

# Vetenschappelijke Raad voor het Regeringsbeleid

**W 26**  
**Determinanten van arbeidsmarktgedrag**

J. M. Bekkering  
J. S. Cramer  
J. A. Doornik

met dank aan M. Dols

Onderzoek in opdracht van

de Wetenschappelijke Raad voor het Regeringsbeleid en  
de Organisatie voor Strategisch Arbeidsmarktonderzoek

's-Gravenhage, februari 1988

Exemplaren van deze uitgave zijn te bestellen bij het  
Distributiecentrum Overheidspublikaties, Postbus 20014,  
2500 EA 's-Gravenhage, door overmaking van f 10,-- op  
giro 751 dan wel schriftelijk of telefonisch (070-789880)  
onder vermelding van titel en ISBN-nummer en het aantal  
gewenste exemplaren.

ISBN: 90 346 1492 1

Publikatie van de Wetenschappelijke Raad voor het  
Regeringsbeleid

TEN GELEIDE

In het kader van de voorbereiding van het WRR-rapport Activerend arbeidsmarktbeleid (december 1987) heeft de Stichting voor Economisch Onderzoek der Universiteit van Amsterdam een onderzoek gedaan naar de determinanten van arbeidsmarktgedrag. De resultaten van het onderzoek, die in februari 1987 zijn opgeleverd, zijn verwerkt in voornoemd rapport.

Prof.dr. W. Albeda

Voorzitter van de projectgroep

Toekomstige ontwikkelingen op de arbeidsmarkt

## VOORWOORD

Voor u ligt het verslag van een onderzoek verricht door de SEO, Stichting voor Economisch Onderzoek der Universiteit van Amsterdam, in opdracht van de Wetenschappelijke Raad voor het Regeringsbeleid en de Organisatie voor Strategisch Arbeidsmarktonderzoek. Het onderzoek betreft de determinanten van het individuele arbeidsmarktgedrag op lange termijn. Speciale aandacht wordt geschonken aan de wederzijdse beïnvloeding door huisgenoten. Het onderzoek is begeleid door een commissie, bestaande uit de volgende personen:

### voor de Wetenschappelijke Raad voor het Regeringsbeleid:

prof. dr. W. Albeda (voorzitter)

dr. W.J. Dercksen

dr. W. van Drimmelen

prof. drs. C.J. van Eijk

drs. F.R. Veeneklaas

drs. K. Vijlbrief (secretaris)

### voor de Organisatie voor Strategisch Arbeidsmarktonderzoek:

dr. ir. J.C. van Ours

## INHOUD

	pag.
1 SAMENVATTING EN CONCLUSIE	1
2 PROBLEEMSTELLING	7
3 OPZET VAN HET ONDERZOEK	11
3.1 Algemeen	11
3.2 De indeling in categorieën	13
3.3 Beperkingen en veronderstellingen	15
4 METHODE	19
4.1 Het logitmodel en enkele eigenschappen	19
4.2 Te verklaren en verklarende variabelen	23
5 ARBEIDSMARKTGEDRAG VAN INDIVIDUEN	29
5.1 Algemeen	29
5.2 Schattingsresultaten	32
5.3 Conclusie	40
6 ARBEIDSMARKTGEDRAG VAN GEZINNEN	43
6.1 Vraag en methode	43
6.2 Huishoudens met twee volwassen personen	46
6.3 Huishoudens met drie volwassen personen	49
6.4 Conclusie	52
LITERATUUR	54
APPENDIX A Het OSA-survey	A1
APPENDIX B Het econometrische model	A10
APPENDIX C Meting van interactie	A14

