

Een onderzoek naar de determinanten van de uitstroom naar werk

Lieven Desmet
Directie Statistieken en Studies
RVA

16 september 2011

Samenvatting

In deze studie wordt onderzocht of (en zo ja in welke mate) bepaalde factoren kunnen verklaren of een werkloze al dan niet werk vindt. De analyse werd uitgevoerd op data van het "RVA panel" en is gebaseerd op een steekproef uit de groep die in de loop van 2007 voor het eerst werkloos werd. Voor deze mensen werd gekeken of zij één jaar later aan het werk zijn en dit wordt gelinkt aan persoonskenmerken, kenmerken van de werkloosheid en familiale achtergrond. De analyse bevestigt een aantal bekende feiten: verschillen tussen de gewesten, grote impact van het studieniveau maar ook van leeftijd en de vergoedingscategorie. De verschillen qua geslacht kunnen in verband gebracht worden met de familiale situatie van de persoon.

Dankwoord

De auteur bedankt Kristel Bogaerts en Karel Van den Bosch van het Centrum voor Sociaal Beleid Herman Deleeck (Universiteit Antwerpen), en Joost Bollens van het HIVA (KULeuven) voor hun opmerkingen en advies.

De onderzoeksdatbank "RVA Panel" is tot stand gekomen dank zij een samenwerking tussen de RVA en het Federaal Wetenschapsbeleid met als doel te voldoen aan de kennisbehoeften van de RVA. De databank is samengesteld met de hulp van onderzoekers van de Universiteit Antwerpen en de KULeuven in het kader van een AGORA project gefinancierd door de POD Wetenschapsbeleid. De gegevens zijn verzameld door de Kruispuntbank van de Sociale Zekerheid. Een woord van dank ook aan Chris Brijs van de KSZ voor de vlotte samenwerking.

Voor het nalezen, hulp bij de vertaling en de opmerkingen is de hulp van de collega's op de RVA ten zeerste geapprecieerd.

1 Onderzoeksopzet en methodologie

1.1 Inleiding en onderzoeksvraag

Voor beleidsmakers is het belangrijk goed inzicht te hebben in de factoren die maken dat een werkloze er in slaagt de werkloosheid te verlaten. We spreken aldus van uitstroom uit de werkloosheid, maar zijn uiteraard vooral geïnteresseerd in de uitstroom naar werk, en niet in de transities naar een toestand van inactiviteit. Het verhogen van de werkzaamheidsgraad is één van de hoofddoelstellingen van de strategie Europa 2020 voor economische groei en werkgelegenheid.

Om onderzoek te verrichten naar de uitstroom naar werk moeten we beschikken over een uitgebreide dataset die niet alleen gegevens over de werkloosheid bevat maar ook gegevens over tewerkstelling, persoonsgegevens en zo meer. Dit is slechts mogelijk door het kruisen van de bestanden van de betalingen voor werkloosheid (RVA) met andere bestanden betreffende de sociale zekerheid. We maken dan ook gebruik van het zogenaamde RVA panel: een geïntegreerde databank speciaal ontworpen voor onderzoek naar de loopbanen van mensen die ooit in aanraking kwamen met de RVA. Deze databank, gebaseerd op een voldoende grote steekproef, is samengesteld in het kader van een samenwerking tussen de RVA, de KSZ en onderzoekers van de KULeuven en Universiteit Antwerpen in een project gefinancierd door de POD Wetenschapsbeleid.

In het RVA panel vinden we niet alleen gegevens over de werkloosheid, zoals de vergoedingscategorie, de toelaatbaarheidsbasis, de hoogte van het dagbedrag en de duur, maar ook persoonsgegevens zoals de woonplaats, het geslacht, de leeftijd, de nationaliteit en gegevens over de familiale situatie van de betrokkene zoals de familieband die deze heeft binnen het gezin, het feit of er kinderen in het gezin zijn (en de leeftijd van het jongste kind) en wat de situatie is van de eventuele partner. Belangrijke gegevens betreffen de zogenaamde socio-economische positie die aangeeft of de persoon werkzoekend is, aan het werk is of inactief is (niet louter gebaseerd op RVA gegevens maar een resultaat afgeleid uit de combinatie van allerlei bronnen betreffende de sociale zekerheid). Bovendien beschikken we ook over gegevens in verband met het studieniveau. Alle gegevens zijn anoniem.

Vermits de gegevens op kwartaalbasis voorhanden zijn (uitgezonderd persoonsgegevens) kunnen we een bepaalde persoon daadwerkelijk doorheen de tijd opvolgen en transities detecteren van een positie *werkloos* naar een positie *werkend*. Op die manier kunnen we dus een uitstroom naar werk vaststellen.

Nu kunnen we de onderzoeksvraag al wat preciezer formuleren: welke kenmerken (van de persoon, van de werkloosheid, familiaal enz.) beïnvloeden (of verklaren) de uitstroom - of niet - naar werk? In welke zin en in welke mate?

Deze ogenschijnlijk eenvoudige vraag kent geen kort en simpel antwoord: het is duidelijk dat meerdere kenmerken tegelijk een rol spelen. Onze aanpak bestaat er

in via een statistische analyse (nl. logistische regressie) de kans op uitstroom naar werk te modelleren als functie van de verschillende variabelen die overeenstemmen met geobserveerde kenmerken. Uit de resultaten van een dergelijke oefening komt dan naar voor welke effecten belangrijk zijn, welke minder belangrijk zijn en in welke zin het effect gaat. Het grote voordeel van deze methode ligt in het feit dat de verschillende effecten tegelijk kunnen bestudeerd worden.

Om de tekst leesbaar te maken voor een breed publiek worden technische details behandeld in de Appendix. Allereerst moeten evenwel een aantal definities en keuzes vastgelegd worden.

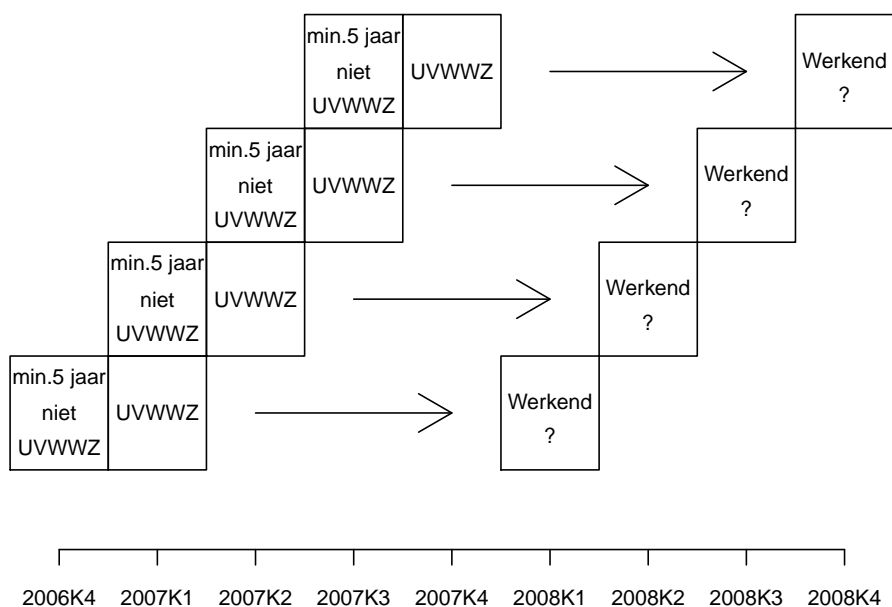
1.2 Onderzoekspopulatie en begrip uitstroompercentage

Allereerst verduidelijken we onze doelgroep: het gaat over personen die enerzijds bij de RVA gekend zijn als vergoede werkloze (meer precies: "UVWWZ" werkzoekende uitkeringsgerechtigde volledige werkloze) en anderzijds bekend staan als werkzoekende (in de nomenclatuur van de socio-economische positie). We zullen nog een aantal andere keuzes moeten maken. De timing is van het grootste belang: wanneer is de betrokkene gekend als werkloze? en op welk later moment gaan we kijken of de persoon is uitgestroomd naar werk of niet? Bovendien is geweten dat de voorgeschiedenis van de werkloze een belangrijke rol speelt.

Om een meer homogene groep te bekomen beschouwen we *alleen personen die geen voorgeschiedenis als werkloze hebben*. Het is geweten dat langdurig werklozen lagere uitstroomkansen hebben, maar het is niet eenvoudig om de duurtijd van een werkloze nauwkeurig te meten, en het is ook geweten dat de duur zelf beïnvloed wordt door andere factoren die misschien niet gemeten worden (omdat naarmate de tijd vordert er een zekere selectie optreedt waardoor mensen met betere kansen uitstromen en mensen met minder gunstige kansen sterker vertegenwoordigd worden in de groep). We merken op dat dit een belangrijke methodologische keuze is: het is ook mogelijk de hele populatie te bekijken van alle werklozen ongeacht hun voorgeschiedenis (en dan zal de duur naar voor komen als belangrijke determinant).

Via de steekproeftrekking in het RVA panel wordt zo goed mogelijk de dynamiek van de populatie werklozen nagebootst. In het bijzonder wordt per kwartaal een steekproef getrokken uit de zogenaamde *nieuwkomers*: dit zijn werklozen (in dat kwartaal) die in de laatste 5 jaar (voorafgaand aan dat kwartaal) niet bekend waren bij de RVA. We concentreren ons op deze groep, wat ook als voordeel heeft dat we weten dat zij ongeveer op hetzelfde ogenblik werkloos geworden zijn.

We voegen gegevens over de werklozen ingestroomd in de vier kwartalen van 2007 samen (dit schakelt ook het seizoen-effect uit) en bekomen aldus een steekproef van 14578 personen. Voor elke werkloze gaan we kijken, 4 kwartalen verder in de tijd, of de betrokkene op dat ogenblik een positie "werkend" inneemt (zie Figuur 1). In dat geval spreken we van een uitstroom naar werk.



