

**Statbel (Direction générale Statistique – Statistics Belgium)**

**Direction Thématique Société**

**Collecte des données**

**Certification et Coordination méthodologique**

**Enquête sur le budget des ménages 2016**

Note méthodologique

**Personne de contact: Kelly Sabbe**

Statbel (Direction générale Statistique - Statistics Belgium)

Chaque jour ouvrable entre 9h et 16h.



## Table des matières

---

1. Introduction .....	4
2. Nombre de ménages participants .....	4
3. Les échantillons d' HBS 2016 T1-T4 .....	5
3.1. Introduction .....	5
3.2. Degré 1 : tirage des unités géographiques (ou UPE) .....	5
3.2.1. Cadre d'échantillonnage des UPE .....	5
3.2.2. Le plan de sondage pour le tirage des UPE .....	7
3.2.3. Répartition des UPE sur l'ensemble de l'année ou du semestre d'enquête .....	9
3.3. Degré 2 : tirage de ménages (USE) pour LFS .....	9
3.3.1. Cadre d'échantillonnage des USE .....	9
3.3.2. Le plan de sondage pour le tirage des ménages de LFS (USE) .....	10
3.4. Degré 3 : tirage des ménages (TSE) pour HBS .....	12
3.4.1. Le cadre d'échantillonnage des TSE, et la sélection des TSE .....	12
3.4.2. Le plan de sondage pour le tirage des ménages de HBS (TSE) .....	15
3.5. Tailles des échantillons .....	17
3.6. Recrutement postal complémentaire .....	21
4. Taux de réponse .....	22
4.1. Nombre de réponses et taux de réponse .....	22
4.2. Calibrage: considérations préparatoires .....	27
5. Calibrage .....	29
5.1. Calibrage de LFS 2016 .....	29
5.1.1. Motivation d'une révision du calibrage de LFS .....	29
5.1.2. Calibrage classique au niveau individuel .....	30
5.1.3. Nouveau calibrage, simultanément au niveau de l'individu et du ménage .....	32
5.2. Calibrage de HBS 2016 .....	35
5.2.1. Modèle de calibrage traditionnel pour HBS .....	36
5.2.2. Extension afin d'atteindre l'équilibre entre les trimestres .....	38
5.2.3. Extension pour réduire un biais .....	40
6. Programmes d'encodage, de contrôle, d'imputation .....	44

6.1.	Modernisation de la collecte des données .....	44
6.2.	Contrôles et imputation .....	45
6.2.1.	Détection des outliers (valeurs aberrantes) .....	45
6.2.2.	Imputation des dépenses énergétiques.....	46
6.2.3.	Revenus .....	51
6.2.4.	Imputed rent .....	52
1.	Quelques définitions .....	54
2.	Composition de l'échantillon de répondants (représentativité).....	55
2.1.	Distributions des ménages.....	55
2.2.	Distributions individuelles.....	56

## 1. Introduction

L'Enquête sur le Budget des Ménages (EBM ; Nous utilisons l'abréviation HBS - Household Budget Survey) a été réformée en 2012 dans le but de produire des chiffres de meilleure qualité, plus adaptés aux demandes de nos principaux utilisateurs et ce, dans de meilleurs délais après la période de collecte de données.

HBS 2014 a été organisée de la même manière que HBS 2012, avec quelques améliorations concernant le questionnaire (interviews réalisées en CAPI).

HBS 2016 a été organisée de la même manière que celle de 2014.

## 2. Nombre de ménages participants

Le nombre de ménages participants à l'enquête (biannuelle) a été de 4490 en 2016.

### 3. Les échantillons d' HBS 2016 T1-T4

#### 3.1. Introduction

Les échantillons de ménages – un par trimestre – pour HBS 2016 T1-T4 sont tirés à partir des échantillons de LFS. Les échantillons de LFS sont des *échantillons à deux degrés*. Pour HBS, une sélection complémentaire de ménages est effectuée dans l'échantillon de LFS. Cette sélection est décrite ci-dessous comme le *troisième degré*. Les échantillons pour 2016 T1-T4 sont donc présentés 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 Registre national (construit, voir paragraphe 3.2.1) et composé d'UPE qui sont des unités géographiques;
- Degré 2: tirage des *unités secondaires d'échantillonnage* (USE) d'un cadre d'échantillonnage dérivé du Registre national composé d'USE, qui sont les ménages *éligibles*, appartenant aux UPE sélectionnées au degré 1;
- Degré 3: tirage des *unités tertiaires d'échantillonnage* (TSE's) de l'échantillon du degré 2 de ménages; les unités tertiaires sont donc aussi des ménages.

Les échantillons de LFS résultent donc des tirages aux degrés 1 et 2. Les échantillons de HBS sont obtenus au degré 3. Nous remarquons que:

- Pour des raisons organisationnelles, les échantillons trimestriels de HBS 2016 T1-T4 sont tirés des échantillons de LFS de deux trimestres (6 mois) antérieurs:
  - HBS 2016 T1 de LFS 2015 T3,
  - HBS 2016 T2 de LFS 2015 T4,
  - HBS 2016 T3 de LFS 2016 T1,
  - HBS 2016 T4 de LFS 2016 T2,
- Sans tenir compte du taux de réponse à LFS quand on sélectionne les ménages pour HBS.

Ce chapitre 3 ne traite que des *échantillons bruts* ou *initiaux*: la non réponse n'est pas abordée ici.

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

##### 3.2.1. *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 nombre défini de ménages, qui représente la *taille du groupe*.

Les UPE des 4 échantillons trimestriels (pour LFS) de l'année d'enquête 2015 ont toutes été tirées au même moment, c'est-à-dire sous la forme d'un seul échantillon. Le cadre d'échantillonnage des UPE est dérivé d'une extraction du registre national datant du 30/08/2014, effectuée le 11/09/2014. Pour l'année d'enquête LFS 2016, en vue de l'introduction du panel LFS à partir du 3<sup>e</sup> trimestre de 2016, seules les UPE des échantillons des 1<sup>er</sup> et 2<sup>e</sup> trimestres ont été tirées en une fois. A cet effet, le cadre d'échantillonnage a été dérivé d'une extraction datant du 10/10/2015, effectuée le 14/10/2015. Remarquez que ces extractions datent de trois à quatre mois avant le début des périodes référence (respectivement 2015 T1-T4 et 2016 T1-T2) des enquêtes LFS. C'est nécessaire car les UPE sélectionnées doivent être connues à temps pour recruter les enquêteurs.

Le cadre d'échantillonnage des UPE (pour une période de référence LFS donnée) est construit de la façon suivante. Les *secteurs statistiques* (au sein d'une commune) sont identifiés par un code à quatre signes (lettres, chiffres, tiret): p.ex. A101, A10-, ...<sup>1</sup> 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-, .... 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 les sections qui contiennent un nombre suffisant de ménages privés pour créer au moins un groupe de ménages pour LFS La taille du groupe est une valeur fixée au préalable. Pour les enquêtes de 2015 et 2016, cette valeur (au moment où les UPE sont construites et tirées) était de 23 en Flandre et en Wallonie et de 26 dans la 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*.<sup>2</sup> Les UPE sont les unités qui ressortent de ce processus. Les UPE "suffisamment grandes" sont finalement celles 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.

---

<sup>1</sup> 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, ...). Voir

[https://statbel.fgov.be/sites/default/files/files/opendata/Statistische%20sectoren/Secteur%20stat-FR\\_tcm326-174181.pdf](https://statbel.fgov.be/sites/default/files/files/opendata/Statistische%20sectoren/Secteur%20stat-FR_tcm326-174181.pdf),  
<https://statbel.fgov.be/fr/open-data/secteurs-statistiques> en [http://nl.wikipedia.org/wiki/Statistische\\_sector](http://nl.wikipedia.org/wiki/Statistische_sector).

<sup>2</sup> Une partie de commune ou lettre (statistique) est une zone géographique 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 partie de commune avant fusion. Voir note de bas de page précédente.

Un code spécial “9999” est utilisé dans les extractions du Registre national pour indiquer que le secteur statistique du domicile du ménage n'est pas connu. Les secteurs *fictifs* 9999 (au sein de chaque commune), et dès lors aussi les sections *fictives* 99 (au sein de chaque commune), ne sont pas retenus dans le cadre d'échantillonnage des UPE.

Enfin, nous remarquons également que des sections insuffisamment grandes, qui subsistent après la méthode de groupement décrite ci-dessus et qui ne peuvent donc pas être fusionnées avec d'autres sections, ne sont pas non plus retenues dans le cadre d'échantillonnage des UPE.

### 3.2.2. Le plan de sondage pour le tirage des UPE

L'échantillon des UPE pour une période de référence LFS donnée 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 (9) communes germanophones. On a donc travaillé avec 12 strates, tant pour LFS 2015 T1-T4 que pour LFS 2016 T1-T2.
- [2] Au sein de chaque strate, les UPE sont classées selon la taille qui est, comme nous l'avons déjà dit, mesurée comme le nombre de ménages privés dans l'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 LFS) 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. (Etant entendu que chaque enquêteur peut éventuellement prendre en charge plusieurs groupes de ménages).

Le tableau 1 A fournit les informations suivantes pour LFS 2015 et LFS 2016: les strates, la taille du groupe en fonction des strates et les nombres de sélections d'UPE par strate et *par trimestre*. Pour LFS 2015, le tableau ne mentionne que les sélections d'UPE pour T3 et T4, mais celles pour T1 et T2 sont identiques à celles pour T3 et T4, si bien que le nombre de sélections d'UPE par strate pour l'ensemble de l'année correspond à quatre fois le nombre par trimestre. Pour LFS 2016, le tableau mentionne les nombres pour T1 et T2. Les nombres pour T3 et T4 sont inférieurs à la suite de l'introduction du panel mais ne sont pas abordés dans ce rapport.

**Tableau 1.A. Stratification des UPE au degré 1**

Strate			Taille du groupe (a priori)	Nombre de sélections d'UPE ou groupes de ménages par trimestre			
				LFS 2015		LFS 2016	
				T3	T4	T1	T2
1	ANT	Anvers	23	70	70	70	70
2	BRU	Région de Bruxelles-Capitale	26	91	91	91	91
3	WVL	Flandre occidentale	23	58	58	58	58
4	OVL	Flandre orientale	23	63	63	63	63
5	HAI	Hainaut	23	62	62	62	62
6	LIE	Liège (francophone)	23	54	54	54	54
7	LIM	Limbourg	23	47	47	47	47
8	LUX	Luxembourg	23	34	34	34	34
9	NAM	Namur	23	36	36	36	36
10	VLB	Brabant flamand	23	54	54	54	54
11	BWA	Brabant wallon	23	34	34	34	34
12	DEU	Communauté germanophone	23	21	21	21	21
<b>Nombre total de groupes par trimestre :</b>				<b>624</b>	<b>624</b>	<b>624</b>	<b>624</b>

[4] Enfin, au sein de chaque strate, on tire un échantillon systématique avec un point de départ aléatoire et des probabilités de sélection proportionnelles à la taille des UPE. La taille de l'échantillon de chaque strate est le nombre de sélections d'UPE *par an* pour 2015 et pour le *premier semestre* pour 2016, tel que discuté sous [3] Quelques caractéristiques d'un tirage de ce type:



- 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 être sélectionnées plusieurs fois.
- 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.

Les grandes UPE peuvent ainsi être sélectionnées plusieurs fois. En d'autres termes, une UPE peut éventuellement correspondre à plusieurs sélections.

### **3.2.3. Répartition des UPE sur l'ensemble de l'année ou du semestre d'enquête**

Les différentes sélections d'UPE pour 2015 sont réparties de manière uniforme entre les 52 (ou 53) semaines de référence de l'ensemble de l'année d'enquête LFS 2015. Les différentes sélections d'UPE pour 2016 T1-T2 sont réparties de manière uniforme entre les 26 semaines de référence au premier semestre 2016. Plusieurs sélections d'une même UPE sont réparties au maximum entre les trimestres d'une année d'enquête LFS. La représentativité des UPE sélectionnées par trimestre est ainsi garantie au maximum.

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

## **3.3. Degré 2 : tirage de ménages (USE) pour LFS**

### **3.3.1. Cadre d'échantillonnage des USE**

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

- Pour LFS 2015 T3 (et HBS 2016 T1) : extraction datant du 16/05/2015
- Pour LFS 2015 T4 (et HBS 2016 T2) : extraction datant du 15/08/2015
- Pour LFS 2016 T1 (et HBS 2016 T3) : extraction datant du 14/11/2015
- Pour LFS 2016 T2 (et HBS 2016 T4) : extraction datant du 05/03/2016

Le cadre d'échantillonnage de ménages (USE) pour un trimestre donné, dérivé de l'extraction correspondante du registre national 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.
- [3] Les ménages, qui ont été sélectionnés dans un échantillon LFS 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 au moins l'âge de 15 ans et au maximum l'âge de 76 ans. 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 3.2.3 pour l'UPE à laquelle le ménage appartient.

### 3.3.2. Le plan de sondage pour le tirage des ménages de LFS (USE)

L'échantillon des ménages de LFS (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 et du nombre de fois que l'UPE est sélectionnée pour le trimestre donné.
- [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), les ménages tirés seront répartis de manière aléatoire et uniforme entre les sélections d'UPE. Un nombre identique de ménages sera de ce fait attribué à chaque sélection d'UPE. On appelle cela un groupe de ménages (LFS) et le nombre de ménages dans le groupe est la taille du groupe.