## 1 SAMENVATTING EN CONCLUSIES

### Opzet en uitkomst

In dit rapport worden twee vragen aan de orde gesteld en naar vermogen beantwoord. De eerste vraag is welke factoren de feitelijke positie op de arbeidsmarkt van een individu bepalen, en wel met name haar of zijn participatie. Deze vraag is de laatste jaren al talloze malen onderzocht, in het bijzonder voor de participatie van (gehuwde) vrouwen en voor het voorkomen van (langdurige) werkloosheid. Het onderhavige onderzoek verschilt in één opzicht van andere studies, en dat is dat het niet gaat om de situatie van het individu op één gegeven moment, maar om de karakteristiek van haar of zijn arbeidsmarktpositie in de afgelopen vijf jaar. Aldus hopen wij de factoren op het spoor te komen die de arbeidsmarktpositie op langere termijn van een individu bepalen, die maken dat sommigen voortdurend werk hebben en dat anderen blijvend van de arbeidsmarkt zijn afgesneden.

Het belangrijkste resultaat van deze analyse is dat het patroon van de lange termijn niet wezenlijk verschilt van dat van een momentopname: dezelfde factoren zijn aan het werk. Bijna alle mannen hebben voortdurend werk; blijvende inactiviteit komt weliswaar voor bij 4% van de steekproef, en meer in gebieden met een hoge werkloosheid, maar de relatie met persoonskenmerken is zwak. Eigenlijk heeft alleen het ontbreken van een partner een duidelijk effect op de kans op blijvende werkloosheid. Bij de vrouwen vinden wij daarentegen een veel lager percentage van participanten, en een veel groter variatie in de participatiegraad naar persoonskenmerken. Aanwezigheid van kinderen, een laag opleidingsniveau en voldoende andere inkomensbronnen in het gezin dragen er alle toe bij dat de vrouw niet werkt, en dan is er ook nog een generatie-effect.

Deze uitkomsten zijn niet verrassend. Zij betekenen voor het beleid (op het eerste gezicht) dat langdurige werkloosheid of voortdurende participatie geen andere oorzaken hebben en geen ander karakter dan de overeenkomstige

korte termijn verschijnselen. Langdurige afwezigheid van de arbeidsmarkt blijft niet minder ongewenst; de langdurige werkloosheid van mannen is uit sociaal oogpunt ongewenst, bij vrouwen gelden emancipatoire overwegingen. Volgens onze uitkomsten bestaat er echter geen aanleiding voor een beleid terzake uit te zien naar specifieke beleidsinstrumenten, afgestemd op een speciale doelgroep. Langdurige werkloosheid kan bestreden worden met dezelfde middelen als kortstondige werkloosheid, want zij heeft dezelfde oorzaken en treft dezelfde mensen. Deze conclusie is niet opzienbarend, maar desondanks niet zonder belang.

De tweede vraag waar wij aandacht aan besteden is of er binnen het huishouden onderlinge beïnvloeding van het individuele arbeidsmarktgedrag plaatsvindt, afgezien van voor de hand liggende inkomenseffecten. Het zou kunnen zijn dat afwezigheid van de arbeidsmarkt als het ware aanstekelijk werkt, dan wel juist compenserende reacties van huisgenoten oproept. Het een noch het ander, blijkt het geval te zijn: buiten de individuele kenmerken die het gedrag beïnvloeden is er geen afzonderlijk gezinseffect, noch in de ene noch in de andere richting. Het ware interessanter geweest als wij wèl een effect hadden gevonden, maar daar staat tegenover dat het resultaat het vertrouwen in de analyse van individuen versterkt. Voor het arbeidsmarktgedrag zijn de huiselijke omstandigheden alleen van belang voorzover zij al expliciet in het model zijn opgenomen: aanwezigheid van ander inkomen, aanwezigheid van een partner, zorgplicht voor een kind. Daarbuiten is er niets aan de hand.

#### Methode

Wij delen de bevolking in in drie categorieën die in een vijfjarige periode in verschillende mate toegang tot de arbeidsmarkt hebben gehad:

- A: personen die voortdurend hebben gewerkt ('PARTICIPEERT');
- B: personen die wisselend gedrag vertonen ('WISSELT');
- C: personen die niet hebben gewerkt ('PARTICIPEERT NIET').