Figuur 1: Timing van in- en uitstroom

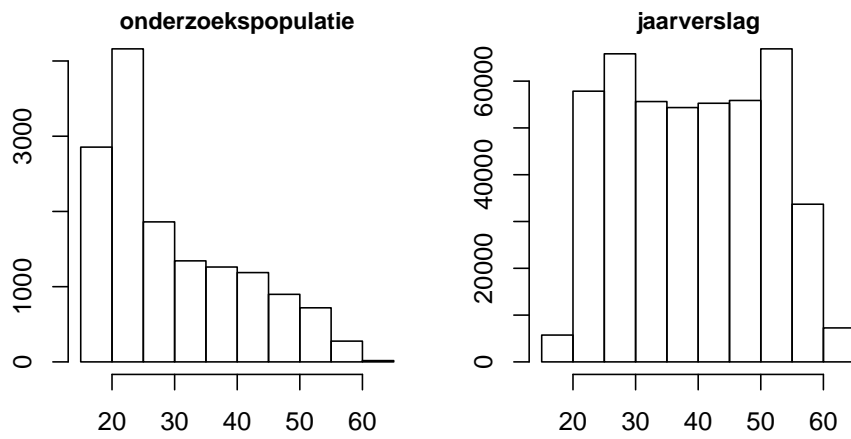
Kort gezegd, we beschouwen de populatie van alle werklozen die in de loop van 2007 ingestroomd zijn in het RVA panel en kijken of zij één jaar na instroom aan het werk zijn. Laten we eerst kijken naar een aantal kenmerken van deze populatie.

1.2.1 Profiel van de onderzoekspopulatie

In de opbouw van de steekproef voor het RVA panel is er voor gezorgd dat de verhoudingen tussen de verschillende RVA statuten en de gewesten gerespecteerd zijn. Wegens de keuzes in Sectie 1.2 is er echter een verschil tussen onze onderzoekspopulatie (nieuwe instroom in 2007) en bijvoorbeeld de populatie van alle UVW WZ gekend in 2007 en besproken in het RVA jaarverslag.

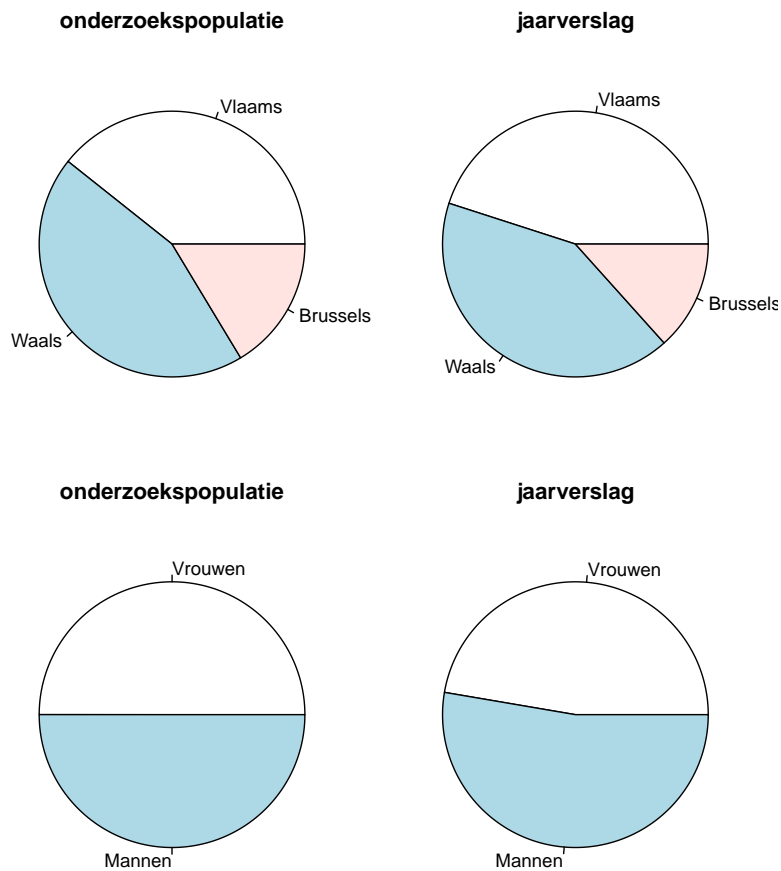
Dit wordt geïllustreerd met een aantal figuren. Het histogram van de leeftijd voor de onderzoekspopulatie, in Figuur 2, toont een asymmetrische verdeling (rechts-scheef), waarbij alle leeftijden vertegenwoordigd zijn maar de meest voorkomende leeftijd ligt vooraan in de twintig (de mediaan bedraagt 26 jaar en het gemiddelde ongeveer 30). Uiteraard komen de schoolverlaters van 2007 in aanmerking om getrokken te worden in onze steekproef (en we zien effectief dat de groepen nieuwkomers in het tweede en derde trimester iets talrijker zijn dan in de overige

trimesters). Het histogram van de leeftijd in de populatie van het jaarverslag is eerder bimodaal: er is ook een grote groep oudere werklozen. Voor wat betreft de duur is het contrast nog groter: in de onderzoekspopulatie bedraagt de mediane duur 2 maanden terwijl die in de populatie van het jaarverslag in de klasse van 24 tot 29 maanden valt (bewerkingen op RVA gegevens voor juni 2007).



Figuur 2: Links: histogram van de leeftijden in de steekproef, rechts: idem voor de populatie van de vergoede werklozen (jaarverslag 2007).

Figuur 3 toont de verschillen naar geslacht en naar gewest: in de onderzoekspopulatie zijn het Brussels Hoofdstedelijk Gewest en het Waals Gewest meer vertegenwoordigd dan in de populatie van alle UVW WZ. Dit kunnen we verklaren door het feit dat deze regio's relatief meer jongere werklozen tellen. Qua geslacht is onze onderzoekspopulatie evenwichtiger dan de populatie van alle UVW WZ waarin de mannen meer vertegenwoordigd zijn.



Figuur 3: Links: onze onderzoekspopulatie, rechts: gegevens jaarverslag 2007. Boven: verdeling naar gewest, onder: verdeling naar geslacht.

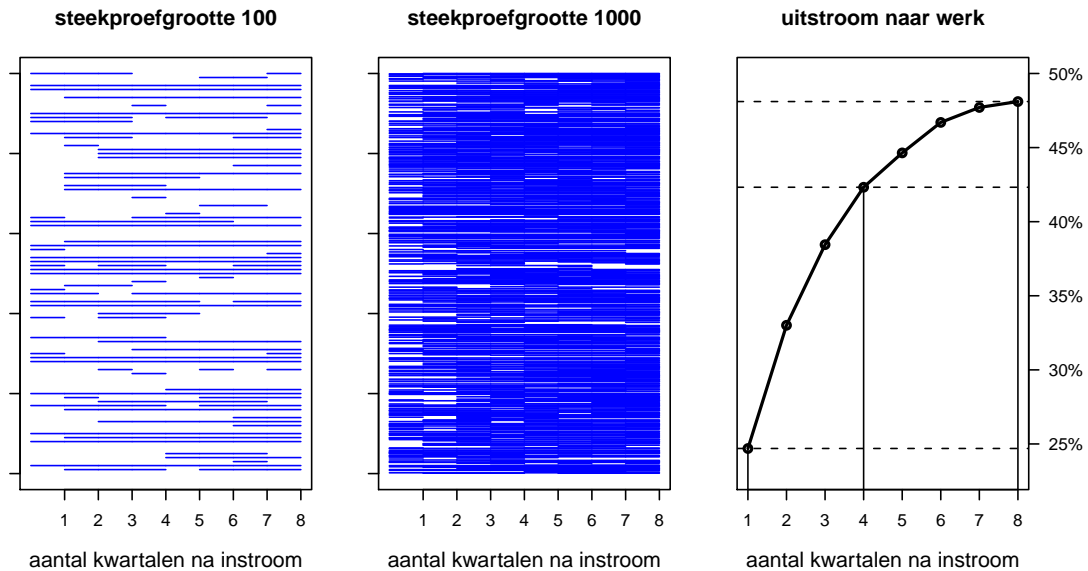
1.2.2 Criteria voor de uitstroom naar werk

Eén jaar na hun instroom zijn er van de 14578 werklozen 6170 aan het werk, zijnde 42,32%, het globaal uitstroompercentage. Dit percentage zouden we kunnen interpreteren als de kans dat een willekeurig iemand uit de steekproef - één jaar na instroom - uitgestroomd is naar werk (en zal een goede benadering zijn voor de uitstroomkans in de volledige onderzoekspopulatie, waarvan we slechts een steekproef kennen). In het vervolg zullen we het verband leggen tussen deze kans en de kenmerken van de persoon. We kunnen dan voor een willekeurige werkloze de uitstroomkans voorspellen op basis van de kenmerken, en aldus het effect van bepaalde kenmerken bespreken.

We merken nog op dat het percentage sterk afhankelijk is van het tijdstip in de opvolging: indien we zouden gaan kijken naar de situatie 1 kwartaal na de instroom

zal deze merkkelijk lager liggen (ong. 25%), indien we gaan kijken na 8 kwartalen zal deze hoger liggen (ong. 48%).

Dit wordt geïllustreerd in Figuur 4.



Figuur 4: Uitstroom naar werk evolueert met de tijd. Loopbanen gedurende de opvolging voor een steekproef van 100 personen (links) en 1000 personen (midden): lijnstukjes geven aan wanneer iemand aan het werk is in de 8 kwartalen na instroom, als de persoon niet gekend is als werkend wordt geen lijnstuk getekend. Rechts: uitstroompercentage stijgt naarmate de tijd vordert in de opvolging.

Anderzijds kijken we in onze definitie slechts naar de toestand op één ogenblik. Als we in dezelfde figuur kijken naar de trajecten die doorlopen worden door een aantal willekeurige werklozen in de 8 kwartalen na hun instroom zien we dat een aantal werklozen onafgebroken aan het werk blijven (volle lijn) of niet uitstromen naar werk (geen lijn getekend) maar ook dat heel wat werklozen een grillige loopbaan doorlopen. Naarmate de tijd vordert zien we wél dat meer personen tegelijk aan het werk zijn. We merken op dat een momentopname, bijvoorbeeld na 4 kwartalen geen volledig beeld geeft.

