

Kwalitatieve werkgelegenheidsdata voor België, een SAM-aanpak voor de periode 1999-2005

Februari 2007

*Vaast Bresseleers, vb@plan.be,
Koen Hendrickx, kh@plan.be,
Bart Hertveldt, bh@plan.be,
Bart Van den Cruyce, bv@plan.be,
Johan Wera, jw@plan.be*

Abstract - In deze paper wordt een methodologie beschreven om een aantal sociale gegevens (zowel administratieve als enquêtegegevens) te enten op de informatie uit de nationale rekeningen, ten einde een kwalitatief beeld te krijgen van de (ontwikkeling van de) werkgelegenheid. Meer bepaald wordt de binnenlandse werkgelegenheid opgesplitst naar persoonskenmerken (geslacht, leeftijdsklasse en opleidingsniveau) en jobkenmerken (statuut en arbeidsregime). Die opdeling gebeurt op een gedetailleerd bedrijfstakniveau en conform de nationale rekeningen voor de jaren 1999 tot 2005. De hier ontwikkelde methodologie en resultaten vormen een eerste stap in de richting van de creatie van een Social Accounting Matrix (SAM) voor België.

Jel Classification - C8, E24, J23

Keywords - Arbeidsmarktdata, Sectorale en intersectorale analyses

Acknowledgements - Vooreerst wensen we de Algemene Directie Statistiek en Economische Informatie (NIS), de Nationale Bank van België en de instellingen voor sociale zekerheid (RSZ, RSZPPO, RSVZ, RVA) te bedanken voor het ter beschikking stellen van de nodige basisdata. Onze dank gaat ook uit naar Jeroen Fiers van het FPB, voor zijn waardevolle commentaren. Tot slot ook een woord van dank aan de leden van de coördinatieceel van het INR, aan wie in februari 2006 een voorlopige versie van de methodologie en de resultaten werd voorgesteld.

Inhoudstafel

Inleiding	1
1. Opdeling van de werknemers naar geslacht, leeftijdsklasse, statuut en arbeidsregime	5
1.1. Overzicht van de opdelingen	5
1.2. Overzicht van de gebruikte basisbronnen	6
1.3. Randtotalen uit de nationale rekeningen	6
1.4. De opdeling van de loontrekkenden naar persoons- en jobkenmerken	6
1.4.1. De reguliere RSZ-werkgelegenheid inclusief diverse correcties	6
1.4.2. De reguliere RSZPPO-werkgelegenheid	9
1.4.3. Jobstudenten in de RSZ-bestanden	10
1.4.4. Jobstudenten in de RSZPPO-bestanden	10
1.4.5. Niet in de RSZ of RSZPPO-bestanden vervatte werknemers	11
2. Opdeling van de zelfstandigen naar geslacht en leeftijdsklasse	12
2.1. De bronnen voor de randtotalen en de behandeling van specifieke groepen	12
2.1.1. De nationale rekeningen	12
2.1.2. De gegevens van het RSVZ	12
2.1.3. De berekening voor vergoede helpers	13
2.1.4. De randtotalen voor niet-vergoede helpers	13
2.1.5. De cijfers voor onthaalouders	14
2.2. Raming van geslacht en leeftijdsstructuur per bedrijfstak voor gewone zelfstandigen en niet-vergoede helpers	14
2.2.1. Algemeen	14
2.2.2. De ordered logit regressietechniek	14
2.2.3. Nested logit regressie	15
2.2.4. Verklarende variabelen in de regressie voor het bepalen van de leeftijdsklasse	18
2.2.5. Raming van de verdeling naar geslacht op basis van EAK-gegevens	20
2.2.6. De weging van de observaties	21
2.3. Resultaten voor verdeling zelfstandigen per bedrijfstak	21
3. Verdere opdeling van werknemers en zelfstandigen naar opleidingsniveau	23
3.1. De bronnen voor een opdeling naar opleidingsniveau	23
3.2. Motivering van de keuze voor regressietechnieken	24
3.3. De gehanteerde regressietechniek	25
3.3.1. De organisatie van de regressies (nested logit structuur)	25
3.3.2. De behandeling van cellen die ontbreken in de EAK	27
3.4. De variabelen in de regressies	27
3.4.1. De afhankelijke variabele	27
3.4.2. De verklarende variabelen	27
3.5. De behandeling van verschillen tussen de EAK en de nationale rekeningen	31
3.5.1. De beheerders van vennootschappen	31
3.5.2. De uitzendkrachten	32
3.5.3. De vergoede helpers	32
3.5.4. Grensarbeid	33
3.6. De resultaten van de SAM	33
3.6.1. Verschillen veroorzaakt door de opname van het jaartal als een continue variabele	33
3.6.2. Verschillen veroorzaakt door een andere weging in de SAM en de EAK	35
Besluit	37
Bijlagen	38

Lijst van tabellen

Tabel 1	Overeenkomst tussen werknemers volgens RSZ en nationale rekeningen per IOKLEMS- branche (2000)	7
Tabel 2	Opsplitsing van de loontrekkenden in branche 75A3 naar persoons- en jobkenmerken: vergelijking per sz-bron (2000)	10
Tabel 3	Aandeel vrouwen per leeftijdsklasse in de EAK, de regressies en de SAM (2000)	21
Tabel 4	Aandeel (in %) in SAM en EAK van combinaties van leeftijdsklasse en statuut (2000)	35
Tabel 5	Opleidingsniveau werkenden in SAM versus EAK (2000)	36

Lijst van figuren

Figuur 1	Schematisch overzicht van de methodologie en de bronnen	3
Figuur 2	Boomstructuur regressies zelfstandigen en niet-vergoede helpers	17
Figuur 3	Evolutie van de statuten tussen 1999 en 2005 in de EAK	29
Figuur 4	Ontwikkeling van het opleidingsniveau van -30 jarige arbeiders	34

Inleiding

Situering en context

De nationale rekeningen geven een beschrijving van de economische processen binnen een land en de economische relaties met het buitenland. Ze vormen de bron bij uitstek om een beeld te krijgen over de *kwantitatieve* ontwikkeling van een aantal sleutelvariabelen (productie, werkgelegenheid...), met als bijkomende voordelen interne consistentie en internationale vergelijkbaarheid. Naast die kwantitatieve ontwikkelingen is het vanuit analyse- en beleidsoogpunt evenzeer van belang een beter inzicht te verwerven in de kwalitatieve aspecten van belangrijke economische grootheden. In deze paper wordt een methodologie beschreven om een aantal sociale gegevens (zowel administratieve als enquêtegegevens) te enten op de informatie uit de nationale rekeningen, ten einde een *kwalitatief* beeld te krijgen van de (ontwikkeling van de) werkgelegenheid.

Met deze werkzaamheden wordt voortgebouwd op de vroegere activiteiten van de INR-coördinatrice, die in 2000 een methodologie ontwikkelde en publiceerde voor het leveren van kwalitatieve data in het kader van ESA Employment over tewerkstelling, gewerkte uren en beloning.¹ In de hier voorliggende paper wordt enkel het luik hoofdelijke werkgelegenheid behandeld. Meer concreet wordt de werkgelegenheid per bedrijfstak opgesplitst naar persoonskenmerken (geslacht, leeftijdsklasse en opleidingsniveau) en jobkenmerken (statuut en arbeidsregime). De resultaten die uit de beschreven methodologie voortvloeien, werden onder meer gebruikt voor het voeden van het Belgische luik van de EUKLEMS-databank.²

Naar een SAM aanpak

De hier ontwikkelde methodologie en resultaten vormen een eerste stap in de richting van de creatie van een Social Accounting Matrix (SAM) voor België. In de handleiding van het ESR 1995 wordt een SAM als volgt gedefinieerd: *“The presentation of ESA-accounts in a matrix which elaborates on the linkages between a supply and use table and sector accounts. SAMs typically focus on the role of people in the economy, which may be reflected by, among other things, extra breakdowns of the household sector and a disaggregated representation of labour markets (i.e. distinguishing various categories of employed persons)”* (Eurostat, 1996).

Uit bovenstaande omschrijving blijkt dat een SAM vertrekt van een matrixpresentatie van de nationale rekeningen, waarbij alle klassieke economische deelprocessen (productie, inkomensvorming, inkomens(her)verdeling, inkomensbesteding en financiering) in één raamwerk worden samengebracht (een zogenaamde NAM, National Accounting Matrix). Een SAM geeft

¹ Zie: NIS, MEZ, Sociale statistieken en nationale rekeningen: naar een SAM aanpak, april 2000.

² Het EUKLEMS-project heeft tot doel analyses uit te voeren met betrekking tot de productiviteit in de Europese Unie op bedrijfstakniveau. Het project wordt gefinancierd door de Europese Commissie en wordt uitgevoerd door een consortium van 15 Europese onderzoeksinstituten (waaronder het FPB).

vervolgens een aantal extra dimensies aan de standaardinformatie uit de nationale rekeningen. Typisch hierbij is dat de nadruk wordt gelegd op de rol van mensen in de economie, wat onder meer tot uiting komt in een opsplitsing van de sector huishoudens naar sociaal-economische deelgroepen en een onderscheid van de werkgelegenheid naar de verschillende categorieën werkenden. Een SAM is derhalve een nuttig analyse-instrument, onder meer voor belangrijke sociale vraagstukken die te maken hebben met de samenstelling van de werkgelegenheid en de werkloosheid.

Basisprincipes van de ontwikkelde methodologie

Bij de in deze paper beschreven methode voor de opsplitsing van de binnenlandse hoofdelijke werkgelegenheid naar kwaliteitskenmerken werd uitgegaan van drie basisprincipes:

- coherentie met de nationale rekeningen;
- mogelijkheid van jaarlijkse updating;
- toepasbaarheid op een sterk gedesaggregeerd werkniveau.

Elk van die basisprincipes zal in het vervolg van de paper meer in detail worden uitgewerkt. Hier beperken we ons tot een algemene bespreking van de principes.

Vooreerst werd gestreefd naar een zo groot mogelijke coherentie met de nationale rekeningen. Hiertoe werd maximaal gebruik gemaakt van dezelfde – voornamelijk administratieve – bronnen als in de nationale rekeningen. Bovendien werd, bij de verwerking van de administratieve bestanden, de toewijzing van NACE-bedrijfstakcodes aan de ondernemingen op identieke wijze gedaan als in de nationale rekeningen en werden, voor zover relevant, bepaalde categorieën werkenden aan de werkgelegenheid volgens de klassieke administratieve bronnen toegevoegd, eruit geschrapt of naar andere bedrijfstakken verschoven. In elk geval worden in de uiteindelijke resultaten de randtotalen voor de werkgelegenheid per statuut (werknemers; zelfstandigen) per bedrijfstak uit de nationale rekeningen gerespecteerd.

Van meet af aan bestond het opzet in het ontwikkelen van een methode die niet enkel toepasbaar is op individuele jaren, maar dupliceerbaar is voor aansluitende jaren. Bij de keuze van de bronnen werd daarom de voorkeur gegeven aan gegevens die jaarlijks (met hetzelfde detail) beschikbaar zijn. Dat sluit uiteraard niet uit dat, bij wijze van ankerpunt of verificatie, ook rekening kan gehouden worden met eenmalige of discontinue informatie. Om uitspraken te kunnen doen over het verloop van de samenstelling van de werkgelegenheid over de tijd, werd bijzondere aandacht besteed aan het vermijden van breuken in de reeksen. De hier beschreven methode kan momenteel worden toegepast op de periode 1999-2005. Retropolaties³ zijn mogelijk op een meer geaggregeerd niveau en op basis van minder gesofistikeerde methodes, maar de beschrijving ervan maakt niet het voorwerp uit van deze paper.

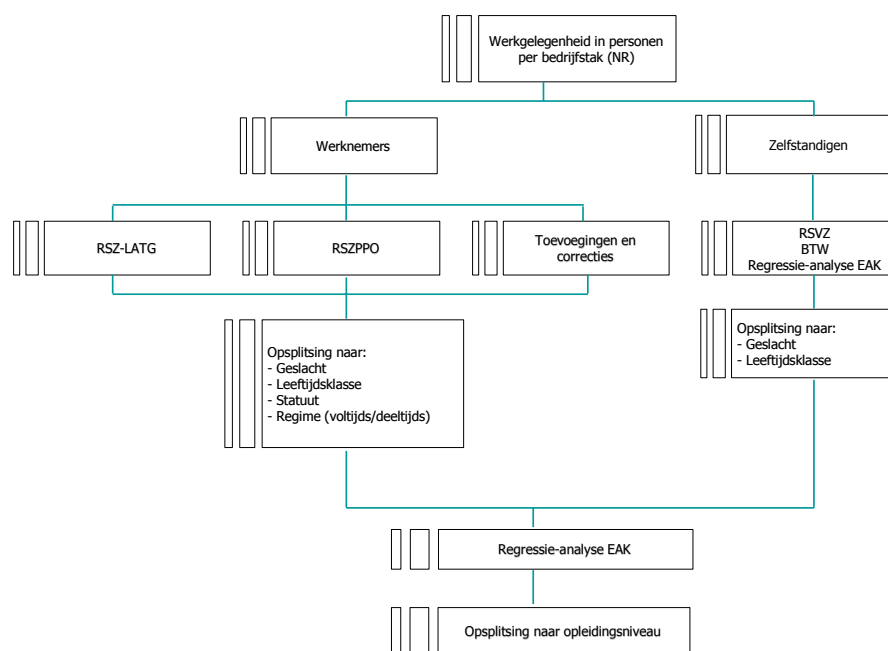
³ Zo worden bijvoorbeeld in het kader van het EUKLEMS-project retropolaties tot 1970 gevraagd.

Er werd voor geopteerd de berekeningen uit te voeren op een sterk gedesaggregeerd niveau. Zo worden op het werkniveau een 150-tal bedrijfstakken onderscheiden (de zogenaamde IOKLEMS-branches, zie bijlage 1). Een fijn niveau van desaggregatie heeft het voordeel meer flexibiliteit te bieden bij de keuze van het aggregatieniveau waarop analyses uitgevoerd worden en resultaten gerapporteerd worden. Bovendien zouden een aantal hypothesen die deel uitmaken van de gekozen methodologie minder goed verdedigbaar geweest zijn op een meer geaggregeerd niveau. De keuze voor een sterke desaggregatie doet evenwel een probleem rijzen in termen van beschikbaarheid en stabiliteit van bepaalde brondata (in het bijzonder van de gegevens uit de enquête naar de arbeidskrachten, EAK). Hieraan werd gedeeltelijk verholpen door een beroep te doen op regressietechnieken, ten einde de robuustheid van de enquêtegegevens te verhogen.

Schematisch overzicht van de methodologie en de bronnen

Onderstaande figuur geeft een schematisch overzicht van de stappen in de gevolgde methodologie en het gebruikte bronnenmateriaal.

Figuur 1 Schematisch overzicht van de methodologie en de bronnen



Bron: FPB.

Uit bovenstaande figuur valt af te leiden dat de opsplitsing van de werknemers naar geslacht, leeftijdsklasse, statuut en arbeidsregime (zie hoofdstuk 1) bijna uitsluitend gebeurt aan de hand van administratieve data (RSZ-LATG en RSZPPO). De opdeling van de zelfstandigen naar geslacht en leeftijdsklasse (zie hoofdstuk 2) geschiedt aan de hand van een combinatie van administratieve bronnen (RSVZ en BTW) en enquêtegegevens (EAK). Voor de opdeling van

werknemers en zelfstandigen naar opleidingsniveau (zie hoofdstuk 3) vormt de EAK de hoofdbron. Vooralsnog is het opleidingsniveau immers niet beschikbaar als informatie in de administratieve bronnen.