**Tableau 1.B. Tailles des groupes lors du tirage des ménages et tailles initiales et totales des échantillons**

Strate			Taille du groupe lors du tirage des MÉN	Nombre de sélections d'UPE ou groupes de ménages par trimestre			
				LFS 2015		LFS 2016	
				T3	T4	T1	T2
1	ANT	Anvers	20	70	70	70	70
2	BRU	Région de Bruxelles-Capitale	23	91	91	91	91
3	WVL	Flandre occidentale	20	58	58	58	58
4	OVL	Flandre orientale	20	63	63	63	63
5	HAI	Hainaut	20	62	62	62	62
6	LIE	Liège (francophone)	20	54	54	54	54
7	LIM	Limbourg	20	47	47	47	47
8	LUX	Luxembourg	20	34	34	34	34
9	NAM	Namur	20	36	36	36	36
10	VLB	Brabant flamand	20	54	54	54	54
11	BWA	Brabant wallon	20	34	34	34	34
12	DEU	Communauté germanophone	23	21	21	21	21
<b>Nombre total de ménages par trimestre :</b>				<b>12816</b>	<b>12816</b>	<b>12816</b>	<b>12816</b>

Au paragraphe 3.2.1, nous avons mentionné la taille du groupe qui est utilisée pour la construction des UPE. 26 pour la Région de Bruxelles-Capitale et 23 pour la Flandre et la Wallonie (voir tableau 1.A). Normalement, nous utilisons les mêmes tailles de groupes pour le tirage des ménages (degré 2) mais, pour les trimestres 2015 T3-T4 et 2016 T1-T2, nous avons finalement dérogé à cette méthode pour des raisons budgétaires. La taille du groupe a été réduite de 3 ménages pour ces périodes, sauf pour la strate de la Communauté germanophone. Dans le tableau 1.B, nous répétons les strates, le nombre de groupes par strate et par trimestre, comme

dans le tableau 1.A, mais la taille des groupes a été adaptée. La dernière ligne donne les tailles totales initiales des échantillons qui en résultent (en nombre de ménages) par trimestre. (Si les tailles initiales des groupes avaient été maintenues, le nombre total de ménages initialement tirés par trimestre aurait été de 14625, comme ce fut le cas, par exemple, en 2015 T1 et T2.) Il est à noter que la taille des groupes ne joue aucun rôle dans la sélection des UPE (en degré 1) ; la taille des groupes joue un rôle dans la construction de la base de sondage des UPE, mais comme la taille des groupes a été réduite entre les degrés 1 et 2, cela n'a aucune incidence négative sur la capacité de créer des groupes à la taille souhaitée (c'est-à-dire, finalement, 20 ou 23).

### 3.4. Degré 3 : tirage des ménages (TSE) pour HBS

#### **3.4.1. Le cadre d'échantillonnage des TSE, et la sélection des TSE**

Pour HBS, en principe, tous les ménages tirés dans le cadre de LFS (degré 2) sont éligibles, à l'exception des ménages des communes germanophones : pour HBS, aucun ménage de la Communauté germanophone n'est sélectionné. Pour chaque trimestre HBS, l'échantillon LFS initial de ménages, à l'exception des communes germanophones, constitue le cadre d'échantillonnage pour le tirage des ménages (TSE) au degré 3.

Pour chacun des trois premiers trimestres de HBS, il a été décidé de n'utiliser que les groupes de ménages tirés pour LFS attribués aux sept premières semaines du trimestre : on s'attendait à observer le taux de réponse d' HBS 2014 pour HBS 2016, et l'objectif pour HBS 2016 était un total de 5.000 ménages répondants. Toutefois, pour le quatrième trimestre, tous les ménages de LFS ont été jugés éligibles. Nous abordons maintenant cela plus en détail pour chacun des quatre trimestres.

Pour le premier trimestre de HBS (2016 T1), c'est-à-dire le trimestre 2015 T3 de LFS, 624 :13 = 48 groupes de ménages ont été attribués à chaque semaine. Sans les 1 ou 2 groupes par semaine pour la Communauté germanophone, cela donne 47 ou 46 groupes de ménages par semaine (semaine 1-7), comme indiqué dans la colonne intitulée " HBS T1" du tableau 2.A. Ces 47 ou 46 groupes par semaine de référence de LFS sont ensuite attribués aux mois de référence de HBS, selon un système qui aboutit finalement à environ 1/3 du nombre total de groupes (327) pour chaque mois de HBS. Ce système consiste à attribuer les groupes des premières semaines de LFS autant que possible au premier mois de HBS, les groupes des dernières semaines de LFS autant que possible au troisième mois de HBS et les groupes des semaines du milieu de LFS le plus possible au deuxième mois de HBS. Par conséquent, les groupes de la troisième semaine de LFS sont attribués aux premier et deuxième mois de HBS selon un ratio d'environ 1:2, et les groupes de la cinquième semaine de LFS sont attribués aux deuxième et troisième mois de HBS selon un ratio d'environ 2:1. Les nombres de groupes qui en résultent pour le premier trimestre de HBS, par combinaison de semaine de LFS et de mois de HBS, est présenté au tableau 2.A dans les colonnes intitulées "Mois HBS 1, 2 ou 3".

Le deuxième volet du tableau 2.A montre comment les groupes de LFS des semaines 40-46 du trimestre 2015 T4 (c'est-à-dire les 7 premières semaines de ce trimestre) ont été attribués aux mois 4, 5 et 6 de HBS. Les chiffres sont ici légèrement inférieurs à ceux du premier volet du

tableau, en raison du fait que les 624 groupes de LFS ont été répartis sur 14 semaines (au lieu des 13 semaines habituelles) au cours du trimestre ( $624 : 14 = 44,57$ , soit 45 ou 44 groupes de LFS par semaine de référence, dont il faut déduire 1 ou 2 groupes par semaine pour la Communauté germanophone).

**Tableau 2.A. Conversion des semaines de référence (LFS) en mois d'enquête (HBS) – Deux premiers trimestres**

LFS 2015 T3 → HBS 2016 T1					LFS 2015 T4 → HBS 2016 T2				
LFS Semaine de référence	Mois HBS			HBS T1	LFS Semaine de référence	Mois HBS			HBS T2
	1	2	3			4	5	6	
27	47			47	40	44			44
28	47			47	41	43			43
29	15	32		47	42	15	28		43
30		47		47	43		43		43
31		31	16	47	44		28	15	43
32			46	46	45			43	43
33			46	46	46			43	43
34				0	47				0
35				0	48				0
36				0	49				0
37				0	50				0
38				0	51				0
39				0	52				0
-				-	53				0
<b>Total</b>	<b>109</b>	<b>110</b>	<b>108</b>	<b>327</b>	<b>Total</b>	<b>102</b>	<b>99</b>	<b>101</b>	<b>302</b>

**Tableau 2.B. Conversion des semaines de référence (LFS) en mois d'enquête (HBS) - Deux derniers trimestres**

LFS 2016 T1 → HBS 2016 T3					LFS 2016 T2 → HBS 2016 T4				
LFS Semaine de référence	Mois HBS			HBS T3	LFS Semaine de référence	Mois HBS			HBS T4
	7	8	9			10	11	12	
1	47			47	14	47			47
2	47			47	15	47			47
3	15	32		47	16	15	32		47
4		47		47	17		47		47
5		31	16	47	18		31	16	47
6			46	46	19			46	46
7			46	46	20			46	46
8				0	21	46			46
9				0	22	46			46
10				0	23		46		46
11				0	24		46		46
12				0	25			46	46
13				0	26			46	46
<b>Total</b>	<b>109</b>	<b>110</b>	<b>108</b>	<b>327</b>	<b>Total</b>	<b>201</b>	<b>202</b>	<b>200</b>	<b>603</b>

Pour le troisième trimestre de HBS (2016 T3), la méthode est entièrement analogue à celle du premier trimestre de HBS : le premier volet du tableau 2B est analogue au premier volet du tableau 2A.

Pour le quatrième trimestre de HBS (2016 T4), la méthodologie était initialement aussi similaire à celle utilisée pour les premier et troisième trimestres, mais il a toutefois été décidé par la suite d'utiliser tous les groupes de LFS 2016 T2. Le volet 2 du tableau 2 B montre comment les groupes de LFS des semaines 21 à 26 ont été attribués aux mois 10, 11 et 12 de HBS. La raison de cette extension de l'échantillon pour HBS 2016 T4 était le taux de réponse plus faible que prévu

(environ 20 %, comme observé en 2014) au cours des premiers trimestres (2016 T1 et T2), combiné à un objectif de 5.000 ménages répondants pour l'année d'enquête 2016 complète.

### 3.4.2. Le plan de sondage pour le tirage des ménages de HBS (TSE)

Pour chaque trimestre de HBS, nous pouvons finalement réorganiser les groupes de ménages sélectionnés, tels qu'indiqués dans les tableaux 2.A et 2.B, selon la strate et le mois de HBS. On aboutit ainsi aux tableaux 3.A et 3.B.

**Tableau 3.A. Stratification du degré 3 et allocation des groupes LFS - Deux premiers trimestres**

		HBS 2016 T1				HBS 2016 T2			
Strate UPE		Mois HBS			T1	Mois HBS			T2
		1	2	3		4	5	6	
1	ANT	14	14	12	40	12	11	12	35
2	BRU	16	16	17	49	16	17	16	49
3	WVL	9	10	11	30	10	8	10	28
4	OVL	12	11	12	35	11	10	9	30
5	HAI	12	12	10	34	10	11	12	33
6	LIE	9	9	11	29	9	10	8	27
7	LIM	9	8	7	24	7	7	8	22
8	LUX	5	6	7	18	7	5	5	17
9	NAM	7	7	7	21	5	6	7	18
10	VLB	9	10	9	28	9	10	9	28
11	BWA	7	7	5	19	6	4	5	15
<b>Total</b>		<b>109</b>	<b>110</b>	<b>108</b>	<b>327</b>	<b>102</b>	<b>99</b>	<b>101</b>	<b>302</b>

Observez maintenant le premier volet du tableau 3.A pour le trimestre HBS 2016 T1. La colonne intitulée "T1" montre le nombre de groupes LFS de ménages sélectionnés au sein des 624 groupes au total (pour le trimestre LFS 2015 T3). La procédure de sélection a été développée au paragraphe précédent 3.4.1. Il s'est avéré que la procédure n'était pas du tout aléatoire. Cependant, nous supposerons dorénavant, par exemple, que les 40 groupes de ménages de la

strate ANT (tableau 3.A, volet 1) ont été tirés de manière aléatoire parmi les 70 groupes de ménages originaux de la strate ANT (tableau 1.A ou tableau 1.B). En effet, ces 40 groupes d'ANT sont tous des groupes d'ANT qui ont été attribués aux 7 premières semaines du trimestre T3 de LFS 2015, mais, comme les 70 groupes d'ANT ont été en principe attribués de façon aléatoire aux 13 semaines du trimestre T3 de 2015, nous pouvons considérer les 40 groupes sélectionnés pour HBS 2016 T1 dans ANT comme un tirage aléatoire parmi 70 groupes. Il en va de même pour chacune des autres strates (BRU, WV, ... BWA) et pour chaque trimestre de HBS.

**Tableau 3.B. Stratification au degré 3 et allocation des groupes LFS - Deux derniers trimestres**

Strate UPE		HBS 2016 T3				HBS 2016 T4			
		Mois HBS			T3	Mois HBS			T4
		7	8	9		10	11	12	
1	ANT	14	14	12	40	24	24	22	70
2	BRU	16	16	17	49	30	30	31	91
3	WV	9	10	11	30	19	20	19	58
4	OVL	12	11	12	35	22	19	22	63
5	HAI	12	12	10	34	20	22	20	62
6	LIE	9	9	11	29	18	17	19	54
7	LIM	9	8	7	24	16	16	15	47
8	LUX	5	6	7	18	11	12	11	34
9	NAM	7	7	7	21	12	11	13	36
10	VLB	9	10	9	28	18	19	17	54
11	BWA	7	7	5	19	11	12	11	34
<b>Total</b>		<b>109</b>	<b>110</b>	<b>108</b>	<b>327</b>	<b>201</b>	<b>202</b>	<b>200</b>	<b>603</b>

Cela signifie que le tirage du degré 3, pour chaque trimestre séparément, peut être vu comme un échantillon en grappes aléatoire stratifié de ménages ; les grappes sont les groupes de ménages de LFS tels que formés au degré 2.

Remarquez que le tirage de HBS 2016 T4 à partir de LFS 2016 T2 est exhaustif.

Les considérations de ce paragraphe concernant le caractère aléatoire du tirage au degré 3 sont importantes pour le calcul correct des pondérations de l'échantillon de ménages dans l'échantillon



HBS (ce qui est important pour le calibrage) et pour l'estimation des variances des indicateurs estimés.

### 3.5. Tailles des échantillons

Les tailles brutes des échantillons, tant en nombre de ménages qu'en nombre de groupes de ménages, par strate (d'UPE) par trimestre, sont résumées dans les tableaux suivants.

