0.50
Total	6.136	100,00	4.921.414	100,00	1.00

6.1.2. Répartitions au niveau individuel

Certaines répartitions au niveau individuel sont comparées aux distributions de la population des sources de données administratives afin de vérifier la représentativité au niveau individuel.

Tableau 41: Répartition selon le sexe des membres des ménages participants

Sexe	EBM 2018				Ratio
	Non pondérée		Pondérée		
	Fréquence	%	Fréquence	%	
Homme	6.905	48,45	5.314.136	48,06	0.99
Femme	7.344	51,53	5.741.977	51,93	1.01
Manquant	2	0,02	333	0,01	0.50
Total	14.251	100,00	11.056.446	100,00	1.00

Remarques

1. Parce que l'EBM n'est pas calibrée au niveau individuel, le total pondéré n'est pas exactement égal au chiffre de la population. Les distributions relatives diffèrent également pour cette même raison.
2. Le poids individuel calibré est égal au poids calibré du ménage auquel appartient l'individu.

Tableau 42: Répartition selon l'âge des membres des ménages participants

Âge	EBM 2018				Ratio
	Non pondérée		Pondérée		
	Fréquence	%	Fréquence	%	
- 24	4.243	29,77	3.481.382	31,49	1.06
26-45	3.691	25,90	3.166.632	28,64	1.11
46-64	3.782	26,54	2.610.485	23,61	0.89
65-74	1.747	12,26	934.107	8,45	0.69
75 +	788	5,53	863.840	7,81	1.41
Total	14.251	100,00	11.056.446	100,00	1.00

Remarques

1. Parce que l'EBM n'est pas calibrée au niveau individuel, le total pondéré n'est pas exactement égal au chiffre de la population. Les distributions relatives diffèrent également pour cette même raison.
2. Le poids individuel calibré est égal au poids calibré du ménage auquel appartient l'individu.

6.2. Traitement des données

Après la collecte des données, celles-ci sont à nouveau contrôlées.

6.2.1. *Identification des valeurs aberrantes*

Pour les dépenses mentionnées dans le carnet ménage, des programmes SAS ont été développés pour le contrôle et l'identification des valeurs aberrantes. Pour l'alimentation et les boissons non alcoolisées, nous avons opté pour une méthode intitulée 'resistant fences method' (RFM). Les données ont au préalable subi une transformation de Box-Cox (avec un paramètre λ qui varie en fonction de la catégorie de produit) puisque la symétrie de la distribution améliore la qualité de la détection avec la méthode RFM. Les valeurs aberrantes sont identifiées et remplacées par la médiane de la catégorie de dépenses.

Dans la méthode RFM univariée, la marge de tolérance est déterminée à l'aide de quartiles:

$$[Q1 - \text{coeff.low} * \max(Q3 - Q1, |a \times Q2|), Q3 + \text{coeff.up} * \max(Q3 - Q1, |a \times Q2|)].$$

Le terme $a \times Q2$ représente une valeur minimale de la distance. Les coefficients (low et up) ont été fixés à 3, comme habituellement dans la pratique et à 0,05 pour le paramètre a.

Pour l'alimentation et les boissons non alcoolisées, les valeurs aberrantes sont détectées automatiquement. Ce n'est pas le cas pour les autres catégories. À l'aide de la méthode RFM, nous corrigeons les dépenses alimentaires ECOICOP 01 par un facteur de 0,6 %, ce qui est raisonnable.

Tableau 43: Différence de montant après l'application de la méthode RFM pour ECOICOP 1 et 2 (non pondéré)

Année	ECOICOP	Montant avant	Montant après application RFM	Variation
2.018	01	16.000.576	11.975.634	4.024.942
	02	1.327.000	1.318.110	8.890

Si nous utilisons la même méthode pour ECOICOP 12 par exemple, 'Soins corporels et services', la baisse des dépenses totales s'établirait à 25 %, ce qui est irréaliste. Prenons l'exemple de la catégorie 06131A 'Lunettes et verres de contact'. Dans cette catégorie, citée par 420 ménages dans notre échantillon, nous trouvons des montants de 18 euros (premier quartile) ou même moins, et des montants supérieurs à 1000 euros. Le maximum pour 2018 est 1215 euros. Ces montants petits et grands sont assez plausibles pour un achat de lunettes. Si nous utilisons la même méthode RFM que pour la ECOICOP01, nous devrions 'corriger' tous les montants supérieurs à 900 euros. La correction consisterait à remplacer ce montant par la médiane des montants rapportés, à savoir 84 euros... Il va sans dire que cette correction réduirait considérablement et de manière totalement erronée le montant total dépensé pour des lunettes.

C'est pourquoi nous préférons contrôler les valeurs données par catégorie de dépenses et remplacer les montants réellement divergents par la médiane, sans effectuer de correction automatique.

Lors de l'enregistrement des dépenses dans un carnet de dépenses pendant un demi-mois, il est difficile d'avoir une idée des dépenses périodiques et importantes (rares ou non) d'un ménage. C'est pourquoi des questions sur ces deux types de dépenses sont reprises dans un questionnaire ménage distinct^{xii}.

Différents contrôles de cohérence sont effectués en comparant les dépenses enregistrées dans le carnet du ménage et celles du questionnaire ménage.

Toutes les dépenses enregistrées sont ramenées sur un demi-mois: les dépenses du carnet restent telles quelles. Les dépenses du questionnaire ménage (questions sur les 4 derniers mois et questions sur la possession des biens durables) sont divisées par 8 et les dépenses périodiques sont divisées par 2. Pour le tableau final, toutes ces dépenses sont recalculées pour refléter les dépenses sur base annuelle, c'est-à-dire multipliées par 24.