1. Opdeling van de werknemers naar geslacht, leeftijdsklasse, statuut en arbeidsregime

Dit hoofdstuk beschrijft de bronnen en methodes die gebruikt worden bij de opdeling van de (binnenlandse) loon- en weddetrekkende werkgelegenheid per bedrijfstak naar geslacht, leeftijdsklasse, statuut en arbeidsregime. Die opdeling gebeurt grotendeels op basis van administratieve data, meer bepaald via microdata afkomstig van de instellingen voor sociale zekerheid (RSZ en RSZPPO), die op het niveau van het individu geanonimiseerd werden. Wegens de verschillen in basisbronnen en methode wordt de indeling naar opleidingsniveau in een apart hoofdstuk behandeld (zie hoofdstuk 3). De beschreven methode werd toegepast op de periode 1999-2005.⁴ In deze oefening werd gestreefd naar een zo groot mogelijke coherentie met de concepten en de cijfers uit de nationale rekeningen, die worden opgesteld volgens het ESR95.⁵

1.1. Overzicht van de opdelingen

De volgende opdelingen worden gemaakt:

- een indeling naar IOKLEMS-bedrijfstakken: 147 branches (zie bijlage 1)
- een opdeling naar de volgende persoonskenmerken:
 - geslacht: man/vrouw (x2)
 - leeftijdsklasse: 15-19, 20-24, ..., 55-59, 60-64 en 65+ (x11)
- een opdeling naar de volgende jobkenmerken⁶:
 - statuut: arbeider/bediende/ambtenaar (x3)
 - arbeidsregime: voltijds/deeltijds (x2)

Daarnaast werd eveneens een indeling naar soort sector (privé/publiek) overeenkomstig de RSZ-definities⁷ doorgevoerd. Die indeling werd toegevoegd om beter de relatie te kunnen leggen met data uit de enquête naar de arbeidskrachten. Hierdoor kunnen immers de contractuelen (arbeiders en bedienden uit de publieke sector) worden onderscheiden van de ambtenaren. Zo kunnen dan de opleidingsniveaus, die worden opgesteld aan de hand van de opdeling naar de andere persoons- en jobkenmerken, betrouwbaarder worden geraamd (zie hoofdstuk 3).

In totaal gaat het hier dus om 264 persoons- en jobkenmerken en 147 branches, of 38 808 mogelijke combinaties. Daarvan zijn slechts een ruime 15 000 relevant. Vooral de RSZ-sector “publiek” is in veel gevallen irrelevant.

⁴ De cijfermatige illustraties die in de tekst werden opgenomen slaan, tenzij anders vermeld, op het jaar 2000.

⁵ Meer bepaald met: INR, Nationale rekeningen, Deel 2 – Gedetailleerde rekeningen en tabellen 1995-2005, november 2006.

⁶ De jobkenmerken zijn gebaseerd op de hoofdactiviteit van de werknemer.

⁷ Die indeling gebaseerd op RSZ-definities verschilt van de indeling in institutionele sectoren volgens de nationale rekeningen.

1.2. Overzicht van de gebruikte basisbronnen

De verdeling van de werknemers naar persoons- en jobkenmerken valt uiteen in vijf deelberekeningen, respectievelijk voor:

- “reguliere” werkgelegenheid in de RSZ-bestanden, inclusief diverse correcties (2 997 000 personen in 2000)
- “reguliere” werkgelegenheid in de RSZPPO-bestanden (293 400 personen in 2000)
- jobstudenten in de RSZ-bestanden (19 600 personen in 2000)
- jobstudenten in de RSZPPO-bestanden (1 200 personen in 2000)
- niet in de RSZ- of RSZPPO-bestanden vervatte werknemers (85 200 personen in 2000).

De aparte behandeling van “reguliere” werkgelegenheid en studentenarbeid was nodig omdat de administratieve bestanden minder gedetailleerde informatie bevatten voor studentenjobs dan voor de overige werkgelegenheid.

1.3. Randtotalen uit de nationale rekeningen

Voor elk van die vijf groepen werknemers zijn (niet-gepubliceerde) randtotalen per bedrijfstak beschikbaar op het werkniveau dat gehanteerd wordt in de nationale rekeningen (een 120-tal branches). Waar nodig werden de randtotalen uit de nationale rekeningen verder opgesplitst tot op IOKLEMS-bedrijfstakniveau aan de hand van de informatie uit de bovenvermelde administratieve basisbronnen. Ook voor elk van de correcties op de reguliere RSZ-tewerkstelling kon een beroep gedaan worden op (niet-gepubliceerde) randtotalen uit de nationale rekeningen.

1.4. De opdeling van de loontrekkenden naar persoons- en jobkenmerken

Zoals hierboven uitgelegd verloopt de berekening afzonderlijk voor elk van de vijf groepen werknemers. Telkens wordt via de relevante basisbron eerst een volledige gekruiste (bedrijfstakken x persoons- en jobkenmerken) raming gemaakt van de werkgelegenheid; vervolgens wordt die raming proportioneel herschaald zodat de gegeven randtotalen voor de werkgelegenheid per bedrijfstak uit de nationale rekeningen gerespecteerd worden. In wat volgt wordt dieper ingegaan op elk van de deelberekeningen.

1.4.1. De reguliere RSZ-werkgelegenheid inclusief diverse correcties

De reguliere werkgelegenheid in de RSZ-bestanden

De LATG-databank (“Loon- en ArbeidsTijdGegevens”) van de Rijksdienst voor Sociale Zekerheid (RSZ) is gebaseerd op de gegevens van de trimestriële aangiften⁸ van de werkgevers aan de RSZ en bevat onder meer informatie over de opdeling naar geslacht, leeftijdsklasse, statuut, ar-

⁸ De LATG-kwartaaldata werden eerst omgezet naar kwartaalgemiddelden (rekenkundig gemiddelde van de eindekwartaalcijfers) en vervolgens naar jaargemiddelden.

beidsregime en soort sector van alle loon- en weddetrekkenden waarvoor sociale zekerheidsbijdragen betaald worden aan de RSZ.

Met het oog op een zo groot mogelijke coherentie met de nationale rekeningen en een maximale homogeniteit van de data over de tijd, werd, bij de verwerking van de RSZ-gegevens, de toewijzing van NACE-bedrijfstakcodes aan de ondernemingen op identieke wijze gedaan als in de nationale rekeningen.

De reguliere werkgelegenheid uit de LATG-databank vertegenwoordigt bijna 90 % van de totale werknemerspopulatie. Voor 78 % van de branches op IOKLEMS-niveau bedraagt de afwijking tussen de bruto RSZ-data en het uiteindelijke bedrijfstaktotaal uit de nationale rekeningen minder dan 5 %. Vooral de industriebranches kunnen goed benaderd worden door LATG-data: voor 54 % van de industriële branches bedraagt de afwijking minder dan 1 %. Bij de dienstenbranches zijn er duidelijk meer afwijkingen, voor 16 % van de dienstenbranches is de afwijking zelfs groter dan 20 %.

Tabel 1 Overeenkomst tussen werknemers volgens RSZ en nationale rekeningen per IOKLEMS-branche (2000)

% afwijking RSZ/NR	Totaal	Primaire sector	Industrie	Diensten
<1%	44%	0%	54%	34%
<5%	78%	33%	89%	70%
<10%	84%	33%	93%	79%
<20%	88%	100%	94%	84%

Bron: RSZ, NBB, berekeningen FPB.

Vervolgens werd, als eerste tussenresultaat, een kruistabel opgesteld met de IOKLEMS-branches en alle nodige persoons- en jobkenmerken. Daarbij werd de groep werknemers waarvoor de leeftijdsklasse in de RSZ-bestanden niet gekend was (4 250 personen in 2000), per branche, sector, geslacht en statuut proportioneel versleuteld over de overige leeftijdscategorieën.

Correcties op de bruto RSZ-data

Bij het opstellen van de werkgelegenheidsreeksen volgens de nationale rekeningen worden voor een aantal categorieën werknemers “correcties” gedaan op de LATG-basisdata. Die correcties kunnen de vorm aannemen van weglatingen, bijschattingen of verschuivingen tussen branches. In de mate van het mogelijke werd in de hier ontwikkelde methodologie eveneens rekening gehouden met de specifieke opdeling naar persoons- en jobkenmerken van die globaal aangebrachte correcties.

Een aantal aanpassingen werden weliswaar buiten beschouwing gelaten. In sommige gevallen was het weinig zinvol om een verdeling naar persoons- en jobkenmerken op te stellen omdat de aanpassingen kwantitatief onbelangrijk waren en/of er van uitgegaan kon worden dat zij geen significant afwijkende verdeling naar kenmerken bevatten dan de rest van de werkgelegenheid

in de betreffende branche. In een aantal andere gevallen beschikten we niet over de data om de correcties op te splitsen naar kenmerken. Telkens wordt dan impliciet verondersteld dat de verdeling naar kenmerken identiek is aan die voor de overige werkgelegenheid in de branche. Hieronder worden de correcties opgesomd waarvoor wél expliciet een aparte verdeling gemaakt werd.

Langdurig zieken

De langdurig zieken die geen loon ontvangen van hun werkgever worden slechts opgenomen in de RSZ-bestanden vanaf 2003. Zij worden in de nationale rekeningen tot de werkgelegenheid gerekend en moeten dus voor de jaren voorafgaand aan 2003 worden toegevoegd aan de RSZ-werkgelegenheidscijfers. Gezien de specifieke kenmerken van deze groep (het aantal arbeiders is hier bijvoorbeeld relatief groot) loonde het de moeite om een aparte verdeling naar kenmerken op te stellen. Daartoe werd eerst een volledige kruistabel voor langdurig zieken volgens IOKLEMS-takken en persoons- en jobkenmerken opgesteld voor het jaar 2004 op basis van de informatie in de RSZ-bestanden. Vervolgens werd voor de periode voorafgaand aan 2003 verondersteld dat – per afzonderlijke cel in de kruising – de voor 2004 geobserveerde proportie langdurig zieken in het totale aantal werknemers eveneens van toepassing was. Dat leverde een eerste raming voor de ontbrekende periode. Die werd ten slotte proportioneel herschaald naar de globale hypothese voor langdurig zieken in de nationale rekeningen voor het betreffende jaar (afgerond 18 000 personen in 2000, bijvoorbeeld).

Ter beschikking gestelde leerkrachten (TBS)

De zogenaamde ter beschikking gestelde, en de facto niet meer werkende, leerkrachten worden in de nationale rekeningen niet tot de werkgelegenheid gerekend (het gaat om ongeveer 14 000 personen in 2000). Zij moeten dus geschrapt worden uit de RSZ-cijfers voor branche 80A3, hoewel zij a priori in de RSZ-bestanden niet onderscheiden kunnen worden van de overige werkgelegenheid. Deze correctie wordt vooral gekenmerkt door een specifieke leeftijdsstructuur (het betreft uitsluitend 55-plussers). De verdeling naar leeftijd en geslacht werd éénmalig (voor het jaar 2000) opgesteld op basis van statistieken voor het schooljaar 1999-2000 van de Vlaamse en Franstalige Gemeenschap. De verdere verdeling per statuut, RSZ-sector en arbeidsregime werd identiek verondersteld aan die van de geobserveerde verdeling van de totale RSZ-werkgelegenheid per geslacht en leeftijdsklasse in de betreffende branche. Voor de overige jaren werd een eerste raming gemaakt via een hypothese van constante aandelen TBS'ers in de werkgelegenheid per kruising. Vervolgens werd proportioneel herschaald naar de globale hypothese voor TBS'ers in de nationale rekeningen voor elk jaar.

Verschuivingen van loontrekkenden van branche 75A3 naar branches 63B3 en 22A3

Op basis van de administratieve gegevens is het niet mogelijk rechtstreeks verdeelsleutels voor branches 63B3 (infrastructuurwerken door de overheid) en 22A3 (het Belgisch Staatsblad) te ramen. Die tewerkstelling is in de RSZ-bestanden onlosmakelijk verbonden met de werkgelegenheid in branche 75A3 (overheidsadministratie). De overhevelingen van werkgelegenheid die

in de nationale rekeningen gebeuren (27 100 personen van branche 75A3 naar branche 63B3 en bijna 300 personen naar branche 22A3 in het jaar 2000) worden bij gebrek aan nadere informatie opgedeeld naar kenmerken aan de hand van de verdeelsleutels voor de totale RSZ-werkgelegenheid in branche 75A3.

Weglating van een aantal categorieën

Een aantal categorieën personen, die in de LATG-data als loon- en weddetrekkenden worden aangemerkt, moeten, overeenkomstig de nationale rekeningen, worden geschrapt:

- vanaf 2003 worden de onthaalouders (branche 85C1) en de lesgevers van middenstandsopleidingen (branche 80A3) opgenomen in de RSZ-tewerkstelling, terwijl die categorieën volgens de nationale rekeningen bij de zelfstandigen moeten worden ondergebracht;
- ten gevolge van de politiehervorming verschuift vanaf 2002 werkgelegenheid van RSZ- naar RSZPPO-aangiften. Gedurende de eerste twee kwartalen van 2002 werd een gedeelte van het personeel echter opgenomen in zowel RSZ- als RSZPPO-bestanden. Die dubbeltelling werd verwijderd uit de RSZ-data.

1.4.2. De reguliere RSZPPO-werkgelegenheid

De RSZPPO-statistieken bevatten de gegevens van de sociale zekerheid voor werknemers tewerkgesteld in provinciale en plaatselijke overheidsdiensten. Het gaat hier voor het jaar 2000 om bijna 9 % van de totale werknemerspopulatie. Omdat die werknemers geconcentreerd zijn in een aantal specifieke branches is het belangrijk ook afzonderlijk met die bron rekening te houden. Ongeveer 220 000 werknemers, dit is 75 % van de populatie, behoren tot branche 75A3 (openbaar bestuur, exclusief defensie en verplichte sociale verzekering) en bijna 50 000 personen horen thuis in branche 85A1 (gezondheidszorg). De rest is verspreid over een dertigtal IOKLEMS-branches.

Zoals bij de verwerking van de RSZ-gegevens, werd een zo groot mogelijke overeenstemming met de branche-indeling uit de nationale rekeningen nagestreefd. Net als de RSZ-bestanden, laten ook de RSZPPO-data toe een opdeling te maken van de werkgelegenheid per branche naar geslacht, leeftijdsklasse, statuut en arbeidsregime. Per definitie behoort in dit geval echter alle werkgelegenheid tot de (administratieve) sector “publiek”.

Onderstaande tabel illustreert (voor het jaar 2000) het belang van de informatie uit de RSZPPO-databank voor bedrijfstak 75A3 (de branche waartoe de meeste RSZPPO-loontrekkenden behoren). Bijna 30 % van het totaal aantal werknemers van branche 75A3 wordt naar persoons- en jobkenmerken opgesplitst op basis van RSZ-data, terwijl de resterende 70 % verdeeld wordt aan de hand van RSZPPO-data. In die laatste groep is het aandeel van de arbeiders merklijk hoger (ten nadele van de ambtenaren), en zijn vrouwen en deeltijders iets sterker vertegenwoordigd.

Tabel 2 Opsplitsing van de loontrekkenden in branche 75A3 naar persoons- en jobkenmerken: vergelijking per sz-bron (2000)

		RSZ-LATG		RSZPPO		TOTAAL	
Geslacht	V	39800	43%	121700	55%	162200	52%
	M	52600	57%	99400	45%	152600	48%
Leeftijd	15-29	13100	14%	34700	16%	49100	16%
	30-49	56200	61%	137100	62%	193300	61%
	50+	23100	25%	49300	22%	72400	23%
Statuut	Arbeider	8000	9%	65000	29%	73400	23%
	Bediende	22500	24%	57800	27%	81100	26%
	Ambtenaar	62000	67%	98300	44%	160300	51%
Regime	Deeltijds	16100	17%	64800	29%	81700	26%
	Voltijds	76300	83%	156300	71%	233100	74%
Totaal		92400	100%	221100	100%	314800	100%

Bron: RSZ, RSZPPO, NBB, berekeningen FPB.

1.4.3. Jobstudenten in de RSZ-bestanden

Ook in dit geval werd de RSZ-informatie afgestemd op de randtotalen per bedrijfstak uit de nationale rekeningen. Via de RSZ-bestanden kunnen dezelfde kruisingen opgesteld worden voor studentenarbeid als voor de reguliere werkgelegenheid, behalve de opsplitsing naar arbeidsregime. Bij gebrek aan informatie werd verondersteld dat het hier exclusief om deeltijders gaat. Bovendien is het zo dat bij studentenarbeid de RSZ-data niet de toestand van de werkgelegenheid op het einde van het trimester bevatten, maar een gecumuleerd aantal personen geven dat tijdens één of ander moment van het kwartaal een studentenjob heeft uitgeoefend⁹. Die cijfers overschatten dus aanzienlijk het aantal personen dat gemiddeld tijdens het kwartaal aan de slag is in studentenjobs; vandaar het belang om voor jobstudenten afzonderlijk – los dus van de reguliere werkgelegenheid – te herschalen naar de gegeven randtotalen per bedrijfstak uit de nationale rekeningen.