**Tableau 4.A. Echantillons bruts de ménages de LFS et d' HBS par rapport à HBS 2016 T1**

LFS 2015 T3				HBS 2016 T1		
Strate	Taille du groupe	Nombre groupes	Nombre MEN	Nombre de groupes LFS	Nombre MEN	
1	ANT	20	70	1400	40	800
2	BRU	23	91	2093	49	1127
3	WVL	20	58	1160	30	600
4	OVL	20	63	1260	35	700
5	HAI	20	62	1240	34	680
6	LIE	20	54	1080	29	580
7	LIM	20	47	940	24	480
8	LUX	20	34	680	18	360
9	NAM	20	36	720	21	420
10	VLB	20	54	1080	28	560
11	BWA	20	34	680	19	380
12	DEU	23	21	483	-	-
<b>Total</b>		<b>624</b>	<b>12816</b>	<b>327</b>	<b>6687</b>	

**Tableau 4.B. Echantillons bruts de ménages de LFS et d' HBS par rapport à HBS 2016 T2**

LFS 2015 T4				HBS 2016 T2		
Strate		Taille du groupe	Nombre groupes	Nombre MEN	Nombre de groupes LFS	Nombre MEN
1	ANT	20	70	1400	35	700
2	BRU	23	91	2093	49	1127
3	WVL	20	58	1160	28	560
4	OVL	20	63	1260	30	600
5	HAI	20	62	1240	33	660
6	LIE	20	54	1080	27	540
7	LIM	20	47	940	22	440
8	LUX	20	34	680	17	340
9	NAM	20	36	720	18	360
10	VLB	20	54	1080	28	560
11	BWA	20	34	680	15	300
12	DEU	23	21	483	-	-
<b>Total</b>			<b>624</b>	<b>12816</b>	<b>302</b>	<b>6187</b>

**Tableau 4.C. Echantillons bruts de ménages de LFS et d' HBS par rapport à HBS 2016 T3**

LFS 2016 T1				HBS 2016 T3		
Strate		Taille du groupe	Nombre groupes	Nombre MEN	Nombre de groupes LFS	Nombre MEN
1	ANT	20	70	1400	40	800
2	BRU	23	91	2093	49	1127
3	WVL	20	58	1160	30	600
4	OVL	20	63	1260	35	700
5	HAI	20	62	1240	34	680
6	LIE	20	54	1080	29	580
7	LIM	20	47	940	24	480
8	LUX	20	34	680	18	360
9	NAM	20	36	720	21	420
10	VLB	20	54	1080	28	560
11	BWA	20	34	680	19	380
12	DEU	23	21	483	-	-
<b>Total</b>			<b>624</b>	<b>12816</b>	<b>327</b>	<b>6687</b>

**Tableau 4.D. Echantillons bruts de ménages de LFS et d' HBS par rapport à HBS 2016 T4**

LFS 2016 T2				HBS 2016 T4		
Strate		Taille du groupe	Nombre groupes	Nombre MEN	Nombre de groupes LFS	Nombre MEN
1	ANT	20	70	1400	70	1400
2	BRU	23	91	2093	91	2093
3	WVL	20	58	1160	58	1160
4	OVL	20	63	1260	63	1260
5	HAI	20	62	1240	62	1240
6	LIE	20	54	1080	54	1080
7	LIM	20	47	940	47	940
8	LUX	20	34	680	34	680
9	NAM	20	36	720	36	720
10	VLB	20	54	1080	54	1080
11	BWA	20	34	680	34	680
12	DEU	23	21	483	-	-
<b>Total</b>			<b>624</b>	<b>12816</b>	<b>603</b>	<b>12333</b>

### 3.6. Recrutement postal complémentaire

La sélection au degré 3 de (groupes de) ménages à partir des échantillons de ménages de LFS, telle que décrite ci-dessus, avait pour objectif final d'atteindre environ 5.000 ménages répondant à HBS, et a implicitement utilisé un taux de réponse (conditionnel) d'environ 20 % observé en 2014. Ceci explique la décision initiale de limiter le recrutement de HBS aux ménages de LFS sélectionnés pour les 7 premières semaines de chaque trimestre. En effet, selon les prévisions, cette procédure devait permettre d'atteindre  $(6687 \times 3 + 6187) \times 20\% \approx 5250$  ménages répondant à HBS.

Après le premier trimestre, il est devenu évident que l'objectif de 5.000 ménages n'était pas réalisable avec le plan d'échantillonnage initial et le taux de réponse prévu de 20 %. Afin de répondre à la pénurie de ménages, il a été décidé de (1) recruter tous les ménages de LFS de 2016 T2 pour le 4ème trimestre de HBS (comme discuté précédemment), et (2) d'organiser un recrutement postal supplémentaire. Les ménages des échantillons de LFS de 2015 T3, 2015 T4 et 2016 T1, appartenant à des groupes des semaines 8 à 13 (ou 14) et ayant effectivement participé à l'enquête LFS, ont été invités par courrier à participer à HBS.

Le tableau 5 présente un aperçu complet des sources des quatre échantillons trimestriels de HBS.

**Tableau 5. Sources et modes de recrutement des échantillons trimestriels de HBS 2016**

<b>Mode de recrutement</b>	<b>HBS 2016 T1</b>	<b>HBS 2016 T2</b>	<b>HBS 2016 T3</b>	<b>HBS 2016 T4</b>
<b>Face à face (FàF)</b>	LFS 2015 T3 s1–s7	LFS 2015 T4 s1–s7	LFS 2016 T1 s1–s7	LFS 2016 T1 s1–s13
<b>Par voie postale</b>	--	--	LFS 2015 T3 s8–s13 LFS 2015 T4 s8–s14	LFS 2016 T1 s8–s13

## 4. Taux de réponse

### 4.1. Nombre de réponses et taux de réponse

L'enquête HBS2016 est une enquête intégrée à l'Enquête sur les Forces de Travail (EFT ou LFS – Labour Force Survey). Nous entendons par *enquête intégrée*, la sélection de ménages parmi les ménages qui ont répondu à une autre enquête – dans le cas du HBS, il s'agit donc de LFS – et qui ont accepté, sur base volontaire, de collaborer à l'enquête HBS après y avoir été invités par un enquêteur LFS lors d'un contact personnel (en face à face) ou par lettre, quelques semaines après la période d'enquête LFS. Le plan de sondage pour l'enquête intégrée a été décrit au chapitre 3.

Comme il est indispensable d'un point de vue pratique de travailler avec des groupes de ménages, un échantillon en grappes simple de ménages a été tiré à partir de l'échantillon LFS (voir paragraphe 3.4.2). Les grappes sont les groupes LFS qui sont sélectionnés de manière aléatoire. L'avantage est que tous les groupes LFS de ménages sont conservés comme groupes HBS. En principe, c'est l'enquêteur LFS qui continue donc de travailler avec un groupe de ménages qui habitent dans une zone géographique relativement restreinte. L'enquêteur HBS est de préférence le même enquêteur que celui qui a réalisé l'enquête LFS. Celui-ci est déjà connu par les ménages et a pris à cœur de faire participer à la nouvelle enquête le plus grand nombre possible de ménages. Toutefois, pour certains groupes, si l'enquêteur de LFS n'était pas disposé à réaliser également l'enquête HBS, il a fallu recruter un autre enquêteur. Le fait que, pour certains groupes, aucun nouvel enquêteur n'a pu être trouvé et qu'un nouvel enquêteur inspire moins confiance qu'un enquêteur déjà connu a exercé, dans une certaine mesure, une influence négative sur le taux de réponse.

Le tableau 6 reprend les nombres et les taux de réponse de HBS 2016. Il convient de noter que le tableau 6 présente les taux de réponse bruts, c'est-à-dire les taux de réponse incluant les groupes de ménages qui n'ont pas été réalisés. Si nous n'incluons pas ces ménages, nous obtenons les taux de réponse nets, qui seront bien sûr supérieurs aux taux bruts.

L'échantillon de HBS se compose uniquement de ménages qui ont participé auparavant à LFS pendant la période de 2015 T3 à 2016 T2, à l'exclusion des ménages des communes germanophones; donc:  $49.332 = 4 \times (12816 - 483)$ , voir tableaux 4.A-D. Le taux de réponse de HBS est une combinaison de 3 facteurs:

- participation à LFS,
- la méthode de recrutement après l'enquête LFS
- participation à HBS.

Le taux de réponse (« fichier HBS complet dans les tableaux 6 et 7 ») par rapport à l'échantillon LFS initial est de 9,1%. Cela signifie que sur 100 ménages qui entraient en ligne de compte pour participer à LFS et ensuite à HBS, 9 ménages ont finalement collaboré jusque la fin et leurs réponses ont été retenues après les contrôles de qualité. Le taux de réponse en 2016 a été beaucoup plus faible que les 17,0 % enregistrés en 2014 en raison du nombre plus élevé de refus et de la méthode mixte de recrutement.

**Tableau 6. Taux de réponse brut HBS par rapport à l'échantillon LFS initial (codes de réponse consolidés)**

	Flandre		Wallonie		Bruxelles		Total	
	N	%	N	%	N	%	N	%
<b>Dossier HBS Complet</b>	2.175	9,3	1.784	10,1	531	6,3	4.490	9,1
<b>Décédé(e)</b>	18	0,1	19	0,1	4	0,0	41	0,1
<b>A déménagé</b>	310	1,3	259	1,5	427	5,1	996	2,0
<b>Ménage collectif</b>	13	0,1	14	0,1	17	0,2	44	0,1
<b>Barrière de la langue</b>	87	0,4	33	0,2	63	0,8	183	0,4
<b>Vacances, voyage d'affaires</b>	280	1,2	223	1,3	168	2,0	671	1,4
<b>Malade, handicapé, dément</b>	157	0,7	74	0,4	25	0,3	256	0,5
<b>Adresse erronée</b>	194	0,8	115	0,7	70	0,8	379	0,8
<b>Refus</b>	16.086	68,9	11.638	66,1	5.630	67,2	33.354	67,6
<b>Pas au domicile</b>	1.601	6,9	1.735	9,9	891	10,6	4.227	8,6
<b>Autre</b>	412	1,8	419	2,4	307	3,7	1.138	2,3
<b>Adresse non visitée</b>	179	0,8	193	1,1	36	0,4	408	0,8
<b>HBS non validée</b>	28	0,1	52	0,3	19	0,2	99	0,2
<b>Aucune information</b>	0	0,0	22	0,0	0	2,2	22	0,0
<b>Pas expulsé</b>	1.820	7,8	1.020	5,8	184	2,2	3.024	6,1
<b>Total</b>	<b>23.360</b>	<b>100,0</b>	<b>17.600</b>	<b>100,0</b>	<b>8.372</b>	<b>100,0</b>	<b>49.332</b>	<b>100,0</b>

Le taux de réponse varie d'une **région** à l'autre. La Wallonie détenait le taux le plus élevé avec 10,1%. Le taux de réponse s'élevait à 9,3% en Flandre et à seulement 6,3% à Bruxelles.

La raison principale de l'attrition était les refus. En cours de processus, 67,6% des ménages ont déclaré qu'ils refusaient de participer à l'enquête. Dans la plupart des cas, le refus est intervenu après l'enquête LFS, ce qui est logique étant donné que LFS est une enquête obligatoire. De plus, 8,6% des ménages n'ont pas été retrouvés à la maison; ils constituent une proportion significative des ménages non contactés. Par manque d'enquêteurs disponibles, 6,1% des adresses n'ont pas pu être attribuées à un enquêteur actif.

Le taux de réponse par rapport à l'échantillon LFS initial diffère selon la **manière dont le recrutement a été effectué**; voir tableau 7.

**Tableau 7. Taux de réponse brut (%) HBS par rapport à l'échantillon LFS initial (codes de réponse consolidés) par mode de recrutement**

	Flandre		Wallonie		Bruxelles		Total	
	FàF	Courrier	FàF	Courrier	FàF	Courrier	FàF	Courrier
<b>Dossier HBS Complet</b>	10,7	6,8	12,4	6,0	7,2	4,7	10,7	6,1
<b>Décédé(e)</b>	0,1	0,1	0,1	0,1	0,0	0,1	0,1	0,1
<b>A déménagé</b>	1,2	1,6	1,5	1,5	5,0	5,4	1,9	2,2
<b>Ménage collectif</b>	0,1	0,0	0,1	0,0	0,3	0,0	0,1	0,0
<b>Barrière de la langue</b>	0,5	0,2	0,2	0,2	0,8	0,7	0,4	0,3
<b>Vacances, voyage d'affaires</b>	1,4	0,8	1,6	0,6	2,5	1,1	1,7	0,8
<b>Malade, handicapé, dément</b>	0,7	0,7	0,5	0,3	0,3	0,3	0,6	0,5
<b>Adresse erronée</b>	0,7	1,0	0,7	0,6	0,6	1,3	0,7	0,9
<b>Refus</b>	67,0	72,2	62,5	72,8	65,2	71,0	65,1	72,2
<b>Pas au domicile</b>	6,9	6,8	10,1	9,3	11,1	9,8	8,8	8,2
<b>Autre</b>	2,1	1,2	2,4	2,3	3,8	3,3	2,5	1,9



<b>Adresse non visitée</b>	0,8	0,7	1,4	0,6	0,4	0,6	0,9	0,6
<b>HBS non validée</b>	0,1	0,1	0,4	0,1	0,3	0,1	0,3	0,1
<b>Aucune information</b>	0,0	0,0	0,2	0,0	0,0	0,0	0,1	0,0
<b>Pas expulsé</b>	7,7	7,9	6,0	5,5	2,5	1,6	6,2	6,0
<b>Total (nombre)</b>	<b>14.980</b>	<b>8.380</b>	<b>11.440</b>	<b>6.160</b>	<b>5.474</b>	<b>2.898</b>	<b>31.894</b>	<b>17.438</b>

Pour une partie de l'échantillon, l'enquête a été réalisée, tout comme en 2014, **en face à face (FàF)** après la clôture de l'enquête LFS. Comme le nombre de recrutements positifs était plus faible qu'en 2014, il a été décidé de recruter aussi parmi les ménages LFS qui n'entraient initialement pas en ligne de compte pour HBS 2016 (les ménages LFS pour les semaines de référence 8 à 13 (ou 14) des trimestres LFS 2015 T3, 2015 T4 et 2016 T1; voir paragraphe 3.6. Ce recrutement s'est effectué par voie **postale**, en envoyant à ces ménages une lettre et un formulaire d'inscription.

