

Statbel (Algemene Directie Statistiek – Statistics Belgium)

Thematische Directie Samenleving

Dataverzameling

Certificatie en Methodologische Coördinatie

Huishoudbudgetonderzoek 2016

Methodologische nota

Contactpersoon: Kelly Sabbe

Statbel (Algemene Directie Statistiek - Statistics Belgium)

Elke werkdag van 9 tot 16 uur.

Inhoudstafel

1. Inleiding	4
2. Aantal deelnemende huishoudens	4
3. De steekproeven voor HBS 2016 T1-T4	5
3.1. Inleiding	5
3.2. Trap 1 : trekking van geografische eenheden (of PSE's).....	5
3.2.1. Het steekproefkader van PSE's	5
3.2.2. Het steekproefplan voor trekking van PSE's	7
3.2.3. Spreiding van de PSE's over het enquêtejaar of -semester	9
3.3. Trap 2 : trekking van huishoudens (SSE's) voor LFS	9
3.3.1. Het steekproefkader van SSE's	9
3.3.2. Het steekproefplan voor trekking van LFS-huishoudens (SSE's)	10
3.4. Trap 3: trekking van huishoudens (TSE's) voor HBS	12
3.4.1. Het steekproefkader van TSE's, en de selectie van TSE's.....	12
3.4.2. Het steekproefplan voor trekking van HBS-huishoudens (TSE's).....	15
3.5. Steekproefgrootten	17
3.6. Bijkomende postale rekrutering.....	21
4. Respons	22
4.1. Responsaantallen en responsgraden.....	22
4.2. Kalibratie: voorbereidende beschouwingen.....	27
5. Kalibratie.....	29
5.1. Kalibratie van LFS 2016.....	29
5.1.1. Motivatie van een herziening van de kalibratie van LFS.....	29
5.1.2. Klassieke kalibratie op individueel niveau	30
5.1.3. Vernieuwde kalibratie, simultaan op individueel en op huishoudniveau	32
5.2. Kalibratie van HBS 2016.....	35
5.2.1. Het traditionele kalibratiemodel voor HBS.....	36
5.2.2. Uitbreiding om evenwicht tussen de kwartalen te realiseren	38
5.2.3. Uitbreiding om een vertekening te reduceren.....	40
6. Coderings-, controle- en imputatieprogramma's.....	44

6.1.	Modernisering van de gegevensverzameling	44
6.2.	Controles en imputatie	45
6.2.1.	Opsporing van outliers (aberrante waarden)	45
6.2.2.	Imputatie van energie-uitgaven	46
6.2.3.	Inkomens	51
6.2.4.	Imputed rent	52
1.	Enkele definities	54
2.	Samenstelling van de steekproef van respondenten (representativiteit)	55
2.1.	Huishoudverdelingen	55
2.2.	Individuele verdelingen	56

1. Inleiding

Het Huishoudbudgetonderzoek (HBO; we gebruiken hierna de Engelse afkorting HBS – Household Budget Survey) werd in 2012 hervormd om cijfers van betere kwaliteit te kunnen produceren, die beter aansluiten bij de vragen van onze belangrijkste gebruikers en die zo snel mogelijk volgen op de periode van de gegevensverzameling.

HBS 2014 werd op dezelfde manier georganiseerd als HBS 2012, met enkele verbeteringen met betrekking tot de vragenlijst (interviews uitgevoerd in CAPI).

HBS 2016 werd georganiseerd op dezelfde manier als in 2014.

2. Aantal deelnemende huishoudens

Het aantal deelnemende huishoudens aan de (tweejaarlijkse) enquête bedroeg 4490 in 2016.

3. De steekproeven voor HBS 2016 T1-T4

3.1. Inleiding

De HBS-steekproeven – een voor elk kwartaal – van huishoudens voor 2016 T1-T4 worden getrokken uit LFS-steekproeven. De LFS-steekproeven zijn *tweetrapssteekproeven*. Voor HBS is er de bijkomende selectie van huishoudens uit de LFS-steekproef; dit wordt hieronder beschreven als een *derde trap*. De HBS-steekproeven voor 2016 T1-T4 worden hier dus voorgesteld als *drietrapsteekproeven*:

- Trap 1: trekking van *primaire steekprofeenheden* (PSE's) uit een uit het Rijksregister afgeleid steekproefkader van PSE's, die (geconstrueerde; zie paragraaf 3.2.1) geografische eenheden zijn;
- Trap 2: trekking van *secundaire steekprofeenheden* (SSE's) uit een uit het Rijksregister afgeleid steekproefkader van SSE's, die de *verkiezbare* huishoudens zijn, behorende tot de in trap 1 geselecteerde PSE's;
- Trap 3: trekking van *tertiaire steekprofeenheden* (TSE's) uit de trap 2-steekproef van huishoudens; tertiaire eenheden zijn dus eveneens huishoudens.

De LFS-steekproeven zijn dus het resultaat van de trekkingen in trap 1 en trap 2; de HBS-steekproeven worden bekomen in trap 3. We merken op dat:

- Om organisatorische redenen, de kwartaalsteekproeven voor HBS 2016 T1-T4 getrokken worden uit de LFS-steekproeven van 2 kwartalen (6 maanden) vroeger:
 - HBS 2016 T1 uit LFS 2015 T3,
 - HBS 2016 T2 uit LFS 2015 T4,
 - HBS 2016 T3 uit LFS 2016 T1,
 - HBS 2016 T4 uit LFS 2016 T2;
- Geen rekening gehouden wordt met de respons voor LFS, wanneer de huishoudens voor HBS worden geselecteerd.

In dit hoofdstuk 3 worden uitsluitend *bruto* of *initiële steekproeven* besproken: non-respons is hier niet aan de orde.

3.2. Trap 1 : trekking van geografische eenheden (of PSE's)

3.2.1. *Het steekproefkader van PSE's*

Het steekproefkader van PSE's in trap 1 is een lijst van geografische eenheden die de zgn. wijken ("statistische secties"), of groeperingen hiervan binnen deelgemeenten ("statistische letters"), zijn. We bespreken dit verderop in deze paragraaf in meer detail. Elke PSE bevat een aantal private

huishoudens (SSE's) dat voldoende is om minstens één zgn. *groep van huishoudens* te vormen (in trap 2); dit aantal noemen we de grootte van de PSE (en het bepaalt de kans dat de PSE getrokken wordt). Elke groep van huishoudens wordt in zijn geheel aan een enkele enquêteur toegewezen, en moet om organisatorische redenen een bepaald aantal huishoudens, de zgn. *groeps grootte*, bevatten.

De PSE's voor de 4 kwartaalsteekproeven (voor LFS) van enquêtejaar 2015, werden alle op hetzelfde moment getrokken, d.w.z. als een enkele steekproef; het steekproefkader van PSE's werd afgeleid uit een extractie uit het Rijksregister d.d. 30/08/2014, aangemaakt op 11/09/2014. Voor LFS-enquêtejaar 2016 werden, met het oog op de invoering van het LFS-panel vanaf het 3^{de} kwartaal van 2016, slechts de PSE's voor de steekproeven van 1^{ste} en 2^{de} kwartaal in een keer getrokken; het steekproefkader daarvoor werd afgeleid uit een extractie d.d. 10/10/2015, aangemaakt op 14/10/2015. Merk op dat deze extracties dateren van drie à vier maanden voor de aanvang van de referentieperioden (2015 T1-T4 en 2016 T1-T2 resp.) van de LFS-enquêtes. Dat is nodig, omdat de geselecteerde PSE's tijdig moeten gekend zijn om enquêteurs te rekruteren.

Het steekproefkader van PSE's (voor een bepaalde LFS-referentieperiode) wordt als volgt geconstrueerd. De *statistische sectoren* (binnen een gemeente) worden geïdentificeerd door een code bestaande uit 4 tekens (letters, cijfers, koppelteken): vb. A101, A10-, ...¹ We zeggen dat de sectoren binnen een gemeente, waarvan de code met dezelfde 2 tekens begint, deel uitmaken van dezelfde (*statistische*) *sectie* (binnen een gemeente); de code van die sectie bestaat bijgevolg uit die eerste 2 tekens; vb. statistische sectie A1 bevat statistische sectoren A101, A10-, Op die manier kunnen we een lijst van secties – die ook wel *wijken* worden genoemd – creëren, en voor elke sectie een indicator bepalen voor zijn grootte, nl. het aantal private huishoudens. Vervolgens wordt gecontroleerd welke secties “voldoende groot” zijn, en welke secties dat niet zijn. “Voldoende groot” zijn die secties die een aantal private huishoudens bevatten dat volstaat om minstens één groep van huishoudens voor LFS te construeren. De groeps grootte is een vooraf vastgelegde waarde; voor de enquêtes van 2015 en 2016 is die waarde (op het moment dat de PSE's worden geconstrueerd en getrokken) 23 in Vlaanderen en Wallonië, en 26 in het Brussels Hoofdstedelijk Gewest. Onvoldoende grote secties worden ofwel samengevoegd tot een voldoende grote groepering van secties, ofwel samengevoegd met de kleinste(n) van de voldoende grote secties. De eventuele groepering van secties gebeurt altijd binnen de zgn. (*statistische*) *letters of deelgemeenten*.² De eenheden die hieruit ontstaan zijn de PSE's. “Voldoende groot” zijn uiteindelijk die PSE's die een aantal private huishoudens bevatten dat volstaat om minstens één groep van huishoudens te construeren. Die uiteindelijke lijst van voldoende grote PSE's is het steekproefkader van PSE's in trap 1 van het steekproefplan.

¹ Een (statistische) sector is een geografisch gebied, dat binnen een gemeente geïdentificeerd is door een code die uit 4 tekens bestaat, waarvan het eerste teken normaal een letter (A, B, ...) is. Zie

https://statbel.fgov.be/sites/default/files/files/opendata/Statistische%20sectoren/Secteurs%20stat-NL_tcm325-174181.pdf,
<https://statbel.fgov.be/nl/open-data/statistische-sectoren> en http://nl.wikipedia.org/wiki/Statistische_sector.

² Een deelgemeente of (statistische) letter is een geografisch gebied, dat binnen een gemeente geïdentificeerd wordt door het eerste teken, nl. een letter (A, B, ...), van alle statistische sectoren die tot de letter of deelgemeente behoren. Zie vorige voetnoot voor referenties.

Een speciale code “9999” wordt in de extracties uit het Rijksregister gebruikt om aan te duiden dat de statistische sector van de woonplaats van een huishouden niet is gekend. De *fictieve* sectoren 9999 (binnen elke gemeente), en bijgevolg ook de *fictieve* secties 99 (binnen elke gemeente), worden niet weerhouden in het steekproefkader van PSE’s.

Ten slotte merken we ook op dat onvoldoende grote secties, die overblijven na de hierboven besproken methode van groepering, en dus niet met andere secties zijn kunnen samengevoegd worden, eveneens niet worden weerhouden in het steekproefkader van PSE’s.

3.2.2. Het steekproefplan voor trekking van PSE’s

De steekproef van PSE’s voor een gegeven LFS-referentieperiode is een *gestratificeerde systematische steekproef* met aselekt startpunt in elk stratum, en met selectiekansen die evenredig zijn aan de grootte van de PSE’s:

- [1] De stratificatie van de PSE’s is gebaseerd op de NUTS 2-indeling van het Rijk. Het Brussels Hoofdstedelijk Gewest en de provincies behalve Luik, vormen elk een stratum. In de provincie Luik worden twee strata bepaald, die overeenkomen met respectievelijk de Franstalige gemeenten en de (9) Duitstalige gemeenten. Er werd dus met 12 strata voor zowel LFS 2015 T1-T4 als voor LFS 2016 T1-T2 gewerkt.
- [2] Binnen elk stratum worden de PSE’s gerangschikt op grootte, die, zoals reeds gezegd, gemeten wordt als het aantal private huishoudens in de PSE.
- [3] Vervolgens wordt voor elk stratum het aantal selecties van PSE’s *per kwartaal* bepaald; dit gebeurt in functie van het gewenste aantal responderende (d.w.z. effectief deelnemende) huishoudens, de verwachte responsgraad (in LFS), en de gewenste omvang van de groepen van huishoudens. Het aantal selecties van PSE’s is het aantal voor het stratum te vormen groepen van huishoudens, die elk aan een enquêteur worden toegewezen. (Met dien verstande dat elke enquêteur evt. meerdere groepen van huishoudens voor zijn of haar rekening kan nemen.)

Tabel 1.A geeft voor LFS 2015 en LFS 2016 volgende informatie: de strata, de groepsgrootte in functie van de strata en de aantallen PSE-selecties per stratum en *per kwartaal*. Voor LFS 2015 zijn enkel de aantallen PSE-selecties voor T3 en T4 vermeld, maar die voor T1 en T2 zijn gelijk aan die voor T3 en T4, zodat het aantal PSE-selecties per stratum voor het hele jaar vier keer het aantal per kwartaal is. Voor LFS 2016 zijn de aantallen voor T1 en T2 vermeld; de aantallen voor T3 en T4 zijn lager in het kader van de invoer van het panel, maar deze doen in dit rapport niet ter zake.

Tabel 1.A. Trap 1-stratificatie van PSE's

Stratum			Groepsgrootte (a priori)	Aantal PSE-selecties of groepen van huishoudens per kwartaal			
				LFS 2015		LFS 2016	
				T3	T4	T1	T2
1	ANT	Antwerpen	23	70	70	70	70
2	BRU	Brussels Hoofdstedelijk Gewest	26	91	91	91	91
3	WVL	West-Vlaanderen	23	58	58	58	58
4	OVL	Oost-Vlaanderen	23	63	63	63	63
5	HAI	Henegouwen	23	62	62	62	62
6	LIE	Luik	23	54	54	54	54
7	LIM	Limburg	23	47	47	47	47
8	LUX	Luxemburg	23	34	34	34	34
9	NAM	Namen	23	36	36	36	36
10	VLB	Vlaams-Brabant	23	54	54	54	54
11	BWA	Waals-Brabant	23	34	34	34	34
12	DEU	Duitstalige Gemeenschap	23	21	21	21	21
Totaal aantal groepen per kwartaal :				624	624	624	624

[4] Binnen elk stratum wordt ten slotte een systematische steekproef met aselect startpunt en met selectiekansen die evenredig zijn aan de grootte van de PSE's getrokken. De grootte van de steekproef voor elk stratum is het in [3] besproken aantal PSE-selecties *per jaar* voor 2015 en voor het *eerste semester* voor 2016. Enkele kenmerken van een dergelijke trekking zijn:

- De grootste PSE's worden met zekerheid getrokken, nl. als de PSE groter is dan de staplengte in de systematische trekking (de selectiekans is dan 1).
- Grotere PSE's kunnen meerdere keren geselecteerd worden.
- Kleinere PSE's worden mogelijk niet getrokken.
- Maar als kleinere, in de sortering opeenvolgende PSE's samen in grootte de staplengte overschrijden, is selectie van een of meerdere onder hen wel gegarandeerd.