1.4.4. Jobstudenten in de RSZPPO-bestanden

Ook in de RSZPPO-bestanden kan studentenarbeid worden onderscheiden van reguliere werkgelegenheid en was een gelijkaardige opdeling naar persoonskenmerken mogelijk, zij het dat in dit geval de opdeling naar statuut ontbreekt in de basisdata. Bij gebrek aan die informatie, en in de wetenschap dat studentenarbeid per definitie nooit onder het ambtenarenstatuut valt, werd de opdeling arbeider/bediende per geslacht, leeftijdsklasse, arbeidsregime en IOKLEMS-branche uit de reguliere RSZPPO-werkgelegenheid exclusief ambtenaren toegepast.

⁹ In de praktijk bevatten, tot en met het jaar 2004, enkel de RSZ-cijfers voor het derde kwartaal studentenjobs.

1.4.5. Niet in de RSZ of RSZPPO-bestanden vervatte werknemers

Voor twee categorieën werknemers die (vrijwel volledig) buiten het toepassingsgebied van RSZ en RSZPPO vallen, worden afzonderlijke verdelingen geschat, nl. de PWA'ers en de huisbedienden.

PWA'ers

PWA'ers worden in de nationale rekeningen bij de tewerkstelling van NACE-branche 74 geteld. Het randtotaal uit de nationale rekeningen (15 600 PWA'ers voor 2000) kon worden opgedeeld naar geslacht en leeftijdsklasse aan de hand van data die ons door de RVA ter beschikking werden gesteld. Verder weten we dat het hier per definitie om deeltijders gaat en veronderstellen we bovendien dat het om arbeiders in de (administratieve) sector "privé" gaat.

Huisbedienden

De administratief gekende werkgelegenheid in branche 95A4 (particuliere huishoudens met werknemers) kan bezwaarlijk representatief worden geacht voor het geheel van de werkgelegenheid dat in de nationale rekeningen tot diezelfde tak wordt gerekend. In 2000 bevatten de RSZ-bestanden 2 600 personen voor die branche, tegenover een globale raming van 69 600 personen in de nationale rekeningen. De RSZ-werkgelegenheid in branche 95A4 werd als dusdanig behouden (binnen de deelberekening die aan bod kwam in 1.4.1), inclusief zijn structuur naar kenmerken. Het restant kan beschouwd worden als een bijschatting voor informele arbeid die in de nationale rekeningen gemaakt wordt. Bij gebrek aan informatie over de verdeling naar persoons- en jobkenmerken van die informele arbeid, werd verondersteld dat de hierboven geraamde verdeling naar kenmerken voor PWA'ers als beste benadering geldt.

2. Opdeling van de zelfstandigen naar geslacht en leeftijdsklasse

In dit hoofdstuk wordt toegelicht hoe zelfstandigen per bedrijfstak ingedeeld worden naar leeftijdsklasse en geslacht voor de periode 1999-2005. Dat gebeurt op basis van geanonimiseerde microgegevens van de jaarlijkse NIS-enquête naar de arbeidskrachten (EAK), met randtotalen per bedrijfstak uit de nationale rekeningen (niet-gepubliceerd detail) en randtotalen per geslacht en leeftijdsklasse uit de statistieken van het Rijksinstituut voor de Sociale Verzekeringen der Zelfstandigen (RSVZ) of de EAK.

Eerst worden de bronnen voor de randtotalen en de behandeling van specifieke groepen beschreven. In sectie 2.2 wordt uitgelegd hoe de verdeling over leeftijdsklassen en geslacht geraamd werd voor gewone zelfstandigen en niet-vergoede helpers. In sectie 2.3 wordt de overgang naar resultaten voor hoofdelijke werkgelegenheid en de equilibrering besproken.

2.1. De bronnen voor de randtotalen en de behandeling van specifieke groepen

2.1.1. De nationale rekeningen

In de nationale rekeningen worden als zelfstandige beschouwd: gewone zelfstandigen, vergoede helpers en niet-vergoede helpers. In 2000 maakten die respectievelijk 84 %, 4 % en 12 % uit van het totaal aantal zelfstandigen. De nationale rekeningen bevatten tijdreeksen voor de zelfstandigen die lopen van 1995 tot en met 2005.¹⁰ De indeling naar bedrijfstakken is in de nationale rekeningen grotendeels gebaseerd op BTW-gegevens.

De gegevens van de nationale rekeningen werden aangewend als totaal voor het aantal gewone zelfstandigen, vergoede helpers en niet-vergoede helpers per bedrijfstak. Daar waar de IOKLEMS-classificatie dit vereiste, werd het aantal zelfstandigen en helpers verder gedesaggregeerd op basis van BTW-statistieken.

2.1.2. De gegevens van het RSVZ

Het RSVZ is een administratieve bron met gegevens over zelfstandigen. In tegenstelling tot de BTW-statistieken geven de RSVZ-data een opdeling naar geslacht en leeftijdsklasse. De RSVZ-statistiek beschrijft de situatie op 31 december van het betreffende jaar.

¹⁰ INR, Nationale rekeningen, Deel 2 – Gedetailleerde rekeningen en tabellen 1995-2005, november 2006.

De RSVZ-data geven ook een opsplitsing naar bedrijfstakken. Er moet evenwel rekening gehouden worden met het feit dat:

1. de bedrijfstakopdeling van de RSVZ-data gebaseerd is op een opdeling in beroepen, die niet altijd eenduidig aan één bepaalde bedrijfstak kunnen toegewezen worden;
2. de nationale rekeningen sinds de methodologische herziening van oktober 2005 gebruik maken van de BTW-gegevens voor de bedrijfstakindeling, en niet meer van de RSVZ-statistieken.

Voor gewone zelfstandigen werden vooralsnog de RSVZ-gegevens enkel gebruikt als globale randtotalen per geslacht en leeftijdsklasse, over alle bedrijfstakken heen, en na herschaling op het totale aantal gewone zelfstandigen uit de nationale rekeningen. De kruising met de bedrijfstakken gebeurt aan de hand van een koppeling met de EAK-gegevens (cf. punt 2.2).

Voor de vergoede helpers, een statuut dat in de EAK niet afgezonderd bestaat, werden de RSVZ-gegevens wel gebruikt om de verdeling naar geslacht en leeftijdsklasse per bedrijfstak te ramen (zie punt 2.1.3).

2.1.3. De berekening voor vergoede helpers

De EAK laat niet toe om vergoede helpers af te zonderen. Vanuit de nationale rekeningen zijn de randtotalen per bedrijfstak gekend. De verdeling van de vergoede helpers naar geslacht en leeftijdsklasse per bedrijfstak wordt geraamd op basis van RSVZ-gegevens (helpers in hoofdberoep). De jaarlijkse RSVZ-publicaties laten toe een onderscheid te maken tussen zes grote bedrijfstakken: landbouw, visserij, nijverheid, handel, vrije beroepen en diensten. De verdere detaillering van die zes bedrijfstakken naar IOKLEMS-bedrijfstakken is louter proportioneel.

Het aantal helpers in hoofdberoep in de RSVZ-bestanden vertoont een breuk in 2003 (toename van 42 400 in 2002 naar 94 300 in 2003). De reden voor die breuk is het (integraal) opnemen van de meewerkende partners van zelfstandigen (voornamelijk vrouwen) in de RSVZ-statistiek. Dankzij afzonderlijke RSVZ-cijfers over de meewerkende partners vanaf 2002, kon de impact van die breuk op de verdeling naar geslacht en leeftijdsklasse geneutraliseerd worden.

2.1.4. De randtotalen voor niet-vergoede helpers

De enquête naar de arbeidskrachten laat toe een onderscheid te maken tussen niet-vergoede helpers en overige zelfstandigen. De groep niet-vergoede helpers is echter vrij klein (zie verder in figuur 3), waardoor er grote jaarlijkse schommelingen optreden in hun verdeling naar leeftijdsklasse en geslacht.

Om die jaarlijkse schommelingen te beperken werd de globale verdeling van niet-vergoede helpers over leeftijdsklassen en geslacht uit de EAK afgevlakt (aan de hand van een Hodrick-Prescott filter). De verdeling over leeftijd en geslacht per bedrijfstak, die beschreven wordt in punt 2.2, wordt nadien door middel van de RAS-techniek aangepast aan die randtotalen.

2.1.5. De cijfers voor onthaalouders

Onthaalouders maken deel uit van IOKLEMS-bedrijfstak 85C1 en worden in de nationale rekeningen als zelfstandigen beschouwd. Het aandeel van deze groep in bedrijfstak 85C1 varieert van 70,5 % in 1999 tot 77,5 % in 2004.

Voor de onthaalouders zijn afzonderlijke data beschikbaar per geslacht en leeftijdsklasse. Gegeven de specificiteit van deze groep (meer dan 99 % zijn vrouwen) werden deze gegevens rechtstreeks gebruikt bij de raming van geslacht en leeftijdsklasse voor bedrijfstak 85C1. De verdeling naar geslacht en leeftijdsklasse voor het geheel van bedrijfstak 85 blijft echter gebaseerd op de EAK-gegevens.

2.2. Raming van geslacht en leeftijdsstructuur per bedrijfstak voor gewone zelfstandigen en niet-vergoede helpers

2.2.1. Algemeen

Het is de bedoeling om per gedetailleerde bedrijfstak het geslacht en 11 leeftijdsklassen te onderscheiden, voor zowel gewone zelfstandigen als niet-vergoede helpers. Dat gebeurt op basis van door het NIS verstrekte microgegevens uit de enquête naar de arbeidskrachten (EAK) voor de periode 1999-2005.

Het jaarlijkse aantal zelfstandigen dat in de EAK bevraagd wordt, schommelt tussen 6 200 en 7 000 in de periode 1999-2005. Dat is te weinig om 11 leeftijdsklassen te onderscheiden op basis van gemiddelden: er duiken dan teveel cellen op met onvoldoende observaties in de EAK.

Daarom wordt gebruik gemaakt van regressietechnieken. In een eerste reeks van regressies worden de 11 leeftijdsklassen geraamd gegeven het geslacht, daarna wordt de verdeling over mannen en vrouwen geschat.

Punten 2 en 3 beschrijven de gehanteerde regressietechniek (ordered logit regressie) en de organisatie van de regressies (nested logit regressie). Punt 4 beschrijft de verklarende variabelen in de regressies voor de leeftijdsklasse. Punt 5 beschrijft de raming van het geslacht. In punt 6 wordt de weging van de observaties besproken.

2.2.2. De ordered logit regressietechniek

De ordered logit techniek houdt er rekening mee dat de afhankelijke variabele een beperkt aantal (k) geordende waarden kan aannemen. De kans op leeftijdsklasse i (met $1 < i < k$) gegeven de vector van verklarende variabelen x , wordt gegeven door de onderstaande formule:

$$\Pr(Y = i|x) = F(\alpha_i + \beta'x) - F(\alpha_{i-1} + \beta'x) \quad (1)$$

Daarbij is $F(\cdot)$ een te bepalen cumulatieve verdelingsfunctie, is β de vector van geschatte coëfficiënten en is α_i de i -de interceptterm in de regressie. Voor de kans op de (alfabetisch of numeriek) laagste leeftijdsklasse valt de term achter het minteken weg.

De kans op de hoogste leeftijdsklasse wordt gegeven door:

$$\Pr(Y = k|x) = 1 - F(\alpha_{k-1} + \beta' x) \quad (2)$$

De logistische verdelingsfunctie gebruikt in de regressies wordt gegeven door:

$$F(z) = \frac{\exp(z)}{1 + \exp(z)} \quad (3)$$

Belangrijk is dat de variabelen in x elk slechts via één coëfficiënt een invloed uitoefenen op de waarde van de afhankelijke variabele (hier de leeftijdsklasse), ongeacht het betreffende niveau. Er is dus geen onafhankelijke schatting per niveau, zoals in een multinomiaal model, maar in de plaats daarvan een opgelegde ordening.

Buiten de intercepttermen zijn er geen coëfficiënten die zouden kunnen toelaten dat een variabele een verschillend effect heeft op de kans op elke leeftijdsklasse. Deze veronderstelling wordt ook de “equal slope assumption” of “proportional odds assumption” genoemd, en kan getoetst worden.

Zij is problematisch als men via deze techniek voor de 11 leeftijdsklassen tegelijk wil schatten. In dergelijke regressies wordt de proportional odds veronderstelling steeds met hoge χ^2 waarden verworpen. Een andere bedrijfstak blijkt niet alleen voor oudere of jongere zelfstandigen te zorgen, maar soms ook voor een combinatie van de twee: zowel meer -30 als meer 50+ jarigen...

Om dit probleem op te vangen werd de leeftijdsklasse bij zelfstandigen geschat via een boomstructuur, bestaande uit 5 regressies. Die wordt uitgelegd in de volgende sectie.

2.2.3. Nested logit regressie

Een nested logit regressie is in feite een ordered logit in meerdere stappen. De onderstaande figuur toont de stappen die gevolgd werden bij de nested logit regressies voor de leeftijdsklasse (dus gegeven het geslacht en het statuut).

In een eerste stap wordt het belangrijkste onderscheid bepaald. Dat is het verschil tussen 50+ en -50 jarigen. Aangezien er slechts twee keuzes zijn, kan hierbij gebruik gemaakt worden van een binaire logit. Alle observaties worden opgenomen in de regressie van stap 1.

Alle andere regressies zijn voorwaardelijk en worden gedaan met een deel van de observaties. De volgende regressie maakt het onderscheid tussen -30 en +30 jarigen, gegeven dat de persoon minder dan 50 jaar is. Uit de eigenschappen van voorwaardelijke verdelingen volgt:

$$\Pr(Y \in [15,29]) = \Pr(Y \in [15,29] | Y \in [15,49]) \bullet \Pr(Y \in [15,49]) \quad (4)$$

Door de kans op een -50 jarige gegenereerd door de eerste regressie te vermenigvuldigen met de voorwaardelijke kans op een -30 jarige die volgt uit de tweede regressie verkrijgt men de onvoorwaardelijke of absolute kans op een -30 jarige.

Die aanpak laat toe de particulariteiten van de leeftjidsverdeling voor alle in de regressies opgenomen variabelen weer te geven. Zijn er bij vrouwen, of in de bouwsector..., relatief veel -30 jarigen en toch evenveel 50+ jarigen dan elders, dan kunnen de twee regressies dat oppikken.

Het nadeel is dat er, in vergelijking tot bij ordered logit waar alles met één regressie kan, evenveel regressies nodig zijn als men flexibiliteit wil inbouwen. Hoe verder men gaat in de boomstructuur, hoe kleiner ook het aantal observaties, waardoor er minder ruimte overblijft voor een zinvolle bedrijfstakopdeling.

Om die reden werd slechts afzonderlijk geschat voor de drie grote leeftjidscategorieën (15-29 / 30-49 / 50+). Daarvoor volstaan twee regressies. Binnen elke grote leeftjidscategorie wordt het verder onderscheid in gedetailleerde leeftjidsklassen gegenereerd door één ordered logit regressie. Die laatste techniek respecteert de EAK-verdeling over de (gedetailleerde) leeftjidsklassen wel voor het totaal, maar niet per bedrijfstak, geslacht, substatuut of jaartal.