Le recrutement en face à face permet d'atteindre le taux de réponse le plus élevé. En moyenne, 10,7% des ménages de l'échantillon LFS ont participé non seulement à LFS mais aussi à HBS. Ce chiffre est inférieur aux 17,0% de 2014, lorsque l'ensemble de l'échantillon HBS avait été recruté de cette manière. Pour les ménages recrutés par courrier, le taux de réponse par rapport à l'échantillon LFS initial d'adresses qui n'entraient initialement pas en ligne de compte pour HBS était de 6,1%, ce qui est comparable aux résultats d'avant 2012, lorsque l'on recrutait uniquement par courrier. Le fait que les ménages ont d'abord participé à LFS a donc eu un effet positif limité sur la volonté de participer à HBS.

Le tableau 8 ci-dessous montre les résultats de l'aspect recrutement, sans tenir compte du résultat final de HBS. N est le nombre de ménages qui ont participé à LFS et ont ensuite répondu « oui » à la question de savoir s'ils souhaitent participer à HBS après LFS, quel que soit le résultat final. Exemples : 7.061 ménage ont participé à LFS et ont ensuite déclaré vouloir participer à HBS lorsque l'enquêteur leur a demandé. Cela représente  $30,2\% \text{ des } 7.061/0,302 \approx 23.381$  ménages qui ont participé à LFS et auxquels la question de la participation à HBS a été posée par l'enquêteur LFS.

Lors du **recrutement en face à face**, 30,2% des ménages ont accepté la proposition de l'enquêteur LFS de participer également à HBS. Cette part était de 36,7% en 2014. La part plus faible s'explique en partie par la diminution de près de moitié de l'incentive pour les ménages participants. Il est frappant de constater les différences relativement importantes entre les régions. En Wallonie, 35,0% des ménages étaient disposés à participer à HBS contre seulement 26,5% en Flandre. Les différences entre régions sont moins grandes dans le cas du **recrutement postal** et en moyenne 11,5% des personnes se sont déclarées disposées à participer à HBS.

**Tableau 8. Recrutement après l'enquête LFS**

	Flandre		Wallonie		Bruxelles		Total	
	N	%	N	%	N	%	N	%
<b>Face à face</b>	2 949	26,5	2 923	35,0	1 189	30,1	7 061	30,2
<b>Par voie postale</b>	751	12,0	526	11,4	219	10,3	1 496	11,5

Les ménages HBS n'ont finalement pas tous participé; voir tableau 9. Le taux de réponse à l'enquêteur HBS s'élevait à 48,4% si le recrutement avait été effectué en face à face (63,4% en 2014) et 71,6% si le ménage était recruté par courrier. Ici aussi, on observe des différences entre les régions. Quel que soit le mode de recrutement, les ménages ayant collaboré jusque la fin sont plus nombreux en Flandre qu'en Wallonie et à Bruxelles. Les différences résultent principalement du plus grand nombre de refus.

**Tableau 9. Taux de réponse brut (%) HBS par rapport aux ménages recrutés par mode de recrutement**

	Flandre		Wallonie		Bruxelles		Total	
	FàF	Courrier	FàF	Courrier	FàF	Courrier	FàF	Courrier
<b>Dossier HBS Complet</b>	54,5	75,5	48,4	70,2	33,3	61,6	48,4	71,6
<b>Décédé(e)</b>	0,1	0,0	0,2	0,2	0,0	0,0	0,1	0,1
<b>A déménagé</b>	0,5	0,4	0,9	0,0	1,1	1,4	0,8	0,4
<b>Ménage collectif</b>	0,0	0,1	0,0	0,0	0,1	0,0	0,0	0,1
<b>Barrière de la langue</b>	0,4	0,1	0,0	0,4	0,6	0,0	0,3	0,2
<b>Vacances, voyage d'affaires</b>	1,1	0,8	1,1	0,6	3,6	3,2	1,5	1,1
<b>Malade, handicapé, dément</b>	1,2	0,8	0,7	0,4	0,8	0,9	0,9	0,7
<b>Adresse erronée</b>	0,1	0,3	0,1	0,0	0,2	0,0	0,1	0,1

<b>Refus</b>	33,9	14,6	40,3	22,8	47,7	23,7	38,8	18,9
<b>Pas au domicile</b>	1,3	0,8	2,6	1,3	5,0	3,2	2,4	1,3
<b>Autre</b>	3,2	1,1	3,2	2,1	4,8	3,2	3,5	1,7
<b>Pas visité</b>	3,0	4,7	1,1	1,1	1,5	0,9	1,9	2,9
<b>Dossier HBS non validé</b>	0,7	0,8	1,5	1,0	1,3	1,8	1,1	1,0
<b>Total (nombre)</b>	<b>2.949</b>	<b>751</b>	<b>2.923</b>	<b>526</b>	<b>1.189</b>	<b>219</b>	<b>7.061</b>	<b>1.496</b>

#### 4.2. Calibrage: considérations préparatoires

Dès le tirage, chaque ménage se voit attribuer un poids d'échantillonnage (*sampling weight*, *design weight*) qui est directement dérivé du plan de sondage et de la composition du cadre d'échantillonnage.<sup>3</sup>

Le plan de sondage implique que, dans chaque strate, tous les ménages sélectionnés ont tous (pratiquement) le même poids d'échantillonnage. Dans le tableau 10, nous donnons pour chaque trimestre le nombre de ménages répondants et la somme des poids d'échantillonnage des ménages répondants.

**Tableau 10. Nombre de ménages répondants et somme de leurs poids d'échantillonnage par trimestre**

	<b>HBS 2016 T1</b>	<b>HBS 2016 T2</b>	<b>HBS 2016 T3</b>	<b>HBS 2016 T4</b>	<b>Total</b>
<b>Nombre de ménages répondants</b>	908	783	1.365	1.434	<b>4.490</b>
<b>Somme des poids d'échantillonnage des ménages répondants</b>	320.388,55	275.587,79	473.681,65	500.050,58	<b>1.569.708,57</b>

Cela démontre que la part des trimestres dans les estimations pondérées pour l'ensemble de l'année d'enquête 2016 serait très déséquilibré, si l'on utilisait les poids d'échantillonnage. Les

<sup>3</sup> Nous ne penchons pas dans cette note sur le calcul des pondérations de l'échantillon. Il suffit de souligner ici que les poids d'échantillonnage sont désormais les poids dérivés directement du plan de sondage (selon la théorie du plan de sondage) mais qui ont aussi déjà été adaptés pour tenir compte, par exemple, de la suppression des ménages des groupes de ménages LFS des communes germanophones.

effets saisonniers pourraient être de cette manière renforcés ou atténués à tort. Nous verrons plus tard comment nous pouvons éviter cela au moyen d'équations de calibrage spéciales. Notons également que le total des poids de l'échantillon de tous les ménages répondants est une nette sous-estimation du nombre total de ménages privés en Belgique ; le premier objectif d'un calibrage est bien sûr d'éliminer ce problème, ce qui signifie que (en moyenne) tous les poids d'échantillon seront relevés.

Pour HBS 2016, aucune correction de la non-réponse n'a été effectuée avant le calibrage comme ce fut le cas pour HBS 2012 et 2014.

## 5. Calibrage

Traditionnellement, le calibrage de l'échantillon de répondants de HBS est réalisé à l'aide de valeurs de référence (benchmarks, totaux de calibrage) estimées à l'aide de LFS. Ces estimations de LFS sont également calculées après calibrage de LFS. Dans cette section, nous aborderons donc d'abord le calibrage de LFS 2016 (paragraphe 5.1). Nous décrivons ensuite le calibrage de LFS (paragraphe 5.2).

### 5.1. [Calibrage de LFS 2016](#)

#### 5.1.1. *Motivation d'une révision du calibrage de LFS*

Traditionnellement, le calibrage de LFS est réalisé tous les trimestres selon un modèle de post-stratification pour les répondants *individuels* de LFS. Les valeurs de référence ici sont les nombres d'*individus* selon le Registre national, dans différentes sous-populations (appelées post-strates qui sont déterminées par les variables de post-stratification ; voir plus loin au paragraphe 5.1.2). Par conséquent, les estimations de LFS sont, dans une certaine mesure, cohérentes avec les chiffres de population au niveau individuel tirés du Registre national. Ce calibrage trimestriel a encore été appliqué pour LFS 2016 T1-T4 pour la publication d'*estimations sur les individus*, comme les chiffres de l'emploi.

Dans le passé, les estimations de LFS *au niveau des ménages*, qui étaient utilisées comme valeurs de référence dans le calibrage de HBS, étaient également calculées sur la base des pondérations résultant de la post-stratification individuelle postérieure de LFS décrite ci-dessus. Cette méthode présente clairement trois inconvénients :

1. Les poids individuels calibrés de LFS doivent être convertis en poids ménage avant de pouvoir calculer les estimations souhaitées au niveau des ménages. Traditionnellement, cela se faisait de manière *ad hoc*, en utilisant comme poids ménage soit le poids de la personne de référence, soit la moyenne des poids des membres du ménage. Cette façon de déterminer le poids des ménages n'a pas le moindre fondement théorique.
2. Les estimations au niveau des ménages ne sont en aucun cas cohérentes avec les chiffres de population au niveau des ménages issus du Registre national. Par exemple, le nombre estimé de ménages pour l'ensemble de la Belgique peut être très différent (positivement ou négativement, et cela peut dépendre de l'année d'enquête) du nombre de ménages selon le Registre national. Les écarts peuvent même augmenter par région, par province, etc. Le nombre de ménages de  $x$  personnes (pour certains  $x = 1, 2, \text{ou } \dots$ ) est un autre exemple et, bien sûr, le problème s'aggrave si on combine la taille du ménage avec la région ou la province. En résumé, pour chaque "domaine" ou sous-population de ménages, l'écart entre l'estimation par LFS du nombre de ménages dans ce domaine et le nombre correspondant issu du Registre national augmentera à mesure que le domaine se réduit.

3. Cette incohérence au niveau des ménages se répercute bien entendu dans les estimations de HBS si les estimations incohérentes de LFS sont utilisées comme valeurs de référence pour le calibrage de HBS.

Ce dernier problème surtout nous oblige à d'abord revoir le calibrage de LFS. Dans le paragraphe 5.1.2, nous allons d'abord donner un rappel du calibrage traditionnel de LFS au niveau individuel. Puis, au paragraphe 5.1.3, nous discuterons du modèle de calibrage rénové de LFS.

### 5.1.2. Calibrage classique au niveau individuel

Pour le calibrage traditionnel de LFS, l'input de l'enquête est le suivant :

1. Une liste de tous les répondants individuels, avec les caractéristiques du domicile, de l'âge et du sexe qui seront intégrées dans le modèle de calibrage.
2. Pour chaque répondant : le poids de l'échantillon (calculé sur la base du plan de sondage et du cadre d'échantillonnage de LFS).

Les variables du domicile, de l'âge et du sexe permettent d'obtenir les trois variables de calibrage ou de post-stratification suivantes :

- STRAT12 : les 12 strates d'échantillonnage pour la sélection des UPE (voir le tableau 1.A au paragraphe 3.2.2) ;
- SEXE : sexe (2 catégories) ;
- AGE : 16 classes d'âge (0-4, 5-9, 10-14, ..., 70-74, 75+).

Le croisement complet de ces trois variables donne  $12 \times 2 \times 16 = 384$  post-strates. Pour chacune de ces post-strates, nous calculons (comptons !) le nombre d'individus issu du Registre national ; nous notons  $N$  pour chacun de ces 384 chiffres de population.

Dans chaque post-strate séparément, nous calculons le poids calibré individuellement comme suit :

- a. Déterminer le total  $\sum d$  des poids de l'échantillon des répondants de la post-strate.
- b. Déterminer le facteur de correction unique (ou *g-weight*)  $g = N / \sum d$  de la post-strate.
- c. Déterminer le poids calibré individuellement  $w = d \times g$  de chaque répondant de la post-strate.

Enfin, nous remarquons que ce modèle est utilisé pour chaque trimestre séparément. C'est notamment nécessaire parce que les estimations importantes de LFS sur des thèmes tels que l'emploi doivent être publiées le plus tôt possible après la fin de la collecte des données pour le trimestre. Cela implique également que nous utilisons chaque fois une nouvelle extraction du Registre national. Pour chaque trimestre, cette extraction est une "photographie" de la population à un moment donné au milieu du trimestre. Cette photographie n'est prise qu'après que les données de LFS pour le trimestre concerné aient été collectées et nettoyées, c'est-à-dire environ un mois après la fin du trimestre. A ce moment, les données du Registre national concernant ce

moment du passé sont également suffisamment mises à jour pour obtenir une image suffisamment correcte de la population.