6.2.2. Imputation des dépenses énergétiques

Le questionnaire ménage a été entièrement rempli sous la forme d'une interview CAPI. Cela signifie que lors de la dernière visite au ménage, l'enquêteur lit les questions sur son écran d'ordinateur et encode immédiatement les réponses données. Il suit l'ordre des questions (routing) tel que programmé à l'avance afin de s'assurer que toutes les questions pertinentes soient posées. Avant 2014, un questionnaire ménage sur papier était utilisé. Dans la pratique, c'était souvent le ménage lui-même qui le remplissait, après quoi l'enquêteur effectuait seulement une vérification. Beaucoup de données étaient dès lors manquantes, notamment sur les dépenses énergétiques. Les incohérences dans les réponses étaient également nombreuses (p. ex. des ménages qui déclarent que leur principale source de chauffage fonctionne au mazout de chauffage, mais ne mentionnent une dépense que pour du gaz naturel). Ces informations manquantes et ces incohérences engendraient beaucoup de travail de correction par la suite. L'introduction des questionnaires ménage CAPI a grandement amélioré la qualité des données, réduisant ainsi le travail de vérification ex post.

Tableau 44: Nombre de données manquantes par catégorie pour les dépenses énergétiques

Dépense	Nombre de données manquantes en 2018
Eau	555
Électricité	82
Gaz (conduite de gaz)	7

Il est également possible d'encoder des factures combinées. Dans une seule facture, par exemple, un ménage paie la facture d'électricité et de gaz, ou la consommation d'électricité et d'eau. Cependant, les ménages et les enquêteurs devaient quand même ventiler les montants autant que possible. Dans la suite de ce document, on appelle également imputation la ventilation d'une telle combinaison de types d'énergie. Même pour les ménages qui n'ont pas fourni d'informations dans leur questionnaire ménage, des sources d'information sur leurs dépenses énergétiques sont disponibles. Les carnets, par exemple, fournissent des informations sur les dépenses pendant la période de référence.

Il a été décidé d'utiliser d'abord les informations des carnets pour estimer les dépenses manquantes, puis d'essayer de ventiler les dépenses mixtes et d'imputer ensuite les dépenses énergétiques encore manquantes en utilisant ensuite un modèle de régression linéaire.

6.2.3. Collecte des informations des carnets

La première étape consiste à utiliser les informations enregistrées dans les carnets pendant la période de référence comme base pour imputer une consommation annuelle manquante dans le questionnaire ménage. Nous ne faisons cette imputation que pour l'eau, le gaz et l'électricité parce que ces dépenses sont effectuées selon une fréquence fixe. En estimant cette périodicité, les dépenses qui ont eu lieu au cours d'un mois donné peuvent être converties en montants annuels. Cette estimation a été faite sur la base de montants moyens basés sur les données des questionnaires. Les données imputées se sont vu attribuer un 'drapeau' :

Tableau 45: Nombre de données imputées des carnets pour les dépenses énergétiques

Catégorie	Drapeau	2018 (n)
Facture d'eau	FL_W	126
Facture d'électricité	FL_E	3
Facture de gaz	FL_G	2

6.2.4. Ventilation de factures groupées

Il y a différentes combinaisons possibles des dépenses énergétiques. La combinaison la plus fréquente est la facturation conjointe de gaz et d'électricité.

Tableau 46: Aperçu des factures groupées pour les dépenses énergétiques

Description	2018 (n)
Eau, gaz et électricité	126
Eau et électricité	34
Eau et gaz	42
Electricité et gaz	2143

Les combinaisons "eau et électricité" et "eau et gaz" ont été ventilées en établissant un modèle de régression qui estime la consommation d'eau. Le modèle de régression a été basé sur les variables indépendantes taille du ménage (exprimée en unités équivalentes), région, présence d'un système de récupération des eaux de pluie et d'une machine à laver. La présence d'un puits n'a pas eu d'incidence significative sur les estimations et n'a pas été incluse dans la régression. La consommation de l'autre source d'énergie (l'électricité et le gaz respectivement) a ensuite été calculée en faisant la différence entre la facture commune et l'estimation de la consommation d'eau.

La combinaison "électricité et gaz" a été ventilée en établissant un modèle de régression qui estime la consommation d'électricité. Le modèle était basé sur la taille du ménage (en unités équivalentes), la région et la présence d'une machine à laver et de panneaux solaires photovoltaïques. La présence d'un réfrigérateur ou congélateur ou l'utilisation de l'électricité pour un chauffage d'appoint n'a pas eu d'incidence significative sur les estimations et n'ont donc pas été retenues pour le modèle de régression. La consommation de gaz a été calculée en faisant la différence entre la facture commune et la consommation d'électricité estimée.

La combinaison "eau, gaz et électricité" a été ventilée en effectuant les deux estimations, d'abord pour l'eau et ensuite pour l'électricité. La consommation de gaz a été ensuite calculée en faisant la différence entre la facture commune et les deux consommations estimées d'eau et d'électricité.

Remarque: toutes les factures communes n'ont pas pu être ventilées de cette façon. Le montant de la consommation restante après déduction de l'estimation de la facture commune était parfois trop faible. Dans ce cas, aucune ventilation n'a été effectuée.

6.2.5. Imputer les données manquantes

Les données encore manquantes ont été complétées par une estimation fondée sur un modèle de régression par catégorie de données. Les dépenses manquantes dans la catégorie BOIS (ECOICOP=04549A) n'ont pas été imputées. Le bois de chauffage peut être obtenu gratuitement ou un stock peut avoir été constitué pour plusieurs années.

Tableau 47: Nombre de données imputées sur la base des modèles de régression pour les dépenses énergétiques

Catégorie	Drapeau	2018 (n)	2016 (n)
Eau	FL_WR	1219	320
Électricité	FL_ER	141	148
Gaz	FL_Gas	203	155
Fioul domestique (mazout)	FL_oil	105	91
Butane	FL_butane	.	11
Propane	FL_propane	.	78
Butane et propane	FL_butprop	9	.
Charbon	FL_coal	1	4
Pellets	FL_pellets	5	9
Pétrole	FL_petroleum	2	.