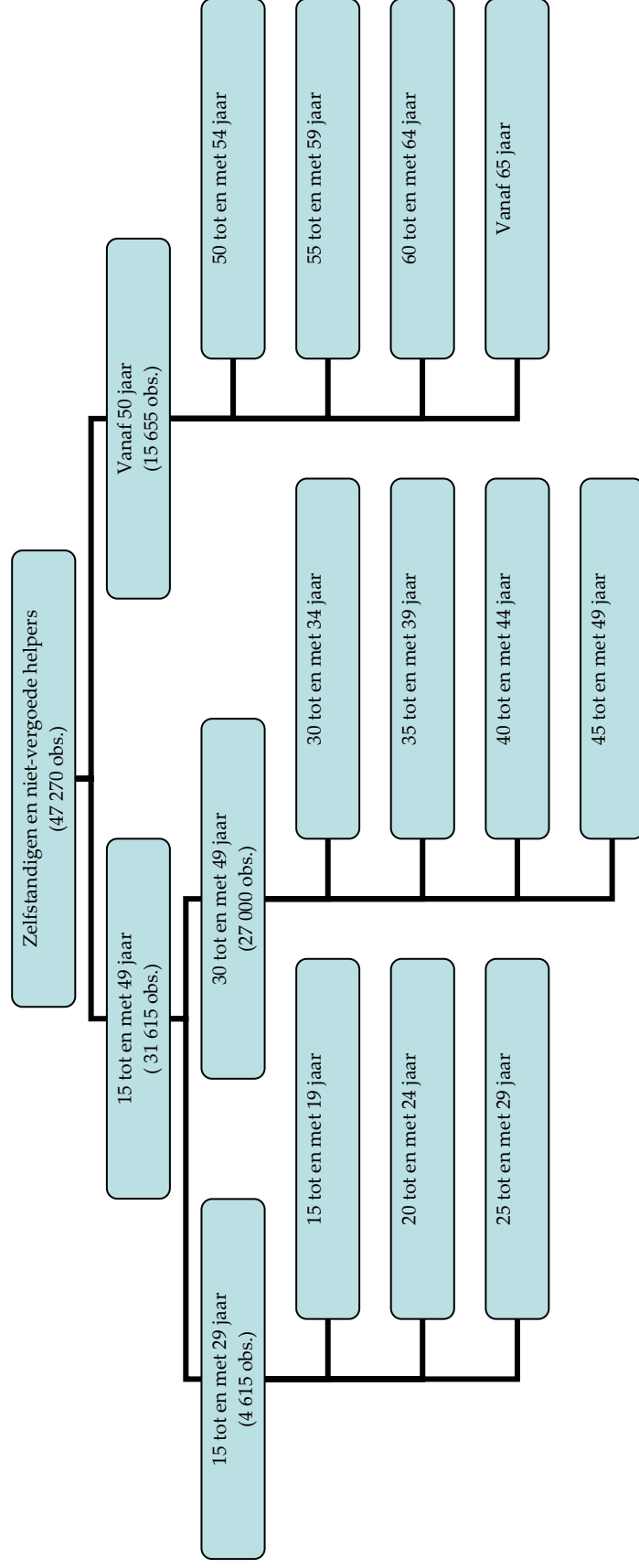
De kansverdeling in de boomstructuur van figuur 1 wordt dus geschat door middel van 5 regressies, waarvan 3 ordered logits en 2 binaire logits.

De nested logit techniek is nog wat complexer dan hierboven beschreven, omdat ook expliciet rekening gehouden werd met de mogelijkheid dat de kansverdelingen binnenin een tak van de in figuur 1 getoonde boomstructuur informatie geeft over de dikte van de tak zélf. Zo kan de aanwezigheid van veel -30 jarigen in een bedrijfstak gepaard gaan met een relatief grote aanwezigheid van -20 of -25 jarigen binnen de groep van -30 jarigen.

Om hiermee rekening te houden werd de aanpak van Greene (2003)¹¹ gevolgd. Daarbij wordt de volgorde van schatten omgedraaid. Eerst worden de meest gedetailleerde voorwaardelijke regressies uitgevoerd.

¹¹ William H. Greene (2003), "Econometric Analysis", Prentice Hall.

Figuur 2 Boomstructuur regressies zelfstandigen en niet-vergoede helpers



Bron: FPB.

Naast een kansverdeling wordt voor elke gedetailleerde regressie per observatie ook een “inclusive value” berekend. Die inclusive value is een transformatie van het vectorproduct $\beta'x$ die als regressor opgenomen wordt in de regressie op een hoger niveau. In navolging van Greene (2003) gebruiken wij de volgende specificatie voor de inclusive value (hier van de -30 jarigen):

$$I_{-30} = \ln(1 + e^{x'\beta_{-30}}) \quad (5)$$

Het vectorproduct $x'\beta$ is altijd positief verbonden met het aandeel van de laagste leeftijdsklassen in de betreffende regressie. De inclusive value van de groep van -30 jarigen, waarvoor de formule hierboven gegeven is, werd berekend op basis van de ordered logit regressie voor de groep van -30 jarigen, en dan opgenomen in de regressie één trap hoger, waar het onderscheid gemaakt wordt tussen 15-29 en 30-49 jarigen. In die regressie werd ook een inclusive value I_{30-49} opgenomen.

De coëfficiënt van I_{-30} bleek significant positief, wat impliceert dat in observaties (= combinaties van bedrijfstakken, geslacht, jaar en substatuut) waarbij de -30 jarigen relatief jong zijn, de kans om een -30 jarige aan te treffen binnen de groep -50 jarigen ook hoger is.

2.2.4. Verklarende variabelen in de regressie voor het bepalen van de leeftijdsklasse

Zoals reeds gezegd, wordt de verdeling over 11 leeftijdsklassen geraamd gegeven het geslacht. Hieronder worden de variabelen die in die regressie voorkomen beschreven. Deze variabelen komen ook terug in de regressie voor de verdeling over het geslacht.

Het effect van het geslacht en het statuut

Het aantal observaties is te klein om een regressie te laten lopen per geslacht of per statuut. In elk van de 5 hierboven beschreven regressies werden daarom 3 wissels opgenomen die toelaten het onderscheid te maken tussen vrouwelijke niet-vergoede helpers, mannelijke niet-vergoede helpers, vrouwelijke gewone zelfstandigen en de basisgroep van mannelijke gewone zelfstandigen.

Het effect van de bedrijfstak

Het effect van de bedrijfstak op de leeftijdsstructuur wordt in elke regressie opgevangen door een klassevariabele (= een reeks wissels). Afhankelijk van het aantal observaties werd een onderscheid gemaakt tussen 31 of 20 bedrijfstakken¹². In de regressie die het onderscheid maakt tussen 50+ en -50 jarigen en in de ordered logit regressie voor 50+ jarigen, werden 31 bedrijfstakken opgenomen. In de andere regressies werd gewerkt met 20 bedrijfstakken.

¹² Beide opdelingen zijn een aggregatie van de 41 bedrijfstakken of subsecties waarnaar later gerefereerd wordt en respecteren de *minimum labour quality industry division* binnen het EUKLEMS-project (zie bijlage 2).

De beheerders van vennootschappen

Voor de zelfstandigen die beheerders van vennootschappen zijn is een speciale behandeling nodig. In de nationale rekeningen wordt deze groep toegewezen aan bedrijfstak NACE-74. In de EAK zitten beheerders in de bedrijfstak die overeenstemt met de hoofdactiviteit van hun onderneming.

In de EAK-gegevens van de periode 1999-2004 kunnen beheerders van vennootschappen onderscheiden worden van niet-beheerders. De verdeling van de beheerders van vennootschappen is verschillend van die van de andere zelfstandigen, zowel naar bedrijfstakken als naar leeftijdsklassen en geslacht (zie verder). Tegelijk blijven beheerders onderling verschillen inzake de verdeling naar geslacht en leeftijdsklasse in functie van de bedrijfstak waarin de activiteit uitgeoefend wordt.

Men zou alle beheerders onmiddellijk kunnen overhevelen naar bedrijfstak NACE-74, maar dan kan een groot aantal observaties niet meer gebruikt worden om het verschil in leeftijd en geslacht tussen bedrijfstakken te bepalen. Daarom werd er voor gekozen beheerders in hun oorspronkelijke bedrijfstak te laten. Het feit dat zij beheerder zijn werd enkel aangegeven door een wissel.

Nadat de regressies uitgevoerd zijn, worden de beheerders van vennootschappen afgezonderd in de extra bedrijfstak voor beheerders van vennootschappen (tak K74v, zie bijlage 2), waarvoor totalen beschikbaar zijn uit de nationale rekeningen.

Vanaf het jaar 2005 wordt geen onderscheid meer gemaakt tussen zelfstandige beheerders en niet-beheerders in de enquête naar de arbeidskrachten. Omdat de regressie de hele periode 1999-2005 omvat, konden de regressieresultaten toch gebruikt worden om een raming te maken voor de verdeling naar geslacht en leeftijd van de niet-beheerders in 2005. Daartoe werd van de wissel niet-beheerder/beheerder een klassevariabele gemaakt die in 2005 niet de waarde nul of één, zoals in de andere jaren, maar een (willekeurige) derde waarde aanneemt¹³. Voor beheerders kan deze methode niet gebruikt worden, omdat niet bekend is in welke bedrijfstakken de beheerders zich bevonden in 2005. Hun verdeling wordt geëxtrapoleerd op basis van die van 2004.

Het effect van de tijd

De evolutie over de tijd wordt opgevangen door het jaartal op te nemen als een continue variabele (met name de natuurlijke logaritme van het jaartal verminderd met 1998). Daarnaast werd ook de groeivoet van het aantal zelfstandigen op het niveau van 41 bedrijfstakken (subsecties genoemd) in de beschouwde periode en een gekruiste variabele van het jaartal en de groeivoet van het aantal zelfstandigen per subsectie opgenomen in de regressies.

¹³ De resultaten voor de niet-beheerders in 2005 zijn die voor 2005 waarbij de klassevariabele op nul gezet is.

De eenvoudige groeivoet vangt overblijvende verschillen in gemiddelde leeftijd op het niveau van de 41 bedrijfstakken op indien zij gerelateerd zijn met de groei. Enkel in de regressie voor het onderscheid tussen -50 en 50+ jarigen, had deze variabele een significant effect. Subsecties met een hogere groei hadden een significant hoger aandeel van -50 jarigen.

De gekruiste variabele voor het jaartal en de groeivoet laat toe na te gaan of een grotere groei van het aantal zelfstandigen een snellere verjonging met zich mee brengt. Zo bleek de kans om een -30 jarige aan te treffen binnen de groep 15-49 jarigen sneller toe te nemen in snel groeiende bedrijfstakken. Tegelijk blijkt in die periode ook het aandeel 50+ jarige zelfstandigen trager af te nemen in bedrijfstakken met een hogere groei.

Een andere variabele laat toe na te gaan of de groep niet-vergoede helpers sneller verjongt/veroudert dan de gewone zelfstandigen. In de periode 1999-2005 nam het aandeel -30 jarigen binnen de groep -50 jarigen af bij niet-vergoede helpers, niet bij de gewone zelfstandigen.

2.2.5. Raming van de verdeling naar geslacht op basis van EAK-gegevens

Ook de verdeling naar geslacht wordt geschat via een regressieanalyse. Omdat er maar twee mogelijkheden zijn, wordt geschat met een binaire logit. De gezamenlijke verdeling van geslacht én leeftijdsklasse wordt verkregen als het product tussen beide kansen. Zo wordt de kans op een vrouw van leeftijdsklasse 25-29 gegeven door de kans op leeftijdsklasse 25-29 gegeven dat het om een vrouw gaat te vermenigvuldigen met de kans op een vrouw.

In de regressie voor het geslacht werd de leeftijdsklasse niet opgenomen. Het aandeel vrouwen per leeftijdsklasse van de EAK wordt door de gecombineerde kansen enkel gerespecteerd voor alle types zelfstandigen en alle jaren in de regressie tezamen.

Binnen een kleinere groep, zoals de niet-vergoede helpers jonger dan 30 jaar, kunnen zo afwijkingen met de EAK-observaties ontstaan. Die afwijkingen kunnen zowel te maken hebben met het beschouwde jaartal¹⁴, als met de structuur opgelegd in de regressies. Dat wordt geïllustreerd in tabel 3.

De hoofdbedoeling van de regressies is echter het ramen van de verschillen in de verdeling naar geslacht en leeftijdsklasse tussen bedrijfstakken. Voor de totale verdeling naar geslacht en leeftijdsklasse van gewone zelfstandigen respecteert de SAM immers de RSVZ-randtotalen; voor de randtotalen naar geslacht en leeftijdsklasse van niet-vergoede helpers werd gebruik gemaakt van een op de EAK gebaseerde afgevlakte tijdreeks (cf. supra). Die randtotalen worden weer gegeven in de laatste kolom van tabel 3.

¹⁴ Het aandeel vrouwen bij de niet-vergoede helpers jonger dan 30 jaar was 69 % in de EAK van 1999, 48 % in de EAK van 2000 en 65 % in die van 2001. Dergelijke schommelingen zijn normaal, gezien de kleine omvang van de groep. Toch wordt het aandeel vrouwen ook systematisch overschat bij niet-vergoede helpers jonger dan 30 door de opgelegde structuur. Dit wordt pas rechtgetrokken bij de aanpassing aan de randtotalen in de SAM.

Tabel 3 Aandeel vrouwen per leeftijdsklasse in de EAK, de regressies en de SAM (2000)

	Aantal EAK-observaties	EAK ongewogen	EAK gewogen	Regressies	SAM randtotalen
Gewone zelfstandigen, niet-beheerders van vennootschappen					
15-29	441	39%	39%	35%	41%
30-49	2476	34%	33%	33%	36%
50+	1405	28%	26%	32%	29%
Gewone zelfstandigen, beheerders van vennootschappen					
15-29	126	29%	27%	22%	25%
30-49	915	22%	21%	21%	23%
50+	412	18%	18%	20%	22%
Niet-vergoede helpers					
15-29	87	48%	52%	84%	64%
30-49	424	91%	90%	85%	92%
50+	352	84%	84%	86%	90%
Totalen					
Niet-beheerders	4322	32%	32%	32%	34%
Beheerders	1453	22%	21%	21%	22%
Niet-vergoede helpers	863	84%	84%	85%	88%

Bron: Berekeningen FPB op basis van EAK-microdata en RSVZ-gegevens.

De in tabel 3 getoonde resultaten voor beheerders en niet-beheerders van vennootschappen zijn een uitsplitsing van de RSVZ-randtotalen voor zelfstandigen op basis van de regressieresultaten. Het verschil tussen beheerders en niet-beheerders is belangrijk. Uit de tabel blijkt duidelijk dat beheerders van vennootschappen in mindere mate vrouwen zijn.

2.2.6. De weging van de observaties

De NIS-databank bevat voor elke observatie ook een gewicht. De weging naar leeftijdsklasse, woonplaats en geslacht wordt gebruikt bij de extrapolatie naar bevolkingstotalen.

In de regressies werd ervoor gekozen om niet-gewogen observaties te gebruiken. Daardoor krijgt elke zelfstandige of vergoede helper in de EAK hetzelfde gewicht bij het bepalen van de coëfficiënten van de effecten van de verklarende variabelen.

De eigenlijke weging in de SAM gebeurt op het moment dat de resultaten aangepast worden aan de nationale rekeningen en de RSVZ-randtotalen (voor leeftijdsklassen en geslacht).

2.3. Resultaten voor verdeling zelfstandigen per bedrijfstak

De kansverdeling op een niveau van 41 bedrijfstakken die volgde uit de EAK-regressies werd toegepast op de totalen van de bedrijfstakken in de nationale rekeningen. Daardoor stemmen de totalen per bedrijfstak automatisch overeen met die in de nationale rekeningen. Er werd een verdere detaillering uitgevoerd tot op het IOKLEMS-niveau, maar die is puur proportioneel.

De globale verdeling naar geslacht en leeftijdsklasse moet echter ook overeenstemmen met de eerder bepaalde globale randtotalen voor zelfstandigen en niet-vergoede helpers. De verschillen

tussen de (bottom up) verkregen globale verdeling over geslacht en leeftijdsklassen en de randtotalen, werden weggewerkt met de RAS-equilibreringstechniek.

3. Verdere opdeling van werknemers en zelfstandigen naar opleidingsniveau

Dit hoofdstuk beschrijft de verdere opdeling van alle werkenden per bedrijfstak in zes opleidingsniveaus: lager, lager secundair, hoger secundair, hoger niet-universitair van het korte en lange type en universitair onderwijs. De opdeling gebeurt op basis van regressietechnieken op de resultaten van de enquête naar de arbeidskrachten voor de periode 1999-2005.

Deze verdeling wordt verbonden met de resultaten over de hoofdelijke werkgelegenheid van werknemers en zelfstandigen per bedrijfstak, geslacht en leeftijdsklasse in de SAM die in hoofdstuk 1 en 2 besproken werden. Daardoor ontstaat een tijdreeks met gegevens over het opleidingsniveau die consistent is met de nationale rekeningen.

3.1. De bronnen voor een opdeling naar opleidingsniveau

Voor de opdeling naar opleidingsniveau van werkenden per bedrijfstak bestaan er in België geen administratieve bestanden. De opdeling is noodzakelijkerwijze gebaseerd op enquêtegegevens. Drie bronnen komen hiervoor in aanmerking.

De eerste is de socio-economische enquête, beter bekend als de volkstelling. Die heeft het voordeel dat ze exhaustief is. Handicaps zijn dat er slechts om de 10 jaar een volkstelling georganiseerd wordt (de laatste in 2001), en dat de verwerking van de gegevens geruime tijd in beslag neemt.