En vue d'autres modèles de calibrage (pour LFS comme pour HBS), qui ne seront plus des modèles "simples" de post-stratification, nous notons le modèle classique comme suit :

$$\{\text{Ind; STRAT12} \times \text{SEX} \times \text{AGE}; -\}$$

Il est important de noter que si une post-strate n'est pas vide dans la population ( $N > 0$ ), elle ne devrait pas l'être non plus dans l'échantillon des répondants ( $\sum d > 0$ ). Cela n'a jamais été un problème pour LFS, parce que l'échantillon des répondants est toujours assez grand (l'échantillon initial est grand et le taux de réponse est suffisamment élevé en raison du caractère obligatoire de l'enquête). En vue d'échantillons plus réduits et d'autres modèles, nous remarquons déjà ici que l'on pourra rapidement passer du modèle  $\{\text{Ind} ; \text{STRAT12} \times \text{SEX} \times \text{AGE} ; -\}$  au modèle par exemple (purement illustratif) :

$$\{\text{Ind; STRAT12} \times (\text{SEX} + \text{AGE}); \text{Lin}\}$$

Nous entendons par là que :

- "Ind" indique que nous appliquons toujours un calibrage à un niveau purement individuel.
- "STRAT12  $\times$  SEXE  $\times$  ÂGE " dans le modèle de post-stratification signifie que nous calibrons l'échantillon sur la distribution jointe de la population (*joint population distribution*) des variables STRAT12, SEXE et AGE. "STRAT12  $\times$  (SEXE + ÂGE) " signifie que nous calibrerions sur les distributions marginales de SEXE et d'AGE dans chaque strate de STRAT12, c'est-à-dire simultanément sur la distribution jointe de STRAT12 et de SEXE et la distribution jointe de STRAT12 et AGE (car STRAT12  $\times$  (SEXE + AGE) = STRAT12  $\times$  SEXE + STRAT12  $\times$  AGE).
- "Lin" indique un certain choix qui peut et doit être fait dans le cas de modèles qui ne sont pas des modèles de post-stratification. "Lin" signifie que le modèle choisi pour les facteurs de correction  $g$  est linéaire dans les variables de calibrage<sup>4</sup>. D'autres choix sont possibles : "Exp" pour une relation exponentielle (ou multiplicatrice), "Logit" pour une relation logistique ou "TrLin" pour une relation linéaire tronquée. On choisit parfois Exp pour garantir des facteurs de correction positifs. Les choix Logit et TrLin sont prévus pour limiter les facteurs de correction vers le haut et vers le bas (par exemple à 0). Les choix Lin, Exp, Logit et TrLin sont également traduits sans ambiguïté en mesures de distance entre les poids initiaux (d'échantillon)  $d$  et les poids calibrés  $w$ . Pour les modèles de post-stratification, ce choix ne joue aucun rôle ! En d'autres termes, chaque choix conduit, dans le cas d'un modèle post-stratification, aux mêmes facteurs de correction  $g$ .

Nous pouvons aussi formuler ce modèle alternatif de la manière suivante:

$$\{\text{Ind; STRAT12; SEX} + \text{AGE}; \text{Lin}\}$$

---

<sup>4</sup> Tout comme un modèle de régression peut être linéaire pour les variables explicatives.

qui indique qu'au sein de chaque strate STRAT12, le calibrage est effectué selon la distribution marginale de la population de SEX et la distribution marginale de la population de AGE. Nous pouvons utiliser cette notation pour le modèle classique de post-stratification:

$$\{Ind; STRAT12; SEX \times AGE; Lin\}$$

Cette notation facilitera le passage à des modèles de calibrage simultané au niveau de l'individu et du ménage (paragraphe 5.1.3) ! Séparer la variable de calibrage STRAT12 lui confère un rôle particulier : le calcul des facteurs de correction peut alors être effectué séparément dans chaque catégorie STRAT12. Une variable qui peut être ainsi isolée s'appelle une *variable de stratification de calibrage* ; les catégories correspondantes de la variable s'appellent les *strates de calibrage*. Notez que les strates de calibrage ne doivent pas nécessairement être les strates de l'échantillon (si elles existent déjà).

Les propriétés importantes du modèle de post-stratification, ainsi que, par exemple, du modèle alternatif qui en est dérivé et qui a été brièvement abordé plus haut, sont :

- I. Les individus d'un même ménage auront généralement des facteurs de correction différents  $g$  si leurs variables de calibrage prennent des valeurs différentes. Comme le plan de sondage de LFS implique que le poids de l'échantillon  $d$  pour chaque individu d'un même ménage ait la même valeur, les poids calibrés  $w$  des individus d'un même ménage différeront donc généralement entre eux.
- II. Par conséquent, aucun poids de ménage calibré n'est déterminé sans ambiguïté. C'est pourquoi nous avons parfois recours à des solutions *ad hoc* pour cet aspect. Cela joue également un rôle dans l'incohérence des estimations de LFS au niveau des ménages, lorsque LFS est calibrée au niveau individuel.
- III. Tous les individus ayant les mêmes valeurs pour toutes les variables de calibrage, quel que soit le ménage auquel ils appartiennent, ont le même facteur de correction  $g$ . Mais cela ne veut pas dire qu'ils auront aussi la même  $w$ , parce qu'ils peuvent avoir une  $d$  différente et parce que  $w = d \times g$ .

Comme nous l'avons mentionné précédemment au paragraphe 5.1.1, la propriété II est la principale raison de la révision du modèle de calibrage de LFS.

### 5.1.3. Nouveau calibrage, simultanément au niveau de l'individu et du ménage

Le nouveau calibrage de LFS, que nous avons effectué pour le calibrage de HBS, utilise les données suivantes de LFS :

1. Nous utilisons toujours une liste de tous les répondants individuels, avec les caractéristiques âge et sexe qui seront intégrées dans le modèle de calibrage.
2. De plus, nous utilisons également une liste de tous les ménages répondants, avec les caractéristiques résidence et nombre de membres de la famille. Il est à noter que cette deuxième liste peut être dérivée de la première (étant donné que tous les membres d'un ménage répondant sont aussi des répondants individuels).



3. Pour chaque ménage répondant : le poids de l'échantillon  $d$  (calculé sur la base du plan de sondage et du cadre d'échantillonnage de LFS). Il est à noter que le plan de sondage de LFS est tel que le poids de l'échantillon de chaque individu ( $\bar{d}$ , disons) dans un ménage donné est égal au poids de l'échantillon du ménage ; donc  $\bar{d} = d$  au sein de chaque ménage. (Ce n'est pas essentiel selon la théorie du calibrage, mais cela facilite la discussion ici).

Comme au paragraphe 5.1.2, les variables individuelles âge et sexe permettent d'obtenir les deux variables de calibrage *individuelles* suivantes :

- SEXE : sexe (2 catégories) ;
- AGE : 16 classes d'âge (0-4, 5-9, 10-14, ..., 70-74, 75+).

et, en outre, les variables "domicile " et " taille du ménage " permettent d'obtenir les deux variables de calibrage du *ménage* suivantes :

- STRAT12 : les 12 strates d'échantillonnage pour la sélection des UPE (voir le tableau 1.A au paragraphe 3.2.2) ;
- HHSIZE: 6 classes de taille (1-, 2-, 3-, 4-, 5- et 6- ou ménage multiple).

A partir de ces variables de calibrage, nous pouvons calculer (compter !) les chiffres suivants à partir du Registre national :

- Par strate (STRAT12), le nombre d'individus dans chaque sous-population d'individus est déterminé par la combinaison SEX et AGE. En d'autres termes, nous continuerons (tout comme dans le modèle classique de post-stratification pour les individus) à utiliser les *nombre d'individus* dans chaque cellule du croisement STRAT12 x SEX x AGE comme valeurs de référence dans le nouveau calibrage.
- Par strate (STRAT12), le nombre de ménages dans chaque sous-population de ménages défini par les catégories HHSIZE. En d'autres termes, nous utiliserons de plus les *nombre de ménages* dans chaque cellule de la STRAT12 x HHSIZE comme valeurs de référence dans le nouveau calibrage.

Nous pouvons encore noter  $N$  pour chacun des 384 chiffres de population individuels, et nous notons maintenant aussi  $M$  pour chacun des  $12 \times 6 = 72$  chiffres de population des ménages. Mais la façon de calculer les facteurs de correction  $g$  n'est plus aussi facile à décrire que pour le modèle classique de post-stratification individuel (paragraphe 5.1.2). Nous nous limitons donc dans ce texte à une description et une interprétation du nouveau modèle.

Le nouveau modèle de calibrage pour LFS 2016 peut être formulé comme suit :

$$\{\text{HH}; \text{STRAT12}; \text{SEX} \times \text{AGE} // \text{HHSIZE}; \text{Logit}\}$$

Nous entendons par là que :

- "HH " indique que nous procédons actuellement à un calibrage *formel* au niveau des ménages, ce qui signifie que les poids dans l'échantillon des ménages sont maintenant corrigés par un facteur  $g$  calculé pour chaque ménage.

- STRAT12 est la variable de stratification du calibrage, ce qui signifie formellement que le calibrage peut être effectué indépendamment dans les strates de calibrage STRAT12.
- "SEXE × ÂGE" avant le signe "/" signifie que l'échantillon de répondants des *individus* est calibré selon la distribution jointe de la population (individuelle) (*joint population distribution*) des variables SEXE et ÂGE. Cela s'effectue à l'intérieur de chaque strate de calibrage STRAT12, et cela implique que nous calibrons en fait toujours les répondants individuels selon la distribution jointe de la population de STRAT12, SEXE et ÂGE, tout comme dans le modèle classique post-stratification.
- "HHSIZE" après le signe "/" signifie que l'échantillon du répondant des *ménages* est calibré en fonction de la distribution de la population des ménages de la variable HHSIZE. Cela s'effectue également à l'intérieur de chaque strate de calibrage STRAT12 et cela implique que nous calibrons en fait les ménages répondants selon la distribution jointe de la population des ménages de STRAT12 et HHSIZE.
- "Logit" signifie que nous avons choisi une relation logistique entre les facteurs de correction  $g$  et les variables de calibrage. Deux paramètres, à savoir la limite inférieure et la limite supérieure pour  $g$ , ont été fixés de manière empirique à 0 et à 10, respectivement. Le choix "Lin" a donné des valeurs négatives pour certains  $g$ , ce qui a dû être corrigé en choisissant l'une des 3 relations alternatives.

Nous avons indiqué plus haut que nous calibrons maintenant *formellement*, d'une part, au niveau des ménages, mais d'autre part, selon les chiffres de la population des ménages  $M$  et les chiffres de la population des individus  $N$ . Comment ces affirmations sont-elles compatibles ? Il s'agit d'une question mathématique que nous ne pouvons pas dévoiler en détail dans ce document. Afin de donner une indication de ce que nous voulons dire exactement, nous soulignons :

- que chaque valeur de référence  $M$  correspond à une équation mathématique qui exprime que la somme de certains poids des ménages calibrés  $w = d \times g$  est égale à  $M$ ; par exemple  $\sum wz = M$  ou  $\sum dgz = M$ , dans laquelle la somme concerne les ménages et dans laquelle  $z$  est un indicateur (variable 0-1) qui indique quels ménages sont effectivement comptés dans la somme ;
- que chaque valeur de référence  $N$  correspond à une équation mathématique qui exprime que la somme de certains poids individuels calibrés  $\bar{w} = \bar{d} \times g$  est égale à  $N$ ; par exemple  $\sum \bar{w}x = N$  ou  $\sum \bar{d}gx = N$ , dans laquelle la somme concerne les individus et dans laquelle  $x$  est un indicateur (variable 0-1) qui indique quels ménages sont effectivement comptés dans la somme ;
- que le " calibrage formel au niveau du ménage " implique que ces dernières équations sont transformées en réécrivant leur terme de gauche sous la forme d'une sommation de ménage :  $\sum(\sum \bar{d}x)g = N$  où la sommation extérieure concerne tous les ménages, et chaque somme intérieure les individus d'un ménage ;
- que l'on peut écrire cette dernière équation comme  $\sum dg(\sum x) = N$  ou  $\sum w(\sum x) = N$ , parce que, pour LFS, tous les individus d'un même ménage ont le même poids d'échantillon  $\bar{d}$ , qui est également égal au poids d'échantillon du ménage  $d$ .

Le nouveau modèle de calibrage comporte donc les équations de calibrage  $\sum dgz = M$  et  $\sum dg(\sum x) = N$ , qui sont formellement du même type : les termes de gauche sont toujours des sommes pondérées (avec poids  $dg = w$ ) des ménages. Les inconnues de ces équations sont les facteurs de correction  $g$ , dont un seul doit être calculé par ménage.

La théorie du calibrage montre que les facteurs de correction  $g$  sont des fonctions

- de la variable de calibrage des ménages HHSIZE,
- et des *agrégations* des variables de calibrage individuelles SEX et AGE.

Concrètement, cela signifie que le facteur de correction  $g$  pour un ménage donné dépend

- de la catégorie HHSIZE à laquelle appartient le ménage,
- et de la composition du ménage, en particulier du nombre de membres du ménage dans chaque catégorie (SEXE  $\times$  ÂGE).

Toutefois, comme le  $g$  pour un ménage donné est aussi le facteur de correction de chaque individu de ce ménage, nous pouvons également formuler cette interprétation comme suit : le facteur de correction  $g$  pour un individu donné dépend

- de la taille du ménage auquel appartient l'individu,
- du *nombre* de personnes avec qui il vit du même sexe et du même groupe d'âge,
- et du nombre de personnes avec qui il vit dans d'autres catégories de sexe et d'âge.

Si les individus de différents ménages ont le même sexe et le même âge, alors ces individus ont le même facteur de correction  $g$  dans le modèle classique de post-stratification, mais ils peuvent se voir attribuer différents facteurs de correction dans le nouveau modèle si les ménages auxquels ils appartiennent ont une taille et/ou une composition différentes !

Enfin, nous remarquons que, pour le calibrage de HBS 2016, le nouveau modèle de calibrage de LFS 2016 n'a été appliqué qu'une seule fois, c'est-à-dire à l'ensemble de l'échantillon des répondants de LFS, et donc pas par trimestre. Les valeurs de référence  $N$  et  $M$  ont été calculées sur la base d'une extraction du Registre national du 1/07/2016.