- On suppose que chaque ménage doit payer une facture d'eau. Les montants manquants sont remplis à cette étape à l'aide du même modèle de régression que celui utilisé pour la ventilation de la facture commune.
- Comme pour la consommation d'eau, nous supposons que chaque ménage participant consomme de l'électricité. La méthode de régression est la même que celle utilisée pour ventiler les factures groupées.
- Un montant est imputé pour le gaz pour tous les ménages qui déclarent se chauffer au gaz et qui n'ont enregistré aucune dépense annuelle de gaz dans le questionnaire ménage ou le carnet. Les variables indépendantes pertinentes sont la taille du ménage (exprimée en unités équivalentes), la région et le fonctionnement du système principal de chauffage au gaz.
- Les montants pour le gasoil de chauffage ont également été estimées sur la base d'un modèle de régression linéaire. Ceci pour tous les ménages qui ont indiqué qu'ils utilisaient du gasoil pour se chauffer, mais qui n'ont pas enregistré de dépenses annuelles en gasoil. Les variables indépendantes pertinentes sont la taille du ménage (exprimée en unités équivalentes), habitant de la Région de Bruxelles-Capitale et le fonctionnement du système principal de chauffage au gasoil.
- Aucun modèle de régression n'a pu être élaboré pour l'imputation du butane en raison de l'absence de données complètes. L'imputation a ensuite été effectuée sur la base des dépenses annuelles moyennes en butane de ces ménages qui avaient déclaré un montant. L'imputation du propane a donc été effectuée sur la base de la moyenne.
- L'imputation du charbon aux ménages qui utilisent le charbon pour se chauffer n'a pas pu être effectuée sur la base d'un modèle de régression. Le nombre de cas complets dans le set de données était en effet trop faible pour cela. L'imputation a dès lors été effectuée sur la base des dépenses annuelles moyennes en charbon de ces ménages qui avaient déclaré une dépense de charbon.
- La consommation de pellets n'a pas pu être prévue de manière significative par un modèle de régression. L'imputation a donc été effectuée sur la base de la moyenne.
- En ce qui concerne la consommation de pétrole, aucune donnée n'était manquante, et donc aucune imputation n'a été nécessaire.

6.3. Loyer imputé

Le 'loyer imputé' est un loyer fictif attribué aux propriétaires et aux ménages qui habitent un logement à titre gratuit.

D'après le SEC (système européen des comptes), l'achat d'une habitation est principalement considéré comme un investissement/une formation de capital, et pas comme une consommation^{xiii}. La propriété de l'habitation fournit un service de logement/hébergement consommé par les ménages au fil du temps. Par conséquent, un prix doit être estimé. Étant donné qu'il n'y a pas de transaction monétaire, cette estimation doit se faire à l'aide d'une imputation. Ce loyer imputé fait partie de la consommation. Dès lors, afin de rendre l'EBM cohérente avec le SEC, les mêmes principes sont appliqués.

Pour EU-SILC, le loyer imputé est obligatoirement transmis à Eurostat depuis 2007. Pour EBM, cette variable est déjà demandée depuis 2005.

Afin d'estimer le prix auquel leur logement pourrait être loué, et donc avoir une idée de l'avantage financier lié au statut de propriétaire, nous utilisons la méthode recommandée par Eurostat: un modèle de Heckman à deux phases. Nous utilisons d'abord une régression probit pour estimer la probabilité d'être locataire en fonction des caractéristiques du ménage et de ses membres (notamment l'âge, le niveau d'instruction, le nombre de membres actifs, etc.) Nous estimons ensuite à l'aide d'une régression le loyer avec les caractéristiques du logement présentes dans l'enquête, telles que le type et l'âge du logement, le nombre de pièces, la présence d'un jardin, d'un garage, d'un balcon, du chauffage central Et naturellement, en tenant compte du résultat de la première régression.

Cette méthode est utilisée par Statbel tant pour l'enquête SILC que pour EBM. Par ailleurs, les résultats d'EBM et de SILC sont comparés pour chaque année d'enquête.

Tableau 48: Validation du loyer réel et imputé par rapport à l'enquête EU-SILC (pondéré)

	EBM 2018 (€)	SILC 2018 (€)
Loyer mensuel par ménage locataire	561	556
Loyer imputé mensuel par ménage propriétaire	728	672

6.3.1. *Extrapolation des données*

Toutes les dépenses ECOICOP sont extrapolées de la même manière, en estimant d'abord les dépenses totales pour l'ensemble de la population par catégorie ECOICOP, et en divisant ensuite ce total par le nombre total de ménages dans la population, c'est-à-dire pour 2018: 4.921.414 ménages.

Étant donné que nous donnons une dépense moyenne pour tous les ménages (pas seulement pour les achats des ménages), le nombre de ménages dans le dénominateur est toujours le même. Il est donc facile de recalculer, sur la base des tableaux que nous fournissons comportant les dépenses moyennes, le montant total des dépenses par catégorie ECOICOP.

7. Publication et diffusion

7.1. [Publication sur le site web](#)

A l'issue de chaque période de collecte de données, les données collectées sont encodées, nettoyées, calibrées, agrégées et contrôlées, de sorte que les résultats puissent être publiés 9 mois après la période de référence.

Les résultats d'EBM sont diffusés sur le site web via des tableaux Excel^{xiv}. Cela concerne:

- Les résultats sur les dépenses par catégorie ECOICOP pour la Belgique et les régions.
- Les résultats sur la possession de biens durables pour la Belgique et les régions.

Les données sont collectées 1 fois tous les 5 ans pour Eurostat (2005 – 2010 – 2015 – 2020 - ...). Eurostat ne publie que quelques années plus tard car la collecte de données n'a pas lieu la même année dans chaque pays.

Les chercheurs peuvent demander des données agrégées (et donc anonymisées). Des microdonnées anonymisées peuvent également être obtenues sous certaines conditions.