De tweede bron is de enquête naar de arbeidskrachten (EAK). Die levert, in tegenstelling tot de volkstelling, jaarlijkse gegevens over het opleidingsniveau op sinds 1985, waardoor het mogelijk is de evolutie te volgen. Een beperking is dat in de EAK slechts 1 % van de beroepsbevolking vertegenwoordigd is.

De derde bron is de enquête naar de structuur en de verdeling van de lonen. Die enquête is momenteel beschikbaar voor de jaren 1999-2004¹⁵ en bevat informatie over meer dan 145 000 werknemers. Ze wordt afgenomen bij 6 000 werkgevers uit de privé-sector met minstens 10 werknemers. De enquête bevat geen cijfers voor zelfstandigen, noch voor werknemers uit de NACE-secties A, B, L, M, N, O en P. Ook door het feit dat enkel werkgevers met minstens 10 werknemers ondervraagd worden, beperkt de representativiteit van deze enquête. De enquête heeft wel het voordeel dat er ook gegevens over de brutolonen opgevraagd werden. Dat is erg nuttig als ook de band tussen opleidingsniveau en lonen moet gelegd worden.

¹⁵ Sedert 1999 is de enquête verplicht. Ook voor vroegere jaren (1966, 1972, 1974, 1978 en 1995) werden bevestigingen gedaan, doch voor een nog beperkter waarnemingsgebied. De vergelijking met de enquêteresultaten van die vroegere jaren is dan ook vrij problematisch.

Van deze drie bronnen wordt de enquête naar de arbeidskrachten verkozen. Dat omdat zij jaarlijks gegevens oplevert en geen bedrijfstakken of categorieën uitsluit. Het probleem van de steekproefgrootte wordt aangepakt door jaartallen te combineren en gebruik te maken van regressietechnieken.

In wat volgt wordt beschreven hoe op basis van de EAK-gegevens een tijdreeks aangemaakt werd met jaarlijkse opleidingsgegevens per bedrijfstak voor de periode 1999-2005.

3.2. Motivering van de keuze voor regressietechnieken

Als men een tijdreeks wil maken op basis van enquêtegegevens over slechts 1 % van de beroepsbevolking, zijn twee benaderingen mogelijk. Ofwel beperkt men het aantal en het detail van de dimensies (bedrijfstakken, leeftijdsklassen,...) in sterke mate en werkt men met eenvoudige gemiddelden; ofwel behoudt men een groot aantal en detail van dimensies, maar legt men structuur op via een aangepaste regressietechniek.

Werken met eenvoudige steekproefgemiddelden heeft het nadeel dat het detail van de bedrijfstakken en het aantal bijkomende dimensies sterk beperkt wordt, waardoor de resultaten minder bruikbaar zijn voor analyse en het detail van de opdeling van loontrekkenden en zelfstandigen uit hoofdstuk 1 en 2 niet volledig gebruikt wordt.

Dat laatste zou jammer zijn, omdat de opdeling naar leeftijdsklasse, geslacht, statuut en arbeidsregime heel wat houvast kan bieden bij de raming van de verdeling over opleidingsniveaus tot op een groot bedrijfstakdetail.

Als men deze verdeling van de werkenden naar statuut, geslacht, leeftijdsklasse en arbeidsregime per bedrijfstak als anker wil gebruiken, dan moet het opleidingsniveau voor alle mogelijke combinaties berekend worden.

Het product van de hierboven voorziene kruisingen per bedrijfstak is niet minder dan $2\ 156 = 7 \times 11 \times 2 \times (3 \times 4 + 2)$, te weten 7 jaren, 11 leeftijdsklassen, 2 geslachten, 4 statuten voor werknemers (arbeiders en bedienden in de privé-sector, contractuelen en ambtenaren in de publieke sector) waarvoor elk 3 arbeidsregimes (voltijds, deeltijds, uitzendkracht) gelden en 2 types zelfstandigen (gewone zelfstandigen en niet-vergoede helpers).

Per bedrijfstak zijn er dus tot 2 156 combinaties of cellen waarvoor een opdeling in opleidingsniveaus moet gegenereerd worden. Zelfs voor een bedrijfstak met een groot aantal werkenden, zoals de kleinhandel, met meer dan 26 700 EAK-respondenten in de periode 1999-2005, is dat veel. Gemiddeld telt een cel dan 12 observaties, wat veel te weinig is voor een opdeling in 6 opleidingsniveaus. Bovendien zijn er combinaties die wel voorkomen in de SAM, maar niet in de EAK.

Regressieanalyse biedt een oplossing voor zowel ontbrekende als weinig frequente combinaties in de EAK, omdat op basis van de coëfficiënten en de verklarende variabelen een verdeling kan gegenereerd worden voor alle mogelijke cellen. De enige voorwaarde is dat de verklarende variabelen in de regressies ook voorkomen in de gegevens over het aantal werknemers en zelfstandigen.

Zo wordt een bestand gegenereerd dat alle kruisingen in de gegevens over de hoofden behoudt. De standaardfouten van de coëfficiënten bieden tegelijk een idee van de statistische kwaliteit van de resultaten.

Bij regressieanalyse wordt er echter ook structuur opgelegd aan de gegevens. Het is noodzakelijk te controleren of de regressieresultaten voldoende overeenstemmen met de EAK. Uit een eerste oefening, waarbij geschat werd met een ordered logit model, bleek het samennemen van verschillende jaren en een 40-tal bedrijfstakken in één regressie geen problemen te stellen. Er waren echter te grote verschillen met de EAK-gemiddelden bij jongere (-30) en oudere (50+) werknemers.

Als antwoord daarop werd een nested logit regressiestructuur aangewend. Die bestaat uit enkele al dan niet voorwaardelijke logit regressies die na elkaar worden uitgevoerd. Wij bespreken eerst deze regressietechniek. Daarna wordt in detail ingegaan op de variabelen in de regressies en de behandeling van enkele verschillen tussen de EAK en de nationale rekeningen.

3.3. De gehanteerde regressietechniek

Eerst wordt de gehanteerde nested logit structuur besproken. Daarna wordt dieper ingegaan op de behandeling van de cellen zonder EAK-observaties.

3.3.1. De organisatie van de regressies (nested logit structuur)

Net zoals dat het geval was voor de leeftijdsklassen bij zelfstandigen, stelt er zich een probleem als men de werkenden wil opdelen in 6 opleidingsniveaus gebruik makend van één ordered logit regressie. Zo blijkt een andere leeftijdsklasse of bedrijfstak vaak niet alleen voor een hoger of lager opleidingsniveau te zorgen, maar ook voor een andere vorm van de verdeling van de opleidingsniveaus (met name meer gemiddeld geschoolden bij -30 jarigen dan in de andere leeftijdsgroepen).

In plaats van de verdeling over zes opleidingsniveaus in één stap te schatten, wordt dat nu gedaan in twee stappen. In de eerste stap wordt het belangrijkste onderscheid geschat, gebruik makend van alle observaties. Wat dat belangrijkste onderscheid is, verschilt in functie van het statuut.

In de periode 1999-2005 is ongeveer 50 % van de arbeiders in de privé-sector ten hoogste lager secundair geschoold en minder dan 5 % hoger geschoold. Daarom is bij arbeiders het onder-

scheid tussen lager versus hoger secundair onderwijs het belangrijkste. Bij bedienden in de privé-sector, werknemers in de publieke sector en zelfstandigen is, afhankelijk van het geslacht, 40 tot 60 % hoger geschoold. In al deze groepen wordt het onderscheid tussen secundair en hoger geschoolden het belangrijkste geacht.

Aangezien er in stap 1 slechts twee mogelijkheden zijn, kan de schatting gebeuren met een binaire logit. In de tweede stap gebeurt één afzonderlijke regressie voor de verdere onderverdeling van de lager opgeleiden en één voor die van de hoger opgeleiden.

Zo wordt voor arbeiders de kans op ten hoogste lager onderwijs, gegeven dat het diploma ten hoogste lager secundair is, geschat. Deze regressie levert een voorwaardelijke kans op. Door die te vermenigvuldigen met de op niveau 1 berekende kans op een diploma van ten hoogste lager secundair verkrijgt men een raming van de absolute kans op een diploma lager onderwijs. Daarnaast wordt ook een ordered logit regressie uitgevoerd die het onderscheid tussen hoger secundair en de (erg kleine) groepen van hoger niet-universitair en universitair geschoolden maakt.

Bij arbeiders komen er in stap 2 respectievelijk 2 en 4 opleidingsniveaus voor. Voor bedienden, zelfstandigen en werknemers in de publieke sector komen er telkens 3 opleidingsniveaus voor. In deze stap wordt altijd gebruik gemaakt van ordered logit. Zo kan een bijkomende stap, die zou leiden tot regressies met een nog kleiner aantal observaties, vermeden worden¹⁶.

Als de regressies van stap 1 en 2 onafhankelijk uitgevoerd worden, komen de resultaten ervan overeen met die van een multinomiale logit. Er werd echter geen onafhankelijke schatting uitgevoerd, maar een nested logit. Daarbij wordt er van uitgegaan dat de kansverdeling van stap 1 verbonden kan zijn met die van stap 2.

Om dat na te gaan werden eerst de regressies van stap 2 uitgevoerd. Dat laat toe om, naast een voorwaardelijke verdeling, ook een *inclusive value* te berekenen. Dat is een monotone transformatie van het vectorproduct $\beta'x$ die kan opgenomen worden in de regressie in stap 1. Door dat te doen hangt die af van de coëfficiënten geschat in stap 2.

Uit de resultaten bleek dat deze inclusive values vaak een significant effect hebben. Zo blijkt voor bedienden te gelden dat in bedrijfstakken waar het aandeel van universitaire binnen de groep van de hoger geschoolden groot is, het globale aandeel van hoger geschoolden in de totale werkgelegenheid van die bedrijfstak groot is. Anderzijds blijkt een hoog aandeel van lager secundair geschoolde bedienden samen te gaan met een klein globaal aandeel van de groep ten hoogste hoger secundair geschoolde bedienden.

¹⁶ Om bruikbare regressieresultaten te verkrijgen met redelijke standaardfouten op het tweede niveau waren ook enkele andere ingrepen nodig. Zo werd voor arbeiders in de privé-sector, werknemers in de publieke sector en zelfstandigen geen afzonderlijke regressie per geslacht uitgevoerd in de voorwaardelijke regressie voor het onderscheid tussen de hoogste 3 of 4 opleidingsniveaus. In de betreffende regressies werden wissels opgenomen voor het geslacht, gekruist met substatuut en enkele bedrijfstakken. Het totaal aantal regressies bleef daardoor beperkt tot 21.

3.3.2. De behandeling van cellen die ontbreken in de EAK

De regressies worden uitsluitend geschat op de observaties van de EAK. De regressieresultaten kunnen echter niet alleen gebruikt worden voor combinaties van geslacht, statuut, leeftijd en bedrijfstak die in de enquête naar de arbeidskrachten voorkomen, maar voor alle combinaties die voorkomen in de SAM.

Om dat mogelijk te maken werd het bestand met de EAK-resultaten uitgebreid met een bestand waarin alle niet in de EAK maar wel in de SAM voorkomende combinaties van verklarende variabelen éénmaal voorkomen. De logistic procedure berekent dan automatisch ook een kansverdeling voor die cellen op basis van de regressieresultaten.

3.4. De variabelen in de regressies

In secties 1 en 2 worden de keuzes i.v.m. de afhankelijke en de verklarende variabelen toegelicht.

3.4.1. De afhankelijke variabele

Het opleidingsniveau wordt onderverdeeld in 6 niveaus: het lager onderwijs, het lager secundair onderwijs, het hoger secundair en postsecundair onderwijs, het hoger niet-universitair onderwijs van het korte type, het hoger niet-universitair onderwijs van het lange type en het universitair onderwijs.

Het postsecundair niet-hoger onderwijs, dat slaat op het 7de jaar dat volgt op het technisch en beroepsonderwijs en ingericht wordt door het secundair onderwijs, werd toegevoegd bij het hoger secundair onderwijs. Met 585 werkenden in de enquête van 2000 is hun aantal te klein om een betrouwbare afzonderlijke raming van het aandeel postsecundair geschoolden te maken.

Het hoger niet-universitair onderwijs van het lange type werd wel als een afzonderlijk niveau behouden. Met 1 596 werkende respondenten met een dergelijk diploma in 2000, is dat de tweede minst vertegenwoordigde opleidingsgroep in de EAK (na het postsecundair niet-hoger onderwijs). De gebruikte regressietechniek maakt het echter mogelijk dit bijkomend onderscheid zonder efficiëntieverlies aan te maken.

3.4.2. De verklarende variabelen

Het opleidingsniveau in de regressies hangt af van een beperkt aantal variabelen. De lijst is beperkt omdat de variabelen zowel moeten voorkomen in de enquête naar de arbeidskrachten als in de resultaten die volgen uit hoofdstuk 1 en 2. Het gaat om:

1. het statuut
2. het geslacht
3. de leeftijdsklasse

4. de bedrijfstak
5. het arbeidsregime
6. de evolutie over de tijd.

De volgende punten geven kort aan hoe met deze variabelen rekening gehouden werd.

Het statuut

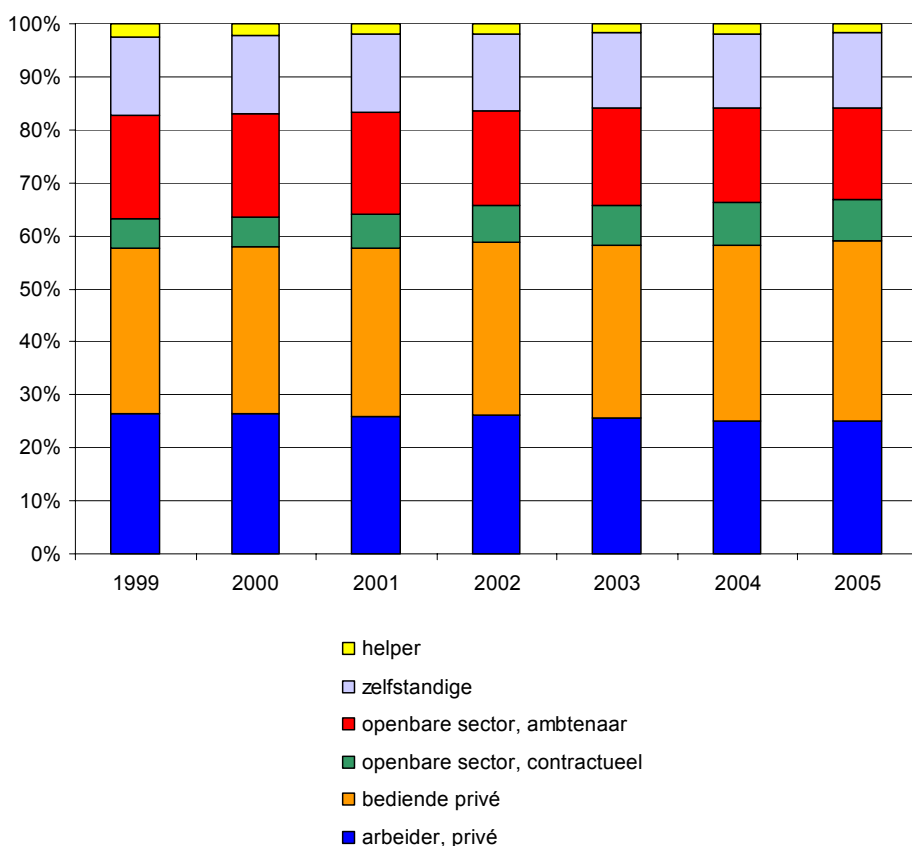
Figuur 3 toont de evolutie van de statuten van de werkenden vertegenwoordigd in de EAK tussen 1999 en 2005.

Er werd beslist om afzonderlijke regressies uit te voeren per statuut. De reden ligt in het feit dat de effecten van bedrijfstak, leeftijd, jaartal... verschillend bleken te zijn tussen arbeiders, bedienden, zelfstandigen en werknemers van overheidsdiensten (zie verder in de bespreking van tabellen 4 en 5).

Terwijl hier zes statuten onderscheiden worden, worden er slechts vier afgescheiden in een eigen regressie. Zoals uit de figuur blijkt, vormen helpers slechts een kleine groep. Daarom worden helpers samen gehouden met gewone zelfstandigen. Wel wordt een wissel opgenomen die aangeeft of het om een helper gaat.