## 5.2. Calibrage de HBS 2016

Enfin, le calibrage de HBS 2016 utilise les données HBS suivantes :

1. Nous utilisons une liste de tous les ménages répondants, avec comme caractéristiques le domicile, le statut professionnel du personne de référence (RP), l'âge du RP, le nombre de membres du ménage et le nombre de membres actifs du ménage.
2. Le trimestre au cours duquel le ménage a participé à l'enquête sera également converti en variables de calibrage, ce qui est une nouveauté par rapport au modèle de calibrage utilisé pour HBS 2014.
3. L'utilisation du revenu total (net imposable) des ménages dans le calibrage est également une nouveauté par rapport au modèle de HBS 2014.

4. Pour chaque ménage répondant : le *poids de l'échantillon*.

Le modèle de HBS 2016 diffère donc considérablement du modèle du HBS 2014. Cela signifie également que les benchmarks ne peuvent pas seulement être calculés à partir de LFS 2016 pondérée : les benchmarks relatifs au revenu seront obtenus à partir d'un couplage entre le registre national et le registre fiscal. De plus, pour certains benchmarks, nous n'aurons pas du tout besoin d'un échantillon ou d'un registre de référence ! La structure du nouveau modèle de calibrage sera expliquée étape par étape dans les sous-paragraphe suivants.

### 5.2.1. Modèle de calibrage traditionnel pour HBS

Nous pouvons affirmer que la partie centrale du modèle de calibrage de HBS 2016 reste le modèle classique ou traditionnel de post-stratification, tel qu'il a été appliqué (par exemple) en 2014, 2012,... Ce modèle peut être formulé comme suit :

{HH; GEWEST; PS; -}

Nous entendons par là que :

- "HH " indique que nous procédons actuellement à un calibrage au niveau du ménage, ce qui signifie que les poids dans l'échantillon des ménages, *corrigés de la non réponse*, sont maintenant corrigés par un facteur *g* calculé pour chaque ménage.
- REGION est la variable de stratification du calibrage, ce qui signifie formellement que le calibrage peut être effectué indépendamment dans les 3 régions.
- "PS" fait référence à 13 post-strates de ménages (à l'intérieur de chaque REGION) déterminées sur la base des caractéristiques des ménages mentionnées ci-dessus :
  - PS = 1 : salarié, -, 1P, -
  - PS = 2 : salarié, -, 2P, <=1A
  - PS = 3 : salarié, -, 2P, 2A
  - PS = 4 : salarié, -, 3+P, <=1A
  - PS = 5 : salarié, -, 3P, 2+A
  - PS = 6 : salarié, -, 4+P, 2+A
  - PS = 7 : indépendant, -, -, -
  - PS = 8 : inactif, <=59, 1P, -
  - PS = 9 : inactif, <=59, 2+P, -
  - PS = 10 : inactif, 60-69, 1P, -
  - PS = 11 : inactif, 60-69, 2+P, -
  - PS = 12 : inactif, 70+, 1P, -
  - PS = 13 : inactif, 70+, 2+P, -

où l'on se réfère successivement au statut socioprofessionnel du RP, à l'âge du RP, au nombre de membres du ménage et au nombre de personnes " actives " dans le ménage.

- " - " signifie qu'il n'est pas nécessaire de faire un choix spécifique quant au lien entre les facteurs de correction et les variables de calibrage. Le modèle est en effet un modèle de post-stratification pour les ménages de HBS.

Avec ce modèle, nous nous calibrons dès lors selon la distribution jointe de la population des ménages de REGIO et de PS. Nous ne pouvons pas calculer cette distribution à partir du Registre national, car certaines caractéristiques sous-jacentes des ménages ne sont pas reprises dans le Registre national. Cependant, nous pouvons les calculer à partir de LFS calibrée.

Grâce au calibrage des estimations de HBS par rapport à celles de LFS, il existe une certaine cohérence entre les estimations de LFS et celles de HBS au niveau des ménages. Mais comme le nouveau modèle de calibrage de LFS implique également une certaine cohérence au niveau des ménages avec le registre national, certaines estimations de HBS seront également cohérentes avec certaines données du Registre national. Nous pouvons conclure en particulier que, grâce au renouvellement du calibrage de LFS 2016, les nombres estimés de ménages dans LFS 2016 et dans HBS 2016 sont égaux entre eux et aux données du Registre national, pour l'ensemble de la Belgique et pour chaque région. Bien que le nouveau modèle de calibrage de LFS et le modèle de calibrage de HBS intègrent également la taille des ménages dans les variables de calibrage, nous ne pouvons pas nous attendre de telles égalités exactes par classe de taille des ménages (par exemple, parce que dans le groupe des ménages ayant un RP indépendant, aucune distinction n'est faite entre les classes de taille des ménages), mais nous pouvons nous attendre à une "quasi cohérence" ici également.

Dans le tableau 11, nous présentons les totaux d'étalonnage par REGIO et POST STRATA, tels qu'ils ont été calculés sous forme d'estimations à partir de LFS 2016, après calibration de LFS comme décrit à la section 5.1.3.

**Tableau 11 Totaux de calage (benchmarks) pour HBS 2016, estimés sur base de LFS 2016**

Poststrate (dans Région)	Région						Total	
	R. Bruxelles-C.		Flandres		Wallonie			
	Poids de calage		Poids de calage		Poids de calage		Poids de calage	
	Somme	%	Somme	%	Somme	%	Som	%
Salarié, 1p	107.626,4	19,8	339.586,5	12,3	178.245,1	11,5	<b>625.457,9</b>	12,9
Salarié, 2p, <=1act	33.088,4	6,1	181.976,4	6,6	114.783,7	7,4	<b>329.848,5</b>	6,8
Salarié, 2p, 2act	26.682,6	4,9	201.880,1	7,3	73.671,9	4,7	<b>302.234,6</b>	6,2
Salarié, 3+p, <=1act	55.425,4	10,2	195.452,6	7,1	137.838,3	8,9	<b>388.716,3</b>	8,0
Salarié, 3p, 2+act	18.864,7	3,5	202.192,2	7,3	91.414,2	5,9	<b>312.471,1</b>	6,4
Salarié, 4+p, 2+act	31.441,5	5,8	336.860,4	12,2	160.233,1	10,3	<b>528.535,0</b>	10,9
Indépendant	60.039,6	11,1	275.424,1	10,0	134.961,5	8,7	<b>470.425,1</b>	9,7
Inactif, <=59, 1p	62.391,0	11,5	131.307,9	4,8	128.919,0	8,3	<b>322.618,0</b>	6,7
Inactif, <=59, 2+p	45.566,2	8,4	81.642,7	3,0	104.281,3	6,7	<b>231.490,2</b>	4,8
Inactif, 60-69, 1p	31.448,9	5,8	160.851,5	5,8	114.383,8	7,4	<b>306.684,2</b>	6,3
Inactif, 60-69, 2+p	13.420,7	2,5	175.444,3	6,4	92.441,2	5,9	<b>281.306,2</b>	5,8
Inactif, 70+, 1p	28.714,0	5,3	177.706,7	6,5	99.524,4	6,4	<b>305.945,1</b>	6,3
Inactif, 70+, 2+p	27.816,6	5,1	293.346,7	10,7	125.590,5	8,1	<b>446.753,9</b>	9,2
<b>Total</b>	<b>542.526,0</b>	<b>100,0</b>	<b>2.753.672,0</b>	<b>100,0</b>	<b>1.556.288,0</b>	<b>100,0</b>	<b>4.852.486,0</b>	<b>100,0</b>

### 5.2.2. Extension afin d'atteindre l'équilibre entre les trimestres

Au paragraphe 4.2, nous avons déjà mentionné que les quatre trimestres de 2016 ne sont pas représentés de manière égale dans l'échantillon total des ménages répondant à l'enquête HBS 2016. Les deux premiers trimestres sont nettement sous-représentés, aussi bien sans pondération (première ligne du tableau 10) qu'après pondération avec les poids dans l'échantillon (deuxième ligne du tableau 10). Après un calibrage au moyen du modèle classique du paragraphe précédent 5.2.1, ce problème n'aurait pas été résolu. En effet, le modèle {HHH ; GEWEST ; PS ; -} implique que tous les ménages répondants ayant les mêmes valeurs pour les variables de calibrage REGION et PS se verrait attribuer le même facteur de correction  $g$ , de sorte que les poids d'échantillon de tous ces ménages, quel que soit le trimestre auquel ils ont participé, seraient ajustés (en pratique : augmentés) exactement de la même manière. Le déséquilibre trimestriel subsisterait après calibrage avec le modèle {HHH ; GEWEST ; PS ; -}, ce qui pourrait encore renforcer ou atténuer les effets saisonniers des estimations pour l'ensemble de l'année 2016.

Le but de l'extension du modèle classique du paragraphe précédent est d'obtenir des poids calibrés  $w$  qui répondent aux équations suivantes :

$$\sum_{R1} w = \sum_{R2} w = \sum_{R3} w = \sum_{R4} w$$

Où  $\sum_{Rt} w$  est la somme des pondérations calibrées pour les ménages ayant participé à HBS au cours du trimestre  $t$ . La première de ces équations,  $\sum_{R1} w = \sum_{R2} w$ , peut être convertie en une équation de calibrage conventionnelle comme suit. Nous définissons les variables  $z_1$  et  $z_2$ :

$$\begin{aligned} z_1 &= 1 \text{ voor elk huishouden dat in T1 aan HBS deelneemt} \\ &= 0 \text{ voor elk huishouden dat in T2, T3 of T4 aan HBS deelneemt} \end{aligned}$$

$$\begin{aligned} z_2 &= 1 \text{ voor elk huishouden dat in T2 aan HBS deelneemt} \\ &= 0 \text{ voor elk huishouden dat in T1, T3 of T4 aan HBS deelneemt} \end{aligned}$$

Cela peut être réécrit comme suit:  $\sum_{R1} w = \sum_{R2} w$ .

$$\sum w \cdot z_1 = \sum w \cdot z_2$$

dans laquelle chacune des sommations concerne *tous* les ménages répondants. L'étape suivante consiste à définir une *variable de contraste*  $z_{12}$  :

$$\begin{aligned} c_{12} = z_1 - z_2 &= +1 \text{ pour elk huishouden dat in T1 aan HBS deelneemt} \\ &= -1 \text{ pour elk huishouden dat in T2 aan HBS deelneemt} \\ &= 0 \text{ pour elk huishouden dat in T3 of T4 aan HBS deelneemt} \end{aligned}$$

Ensuite, l'équation précédente peut être réécrite sous la forme :

$$\sum w \cdot c_{12} = 0$$

Cette équation a la forme habituelle d'une équation de calibrage avec comme variable de calibrage la *variable de contraste*  $c_{12}$  et le benchmark ou le total de calibrage 0. Pour le calcul de tels benchmarks, nous n'avons bien entendu pas besoin d'un échantillon ou d'un registre de référence !

Cette équation nous a permis de contraster les trimestres T1 et T2. De la même manière, nous pouvons contraster les trimestres T1 et T3, et T1 et T4, c'est-à-dire en introduisant les *variables de contraste*  $c_{13}$  et  $c_{14}$  correspondantes. Les équations initiales  $\sum_{R1} w = \sum_{R2} w = \sum_{R3} w = \sum_{R4} w$  peuvent donc être reformulées sous la forme d'un système de trois équations  $\sum w \cdot c_{12} = 0$ ,  $\sum w \cdot c_{13} = 0$  et  $\sum w \cdot c_{14} = 0$ .

Si nous remarquons maintenant que nous pouvons aussi formuler le modèle classique {HH; REGION; PS; -} sous la forme de {HH ; REGION x PS ; -}, il est facile de comprendre qu'un modèle étendu peut être formulé comme :

$$\{\text{HH; GEWEST} \times \text{PS} + \text{C12} + \text{C13} + \text{C14}; \text{TrLin}\}$$

Ici, "TrLin" fait référence à la méthode linéaire tronquée (voir le paragraphe 5.1.2 concernant ce choix et d'autres choix de la "méthode" ; et nous avons finalement fixé les limites inférieure et supérieure des facteurs de correction (*g-weights*) à 0,10 et 100, respectivement).

Ce n'est pas le même modèle que :

$$\begin{aligned} &\{\text{HH; REGION; PS} + \text{C12} + \text{C13} + \text{C14}; \text{TrLin}\} \\ &= \{\text{HH; REGION} \times (\text{PS} + \text{C12} + \text{C13} + \text{C14}); \text{TrLin}\} \\ &= \{\text{HH; REGION} \times \text{PS} + \text{REGION} \times (\text{C12} + \text{C13} + \text{C14}); \text{TrLin}\} \\ &= \{\text{HH; REGION} \times \text{PS} + \text{REGION} \times \text{C12} + \text{REGION} \times \text{C13} + \text{REGION} \times \text{C14}; \text{TrLin}\} \end{aligned}$$

où l'équilibre entre les trimestres est atteint non seulement au niveau global, mais aussi par région. Il est à noter que le modèle {HHH ; REGION x (PS + C12 + C13 + C14) ; TrLin} est plus restrictif que le modèle {HHH ; REGION x PS + C12 + C13 + C14 ; TrLin} ; et que les limites de contraste au niveau des régions impliquent des limites de contraste au niveau global.

Un modèle encore plus restrictif que {HH; REGION × (PS + C12 + C13 + C14); TrLin} et aussi encore plus restrictif que {HH; REGION × PS + C12 + C13 + C14 ; TrLin}, et donc le modèle

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

où l'équilibre entre les trimestres est visé non seulement au niveau global et par région, mais aussi par province.

Après évaluation de ces modèles, qui visent tous à rétablir l'équilibre entre les trimestres, ce dernier modèle a été retenu, car il est le plus proche de la représentation correcte.