7.2. [Dépenses courantes / saisonnières](#)

Les tableaux Excel montrent par produit ou catégorie ECOICOP les dépenses annuelles moyennes pour un ménage de la population belge de ménages privés, y compris les ménages qui n'ont pas effectué d'achat dans cette catégorie ECOICOP. L'estimation de ces moyennes est abordée au par. 5.4.1 et au par. 5.4.2. En raison de l'inclusion des ménages qui n'ont pas enregistré d'achats, certains résultats semblent étranges ou pas directement utilisables. Un problème similaire se présente pour l'estimation de la proportion de ménages qui effectuent un achat dans une certaine catégorie ECOICOP (par. 5.4.3).

Les ménages remplissent quotidiennement un carnet de dépenses pendant un demi-mois. Ce carnet est accompagné d'un questionnaire ménage. Le carnet et le questionnaire ménage reprennent tous les deux tant les dépenses courantes que saisonnières.

L'estimation des dépenses annuelles totales et moyennes sur la base de l'EBM repose sur la supposition que le comportement du ménage pendant une certaine période de référence (d'un demi-mois) reste le même toute l'année, en d'autres termes, que ces dépenses x_i^0 effectuées pendant la période de référence multipliées par 24 fournissent le total x_i de ses dépenses pendant l'année. Pour les dépenses courantes, c'est une hypothèse plausible. En effet, pour certains ménages répondants, x_i sera une bonne approximation des dépenses annuelles totales, alors que pour d'autres ménages, on obtiendra soit une surestimation (si par hasard une dépense supérieure à la moyenne est effectuée pendant la période de référence), soit une sous-estimation (si par hasard une dépense inférieure à la moyenne (ou aucune dépense!) a été effectuée pendant la période de référence) des dépenses annuelles totales x_i . Comme les ménages sont répartis assez uniformément sur les 24 périodes de référence, un nombre suffisamment important de ménages peut être considéré ensemble pour donner une bonne image des dépenses courantes totales et moyennes annuelles. Les dépenses saisonnières sont également traitées officiellement de cette manière.

Les dépenses moyennes de tous les ménages (y compris de ceux qui n'en ont pas effectuées) telles que publiées sur le site web sont la meilleure façon d'estimer les dépenses pour une catégorie de produits particulière. Nous démontrons ci-dessous que cela s'applique également à (certains) produits saisonniers ;

d'autre part, nous montrons également que l'estimation des proportions et du nombre de ménages qui effectuent des dépenses pour un produit saisonnier est problématique.

Nous l'illustrons cela à l'aide des dépenses pour un sapin de Noël (naturel) (catégorie de produits fictive, car un 'sapin de Noël naturel' fait partie de la catégorie ECOICOP 09322F). Supposons:

- H1: que les ménages répondants qui ont acheté un sapin de Noël, ne l'ont fait que pendant la première moitié de décembre;
- H2: que précisément 50 % des ménages répondants ont acheté un sapin de Noël, avec la première moitié de décembre comme période de référence;
- H3: que chacun de ces ménages ont payé $x_i^0 = €50,00$ ($x_i^0 = €0,00$ pour tous les autres ménages répondants).

Des hypothèses de travail supplémentaires, uniquement pour simplifier les calculs ci-dessous, sont

- H4: que tous les ménages répondants ont le même poids calibré ajusté $\tilde{w}_i = \tilde{w}$;
- H5: qu'exactly le même nombre de ménages répondants a été obtenu pour l'ensemble des 24 périodes de référence.

Il ressort immédiatement de l'hypothèse H4 que $N = n\tilde{w}$ (n est le nombre total de ménages répondants; N est le nombre total de ménages dans la population étudiée). Il ressort de l'hypothèse 5 que n est un multiple de 24, que $n/24$ ménages répondants ont la première moitié de décembre comme période de référence, et il ressort des hypothèses H1 et H2 que $n/48$ ménages répondants ont effectivement acheté un sapin de Noël. La dépense moyenne pour un sapin de Noël pour l'ensemble de la population de ménages privés belges est alors estimée comme suit:

$$\begin{aligned}\hat{X} &= \frac{1}{N} \sum_{i=1}^n \tilde{w}_i x_i \\ &= \frac{1}{n\tilde{w}} \sum_{i=1}^n \tilde{w}_i \times (24 \times x_i^0) = \frac{1}{n} \times 24 \times \sum_{i=1}^n x_i^0 \\ &= \frac{1}{n} \times 24 \times \left(50 \times \frac{n}{48} + 0 \times \frac{47n}{48} \right) \\ &= 24 \times 50 \times \frac{1}{48} = 25\end{aligned}$$

Et la dépense annuelle totale estimée pour les sapins de Noël dans la population belge de ménages privés est de $25N$. Cette moyenne et le total qui en découle sont des estimations parfaitement acceptables, sous les hypothèses H1-H3 (complétées par les hypothèses de travail H4-H5).

L'estimation de la proportion et du nombre de ménages privés belges qui achètent un sapin de Noël est en revanche incertaine si l'on utilise les formules générales. En effet, de nouveau sous les hypothèses H1-H5, la variable dummy x , qui indique si un ménage répondant i achète ou non un sapin de Noël, a la valeur

$$\begin{aligned}x_i &= 1 \quad \text{voor } n/48 \text{ huishoudens die wel de aankoop van een kerstboom registreren} \\ &= 0 \quad \text{voor } 47n/48 \text{ huishoudens die geen aankoop van een kerstboom registreren}\end{aligned}$$

Par conséquent, la proportion est estimée comme suit:

$$\begin{aligned}\hat{X} &= \frac{1}{N} \sum_{i=1}^n \tilde{w}_i x_i = \frac{1}{n\tilde{w}} \sum_{i=1}^n \tilde{w} x_i = \frac{1}{n} \sum_{i=1}^n x_i \\ &= \frac{1}{n} \times \frac{n}{48} = \frac{1}{48} = 2,08\%\end{aligned}$$

Et le nombre estimé de ménages privés belges qui achètent un sapin de Noël est de $\frac{1}{48}N$. Cette proportion et le nombre qui en découle sont absolument inacceptables: nous attendons que 50 % de tous les ménages de la population achètent un sapin de Noël, car le 'comportement' à l'égard de l'achat d'un sapin de Noël qui est démontré par les ménages répondants, avec la première moitié de décembre comme période de référence, est supposé pouvoir être projeté aux autres ménages répondants et à l'ensemble de la population.