De helpers in figuur 3 omvatten enkel de niet-vergoede helpers. De vergoede helpers (voor een omschrijving zie hoofdstuk 2) zijn in de EAK niet te onderscheiden van de arbeiders of bedienden in de privé-sector. Daarom worden voor de vergoede helpers dezelfde regressieresultaten toegepast als voor arbeiders en bedienden. De vergoede helpers worden daartoe eerst opgesplitst in arbeiders en bedienden conform het aandeel van arbeiders en bedienden in de betreffende cel, gedefinieerd aan de hand van jaartal, bedrijfstak, geslacht en leeftijdsklasse.

Ook voor contractuelen in de publieke sector wordt geen afzonderlijke regressie gedaan. Die groep werd samengenomen met de statutairen. Net zoals bij de helpers wordt via een wissel het specifieke effect van het statuut van contractueel op het opleidingsniveau gemeten. Bovendien wordt een gekruiste variabele van de wissel voor contractuelen met die voor de leeftijd ingevoerd. Daarmee kan een verschil in effect van de leeftijd op het opleidingsniveau tussen contractuelen en statutairen opgevangen worden.

Figuur 3 Evolutie van de statuten tussen 1999 en 2005 in de EAK

Bron: Berekeningen FPB op basis van EAK.

Het geslacht

Naast 4 statuten, worden er ook afzonderlijke regressies uitgevoerd per geslacht. Uit de resultaten blijkt dat de intercepttermen en de effecten van leeftijd, jaartal en de overige verklarende variabelen meestal significant verschillen tussen mannen en vrouwen binnen één statuut.

Het onderscheid naar geslacht brengt het totaal aantal regressies op 8. Dat aantal wordt vermenigvuldigd met 3 bij de toepassing van de nested logit techniek, wat leidt tot 24 regressies. Omdat de data dat aantal regressies niet altijd aankunnen, werden mannen en vrouwen in bepaalde regressies toch samen geschat.

De leeftijdsklasse

De EAK deelt de personen op in leeftijdsklassen van 5 jaar. Vanaf de leeftijd van 15 jaar tot en met 64 jaar zijn er 10 leeftijdsklassen (NIS-klasse 4 tot en met 13). Voor de werkenden vanaf 65 jaar werd één extra leeftijdsklasse gecreëerd (klasse 14). Per regressie zouden dus 11 coëfficiënten kunnen geschat worden. Om het aantal vrijheidsgraden hoog te houden en gekruiste effecten te kunnen onderzoeken werd het effect van de leeftijdsklasse echter samengevat door 5 wissels en één numerieke variabele.

Het verwachte negatieve effect van een hogere leeftijd op het opleidingsniveau wordt opgevangen door een numerieke variabele: de natuurlijke logaritme van de NIS-leeftijdsklasse. Er werden ook wissels voor de leeftijdsklassen 15-19, 20-24, 25-30 en 50-60, 60+ ingebracht, omdat deze wissels vaak significant verschillende coëfficiënten hebben. In de regressie voor werknemers uit de publieke sector werd de numerieke variabele ook gekruist met de wissel voor contractuelen.

De bedrijfstak

De hoofdactiviteit van werknemers en zelfstandigen wordt in de EAK weergegeven aan de hand van de NACE-code op 2 cijfers. Omdat sommige van deze NACE 2-digit bedrijfstakken te weinig vertegenwoordigd zijn in de enquête, werden zij geaggregeerd tot 41 bedrijfstakken, die verder subsecties genoemd worden. In bijlage 2 wordt een lijst van subsecties en de overeenstemmende NACE 2-digit codes gegeven.

Voor de zelfstandigen en de werknemers in de openbare sector werd telkens een meer geaggregeerde opdeling bestaande uit resp. 30 en 17 bedrijfstakken gebruikt om te grote standaardfouten te vermijden. In sommige regressies voor vrouwelijke zelfstandigen werd gewerkt met een opdeling in 20 bedrijfstakken.

Het arbeidsregime

De volgende verklarende variabele is het arbeidsregime. Die variabele wordt alleen gebruikt in de regressies voor werknemers.

Er wordt een onderscheid gemaakt tussen drie soorten arbeidsregime: voltijdse tewerkstelling, deeltijdse tewerkstelling en tewerkstelling als uitzendkracht of PWA. In de EAK zijn er zowel vragen opgenomen die peilen naar het tijdelijk karakter van het uitgevoerde werk, als naar de gepresterde uren tijdens de referentieweek, zodat deze bijkomende opdeling kan toegevoegd worden aan de resultaten voortvloeiend uit hoofdstukken 1 en 2.

In de regressies kan de variabele arbeidsregime dus drie waarden aannemen. De variabele wordt, net zoals de variabele bedrijfstak, behandeld als een klassevariabele. Er zijn bijgevolg twee geschatte coëfficiënten, met de voltijds werkende niet-uitzendkrachten als basisgroep.

De evolutie over de tijd

De regressies hebben slechts betrekking op een periode van 7 jaar. De verschillen tussen jaren blijken duidelijk minder significant dan de verschillen tussen geslacht, leeftijdsgroepen, statuten en bedrijfstakken. Dat resultaat verantwoordt de optie om de gegevens van meerdere jaren samen te nemen om de structurele verschillen preciezer te kunnen bepalen.

Om de data ook bruikbaar te maken als tijdreeks is het echter nodig om ook aandacht te hebben voor de evolutie, ondanks de korte tijdsperiode.

Een eerdere oefening waarbij het jaartal als een klassevariabele werd opgenomen en niet gekruist werd met andere variabelen leidde tot een te grote gevoeligheid aan jaarlijkse schommelingen in de EAK-data. Tegelijk bleek de evolutie van het opleidingsniveau in de SAM in conflict met die in enkele (grotere) bedrijfstakken in de EAK.

Om te grote jaarlijkse schommelingen te vermijden, werd het jaartal als een continue variabele opgenomen: met name de natuurlijke logaritme van het jaartal verminderd met 1998.

Ten einde beter de evolutie per grote leeftijdsklasse weer te geven werd ook een gekruiste variabele jaartal - grote leeftijdsklasse (-30, 30-49, 50+) opgenomen in de regressies. In de regressies in de eerste stap (zie verder) werden bovendien vijf variabelen opgenomen die het jaartal kruisen met vijf groepen van bedrijfstakken.

Zo wordt een eigen evolutie mogelijk gemaakt voor landbouw, visserij, delfstoffen en industrie (A+B+C+D), energie- en bouwsector (E+F), handel, horeca en diensten aan particuliere huishoudens (G+H+P), financiële en zakelijke diensten (J+K) en onderwijs en gezondheidszorg (M+N). De vervoersdiensten, administratie en overige diensten (I+L+O) dienen als basisgroep.

3.5. De behandeling van verschillen tussen de EAK en de nationale rekeningen

3.5.1. De beheerders van vennootschappen

In de EAK worden de zelfstandigen die beheerder zijn van een vennootschap, toegevoerd aan de bedrijfstak van de activiteit die zij uitoefenen. In de nationale rekeningen worden al die beheerders samengebracht in de bedrijfstak NACE-74.

Voor de periode 1999-2004 is het mogelijk in de EAK de beheerders te onderscheiden van de niet-beheerders. In plaats van onmiddellijk alle beheerders toe te wijzen aan bedrijfstak NACE-74, werden zij in hun bedrijfstak van oorsprong gelaten gedurende de regressie. Het feit dat zij beheerder zijn, werd aangegeven via een wissel. Zo blijven de beheerders, die beter opgeleid zijn, maar waarvan het opleidingsniveau ook sterk verschilt in functie van de uitgeoefende activiteit, nuttig bij het bepalen van het onderscheid tussen bedrijfstakken. Tegelijk wordt hun opleidingsniveau gemiddeld juist geraamd. Die aanpak verbeterde de kwaliteit van de

regressies in vergelijking met een situatie waarin geen onderscheid gemaakt werd tussen beheerders en niet-beheerders.

Nadat de regressies uitgevoerd werden, werden de beheerders van vennootschappen afgezonderd in de extra bedrijfstak voor beheerders van vennootschappen (tak K74v, zie bijlage 2), waarvoor totalen beschikbaar zijn in de SAM.

Vanaf 2005 is het niet meer mogelijk om in de EAK het onderscheid te maken tussen beheerders en niet-beheerders. Dit bemoeilijkt de vergelijkbaarheid met de cijfers van de nationale rekeningen.

De regressietechniek biedt een (voorlopige) oplossing, door in het jaar 2005 de wissel voor beheerders een derde waarde (niet nul of één) te laten aannemen. Daardoor kunnen op basis van de regressieresultaten rechtstreeks ramingen voor het opleidingsniveau van niet-beheerders in 2005 gegenereerd worden. De resultaten voor de beheerders in 2005 worden afzonderlijk geraamd op basis van de opleidingsniveaus per leeftijdsklasse en geslacht in 2004.

3.5.2. De uitzendkrachten

Het verschil tussen uitzendkrachten en niet-uitzendkrachten is niet alleen belangrijk omdat het vaak leidt tot significante verschillen in opleidingsniveau¹⁷, maar ook omdat uitzendkrachten in de nationale rekeningen samengebracht worden in bedrijfstak NACE -74. In de EAK daarentegen maken zij meestal gewoon deel uit van de bedrijfstak waarin de arbeid verricht werd.

Dit probleem is gelijkaardig aan dat van de beheerders van vennootschappen en werd op een gelijkaardige wijze aangepakt: in de regressies werden de uitzendkrachten in hun sector van tewerkstelling gelaten; de wissel voor uitzendkrachten laat toe om ze nadien over te brengen naar bedrijfstak NACE-74 (specifieke tak K74e, zie bijlage 2). In tegenstelling tot bij de beheerders van vennootschappen, werd het onderscheid uitzendkrachten/niet-uitzendkrachten wél behouden in de EAK van 2005.

3.5.3. De vergoede helpers

De vergoede helpers (die in 2000 ongeveer 4 % uitmaakten van het totaal aantal zelfstandigen) kunnen niet afgezonderd worden in de EAK. Daarom kregen ze per bedrijfstak, leeftijdsklasse en geslacht een opleidingsniveau toegewezen dat berekend werd als gewogen gemiddelde van het opleidingsniveau van de gewone zelfstandigen en de niet-vergoede helpers.

¹⁷ Bij bedienden en werknemers uit de publieke sector zijn uitzendkrachten minder vaak hoog geschoold, bij mannelijke arbeiders geldt het tegenovergestelde: uitzendkrachten zijn er vaker hoger geschoold.

3.5.4. Grensarbeid

De hier ontwikkelde methodologie heeft tot doel een opdeling te maken van de *binnenlandse* werkgelegenheid volgens de nationale rekeningen. Daarom werden uitgaande grensarbeiders uitgesloten uit de EAK-observaties. De binnenkomende grensarbeid (van niet-ingezetenen in België) is niet gedekt door de EAK, maar maakt wel deel uit van de binnenlandse werkgelegenheid volgens de nationale rekeningen. Dit blijft dus een verschil waarvoor onmogelijk kon gecorrigeerd worden.

3.6. De resultaten van de SAM

De verdeling over 6 opleidingsniveaus in de SAM wordt verkregen door de gegeneerde kansverdelingen per combinatie van geslacht, leeftijdsklasse, statuut, bedrijfstak en arbeidsregime te verbinden met de resultaten uit hoofdstukken 1 en 2.

In dit deel worden de resultaten van de SAM besproken door ze te vergelijken met zowel gewogen als ongewogen EAK-gemiddelden voor de periode 1999-2005. Er is aandacht voor twee oorzaken van verschillen. Verschillen die het gevolg zijn van de opgelegde “smoothing” en verschillen die veroorzaakt zijn door een verschillende weging van cellen in de SAM en de EAK.

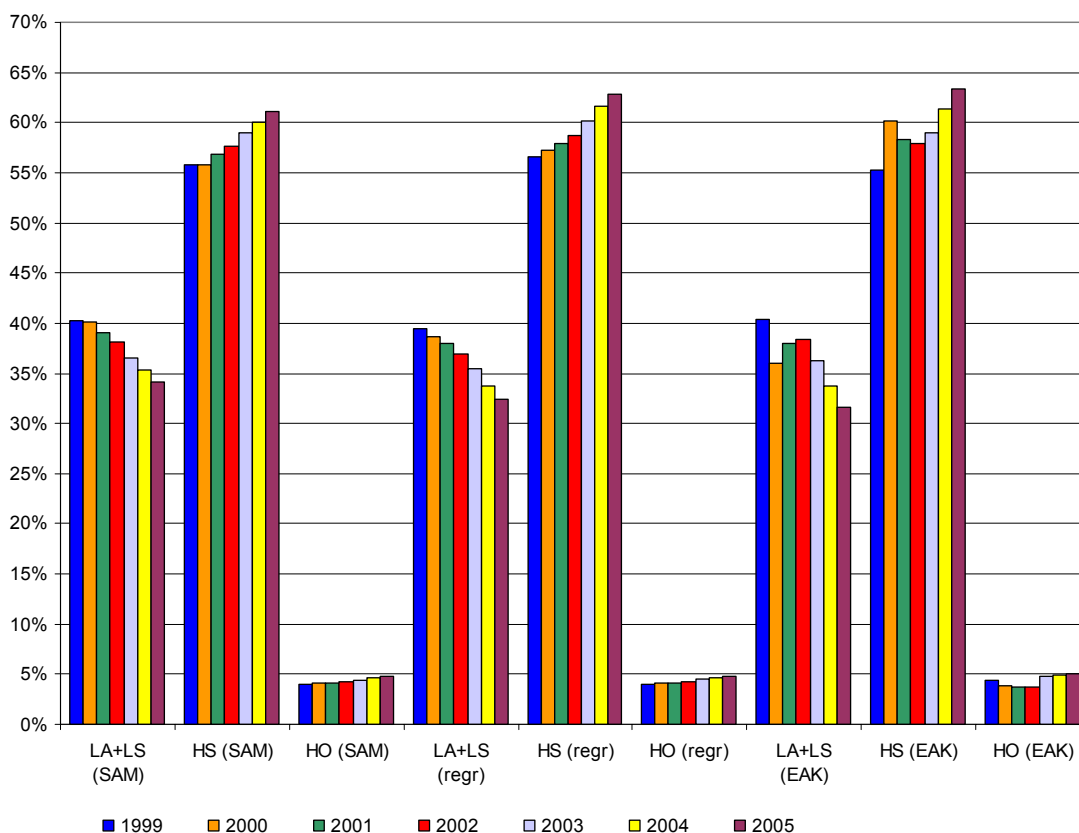
3.6.1. Verschillen veroorzaakt door de opname van het jaartal als een continue variabele

Omdat het jaartal niet als een klassevariabele werd opgenomen, maar als een continue variabele, komt de gemiddelde kansverdeling uit een regressie voor een gegeven jaar niet meer exact overeen met die in de EAK voor hetzelfde jaar.

Dat is ook niet geboden, gegeven de deels ongewenste schommelingen die zich van jaar tot jaar in de EAK voordoen. Dat wordt geïllustreerd voor de groep van arbeiders jonger dan 30 jaar aan de hand van figuur 4, waar de verdeling over drie opleidingsniveaus in de SAM, de EAK-regressies en de (ongewogen) EAK-observaties naast elkaar gezet wordt. De zes opleidingsniveaus werden er gegroepeerd tot drie: lager en lager secundair (LA+LS), hoger secundair (HS) en hoger (HO).

De groep van arbeiders jonger dan 30 jaar wordt jaarlijks door ongeveer 2 800 personen vertegenwoordigd in de EAK. Rechts in de figuur kan gezien worden dat er zich bij deze groep schommelingen in het aandeel hoger secundair versus lager secundair of lager geschoolden voordoen die gaan tot 5 %-punt per jaar.

Figuur 4 Ontwikkeling van het opleidingsniveau van -30 jarige arbeiders



Bron: EAK, nationale rekeningen, berekeningen FPB.

Alhoewel de EAK-evolutie op zich over een langere periode wel relevante informatie geeft, zijn de schommelingen van jaar tot jaar meer toe te schrijven aan toeval. De EAK-gemiddelden in figuur 4 zijn ongewogen. De schommelingen zijn echter even groot voor gewogen EAK-observaties.