### **5.2.3. Extension pour réduire un biais**

Une évaluation plus approfondie du modèle {HH; REGION × PS + PROV × (C12 + C13 + C14); TrLin} s'est révélée ne pas répondre pleinement aux attentes (éventuelles) des utilisateurs. En effet, ce modèle a donné une surestimation frappante du pourcentage de propriétaires (au niveau des ménages). Le tableau 12 compare les estimations de la distribution relative de la variable du statut de propriété dans HBS 2012, 2014 et 2016 - quand on utilise le modèle classique pour 2012 et 2014 et le modèle plus détaillé avec des équilibres trimestriels pour 2016 - avec des estimations basées sur l'enquête SILC 2014-2016.



**Tableau 12 Estimations du statut de propriété issues de HBS et de SILC**

	HBS 2012	HBS 2014	HBS 2016	SILC 2014	SILC 2015	SILC 2016
	Modèle classique	Modèle classique	Modèle classique, avec équilibre entre les trimestres	Calibrage usuel		
<b>Propriétaire</b>	68,2%	70,0%	72,4%	66,1%	66,2%	66,7%
<b>Locataire</b>	30,6%	29,0%	26,4%	32,2%	32,3%	31,9%
<b>Inconnu</b>	1,2%	1,0%	1,2%	1,7%	1,5%	1,4%

Nous avons donc essayé d'utiliser le statut de propriété (propriétaire ou locataire) comme variable de calibrage supplémentaire ; cette variable fait l'objet d'une interrogation dans HBS. Cependant, nous n'avons pas trouvé de bons benchmarks pour cette variable. Comme proxy (ou prédicteur) du statut de propriété, nous avons donc utilisé le revenu familial : dans les classes de revenu plus élevées, nous trouvons, en pourcentage, plus de propriétaires de logements. Ce revenu (net imposable) du ménage peut être obtenu à partir d'un lien entre le Registre national et le registre fiscal pour chaque ménage répondant et pour chaque ménage de la population, si bien que des benchmarks appropriés peuvent également être déterminés.

Bien que nous n'ayons pas utilisé le revenu familial comme variable numérique dans le modèle de calibrage, nous avons effectué une répartition par classes de cette variable, de la manière suivante.

Tout d'abord, nous constatons que, selon le registre fiscal, certains ménages (privés) de la population n'ont pas de revenu (net imposable). La première classe de revenu se voit donc attribuer l'étiquette "manquant" et le code 1. Les 4 quintiles de revenu  $K_1, K_2, K_3, K_4$  ont ensuite été déterminés sur la base de la (sous-)population des ménages qui ont un revenu. Ces quintiles déterminent 5 classes de revenu auxquelles nous donnons les codes respectifs 2, 3, 4, 5 et 6. L'inclusion des ménages sans revenu (imposable) a pour conséquence que, si 100  $m$  est le pourcentage des ménages de la classe de revenu 1 (c'est-à-dire "manquants"), le pourcentage dans chacune des classes de 2 à 6 est égal à  $(100 - 100m)/5$ ; si  $m = 0$ , alors chaque classe de 2 à 6 contiendra exactement 20% des ménages. De cette façon, nous déterminons la classe de revenu (1, 2, ..., ou 6) de chacun des ménages répondants (en la couplant au Registre national détaillé) ; nous appelons cette variable INK\_5. Cette nouvelle variable peut ensuite être reprise en tant que variable de calibrage catégorielle dans le dernier modèle de calibrage présenté au paragraphe précédent :

$$\{HH; REGION \times PS + PROV \times (C12 + C13 + C14) + INK\_5; TrLin\}$$

Notons que les 6 benchmarks, qui appartiennent au dernier terme, INK\_5, de ce modèle peuvent être calculés en multipliant le nombre total de ménages dans la population - qui est estimé à partir de LFS 2016 calibrée et est (implicitement) un benchmark qui appartient au terme REGION x PS - par les proportions  $m$  ou  $(1 - m)/5$ .

Ce dernier modèle calibre donc, entre autres, selon la distribution globale des revenus dans la population des ménages privés belges, cette distribution des revenus étant résumée par le pourcentage de ménages sans revenus et les quintiles des ménages avec revenus. Nous pouvons affiner ce modèle en résumant de la même manière la répartition du revenu des ménages par région. Si  $K_{j1}, K_{j2}, K_{j3}, K_{j4}$  sont les quatre quintiles de revenu de la population des ménages ayant un revenu dans la région  $j$  ( $j = 1, 2, 3$ ), nous construisons une variable catégorielle INK\_5\_G comme suit :

- INK\_5\_G =
- 1 pour les ménages sans revenu
  - 2 pour les ménages de la région  $j$  ( $j = 1, 2, 3$ ) et dont le revenu n'excède pas  $K_{j1}$
  - 3 pour les ménages de la région  $j$  ( $j = 1, 2, 3$ ) et dont le revenu excède  $K_{j1}$  mais pas  $K_{j2}$
  - 4 pour les ménages de la région  $j$  ( $j = 1, 2, 3$ ) et dont le revenu excède  $K_{j2}$  mais pas  $K_{j3}$
  - 5 pour les ménages de la région  $j$  ( $j = 1, 2, 3$ ) et dont le revenu excède  $K_{j3}$  mais pas  $K_{j4}$
  - 6 pour les ménages de la région  $j$  ( $j = 1, 2, 3$ ) et dont le revenu excède  $K_{j4}$

On obtient alors le modèle de calibrage suivant :

$$\{HH; \text{REGION} \times \text{PS} + \text{PROV} \times (\text{C12} + \text{C13} + \text{C14}) + \text{REGION} \times \text{INK}_5\_G; \text{TrLin}\}$$

Le terme REGION x INK\_5\_G détermine  $3 \times 6 = 18$  nouvelles équations de calibrage. Les benchmarks (ou totaux de calibrage) sont les nombres de ménages dans la population par région et par catégorie d'INK\_5\_G, calculés à partir des nombres de ménages par région (qui sont implicitement déjà des benchmarks pour le terme REGION x PS) et les proportions  $m_j$  ( $j = 1, 2, 3$ ) des ménages sans revenu dans les trois régions.

De la même façon, nous pouvons construire un modèle de calibrage encore plus raffiné qui tient compte de la distribution des revenus par province :

$$\{HH; \text{REGION} \times \text{PS} + \text{PROV} \times (\text{C12} + \text{C13} + \text{C14}) + \text{PROV} \times \text{INK}_5\_P; \text{TrLin}\}$$

La variable de calibrage catégorielle INK\_5\_P est construite de la même manière que la variable INK\_5\_G ; bien entendu, la répartition des revenus par province  $p$  doit être résumée au préalable à l'aide des ratios  $m_p, K_{p1}, K_{p2}, K_{p3}, K_{p4}$  ( $p = 1, \dots, 11$ ).

Le tableau 13 montre qu'avec le nouveau modèle de calibrage pour 2016, qui a également été utilisé pour la validation de HBS 2014, les estimations de la répartition du statut de propriété sont plus conformes aux estimations fondées sur l'enquête SILC (tableau 12) et HBS 2012.

**Tableau 13 Estimations du statut de propriété à partir de HBS selon différents modèles de calibrage**

	HBS 2012	HBS 2014	HBS 2016	HBS 2014	HBS 2016
	<b>Modèle classique</b>	<b>Modèle classique</b>	<b>Modèle classique, avec équilibre entre les trimestres</b>	<b>Nouveau calibrage</b>	
<b>Propriétaire</b>	68,2%	70,0%	72,4%	67,3%	68,6%
<b>Locataire</b>	30,6%	29,0%	26,4%	31,6%	30,1%
<b>Inconnu</b>	1,2%	1,0%	1,2%	1,1%	1,3%

## 6. Programmes d'encodage, de contrôle, d'imputation

Les ménages participants doivent compléter un questionnaire avec l'aide de l'enquêteur et un carnet de dépenses dans lequel ils enregistrent toutes leurs dépenses pendant un mois. Dans le passé, l'enquête sur le budget des ménages était une enquête exclusivement papier. Tant les enquêteurs que les ménages ne remplissaient les questionnaires et les carnets que sur papier. Cette méthode de travail nécessite beaucoup de travail d'encodage par la suite et offre peu de possibilités d'effectuer des contrôles de qualité pendant le travail de terrain. Par ailleurs, les ménages demandaient de plus en plus souvent de pouvoir remplir les carnets par voie électronique.

### 6.1. Modernisation de la collecte des données

L'usage du CAPI (Computer Assisted Personal Interviewing) pour l'interview directe des membres du ménage par l'enquêteur, avec l'aide d'un PC-tablet et d'un programme d'encodage adapté (routing adapté à chaque situation, pas de travail de ré-encodage, moins d'erreurs, ...) était généralisé en 2016.

En ce qui concerne le remplissage du carnet de dépenses, les ménages se sont vu proposer, à côté d'une version papier 'standard' (PPSI, ou Paper-and-pencil Self-administered Interviewing), la possibilité d'utiliser une application 'websurvey' développée en Java (CAWI, ou Computer Assisted Web Interviewing). En 2016, 55 % des ménages ont choisi d'utiliser la version électronique du carnet (30 % en 2014).

A partir de l'enquête HBS 2012, le contenu des questionnaires et des carnets a été simplifié, allégeant la charge pesant sur les répondants.

Le recrutement via LFS présente l'avantage que de très nombreuses informations sur les individus du ménage sont déjà disponibles. Ces informations ne doivent plus être demandées dans les questionnaires. On ne demande que de confirmer les données déjà obtenues. La modernisation a ouvert d'autres possibilités d'amélioration de la qualité du travail de terrain. A partir de 2012, les enquêteurs ont dû documenter chaque contact avec les répondants dans un formulaire de contact (dates des prises de contact, modes de prise de contact, résultat de chaque tentative de contact). Comme ces formulaires de contact sont remplis directement sur support électronique et transmis régulièrement au serveur, le déroulement du travail de terrain peut être suivi de près.

La méthode d'enregistrement des quantités a aussi été améliorée dans les carnets en 2012, ce qui facilitera la livraison des données à Eurostat et aux autres utilisateurs.

Le questionnaire en CAPI suit également un certain routing. Les questions qui ne sont pertinentes que dans une certaine situation ne sont posées qu'aux personnes dans cette situation. Par exemple, le montant du loyer n'est demandé qu'aux locataires et les dépenses annuelles en mazout de chauffage qu'aux personnes se chauffant au mazout. Avec le questionnaire papier en 2012, cette possibilité n'existait pas. Cela engendrait une grosse perte de temps pour le nettoyage des résultats par la suite.

## 6.2. Contrôles et imputation

### 6.2.1. *Détection des outliers (valeurs aberrantes)*

Pour les dépenses indiquées dans le carnet, des programmes SAS de contrôle et de détection des outliers ont été développés. Plusieurs méthodes ont été testées. Pour les dépenses alimentaires et les boissons non alcoolisées, nous avons finalement opté pour la méthode dite de 'resistant fences method' (RFM). Dans cette méthode univariée, l'intervalle de tolérance est défini à partir des quartiles :  $[ Q1 - \text{coeff.low} * \max(Q3 - Q1, |a \times Q2|), Q3 + \text{coeff.up} * \max(Q3 - Q1, |a \times Q2|) ]$ . Le terme  $a \times Q2$  assure une valeur minimale de la distance. Les coefficients (low et up) ont été fixés à 3, comme habituellement dans la pratique et 0.05 pour le paramètre a.

Les données ont au préalable subi une transformation de Box-Cox (avec paramètre  $\lambda$  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.

Nous avons identifié les montants aberrants par code COICOP\_HBS\_BE détaillé et avons remplacé les montants aberrants par la médiane des dépenses données.

Impact sur le total des dépenses (montants non pondérés) : diminution de 0.6 % pour les dépenses alimentaires et boissons non alcoolisées.

CD_COICOP	Amount	Amount after processing	Difference
1	1 779 066	1 768 368	10 698

Pendant un mois, il est difficile de se faire une idée des dépenses périodiques et des grosses (et donc rares) dépenses d'un ménage. Des questions concernant ces deux types de dépenses figurent donc dans un questionnaire à part, le questionnaire "ménage". Celui-ci comprend également des questions sur la composition du ménage, son habitation (confort, âge, type d'habitation, type de chauffage...), les dépenses groupées (gaz + électricité, par exemple), les consommations d'eau, de gaz, d'électricité, de mazout ; les assurances, le statut, la profession et le niveau d'étude des membres du ménage ainsi que leurs revenus.

Différents contrôles de cohérence ont ainsi pu être réalisés, en comparant le carnet et le questionnaire.

Voici, dans le paragraphe 6.2.2, un exemple de traitement de ces nombreuses informations dans la correction et l'imputation des dépenses d'énergie. Les autres dépenses périodiques, comme les loyers et charges, ainsi que les dépenses pour des biens durables ont également fait l'objet de contrôles approfondis et de traitements spécifiques.

Note importante :

Toutes les dépenses sont ramenées à un mois : les dépenses du carnet restent telles quelles, les dépenses du questionnaire rétroactif (4 derniers mois), avec les biens durables dont les voitures, sont divisées par 4 et les dépenses périodiques sont divisées par la périodicité.

Pour le tableau final, toutes ces dépenses sont annualisées, donc multipliées par 12.