Enfin, les dépenses annuelles moyennes estimées pour un sapin de Noël pour les ménages qui font effectivement cette dépense, sont égales à $25N / \left(\frac{1}{48}N\right) = 25 \times 48 = \text{€}1.200$. Il s'agit clairement d'une surestimation (nous attendons en effet une moyenne de 50,00€!), ce qui est naturellement la conséquence du fait que le dénominateur $\frac{1}{48}N$ est une nette sous-estimation du nombre de ménages qui achètent effectivement un sapin de Noël.

Conclusion: nous ne publions pas les dépenses moyennes par ménage avec dépense effective ni la part des ménages avec dépense effective, car bien que cela fournisse des informations intéressantes sur certains biens de consommation quotidiens dont la consommation est répartie équitablement, ce n'est pas le cas pour tous les biens (même l'alimentation peut avoir un caractère saisonnier).

7.3. Microdonnées disponibles pour la recherche

Il existe deux niveaux de variables dans l'EU-HBS: les variables au niveau du ménage et les variables au niveau individuel. Vous trouverez ci-dessous la liste des variables par niveau dans l'enquête.

Tableau 49: Données au niveau du ménage et au niveau individuel

Données au niveau du ménage	
DONNEES DE BASE	Données de base sur le ménage
LOGEMENT	Type de logement, statut de propriété et caractéristiques du logement
	Equipements du logement
	Coût de l'hébergement
VOITURES	Voiture principale, nombre de voitures et nombre de voitures à disposition de l'employeur
BIENS DURABLES	Possession de biens durables
DÉPENSES	Dépenses des ménages selon la nomenclature ECOICOP d'Eurostat.
	Les quantités sont demandées mais ne sont exploitées que comme variables de contrôle. La qualité est encore insuffisante pour les publier.
Données au niveau individuel	
DONNEES DE BASE	Données individuelles de base
	Données démographiques
REVENUS	Données de base sur le revenu déclaré des ménages

Un manuel des codes fournit des informations plus détaillées sur les fichiers et les variables existants. Il se trouve dans le catalogue des données^{xv}.

Les microdonnées belges peuvent être demandées sur le site de Statbel via les étapes décrites dans la rubrique « Microdonnées pour la recherche »^{xvi} et les fichiers européens auprès d'Eurostat^{xvii}

Annexes

Définitions

L'**Enquête sur le budget des ménages^{xviii} (EBM)** est une enquête nationale axée sur les dépenses supportées par les ménages en vue de l'acquisition de biens et de services. Elle dresse un tableau des conditions de vie dans l'Union européenne (UE). Elle est menée par chaque État membre et est utilisée pour définir la pondération d'importants indicateurs macroéconomiques, comme les indices des prix à la consommation (qui mesurent l'inflation) et la comptabilité nationale.

Les données des enquêtes sont ventilées en fonction des caractéristiques des ménages: revenus, caractéristiques socioéconomiques, taille et composition, degré d'urbanisation et région.

La collecte des données implique la combinaison d'un ou plusieurs entretiens et d'un journal complété par les ménages quotidiennement.

L'unité de base pour la collecte et l'analyse des données dans l'EBM est le ménage. Il importe d'identifier la **personne de référence** (le chef du ménage, généralement) dont les caractéristiques personnelles pourront être utilisées à des fins de classement et d'analyse des informations relatives au ménage. Les résultats sont classés et présentés en fonction de la catégorie socioéconomique, de l'emploi exercé et de la situation professionnelle, du revenu, du sexe et de l'âge de la personne de référence. La **personne de référence selon EBM** est la personne qui participe pour la plus grande part aux revenus du ménage. Cette définition de la **personne de référence** (chef de ménage) ne correspond pas nécessairement à celle du RN.

Les dépenses supportées par les ménages aux fins de l'acquisition de biens et de services sont consignées au prix réellement payé, y compris les impôts indirects (TVA et accises) à charge de l'acheteur.

Dans le cadre des enquêtes sur les conditions sociales et le revenu, comme EBM, le **ménage^{xix}** est défini comme une unité budgétaire ou, dans la pratique, comme une unité sociale:

- ayant des structures communes,
- partageant les dépenses domestiques et les besoins quotidiens,
- vivant dans une résidence commune.

Il se compose d'une personne vivant seule ou d'un groupe de personnes non obligatoirement liées par un lien de parenté résidant à la même adresse et consommant collectivement certains biens ou services, c'est-à-dire partageant au moins un repas par jour ou partageant un salon ou une pièce de séjour.

Il existe également, par opposition aux ménages privés, des ménages collectifs ou ménages institutionnels, par exemple: les hôpitaux, les maisons de retraite, les établissements d'hébergement collectifs, les prisons, les casernes militaires, les institutions religieuses, les internats et les résidences mises à disposition des travailleurs par les employeurs. Ceux-ci ne sont pas interrogés dans le cadre de l'EBM.

Le **revenu disponible équivalent^{xx}** correspond au revenu total d'un ménage, après impôt et autres déductions, disponible en vue d'être dépensé ou épargné, divisé par le nombre de membres du ménage converti en équivalents adultes. L'équivalence entre les membres du ménage est obtenue par pondération en fonction de l'âge, à partir de l'échelle d'équivalence modifiée de l'OCDE.^{xxi}

Le revenu disponible équivalent est calculé en trois étapes:

1. D'abord, tous les revenus monétaires perçus par chaque membre du ménage, peu importe leur origine, sont additionnés. Il s'agit des revenus du travail et des placements et des prestations sociales, auxquels s'ajoute tout autre revenu perçu par le ménage, tous nets d'impôts et de cotisations sociales.
2. Ensuite, afin de prendre en compte les différences de taille et de composition des ménages, le revenu total (net) du ménage est divisé par le nombre d'équivalents adultes au moyen d'une échelle standard: l'échelle modifiée de l'OCDE. Cette échelle assigne une pondération à tous les membres du ménage (dont l'addition constitue la taille équivalente du ménage):
 - a. 1,0 au premier adulte;
 - b. 0,5 à chaque membre âgé de 14 ans et plus;
 - c. 0,3 aux enfants de moins de 14 ans.
3. Enfin, le chiffre obtenu, appelé «revenu disponible équivalent», est distribué à parts égales entre chaque membre du ménage.