De gegevens waar we mee werken zijn de laatst gekende gegevens in het RVA panel: personen die werkloos werden in 2007 en eventueel uitstromen in 2008 of 2009. Uiteraard is het de bedoeling de studie te actualiseren zodra recentere gegevens beschikbaar zijn. De timing van deze studie is in elk geval interessant omdat we er mogen van uit gaan dat de crisis nog niet zo veel invloed heeft, qua conjunctuur is het een relatief gunstige periode.

2 Determinanten van de uitstroom naar werk

Vooraleer een bespreking te maken van de determinanten moeten we beschikken over een degelijk model dat enerzijds goed bij de data past en anderzijds niet te ingewikkeld is zodat het interpreteerbaar blijft. Het is niet zinvol alle variabelen waar men over beschikt op te nemen in het model, alleen al omdat bepaalde variabelen ongeveer dezelfde informatie bevatten wat de berekeningen in de war stuurt.

De keuze van het model is een onderzoek op zich en hangt af van meerdere overwegingen zoals voorafgaand inzicht in de factoren waarvan geweten is dat ze bepalend zijn, gecombineerd met een evaluatie van de statistische maten die weergeven hoe waarschijnlijk het model is (gegeven de data) en hoe goed het de data verklaart.

Soms beïnvloeden effecten mekaar: zo is het mogelijk dat de invloed van bijvoorbeeld de vergoedingscategorie anders is voor mannen dan voor vrouwen. In dát geval kan het zinvol zijn een interactie-effect op te nemen.

Voor technische details verwijzen we naar de Appendix waar de variabelenlijst en de modelselectie besproken worden. Uiteindelijk werden twee verschillende modellen weerhouden: een basismodel dat zoveel mogelijk alle relevante effecten (maar zonder overvloedige variabelen) bevat maar gebruik maakt van de RVA vergoedingscategorie en niet de huishoudpositie (positie die de rechthebbende inneemt in het gezin), en een alternatief model waarin de huishoudpositie een belangrijke rol speelt (om toch een goed beeld te krijgen van de invloed van gezinssituatie).

2.1 Bespreking van het basismodel

Het basismodel dat weerhouden is bevat als effecten het studieniveau, het gewest, de vergoedingscategorie, de leeftijd, de hoogte van het dagbedrag, de nationaliteit, het geslacht, het gegeven of er een werkende partner is, de toelaatbaarheidsbasis, de leeftijdsklasse van het jongste kind (zo er kinderen zijn) en tenslotte het interactie-effect tussen geslacht en vergoedingscategorie.

Dit is de volgorde waarin de effecten gerangschikt zijn volgens dalende waarschijnlijkheid (zie ook Tabel 11), wat iets zegt over het relatieve belang van de effecten. Zonder dit al te strikt te nemen mogen we in elk geval zeggen dat studieniveau en gewest de belangrijkste effecten zijn in ons model, gevolgd door leeftijd en vergoedingscategorie.

Het is mogelijk op basis van de parameterschattingen een interpretatie te maken van de effecten van de verschillende determinanten: in Tabel 12 worden schattingen gegeven van de parameters en hun significantie. Daar zien we bijvoorbeeld een sterk significant effect in de variabele "gewest" met negatieve coëfficiënten voor de

categorieën Wallonië en Brussel (t.o.v. de referentiecategorie Vlaanderen), wat betekent dat de uitstroom naar werk in het Waals en het Brussels Gewest beduidend lager ligt dan in het Vlaams Gewest. Op basis van de coëfficiënten kunnen we echter ook de geschatte uitstroomkans berekenen voor een aantal concrete gevallen en dit vergemakkelijkt de interpretatie van de effecten (voor een voorbeeldberekening, zie Sectie 4.3.3). Dit is de aanpak in deze Sectie.

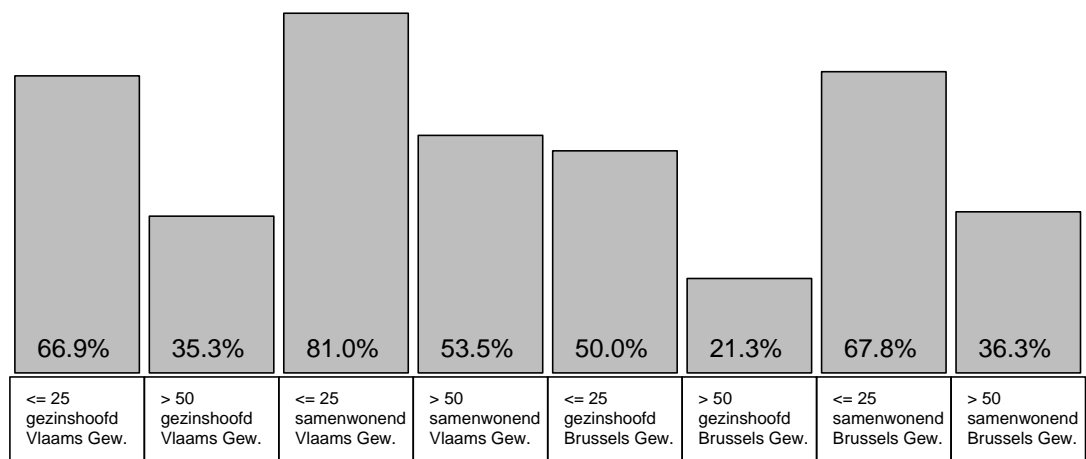
Om het effect van de belangrijkste determinanten studieniveau, gewest, vergoedingscategorie en leeftijdsklasse te bespreken kijken we naar een aantal concrete gevallen waar combinaties gemaakt worden van waarden die deze kenmerken kunnen aannemen en hiervoor wordt de uitstroomkans geschat (zie Tabel 1). *We zullen in alle voorbeelden de kansen voor de drie gewesten geven. Waarden van de andere variabelen die wegens plaatsgebrek niet in de tabel zitten komen zoveel mogelijk verderop nog aan bod.*

| studie-niveau | vergoedings-categorie | werkende partner | leeftijd (klasse) | kans per gewest | | |
|---------------|-----------------------|------------------|-------------------|-----------------|-------|----------|
| | | | | Vlaams | Waals | Brussels |
| hoog | gezinshoofd | neen | ≤ 25 | 66,9% | 53,6% | 50,0% |
| hoog | gezinshoofd | neen | > 50 | 35,3% | 23,8% | 21,3% |
| hoog | samenwonend | ja | ≤ 25 | 81,0% | 70,9% | 67,8% |
| hoog | samenwonend | ja | > 50 | 53,5% | 39,7% | 36,3% |
| midden | gezinshoofd | neen | ≤ 25 | 52,7% | 38,9% | 35,5% |
| midden | gezinshoofd | neen | > 50 | 23,1% | 14,7% | 13,0% |
| midden | samenwonend | ja | ≤ 25 | 70,1% | 57,3% | 53,7% |
| midden | samenwonend | ja | > 50 | 38,8% | 26,6% | 23,9% |
| laag | gezinshoofd | neen | ≤ 25 | 43,3% | 30,4% | 27,4% |
| laag | gezinshoofd | neen | > 50 | 17,1% | 10,5% | 9,3% |
| laag | samenwonend | ja | ≤ 25 | 61,7% | 47,9% | 44,3% |
| laag | samenwonend | ja | > 50 | 30,3% | 19,9% | 17,7% |

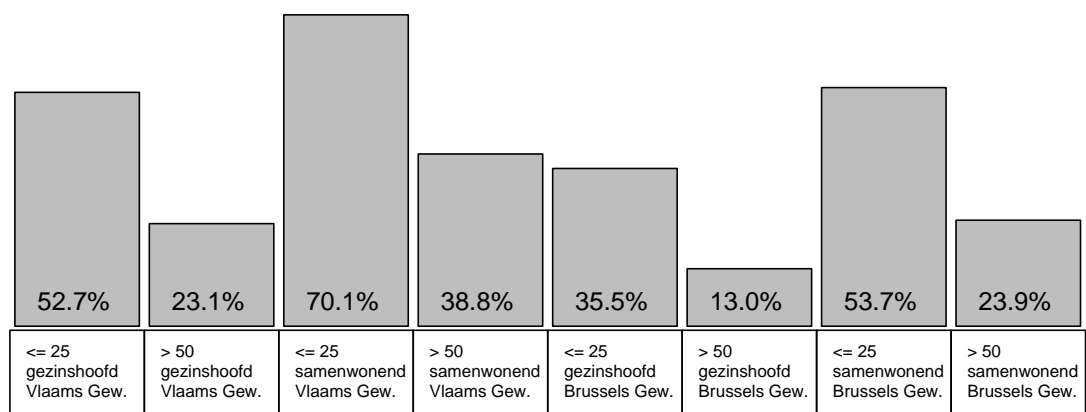
Tabel 1: Effect van de belangrijkste determinanten

In deze voorbeelden gaat het - voor de overige kenmerken - over een Belg, van het mannelijk geslacht, toelaatbaar na arbeid, met een middelhoge uitkering en zonder kinderen (andere situaties worden weerom verder besproken). Indien de RVA vergoedingscategorie "samenwonend" is wordt tegelijk ook het effect van een werkende partner verondersteld.