De grotere PSE's kunnen op deze manier meerdere keren worden geselecteerd; m.a.w. met een PSE kunnen eventueel meerdere selecties overeenkomen.

3.2.3. Spreiding van de PSE's over het enquêtejaar of -semester

De verschillende selecties van PSE's voor 2015 worden uniform gespreid over de 52 (of 53) referentieweken in het hele LFS-enquêtejaar 2015. De verschillende selecties van PSE's voor 2016 T1-T2 worden uniform gespreid over de 26 referentieweken in het eerste semester van 2016. Meerdere selecties van eenzelfde PSE worden maximaal over de kwartalen van een LFS-enquêtejaar verdeeld. Op die manier wordt de representativiteit van de geselecteerde PSE's per kwartaal maximaal gegarandeerd.

Hierna is elke PSE-selectie, en dus elke groep van (nog te selecteren) huishoudens, definitief gelinkt aan een referentieweek.

3.3. Trap 2 : trekking van huishoudens (SSE's) voor LFS

3.3.1. Het steekproefkader van SSE's

De steekproeven van huishoudens worden kwartaal na kwartaal getrokken, en een steekproefkader van huishoudens wordt elk kwartaal opnieuw geconstrueerd, telkens op basis van een zo recent mogelijke extractie uit het Rijksregister:

- voor LFS 2015 T3 (en HBS 2016 T1) : extractie d.d. 16/05/2015
- voor LFS 2015 T4 (en HBS 2016 T2) : extractie d.d. 15/08/2015
- voor LFS 2016 T1 (en HBS 2016 T3) : extractie d.d. 14/11/2015
- voor LFS 2016 T2 (en HBS 2016 T4) : extractie d.d. 05/03/2016

Het steekproefkader van huishoudens (SSE's) voor een gegeven kwartaal, afgeleid uit de overeenkomstige extractie uit het Rijksregister waarin alle Belgische huishoudens kunnen gevonden worden, bestaat uit de huishoudens die voldoen aan de volgende criteria:

- [1] Enkel de huishoudens die behoren tot de in trap 1 geselecteerde PSE's worden weerhouden.

- [2] We beperken ons tot private huishoudens; collectieve huishoudens zoals rusthuizen, gevangenissen, ... zijn dus uitgesloten.
- [3] Huishoudens die geselecteerd zijn geweest in de LFS-steekproef voor een van de voorbije 10 kwartalen, komen niet meer in aanmerking voor selectie.
- [4] Minstens één lid van het huishouden heeft minstens de leeftijd van 15 jaar en hoogstens de leeftijd van 76 jaar bereikt. Hiertoe worden de leeftijden van de huishoudleden berekend op de zondag van de referentieweek (d.i. de laatste dag van de referentieweek) die volgens paragraaf 3.2.3 bepaald is voor de PSE waartoe het huishouden behoort.

3.3.2. *Het steekproefplan voor trekking van LFS-huishoudens (SSE's)*

De steekproef van LFS-huishoudens (SSE's) voor een gegeven kwartaal, uit het voor dat kwartaal geconstrueerde steekproefkader van SSE's, is een *gestratificeerde enkelvoudige aselechte steekproef*:

- [1] De strata zijn de in trap 1 voor het kwartaal geselecteerde *verschillende* PSE's.
- [2] Het aantal huishoudens dat in elk stratum of elke PSE geselecteerd wordt, is het product van de groepsgrootte voor de PSE en het aantal keren dat de PSE geselecteerd is voor het gegeven kwartaal.
- [3] Binnen elk stratum of elke PSE worden de huishoudens enkelvoudig aselekt getrokken.

Als een PSE meerdere keren werd geselecteerd (voor het gegeven kwartaal), dan zullen de getrokken huishoudens aselekt en uniform over de PSE-selecties verdeeld worden. Aan elke PSE-selectie zal daardoor eenzelfde aantal huishoudens worden toegewezen; dat wordt een (LFS-)groep van huishoudens genoemd, en het aantal huishoudens in de groep is de groepsgrootte.

Tabel 1.B. Groepsgrootten bij trekking van huishoudens, en totale initiële steekproefgrootten

Stratum			Groepsgroote bij trekking van HHs	Aantal PSE-selecties of groepen van huishoudens per kwartaal			
				LFS 2015		LFS 2016	
				T3	T4	T1	T2
1	ANT	Antwerpen	20	70	70	70	70
2	BRU	Brussels Hoofdstedelijk Gewest	23	91	91	91	91
3	WVL	West-Vlaanderen	20	58	58	58	58
4	OVL	Oost-Vlaanderen	20	63	63	63	63
5	HAI	Hainaut	20	62	62	62	62
6	LIE	Luik	20	54	54	54	54
7	LIM	Limburg	20	47	47	47	47
8	LUX	Luxemburg	20	34	34	34	34
9	NAM	Namen	20	36	36	36	36
10	VLB	Vlaams-Brabant	20	54	54	54	54
11	BWA	Waals-Brabant	20	34	34	34	34
12	DEU	Duitstalige Gemeenschap	23	21	21	21	21
Totaal aantal huishoudens per kwartaal :				12816	12816	12816	12816

In paragraaf 3.2.1 hebben we de groepsgroote vermeld die gebruikt wordt bij constructie van de PSE's: 26 voor het Brussels Hoofdstedelijke Gewest en 23 voor Vlaanderen en Wallonië (zie tabel 1.A). Normaal gezien worden dezelfde groepsgrootten gebruikt bij de trekking van de huishoudens (trap 2), maar voor de kwartalen 2015 T3-T4 en 2016 T1-T2 werd daar in laatste instantie van afgeweken om budgettaire redenen. De groepsgroote werd voor die perioden verminderd met 3 huishoudens, behalve voor het stratum Duitstalige Gemeenschap. In tabel 1.B herhalen we de strata, de aantallen groepen per stratum en kwartaal, zoals in tabel 1.A, maar de groepsgrootten zijn aangepast; in de laatste lijn worden de resulterende totale initiële

steekproefgrootten (in aantallen huishoudens) per kwartaal gegeven. (Als de oorspronkelijke groepsgrootten zouden zijn behouden, dan zou het totaal aantal initieel getrokken huishoudens per kwartaal gelijk zijn aan 14625, zoals bijvoorbeeld het geval was in 2015 T1 en T2.) Merk op dat de groeps grootte geen rol speelt in de selectie van de PSE's (in trap 1); de groeps grootte speelt wel een rol in de constructie van het PSE-steekproefkader, maar omdat tussen trap 1 en trap 2 de groeps grootte werd verlaagd, heeft dit geen negatieve impact op de mogelijkheid om groepen van de gewenste omvang (uiteindelijk dus 20 of 23) te maken.

3.4. Trap 3: trekking van huishoudens (TSE's) voor HBS

3.4.1. Het steekproefkader van TSE's, en de selectie van TSE's

Voor HBS zijn in principe alle huishoudens die getrokken werden voor LFS (in trap 2) verkiesbaar, met uitzondering van de huishoudens in Duitstalige gemeenten: voor HBS worden geen huishoudens in de Duitstalige Gemeenschap geselecteerd. Voor elk HBS-kwartaal is dus de initiële LFS-steekproef van huishoudens, uitgezonderd deze in Duitstalige gemeenten, het steekproefkader voor trekking van huishoudens (TSE's) in trap 3.

Voor elk van de eerste 3 HBS-kwartalen werd beslist slechts gebruik te maken van de voor LFS getrokken groepen van huishoudens die toegewezen werden aan de eerste 7 weken van het kwartaal: er werd verwacht dat de responsgraad van HBS 2014 ook voor HBS 2016 zou worden geobserveerd, en in totaal werden 5000 responderende huishoudens nagestreefd voor HBS 2016. Voor het vierde kwartaal werden dan weer wel alle LFS-huishoudens als verkiesbaar beschouwd. We bespreken dit nu voor elk van de 4 kwartalen in meer detail.

Voor het eerste HBS-kwartaal (2016 T1), d.i. kwartaal 2015 T3 voor LFS, werden aan elke week $624 : 13 = 48$ groepen van huishoudens toegewezen. Zonder de 1 of 2 groepen per week voor de Duitstalige Gemeenschap, geeft dat 47 of 46 groepen van huishoudens per week (week 1-7), zoals is aangegeven in de kolom met hoofding "HBS T1" in tabel 2.A. Die 47 of 46 groepen per LFS-referentieweek worden vervolgens aan HBS-referentiemaanden toegewezen, volgens een systeem dat uiteindelijk resulteert in ongeveer 1/3 van het totaal aantal groepen (327) voor elke HBS-maand. Dit systeem houdt in dat de groepen voor de eerste LFS-weeken zoveel mogelijk aan de eerste HBS-maand worden toegewezen, dat de groepen voor de laatste LFS-weeken zoveel mogelijk aan de derde HBS-maand worden toegewezen, en dat de groepen voor de middelste LFS-weeken zoveel mogelijk aan de tweede HBS-maand worden toegewezen. Een gevolg is dat de groepen voor de derde LFS-week aan de eerste en de tweede HBS-maand worden toegewezen in een verhouding van ongeveer 1:2, en dat de groepen voor de vijfde LFS-week aan de tweede en de derde HBS-maand worden toegewezen in een verhouding van ongeveer 2:1. De resulterende aantallen groepen voor het eerste HBS-kwartaal, per combinatie van LFS-week en HBS-maand, zijn gepresenteerd in tabel 2.A in de kolommen met hoofding "HBS-maand 1, 2 of 3".

Het tweede luik van tabel 2.A laat zien hoe de LFS-groepen van weken 40-46 in kwartaal 2015 T4 (dus de eerste 7 weken van dat kwartaal) werden toegewezen aan de HBS-maanden 4, 5 en 6. De aantallen liggen hier iets lager dan in het eerste luik van de tabel, als gevolg van het feit dat

de 624 LFS-groepen verdeeld werden over 14 (i.p.v. de gebruikelijke 13) weken in het kwartaal (624 : 14 = 44,57, wat 45 of 44 LFS-groepen per referentieweek impliceert, te verminderen met de 1 of 2 groepen per week voor de Duitstalige Gemeenschap).

Tabel 2.A. Omzetting van referentieweken (LFS) in enquêtemaanden (HBS) – Eerste twee kwartalen

LFS 2015 T3 → HBS 2016 T1					LFS 2015 T4 → HBS 2016 T2				
LFS Ref.week	HBS-maand			HBS T1	LFS Ref.week	HBS-maand			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
Totaal	109	110	108	327	Totaal	102	99	101	302

Tabel 2.B. Omzetting van referentieweken (LFS) in enquêtemaanden (HBS) - Laatste twee kwartalen

LFS 2016 T1 → HBS 2016 T3					LFS 2016 T2 → HBS 2016 T4				
LFS Ref.week	HBS-maand			HBS T3	LFS Ref.week	HBS-maand			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
Totaal	109	110	108	327	Totaal	201	202	200	603

Voor het derde HBS-kwartaal (2016 T3) is de werkwijze volledig analoog aan deze voor het eerste HBS-kwartaal: het eerste luik van tabel 2.B is analoog aan het eerste luik van tabel 2.A.

Voor het vierde HBS-kwartaal (2016 T4) was oorspronkelijk de werkwijze eveneens analoog aan deze voor het eerste en het derde kwartaal, maar achteraf werd toch beslist om alle LFS-groepen van LFS 2016 T2 te gebruiken. Luik 2 van tabel 2.B laat zien hoe de LFS-groepen voor weken 21-26 aan de HBS-maanden 10, 11 en 12 werden toegewezen. De reden voor deze uitbreiding van de steekproef voor HBS 2016 T4, was de lager dan verwachte respons (ongeveer 20%, zoals geobserveerd in 2014) in de eerste kwartalen (2016 T1 en T2), in combinatie met het streefcijfer van 5000 responderende huishoudens voor het volledige enquêtejaar 2016.

3.4.2. Het steekproefplan voor trekking van HBS-huishoudens (TSE's)

Voor elk HBS-kwartaal kunnen we de geselecteerde groepen van huishoudens, zoals aangegeven in tabellen 2.A en 2.B, ten slotte herschikken volgens stratum en HBS-maand. Dat leidt tot de tabellen 3.A en 3.B.

Tabel 3.A. Trap 3-stratificatie en allocatie van LFS-groepen
- Eerste twee kwartalen

		HBS 2016 T1				HBS 2016 T2			
PSE-stratum		HBS-maand			T1	HBS-maand			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
Totaal		109	110	108	327	102	99	101	302

Beschouw nu in tabel 3.A het eerste luik, voor HBS-kwartaal 2016 T1. De kolom met hoofding "T1" geeft het aantal LFS-groepen van huishoudens die geselecteerd werden uit de in totaal 624 groepen (voor LFS-kwartaal 2015 T3). De selectieprocedure werd uiteengezet in vorige paragraaf 3.4.1. Die procedure leek helemaal niet aselekt te zijn. Toch zullen we vanaf nu aannemen dat bijvoorbeeld de 40 groepen van huishoudens voor stratum ANT (tabel 3.A, luik 1), aselekt werden getrokken uit de oorspronkelijk 70 groepen van huishoudens in stratum ANT (tabel 1.A of tabel 1.B). Immers, die 40 groepen voor ANT zijn alle groepen voor ANT die aan de eerste 7 weken in LFS-kwartaal 2015 T3 werden toegewezen, maar, vermits de 70 groepen voor ANT in wezen

willekeurig aan de 13 weken in kwartaal 2015 T3 werden toegewezen, kunnen we de 40 voor HBS 2016 T1 in ANT geselecteerde groepen als een aselechte trekking uit 70 groepen opvatten. Hetzelfde geldt voor elk ander stratum (BRU, WV, ... BWA) en voor elk HBS-kwartaal.