On entend par **ménage de type EFT** (dans la base de sondage), un ménage dont au moins un membre appartient à la classe d'âge des 15-76⁹, à savoir un ménage qui serait aussi éligible pour le tirage d'EFT. Le critère d'âge doit être valable au moment du tirage. On entend par **ménage âgé** (dans la base de sondage) un ménage dont tous les membres ont au moins 77 ans au moment du tirage. Dans un premier temps, nous parlons de ménages de type EFT et de ménages âgés dans le contexte du tirage supplémentaire du RN (source 3). Mais les ménages qui ont déjà été tirés pour EFT ou recrutés pour EBM 2016 via des échantillons EFT peuvent également être considérés comme des ménages de type EFT.

La **Classification of Individual Consumption according to Purpose (COICOP)** est une nomenclature des dépenses de consommation gérée par les Nations unies (COICOP (ONU)). La **European Classification of Individual Consumption according to Purpose (ECOICOP)^{xxii}** est l'adaptation européenne de cette nomenclature et compte 5 niveaux. Cette nomenclature est à la base des nomenclatures utilisées pour les dépenses de consommation notamment dans les comptes nationaux^{xxiii}, l'indice des prix à la consommation^{xxiv} et EBM. L'ECOICOP se compose de douze divisions décrivant la consommation individuelle des ménages, divisées en 47 groupes et 117 classes. La nomenclature est ensuite sous-divisée en sous-classes. Pour la Belgique, un sixième niveau est ajouté à l'ECOICOP: elle devient donc la COICOP-BE. Les données pour la Belgique sont publiées jusqu'au sixième niveau, et jusqu'au quatrième niveau à l'échelle régionale, en raison de la faible consommation dans certaines catégories. La nomenclature COICOP des Nations Unies fait actuellement l'objet d'une mise à jour.^{xxv}

⁹ Si nous parlons d'une personne du groupe d'âge A-B, cela signifie que cette personne a au moins A ans, mais n'a pas encore B+1 ans.

L'échantillonnage à deux degrés en bref – Approche des poids de sondage pour les USE

Cadre théorique

Considérons une base de sondage (c-à-d une liste) de N unités statistiques qui joueront le rôle d'USE (unités secondaires d'échantillonnage). Supposons que la base de sondage soit divisée en ce que nous appelons ci-après des UPE (unités primaires d'échantillonnage). Un échantillonnage à deux degrés dans ce texte se compose comme suit:

- Degré 1: tirage systématique des UPE avec éventuellement une sélection multiple d'UPE suffisamment importantes;
- Degré 2: tirage aléatoire des USE parmi les UPE tirées précédemment.

Les tailles des UPE dans la population, les nombres réalisés de sélections d'UPE et les nombres réalisés d'USE sélectionnées peuvent être résumés comme suit:

Tableau 50: Tailles des UPE dans la population, nombres réalisés de sélections d'UPE et nombres réalisés d'USE sélectionnées

Numérotation UPE	Taille UPE	Nombre de sélections d'UPE (Degré 1)	Nombre d'USE sélectionnées (Degré 2)
1	N_1	m_1	$m_1 g$
2	N_2	m_2	$m_2 g$
3	N_3	m_3	$m_3 g$
...
h	N_h	m_h	$m_h g$
...
H	N_H	m_H	$m_H g$
Totaux:	$N = \sum_{h=1}^H N_h$	$m = \sum_{h=1}^H m_h$	mg

Les tailles d'UPE N_h sont déterminées sur la base d'une extraction du RN. Pour une population cible choisie, les USE éligibles sont identifiées et rassemblées dans la base de sondage des USE. Dans cette base de sondage, les USE sont regroupées en UPE d'une manière spécifique. Chaque N_h est le nombre d'USE éligibles dans l'UPE h . Le total N est le nombre total d'USE éligibles dans la population cible ou l'ampleur totale de la base de sondage.

Le nombre de fois que l'UPE h est tirée au degré 1 est une variable aléatoire m_h que nous appelons la cardinalité de l'UPE. La somme m des cardinalités est déterminée au préalable: il s'agit du nombre de groupes d'USE que nous souhaitons créer. (Il convient de noter que le nombre d'UPE *différentes* tirées est le nombre d'UPE h pour lequel $m_h > 0$. Ce nombre est également une variable aléatoire avec une valeur réalisée pouvant être inférieure à m . Cela n'est toutefois pas pertinent dans le contexte actuel.)

Afin d'obtenir des formules utilisables pour les probabilités de sélection et les poids de sondage pour les USE, nous faisons les hypothèses de simplification suivantes:

- Un nombre fixe g d'USE est tiré par sélection d'une UPE, au degré 2.

- Les tirages aux degrés 1 et 2 sont basés sur une seule extraction du RN. En d'autres termes, sur une liste d'USE divisée en UPE.

La probabilité de sélection de chaque USE dans la base de sondage est donc mg/N , de sorte que chacune des mg USE sélectionnées ait un même poids de sondage $N/(mg)$.