In de eerste fase van het onderzoek (de analysefase) formuleren en schatten wij een model voor het arbeidsmarktgedrag van personen, waarbij dit gedrag wordt verklaard uit persoonsgebonden kenmerken. Het arbeids-

marktgedrag van huisgenoten komt pas aan de orde in de tweede fase (de simulatiefase). Wij voegen de individuen dan samen tot huishoudens, en met behulp van het in de analysefase ontwikkelde model berekenen wij hoe de steekproef van huishoudens er uit zou zien indien huisgenoten elkaar niet zouden beïnvloeden. Wij confronteren dit beeld met de werkelijkheid en kunnen zo de invloed van huisgenoten vaststellen ('voorbeeldwerking' of 'compensatie').

### Data

Wij maken gebruik van het arbeidsmarktsurvey van OSA, de Organisatie voor Strategisch Arbeidsmarktonderzoek. Dit is een enquête onder 2132 huishoudens met 6348 individuen, medio 1985 door Intomart in opdracht van OSA gehouden. In het vraaggesprek is een passage opgenomen over de arbeidsmarktgeschiedenis over de afgelopen vijf jaar, die een indruk geeft van de min of meer permanente of blijvende arbeidsmarktpositie van een persoon. Ook is het mogelijk de individuele gegevens naar huishoudens bijeen te voegen.

Het OSA-survey heeft een retrospectief karakter. Het is de bedoeling dat het zal uitgroeien tot een longitudinaal panel; onlangs zijn daartoe alle respondenten voor de tweede keer benaderd.

Het onderhavige onderzoek is uitgevoerd op een deelsteekproef uit het OSA-bestand, waarvoor alle benodigde gegevens (zoals inkomen, uurloon) bekend waren. In de analysefase zijn in totaal 2972 individuen betrokken (1586 vrouwen en 1386 mannen). Voor de simulatiefase kwamen 1629 gezinnen in aanmerking, maar om praktische redenen is deze fase van het onderzoek beperkt tot 1057 gezinnen van 2 personen en 163 gezinnen van 3 personen. Het aantal grotere huishoudens in de steekproef was te klein om zinvolle uitspraken te doen.

### Model

Het multinomiale logit model is bijzonder geschikt voor de analyse van waarnemingen die in meerdere categorieën zijn ingedeeld. De te verklaren variabele in het model is de verdeling van de individuen over



deze categorieën, in ons geval over 3 categorieën van arbeidsmarktgedrag ('PARTICIPEERT', 'WISSELT' of 'PARTICIPEERT NIET').

Als verklarende variabelen nemen wij kenmerken van het individu op, zoals geslacht en leeftijd. Op basis van individuele waarnemingen van de waarden van de verklarende variabelen en van de arbeidsmarktpositie kunnen wij de parameters schatten die de kans bepalen dat een persoon, gegeven haar of zijn kenmerken, tot één van de drie categorieën behoort.

In ons model zijn de volgende verklarende variabelen opgenomen:

GSL geslacht

LFT leeftijd op 1-1-1985

VKD verzorgt kind (ja = 1, nee = 0)

NIV niveau van de hoogst gevolgde opleiding

INA niet-arbeidsinkomen

ULN uurloon

BSP woont samen met partner (ja = 1, nee = 0)

BSS gescheiden (ja = 1, nee = 0)

RWL werkloosheidspercentage in provincie

UST woont in stad (ja = 1, nee = 0)

#### Resultaten van de analysefase

In de analysefase wordt het individuele arbeidsaanbod geanalyseerd. Het model is voor mannen en vrouwen apart geschat, omdat het voor de hand ligt dat de effecten voor mannen en vrouwen verschillend zijn. Het algemene beeld dat uit de schattingen naar voren komt is dat het arbeidsmarktgedrag van mannen minder variatie vertoont en minder gevoelig is voor veranderingen in de omstandigheden, dan dat van vrouwen. Dit blijkt ook uit het feit dat het overgrote deel van de mannen de gehele vijfjarige periode heeft geparticipeerd. Bij de vrouwen zien we meer spreiding over de drie categorieën.