Het logaritmisches verband tussen het jaartal en het opleidingsniveau in de regressie zorgt voor een uitvlakking van de jaarlijkse schommelingen met het behoud van de trendmatige evolutie die aangegeven wordt door de EAK-gegevens. Voor een gegeven jaar valt de kansverdeling die rechtstreeks volgt uit de regressie (weergegeven in het midden van figuur 4) niet meer samen met die EAK-observaties, maar wel voor het geheel van de beschouwde periode.

Links in figuur 4 wordt ook de verdeling in de SAM opgenomen. Die verschilt bij deze groep arbeiders nauwelijks van de verdeling die volgt uit de regressies. Dat kan wel het geval zijn, omdat de vertegenwoordiging in de SAM van de bedrijfstakken, geslachten en (deel)leeftijdsgroepen anders is dan in de EAK.

Dat de verschillen hier gering zijn, heeft juist te maken met het feit dat binnen één statuut en grote leeftijdsklasse vergeleken wordt. Uit volgend deel blijkt dat de EAK en de SAM juist in samenstelling verschillen wat betreft statuten en leeftijdsklassen.

3.6.2. Verschillen veroorzaakt door een andere weging in de SAM en de EAK

Tabel 4 toont de samenstelling van het aantal werkenden naar leeftijdsklasse en statuut in de SAM en de EAK voor het jaar 2000. De EAK-gegevens worden er zowel zonder als mét NIS-weging voorgesteld.

Onderaan de tabel blijkt dat arbeiders, met 26,4 %, minder sterk vertegenwoordigd zijn bij de EAK-respondenten dan in de SAM, waar zij 30,5 % van de werkenden uitmaken in 2000. Bedienden in de privé-sector daarentegen zijn met 31,5 % sterker vertegenwoordigd bij de EAK-respondenten dan in de SAM (28,8 %). Ook de andere categorieën (werknemers van de openbare sector, niet-vergoede helpers en gewone zelfstandigen) zijn iets sterker vertegenwoordigd bij de EAK-respondenten dan in de SAM.

Tabel 4 Aandeel (in %) in SAM en EAK van combinaties van leeftijdsklasse en statuut (2000)

	Arbeiders, privé- sector	Bedienden, privé- sector	Openbare sector	Gewone zelfstandigen	Niet- vergoede helpers	Vergoede helpers ²	Totaal
SAM							
-30	9,4	8,3	3,9	1,6	0,2	0,1	23,6
30-49	17,4	16,3	14,7	7,8	1,1	0,4	57,8
50+	3,6	4,1	5,2	4,9	0,7	0,1	18,6
EAK, ongewogen							
-30	7,4	7,8	3,3	1,5	0,2		20,2
30-49	15,3	18,6	15,7	8,7	1,1		59,4
50+	3,7	5,2	6,0	4,7	0,9		20,5
EAK, gewogen ¹							
-30							22,6
30-49							59,9
50+	3,4	4,3	5,2	3,8	0,8		17,5
Totaal in SAM	30,5	28,8	23,7	14,3	2,0	0,6	100
Totaal in EAK, ongewogen	26,4	31,5	25,0	14,9	2,2		100
Totaal in EAK, gewogen	27,9	32,0	24,4	13,7	2,1		100

1 Berekend op basis van tabellen T 1.001 en T 2.001, in: NIS (2001), "Werkgelegenheid en werkloosheid, Enquête naar de Arbeidskrachten 2000".

2 0,6 % van de werkenden in de SAM zijn vergoede helpers. In de EAK zijn die niet afzonderlijk beschikbaar, maar opgenomen in de arbeiders en bedienden van de privé-sector.

Bron: Berekeningen FPB op basis van EAK en nationale rekeningen.

De weging naar provincie, leeftijdsklasse en geslacht uitgevoerd door het NIS zorgt voor een stijging van het aandeel arbeiders (van 26,4 % naar 27,9 %), maar die is onvoldoende om te komen tot de SAM-aandelen.

Bekijkt men tabel 4 in functie van de leeftijdsklassen, dan blijkt dat de -30 jarigen minder sterk vertegenwoordigd zijn in de EAK dan in de SAM (20,2 % versus 23,6 %), terwijl de 50+ jarigen sterker vertegenwoordigd zijn in de EAK dan in de SAM (20,5 % versus 18,6 %). Ook dit verschil in samenstelling vermindert via de NIS-weging; het aandeel van de 50+ jarigen wordt (met 17,5 %) na weging zelfs kleiner dan in de SAM.

Gezien het belang van zowel het statuut als de leeftijdsklasse voor het opleidingsniveau, heeft het aandeel van elk van deze cellen een groot belang bij het bepalen van het opleidingsniveau van de werkenden, dit zowel voor het totaal als op het niveau van de bedrijfstakken.

In tabel 5 wordt het opleidingsniveau van alle werkenden in de SAM vergeleken met twee EAK-gemiddelden. Eén zonder weging en één met NIS-weging. Alhoewel er in de SAM meer lager en lager secundair opgeleiden zijn dan in de EAK, zijn de globale verschillen eerder klein. Dat komt ook omdat het kleiner aandeel arbeiders en het kleiner aandeel -30 jarigen onder de EAK-respondenten een tegengesteld effect hebben op het opleidingsniveau.

Tabel 5 Opleidingsniveau werkenden in SAM versus EAK (2000)

	EAK ongewogen	EAK gewogen ¹	SAM-cellen mét EAK-observaties	SAM-cellen zonder EAK-observaties	SAM
Lager	10,8%	10,2%	11,5%	14,2%	11,6%
Lager secundair	20,4%	20,1%	21,4%	22,8%	21,4%
Hoger secundair	35,6%	36,5%	36,1%	39,7%	36,1%
Hoger korte type	18,2%	18,2%	16,4%	11,2%	16,3%
Hoger lange type	3,9%	4,1%	3,6%	2,8%	3,6%
Universitair	11,2%	10,8%	11,0%	9,3%	10,9%
Aantal werkenden	38 900	4 092 700	4 022 100	69 200	4 091 300
			98,3%	1,7%	100,0%

¹ Berekend op basis van tabel T 2.017, in: NIS (2001), "Werkgelegenheid en werkloosheid, Enquête naar de Arbeidskrachten 2000"

Bron: Berekeningen FPB op basis van EAK en nationale rekeningen.

Tabel 5 toont ook het verschil in opleidingsniveau tussen de SAM-cellen mét EAK-observaties en die zonder EAK-observaties. Dankzij de regressietechnieken konden immers ook resultaten verkregen worden voor die "lege" cellen. 1,7 % van de hoofden in de SAM behoort tot een combinatie van bedrijfstak, leeftijdsklasse, geslacht,... die niet voorkwam in de EAK van 2000.

Men kan vaststellen dat de geschatte opleidingsniveaus voor de cellen zonder EAK-observaties lager liggen dan die in de cellen waarvoor wel EAK-observaties bestaan. Dat impliceert mogelijk dat lager geschoolden ondervertegenwoordigd zijn in de EAK. De toegepaste methodologie laat toe hiervoor (alvast gedeeltelijk) te corrigeren. De impact ervan is al bij al vrij klein.

Besluit

In deze paper werd beschreven hoe de binnenlandse hoofdelijke werkgelegenheid opgedeeld kan worden naar persoonskenmerken (geslacht, leeftijdsklasse en opleidingsniveau) en jobkenmerken (statuut en arbeidsregime). Die opdeling gebeurde op een gedetailleerd bedrijfstakniveau en conform de nationale rekeningen voor de jaren 1999 tot 2005.

Hiermee werd slechts een eerste stap gezet om aan de werkgelegenheidsgegevens uit de nationale rekeningen een kwalitatieve dimensie toe te voegen, wat uiteindelijk kan leiden tot de creatie van een volwaardige SAM (Social Accounting Matrix) voor België. De volgende logische stappen omvatten gelijkaardige werkzaamheden voor lonen en gewerkte uren.

Waar mogelijk werd gebruik gemaakt van administratieve gegevens (RSZ, RSZPPO, RVA, RSVZ en BTW), overeenkomstig de methodologie die gevolgd wordt bij het opstellen van de nationale rekeningen. Waar geen administratieve bestanden beschikbaar waren (met name voor de opdeling naar opleidingsniveau en voor de verdeling van de zelfstandigen per bedrijfstak naar leeftijdsklasse en geslacht) werd een beroep gedaan op gegevens van de enquête naar de arbeidskrachten (EAK). Zowel bij het verzoenen van administratieve gegevens met randtotalen uit de nationale rekeningen als bij het gebruik van een enquête die slechts 1 % van de beroepsbevolking bevraagt, duiken specifieke methodologische problemen op. De gekozen methodologie tracht voor al die problemen zo bevredigend mogelijke oplossingen te bieden.

Zo konden bijvoorbeeld, dankzij aangepaste regressietechnieken, de EAK-data gebruikt worden voor een groter aantal dimensies en op een groter detail dan mogelijk zou geweest zijn bij het gebruik van eenvoudige steekproefgemiddelden. Een bijkomend voordeel van de gehanteerde regressietechnieken is dat soms onrealistisch heftige jaarlijkse schommelingen werden uitgevlakt. Tegelijk laat de gekozen methode toe te corrigeren voor de kleinere vertegenwoordiging van arbeiders en jongeren in de EAK dan in de administratieve bestanden.

Het eindresultaat van deze oefening is een bestand dat toelaat diverse job- en persoonskenmerken op een gedetailleerd bedrijfstakniveau te kruisen en de evolutie ervan na te gaan over de periode 1999-2005. Men dient er zich van bewust te zijn dat men bij het uitvoeren van multidimensionele kruisingen kan stoten op de beperkingen van de gehanteerde methode, bijvoorbeeld als gevolg van de structuur opgelegd in de regressies.

Bijlagen

Bijlage 1: De IOKLEMS-classificatie

IOKLEMS	EUKLEMS	NACE	Omschrijving
01A1	1	01.1	Akkerbouw en tuinbouw
		01.2	Veeteelt
		01.3	Gemengd bedrijf
		01.4	Diensten verwant aan de landbouw en de veeteelt
		01.5	Jacht
02A1	2	02.0	Bosbouw, bosexploitatie en aanverwante diensten
05A1	B	05.0	Visserij en het kweken van vis en schaal- en schelpdieren
10A1	10	10.1	Winning van steenkool en vervaardiging van steenkoolagglomeraten
		10.2	Winning van bruinkool en vervaardiging van bruinkoolagglomeraten
		10.3	Winning van turf en vervaardiging van turfagglomeraten
11A1	11	11.1	Winning van aardolie en aardgas
		11.2	Diensten verwant aan de aardolie- en aardgaswinning
12A1	12	12.0	Winning van uranium- en thoriumerts
13A1	13	13.1	Winning van ijzererts
		13.2	Winning van non-ferro metaalerts
14A1	14	14.1	Winning van steen
		14.2	Winning van zand en klei
		14.3	Winning van mineralen voor de chemische en de kunstmestindustrie
		14.4	Productie van zout
		14.5	Overige winning van delfstoffen, n.e.g.
15A1	15	15.1	Productie en verwerking van vlees en vleesproducten
15B1	15	15.2	Verwerking en conservering van vis en vervaardiging van visproducten
15C1	15	15.3	Verwerking en conservering van groenten en fruit
15D1	15	15.4	Vervaardiging van plantaardige en dierlijke oliën en vetten
15E1	15	15.5	Zuivelnijverheid
15F1	15	15.6	Maalderijen en vervaardiging van zetmeel en zetmeelproducten
15G1	15	15.7	Vervaardiging van diervoeders
15H1	15	15.81	Brood en vers banketbakkerswerk
		15.82	Beschuit en koekjes
15I1	15	15.83	Suiker
		15.84	Chocolade en suikerwerk
15J1	15	15.85	Deegwaren
		15.86	Koffie en thee
		15.87	Specerijen, sausen, kruiden
		15.88	Voedingspreparaten, dieetvoeding
		15.89	Overige voedingsmiddelen
15K1	15	15.91	Gedistilleerde alcoholische dranken
		15.92	Ethylalcohol
		15.93	Wijn
		15.94	Cider e.d.
		15.95	Andere gedistilleerde dranken
		15.96	Brouwerijen
		15.97	Mouterijen
15L1	15	15.98	Vervaardiging van mineraalwater en frisdranken

IOKLEMS	EUKLEMS	NACE	Omschrijving
16A1	16	16.0	Vervaardiging van tabaksproducten
17A1	17	17.1	Bewerken en spinnen van textielvezels
		17.2	Weven van textiel
		17.3	Textielveredeling
17B1	17	17.4	Vervaardiging van geconfectioneerde artikelen van textiel, exclusief kleding
		17.5	Vervaardiging van overige textielproducten
		17.6	Vervaardiging van gebreide en gehaakte stoffen
		17.7	Vervaardiging van gebreide en gehaakte artikelen
18A1	18	18.1	Vervaardiging van kleding van leer
		18.2	Vervaardiging van overige kleding
		18.3	Bontnijverheid
19A1	19	19.1	Looien en bereiden van leer
		19.2	Vervaardiging van koffers, tassen en dergelijke en van overige artikelen van leer
		19.3	Vervaardiging van schoeisel
20A1	20	20.1	Zagen en schaven van hout, impregneren van hout
		20.2	Vervaardiging van panelen en platen van hout
		20.3	Vervaardiging van schrijn- en timmerwerk
		20.4	Vervaardiging van houten emballage
		20.5	Vervaardiging van overige artikelen van hout, van kurk en riet en van vlechtwerk
21A1	21	21.1	Vervaardiging van papierpulp, papier en karton
		21.2	Vervaardiging van artikelen van papier en karton
22A1	221	22.1	Uitgeverijen, zonder staatsblad
22A3	221	22.1	Het staatsblad
22B1	22x	22.2	Drukkerijen en aanverwante diensten
		22.3	Reproductie van opgenomen media
23A1A	23	23.1	Vervaardiging van cokesovenproducten
23A1B	23	23.2	Vervaardiging van geraffineerde aardolieproducten
23A1C	23	23.3	Bewerking van splijt- en kweekstoffen
24A1	24x	24.1	Vervaardiging van chemische basisproducten
24B1	24x	24.2	Vervaardiging van verdelingsmiddelen en van chemische producten voor de landbouw
24C1	24x	24.3	Vervaardiging van verf, vernis en drukinkt
24D1	244	24.4	Farmaceutische nijverheid
24E1	24x	24.5	Vervaardiging van zeep, was- en poetsmiddelen, parfums en cosmetische artikelen
24F1	24x	24.6	Vervaardiging van overige chemische producten
24G1	24x	24.7	Vervaardiging van synthetische en kunstmatige vezels
25A1	25	25.1	Rubbernijverheid
25B1	25	25.2	Vervaardiging van producten van kunststof
26A1	26	26.1	Vervaardiging van glas en glaswerk
26B1	26	26.2	Vervaardiging van keramische producten, exclusief tegels, dakpannen en bakstenen
		26.3	Vervaardiging van keramische tegels en plavuizen
		26.4	Vervaardiging van dakpannen, bakstenen en overige producten voor de bouw van gebakken klei
		26.5	Vervaardiging van cement, kalk en gips
26C1	26	26.5	Vervaardiging van cement, kalk en gips
26D1	26	26.6	Vervaardiging van artikelen van beton, gips en cement
		26.7	Bewerken van natuursteen
		26.8	Vervaardiging van overige niet-metaalhoudende minerale producten
27A1	27	27.1	Vervaardiging van ijzer en staal en van ferro-legeringen (EGKS)