Évolution de la base de sondage entre le degré 1 et le degré 2

En pratique (pour EFT), les UPE et les USE sont généralement tirées avec un intervalle de plusieurs mois, et différentes extractions du RN sont utilisées pour ces tirages. Cela complique le plan de sondage et le calcul des probabilités de sélection et des poids d'échantillon. Pour arriver aux formules ci-dessus, et donc calculer des poids de sondage *approximatifs*, nous faisons une hypothèse simplificatrice supplémentaire:

- L'évolution de la base de sondage entre le tirage des UPE au degré 1 et le tirage ultérieur des USE au degré 2, qui est en pratique nécessaire, est telle que tous les N_h ainsi que les N varient peu.

Compte tenu de cette hypothèse, les probabilités de sélection mg/N et les poids de sondage $N/(mg)$ peuvent être bien estimés en utilisant la valeur de N , qui peut être déterminée à partir de l'une ou l'autre des deux bases de sondage.

Si (en pratique) l'intervalle de temps entre le tirage au degré 1 et au degré 2 devient plus important, alors, compte tenu de la croissance de la population belge, la valeur de N peut augmenter considérablement. Dans le même temps, les valeurs N_h changeront également (augmentation pour la plupart des h , mais une diminution pour certains h n'est pas exclue). (Notez que nous laissons H inchangé). Dans ce cas, nous faisons l'hypothèse supplémentaire suivante:

- L'évolution de la base de sondage entre le tirage des UPE au degré 1 et le tirage ultérieur des USE au degré 2, qui est en pratique nécessaire, est telle que tous les N_h varient du même facteur. La valeur de N varie également de ce même facteur.

Compte tenu de cette hypothèse, les poids de sondage sont de préférence estimés en utilisant dans la formule $N/(mg)$ la valeur de N qui peut être déterminée à partir de la base de sondage utilisée au degré 2. (Cela revient à prétendre que lors du tirage des USE, les UPE sont également tirées à nouveau de la nouvelle base de sondage avec le même résultat, c'est-à-dire les mêmes valeurs pour les cardinalités m_h , que lors du tirage initial des UPE.)

Extension d'un échantillon à deux degrés avec un tirage supplémentaire d'USE

Considérons un échantillonnage à deux degrés réalisé exactement comme décrit formellement à l'annexe 'Cadre théorique' et qui utilise donc une seule base de sondage pour les tirages aux degrés 1 et 2. Imaginons que cet échantillon initial soit complété (ultérieurement) en tirant, par UPE h sélectionnée, $m_h g'$ USE supplémentaires, ce qui implique que les $m_h g$ USE tirées précédemment ne sont pas à nouveau tirées. Supposons également que la même base de sondage initiale soit utilisée. L'ensemble de ces tirages peut alors être compris comme un échantillonnage à deux degrés comme décrit à l'annexe 'Cadre théorique' dans lequel g est remplacé par $g + g'$. En d'autres termes, la taille du groupe est étendue de g à $g + g'$ USE. Les poids de sondage des USE dans l'échantillon à deux degrés complété sont alors .

Dans la pratique (pour EFT et EBM), les tirages initiaux et supplémentaires peuvent être assez éloignés dans le temps. Dans ce cas, il peut être indiqué pour le tirage supplémentaire de créer une nouvelle base de

sondage à partir du RN. Si l'hypothèse simplificatrice de l'annexe précédente 'Évolution de la base de sondage entre le degré 1 et le degré 2' est retenue, nous estimerons les poids de sondage pour l'échantillon complémentaire en dérivant la valeur N de la nouvelle base de sondage dans la formule du paragraphe précédent.

Situation particulière: UPE trop petites

Dans la situation simplificatrice (théorique) de l'annexe 'L'échantillonnage à deux degrés en bref – Approche des poids de sondage pour les USE', il peut apparaître au degré 2 qu'une USE h , qui a été tirée $m_h > 0$ fois au degré 1, soit « trop petite », à savoir que $m_h g > N_h$. On peut démontrer alors que $m_h - 1$ groupes de taille g peuvent être formés (et que $(m_h - 1)g \leq N_h$), mais que le « dernier » groupe m_h ne contiendra que $N_h - (m_h - 1)g < g$ USE. Il convient de noter qu'une UPE « trop petite » est tirée de manière exhaustive au degré 2.

Dans ce cas, nous calculerons les poids de sondage en remplaçant au dénominateur la taille totale *prévue* de l'échantillon mg par la taille totale *réalisée* que nous noterons n_{rel} . Le poids de sondage pour les USE peut être noté N/n_{rel} . Il convient de noter que n_{rel} est la somme des nombres réalisés d'USE $n_{h,rel}$ spécifiques à l'UPE, dont certains peuvent être inférieurs aux nombres prévus $m_h g$. Par ailleurs, $n_{h,rel} = N_h$ est comme $m_h g > N_h$.

Enfin, il convient de noter que, dans la pratique (EFT et EBM), le cas des USE "trop petites" peut se produire à n'importe quel degré du tirage : lors du tirage des UPE (bien que le problème ne se manifeste pas directement) ; non pas lors du tirage des UPE, mais lors du tirage différé des USE lorsqu'une nouvelle base de sondage est utilisée ; lors d'un tirage supplémentaire des USE lorsqu'une nouvelle base de sondage est utilisée.

La formule (d'approximation) N/n_{rel} pour les poids de sondage, dans laquelle la valeur de N est déterminée de manière appropriée (voir annexe 'Évolution de la base de sondage entre le degré 1 et le degré 2' et annexe 'Extension d'un échantillon à deux degrés avec un tirage supplémentaire d'USE') est par conséquent généralement utilisable.