### 6.2.2. Imputation des dépenses énergétiques

Le questionnaire ménage est rempli entièrement sous la forme d'une interview CAPI. 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 que toutes les questions pertinentes soient certainement posées. Avant 2014, un questionnaire ménage sur papier était utilisé. Dans la pratique, il était souvent rempli par le ménage lui-même, après quoi l'enquêteur effectuait une simple vérification. Beaucoup de données étaient dès lors manquantes, notamment sur les dépenses énergétiques. Les incohérences dans les réponses étaient également nombreuses (les ménages déclaraient, p. ex., que leur principale source de chauffage fonctionnait au mazout de chauffage, mais ne mentionnaient que des dépenses en gaz naturel). Ces informations manquantes et ces incohérences engendraient beaucoup de travail de correction par la suite. L'introduction des questionnaires CAPI a grandement amélioré la qualité des données, réduisant ainsi le travail de vérification par la suite.

<i>Dépense</i>	<i>Nombre de données manquantes en 2016</i>	<i>Nombre de données manquantes en 2014</i>	<i>Nombre de données manquantes en 2012</i>
<i>Eau</i>	76	92	666
<i>Électricité</i>	41	75	1302
<i>Gaz (conduite de gaz)</i>		27	1223

Il est également possible d'encoder des factures combinées. Un ménage paie, par exemple, la facture d'électricité et de gaz, ou la consommation d'électricité et d'eau via une seule facture. Cependant, les ménages et les enquêteurs doivent malgré tout ventiler autant que possible les montants. Dans la suite de ce document, on appelle aussi imputation la ventilation de ce genre de combinaison de types d'énergie.

La méthode d'imputation reste la même qu'en 2014. Des sources d'information sur les dépenses énergétiques sont disponibles même pour les ménages qui n'ont pas fourni d'informations dans le questionnaire. Les carnets, par exemple, fournissent des informations sur les dépenses pendant le mois de référence. Les dépenses d'énergie peuvent avoir été effectuées au cours de ce mois.

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 un modèle de régression linéaire.

**Ronde 1** Collecte d'informations dans les carnets afin de les utiliser pour les dépenses annuelles.

La première étape consiste à utiliser les informations enregistrées dans les carnets pendant le mois de référence comme base pour imputer une consommation annuelle manquante dans le questionnaire. 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.

Un "flag" a été attribué aux données imputées :

<i>Facture d'eau</i>	<i>FL_W</i>	<i>N=20</i>
<i>Facture d'électricité</i>	<i>FL_E</i>	<i>N=9</i>
<i>Facture de gaz</i>	<i>FL_G</i>	<i>N=6</i>

**Ronde 2** Ventilation de factures groupées

Différentes combinaisons sont possibles de les dépenses énergétiques. La combinaison la plus fréquente est la facturation conjointe de gaz et d'électricité.

<b>Description</b>	<b>Nombre</b>
<i>Eau, gaz et électricité</i>	<i>N=38</i>
<i>Eau, gaz et électricité</i>	<i>N=28</i>
<i>Eau et gaz</i>	<i>N=38</i>
<i>Electricité et gaz</i>	<i>N=1420</i>

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/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.

### **Ronde 3** Imputer les données manquantes au moyen de modèles de régression

Les données encore manquantes ont été complétées par une estimation fondée sur un modèle de régression.

#### 1. EAU

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.

Les données manquantes ont été estimées et se sont vu attribuer un 'flag':

<i>FL_WR</i>	<i>N=320</i>
--------------	--------------

#### 2. ÉLECTRICITÉ

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.

Les estimations se sont vu attribuer un 'flag' :

<i>FL_ER</i>	<i>N=148</i>
--------------	--------------

#### 3. GAZ

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 ou le carnet.



Les variables indépendantes pertinentes sont la taille du ménage (exprimée en unités équivalentes), la région et/ou le fonctionnement du système principal de chauffage au gaz. Les montants imputés se sont vu attribuer un 'flag' :

<i>FL_gas</i>	<i>N=155</i>
---------------	--------------

#### 4. GASOIL DE CHAUFFAGE

Les montants pour le gasoil de chauffage ont également été estimés sur la base d'un modèle de régression linéaire. Ils ont été estimés 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/ou le fonctionnement du système principal de chauffage au gasoil.

Les montants imputés se sont vu attribuer un 'flag' :

<i>FL_oil</i>	<i>N=91</i>
---------------	-------------

#### 5. BUTANE

11 données manquaient pour le butane. Toutefois, aucun modèle de régression n'a pu être élaboré pour l'imputation 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.

Les ménages pour lesquels une imputation a été effectuée se sont vu attribuer un 'flag' :

<i>FL_butane</i>	<i>N=11</i>
------------------	-------------

#### 6. GAZ PROPANE

En 2016, un problème est apparu dans la mesure des dépenses en gaz propane. En raison d'une erreur de routing dans le questionnaire, on n'a pas demandé aux ménages qui se chauffaient au propane quel était le montant annuel de leurs achats de propane. Les ménages concernés ont été contactés par la suite pour leur demander les informations manquantes. Cependant, nous ne sommes pas parvenus à contacter tout le monde. Ces montants ont donc dû être imputés. Cependant, aucun modèle de régression approprié n'a pu être trouvé. L'imputation a donc été effectuée sur la base de la moyenne.

Les valeurs estimées du propane se sont vu attribuer un "flag" :

<i>FL_propane</i>	<i>N=78</i>
-------------------	-------------

## 7. CHARBON

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 des ménages qui avaient déclaré des dépenses en charbon.

Les estimations se sont vu attribuer un 'flag' :

<i>FL_coal</i>	<i>N=4</i>
----------------	------------

## 8. PELLETS

La consommation pellets n'a pas pu être estimée de manière significative par un modèle de régression. L'imputation a donc été effectuée sur la base de la moyenne.

Les estimations se sont vu attribuer un 'flag' correspondant:

<i>FL_pellets</i>	<i>N=9</i>
-------------------	------------

## 9. PETROLE

Aucune donnée manquante, donc aucune imputation nécessaire.

Remarque : Les dépenses manquantes dans la catégorie BOIS (COICOP=04549A) n'ont pas été imputées. Le bois de chauffage peut être obtenu gratuitement ou un stock peut avoir été constitué pour plusieurs années.

### 6.2.3. Revenus

Jusqu'en 2010, les revenus étaient comme les dépenses remplis de manière exhaustive dans le carnet. Dès 2012, pour simplifier le travail des répondants, les revenus sont collectés dans le questionnaire, de manière plus globale. Ceci garantit toujours la possibilité de comparer les dépenses des ménages en fonction de leur niveau de vie (quartiles de revenus par exemple), mais des tableaux détaillés concernant la composition du revenu ne seront plus produits sur base de l'enquête HBS.

Pour toute étude concernant les revenus, nous recommandons l'utilisation de l'enquête EU-SILC (enquête communautaire sur les revenus et les conditions de vie) également organisée par la DGSIE.

Les revenus ont fait l'objet de contrôles approfondis (détection d'outliers, contrôle des ménages qui n'avaient pas rempli l'information correctement : questionnaire vide, revenu trop bas ou incohérent avec les informations remplies concernant le statut socio-professionnel, le secteur d'activité, ...)

En 2014, certaines questions ont été ajoutées dans le questionnaire concernant les revenus de remboursement reçus au niveau du ménage.

#### *Validation des revenus*

<b>Weighted Basic Statistical Measures (Location &amp; Variability)</b>	<b>HBS 2016</b>	<b>EU-SILC 2016</b>
Mean	36816.25	38184.78
Std Deviation	1010177	1021175
Median	31239	31426.83
Variance	1020460000000	685831000000
Mode	19200	13200
Range	1666798	1117902
Interquartile Range	28140	32245

*Distribution des revenus*

Income class	Income EU-SILC 2016		Income HBS 2016	
	Percent	Cumulative Percent	Percent	Cumulative Percent
<b>0-250</b>	0.57	0.57	0.27	0.27
<b>250-500</b>	0.46	1.03	0.20	0.46
<b>500-1000</b>	4.79	5.81	3.39	3.85
<b>1000-1500</b>	16.75	22.56	15.3	19.15
<b>1500-2000</b>	14.00	36.56	16.37	35.53
<b>2000-2500</b>	11.16	47.72	11.98	47.5
<b>2500-3000</b>	9.48	57.21	10.51	58.01
<b>3000-5000</b>	25.66	82.86	29.70	87.71
<b>5000+</b>	17.14	100.00	12.29	100.00

#### 6.2.4. Imputed rent

Le 'imputed rent' est un loyer fictif attribué aux propriétaires, ainsi qu'aux ménages occupant un logement à titre gratuit.

Pour estimer le prix auquel leur logement serait loué et donc avoir une idée de l'avantage financier lié à leur statut de propriétaire, nous utilisons la méthode préconisée par Eurostat: un modèle de Heckman à deux phases. Nous utilisons d'abord une régression probit pour estimer la probabilité d'être locataire selon les caractéristiques du ménage et de ses membres (dont l'âge, le niveau d'éducation, le nombre d'actifs, ...). Ensuite, nous estimons par régression le loyer avec de nombreuses caractéristiques du logement présentes dans l'enquête comme 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 bien sûr tenant compte du résultat de la première régression.

*Validation loyers réels et loyers imputés par rapport à l'enquête EU-SILC 2016*

	<b>SILC 2016</b>	<b>HBS 2016</b>
Loyer mensuel par ménage locataire (€)	544	533
Imputed rent mensuel par ménage propriétaire(€)	637	676

# Appendices

---

## 1. Quelques définitions

Le **ménage** est constitué, soit par une personne vivant habituellement seule, soit par deux ou plusieurs personnes qui, unies ou non par des liens de parenté, occupent habituellement un même logement et y vivent en commun.

La **personne de référence selon HBS** 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 **Registre national**.

L'**unité de consommation modifiée** est une échelle d'équivalence qui est appliquée pour adapter les dépenses de consommation en fonction de la taille et de la composition du ménage. Un coefficient de 1 est attribué au premier adulte, de 0,5 aux autres personnes de plus de 13 ans et de 0,3 aux enfants de 13 ans ou moins (échelle modifiée de l'OCDE).

## 2. Composition de l'échantillon de répondants (représentativité)

### 2.1. Distributions des ménages

#### *Distribution par type de taille de ménage*

MS_Size	HBS 2016			
	Unweighted		Weighted	
	Freq.	%	Freq.	%
1	1239	27.59	1646832	33.94
2	1546	34.43	1516933	31.26
3	742	16.53	751574.2	15.49
4	682	15.19	658140.4	13.56
5	214	4.77	206393.2	4.25
6	52	1.16	55747.01	1.15
7	9	0.2	10986.88	0.23
8	3	0.07	2738.95	0.06
9	2	0.04	2838.752	0.06
10	1	0.02	302.8624	0.01
11 +	0	0	1646832	33.94
<b>Total</b>	4490	100	4852487	100

#### *Remarques*

- 1) Dans l'échantillon des ménages répondants, 9 est la taille de ménage la plus élevée. Elle atteint 31 dans la population !
- 2) La somme des pondérations (4,852,487) est égale au chiffre de la population, car tant LFS que HBS sont calibrées sur le nombre global de ménages.
- 3) La distribution pondérée et la distribution de la population de MS\_Size diffèrent, malgré et en raison d'un calibrage limité selon la taille de ménage dans HBS.

**Distribution du nombre d'actifs par ménage**

MS_Workers	HBS 2016					
	Unweighted			Weighted		
	Freq.	%	Cum. %	Freq.	%	Cum. %
0	1550	34.52	34.52	1799993	37.09	37.09
1	1444	32.16	66.68	1644850	33.9	70.99
2	1397	31.11	97.8	1321834	27.24	98.23
3	90	2	99.8	79099.34	1.63	99.86
4	9	0.2	100	6710.483	0.14	100
5 +				0	0	
<b>Total</b>	4490	100		4852487	100	

2.2. [Distributions individuelles](#)

**Distribution du genre des membres des ménages répondants**

CD_Gender	HBS 2016			
	Unweighted		Weighted	
	Freq.	%	Freq.	%
Homme	5230	48.51	5299402	47.91
Femme	5552	51.49	5762425	52.09
<b>Total</b>	10782	100	11061826	

*Remarque*

Le sexe est une variable de calibrage pour LFS, mais pas pour HBS. Par conséquent, la somme des pondérations individuelles de HBS (11,061,826) n'est pas égale au chiffre de la population.



***Distribution de l'âge des membres des ménages répondants***

MS_Age	<b>HBS 2016</b>			
	<b>Unweighted</b>		<b>Weighted</b>	
	<b>Freq.</b>	<b>%</b>	<b>Freq.</b>	<b>%</b>
- 25	3444	31.94	3533290	31.94
26-45	2976	27.6	3041867	27.5
46-64	3151	29.22	2911568	26.32
65-74	998	9.26	1266899	11.45
75 +	213	1.98	308201.6	2.79
<b>Total</b>	10782	100	11061826	100

***Distribution du statut socio-professionnel des membres des ménages  
répondants***

Socstat	HBS 2016					
	Unweighted			Weighted		
	Freq.	%	Cum. %	Freq.	%	Cum. %
Indépendant (y compris aidant non rémunéré)	373	3.51	3.51	648737.6	5.97	5.97
Salarié	4001	37.7	41.22	3711362	34.15	40.11
Chômeur	456	4.3	45.51	449621.5	4.14	44.25
Pensionné/Prépensionné	1783	16.8	62.32	1921385	17.68	61.93
Invalide	448	4.22	66.54	455481.5	4.19	66.12
S'occupe de son ménage	313	2.95	69.49	338918.6	3.12	69.24
Bébé, écolier ou étudiant	3095	29.17	98.65	3180745	29.26	98.5
Autres non actifs	143	1.35	100	163017.3	1.5	100
<b>Total</b>	10612	100		10869268	100	

Number of missing values: 170

*Remarque*

La différence entre le chiffre de population 10,869,268 dans ce tableau et le chiffre de population est due au nombre de valeurs manquantes pour la variable Socstat.