Ook overigens zijn de schattingsresultaten plausibel. De participatiekans is voor de middelbare leeftijdsgroepen hoger dan voor jongeren en voor ouderen. Jongeren (vooral vrouwen) vallen relatief vaak in de categorie 'WISSELT', en ouderen in 'PARTICIPEERT NIET'. De aanwezigheid van een kind

heeft bij vrouwen meer effect dan bij mannen; de kans op 'PARTICIPEERT' neemt dan sterk af en de kans op 'PARTICIPEERT NIET' neemt sterk toe. Bij mannen zouden we het tegengestelde verwachten, maar dit is niet het geval; de invloed op de kansen is overigens maar klein.

Een hoger opleidingsniveau leidt bij mannen tot een wat hogere kans op participatie, vooral ten koste van de categorie 'WISSELT'. Bij vrouwen is er een veel sterker effect, en wel een daling van de kans op 'PARTICIPEERT NIET', vooral ten bate van 'WISSELT'. Men kan dus stellen dat een hogere opleiding de kans op participatie in het algemeen vergroot, maar de precieze uitwerking verschilt naar geslacht.

Het uurloon blijkt bij mannen geen significante invloed te hebben op het arbeidsmarktgedrag. Bij vrouwen daarentegen heeft een hoog uurloon een positief effect op de participatie.

De burgerlijke staat is de enige variabele die het arbeidsmarktgedrag van mannen sterker beïnvloedt dan dat van vrouwen; mannen met partner participeren meer. Bij vrouwen treedt hetzelfde verschijnsel op, maar dan zwakker. Dit laatste lijkt vreemd, omdat bekend is dat gehuwde vrouwen minder werken dan ongehuwde. De verklaring is dat kennelijk niet de burgerlijke staat zelf, maar de verschillen in omstandigheden die daarmee gepaard gaan (kinderen, inkomen echtgenoot) de oorzaak vormen van het verschil in gedrag. Met andere woorden, gehuwde en ongehuwde vrouwen gedragen zich hetzelfde in vergelijkbare situaties.

Het werkloosheidspercentage is als verklarende variabele opgenomen om het zgn. 'discouraged worker effect' te kunnen meten. Dit is het verschijnsel dat bij hoge werkloosheid mensen die eigenlijk zouden willen participeren die niet doen omdat zij verwachten toch geen werk te zullen vinden. In regio's met hoge werkloosheid blijkt de participatie inderdaad te dalen; ontmoedigde vrouwen stromen vooral naar de categorie 'PARTICIPEERT NIET' en mannen naar 'WISSELT'.

#### Resultaten van de simulatiefase

Duidelijke aanwijzingen voor wederzijdse beïnvloeding door huisgenoten hebben wij niet gevonden. Voor zover er sprake is van beïnvloeding is deze positief ('voorbeeldwerking') en manifesteert hij zich alleen bij de twee persoons huishoudens. Dit effect is echter heel zwak, en het lukt niet het

nader te lokaliseren.

De analyse van de driepersoons huishoudens wordt gehinderd door het feit dat relatief veel gezinnen in deze groep één of meer personen bevatten die hun arbeidsmarktcategory niet vrij kunnen kiezen (meestal een student of scholier). Aanwijzingen voor enige wederzijdse beïnvloeding zijn er in deze groep niet.

Achteraf gezien is dit resultaat niet verbazingwekkend wanneer wij bedenken dat een typisch tweepersoons huishouden bestaat uit twee volwassenen en een typisch driepersoons huishouden uit twee volwassenen met een studierend kind. Indien er sprake is van beïnvloeding door huisgenoten zal deze tussen partners onderling anders zijn dan tussen ouders en kinderen. De eerste soort beïnvloeding komt in onze data blijkbaar voor, zij het in een zwakke vorm, de tweede soort niet.