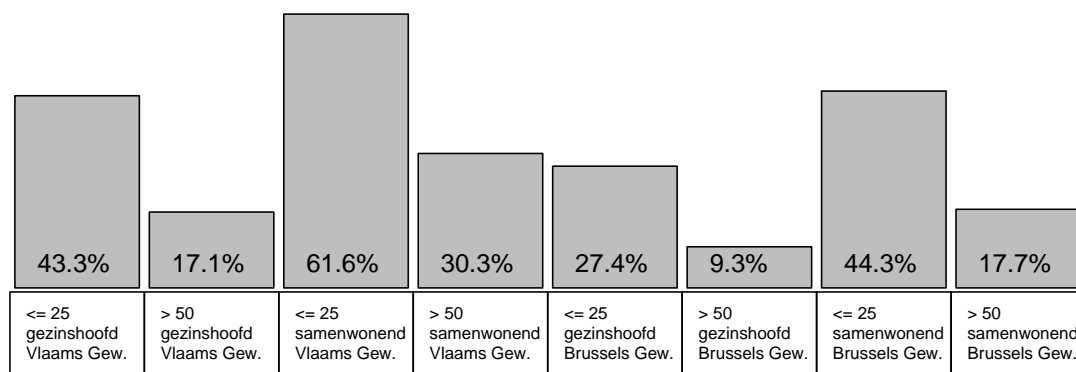
Om een aantal voorbeelden aanschouwelijk te maken worden de kansen grafisch voorgesteld in Figuur 5, Figuur 6 en Figuur 7 respectievelijk.



Figuur 5: Hoogopgeleiden: geschatte uitstroomkansen voor verschillende combinaties van leeftijdsklasse, gewest en vergoedingscategorie.



Figuur 6: Middengroep: geschatte uitstroomkansen voor verschillende combinaties van leeftijdsklasse, gewest en vergoedingscategorie.



Figuur 7: Laagopgeleiden: geschatte uitstroomkansen voor verschillende combinaties van leeftijdsklasse, gewest en vergoedingscategorie.

Uit deze voorbeelden blijkt meteen het belangrijk effect van het scholingsniveau: waar de kansen voor een laaggeschoolde grosso modo tussen de 10% en de 60% liggen is dat bij de groep hooggeschoolden tussen de 20% en de 80%. Indien de studiecodel een leercontract betreft ligt de uitstroomkans iets hoger dan voor iemand met een opleiding op niveau secundair.

Ook blijken uit Tabel 1 grote verschillen tussen de gewesten: in het Vlaams Gewest ligt de uitstroomkans duidelijk hoger dan in het Brussels Hoofdstedelijk Gewest. In het Waals Gewest liggen de kansen iets beter dan in het Brussels Hoofdstedelijk Gewest, maar de kloof met het Vlaams Gewest blijft belangrijk. Indien we op niveau arrondissement zouden kijken zullen verschillen binnen de gewesten naar boven komen, maar daarvoor beschikken we over iets te weinig data.

Voor het Brussels Gewest worden vaak volgende elementen gegeven ter verklaring: een belangrijke mismatch tussen de arbeidsmarkt in het gewest (gekenmerkt door vaak hooggekwalificeerde jobs en een groot aantal pendelaars van buiten het gewest) en het profiel van de lokale bevolking.

Het effect van de vergoedingscategorie in Tabel 2 is duidelijk: samenwonenden hebben de hoogste uitstroomkansen terwijl gezinshoofden de laagste kansen hebben. Alleenstaanden positioneren zich op een iets hoger niveau dan de gezinshoofden. Verderop zal blijken dat het feit dat de rechthebbende een werkende partner heeft ook de kansen op uitstroom verhoogt.

Ondanks het feit dat samenwonenden kunnen rekenen op de inbreng van een andere kostwinner voor het gezinsinkomen, zien we toch een grotere uitstroomkans

wat waarschijnlijk te maken heeft met de lagere uitkeringen voor samenwonenden. Mogelijk hebben zij door het feit dat zij er niet alleen voor staan in het huishouden ook meer faciliteiten om uit te gaan werken. Bovendien is het zo dat een aantal van de samenwonenden inwonende kinderen zijn (zie verder).

In de onderstaande tabel gaat het over een mannelijke rechthebbende. Omdat er een interactie is tussen het effect van de vergoedingscategorie en het geslacht bekijken we dit verderop opnieuw.

| studie-niveau | vergoedings-categorie | werkende partner | leeftijd (klasse) | kans per gewest | | |
|---------------|-----------------------|------------------|-------------------|-----------------|-------|----------|
| | | | | Vlaams | Waals | Brussels |
| laag | samenwonend | ja | ≤ 25 | 61,7% | 47,9% | 44,3% |
| laag | alleenstaand | neen | ≤ 25 | 43,4% | 30,5% | 27,5% |
| laag | gezinshoofd | neen | ≤ 25 | 43,3% | 30,4% | 27,4% |
| laag | gezinshoofd | neen | $> 25, \leq 35$ | 40,1% | 27,7% | 24,9% |
| laag | gezinshoofd | neen | $> 35, \leq 50$ | 31,5% | 20,8% | 18,5% |
| laag | gezinshoofd | neen | > 50 | 17,1% | 10,5% | 9,3% |

Tabel 2: Effect van de vergoedingscategorie en de leeftijd

Het effect van de leeftijd, eveneens in Tabel 2, is ook duidelijk: naarmate de leeftijd stijgt verminderen de kansen. Opvallend is dat het verschil tussen de klassen jonger dan 25 en de klasse tussen 25 en 35 klein is (het kan financieel interessant zijn een jongere aan te werven terwijl een werknemer tussen 25 en 35 waarschijnlijk meer ervaring heeft en productiever is). Tussen 35 en 50 liggen de kansen al merkkelijk lager en boven de 50 zelfs zeer laag.

In eerste instantie kan men het effect van het geslacht apart onderzoeken. In dát geval blijkt dat de vrouwen systematisch een iets lagere uitstroomkans hebben. Aangezien we vermoeden dat dit gerelateerd kan worden aan de gezinssituatie bekijken we dit liever samen met de vergoedingscategorie, het blijkt dan dat het effect van de vergoedingscategorie afhangt van het geslacht.

In Tabel 3 komt het interactie effect tussen het geslacht en de vergoedingscategorie duidelijk naar voor: bij de vrouwen betekent een positie gezinshoofd een veel lagere uitstroomkans dan bij de mannen (er is bij de mannen nauwelijks verschil tussen de posities alleenstaand en gezinshoofd), een samenwonende vrouw heeft ook een iets lagere uitstroomkans dan een samenwonende man. Uit kruistabellen blijkt ook duidelijk dat vrouwelijke gezinshoofden, relatief vaker dan mannelijke gezinshoofden, kinderen ten laste hebben (respectievelijk 82,5% en 55%).

| studie-niveau | vergoedings-categorie | leeftijd (klasse) | werkende partner | geslacht | kans per gewest | | |
|---------------|-----------------------|-------------------|------------------|----------|-----------------|-------|----------|
| | | | | | Vlaams | Waals | Brussels |
| hoog | gezinshoofd | $> 25, \leq 35$ | neen | man | 63,9% | 50,4% | 46,8% |
| hoog | samenwonend | $> 25, \leq 35$ | ja | man | 78,9% | 68,1% | 64,9% |
| hoog | alleenstaand | $> 25, \leq 35$ | neen | man | 64,1% | 50,5% | 46,9% |
| hoog | gezinshoofd | $> 25, \leq 35$ | neen | vrouw | 53,6% | 39,8% | 36,4% |
| hoog | samenwonend | $> 25, \leq 35$ | ja | vrouw | 74,5% | 62,5% | 59,1% |
| hoog | alleenstaand | $> 25, \leq 35$ | neen | vrouw | 65,3% | 51,9% | 48,3% |

Tabel 3: Interactie effect tussen geslacht en vergoedingscategorie

Naast de boven reeds besproken effecten kunnen we het nog hebben over bijkomende effecten die uitgaan van het arbeidsverleden, de hoogte van het dagbedrag, de nationaliteit en de familiale situatie.

Zoals verwacht speelt ook de toelaatbaarheidsbasis een rol: toelaatbaarheid na arbeid geeft hogere uitstroomkansen dan toelaatbaarheid na studies, zie Tabel 4. We moeten wel opmerken dat de leeftijd een belangrijkere rol kan spelen. Aangezien meer dan 90% van de werklozen toelaatbaar na studies jonger dan 25 zijn hebben zij over het algemeen vrij hoge kansen. Als men echter werklozen van dezelfde leeftijdsgroep vergelijkt geeft toelaatbaarheid na arbeid steeds een hogere uitstroomkans, wat te verwachten is gezien het gaat om mensen met werkervaring die bovendien relatief recent is.

| studie-niveau | vergoedings-categorie | leeftijd (klasse) | geslacht | arbeids-verleden | kans per gewest | | |
|---------------|-----------------------|-------------------|----------|------------------|-----------------|-------|----------|
| | | | | | Vlaams | Waals | Brussels |
| Midden | alleenstaand | ≤ 25 | man | na studies | 45,0% | 31,9% | 28,8% |
| Midden | alleenstaand | ≤ 25 | man | na arbeid | 52,8% | 39,0% | 35,6% |

Tabel 4: Effect van het arbeidsverleden

Voor wat betreft de nationaliteit valt op dat de Belgen een duidelijk hogere uitstroomkans hebben dan niet-belgen, zie Tabel 5. Het blijkt dat er geen groot verschil bestaat tussen EU onderdanen (buiten België) en andere buitenlanders. Wél kunnen we veronderstellen dat er onder de groep buitenlanders heel uiteenlopende situaties kunnen voorkomen wat de bespreking bemoeilijkt. In elk geval kunnen we opmerken dat de buitenlanders (zeker buiten de EU) qua studieniveau minder goed scoren dan de belgen. Dit blijkt ook uit Tabel 6.

| studie-niveau | vergoedings-categorie | leeftijd (klasse) | geslacht | nationaliteit | kans per gewest | | |
|---------------|-----------------------|-------------------|----------|---------------|-----------------|-------|----------|
| | | | | | Vlaams | Waals | Brussels |
| Midden | Samenwonend | $> 25, \leq 35$ | vrouw | Belg | 61,6% | 47,9% | 44,3% |
| Midden | Samenwonend | $> 25, \leq 35$ | vrouw | EU | 53,0% | 39,2% | 35,8% |
| Midden | Samenwonend | $> 25, \leq 35$ | vrouw | overig | 51,7% | 38,0% | 34,7% |

Tabel 5: Effect van de nationaliteit

| | laag | midden | hoog | |
|-----------------|--------|--------|--------|---------|
| België | 13,55% | 62,89% | 23,56% | 100,00% |
| Overige EU | 23,94% | 54,77% | 21,29% | 100,00% |
| Overige niet EU | 39,21% | 49,15% | 11,65% | 100,00% |

Tabel 6: Studieniveau naar nationaliteit

Meer diepgaand onderzoek naar de specifieke situatie van allochtonen is gebeurd door Okkerse en Termote (2004). Hierin komt onder meer naar voor dat de allochtone vrouwen een veel lagere activiteitsgraad hebben dan de allochtone mannen. Uit het onderzoek van Vertommen en Martens (2005) is gebleken dat genaturaliseerde belgen van allochtone origine vaak nog een inferieure positie op de arbeidsmarkt bekleden.