Tabel 3.B. Trap 3-stratificatie en allocatie van LFS-groepen
- *Laatste twee kwartalen*

		HBS 2016 T3				HBS 2016 T4			
PSE-stratum		HBS-maand			T3	HBS-maand			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
Totaal		109	110	108	327	201	202	200	603

Dit betekent dat de trekking in trap 3, voor elk kwartaal afzonderlijk, kan opgevat worden als een *gestratificeerde aselechte clustersteekproef* van huishoudens; de clusters zijn LFS-groepen van huishoudens zoals gevormd in trap 2.

Merk op dat de trekking voor HBS 2016 T4 uit LFS 2016 T2 exhaustief is.

De beschouwingen in deze paragraaf met betrekking tot het random of aselechte karakter van de trekking in trap 3, zijn belangrijk voor de correcte berekening van steekproefgewichten voor de huishoudens in de HBS-steekproef (wat belangrijk is voor de kalibratie) en voor de schatting van varianties van geschatte indicatoren.

3.5. Steekproefgrootten

De bruto steekproefgrootten, zowel in aantallen groepen van huishoudens als in aantallen huishoudens, naar (PSE-) stratum per kwartaal, zijn samengevat in volgende tabellen.

Tabel 4.A. Bruto LFS- en HBS-steekproeven van huishoudens t.b.v. HBS 2016 T1

LFS 2015 T3				HBS 2016 T1		
Stratum		Groeps- grootte	Aantal groepen	Aantal HH's	Aantal LFS- groepen	Aantal HH's
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	-	-
Totaal			624	12816	327	6687

Tabel 4.B. Bruto LFS- en HBS-steekproeven van huishoudens t.b.v. HBS 2016 T2

LFS 2015 T4				HBS 2016 T2		
Stratum		Groeps- grootte	Aantal groepen	Aantal HH's	Aantal LFS- groepen	Aantal HH's
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	-	-
Totaal			624	12816	302	6187

Tabel 4.C. Bruto LFS- en HBS-steekproeven van huishoudens t.b.v. HBS 2016 T3

LFS 2016 T1				HBS 2016 T3		
Stratum		Groeps- grootte	Aantal groepen	Aantal HH's	Aantal LFS- groepen	Aantal HH's
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	-	-
Totaal			624	12816	327	6687

Tabel 4.D. Bruto LFS- en HBS-steekproeven van huishoudens t.b.v. HBS 2016 T4

LFS 2016 T2				HBS 2016 T4		
Stratum		Groeps-grootte	Aantal groepen	Aantal HH's	Aantal LFS-groepen	Aantal HH's
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	-	-
Totaal			624	12816	603	12333

3.6. Bijkomende postale rekrutering

De selectie in trap 3 van (groepen van) huishoudens uit de LFS-steekproeven van huishoudens, zoals hierboven beschreven, had als doelstelling uiteindelijk ongeveer 5000 responderende HBS-huishoudens te realiseren, en maakte impliciet gebruik van een in 2014 geobserveerde (conditionele) responsgraad van ongeveer 20%. Dit verklaart de aanvankelijke beslissing om de rekrutering voor HBS te beperken tot de LFS-huishoudens die voor de eerste 7 weken in elk kwartaal werden geselecteerd. Immers, die procedure zou volgens de verwachtingen moeten leiden tot $(6687 \times 3 + 6187) \times 20\% \approx 5250$ responderende HBS-huishoudens.

Na het eerste kwartaal werd duidelijk dat het objectief van 5.000 huishoudens niet haalbaar was met het initiële steekproefdesign en de verwachte responsgraad van 20%. Om op het tekort aan huishoudens in te spelen, werd beslist om (1) voor het 4^{de} HBS-kwartaal alle LFS-huishoudens van 2016 T2 te rekruteren (zoals eerder reeds werd besproken), en (2) een extra postale rekrutering te organiseren. Huishoudens uit de LFS-steekproeven van 2015 T3, 2015 T4 en 2016 T1, die behoorden tot groepen uit week 8 t/m 13 (of 14) en die effectief meewerkten aan de LFS-enquête, werden postaal gevraagd deel te nemen aan HBS.

Een volledig overzicht van de bronnen voor de 4 trimestriële HBS-steekproeven wordt in tabel 5 gepresenteerd.

Tabel 5. Bronnen en wijze van rekrutering voor de kwartaalsteekproeven van HBS 2016

Wijze van rekrutering	HBS 2016 T1	HBS 2016 T2	HBS 2016 T3	HBS 2016 T4
Face-to-face (FtF)	LFS 2015 T3 w1-w7	LFS 2015 T4 w1-w7	LFS 2016 T1 w1-w7	LFS 2016 T2 w1-w13
Postaal	--	--	LFS 2015 T3 w8-w13 LFS 2015 T4 w8-w14	LFS 2016 T1 w8-w13

4. Respons

4.1. Responsaantallen en responsgraden

HBS 2016 is een enquête die geïntegreerd werd in de Enquête naar de Arbeidskrachten (EAK, of LFS – Labour Force Survey). Wij verstaan onder *geïntegreerde enquête* de selectie van huishoudens onder de huishoudens die geantwoord hebben op een andere, voorafgaande, enquête – in het geval van HBS is dat dus LFS – en die op vrijwillige basis hebben toegestemd om mee te werken aan HBS, nadat ze hiervoor werden uitgenodigd door een LFS-enquêteur tijdens een persoonlijk contact (face-to-face) of per brief, enkele weken na de periode van de LFS. Het steekproefplan voor de geïntegreerde enquête werd beschreven in hoofdstuk 3.

Aangezien het praktisch gezien essentieel is om met groepen huishoudens te werken, werd er een enkelvoudige clustersteekproef van huishoudens getrokken uit de LFS-steekproef (zie paragraaf 3.4.2). De clusters zijn de LFS-groepen die op een willekeurige manier werden geselecteerd. Het voordeel is dat LFS-groepen van huishoudens behouden worden als HBS-groepen. In principe is het dus de LFS-enquêteur die blijft werken met een groep huishoudens die in een relatief beperkt geografisch gebied wonen. De HBS-enquêteur is bij voorkeur dezelfde enquêteur als diegene die de LFS heeft uitgevoerd. Die enquêteur is al bekend bij de huishoudens en heeft zo veel mogelijk huishoudens ertoe aangezet om deel te nemen aan de nieuwe enquête. Voor sommige groepen echter, wanneer de LFS-enquêteur niet bereid was ook de HBS-enquête uit te voeren, moest een andere enquêteur gerekruteerd worden. Het feit dat voor sommige groepen geen nieuwe enquêteur werd gevonden, en dat een nieuwe enquêteur minder vertrouwen inboezemt dan een reeds gekende, heeft de responsgraad in zekere mate negatief beïnvloed.

De responsaantallen en de responspercentages voor HBS 2016 zijn in tabel 6 samengebracht. Merk op dat in tabel 6 de *bruto responspercentages* worden weergegeven, d.w.z. de respons met inbegrip van de groepen van huishoudens die “niet zijn uitgezet”. Als we deze huishoudens niet opnemen, bekomt men de *netto responspercentages*, die uiteraard hoger zullen liggen dan de bruto percentages.

De steekproef voor HBS bestaat uitsluitend uit huishoudens die eerder deelnamen aan LFS in de periode 2015 T3 t.e.m. 2016 T2, met uitsluiting van huishoudens in Duitstalige gemeenten; dus: $49.332 = 4 \times (12816 - 483)$, zie tabellen 4.A-D. De respons van HBS is een combinatie van 3 factoren:

- deelname aan LFS,
- de rekruteringsmethode na de LFS enquête,
- deelname aan HBS.

De respons (“volledig HBS dossier” in tabellen 6 en 7) ten opzichte van de initiële LFS-steekproef bedraagt 9,1%. D.w.z. dat van elke 100 huishoudens die in aanmerking kwamen om deel te nemen aan LFS en nadien HBS, er uiteindelijk 9 huishoudens tot het einde hebben meegewerkt en hun antwoorden weerhouden werden na kwaliteitscontroles. De respons in 2016 ligt door o.a.

het hoger aantal weigeringen en de gemengde rekruteringsmethode veel lager dan de 17,0% in 2014.

Tabel 6. Bruto respons HBS t.o.v. de initiële LFS steekproef (geconsolideerde responscodes)

	Vlaanderen		Wallonië		Brussel		Totaal	
	N	%	N	%	N	%	N	%
Volledig HBS dossier	2.175	9,3	1.784	10,1	531	6,3	4.490	9,1
Overleden	18	0,1	19	0,1	4	0,0	41	0,1
Verhuisd	310	1,3	259	1,5	427	5,1	996	2,0
Collectief huishouden	13	0,1	14	0,1	17	0,2	44	0,1
Taalbarrière	87	0,4	33	0,2	63	0,8	183	0,4
Vakantie, zakenreis	280	1,2	223	1,3	168	2,0	671	1,4
Ziek, gehandicapt, dement	157	0,7	74	0,4	25	0,3	256	0,5
Fout adres	194	0,8	115	0,7	70	0,8	379	0,8
Weigering	16.086	68,9	11.638	66,1	5.630	67,2	33.354	67,6
Niet thuis	1.601	6,9	1.735	9,9	891	10,6	4.227	8,6
Andere	412	1,8	419	2,4	307	3,7	1.138	2,3
Adres niet bezocht	179	0,8	193	1,1	36	0,4	408	0,8
Niet gevalideerd HBS	28	0,1	52	0,3	19	0,2	99	0,2
Geen informatie	0	0,0	22	0,0	0	2,2	22	0,0
Niet uitgezet	1.820	7,8	1.020	5,8	184	2,2	3.024	6,1
Totaal	23.360	100,0	17.600	100,0	8.372	100,0	49.332	100,0

De respons verschilt van **regio** tot regio en was het hoogst in Wallonië waar 10,1% van de huishoudens meewerkte. In Vlaanderen bedraagt de respons 9,3% en in Brussel slechts 6,3%.

De belangrijkste reden voor uitval zijn de weigeringen. In de loop van het proces lieten 67,6% van de huishoudens weten dat ze weigerden om deel te nemen aan de enquête. De weigering vond in de meeste gevallen plaats na de LFS-enquête, wat logisch is aangezien LFS een verplichte enquête is. Verder werden 8,6% van de huishoudens niet thuis gevonden; deze vormen een belangrijk deel van de niet gecontacteerde huishoudens. Door een gebrek aan beschikbare interviewers kon 6,1% van de adressen niet toegewezen worden aan een actieve enquêteur.

De respons ten opzichte van de initiële LFS-steekproef verschilt naargelang de **wijze waarop gerekruteerd werd**; zie tabel 7.

Tabel 7. Bruto responsgraad (%) HBS t.o.v. de initiële LFS steekproef (geconsolideerde responscodes) per rekruteringsmodus

	Vlaanderen		Wallonië		Brussel		Totaal	
	FtF	Post	FtF	Post	FtF	Post	FtF	Post
Volledig HBS dossier	10,7	6,8	12,4	6,0	7,2	4,7	10,7	6,1
Overleden	0,1	0,1	0,1	0,1	0,0	0,1	0,1	0,1
Verhuisd	1,2	1,6	1,5	1,5	5,0	5,4	1,9	2,2
Collectief huishouden	0,1	0,0	0,1	0,0	0,3	0,0	0,1	0,0
Taalbarrière	0,5	0,2	0,2	0,2	0,8	0,7	0,4	0,3
Vakantie, zakenreis	1,4	0,8	1,6	0,6	2,5	1,1	1,7	0,8
Ziek, gehandicapt, dement	0,7	0,7	0,5	0,3	0,3	0,3	0,6	0,5
Fout adres	0,7	1,0	0,7	0,6	0,6	1,3	0,7	0,9
Weigering	67,0	72,2	62,5	72,8	65,2	71,0	65,1	72,2
Niet thuis	6,9	6,8	10,1	9,3	11,1	9,8	8,8	8,2
Andere	2,1	1,2	2,4	2,3	3,8	3,3	2,5	1,9
Adres niet bezocht	0,8	0,7	1,4	0,6	0,4	0,6	0,9	0,6

Niet gevalideerd HBS	0,1	0,1	0,4	0,1	0,3	0,1	0,3	0,1
Geen informatie	0,0	0,0	0,2	0,0	0,0	0,0	0,1	0,0
Niet uitgezet	7,7	7,9	6,0	5,5	2,5	1,6	6,2	6,0
Totaal (aantal)	14.980	8.380	11.440	6.160	5.474	2.898	31.894	17.438

Bij een deel van de steekproef gebeurde dat net als in 2014 **face-to-face (FtF)** na het afronden van de LFS-enquête. Omdat het aantal positieve rekruteringen lager lag dan in 2014, werd beslist om ook bij de positieve LFS-huishoudens die oorspronkelijk niet in aanmerking kwamen voor HBS 2016 (nl. de LFS-huishoudens voor referentieweken 8 t.e.m. 13 (of 14) in de LFS-kwartalen 2015 T3, 2015 T4 en 2016 T1) te rekruteren; zie paragraaf 3.6. Dit gebeurde **postaal** door deze huishoudens een brief en een inschrijvingsformulier te sturen.

De respons is het hoogst indien de rekrutering face-to-face is gebeurd. Gemiddeld werkte dan 10,7% van de huishoudens uit de LFS-steekproef niet alleen mee aan LFS, maar ook aan HBS. Dat cijfer ligt lager dan de 17,0% uit 2014, toen de volledige HBS-steekproef op deze wijze gerekruteerd werd. Bij postaal gerekruteerde huishoudens bedroeg de respons t.o.v. de initiële LFS-steekproef 6,1%, wat vergelijkbaar is met de resultaten van vóór 2012 toen er uitsluitend postaal werd gerekruteerd. Het feit dat de huishoudens hier aanvankelijk ook aan LFS hebben deelgenomen heeft dus amper een positief effect op de bereidheid om ook deel te nemen aan HBS.