## 2 PROBLEEMSTELLING

Het inzicht in de factoren die het arbeidsaanbod bepalen is de laatste jaren sterk gestegen dankzij een aantal onderzoeken waarin het individuele arbeidsaanbod vanuit micro-economisch gezichtspunt wordt verklaard (Siegers (1985), Hartog en Theeuwes (1983)). Heel in het kort gezegd komt deze benadering erop neer dat een model wordt geformuleerd waarin het individuele arbeidsaanbod wordt verklaard uit kenmerken van het individu, zoals geslacht, leeftijd en opleiding. De invloed van deze factoren komt tot uiting in coëfficiënten, die kunnen worden geschat wanneer men beschikt over de juiste techniek en de juiste soort data. Zijn de parameters eenmaal geschat dan kan het model gebruikt worden voor allerlei simulaties waarmee bijvoorbeeld de gevolgen van bepaalde beleidsmaatregelen kunnen worden berekend (Siegers (1983); Bekkering, Grift en Siegers (1986)).

Voor dit soort onderzoeken maakt men tot voor kort vooral gebruik van enquêtes, waarin van de arbeidsmarktpositie van de ondervraagde een momentopname wordt genomen. Tegenwoordig ontstaat er echter meer belangstelling voor het verloop van arbeidsmarktprocessen in de tijd. De reden hiervoor is de groeiende bezorgdheid over het probleem van de langdurige werkloosheid. Het vermoeden bestaat dat sommige bevolkingsgroepen ook bij een aantrekkende economie van het arbeidsproces uitgesloten zullen blijven. Dit zal leiden tot het ontstaan van een harde kern van langdurig werklozen. Om te weten of deze ontwikkeling zal optreden en wat men daartegen zou kunnen doen is inzicht in de omvang en de oorzaken van langdurige werkloosheid onontbeerlijk.

De bestaande analyses bieden hiervoor te weinig aanknopingspunten, omdat zij slechts uitgaan van de situatie op een bepaald tijdstip of te zeer geaggregeerde grootheden hanteren. Een eenvoudig getallenvoorbeeld kan dit laatste verduidelijken. Wanneer van een groep van 1000 personen er 100 het gehele jaar door werkloos zijn, en 400 ieder slechts drie maanden van het jaar, dan is het gemiddelde werkloosheidspercentage over het gehele jaar 20%. Ditzelfde percentage wordt echter ook gerealiseerd wanneer 400 personen de helft van het jaar werkloos zijn. Het werkloosheidspercentage

zegt derhalve op zichzelf niets over het vóórkomen van langdurige werkloosheid.

Om hier meer over te weten te komen moet men beschikken over speciaal op het probleem toegesneden gegevens en analysetechnieken. Het meest geschikt is een longitudinale enquête, dat betekent dat van dezelfde groep mensen gedurende langere tijd de arbeidsmarktsituatie is geregistreerd. Een voorbeeld is de studie van Van Opstal en Theeuwes (1985). Zij volgen een groep jeugdige werklozen gedurende een periode van 7 maanden. Hun voornaamste conclusie is echter dat deze onderzoeksperiode te kort is, omdat zij te weinig waarnemingen hebben van voltooide werkloosheidsduren.