Références

Arrêté royal du 17 janvier 1996.

http://www.ejustice.just.fgov.be/cgi_loi/change_lg.pl?language=fr&la=F&cn=1996011731&table_name=loi

Arrêté royal du 17 mai 2006.

http://www.ejustice.just.fgov.be/cgi/article_body.pl?language=fr&caller=summary&pub_date=06-06-14&numac=2006011243

Statbel (2017a) *Huishoudbudgetonderzoek 2016 – Methodologische Nota*, Statbel, Brussel.

https://statbel.fgov.be/sites/default/files/files/documents/Huishoudens/10.1%20Huishoudbudget/Methodologische%20nota%20HBS%202016%20NL_finaal.pdf

Statbel (2017b) *Enquête sur le budget des ménages 2016 – Note Méthodologique*, Statbel, Brussel.

https://statbel.fgov.be/sites/default/files/files/documents/Huishoudens/10.1%20Huishoudbudget/Note%20methodologique%20HBS%202016%20FR_final.pdf

Statbel/Delclite (2020) *Modification de la durée de l'enquête sur le budget des ménages - analyse de la transition à 15 jours*, Statbel, Bruxelles. - en cours de révision -

Termote, A. en Depickere, A. (2018a) *De hervormde enquête naar de arbeidskrachten in 2017*, *Statbel Analyse*, nr. 4.

https://statbel.fgov.be/sites/default/files/Over_Statbel_FR/Analyse_eak_2017_nl_20181220.pdf

Termote, A. et Depickere, A. (2018b) *Réforme de l'enquête sur les forces de travail en 2017*, *Statbel Analyse*, no. 4.

https://statbel.fgov.be/sites/default/files/Over_Statbel_FR/Analyse_eak_2017_fr_20181220.pdf

Vanderhoeft, C., Depickere, A. en Termote, A. (2021) *Calibrage de l'enquête à panel sur les forces de travail depuis 2017*, *Statbel Analyse*, nr. XX. - en cours de révision -

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

Références aux pages web pertinentes

i Page FAQ d'EBM:

<https://statbel.fgov.be/fr/ebm-faq>

ii Secteurs statistiques, Vade-mecum:

https://statbel.fgov.be/sites/default/files/files/opendata/Statistische%20sectoren/Secteur%20stat-FR_tcm326-174181.pdf.

iii Open data secteurs statistiques:

<https://statbel.fgov.be/fr/open-data/secteurs-statistiques>

iv Description secteurs statistiques:

http://nl.wikipedia.org/wiki/Statistische_sector

v Carnet des dépenses en ligne

<https://hbs.statdata.be/hbs/login.xhtml?jsessionid=59EB4B588AE1EA4F5B93FE1BC87F9728.worker12>

vi Carnet des dépenses papier:

https://statbel.fgov.be/sites/default/files/files/documents/Huishoudens/10.1%20Huishoudbudget/carnet2016_FR.pdf

vii Questionnaire ménage:

https://statbel.fgov.be/sites/default/files/files/documents/Huishoudens/10.1%20Huishoudbudget/Questionnaire_m%C3%A9nage_2018.pdf

viii Questionnaire ménage:

https://statbel.fgov.be/sites/default/files/files/documents/Huishoudens/10.1%20Huishoudbudget/Questionnaire_m%C3%A9nage_2018.pdf

ix EUROSTAT RAMON - Reference And Management Of Nomenclatures – European Classification of Individual Consumption according to Purpose (ECOICOP):

https://ec.europa.eu/eurostat/ramon/nomenclatures/index.cfm?TargetUrl=LST_CLS_DLD&StrNom=COICOP_5&StrLanguageCode=FR&StrLayoutCode=HIERARCHIC

x EU-SILC (European Union Statistics on Income and Living Conditions):

<https://statbel.fgov.be/fr/themes/menages/pauvrete-et-conditions-de-vie>

xi Chiffres de l'enquête sur le budget des ménages:

<https://statbel.fgov.be/fr/themes/menages/budget-des-menages/plus>

xii Questionnaire ménage:

https://statbel.fgov.be/sites/default/files/files/documents/Huishoudens/10.1%20Huishoudbudget/Questionnaire_m%C3%A9nage_2018.pdf

xiii Household Budget Surveys in the EU: Methodology and recommendations for harmonisation – 2003:

<https://ec.europa.eu/eurostat/documents/3859598/5875361/KS-BF-03-003-EN.PDF/42a95cc0-cb48-48c7-8d3a-dfc5fa265eff?version=1.0>

xiv Chiffres de l'enquête sur le budget des ménages:

<https://statbel.fgov.be/fr/themes/menages/budget-des-menages/plus>

xv Catalogue des données EBM pour demander les microdonnées pour la recherche:

<https://statbel.fgov.be/fr/propos-de-statbel/vie-privee/microdonnees-pour-la-recherche/catalogue-des-donnees>

xvi Demander des microdonnées pour la recherche

<https://statbel.fgov.be/fr/propos-de-statbel/vie-privee/microdonnees-pour-la-recherche>

xvii Microdonnées européennes pour EBM 2005, 2010 et 2015:

-
- <https://ec.europa.eu/eurostat/fr/web/microdata/household-budget-survey>
- xxiii Eurostat Statistics Explained: Glossaire: Enquête sur le budget des ménages (EBM):
[https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:Household_budget_survey_\(HBS\)/fr](https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:Household_budget_survey_(HBS)/fr)
- xxix Eurostat Statistics Explained: Glossaire: Ménage – statistiques sociales
https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:Household_social_statistics/fr
- xxx Eurostat Statistics Explained: Glossaire: Revenu disponible équivalent
https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:Equivalent_disposable_income/fr
- xxxi Adjusting household incomes: equivalence scales:
<http://www.oecd.org/els/soc/OECD-Note-EquivalenceScales.pdf>
- xxii EUROSTAT RAMON - Reference And Management Of Nomenclatures – Metadata European Classification of Individual Consumption according to Purpose (ECOICOP):
https://ec.europa.eu/eurostat/ramon/nomenclatures/index.cfm?TargetUrl=LST_NOM_DTL&StrNom=COICOP_5
- xxiii Banque Nationale de Belgique: Comptes nationaux/régionaux:
<https://www.nbb.be/fr/statistiques/comptes-nationauxregionaux>
- xxiv Statbel – Indice des prix à la consommation:
<https://statbel.fgov.be/fr/themes/prix-la-consommation/prix-la-consommation>
- xxv United Nations Statistical Division – COICOP revision:
https://unstats.un.org/unsd/class/revisions/coicop_revision.asp