Hieronder in tabel 8 staan de resultaten van het rekruteringsaspect, zonder rekening te houden met het uiteindelijke HBS-resultaat. N is het aantal huishoudens dat deelnam aan LFS en vervolgens “ja” antwoordde op de vraag of men na LFS ook wenst deel te nemen aan HBS, ongeacht het uiteindelijke resultaat. Bijvoorbeeld: 7.061 huishoudens hebben deelgenomen aan LFS, en hebben nadien gezegd deel te willen nemen aan HBS toen hun dat door de LFS-enquêteur werd gevraagd; dat is $30,2\%$ van de $7.061/0,302 \approx 23.381$ huishoudens die aan LFS hebben deelgenomen en aan wie de vraag voor deelname aan HBS door de LFS-enquêteur werd gesteld.

Bij de **face-to-face rekrutering** ging dus 30,2% van de huishoudens in op het voorstel van de LFS-enquêteur om ook deel te nemen aan HBS. In 2014 bedroeg dit aandeel 36,7%. Het lagere aandeel kan deels verklaard worden door het quasi halveren van de incentive voor de deelnemende huishoudens. Opvallend zijn de relatief grote verschillen tussen de regio's. In Wallonië was 35,0% bereid om deel te nemen aan HBS, terwijl dit in Vlaanderen slechts 26,5% bedroeg. Bij de **postale rekrutering** zijn de verschillen tussen de regio's minder groot en verklaarde gemiddeld 11,5% zich bereid om mee te werken aan HBS.

Tabel 8. Rekrutering na LFS enquête

	Vlaanderen		Wallonië		Brussel		Totaal	
	N	%	N	%	N	%	N	%
Face-to-face	2.949	26,5	2.923	35,0	1.189	30,1	7.061	30,2
Postaal	751	12,0	526	11,4	219	10,3	1.496	11,5

Niet alle gerekruteerde HBS-huishoudens werkten uiteindelijk mee; zie tabel 9. De respons bij de HBS-enquêteur bedroeg 48,4% indien de rekrutering face-to-face gebeurde (63,4% in 2014) en 71,6% indien het huishouden postaal werd gerekruteerd. Ook hier zijn er regionale verschillen. Ongeacht de rekruteringsmodus werkten in Vlaanderen meer huishoudens tot het einde mee dan in Wallonië en Brussel. De verschillen zijn hoofdzakelijk een gevolg van het hoger aantal weigeringen.

Tabel 9. Brutoresponsgraad (%) HBS t.o.v. gerekruteerde huishoudens per rekruteringsmodus

	Vlaanderen		Wallonië		Brussel		Totaal	
	FtF	Post	FtF	Post	FtF	Post	FtF	Post
Volledig HBS dossier	54,5	75,5	48,4	70,2	33,3	61,6	48,4	71,6
Overleden	0,1	0,0	0,2	0,2	0,0	0,0	0,1	0,1
Verhuisd	0,5	0,4	0,9	0,0	1,1	1,4	0,8	0,4
Collectief huishouden	0,0	0,1	0,0	0,0	0,1	0,0	0,0	0,1
Taalbarrière	0,4	0,1	0,0	0,4	0,6	0,0	0,3	0,2
Vakantie, zakenreis	1,1	0,8	1,1	0,6	3,6	3,2	1,5	1,1
Ziek, gehandicapt, dement	1,2	0,8	0,7	0,4	0,8	0,9	0,9	0,7
Fout adres	0,1	0,3	0,1	0,0	0,2	0,0	0,1	0,1
Weigering	33,9	14,6	40,3	22,8	47,7	23,7	38,8	18,9
Niet thuis	1,3	0,8	2,6	1,3	5,0	3,2	2,4	1,3

Andere	3,2	1,1	3,2	2,1	4,8	3,2	3,5	1,7
Niet bezocht	3,0	4,7	1,1	1,1	1,5	0,9	1,9	2,9
Niet gevalideerd HBS dossier	0,7	0,8	1,5	1,0	1,3	1,8	1,1	1,0
Totaal (aantal)	2.949	751	2.923	526	1.189	219	7.061	1.496

4.2. Kalibratie: voorbereidende beschouwingen

Reeds bij de trekking krijgt elk huishouden een zogenaamd steekproefgewicht (*sampling weight*, *design weight*), d , zeg, dat onmiddellijk volgt uit het steekproefplan en de samenstelling van het steekproefkader.³

Het steekproefplan impliceert dat in elk stratum alle geselecteerde huishoudens (nagenoeg) hetzelfde steekproefgewicht hebben. In tabel 10 geven we voor elk kwartaal het aantal responderende huishoudens en de som van de steekproefgewichten voor de responderende huishoudens.

Tabel 10. Aantal responderende huishoudens en hun som van steekproefgewichten per kwartaal

	HBS 2016 T1	HBS 2016 T2	HBS 2016 T3	HBS 2016 T4	Totaal
Aantal responderende huishoudens	908	783	1.365	1.434	4.490
Som van de steekproefgewichten voor de responderende huishoudens	320.388,55	275.587,79	473.681,65	500.050,58	1.569.708,57

³ We gaan in deze nota niet verder in op de berekening van de steekproefgewichten. Het volstaat hier op te merken dat de steekproefgewichten die we vanaf nu bedoelen, de gewichten zijn die rechtstreeks (volgens de theorie van steekproefplanning) uit het steekproefplan volgen, maar ook reeds aangepast zijn voor bijvoorbeeld het wegvallen van de LFS-groepen van huishoudens in Duitstalige gemeenten.

Dit toont aan dat het aandeel van de kwartalen in gewogen schattingen voor het hele onderzoeksjaar 2016, als gebruikgemaakt zou worden van de steekproefgewichten, erg onevenwichtig zou zijn. Seizoenseffecten zouden op die manier ofwel onterecht versterkt, ofwel onterecht afgezwakt kunnen worden. We zullen later zien hoe we dit via speciale kalibratievergelijkingen kunnen vermijden. Merk ook op dat het totaal van de steekproefgewichten over alle responderende huishoudens een sterke onderschatting van het totaal aantal private huishoudens in België is; het eerste doel van een kalibratie is uiteraard dit probleem weg te werken, wat betekent dat (gemiddeld) alle steekproefgewichten zullen opgehoogd worden.

Er werd voor HBS 2016 geen aan de eigenlijke kalibratie voorafgaande nonresponscorrectie doorgevoerd, zoals wel het geval was voor HBS 2012 en 2014.

5. Kalibratie

De kalibratie van de respondentensteekproef van HBS gebeurt traditioneel met behulp van referentiewaarden (benchmarks, kalibratietotalen) die op basis van LFS worden geschat. Die LFS-schattingen worden zelf ook berekend na kalibratie van LFS. In dit hoofdstuk zullen we daarom eerst de kalibratie van LFS 2016 behandelen (paragraaf 5.1), om daarna de kalibratie van HBS te beschrijven (paragraaf 5.2).

5.1. [Kalibratie van LFS 2016](#)

5.1.1. *Motivatie van een herziening van de kalibratie van LFS*

De kalibratie van LFS gebeurt traditioneel elk trimester volgens een post-stratificatiemodel voor de *individuele* LFS-respondenten. De referentiewaarden zijn hier aantallen *individuen* volgens het Rijksregister, in verschillende subpopulaties (de zgn. post-strata die bepaald worden door de post-stratificatievariabelen; zie verder in paragraaf 5.1.2). Daardoor zijn de LFS-schattingen in zekere mate coherent met populatiecijfers op individueel niveau uit het Rijksregister. Die kalibratie, per kwartaal, werd nog steeds toegepast voor LFS 2016 T1-T4 met het oog op de publicatie van *schattingen over individuen*, zoals bv. tewerkstellingscijfers.

In het verleden werden LFS-schattingen *op niveau van huishoudens*, die als referentiewaarden in de kalibratie van HBS werden gebruikt, ook berekend op basis van gewichten die resulteerden uit de zonet beschreven individuele post-stratificatie van LFS. Die werkwijze heeft duidelijk drie nadelen:

1. De individuele gekalibreerde LFS-gewichten moeten omgezet worden naar huishoudgewichten, alvorens de gewenste schattingen op niveau van huishoudens te kunnen berekenen. Dat gebeurde traditioneel op een *ad hoc* manier, door als huishoudgewicht ofwel het gewicht van de referentiepersoon, ofwel het gemiddelde van de gewichten van de leden van het huishouden te nemen. Deze manier om huishoudgewichten te bepalen heeft geen enkele theoretisch gefundeerde verklaring.
2. De schattingen op niveau van huishoudens, zijn in geen enkel opzicht noodzakelijk coherent met populatiecijfers op niveau van huishoudens uit het Rijksregister. Zo kan bijvoorbeeld het geschatte aantal huishoudens voor heel België sterk verschillen (in positieve of negatieve zin, en dat mogelijk afhankelijk van het onderzoeksjaar) van het aantal huishoudens volgens het Rijksregister. Per gewest, per provincie enz. kunnen die afwijkingen zelfs toenemen. Een ander voorbeeld betreft het aantal *x*-persoons-huishoudens (voor bepaalde $x = 1, 2, \text{ of } \dots$), en het probleem wordt natuurlijk groter als we huishoudgrootte combineren met gewest of provincie. Kortom, voor elk “domein” of subpopulatie van huishoudens zal de afwijking tussen de LFS-schatting van het aantal huishoudens in dat domein en het overeenkomstige aantal uit het Rijksregister groter worden naarmate het domein kleiner wordt.

3. Die incoherentie op huishoudniveau zet zich uiteraard door naar HBS-schattingen als de incoherente LFS-schattingen als referentiewaarden in de kalibratie van HBS worden gebruikt.

Vooraf dit laatste probleem dwingt ons in eerste instantie de kalibratie van LFS te herzien. In paragraaf 5.1.2 zullen we eerst de traditionele kalibratie van LFS op individueel niveau in herinnering brengen. Daarna, in paragraaf 5.1.3, behandelen we het vernieuwde kalibratiemodel voor LFS.

5.1.2. Klassieke kalibratie op individueel niveau

Voor de traditionele kalibratie van LFS is de input uit de enquête als volgt:

1. Een lijst van alle individuele respondenten, met de kenmerken woonplaats, leeftijd en geslacht die in het kalibratiemodel zullen geïntegreerd worden.
2. Voor elke respondent: het steekproefgewicht (berekend uit steekproefplan en -kader voor LFS).

De variabelen woonplaats, leeftijd en geslacht leiden tot de volgende drie kalibratie- of post-stratificatievariabelen:

- STRAT12: de 12 steekproefstrata voor selectie van de PSE's (zie tabel 1.A in paragraaf 3.2.2);
- SEX: geslacht (2 categorieën);
- AGE: 16 leeftijdsklassen (0-4, 5-9, 10-14, ..., 70-74, 75+).

De volledige kruising van deze drie variabelen geeft $12 \times 2 \times 16 = 384$ post-strata. Voor elk van deze post-strata berekenen (tellen!) we uit het Rijksregister het aantal individuen; we noteren N voor elk van deze 384 populatiecijfers.

In elk post-stratum apart berekenen we het individueel gekalibreerd gewicht als volgt:

- a. Bepaal het totaal $\sum d$ van de steekproefgewichten van de respondenten in het post-stratum.
- b. Bepaal de unieke correctiefactor (of *g-weight*) $g = N / \sum d$ voor het post-stratum.
- c. Bepaal voor elke respondent in het post-stratum het individueel gekalibreerde gewicht $w = d \times g$.

Ten slotte merken we op dat dit model voor elk kwartaal apart wordt toegepast. Dit is o.m. nodig omdat zo snel mogelijk na de beëindiging van de gegevensverzameling voor het kwartaal, belangrijke LFS-schattingen over thema's zoals tewerkstelling moeten gepubliceerd worden. Dat impliceert ook dat we telkens een nieuwe extractie uit het Rijksregister gebruiken. Voor elk kwartaal is die extractie een "foto" van de populatie op een tijdstip in het midden van het kwartaal. Die foto wordt pas genomen nadat de LFS-gegevens voor het betreffende kwartaal zijn verzameld en opgekuist, m.a.w. ongeveer een maand na afloop van het kwartaal. Op dat moment zijn ook

de gegevens in het Rijksregister over dat tijdstip in het verleden voldoende geüpdatet om een voldoende correcte foto van de populatie te bekomen.

Met het oog op andere kalibratiemodellen (zowel voor LFS als voor HBS), modellen die geen “eenvoudige” post-stratificatiemodellen meer zullen zijn, noteren we het klassieke model als volgt:

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

Het is belangrijk op te merken dat als een post-stratum in de populatie niet leeg is (d.w.z. $N > 0$), het in de respondentensteekproef ook niet leeg mag zijn (d.w.z. $\sum d > 0$). Voor LFS is dat nooit een probleem geweest, omdat de respondentensteekproef steeds groot genoeg blijkt te zijn (de initiële steekproef is groot, en de respons is voldoende hoog door het verplichte karakter van de enquête).