IOKLEMS	EUKLEMS	NACE	Omschrijving
		27.2	Vervaardiging van buizen
27B1A	27	27.3	Eerste verwerking van staal en productie van niet-EGKS-ferrolegeringen
27B1B	27	27.4	Productie van non-ferro metalen
27B1C	27	27.5	Gieten van metalen
28A1	28	28.1	Vervaardiging van metalen constructiewerken
		28.2	Vervaardiging van metalen recipiënten; vervaardiging van radiatoren en ketels voor centrale verwarming
		28.3	Vervaardiging van stoomketels
		28.4	Smeden, persen, stampen en profielwalsen van metaal; poedermetallurgie
28B1	28	28.5	Oppervlaktebehandeling en bekleding van metaal; algemene metaalbewerking
28C1	28	28.6	Vervaardiging van scharen, messen, bestekken, gereedschap en ijzervaren
		28.7	Vervaardiging van overige producten van metaal
29A1	29	29.1	Vervaardiging van motoren en mechanisch drijfwerk, exclusief motoren voor luchtvaartuigen, motorvoertuigen en -rijwielen
29B1	29	29.2	Vervaardiging van machines voor algemeen gebruik
29C1	29	29.3	Vervaardiging van machines voor de landbouw en de bosbouw
		29.4	Vervaardiging van gereedschapswerktuigen
		29.5	Vervaardiging van overige machines voor specifieke doeleinden
		29.6	Vervaardiging van wapens en munitie
29D1	29	29.7	Vervaardiging van huishoudapparaten
30A1	30	30.0	Vervaardiging van kantoormachines en computers
31A1I	313	31.3	Vervaardiging van geïsoleerde kabels en draad
31A1N	31x	31.1	Vervaardiging van elektromotoren en van elektrische generatoren en transformatoren
		31.2	Vervaardiging van schakel- en verdeelinrichtingen
31B1	31x	31.4	Vervaardiging van accumulatoren en elektrische batterijen
		31.5	Vervaardiging van elektrische lampen en verlichtingsapparaten
		31.6	Vervaardiging van elektrische benodigdheden
32A1I	321	32.1	Vervaardiging van elektronische onderdelen
32A1J	322	32.2	Vervaardiging van zend- en transmissieapparatuur
32A1K	323	32.3	Vervaardiging van audio- en videoapparatuur
33A1I	331t3	33.1	Vervaardiging van medische apparatuur en instrumenten en van orthopedische artikelen
		33.2	Vervaardiging van wetenschappelijke en technische instrumenten
		33.3	Vervaardiging van controleapparatuur van industriële processen
33A1N	334t5	33.4	Vervaardiging van optische instrumenten en van foto- en filmapparatuur
		33.5	Vervaardiging van uurwerken
34A1	34	34.1	Vervaardiging en assemblage van auto's
34B1	34	34.2	Vervaardiging van carrosserieën, aanhangwagens en caravans
		34.3	Vervaardiging van onderdelen en accessoires voor motorvoertuigen en motoren daarvan
35A1A	351	35.1	Scheepsbouw en -reparatie
35A1B	35x	35.2	Vervaardiging van rollend materieel voor spoor- en tramwegen
35A1C	353	35.3	Vervaardiging van lucht- en ruimtevaartuigen
35B1	35x	35.4	Vervaardiging van motorrijwielen en rijwielen
		35.5	Vervaardiging van overige transportmiddelen, n.e.g.
36A1	36	36.1	Vervaardiging van meubels
36B1	36	36.2	Bewerking van edelstenen en vervaardiging van juwelen
36C1	36	36.3	Vervaardiging van muziekinstrumenten
		36.4	Vervaardiging van sportartikelen
		36.5	Vervaardiging van spellen en speelgoed

IOKLEMS	EUKLEMS	NACE	Omschrijving
		36.6	Overige industrie
37A1	37	37.1	Recuperatie van recycleerbaar metaalafval
		37.2	Recuperatie van overig recycleerbaar afval
40A1A	40x	40.1	Productie en distributie van elektriciteit
		40.3	Distributie van stoom en warm water; vervaardiging van ijs, niet bestemd voor consumptie
40A1B	402	40.2	Productie en distributie van gas
41A1	41	41.0	Winning, zuivering en distributie van water
45A1	F	45.1	Het bouwrijp maken van terreinen
45B1A	F	45.21	Alg. bouwkunde en civieltechnische werken
45B1B	F	45.22	Dakbedekking en bouw van dakconstructies
45C1A	F	45.23	Aanleg van spoorwegen, wegen, straten, vliegvelden en sportaccommodaties
45C1B	F	45.24	Waterbouw
45C1C	F	45.25	Overige werkzaamheden in de bouw
45D1	F	45.3	Bouwinstallatie
45E1	F	45.4	Afwerking van gebouwen
		45.5	Verhuur van machines voor de bouwnijverheid met bedieningspersoneel
50A1	50	50.1	Handel in auto's
		50.2	Onderhoud en reparatie van auto's
		50.3	Handel in onderdelen en accessoires van auto's
		50.4	Handel in en reparatie van motorrijwielen
50B1	50	50.5	Kleinhandel in motorbrandstoffen
51A1A	51	51.51	Groothandel in brandstoffen
51A1B	51	51.1	Handelsbemiddeling
		51.2	Groothandel in landbouwproducten en levende dieren
		51.3	Groothandel in voedings- en genotmiddelen
		51.4	Groothandel in huishoudelijke artikelen
		51.52	Groothandel in metalen en metaalertsen
		51.53	Groothandel in hout, verf, vernis en bouwmaterialen, inclusief sanitair
		51.54	Groothandel in ijzerwaren, loodgietersmateriaal en verwarmingsinstallaties
		51.55	Groothandel in chemische producten
		51.56	Groothandel in overige intermediaire producten
		51.57	Groothandel in afval en schroot
		51.6	Groothandel in machines, apparaten en toebehoren
		51.7	Overige groothandel
		52A1	52
52.2	Kleinhandel in voedings- en genotmiddelen in gespecialiseerde winkels		
52.3	Kleinhandel in farmaceutische producten, reukwerk en cosmetica		
52.4	Overige gespecialiseerde kleinhandel in nieuwe artikelen in winkels		
52.5	Kleinhandel in tweedehandsgoederen en antiques in winkels		
52.6	Kleinhandel, niet in winkels		
52.7	Reparatie van consumentenartikelen		
55A1	H	55.1	Hotels
		55.2	Overige accommodaties voor kortstondig verblijf
55B1	H	55.3	Restaurants
		55.4	Drankgelegenheden
		55.5	Kantines en catering
60A1	60	60.1	Vervoer per spoor
60B1	60	60.21	Personenvervoer te land volgens dienstregeling

IOKLEMS	EUKLEMS	NACE	Omschrijving
		60.22	Taxi's
		60.23	Overig personenvervoer
60B3	60	60.21	Gewestelijke vervoersmaatschappijen: De Lijn, MIVB, Sociétés Régionales Wallonnes des Transports
60C1	60	60.24	Goederenvervoer over de weg en verhuisdiensten
		60.3	Vervoer via pijpleidingen
61A1	61	61.1	Zee- en kustvaart
61B1	61	61.2	Binnenvaart
62A1	62	62.1	Luchtvaart volgens dienstregeling
		62.2	Luchtvaart zonder dienstregeling
		62.3	Ruimtevaart
63A1	63	63.3	Reisbureaus en touroperators
63B1	63	63.1	Vrachtbehandeling en opslag
		63.2	Overige vervoerondersteunende activiteiten
		63.4	Organisatie van het vrachtvervoer
63B3	63	63.2	Exploitatie en onderhoud van autowegen en waterwegen door gemeentelijke, provinciale en regionale overheidsdiensten
64A1	64	64.1	Postactiviteiten
64B1	64	64.2	Telecommunicatie
65A2	65	65.1	Geldscheppende financiële instellingen
		65.2	Overige financiële instellingen
66A2	66	66.0	Verzekeringswezen
67A1	67	67.1	Hulpbedrijven i.v.m. financiële instellingen
		67.2	Hulpbedrijven i.v.m. het verzekeringswezen
70A1A	70imp	n.v.t.	Toegerekende huren
70A1B	70x	70.1	Projectontwikkeling en handel in onroerend goed voor eigen rekening
		70.2	Verhuur van eigen onroerend goed
		70.3	Bemiddeling in en beheer van onroerend goed voor rekening van derden
71A1	71	71.1	Verhuur van auto's
		71.2	Verhuur van overige transportmiddelen
71B1	71	71.3	Verhuur van machines en werktuigen
		71.4	Verhuur van overige roerende goederen
72A1A	72	72.1	Computeradviesbureaus
		72.2	Realisatie van programma's en gebruiksklare systemen
		72.3	Gegevensverwerking
		72.4	Databanken
72A1B	72	72.5	Onderhoud en reparatie van computers en van kantoormachines
		72.6	Overige activiteiten in verband met computers
73A1	73	73.1	Natuurwetenschappelijk speur- en ontwikkelingswerk, markt
		73.2	Speur- en ontwikkelingswerk i.v.m. sociale en geesteswetenschappen, markt
73A5	73	73.1	Natuurwetenschappelijk speur- en ontwikkelingswerk, niet-markt
		73.2	Speur- en ontwikkelingswerk i.v.m. sociale en geesteswetenschappen, niet-markt
74A1	741t4	74.11	Rechtskundige dienstverlening
		74.12	Accountants, boekhouders en belastingconsulenten
		74.13	Markt- en opinieonderzoeksbureaus
74B1	741t4	74.14	Advies bedrijfsvoering en beheer
		74.15	Managementactiviteiten van holdings en coördinatiecentra
74C1	741t4	74.2	Technisch advies, architecten en ingenieurs
		74.3	Technische testen en analyses
74D1	741t4	74.4	Reclamewezen

IOKLEMS	EUKLEMS	NACE	Omschrijving
74E1	745t8	74.5	Selectie en terbeschikkingstelling van personeel
74F1	745t8	74.6	Opsporings- en beveiligingsdiensten
		74.7	Industriële reiniging
		74.8	Diverse dienstverlening hoofdzakelijk aan bedrijven
75A3	L	75.1	Algemeen economisch en sociaal bestuur
		75.21	Buitenlandse zaken
		75.23	Justitie
		75.24	Openbare veiligheid
		75.25	Civiele bescherming
75B3	L	75.22	Defensie
75C3	L	75.3	Verplichte sociale verzekering
80A1	M	80.1	Basisonderwijs, markt
		80.2	Voortgezet onderwijs, markt
		80.3	Hoger onderwijs, markt
		80.4	Volwasseneducatie en overige vormen van onderwijs, markt
80A3	M	80.1	Basisonderwijs, openbaar
		80.2	Voortgezet onderwijs, openbaar
		80.3	Hoger onderwijs, openbaar
		80.4	Volwasseneducatie en overige vormen van onderwijs, openbaar
80A5	M	80.1	Basisonderwijs, niet-markt
		80.2	Voortgezet onderwijs, niet-markt
		80.3	Hoger onderwijs, niet-markt
		80.4	Volwasseneducatie en overige vormen van onderwijs, niet-markt
85A1	N	85.1	Gezondheidszorg
85B1	N	85.2	Veterinaire diensten
85C1	N	85.3	Maatschappelijke dienstverlening, markt
85C5	N	85.3	Maatschappelijke dienstverlening, niet-markt
90A1	90	90.0	Afvalwater- en afvalverzameling; straatreiniging, markt
90A3	90	90.0	Afvalwater- en afvalverzameling; straatreiniging, openbaar
91A1	91	91.11	Bedrijfs- en werkgeversorganisaties
		91.12	Beroepsorganisaties, markt
		91.3	Overige verenigingen, markt
91A5	91	91.12	Beroepsorganisaties, niet-markt
		91.2	Vakverenigingen
		91.3	Overige verenigingen, niet-markt
92A3	921t2	92.2	Openbare omroepen
92A1	921t2	92.1	Activiteiten op het gebied van film en video, markt
		92.2	Radio en televisie, markt
92B1	923t7	92.3	Overige activiteiten op het gebied van amusement, markt
92B5	923t7	92.3	Overige activiteiten op het gebied van amusement, niet-markt
92C1	923t7	92.4	Persagentschappen
		92.5	Overige culturele activiteiten, markt
92C5	923t7	92.5	Overige culturele activiteiten, niet-markt
92D1	923t7	92.6	Sport markt
		92.7	Overige recreatie, markt
92D5	923t7	92.6	Sport niet-markt
		92.7	Overige recreatie, niet-markt
93A1	93	93.0	Overige diensten
95A4	P	95.0	Particuliere huishoudens met werknemers

Bijlage 2: Bedrijfstakindeling voor de opdelingen op basis van EAK-gegevens

Subsectie* werknemers publieke sector (22)	Subsectie* zelfstandigen (31)	Subsectie* werknemers privé-sector (41)	NACE 2-digiet	Omschrijving subsectie / minimum labour quality industry division **
A+B	A+B	A+B	1, 2, 5	Agriculture, hunting, forestry & fishing Landbouw, jacht, bosbouw en visvangst
DD/DJ+C	DF+DG+CA	DF+DG+CA	10, 11, 12	Other primary production Energiehoudende delfstoffen
DD/DJ+C	DH+DI+CB	CB	13, 14	Winning van niet-energiehoudende delfstoffen
DA	DA	DA	15, 16	Consumer manufacturing Voeding en genotsmiddelen
DB+DC	DB+DC	DB+DC	17, 18, 19	Textiel en textielproducten; leder en schoeisel
DD/DJ+C	DD+DE	DD+DE21	20, 21	Intermediate manufacturing Hout- en papierindustrie
DD/DJ+C	DD+DE	DE22	22	Uitgeverijen en drukkerijen
DD/DJ+C	DF+DG+CA	DF+DG+CA	23, 24	Petrochemie & chemie
DD/DJ+C	DH+DI+CB	DH	25	Rubber en kunststof
DD/DJ+C	DH+DI+CB	DI	26	Overige niet metaalhoudende minerale producten
DD/DJ+C	DJ	DJ27	27	Metaal in primaire vormen
DD/DJ+C	DJ	DJ28	28	Producten van metaal
DK/DM	DK+DM	DK	29	Investment manufacturing Machines, apparaten en werktuigen
DK/DM	DL	DL32	32	Audio, video en telecommunicatieapparatuur
DK/DM	DL	Dlrest	30, 31, 33	Rest elektrische & elektronische machines en apparaten
DK/DM	DK+DM	DM	34, 35	Transportmiddelen
DN	DN	DN	36, 37	Other production Overige industrie
E	E	E	40, 41	Elektriciteit, gas & water
F	F	F	45	Construction Bouwnijverheid
G50	G50	G50	50	Distributive trade Verkoop, onderhoud en reparatie van motorvoertuigen en motorrijwielen; detailhandel in motorbrandstoffen
G51/52	G51	G51	51	Groothandel en handelsbemiddeling
G51/52	G52	G52	52	Kleinhandel en reparatie van consumentenartikelen
H	H	H	55	Hotels en restaurants
I60/63	I60	I60	60	Transport Vervoer te land & via pijpleidingen
I60/63	I61-63	I61+62	61, 62	Vervoer over water en luchtvaart
I60/63	I61-63	I63	63	Vervoerondersteunende activiteiten; reisbureaus
I64	I64	I64	64	Communications Post en telecommunicatie
J	J	J	65, 66, 67	Financial intermediation Financiële instellingen
K70/71	K70/71	K70/71	70, 71	Real estate and business activities Verhuur & immobiëlen

Subsectie* werknemers publieke sector (22)	Subsectie* zelfstandigen (31)	Subsectie* werknemers privé-sector (41)	NACE 2-digitaal	Omschrijving subsectie / minimum labour quality industry division **
K72/73	K72/73	K72	72	Activiteiten i.v.m. computer
K72/73	K72/73	K73	73	Speur- en ontwikkelingswerk
K74	K74	K74	74	Overige zakelijke diensten, behalve bestuurders van vennootschappen, uitzendkrachten en PWA
K74e		K74e	74	Uitzendkrachten en werknemers Plaatselijke WerkgelegenheidsAgentschappen
	K74v	K74v	74	Bestuurders van vennootschappen
L	L75	L75	75	Public administration Openbaar bestuur en defensie
M	M80	M80	80	Education Onderwijs
N	N85	N85	85	Health Gezondheidszorg en maatschappelijke dienstverlening
O+P	O90	O90	90	Community social and personal services Inzameling en verwerking van afvalwater en afval
O+P	O91	O91	91	Diverse verenigingen
O+P	O92	O92	92	Recreatie, cultuur en sport
O+P	O93+P	O93	93	Overige diensten
O+P	O93+P	P	95	Huishoudens als werkgever van huishoudelijk personeel

* De "subsectie" is het meest gedetailleerde bedrijfstakniveau waarop een onderscheid naar opleidingsniveau gemaakt wordt in de uitgevoerde EAK-regressies. Dit verschilt tussen werknemers in de privé-sector, zelfstandigen en werknemers in de publieke sector.

** De Minimum Labour Quality Industry Division is de bedrijfstakindeling die als minimaal beschouwd wordt voor het verzamelen van kwalitatieve arbeidsgegevens binnen het EUKLEMS-project. Die opdeling in 16 bedrijfstakken wordt altijd gerespecteerd binnen de regressies, op de winning van energiehoudende delfstoffen na, waar geopteerd werd voor een gelijkgeschakeling van de kwalitatieve gegevens met de petrochemie.