Voor wat betreft de familiale situatie zijn er mogelijk belangrijke effecten in verband met het hebben van een werkende partner en het effect van het hebben van kinderen. De positie die de rechthebbende inneemt in het gezin wordt verderop besproken.

Het is uiteraard zo dat er een verband bestaat tussen de vergoedingscategorie en het hebben van een werkende partner. Toch moeten we voorzichtig zijn: terwijl de overgrote meerderheid van de werklozen bij wie een werkende partner bekend is ook de vergoedingscategorie samenwonend hebben (meer dan 91%), geldt omgekeerd niet dat de meeste samenwonenden een werkende partner hebben (slechts 35,0% van de werklozen in vergoedingscategorie samenwonend heeft een werkende partner, overigens hebben heel wat samenwonenden de gezinspositie kind, zo'n 50,7%).

Het is dus zinvol het effect van een werkende partner als apart effect te beschouwen, en er lijkt een duidelijk positief effect uit te gaan van het hebben van een werkende partner. Enerzijds gaat het dus om mensen die waarschijnlijk het financieel minder gunstige statuut van samenwonende hebben, anderzijds betekent het hebben van een werkende partner dat de betrokkene waarschijnlijk als partner in een gezin met of zonder kinderen meer verantwoordelijk is voor het gezinsinkomen (in vergelijking met de situatie dat de betrokkene bijvoorbeeld een inwonend kind is).

Voor wat betreft het hebben van kinderen is het effect afhankelijk van de leeftijd van die kinderen, zoals getoond in Tabel 7. Waar het hebben van zeer jonge kinderen de uitstroomkans lijkt te verminderen, zal het hebben van oudere kinderen (3-12 jaar) deze eerder verhogen (hoewel dit laatste niet significant is), steeds ten opzichte van de situatie dat er geen kinderen zijn. Om het voorbeeld realistisch te houden is de leeftijd van de moeder ook een klasse verhoogd naarmate de kinderen ouder worden.

| studie-niveau | vergoedings-categorie | leeftijd (klasse) | geslacht | leeftijd jongste kind | kans per gewest | | |
|---------------|-----------------------|-------------------|----------|-----------------------|-----------------|----------|-------|
| | | | | | Vlaams | Brussels | Waals |
| Midden | samenwonend | $> 25, \leq 35$ | vrouw | geen | 61,6% | 47,9% | 44,3% |
| Midden | samenwonend | $> 25, \leq 35$ | vrouw | 0-2 | 55,5% | 41,6% | 38,2% |
| Midden | samenwonend | $> 25, \leq 35$ | vrouw | 3-11 | 62,2% | 48,5% | 44,9% |
| Midden | samenwonend | $> 35, \leq 50$ | vrouw | 12-17 | 57,1% | 43,2% | 39,7% |
| Midden | samenwonend | $> 35, \leq 50$ | vrouw | 18-24 | 51,9% | 38,2% | 34,9% |

Tabel 7: Effect van kinderlast in het gezin, volgens leeftijd jongste kind

Rest nog één enkele determinant te bespreken: de hoogte van het dagbedrag. Het is duidelijk dat deze gedeeltelijk overlapt met andere kenmerken en het is ook moeilijk een objectieve indeling te vinden. Desalniettemin is er een significant positief effect dat uitgaat van de categorie van het hoogste dagbedrag, zie Tabel 8. We kunnen veronderstellen dat de hoogte van het dagbedrag de hoogte van het loonplafond weerspiegelt (eerste vergoedbaarheidsperiode).

| studie-niveau | vergoedings-categorie | leeftijd (klasse) | geslacht | dagbedrag (klasse) | kans per gewest | | |
|---------------|-----------------------|-------------------|----------|--------------------|-----------------|-------|----------|
| | | | | | Vlaams | Waals | Brussels |
| Midden | alleenstaand | $> 35, \leq 50$ | man | laagste | 38,8% | 26,7% | 24,0% |
| Midden | alleenstaand | $> 35, \leq 50$ | man | midden | 40,4% | 28,0% | 25,3% |
| Midden | alleenstaand | $> 35, \leq 50$ | man | hoogste | 48,2% | 34,8% | 31,7% |

Tabel 8: Effect van de hoogte van het dagbedrag

Dit kunnen we als volgt interpreteren: wie een hoog loon had staat waarschijnlijk sterker op de arbeidsmarkt (hoger opgeleid, een betere kwalificatie, werkervaring) en zal een grotere daling van het inkomen voelen wat een bijkomende stimulans is om terug te keren naar een werkende positie.

2.2 Bespreking van het alternatief model

Bij de modelselectie is gebleken dat de voor handen zijnde variabele "huishoudpositie" best niet opgenomen wordt in het model, aangezien de informatie grotendeels in andere variabelen vervat zit, onder meer in de vergoedingscategorie. Nochtans is het interessant een apart model te schatten waarin de huishoudpositie wél een belangrijke rol kan spelen door de vergoedingscategorie weg te laten alsook de gegevens over partner en kinderen.

Tabel 9 is gebaseerd op dit alternatief model (meer details in Sectie 4.3) dat iets minder goed scoort qua waarschijnlijkheid en verklaringskracht maar wel het voordeel heeft een interpretatie naar huishoudpositie mogelijk te maken. Voor de selectie van de variabelen en parameterschattingen verwijzen we naar Tabel 15 en Tabel

16 respectievelijk. Het ligt voor de hand hier ook de interactie tussen geslacht en huishoudpositie te modelleren.

| studie-niveau | huishoud-positie | leeftijd (klasse) | geslacht | kans per gewest | | |
|---------------|-------------------------|-------------------|----------|-----------------|-------|----------|
| | | | | Vlaams | Waals | Brussels |
| midden | alleenstaand | $> 25, \leq 35$ | man | 54,7% | 40,2% | 36,3% |
| midden | kind | ≤ 25 | man | 55,9% | 41,3% | 37,4% |
| midden | hoofd eenoudergezin | $> 25, \leq 35$ | man | 45,9% | 32,1% | 28,6% |
| midden | partner met kinderen | $> 25, \leq 35$ | man | 68,1% | 54,3% | 50,1% |
| midden | partner zonder kinderen | $> 25, \leq 35$ | man | 66,9% | 52,9% | 48,7% |
| midden | alleenstaand | $> 25, \leq 35$ | vrouw | 55,4% | 40,8% | 36,9% |
| midden | kind | ≤ 25 | vrouw | 56,0% | 41,4% | 37,5% |
| midden | hoofd eenoudergezin | $> 25, \leq 35$ | vrouw | 45,2% | 31,5% | 28,0% |
| midden | partner met kinderen | $> 25, \leq 35$ | vrouw | 54,9% | 40,4% | 36,5% |
| midden | partner zonder kinderen | $> 25, \leq 35$ | vrouw | 57,2% | 42,7% | 38,7% |

Tabel 9: Effect van de huishoudpositie

Wat opvalt is de lage uitstroomkans voor de positie "hoofd eenoudergezin", en dit zowel voor mannen als voor vrouwen (we merken wel op dat in 86% van de eenoudergezinnen het gezinshoofd een vrouw is). Voor alleenstaanden is er ook geen groot verschil tussen mannen en vrouwen. Opvallend is wel dat de vrouwelijke partners met of zonder kinderen een gelijkaardige uitstroomkans hebben terwijl die bij de mannen een heel stuk hoger ligt. Bij mannen lijkt het hebben van kinderen de kans iets te verhogen, bij vrouwen zal die een stukje verlagen. Dit alles wijst op eerder traditionele patronen. Tenslotte, de kinderen hebben een iets hogere uitstroomkans dan alleenstaanden, maar merk op dat in het voorbeeld een lagere leeftijdsklasse genomen is.

3 Conclusies en verder onderzoek

3.1 Algemene besluiten

Hoewel de onderzoeksvraag niet nieuw is blijft ze actueel en het is voor het eerst, dank zij het project van het RVA panel, dat de RVA zelf over data beschikt om dit soort onderzoek te doen.

De conclusies liggen meestal in de lijn van bekende feiten, maar blijven vermeldenswaard:

- Er zijn belangrijke verschillen tussen de gewesten. De kansen liggen het hoogst in het Vlaams gewest, in het Brussels Hoofdstedelijk Gewest liggen die het laagst.

- Het studieniveau komt naar voor als belangrijke determinant. Hoe hoger opgeleid hoe beter de kansen. Uiteraard blijft het zo dat mensen met een veelgevraagde kwalificatie ook zeer goede kansen hebben.
- De vergoedingscategorie speelt een belangrijke rol: de samenwonenden hebben de hoogste uitstroomkansen, de gezinshoofden de laagste.
- De kansen verminderen met de leeftijd, maar dit is zeker niet het belangrijkste effect.

Het blijft uiteraard moeilijk om alle effecten te ontwarren, vaak spelen de effecten op mekaar in. Zo hebben we vastgesteld dat het effect van de vergoedingscategorie voor mannen en vrouwen verschillend is: vrouwelijke gezinshoofden hebben een veel lagere uitstroomkans mannelijke en een samenwonende vrouw heeft ook een iets lagere uitstroomkans dan een samenwonende man.