Met het oog op kleinere steekproeven en andere modellen merken we hier reeds op dat van het model $\{\text{Ind; STRAT12} \times \text{SEX} \times \text{AGE; -}\}$ vlot zou kunnen overgestapt worden op bijvoorbeeld (en louter illustratief) het model

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

Hiermee bedoelen we het volgende:

- “Ind” wijst erop dat we nog steeds een kalibratie op zuiver individueel niveau toepassen.
- “STRAT12 \times SEX \times AGE” in het post-stratificatiemodel betekent dat we de steekproef kalibreren op de gezamenlijke populatieverdeling (*joint population distribution*) van de variabelen STRAT12, SEX en AGE. “STRAT12 \times (SEX + AGE)” betekent dat we zouden kalibreren op de marginale verdelingen van SEX en AGE in elk stratum STRAT12, of m.a.w. simultaan op de gezamenlijke verdeling van STRAT12 en SEX en de gezamenlijke verdeling van STRAT12 en AGE (want STRAT12 \times (SEX + AGE) = STRAT12 \times SEX + STRAT12 \times AGE).
- “Lin” wijst op een bepaalde keuze die men kan en moet maken in geval van modellen die geen post-stratificatiemodellen zijn. Met “Lin” bedoelen we dat het gekozen model voor de correctiefactoren g lineair is in de kalibratievariabelen⁴. Andere keuzen zijn mogelijk: “Exp” voor een exponentieel (of multiplicatief) verband, “Logit” voor een logistisch verband of “TrLin” voor een afgeknot (of getrunceerd) lineair verband. De keuze Exp wordt soms gemaakt om positieve correctiefactoren te garanderen, de keuzen Logit en TrLin zijn voorzien om de correctiefactoren zowel naar boven als naar beneden (bv. op 0) te begrenzen. De keuzen Lin, Exp, Logit en TrLin worden ook eenduidig vertaald naar bepaalde afstandsmaten tussen initiële (steekproef-)gewichten d en gekalibreerde gewichten w . Voor post-stratificatiemodellen speelt deze keuze helemaal geen rol! M.a.w. elke keuze leidt, in geval van een post-stratificatiemodel, tot dezelfde correctiefactoren g .

Dit alternatieve model kunnen we ook als volgt formuleren:

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

⁴ Net zoals een regressiemodel lineair kan zijn in de verklarende variabelen.

waarmee we aangeven dat binnen ieder stratum STRAT12 de kalibratie gebeurt naar de marginale populatieverdeling van SEX en de marginale populatieverdeling van AGE. Deze notatie kunnen we dan ook gebruiken voor het klassieke post-stratificatiemodel:

{Ind; STRAT12; SEX × AGE; Lin}

Deze notatie zal het gemakkelijker maken over te stappen naar modellen voor simultane kalibratie op individueel en huishoudniveau (paragraaf 5.1.3)! Het “afzonderen” van de kalibratievariabele STRAT12 geeft deze een speciale rol: de berekening van de correctiefactoren kan dan in elke categorie van STRAT12 apart gebeuren. Een variabele die we op dergelijke manier kunnen afzonderen, noemen we *kalibratiestratificatievariabele*; de overeenkomstige categorieën van de variabele noemen we de *kalibratiestrata*. Merk op dat de kalibratiestrata niet noodzakelijk de steekproefstrata (als die al bestaan) moeten zijn.

Belangrijke eigenschappen van het post-stratificatiemodel, en ook bv. van het daaruit afgeleide en hierboven bondig besproken alternatieve model, zijn:

- I. Individuen in eenzelfde huishouden zullen in het algemeen verschillende correctiefactoren g hebben als hun kalibratievariabelen verschillende waarden aannemen. Omdat het steekproefplan voor LFS impliceert dat het steekproefgewicht d voor elk individu van eenzelfde huishouden dezelfde waarde heeft, zullen dus in het algemeen de gekalibreerde gewichten w van de individuen in eenzelfde huishouden onderling verschillen.
- II. Hieruit volgt dat er geen eenduidig bepaald gekalibreerd huishoudgewicht is. En daarom neemt men wel eens zijn toevlucht tot *ad hoc* oplossingen wat dit aspect betreft. Dit speelt ook mee in het niet-coherent zijn van LFS-schattingen op huishoudniveau, als LFS op individueel niveau is gekalibreerd.
- III. Alle individuen met dezelfde waarden voor alle kalibratievariabelen, ongeacht tot welk huishouden ze behoren, hebben dezelfde correctiefactor g . Maar dat betekent niet dat ze ook dezelfde w zullen hebben, omdat ze een verschillende d kunnen hebben en omdat $w = d \times g$.

Eigenschap II is, zoals reeds eerder in paragraaf 5.1.1 werd aangehaald, de voornaamste reden tot herziening van het kalibratiemodel voor LFS.

5.1.3. Vernieuwde kalibratie, simultaan op individueel en op huishoudniveau

De nieuwe kalibratie van LFS, die we met het oog op de kalibratie van HBS hebben uitgevoerd, maakt gebruik van volgende LFS-gegevens:

1. Nog steeds gebruiken we een lijst van alle individuele respondenten, met de kenmerken leeftijd en geslacht die in het kalibratiemodel zullen geïntegreerd worden.
2. Daarnaast gebruiken we ook een lijst van alle responderende huishoudens, met de kenmerken woonplaats en aantal gezinsleden. Merk op dat deze tweede lijst uit de eerste

kan afgeleid worden (gegeven dat alle leden van een responderend huishouden ook individuele respondenten zijn).

3. Voor elk responderend huishouden: het steekproefgewicht d (berekend uit steekproefplan en -kader voor LFS). Merk op dat het steekproefplan voor LFS zo is dat het steekproefgewicht voor elk individu (\bar{d} , zeg) in een bepaald huishouden gelijk is aan het steekproefgewicht voor het huishouden; dus: $\bar{d} = d$ binnen elk huishouden. (Dat is niet essentieel in kalibratietheorie, maar maakt de bespreking hier wel eenvoudiger.)

Net zoals in paragraaf 5.1.2, leiden de individuele variabelen leeftijd en geslacht tot de volgende twee *individuele* kalibratievariabelen:

- SEX: geslacht (2 categorieën);
- AGE: 16 leeftijdsklassen (0-4, 5-9, 10-14, ..., 70-74, 75+);

en bovendien geven de huishoudvariabelen woonplaats en huishoudgrootte aanleiding tot volgende twee *huishoud*kalibratievariabelen:

- STRAT12: de 12 steekproefstrata voor selectie van de PSE's (zie tabel 1.A in paragraaf 3.2.2);
- HHSIZE: 6 grootteklassen (1-, 2-, 3-, 4-, 5- en 6- of meerpersoonshuishoudens).

Overeenkomstig deze kalibratievariabelen kunnen we uit het Rijksregister volgende cijfers berekenen (tellen!):

- Per stratum (STRAT12), het aantal individuen in elke subpopulatie van individuen bepaald door de combinatie van SEX en AGE. M.a.w. we zullen nog steeds (dus net zoals in het klassieke post-stratificatiemodel voor individuen) de *aantallen individuen* in elke cel van de kruising STRAT12 × SEX × AGE gebruiken als referentiewaarden in de vernieuwde kalibratie.
- Per stratum (STRAT12), het aantal huishoudens in elke subpopulatie van huishoudens bepaald door de categorieën van HHSIZE. M.a.w. we zullen nu bovendien de *aantallen huishoudens* in elke cel van de kruising STRAT12 × HHSIZE gebruiken als referentiewaarden in de vernieuwde kalibratie.

We kunnen nog steeds N noteren voor elk van de 384 individuele populatiecijfers, en we noteren nu bovendien M voor elk van de $12 \times 6 = 72$ huishoudpopulatiecijfers. Maar de manier om de correctiefactoren g te berekenen is nu niet meer zo eenvoudig te beschrijven als voor het klassieke individuele post-stratificatiemodel (paragraaf 5.1.2). We beperken ons daarom in deze tekst tot een beschrijving en interpretatie van het vernieuwde model.

Het vernieuwde kalibratiemodel voor LFS 2016 kan als volgt worden geformuleerd:

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

Hiermee bedoelen we het volgende:

- “HH” wijst erop dat we nu *formeel* een kalibratie op huishoudniveau uitvoeren, waarmee we bedoelen dat nu de steekproefgewichten voor de huishoudens worden gecorrigeerd met een factor g die berekend wordt voor elk huishouden.
- STRAT12 is de kalibratiestratificatievariabele, wat formeel wil zeggen dat de kalibratie onafhankelijk in de kalibratiestrata STRAT12 kan uitgevoerd worden.
- “SEX × AGE” vóór het teken “//” betekent dat de respondentensteekproef van *individuen* gekalibreerd wordt op de (individuele) gezamenlijke populatieverdeling (*joint population distribution*) van de variabelen SEX en AGE. Dit gebeurt dus binnen elk kalibratiestratum STRAT12, en dat impliceert dat we eigenlijk nog steeds de individuele respondenten kalibreren naar de gezamenlijke populatieverdeling van STRAT12, SEX en AGE, net zoals in het klassieke post-stratificatiemodel.
- “HHSIZE” na het teken “//” betekent dat de respondentensteekproef van *huishoudens* gekalibreerd wordt op de huishoudpopulatieverdeling van de variabele HHSIZE. Dit gebeurt eveneens binnen elk kalibratiestratum STRAT12, en dat impliceert dat we eigenlijk de responderende huishoudens kalibreren naar de gezamenlijke huishoudpopulatieverdeling van STRAT12 en HHSIZE.
- “Logit” betekent dat we voor een logistisch verband tussen de correctiefactoren g en de kalibratievariabelen hebben gekozen. Twee parameters, nl. de ondergrens en de bovengrens voor de g , werden door *trial and error* vastgelegd op 0 en 10, respectievelijk. De keuze “Lin” leverde negatieve waarden voor sommige g op, wat dus moest verholpen worden door keuze van een van de 3 alternatieve verbanden.

We hebben hierboven gesteld dat we nu enerzijds *formeel* kalibreren op huishoudniveau, maar anderzijds zowel naar huishoudpopulatiecijfers M als naar individuele populatiecijfers N . Hoe zijn deze beweringen verenigbaar? Dat is een mathematische kwestie, die we in deze nota niet in detail uit de doeken kunnen doen. Om toch enigszins aan te geven wat we precies bedoelen, merken we op:

- dat met elke referentiewaarde M een wiskundige vergelijking overeenkomt die uitdrukt dat de som van zekere gekalibreerde huishoudgewichten $w = d \times g$ gelijk is aan M ; bv. $\sum wz = M$ of $\sum dgz = M$, waarin de sommatie over huishoudens gaat en waarin z een indicator (0-1 variabele) is die aangeeft welke huishoudens effectief meetellen in de som;
- dat met elke referentiewaarde N een wiskundige vergelijking overeenkomt die uitdrukt dat de som van zekere individuele gekalibreerde gewichten $\bar{w} = \bar{d} \times g$ gelijk is aan N ; bv. $\sum \bar{w}x = N$ of $\sum \bar{d}gx = N$, waarin de sommatie over individuen gaat en waarin x een indicator (0-1 variabele) is die aangeeft welke individuen effectief meetellen in de som;
- dat “formeel kalibreren op huishoudniveau” impliceert dat deze laatste vergelijkingen worden getransformeerd door hun linker lid als een sommatie over huishoudens te herschrijven: $\sum(\sum \bar{d}x)g = N$, waarin de buitenste sommatie over alle huishoudens gaat, en elke binnenste som over de individuen binnen een huishouden;

- dat we deze laatste vergelijking nog kunnen schrijven als $\sum dg(\sum x) = N$ of $\sum w(\sum x) = N$, omdat voor LFS alle individuen binnen hetzelfde huishouden hetzelfde steekproefgewicht \bar{d} hebben, dat ook gelijk is aan het huishoudsteekproefgewicht d .

Het vernieuwde kalibratiemodel heeft dus de kalibratievergelijkingen $\sum dgz = M$ en $\sum dg(\sum x) = N$, die formeel van hetzelfde type zijn: de linkerleden zijn telkens gewogen sommen (met gewichten $dg = w$) over huishoudens. De onbekenden in deze vergelijkingen zijn de correctiefactoren g , waarvan er slechts één per huishouden dient berekend te worden.

Kalibratietheorie leert dat de correctiefactoren g functies zijn

- van de huishoudkalibratievariabele HHSIZE,
- en van *aggregaties* van de individuele kalibratievariabelen SEX en AGE.

Concreet betekent dit dat de correctiefactor g voor een bepaald huishouden afhangt

- van de HHSIZE-categorie waartoe het huishouden behoort,
- en van de samenstelling van het huishouden, meer bepaald van het aantal huishoudleden in elke (SEX x AGE)-categorie.

Vermits echter de g voor een bepaald huishouden ook de correctiefactor is voor elk individu in dat huishouden, kunnen we deze interpretatie ook formuleren als volgt: de correctiefactor g voor een bepaald individu hangt af

- van de grootte van het huishouden waartoe het individu behoort,
- van het *aantal* huisgenoten van hetzelfde geslacht en in dezelfde leeftijdscategorie,
- en van het *aantal* huisgenoten in andere geslachts- en leeftijdscategorieën.

Als individuen in verschillende huishoudens hetzelfde geslacht en dezelfde leeftijd hebben, dan hebben die individuen dezelfde correctiefactor g in het klassieke post-stratificatiemodel, maar ze kunnen verschillende correctiefactoren krijgen in het vernieuwde model als de huishoudens waartoe ze behoren verschillende grootte en/of samenstelling hebben!

Ten slotte merken we op dat, ten behoeve van de kalibratie van HBS 2016, het vernieuwde kalibratiemodel voor LFS 2016 slechts één keer werd toegepast, d.w.z. op de gehele LFS-steekproef van respondenten, en dus niet per kwartaal. De referentiewaarden N en M werden berekend op basis van een extractie uit het Rijksregister d.d. 1/07/2016.

5.2. [Kalibratie van HBS 2016](#)

De kalibratie van HBS 2016 ten slotte, maakt gebruik van volgende HBS-gegevens:

1. We gebruiken een lijst van alle responderende huishoudens, met de kenmerken woonplaats, professioneel statuut van de referentiepersoon (RP), leeftijd van de RP, aantal gezinsleden en aantal actieve gezinsleden.

2. Ook het kwartaal waarin het huishouden meewerkte aan de enquête zal omgezet worden naar kalibratievariabelen; dit is nieuw t.o.v. het kalibratiemodel dat voor HBS 2014 werd gebruikt.
3. Nieuw t.o.v. het model voor HBS 2014 is ook het gebruik van het totale (netto belastbare) inkomen voor de huishoudens in de kalibratie.
4. Voor elk responderend huishouden: het *steekproefgewicht*.