De onderhavige studie kan gezien worden als een verkenning, die een aanvulling kan zijn op een duuranalyse. Het onderwerp van studie is het arbeidsmarktgedrag over een periode van vijf jaar. Wij delen personen in in drie groepen, die op grond van hun geschiedenis gedurende de afgelopen vijf jaar in verschillende mate toegang hebben tot de arbeidsmarkt. Over een zo lange termijn gemeten verschillen tussen individuen dragen een min of meer structureel karakter. De vraagstelling is welke kenmerken van een persoon bepalen in welke groep zij of hij terecht komt. Vervolgens verschuift het accent van de studie van de kenmerken van het individu naar de invloed van de omgeving. Hierover is nog zeer weinig bekend, terwijl het probleem een belangrijke rol speelt in de discussie over langdurige werkloosheid. Deze is immers niet gelijkmatig over het land gespreid, maar komt vooral voor in bepaalde probleemgebieden zoals sommige wijken in de grote steden. Ook wordt wel aangenomen dat kinderen van langdurig werklozen een grotere kans hebben om zelf werkloos te worden. Maar komt dit uitsluitend doordat zij bepaalde kenmerken (opleiding, woonplaats) gemeenschappelijk hebben, of wordt dit effect versterkt door de 'voorbeeldwerking' binnen het gezin? Of is er geen sprake van voorbeeldwerking en vindt er zelfs 'compensatie' plaats, dat wil zeggen dat het werklooszijn van één huisgenoot de kans vergroot dat anderen werken? Op dit soort vragen proberen wij een antwoord te vinden. Wij beperken ons daarbij tot de onderlinge beïnvloeding door huisgenoten, dus door leden van hetzelfde huishouden, door gedrag, dus los

van de financiële gevolgen die de arbeidsmarktstatus van de één binnen een huishouden voor de anderen heeft.

Merk op dat wij de invloed van de omgeving beperken tot de directe omgeving, namelijk de huisgenoten. Deze beperking kan betwist worden; misschien is het beter om ervan uit te gaan dat de invloed van de omgeving werkt via de "sociale groep" waartoe een individu zichzelf rekent.\*

Kort samengevat concentreert het onderzoek zich op de volgende twee vragen:

- welke kenmerken van een individu bepalen haar of zijn structurele arbeidsmarktgedrag?;
- is er daarbij sprake van onderlinge beïnvloeding door huisgenoten en zo ja, in welke richting?

Beide vragen worden onderzocht aan de hand van het arbeidsmarktsurvey van OSA, de Organisatie voor Strategisch Arbeidsmarktonderzoek. Dit is een enquête onder 2132 huishoudens met 6348 individuen, in 1985 door Intomart in opdracht van OSA gehouden. In het vraaggesprek is een passage opgenomen over de arbeidsmarktgeschiedenis van het individu over de afgelopen vijf jaar, die een indruk geeft van haar of zijn min of meer permanente of blijvende arbeidsmarktpositie. Ook is het mogelijk de individuele gegevens naar huishoudens bijeen te voegen. Ten slotte biedt het materiaal uitzicht op vervolgonderzoek, mocht daartoe aanleiding zijn, omdat dezelfde steekproef onlangs opnieuw is ondervraagd.

Bij de beantwoording van de twee bovengenoemde vragen hebben wij ons beperkt tot traditionele analysemethoden. Naar nieuwe technische of theoretische vondsten is niet gestreefd. Alleen bij de beantwoording van de tweede vraag zijn nieuwe wegen ingeslagen.

---

\* zie Kapteyn en Woittiez (1986) voor een modelmatige uitwerking hiervan.



### 3 OPZET VAN HET ONDERZOEK

#### 3.1 Algemeen

Het onderzoek gaat over de verdeling van de bevolking over categorieën die in verschillende mate deel hebben aan de arbeidsmarkt, en de vraagstelling is welke kenmerken van een persoon bepalen in welke categorie zij of hij terecht komt. Het OSA-survey is zeer geschikt als basis voor dit onderzoek, omdat het inzicht verschaft in de structurele arbeidsmarktpositie van individuen, gemeten over een vijfjarige periode, en omdat het in beginsel mogelijk is om de individuen waarover gegevens beschikbaar zijn tot huishoudens samen te voegen.

Wij onderscheiden drie categorieën van arbeidsmarktgedrag:

- personen die voortdurend betaalde arbeid verrichten ('PARTICIPEERT');
- personen die wisselend gedrag vertonen ('WISSELT');
- personen die in het geheel niet hebben gewerkt ('PARTICIPEERT NIET').