Die interactie was ook relevant in ons onderzoek naar de huishoudpositie: in de huishoudpositie "partner" hebben de vrouwen duidelijk lagere uitstroomkansen dan de mannen, het hebben van kinderen verhoogt de uitstroomkansen voor de mannen maar verlaagt die bij de vrouwen. Wie aan het hoofd staat van een eenoudergezin heeft ook een duidelijk lagere uitstroomkans.

Men mag dus niet zomaar zeggen dat er een zuiver effect is van het geslacht, dit moet eerder in verband gebracht worden met de gezinssituatie.

Het hebben van een werkende partner is steeds positief. Overige effecten zijn nog:

- Het netto effect van het hebben van een arbeidsverleden is positief.
- Belgen hebben betere kansen dan niet-belgen maar er lijkt geen groot verschil tussen EU en niet-EU nationaliteiten (globaal genomen).
- De hoogte van het dagbedrag heeft een positief effect, dit lijkt dus de uitstroom niet te belemmeren.

3.2 Verder onderzoek

In deze studie is gepoogd een algemeen beeld te geven van de factoren die bepalen of de populatie van de nieuw ingestroomde werklozen in 2007 één jaar na de instroom aan het werk is. Allereerst is het de bedoeling de analyse te herhalen met meer recente gegevens zodra die beschikbaar zijn.

Anderzijds kan de methodologie verfijnd worden: zoals uitgelegd is in Sectie 1.2.2 wordt in ons model slechts rekening gehouden met de positie werkend (of niet) op één tijdstip (namelijk na één jaar) terwijl die positie gekend is op 8 verschillende tijdstippen. Er bestaan discrete duurtijd modellen (ook gebruikt in de biostatistiek

in het kader van survival analysis) die rekening houden met de gegevens gedurende de 8 kwartalen dat de persoon opgevolgd wordt. Bovendien houden deze rekening met het feit dat de uitstroom mogelijk niet geobserveerd wordt (censoring). Voor meer details, zie bijvoorbeeld Singer en Willett (2007).

Uiteraard geeft een studie niet alleen antwoorden, of elementen van antwoord, maar roept zij even goed nieuwe vragen op. Zo kunnen we ons afvragen hoe het zit met de kwaliteit en duurzaamheid van de betrekking (sector, salaris, ...). In Sectie 1.2.2 is immers gebleken dat heel wat mensen een grillige loopbaan hebben.

We kunnen ook de populatie verruimen en ook werklozen opvolgen die mogelijk al een zekere voorgeschiedenis hebben. Hier zal de duur als belangrijke factor naar voor komen. Het in rekening brengen hiervan brengt enkele specifieke complicaties met zich mee, zie bijvoorbeeld Heylen (2010) voor een bespreking hiervan.

Een voor de hand liggend verder onderzoek betreft het opsplitsen van de analyse van de studie per geslacht en gewest.

4 Technische en methodologische appendix

4.1 RVA panel steekproeftrekking

Het RVA panel is gebaseerd op een steekproef uit de populatie van alle RVA uitkeringstrekkers. Voor die populatie worden bijkomende gegevens over sociale zekerheid, tewerkstelling en gezinsgegevens toegevoegd uit het datawarehouse Arbeidsmarkt en Sociale Bescherming van de Kruispuntbank Sociale Zekerheid.

Zonder in alle details te gaan is het nuttig een korte beschrijving te geven van het steekproefmechanisme omdat dit toelaat de methodologie van deze studie beter te begrijpen.

Allereerst geldt het principe dat alleen RVA begunstigen in het panel opgenomen kunnen worden. Dit kan op verschillende momenten:

- Eerste trimester van 2000 (initiële steekproef).
 - 33,33 % toevalssteekproef uit de *stock* van alle RVA begunstigen.
 - Gestratificeerd volgens arrondissement en RVA statuut (hergroepering).
- Een later trimester, vanaf het tweede trimester van 2000 (we spreken van een *bijtrekking*).
 - 25% toevalssteekproef uit de *flow* (nieuwe instroom: personen die in dát trimester RVA begunstigde waren, maar niet bekend waren bij de RVA in de periode vóór dat trimester (voorgeschiedenis vanaf 2000 gedurende maximum 5 jaar).
 - Gestratificeerd volgens gewest en RVA statuut.

Ten tweede geldt dat elke persoon in de steekproef minimum 5 jaar opgevolgd wordt en minimum 36 maanden ná het verbreken van de band met de RVA. Na deze periode zal de persoon uitstromen uit het panel (maar deze kan eventueel wel later in aanmerking komen voor een bijtrekking), en dit gebeurt tevens als de persoon zou overlijden of de leeftijd van 65 jaar bereikt.

De opvolging betekent dat per kwartaal (of jaarlijks indien de gegevens niet per kwartaal beschikbaar zijn) gegevens ingevuld worden vanaf de instroom tot zo recent mogelijk als voorhanden in het datawarehouse.

De RVA rechthebbende is de sleutelpersoon waaraan alle gegevens gerelateerd worden via een geanonimiseerd (fictief) nummer dat gebruikt wordt als sleutelvariabele.

Om de gezinssituatie te kunnen beschrijven wordt de methodologie van het datawarehouse gevolgd: een gezin wordt verondersteld gevormd te zijn rond de identiteit

van de zogenaamde referentiepersoon voor het gezin (als zodanig gekend door de administratieve overheid). We kennen de familieband van de RVA begunstigde ten opzichte van deze referentiepersoon (samengevat in de zogenaamde "huishoudpositie") maar de familiebanden met de andere gezinsleden blijven onbekend. In de meeste gevallen kunnen we door voldoende gegevens te combineren echter wel een goed beeld hebben van de socio-economische achtergrond van de betrokkene. Naast de referentiepersoon van het gezin wordt in het datawarehouse ook afgeleid, weerom op basis van geregistreerde administratieve gegevens, of de RVA begunstigde een partner heeft.

Om het verband te leggen met ons onderzoek merken we op:

- Onze onderzoekspopulatie is gebaseerd op de RVA begunstigten van de 4 *bijtrekkingen* van 2007.
- We beperken ons tot RVA begunstigten in een statuut uitkeringsgerechtigd volledig werkloos en werkzoekend (in de nomenclatuur, zie verder).

Niet alle variabelen in het RVA panel zijn relevant voor deze studie. In de volgende paragraaf geven we een overzicht van de gebruikte variabelen.

4.2 Beschrijving variabelen voor de logistische regressie

Alle gebruikte variabelen zijn afkomstig uit het Datawarehouse Arbeidsmarkt en Sociale Bescherming, of daarvan afgeleid. Voor details daarover verwijzen we naar de relevante documentatie (zie bijvoorbeeld website van de KSZ, <http://www.ksz-bcss.fgov.be/>)

Hoewel het mogelijk is met continue variabelen te werken geven we er de voorkeur aan alles in te delen in categorieën (minder strenge veronderstelling in het logistisch regressiemodel).

In Tabel 10 staan alle mogelijke verklarende variabelen alsook de binaire verklaarde variabele "uitstroom" opgelijst, samen met hun eventuele indeling in categorieën.

De nomenclatuur van de socio-economische positie zelf staat niet in de lijst maar is de cruciale variabele in het panel omdat het een vrij gedetailleerde samenvatting geeft van de socio-economische positie van de persoon op de laatste dag van het kwartaal (met hoofdindeling "werkend", "werkzoekend", "inactief" en "ander"). Hieruit wordt afgeleid of iemand werkzoekend of werkend is. Op basis van de nomenclatuur kunnen we dus de uitstroom bepalen na 4 kwartalen en ook of de eventuele partner werkend is. Dit vereist wel dat de relatie tussen de partners officieel is en geregistreerd is. De betrouwbaarheid van dit gegeven zal dus niet optimaal zijn. Aangezien er geen significante verschillen bleken te zijn tussen

| Concept en oorsprong | Variabele | Waarden |
|--|-----------------------|--|
| Vergoedingscategorie Fiche61 (RVA) hergroepering | "vergcatt" | Alleenstaand Samenwonend Gezinshoofd |
| Toelaatbaarheidsbasis (arbeidsverleden) Fiche7 (RVA) hergroepering | "arbeidsverleden" | 1: Na arbeid 0: Na studies |
| Hoogte dagbedrag Dgndmnd_klasse (RVA) indeling | "bedrk" | 0: laagste (\leq 25ste percentiel) 1: middengroep 2: hoogste ($>$ 75ste percentiel) |
| Woonplaats (gewest) Arrondissement (rijksregister) hergroepering | "gewest" | B: Brussels Hoofdstedelijk Gewest V: Vlaams Gewest W: Waals Gewest |
| Geslacht (rijksregister) | "geslacht" | 0: Man 1: Vrouw |
| Leeftijd (in jaren) indeling | "leeftk4" | 0: ≤ 25 1: $> 25, \leq 35$ 2: $> 35, \leq 50$ 3: > 50 |
| Nationaliteit hergroepering landcode (rijksregister) | "nat" | 0: Belgen 1: EU buiten België 2: Overige landen |
| Huishoudpositie (t.o.v. referentiepersoon) (rijksregister) LIPRO | "hhpos" | Alleenstaand Kind Hoofd eenoudergezin Partner met kinderen Partner zonder kinderen |
| Leeftijd jongste kind (jaren) klassen | "jongste_kind_leeftk" | 0: Geen of ouder dan 25 1: 0-2 2: 3-11 3: 12-17 4: 18-24 |
| Werkende partner nomenclatuur en rijksregister | "Wptn" | 0: Partner onbekend/ niet-werkend 1: Partner werkend |
| Uitstroom (4 kwartalen na de instroom) | "Wout" | 0: niet-werkend 1: werkend |
| Studieniveau (eerste van 7 posities) | "stuck" | 0: Laag (lager + eerste graad sec.) 1: Midden (secundair onderwijs) 2: Hoog (hoger onderwijs) 3: (leercontracten) |

Tabel 10: Lijst van variabelen

de situaties waar geen partner aanwezig en de gevallen dat de partner een positie "werkzoekend", "inactief" of "ander" inneemt wordt alleen het onderscheid gemaakt tussen de situatie "er is een werkende partner" en alle andere gevallen (daarom definiëren we een binaire variabele "Wptn").