Het model voor HBS 2016 wijkt dus substantieel af van het model voor HBS 2014. Een en ander heeft ook tot gevolg dat de benchmarks niet alleen uit de gewogen LFS 2016 kunnen berekend worden: de benchmarks in verband met inkomen zullen uit een koppeling van het Rijksregister met het fiscale register worden bekomen. Bovendien zullen we voor bepaalde benchmarks helemaal geen referentiesteekproef of -register nodig hebben! De opbouw van het nieuwe kalibratiemodel wordt in volgende subparagrafen stap voor stap uit de doeken gedaan.

5.2.1. Het traditionele kalibratiemodel voor HBS

We kunnen stellen dat het centrale deel van het kalibratiemodel voor HBS 2016 nog steeds het klassieke of traditionele post-stratificatiemodel is, zoals werd toegepast voor (bijvoorbeeld) 2014, 2012, ... Dat model kan als volgt worden geformuleerd:

$$\{HH; GEWEST; PS; -\}$$

Hiermee bedoelen we het volgende:

- “HH” wijst erop dat we een kalibratie op huishoudniveau uitvoeren, waarmee we bedoelen dat de *voor non-respons gecorrigeerde* steekproefgewichten voor de huishoudens worden gecorrigeerd met een factor g die berekend wordt voor elk huishouden.
- GEWEST is de kalibratiestratificatievariabele, wat formeel wil zeggen dat de kalibratie onafhankelijk in de 3 gewesten kan uitgevoerd worden.
- “PS” verwijst naar 13 post-strata van huishoudens (binnen elk GEWEST) die bepaald worden aan de hand van de hierboven vermelde huishoudkenmerken:
 - PS = 1 : loontrekkende, -, 1P, -
 - PS = 2 : loontrekkende, -, 2P, <=1A
 - PS = 3 : loontrekkende, -, 2P, 2A
 - PS = 4 : loontrekkende, -, 3+P, <=1A
 - PS = 5 : loontrekkende, -, 3P, 2+A
 - PS = 6 : loontrekkende, -, 4+P, 2+A
 - PS = 7 : zelfstandige, -, -, -
 - PS = 8 : niet-actieve, <=59, 1P, -
 - PS = 9 : niet-actieve, <=59, 2+P, -

- PS = 10 : niet-actieve, 60-69, 1P, -
- PS = 11 : niet-actieve, 60-69, 2+P, -
- PS = 12 : niet-actieve, 70+, 1P, -
- PS = 13 : niet-actieve, 70+, 2+P, -

waarbij we refereren naar, achtereenvolgens: het socio-professioneel statuut van de RP, de leeftijd van de RP, het aantal gezinsleden en het aantal “actieven” in het huishouden.

- “-” betekent dat we geen specifieke keuze voor het verband tussen de correctiefactoren en de kalibratievariabelen hoeven te maken. Het model is immers een post-stratificatiemodel voor de HBS-huishoudens.

Met dit model kalibreren we bijgevolg naar de gezamenlijke huishoudpopulatieverdeling van REGIO en PS. Die verdeling kunnen we niet uit het Rijksregister berekenen, omdat sommige onderliggende huishoudkenmerken niet in het Rijksregister zijn opgenomen. We kunnen die echter wel uit de gekalibreerde LFS berekenen.

Door die kalibratie van HBS naar LFS-schattingen is er op huishoudniveau een zekere coherentie tussen LFS- en HBS-schattingen. Maar omdat het vernieuwde LFS-kalibratiemodel ook zekere coherentie op huishoudniveau met het Rijksregister impliceert, zullen zekere HBS-schattingen eveneens met zekere Rijksregister-gegevens coherent zijn. Meer bepaald kunnen we besluiten dat, dankzij de vernieuwing van de kalibratie van LFS 2016, de geschatte aantallen huishoudens in LFS 2016 en in HBS 2016 gelijk zijn aan elkaar en aan Rijksregistergegevens, voor heel België en voor elk gewest. Ondanks het feit dat in zowel het vernieuwde kalibratiemodel voor LFS als in het kalibratiemodel voor HBS ook huishoudgrootte in de kalibratievariabelen is geïntegreerd, kunnen we geen dergelijke exacte gelijkheden per huishoudgrootteklasse verwachten (bijvoorbeeld omdat binnen de groep van huishoudens met zelfstandige RP geen onderscheid gemaakt wordt in huishoudgrootteklassen), maar toch zullen we ook hier een soort “bijna-coherentie” mogen verwachten.

In tabel 11 presenteren we de kalibratietotalen naar REGIO en PS, zoals berekend als schattingen uit LFS 2016, na kalibratie van LFS zoals beschreven in paragraaf 5.1.3.

Tabel 11 Kalibratietotalen (benchmarks) voor HBS 2016, geschat op basis van LFS 2016

Post-stratum (binnen Regio)	Regio						Totaal	
	Brussels H.G.		Vlaanderen		Wallonië			
	Gekalibreerd gewicht		Gekalibreerd gewicht		Gekalibreerd gewicht		Gekalibreerd gewicht	
	Som	%	Som	%	Som	%	Som	%
Loontrekkende, 1p	107.626,4	19,8	339.586,5	12,3	178.245,1	11,5	625.457,9	12,9
Loontrekkende, 2p, <=1act	33.088,4	6,1	181.976,4	6,6	114.783,7	7,4	329.848,5	6,8
Loontrekkende, 2p, 2act	26.682,6	4,9	201.880,1	7,3	73.671,9	4,7	302.234,6	6,2
Loontrekkende, 3+p, <=1act	55.425,4	10,2	195.452,6	7,1	137.838,3	8,9	388.716,3	8,0
Loontrekkende, 3p, 2+act	18.864,7	3,5	202.192,2	7,3	91.414,2	5,9	312.471,1	6,4
Loontrekkende, 4+p, 2+act	31.441,5	5,8	336.860,4	12,2	160.233,1	10,3	528.535,0	10,9
Zelfstandige	60.039,6	11,1	275.424,1	10,0	134.961,5	8,7	470.425,1	9,7
Niet-actieve, <=59, 1p	62.391,0	11,5	131.307,9	4,8	128.919,0	8,3	322.618,0	6,7
Niet-actieve, <=59, 2+p	45.566,2	8,4	81.642,7	3,0	104.281,3	6,7	231.490,2	4,8
Niet-actieve, 60-69, 1p	31.448,9	5,8	160.851,5	5,8	114.383,8	7,4	306.684,2	6,3
Niet-actieve, 60-69, 2+p	13.420,7	2,5	175.444,3	6,4	92.441,2	5,9	281.306,2	5,8
Niet-actieve, 70+, 1p	28.714,0	5,3	177.706,7	6,5	99.524,4	6,4	305.945,1	6,3
Niet-actieve, 70+, 2+p	27.816,6	5,1	293.346,7	10,7	125.590,5	8,1	446.753,9	9,2
Totaal	542.526,0	100,0	2.753.672,0	100,0	1.556.288,0	100,0	4.852.486,0	100,0

5.2.2. Uitbreiding om evenwicht tussen de kwartalen te realiseren

In paragraaf 4.2 hebben we reeds vermeld dat de 4 kwartalen van 2016 niet gelijk zijn vertegenwoordigd in de totale HBS 2016-steekproef van responderende huishoudens. De eerste twee kwartalen zijn duidelijk ondervertegenwoordigd, en dit zowel zonder weging (eerste lijn in tabel 10), als na weging met de steekproefgewichten (tweede lijn in tabel 10). Na een kalibratie met het klassieke model uit vorige paragraaf 5.2.1 zou dit probleem niet weggewerkt zijn. Immers, het model {HH; GEWEST; PS; -} impliceert dat alle responderende huishoudens met dezelfde waarden voor de kalibratievariabelen GEWEST en PS, dezelfde correctiefactor g zouden krijgen, zodat de steekproefgewichten voor al deze huishoudens, ongeacht het kwartaal waarin ze deelnamen, op precies dezelfde manier zouden worden aangepast (in de praktijk: opgehoogd). Na kalibratie met model {HH; GEWEST; PS; -} zou het onevenwicht tussen de kwartalen blijven bestaan, met bijgevolg nog steeds een mogelijke versterking of afzwakking van seizoenseffecten in schattingen voor het hele jaar 2016.

Het doel van de uitbreiding van het klassieke model uit vorige paragraaf is het bekomen van gekalibreerde gewichten w die voldoen aan de volgende vergelijkingen:

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

waarin $\sum_{Rt} w$ de som is van de gekalibreerde gewichten voor de huishoudens die in het t -de kwartaal hebben deelgenomen aan HBS. De eerste van deze vergelijkingen, $\sum_{R1} w = \sum_{R2} w$, kan als volgt naar een gebruikelijke kalibratievergelijking worden omgezet. We definiëren de variabelen z_1 en z_2 :

$$\begin{aligned}
 z_1 &= 1 \text{ voor elk huishouden dat in T1 aan HBS deelneemt} \\
 &= 0 \text{ voor elk huishouden dat in T2, T3 of T4 aan HBS deelneemt} \\
 z_2 &= 1 \text{ voor elk huishouden dat in T2 aan HBS deelneemt} \\
 &= 0 \text{ voor elk huishouden dat in T1, T3 of T4 aan HBS deelneemt}
 \end{aligned}$$

Dan kan $\sum_{R1} w = \sum_{R2} w$ herschreven worden als:

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

waarin elk van de sommaties over *alle* responderende huishoudens gaat. De volgende stap bestaat erin een zgn. *contrastvariabele* z_{12} te definiëren:

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

Dan kan voorgaande vergelijking herschreven worden als:

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

Deze vergelijking heeft de gebruikelijke vorm van een kalibratievergelijking met als kalibratievariabele de *contrastvariabele* c_{12} en benchmark of kalibratietotaal 0. Voor berekening van dergelijke benchmarks hebben we natuurlijk geen referentiesteekproef of -register nodig!

Met deze vergelijking hebben we kwartalen T1 en T2 gecontrasteerd. Op analoge manier kunnen we kwartalen T1 en T3, en T1 en T4 contrasteren, d.w.z. door invoering van overeenkomstige *contrastvariabelen* c_{13} en c_{14} . De oorspronkelijke vergelijkingen $\sum_{R1} w = \sum_{R2} w = \sum_{R3} w = \sum_{R4} w$ kunnen bijgevolg geherformuleerd worden als een stelsel van drie vergelijkingen $\sum w \cdot c_{12} = 0$, $\sum w \cdot c_{13} = 0$ en $\sum w \cdot c_{14} = 0$.

Als we nu opmerken dat we het klassieke model {HH; GEWEST; PS; -} ook kunnen formuleren als {HH; GEWEST × PS; -}, dan kan gemakkelijk ingezien worden dat een uitgebreid model kan geformuleerd worden als:

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

Hierin verwijst “TrLin” naar de getrunceerd lineaire methode (zie paragraaf 5.1.2 voor deze en andere keuzen van de “methode”; en we hebben uiteindelijk de onder- en bovengrens voor de correctiefactoren (*g*-weights) vastgelegd op 0,10 en 100, respectievelijk).

Dit is niet hetzelfde model als:

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

waarin het evenwicht tussen de kwartalen niet alleen globaal, maar ook per gewest wordt beoogd. Merk op dat model {HH; GEWEST × (PS + C12 + C13 + C14); TrLin} restrictiever is dan model

{HH; GEWEST × PS + C12 + C13 + C14; TrLin}, en dat de contrastbeperkingen op niveau van de gewesten de contrastbeperkingen op globaal niveau impliceren.

Een model dat nog restrictiever is dan {HH; GEWEST × (PS + C12 + C13 + C14); TrLin}, en dus ook restrictiever dan {HH; GEWEST × PS + C12 + C13 + C14; TrLin}, is het model

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

waarin het evenwicht tussen de kwartalen niet alleen globaal en per gewest, maar ook per provincie wordt beoogd.

Na evaluatie van deze modellen, die alle het herstellen van het evenwicht tussen de kwartalen beogen, werd gekozen voor dit laatste model, omdat dit het meest de correcte weergave benadert.

5.2.3. *Uitbreiding om een vertekening te reduceren*

Een diepgaandere evaluatie van het model {HH; GEWEST × PS + PROV × (C12 + C13 + C14); TrLin} bleek toch niet helemaal te voldoen aan (mogelijke) verwachtingen van de gebruikers. Immers, dat model leverde een frappante overschatting op van het percentage eigenaars van woningen (op niveau van huishoudens): tabel 12 vergelijkt de schattingen van de relatieve verdeling van de variabele eigendomsstatuut voor HBS 2012, 2014 en 2016 – wanneer gebruikgemaakt wordt van het klassieke model voor 2012 en 2014 en van het uitgebreidere model met evenwichten tussen kwartalen voor 2016 – met schattingen op basis van SILC 2014-2016.

Tabel 12 Schattingen van eigendomsstatuut uit HBS en uit SILC

	HBS 2012	HBS 2014	HBS 2016	SILC 2014	SILC 2015	SILC 2016
	Klassiek model	Klassiek model	Klassiek model, met evenwicht tussen kwartalen	Gebruikelijke kalibratie		
Eigenaar	68,2%	70,0%	72,4%	66,1%	66,2%	66,7%
Huurder	30,6%	29,0%	26,4%	32,2%	32,3%	31,9%
Onbekend	1,2%	1,0%	1,2%	1,7%	1,5%	1,4%

We poogden daarom als extra kalibratievariabele het eigendomsstatuut (eigenaar of huurder) te gebruiken; deze variabele wordt bevraagd in HBS. Voor deze variabele vonden we echter geen goede benchmarks. Als proxy (of predictor) voor eigendomsstatuut hebben we daarom het gezinsinkomen gebruikt: in hogere inkomensklassen vinden we procentueel meer eigenaars van woningen. Dat (netto belastbaar) gezinsinkomen kunnen we uit een koppeling van het Rijksregister met het fiscale register bekomen voor ieder responderend huishouden, en voor ieder huishouden in de populatie, zodat ook geschikte benchmarks kunnen bepaald worden.

We hebben weliswaar niet het gezinsinkomen als numerieke variabele in het kalibratiemodel gebruikt, maar zijn overgestapt op een klassenindeling van deze variabele, als volgt.

Vooreerst merken we op dat, volgens het fiscale register, sommige (private) huishoudens in de populatie geen (netto belastbaar) inkomen hebben. De eerste inkomensklasse krijgt daarom het label “missing” en de code 1. Vervolgens werden op basis van de (sub)populatie van huishoudens die wel een inkomen hebben, de 4 inkomenskwintielen K_1, K_2, K_3, K_4 bepaald. Die kwintielen bepalen 5 inkomensklassen, die we de respectieve codes 2, 3, 4, 5 en 6 geven. Het incalculeren van huishoudens zonder (belastbaar) inkomen, heeft voor gevolg dat, als $100m$ het percentage huishoudens in inkomensklasse 1 (d.i. “missing”) is, het percentage in elk van de klassen 2 t.e.m. 6, gelijk is aan $(100 - 100m)/5$; als $m = 0$, dan zal elk van de klassen 2 t.e.m. 6 precies 20% van de huishoudens bevatten. Op die manier bepalen we voor elk van de responderende huishoudens (door koppeling met het uitgebreide Rijksregister), de inkomensklasse (1, 2, ... of 6); die variabele noemen we INK_5. Deze nieuwe variabele kan vervolgens als categoriale kalibratievariabele in het laatste kalibratiemodel dat in vorige paragraaf werd geïntroduceerd worden opgenomen:

$$\{HH; GEWEST \times PS + PROV \times (C12 + C13 + C14) + INK_5; TrLin\}$$

Merk op dat de 6 benchmarks die bij de laatste term, INK_5, van dit model horen, kunnen berekend worden door het totaal aantal huishoudens in de populatie – dat geschat wordt op basis van de gekalibreerde LFS 2016, en dat (impliciet) een benchmark is die bij de term $GEWEST \times PS$ hoort – te vermenigvuldigen met de proporties m of $(1 - m)/5$.

Dit laatste model kalibreert dus o.a. naar de globale inkomensverdeling in de Belgische populatie van private huishoudens, waarbij die inkomensverdeling werd samengevat door het percentage huishoudens zonder inkomen, en de kwintielen voor de huishoudens met een inkomen. We kunnen dit model verfijnen door per gewest de verdeling van het gezinsinkomen op een gelijkaardige manier samen te vatten. Als $K_{j1}, K_{j2}, K_{j3}, K_{j4}$ de vier inkomenskwintielen zijn voor de populatie van huishoudens met inkomen in gewest j ($j = 1, 2, 3$), dan construeren we een categoriale variabele INK_5_G als volgt:

- INK_5_G =
- 1 voor huishoudens zonder inkomen
 - 2 voor huishoudens in gewest j ($j = 1, 2, 3$) en met inkomen ten hoogste K_{j1}
 - 3 voor huishoudens in gewest j ($j = 1, 2, 3$) en met inkomen hoger dan K_{j1} en ten hoogste K_{j2}
 - 4 voor huishoudens in gewest j ($j = 1, 2, 3$) en met inkomen hoger dan K_{j2} en ten hoogste K_{j3}
 - 5 voor huishoudens in gewest j ($j = 1, 2, 3$) en met inkomen hoger dan K_{j3} en ten hoogste K_{j4}
 - 6 voor huishoudens in gewest j ($j = 1, 2, 3$) en met inkomen hoger dan K_{j4}

Dit leidt dan tot volgend kalibratiemodel:

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

De term $GEWEST \times INK_5_G$ bepaalt $3 \times 6 = 18$ nieuwe kalibratievergelijkingen. De benchmarks (of kalibratietotalen) zijn de aantallen huishoudens in de populatie per gewest en per categorie van INK_5_G, die berekend worden uit de aantallen huishoudens per gewest (die impliciet ook reeds benchmarks zijn bij de term $GEWEST \times PS$) en de proporties m_j ($j = 1, 2, 3$) van huishoudens zonder inkomen in de drie gewesten.

Op analoge manier kunnen we een nog verfijnder kalibratiemodel construeren, waarin met de inkomensverdeling per provincie wordt rekening gehouden:

$$\{HH; GEWEST \times PS + PROV \times (C12 + C13 + C14) + PROV \times INK_5_P; TrLin\}$$

De categoriale kalibratievariabele INK_5_P wordt hierbij geconstrueerd op een manier die analoog is aan de manier waarop INK_5_G werd geconstrueerd; vooraf moet natuurlijk de inkomensverdeling per provincie p worden samengevat via de kengetallen $m_p, K_{p1}, K_{p2}, K_{p3}, K_{p4}$ ($p = 1, \dots, 11$).

Tabel 13 laat zien dat met het vernieuwde kalibratiemodel voor 2016, dat ter validering ook werd toegepast voor HBS 2014, de schattingen voor de verdeling van eigendomsstatuut beter aansluiten bij de schattingen op basis van SILC (tabel 12) en HBS 2012.

Tabel 13 Schattingen van eigendomsstatuut uit HBS, volgens verschillende kalibratiemodellen

	HBS 2012	HBS 2014	HBS 2016	HBS 2014	HBS 2016
	Klassiek model	Klassiek model	Klassiek model, met evenwicht tussen kwartalen	Vernieuwde kalibratie	
Eigenaar	68,2%	70,0%	72,4%	67,3%	68,6%
Huurder	30,6%	29,0%	26,4%	31,6%	30,1%
Onbekend	1,2%	1,0%	1,2%	1,1%	1,3%

6. Coderings-, controle- en imputatieprogramma's

Deelnemende huishoudens moeten een vragenlijst invullen met behulp van de enquêteur en een uitgavenboekje waarin zij al hun uitgaven gedurende één maand registreren. In het verleden was het huishoudbudgetonderzoek een papieren enquête. Zowel enquêteurs als huishoudens vulden vragenlijsten en boekjes alleen op papier in. Deze manier van werken vereist veel codeerwerk achteraf en biedt weinig mogelijkheden om kwaliteitscontroles uit te voeren tijdens het veldwerk. Bovendien waren de huishoudens steeds meer vragende partij om de boekjes elektronisch in te vullen.

6.1. Modernisering van de gegevensverzameling

Er werd beslist om CAPI (Computer Assisted Personal Interviewing) te gebruiken, waarbij de enquêteur de leden van het huishouden rechtstreeks interviewt met behulp van een tablet-pc en een aangepast coderingsprogramma (aangepaste routing bij elke situatie, geen codeerwerk meer, minder fouten...). Dit wordt op grote schaal gebruikt sinds 2016.

Voor het invullen van het uitgavenboekje hadden de huishoudens de keuze tussen een standaardversie op papier (PPSI, of Paper-and-pencil Self-administered Interviewing) en een websurvey-applicatie in Java (CAWI, of Computer Assisted Web Interviewing). In 2016 heeft 55% van de huishoudens gekozen voor de elektronische versie van het boekje (30% in 2014).

Vanaf het huishoudbudgetonderzoek 2012 werd de inhoud van de vragenlijsten en de huishoudboekjes vereenvoudigd, waardoor de last op de respondenten verminderde.

De aanwerving via LFS heeft als voordeel dat er al veel informatie over personen in het huishouden beschikbaar is. Die informatie dient niet langer in de vragenlijsten te worden opgevraagd. Er wordt enkel gevraagd om de reeds verkregen gegevens te bevestigen. De modernisering heeft gezorgd voor andere mogelijkheden om de kwaliteit van het veldwerk te verbeteren. Vanaf 2012 moesten de enquêteurs elk contact met de respondenten documenteren in een contactformulier (data van contactnames, contactmethodes, resultaat van elke contactpoging). Aangezien deze contactformulieren rechtstreeks op elektronische dragers worden ingevuld en regelmatig naar de server worden doorgestuurd, kan de voortgang van het veldwerk op de voet worden gevolgd.

De registratiemethode van de hoeveelheden werd in 2012 ook verbeterd in de boekjes, wat de levering van gegevens aan Eurostat en andere gebruikers zal vergemakkelijken.

De vragenlijst in CAPI is ook gebonden aan een bepaalde routing, dit wil zeggen dat vragen die enkel in een bepaalde situatie relevant zijn, ook enkel aan mensen in deze situatie worden gesteld. Bijvoorbeeld: enkel huurders krijgen de vraag naar hoeveel huur ze betalen; enkel indien er verwarmd wordt op mazout, wordt er een vraag gesteld naar de jaarlijkse uitgave aan mazout. Met de papieren vragenlijst in 2012 bestond deze mogelijkheid niet, waardoor er veel tijd verloren ging aan de cleaning van de resultaten achteraf.

6.2. Controles en imputatie

6.2.1. Opsporing van outliers (aberrante waarden)

Voor de in het boekje vermelde uitgaven werden er SAS-programma's voor de controle en de opsporing van outliers ontwikkeld. Er werden verschillende methodes getest. Voor voedsel en niet-alcoholische dranken kozen we uiteindelijk voor de zogenaamde 'resistant fences method' (RFM). In deze univariate methode wordt het tolerantie-interval bepaald aan de hand van de kwartielen: $[Q1 - \text{coeff.low} * \max(Q3 - Q1, |a \times Q2|), Q3 + \text{coeff.up} * \max(Q3 - Q1, |a \times Q2|)]$. De term $a \times Q2$ staat voor een minimale waarde van de afstand. De coëfficiënten (low en up) werden op 3 gezet, zoals gebruikelijk in de praktijk en 0,05 voor parameter a.

De gegevens hebben eerder een Box-Cox transformatie ondergaan (met een parameter λ die varieert in functie van de productcategorie) aangezien de symmetrie van de verdeling de kwaliteit van de detectie met de RFM-methode verbetert.

We hebben aberrante waarden geïdentificeerd aan de hand van een gedetailleerde COICOP_HBS_BE code en hebben de aberrante waarden vervangen door de mediaan van de gegeven uitgaven.

Impact op de totale uitgaven (ongewogen bedragen): 0,6% minder uitgaven voor voeding en niet-alcoholische dranken.

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

Gedurende een maand is het moeilijk om een idee te krijgen van de periodieke uitgaven en de grote (en dus zeldzame) uitgaven van een huishouden. Vragen over deze twee soorten uitgaven zijn daarom opgenomen in een aparte vragenlijst, de vragenlijst "huishouden". Het omvat ook vragen over de samenstelling van het huishouden, de huisvesting (comfort, leeftijd, type woning, type verwarming, ...), gegroepeerde uitgaven (gas + elektriciteit bijvoorbeeld), het verbruik van water, gas, elektriciteit, stookolie; verzekeringen, status, beroep en opleidingsniveau van de leden van het huishouden en hun inkomens.

Er werden dus verschillende coherentiecontroles uitgevoerd door het boekje en de vragenlijst met elkaar te vergelijken.

In paragraaf 6.2.2 wordt een voorbeeld gegeven van hoe deze talrijke informatie kan worden gebruikt om de energiekosten te corrigeren en te imputeren. Andere periodieke uitgaven zoals huur en lasten, alsmede uitgaven voor duurzame goederen, zijn grondig gecontroleerd en specifiek verwerkt.

Belangrijke opmerking:

Alle uitgaven worden teruggebracht tot één maand: de uitgaven van het boekje blijven zoals ze zijn, de uitgaven van de vragenlijst met terugwerkende kracht (laatste 4 maanden) met de duurzame goederen waaronder auto's worden gedeeld door 4 en de periodieke uitgaven worden gedeeld door de periodiciteit.

Voor de definitieve tabel worden al deze uitgaven op jaarbasis berekend, d.w.z. vermenigvuldigd met 12.

6.2.2. Imputatie van energie-uitgaven

De huishoudvragenlijst wordt volledig afgenomen in een CAPI-interview. Dit betekent dat de enquêteur tijdens het laatste bezoek aan het huishouden de vragen afleest van zijn computerscherm en de gegeven antwoorden meteen codeert. Hij volgt hierbij de volgorde van de vragen (routing) zoals op voorhand geprogrammeerd zodat alle relevante vragen zeker gesteld worden. Vóór 2014 werd er gewerkt met een huishoudvragenlijst op papier. Dit werd in de praktijk vaak ingevuld door het huishouden zelf, waarna de enquêteur dit enkel nakeek. Dit zorgde ervoor dat er veel gegevens ontbraken, vooral op de energie-uitgaven. Er waren ook veel inconsistenties in de antwoorden (bv. zeggen dat hun voornaamste verwarmingsbron op mazout draait, maar enkel een uitgave voor aardgas vermelden). Deze ontbrekende informatie en inconsistenties zorgden voor veel verbeterwerk achteraf. Door het invoeren van de CAPI-vragenlijsten is de kwaliteit van de gegevens sterk verbeterd, waardoor er minder werk was achteraf bij de controle van de gegevens.

<i>Uitgave</i>	<i>Aantal missings in 2016</i>	<i>Aantal missings in 2014</i>	<i>Aantal missings in 2012</i>
<i>Water</i>	76	92	666
<i>Elektriciteit</i>	41	75	1302
<i>Gas (aardgasleiding)</i>		27	1223

Er wordt ook de mogelijkheid geboden om gecombineerde facturen in te geven. In één factuur betaalt een gezin bijvoorbeeld de elektriciteits- en de gasrekening, of het elektriciteits- en het waterverbruik. Aan de gezinnen en enquêteurs werd wel nog steeds gevraagd om zoveel mogelijk op te splitsen. Het opsplitsen van een dergelijke combinatie van energietypes noemen we in het vervolg van dit document ook een imputatie.

De methode waarop de imputaties gebeurden blijft dezelfde als in 2014. Zelfs voor gezinnen die geen informatie opgaven in hun vragenlijst, zijn er informatiebronnen over hun energie-uitgaven beschikbaar. Zo zijn er de dagboeken die informatie geven over de uitgaven tijdens de referentiemaand. Uitgaven voor energie kunnen in deze maand hebben plaatsgevonden.

Er werd voor gekozen om eerst de informatie uit de dagboeken aan te wenden om de ontbrekende uitgaven in te schatten, daarna de gemengde uitgaven trachten op te splitsen en om de dan nog ontbrekende energie-uitgaven te imputeren a.d.h.v. een lineair regressiemodel.