In principe worden de verklarende variabelen bepaald op het ogenblik dat de persoon instroomt als werkloze. Gaat het over maandgegevens (zoals RVA) dan wordt

de laatste maand van het kwartaal beschouwd, gaat het over rijksregistergegevens dan wordt de toestand aan het eind van 2007 gerekend.

Zoals hierboven opgemerkt moet men voorzichtig zijn met het interpreteren van de gegevens over de gezinssituatie. Indien de rechthebbende de huishoudpositie "partner met kinderen" inneemt kunnen we veronderstellen dat de kinderen aanwezig in het gezin ten laste zijn van de rechthebbende (wetende dat dit in de meeste gevallen zal kloppen). Indien de rechthebbende echter zélf bekend staat als "kind" (ten opzichte van de referentiepersoon) mogen we er niet van uitgaan dat de jonge kinderen in het gezin ten laste zouden zijn van de rechthebbende, vaak zal het gaan om broertjes of zusjes. Indien de rechthebbende zelf de positie kind inneemt zullen we de hypothese maken dat de kinderen in het gezin in regel niet ten laste van de rechthebbende zijn, tenzij die persoon in de vergoedingscategorie "gezinshoofd" zit (zo'n 300 gevallen). (Concreet: variabele "jongste_kind_leeftk" altijd op 0 gezet indien "hhpos"="kind", tenzij "vergc"="gezinshoofd")

4.3 Modelselectie

In de lijst van mogelijke verklarende variabelen is er geen enkele die we a priori kunnen uitsluiten. Wél is het zo dat er duidelijk overlappingen zijn, zo bijvoorbeeld levert de huishoudpositie niet zo veel nieuws als we beschikken over vergoedingscategorie, de aanwezigheid van kinderen in het gezin en het feit of een werkende partner aanwezig is of niet.

Indien deze vier variabelen samen voorkomen in een model blijkt het dat de huishoudpositie niet meer significant is. Daarom wordt in een basismodel de variabele "hhpos" weggelaten. Deze wordt wél in een alternatief model bekeken.

De criteria die gehanteerd worden om modellen te evalueren zijn de *deviance* (statistische waarschijnlijkheid van het model, zie bijv. Agresti (2003)) en de kwaliteit van de classificatie van het model. Deze classificatie gebeurt als volgt: we gebruiken de geschatte coëfficiënten om per case de uitstroomkans te schatten en indien deze (strikt) groter is dan de globale uitstroomkans voorspelt het model een uitstroom, indien deze lager is voorspelt het model een niet-uitstroom (met uitstroom bedoelen we steeds deze *naar werk*). De kwaliteit van het model blijkt dan uit de mate waarin de voorspelde uitkomsten overeenstemmen met de werkelijk geobserveerde.

We kunnen dan het percentage correct voorspelde uitkomsten, de specificiteit (% correct voorspelde niet-uitstroom), de sensitiviteit (% correct voorspelde uitstroom), het percentage vals positieve (% ten onrechte voorspelde uitstroom) en het percentage vals negatieve (% ten onrechte voorspelde niet-uitstroom) voorspellingen berekenen.

Eens de hoofdeffecten vastliggen blijkt het toevoegen van interacties niet zozeer de voorspellingswaarde van het model maar wél de waarschijnlijkheid van het

model te verbeteren. Het is mogelijk de software automatisch te laten zoeken naar interacties maar dit levert een model dat praktisch onleesbaar wordt qua interpreteerbaarheid. We geven er de voorkeur aan één significante interactie toe te voegen die ook een duidelijke interpretatie heeft: de interactie tussen geslacht en vergoedingscategorie. Enkele andere voor de hand liggende interacties zijn deze tussen huishoudpositie en leeftijd jongste kind en tussen nationaliteit en gewest (deze bleken niet zeer sterk te zijn).

4.3.1 Basismodel

In het basismodel worden volgende variabelen ingevoerd (we geven de referentie-categorie tussen haakjes):

gewest ("Vlaams Gewest"), studieniveau ("laag"), vergoedingscategorie ("alleenstaand"), leeftijd (≤ 25), nationaliteit ("belg"), hoogte dagbedrag ("laagste"), werkende partner ("geen"), leeftijd jongste kind ("geen"), geslacht ("man"), arbeidsverleden ("na studies").

De *deviance* geeft de mate van waarschijnlijkheid van het effect. Hoe hoger de deviance, hoe belangrijker het effect. Deze informatie zit in de *analysis of variance* tabel, zie Tabel 11. Op basis van deze tabel blijkt dat alle variabelen een significant effect hebben. Dit wordt weergegeven met de codes in de laatste kolom die staan voor het significantieniveau: $0 \leq \text{'***'} < 0.001 \leq \text{'**'} < 0.01 \leq \text{'*'} < 0.05 \leq \text{'.'} < 0.1$

| | Df | Deviance | Resid. Df | Resid. Dev | $P(> Chi)$ | |
|-----------------------------|----|----------|-----------|------------|--------------|-----|
| NULL | | | 14577 | 19865 | | |
| factor(gewest) | 2 | 404.83 | 14575 | 19460 | < 2.2e-16 | *** |
| factor(leeftk4) | 3 | 224.26 | 14572 | 19235 | < 2.2e-16 | *** |
| geslacht | 1 | 14.77 | 14571 | 19221 | 0.0001212 | *** |
| factor(bedrk) | 2 | 30.84 | 14569 | 19190 | 2.011e-07 | *** |
| factor(vergcat) | 2 | 283.09 | 14567 | 18907 | < 2.2e-16 | *** |
| factor(nat) | 2 | 71.21 | 14565 | 18836 | 3.451e-16 | *** |
| arbeidsverleden | 1 | 7.05 | 14564 | 18828 | 0.0079084 | ** |
| factor(stuck) | 3 | 294.73 | 14561 | 18534 | < 2.2e-16 | *** |
| Wptn | 1 | 30.14 | 14560 | 18504 | 4.018e-08 | *** |
| factor(jongste_kind.leeftk) | 4 | 22.30 | 14556 | 18481 | 0.0001744 | *** |
| geslacht:factor(vergcat) | 2 | 16.64 | 14554 | 18465 | 0.0002435 | *** |

Tabel 11: ANOVA tabel voor basismodel.

Qua interpretatie zijn we vooral geïnteresseerd in de parameterschattingen, samengevat in Tabel 12.

Alvorens deze te bekijken brengen we het algemene logistische regressie model in herinnering:

$$\log\left(\frac{p}{1-p}\right) = \alpha + \beta_1 x_1 + \dots + \beta_k x_k$$

Hierin staat p voor de verwachtingswaarde voor de uitstroom naar werk (variabele "Wout"), beschouwd als binaire toevalsvariabele. Het linkerlid van de vergelijking is de *logit* transformatie van p en de basisonderstelling is dat deze gehoorzaamt aan het lineair model in het rechterlid. Dit lineair model is een som van een constante term (of *intercept*) en termen $\beta_i x_i$ waar β_i de coëfficiënt is die hoort bij het effect x_i . De schattingen voor de α en voor de β_i vinden we in de tweede kolom van de tabel. De intercept geeft een schatting voor α , de overige waarden stemmen overeen met de β_i . Voor binaire variabelen staat er één coëfficiënt die hoort bij de waarde 1 (waarde 0 is de referentie), voor categorische variabelen met meer dan twee categorieën hoort een coëfficiënt bij elke waarde verschillend van de referentie categorie. Het teken (+ of $-$) van de coëfficiënt geeft aan of de categorie aanleiding geeft tot een hogere, resp. een lagere logit (en dus kans). Hoe kleiner de standaardfout (derde kolom), hoe beter de significantie (hoe lager de kans in de vijfde kolom) en hoe sterker het effect. Dit wordt weergegeven met de codes in de laatste kolom die staan voor het significantieniveau, met een code zoals boven.

| | Estimate | Std. Error | z value | $Pr(> z)$ | |
|-------------------------------------|-----------|------------|---------|-------------|-----|
| (Intercept) | -0.626718 | 0.091206 | -6.871 | 6.35e-12 | *** |
| factor(gewest)B | -0.702862 | 0.054511 | -12.894 | < 2e-16 | *** |
| factor(gewest)W | -0.558702 | 0.039613 | -14.104 | < 2e-16 | *** |
| factor(leeftk4)1 | -0.129889 | 0.059044 | -2.200 | 0.027815 | * |
| factor(leeftk4)2 | -0.507433 | 0.067229 | -7.548 | 4.43e-14 | *** |
| factor(leeftk4)3 | -1.309116 | 0.093739 | -13.966 | < 2e-16 | *** |
| geslacht | 0.055753 | 0.081495 | 0.684 | 0.493896 | |
| factor(bedrk)1 | 0.048225 | 0.076459 | 0.631 | 0.528216 | |
| factor(bedrk)2 | 0.366653 | 0.102735 | 3.569 | 0.000358 | *** |
| factor(vergcat)gezinshoofd | -0.004719 | 0.089224 | -0.053 | 0.957821 | |
| factor(vergcat)samenwonend | 0.428101 | 0.072959 | 5.868 | 4.42e-09 | *** |
| factor(nat)1 | -0.355707 | 0.075428 | -4.716 | 2.41e-06 | *** |
| factor(nat)2 | -0.406126 | 0.081452 | -4.986 | 6.16e-07 | *** |
| arbeidsverleden | 0.312315 | 0.078005 | 4.004 | 6.23e-05 | *** |
| factor(stuck)1 | 0.377522 | 0.054813 | 6.887 | 5.68e-12 | *** |
| factor(stuck)2 | 0.973637 | 0.063128 | 15.423 | < 2e-16 | *** |
| factor(stuck)3 | 0.448160 | 0.114261 | 3.922 | 8.77e-05 | *** |
| Wptn | 0.312938 | 0.052873 | 5.919 | 3.25e-09 | *** |
| factor(jongste_kind_leeftk)1 | -0.253844 | 0.067751 | -3.747 | 0.000179 | *** |
| factor(jongste_kind_leeftk)2 | 0.025561 | 0.072405 | 0.353 | 0.724063 | |
| factor(jongste_kind_leeftk)3 | 0.187604 | 0.096445 | 1.945 | 0.051752 | . |
| factor(jongste_kind_leeftk)4 | -0.018996 | 0.103630 | -0.183 | 0.854559 | |
| geslacht:factor(vergcat)gezinshoofd | -0.485615 | 0.126349 | -3.843 | 0.000121 | *** |
| geslacht:factor(vergcat)samenwonend | -0.303691 | 0.092354 | -3.288 | 0.001008 | ** |

Tabel 12: Parameterschattingen voor basismodel

Opvallend in Tabel 12 is het feit dat de coëfficiënt voor geslacht niet significant is. Dit is niet in tegenspraak met de gegevens van Tabel 11: er is immers een interactie-effect tussen geslacht en vergoedingscategorie en daar hebben we wél significante

coëfficiënten, bijvoorbeeld een negatieve voor vrouwelijke gezinshoofden.

In verband met de kwaliteit van de classificatie kunnen we volgende tabellen geven: Tabel 13 geeft de aantallen geobserveerde en voorspelde uitstroom of niet. Tabel 14 geeft de bovenvermelde percentages.

| effectief 1 | effectief 0 | voorspeld 1 | voorspeld 0 |
|-------------|-------------|-------------|-------------|
| 6170 | 8408 | 6834 | 7734 |

Tabel 13: Voorspellingen

| sensitiviteit | specificiteit | vals positief | vals negatief | correct voorspeld |
|---------------|---------------|---------------|---------------|-------------------|
| 61,5% | 63,9% | 44,4% | 30,6% | 62,9% |

Tabel 14: classificatie op basis van globaal uitstroompercentage 42.32%

4.3.2 Alternatief model

Zoals reeds gezegd bevat het alternatief model de gegevens over de huishoudpositie (variabele "hhpos", met referentiewaarde "alleenstaande") maar niet de variabelen "vergcst", "Wptn" noch "jongste_kind_leeftk". We beschouwen, zoals in het basismodel, wél het interactie-effect tussen geslacht en huishoudpositie.

Tabellen 15 en 16 geven respectievelijk de ANOVA tabel en de parameterschattingen voor dit model. Tabellen 17 en 18 bevatten de gegevens over de classificatie met dit model.

| | Df | Deviance | Resid. Df | Resid. Dev | $P(> Chi)$ | |
|-----------------------------|----|----------|-----------|------------|--------------|-----|
| NULL | | | 14577 | 19865 | | |
| factor(gewest) | 2 | 404.83 | 14575 | 19460 | <2.2e-16 | *** |
| factor(leeftk4) | 3 | 224.26 | 14572 | 19235 | < 2.2e-16 | *** |
| geslacht | 1 | 14.77 | 14571 | 19221 | 0.0001212 | *** |
| factor(bedrk) | 2 | 30.84 | 14569 | 19190 | 2.011e-07 | *** |
| factor(hhpos) | 4 | 121.26 | 14565 | 19069 | < 2.2e-16 | *** |
| factor(nat) | 2 | 91.72 | 14563 | 18977 | < 2.2e-16 | *** |
| arbeidsverleden | 1 | 49.40 | 14562 | 18927 | 2.088e-12 | *** |
| factor(stuck) | 3 | 329.32 | 14559 | 18598 | < 2.2e-16 | *** |
| factor(jongste_kind_leeftk) | 4 | 32.10 | 14555 | 18566 | 1.825e-06 | *** |
| geslacht:factor(hhpos) | 4 | 47.49 | 14551 | 18519 | 1.205e-09 | *** |

Tabel 15: ANOVA tabel voor alternatief model

| | Estimate | Std. Error | z value | Pr(> z) | |
|---|----------|------------|---------|-----------|-----|
| (Intercept) | -0.43537 | 0.08990 | -4.843 | 1.28e-06 | *** |
| factor(gewest)B | -0.75275 | 0.05443 | -13.829 | < 2e-16 | *** |
| factor(gewest)W | -0.58694 | 0.03956 | -14.836 | < 2e-16 | *** |
| factor(leeftk4)1 | -0.10911 | 0.06158 | -1.772 | 0.076430 | . |
| factor(leeftk4)2 | -0.54137 | 0.07000 | -7.734 | 1.04e-14 | *** |
| factor(leeftk4)3 | -1.41851 | 0.09887 | -14.347 | < 2e-16 | *** |
| geslacht | 0.02653 | 0.07962 | 0.333 | 0.738919 | |
| factor(bedrk)1 | -0.24013 | 0.06648 | -3.612 | 0.000304 | *** |
| factor(bedrk)2 | -0.08258 | 0.08663 | -0.953 | 0.340454 | |
| factor(hhpos)hoofd eenoudergezin | -0.35342 | 0.22755 | -1.553 | 0.120393 | |
| factor(hhpos)kind | -0.06168 | 0.09179 | -0.672 | 0.501604 | |
| factor(hhpos)partner met kinderen | 0.56884 | 0.12217 | 4.656 | 3.22e-06 | *** |
| factor(hhpos)partner zonder kinderen | 0.51369 | 0.08978 | 5.722 | 1.05e-08 | *** |
| factor(nat)1 | -0.41245 | 0.07545 | -5.467 | 4.59e-08 | *** |
| factor(nat)2 | -0.53558 | 0.08127 | -6.590 | 4.40e-11 | *** |
| arbeidsverleden | 0.55140 | 0.07227 | 7.630 | 2.36e-14 | *** |
| factor(stuck)1 | 0.42217 | 0.05468 | 7.720 | 1.16e-14 | *** |
| factor(stuck)2 | 1.04222 | 0.06309 | 16.519 | < 2e-16 | *** |
| factor(stuck)3 | 0.49819 | 0.11413 | 4.365 | 1.27e-05 | *** |
| factor(jongste_kind_leeftk)1 | -0.21294 | 0.11575 | -1.840 | 0.065826 | . |
| factor(jongste_kind_leeftk)2 | 0.04341 | 0.10026 | 0.433 | 0.665055 | |
| factor(jongste_kind_leeftk)3 | 0.17094 | 0.09487 | 1.802 | 0.071578 | . |
| factor(jongste_kind_leeftk)4 | 0.21727 | 0.08566 | 2.536 | 0.011200 | * |
| geslacht:factor(hhpos)hoofd eenoudergezin | -0.05378 | 0.23123 | -0.233 | 0.816073 | |
| geslacht:factor(hhpos)kind | -0.02163 | 0.09819 | -0.220 | 0.825682 | |
| geslacht:factor(hhpos)partner met kinderen | -0.58695 | 0.11020 | -5.326 | 1.00e-07 | *** |
| geslacht:factor(hhpos)partner zonder kinderen | -0.43753 | 0.12370 | -3.537 | 0.000404 | *** |

Tabel 16: Parameterschattingen voor alternatief model

| effectief 1 | effectief 0 | voorspeld 1 | voorspeld 0 |
|-------------|-------------|-------------|-------------|
| 6170 | 8408 | 6866 | 7712 |

Tabel 17: Voorspellingen

| sensitiviteit | specificiteit | vals positief | vals negatief | correct voorspeld |
|---------------|---------------|---------------|---------------|-------------------|
| 61,4% | 63,4% | 44,8% | 30,9% | 62,5% |

Tabel 18: classificatie op basis van globaal uitstroompercentage 42.32%

4.3.3 Typevoorbeeld berekening

Met de gegevens van het basismodel berekenen we de kans voor het eerste voorbeeld van Figuur 7. We geven in Tabel 19 een overzicht van alle kenmerken en plaatsen de bijhorende coëfficiënt in de laatste kolom: indien het kenmerk de referentie is wordt een nul gerekend. Als we de som maken van alle coëfficiënten én het

intercept bekomen we de logit van p en deze bedraagt -0.2709 ; tenslotte kunnen we p halen uit de omgekeerde transformatie: $p = \frac{\exp(-0.2709)}{1+\exp(-0.2709)} = 0.4327$

| Variabele | waarde | parameter |
|-----------------------|----------------|----------------|
| Vergoedingscategorie | Gezinshoofd | -0.0047 |
| Toelaatbaarheidsbasis | 1: Na arbeid | 0.3123 |
| Hoogte dagbedrag | 1: middengroep | 0.04822 |
| Gewest | Vlaams Gewest | 0 (referentie) |
| Geslacht | man | 0 (referentie) |
| Leeftijd | 0: ≤ 25 | 0 (referentie) |
| Nationaliteit | 0: Belg | 0 (referentie) |
| Leeftijd jongste kind | geen | 0 (referentie) |
| Werkende partner | geen | 0 (referentie) |
| Studieniveau | laag | 0 (referentie) |
| Intercept | | -0.6267 |
| logit | | -0.2709 |

Tabel 19: Gegevens voorbeeld en berekening

Referenties

- [1] A. Agresti, *An introduction to Categorical Data Analysis* (New York: Wiley, 2007).
- [2] V. Heylen, *Duration dependence in Flemish unemployment data* (Leuven, WSE Report, 2010).
- [3] L. Okkerse en A. Termote, *Hoe vreemd is vreemd op de arbeidsmarkt, over de allochtone arbeidskrachten in België* (NIS, Brussel: Statistische studie 111, 2004).
- [4] J. Singer en J. Willett, *Applied Longitudinal Data Analysis* (New York: Oxford University Press, 2003).
- [5] S. Vertommen en A. Martens, *Allochtone werknemers op lokale arbeidsmarkten* (Leuven, Over.Werk, Tijdschrift van het Steunpunt WAV, 2 (2005)).