Ronde 1 Verzamelen van info uit de dagboeken om te gebruiken in de jaarlijkse uitgaven

De eerste stap bestaat erin dat de informatie die in de dagboeken werd genoteerd tijdens de referentiemaand zelf wordt gebruikt als basis voor een imputatie van een ontbrekend jaarverbruik in de vragenlijst. We doen deze imputatie enkel voor water, gas en elektriciteit omdat deze uitgaven met een vaste regelmaat gebeuren. Door het inschatten van deze periodiciteit kunnen de uitgaven die in een bepaalde maand gebeurden, omgezet worden naar jaarbedragen. Deze inschatting werd gemaakt aan de hand van gemiddelde bedragen op basis van gegevens in de vragenlijsten.

De gegevens die geïmputeerd werden, kregen een 'flag' mee:

<i>Waterfactuur</i>	<i>FL_W</i>	<i>N=20</i>
<i>Elektriciteitsfactuur</i>	<i>FL_E</i>	<i>N=9</i>
<i>Gasfactuur</i>	<i>FL_G</i>	<i>N=6</i>

Ronde 2 Opsplitsen van de gegroepeerde facturen

Er zijn verschillende combinaties mogelijk van de energie-uitgaven. De meest voorkomende combinatie is het samen factureren van gas en elektriciteit.

<i>Omschrijving</i>	<i>Aantal</i>
<i>Water, gas en elektriciteit</i>	<i>N=38</i>
<i>Water en elektriciteit</i>	<i>N=28</i>
<i>Water en gas</i>	<i>N=38</i>
<i>Elektriciteit en gas</i>	<i>N=1420</i>

De combinaties "water en elektriciteit" en "water en gas" werden opgesplitst door een regressiemodel op te stellen dat een schatting maakt van het waterverbruik. Het regressiemodel was gebaseerd op de onafhankelijke variabelen gezinsgrootte (uitgedrukt in equivalente eenheden), regio, aanwezigheid van een regenwaterrecuperatiesysteem en een wasmachine. De aanwezigheid van een waterput maakte geen significant verschil op de schattingen en werd ook niet opgenomen in de regressie. Het verbruik van de andere energiebron (respectievelijk elektriciteit en gas) werd dan berekend door het verschil te maken van de gemeenschappelijke factuur en de schatting voor het waterverbruik.

De combinatie “elektriciteit en gas” werd opgesplitst door een regressiemodel op te stellen dat een schatting maakt van het elektriciteitsverbruik. Het model werd gebaseerd op de gezinsgrootte (in equivalente eenheden), de regio en de aanwezigheid van een wasmachine en fotovoltaïsche zonnepanelen. De aanwezigheid van een frigo/diepvries of het gebruik van elektriciteit als additionele verwarming maakte geen significant verschil op de schattingen en deze werden dan ook niet weerhouden voor het regressiemodel. Het verbruik voor gas werd berekend door het verschil te maken van de gemeenschappelijke factuur en het geschatte elektriciteitsverbruik.

De combinatie “water, gas en elektriciteit” werd opgesplitst door de twee schattingen, eerst die van water en dan die van elektriciteit uit te voeren. Het verbruik voor gas werd vervolgens berekend door het verschil te maken van de gemeenschappelijke factuur en de twee geschatte verbruiken voor water en elektriciteit.

Opmerking: Niet alle gemeenschappelijke facturen konden op deze manier worden gesplitst. Het bedrag voor het overblijvende verbruik nadat de schatting werd afgetrokken van de gemeenschappelijke factuur was soms te klein. In dit geval werd er geen opsplitsing gemaakt.

Ronde 3 Ontbrekende gegevens imputeren gebruikmakend van regressiemodellen

De nu nog ontbrekende gegevens werden aangevuld door middel van een schatting op basis van een regressiemodel.

1. WATER

Er wordt verondersteld dat elk huishouden een waterfactuur moet betalen. De ontbrekende bedragen worden in deze stap ingevuld op basis van hetzelfde regressiemodel als gebruikt bij de splitsing van de gemeenschappelijke factuur.

De ontbrekende gegevens werden geschat en kregen een ‘flag’:

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

2. ELEKTRICITEIT

Net als bij waterverbruik, veronderstellen we een elektriciteitsverbruik voor elk deelnemend huishouden. De regressiemethode is dezelfde als deze die werd gebruikt bij het splitsen van de gegroepeerde facturen.

De schattingen kregen een ‘flag’ mee:

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

3. GAS

Er wordt een bedrag voor gas geïmputeerd voor alle gezinnen die zeggen met gas te verwarmen en die geen jaarlijkse uitgave aan gas noteerden in de vragenlijst of het dagboek.

De relevante onafhankelijke variabelen zijn gezinsgrootte (uitgedrukt in equivalente eenheden), regio en of de hoofdverwarming op gas werkt. De geïmputeerde bedragen kregen een 'flag' mee:

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

4. STOOKOLIE

De bedragen voor stookolie werden ook geschat op basis van een lineair regressiemodel. Dit voor alle gezinnen die aangaven dat ze stookolie gebruikten om te verwarmen, maar die geen jaarlijkse uitgave voor stookolie noteerden.

De gebruikte onafhankelijke variabelen zijn gezinsgrootte (uitgedrukt in equivalente eenheden), inwoner van het Brusselse Gewest en of de hoofdverwarming op mazout werkte.

De geïmputeerde bedragen kregen een 'flag':

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

5. BUTAAN

Voor butaan ontbraken er 11 gegevens. Voor de imputatie kon echter geen regressiemodel opgesteld worden wegens het gebrek aan volledige gegevens. Er werd dan geïmputeerd op basis van de gemiddelde jaaruitgaven aan butaan voor deze gezinnen die een bedrag hadden opgegeven.

De gezinnen waar een imputatie werd uitgevoerd, kregen een 'flag' mee:

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

6. PROPAANGAS

In 2016 bleek een probleem bij de meting van de uitgaven voor propaangas. Door een fout in de routing van de vragenlijst, werd er bij huishoudens die verwarmden op propaan, niet gevraagd naar het jaarlijkse bedrag voor de aankoop van propaan. De betrokken huishoudens werden nadien gecontacteerd om de ontbrekende informatie op te vragen. Het lukte echter niet om iedereen te contacteren. Die bedragen moesten dan geïmputeerd worden. Er kon echter geen passend regressiemodel gevonden worden. De imputatie gebeurde dan ook o.b.v. het gemiddelde.

De geschatte waarden voor propaan kregen een 'flag' mee:

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

7. STEENKOOL

De imputatie voor steenkool bij die gezinnen die steenkool gebruiken om te verwarmen, kon niet gebeuren op basis van een regressiemodel. Hiervoor waren er immers te weinig volledige cases aanwezig in de dataset. Er werd daarom geïmputeerd op basis van een gemiddelde uitgave aan steenkool voor deze gezinnen die een uitgave voor steenkool hadden geregistreerd.

De schattingen kregen een 'flag':

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

8. PELLETS

Het verbruik aan pellets kon niet significant voorspeld worden door een regressiemodel. De imputatie gebeurde daarom op basis van het gemiddelde.

De schattingen kregen een bijhorende 'flag':

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

9. PETROLEUM

Geen ontbrekende gegevens, dus geen imputatie nodig.

Opmerking: De ontbrekende uitgaven in de categorie HOUT (COICOP=04549A) werden niet geïmputeerd. Brandhout kan immers gratis worden verkregen, of er kan in het verleden een voorraad aangelegd geweest zijn voor meerdere jaren.

6.2.3. Inkomens

Tot 2010 werden de inkomens net als de uitgaven op een exhaustieve manier ingevuld in het boekje. Sinds 2012 worden de inkomens op een meer globale manier verzameld in een vragenlijst om de werklast van de respondenten te beperken. Deze werkwijze biedt de mogelijkheid om de uitgaven van huishoudens in functie van hun levensstandaard te vergelijken (bv. inkomenskwartielen), maar gedetailleerde tabellen over de samenstelling van de inkomens zullen niet meer gebaseerd zijn op HBS.

Voor alle studies over de inkomens bevelen wij het gebruik van de EU-SILC-enquête aan (communautaire enquête naar de inkomens en levensomstandigheden), die ook door Statbel wordt georganiseerd.

De inkomens werden grondig gecontroleerd (opsporing van outliers, controle van huishoudens die de informatie niet correct hadden ingevuld: lege vragenlijst, inkomen te laag of niet in overeenstemming met de ingevulde informatie over de socio-professionele status, activiteitssector, ...)

In 2014 werden bepaalde vragen toegevoegd aan de vragenlijst over de terugbetalingsinkomsten die ontvangen werden op het niveau van het huishouden.

Validatie van de inkomens

Weighted Basic Statistical Measures (Location & Variability)	EBM 2016	EU-SILC 2016
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

Verdeling van de inkomsten

Income class	Income EU-SILC 2016		Income EBM 2016	
	Percent	Cumulative Percent	Percent	Cumulative Percent
0-250	0.57	0.57	0.27	0.27
250-500	0.46	1.03	0.20	0.46
500-1000	4.79	5.81	3.39	3.85
1000-1500	16.75	22.56	15.3	19.15
1500-2000	14.00	36.56	16.37	35.53
2000-2500	11.16	47.72	11.98	47.50
2500-3000	9.48	57.21	10.51	58.01
3000-5000	25.66	82.86	29.70	87.71
5000+	17.14	100.00	12.29	100.00

6.2.4. Imputed rent

De 'imputed rent' is een fictieve huurprijs toegewezen aan de eigenaars en aan de huishoudens die een woning gratis bewonen.

Om de prijs te schatten waartegen hun woning gehuurd zou kunnen worden en om zo dus een idee te krijgen van het financiële voordeel dat verbonden is aan hun eigendomsstatus, gebruiken we de door Eurostat aanbevolen methode: een Heckman-model in twee fasen. We gebruiken eerst een probit-regressie om de waarschijnlijkheid om huurder te zijn in te schatten op basis van de kenmerken van het huishouden en zijn leden (waaronder leeftijd, opleidingsniveau, aantal beroepsactiviteiten, ...). Vervolgens schatten we door middel van regressie de huurprijs met vele kenmerken van de woning die in de enquête aanwezig is, zoals het type en de leeftijd van de woning, het aantal kamers, de aanwezigheid van een tuin, een garage, een balkon, centrale verwarming, en natuurlijk rekening houdend met het resultaat van de eerste regressie.

Validatie van de werkelijke en geïmputeerde huur ten opzichte van de EU-SILC 2016-enquête

	SILC 2016	HBS 2016
Maandelijkse huur per huurdershuishouden (€)	544	533
Imputed rent maandelijks per eigenaarshuishouden (€)	637	676

Bijlagen

1. Enkele definities

Een **huishouden** bestaat ofwel uit een alleenstaande, veelal alleen levende persoon, ofwel uit twee of meer personen die, al dan niet door verwantschap aan elkaar verbonden, gewoonlijk in één en dezelfde woning verblijven en er gezamenlijk leven.

De **referentiepersoon volgens HBS** is de persoon die zorgt voor het grootste deel van het inkomen van het huishouden. Deze definitie van **referentiepersoon** (gezinshoofd) komt niet noodzakelijk overeen met deze **volgens het Rijksregister**.

De **gemodificeerde consumptie-eenheid** is een equivalentieschaal die wordt toegepast om de consumptie-uitgaven aan te passen in functie van de omvang en de samenstelling van het huishouden. Een coëfficiënt 1 wordt toegekend aan de eerste volwassene, een coëfficiënt 0,5 aan de overige personen ouder dan 13 jaar, en een coëfficiënt 0,3 aan kinderen jonger dan 13 jaar (aangepaste OESO-schaal).

2. Samenstelling van de steekproef van respondenten (representativiteit)

2.1. Huishoudverdelingen

Verdeling naar type type huishoudgrootte

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
Total	4490	100	4852487	100

Opmerkingen

- 1) In de steekproef van responderende huishoudens is 9 de hoogste huishoudgrootte, in de populatie is dat 31!
- 2) De som van de gewichten (4,852,487) is gelijk aan het populatiecijfer, omdat zowel LFS als HBS gekalibreerd worden op het globaal aantal huishoudens.
- 3) De gewogen en de populatieverdeling van MS_Size verschillen, ondanks en vanwege een beperkte kalibratie naar huishoudgrootte in HBS.

Verdeling van het aantal actieven per huishouden

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	
Total	4490	100		4852487	100	

2.2. [Individuele verdelingen](#)

Geslachtsverdeling van de leden van de deelnemende huishoudens

CD_Gender	HBS 2016			
	Unweighted		Weighted	
	Freq.	%	Freq.	%
Man	5230	48.51	5299402	47.91
Vrouw	5552	51.49	5762425	52.09
Total	10782	100	11061826	

Opmerking

Geslacht is een kalibratievariabele voor LFS, maar niet voor HBS. Daardoor is de som van de individuele HBS-gewichten (11,061,826) niet gelijk aan het populatiecijfer.

Verdeling van de leeftijd van de leden van de deelnemende huishoudens

MS_Age	<i>HBS 2016</i>			
	<i>Unweighted</i>		<i>Weighted</i>	
	Freq.	%	Freq.	%
- 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
Total	10782	100	11061826	100

Verdeling van het socio-professionele statuut van de leden van de deelnemende huishoudens

Socstat	HBS 2016					
	Unweighted			Weighted		
	Freq.	%	Cum. %	Freq.	%	Cum. %
Zelfstandige (inclusief meewerkend familielid)	373	3.51	3.51	648737.6	5.97	5.97
Werknemer	4001	37.7	41.22	3711362	34.15	40.11
Werkloos	456	4.3	45.51	449621.5	4.14	44.25
Gepensioneerd/Bruggepensioneerd	1783	16.8	62.32	1921385	17.68	61.93
Invalide	448	4.22	66.54	455481.5	4.19	66.12
Huisman of -vrouw (verzorgt het huishouden)	313	2.95	69.49	338918.6	3.12	69.24
Baby, scholier of student	3095	29.17	98.65	3180745	29.26	98.5
Andere niet-actieven	143	1.35	100	163017.3	1.5	100
Totaal	10612	100		10869268	100	

Number of missing values: 170

Opmerking

Verschil tussen het populatiecijfer 10,869,268 in deze tabel en het populatiecijfer is te wijten aan het aantal *missing values* voor de variabele Socstat.