Op de precieze definitie van deze categorieën komen wij in het volgende terug. De verdeling van personen over de drie categorieën hangt af van een aantal kenmerken. Deze kenmerken vallen uiteen in drie groepen:

- onveranderlijke factoren zoals leeftijd, geslacht;
- factoren die min of meer door beleid kunnen worden beïnvloed, zoals opleiding of inkomen;
- de omgeving van de persoon (het gedrag van huisgenoten).

Het verschil tussen de eerste twee groepen factoren is eigenlijk alleen voor het beleid van belang; in de analyse worden zij gelijk behandeld. De derde groep variabelen onderscheidt zich door het kwalitatieve karakter: de zuiver persoonsgebonden kenmerken kunnen meestal eenvoudig in één getal worden uitgedrukt, maar het gedrag van de huisgenoten heeft vele dimensies. Bovendien hoeft het effect ervan niet onder alle omstandigheden hetzelfde te zijn: soms kan er 'compensatie' optreden, en in andere gevallen juist



'voorbeeldwerking'. Dit onderscheid is echter te fijn om zich voor empirisch onderzoek te lenen.

Het onderzoek wordt in twee fasen uitgevoerd:

- 1 de analysefase. In deze fase formuleren en schatten wij een model voor het arbeidsmarktgedrag van personen, waarbij dit gedrag wordt verklaard uit zuiver persoonsgebonden kenmerken. In deze fase wordt het arbeidsmarktgedrag van huisgenoten buiten beschouwing gelaten, maar hun inkomen niet.
  
- 2 de simulatiefase. De te onderzoeken eenheid wordt nu het huishouden in plaats van de persoon. Met behulp van het in de analysefase ontwikkelde model berekenen wij hoe de steekproef van huishoudens eruit zou zien, indien huisgenoten elkaar niet beïnvloeden. Wij confronteren dit beeld met de werkelijkheid en kunnen op deze manier de invloed van huisgenoten vaststellen ('voorbeeldwerking' of 'compensatie').

Kort gezegd houdt onze benadering in dat wij het arbeidsmarktgedrag van individuen in eerste instantie verklaren uit individuele kenmerken, met behulp van een geschikt model. Vervolgens maken wij de veronderstelling dat afwijkingen tussen het door het model voorspelde gedrag en het waargenomen gedrag uitsluitend worden veroorzaakt door de onderlinge beïnvloeding van gezinsleden. Wij analyseren dit door na te gaan of personen uit één huishouden opvallend vaak hetzelfde of juist opvallend vaak verschillend gedrag vertonen, gegeven hun kenmerken.

Het in de eerste fase gebruikte statistische model is het multinomiale logitmodel. Op de modelspecificatie wordt uitgebreid ingegaan in Hoofdstuk 4. Hier volstaan wij met erop te wijzen dat de te verklaren variabele bestaat uit de kansverdeling van personen over de drie bovengenoemde categorieën; als verklarende variabelen nemen wij zulke kenmerken op als leeftijd, opleiding, etc.

### 3.2 De indeling in categorieën

De drie categorieën van arbeidsmarktgedrag worden zodanig vastgesteld, dat zij het structurele gedrag weergeven. Op deze manier kunnen wij de invloed van allerlei toevallige factoren (korte ziekte, frictiewerkloosheid) beperkt houden.

Toch zitten er nog wat haken en ogen aan de gekozen benadering, omdat wij impliciet veronderstellen dat personen vrijelijk kunnen kiezen in welke categorie zij thuis horen, en dat hun keuze door beleid kan worden beïnvloed. De werkelijkheid is natuurlijk anders: voor bepaalde personen is het aantal vrij te kiezen alternatieven beperkt.

Om hiermee rekening te houden hebben wij als vierde categorie de groep 'geen keuze' opgenomen. De personen in deze groep participeren meestal niet op de arbeidsmarkt, maar afgezien daarvan hebben zij weinig gemeen: (sommige) invaliden, dagstudenten, VUT-ters en dienstplichtigen worden allen geacht geen keuze te hebben. Bij dagstudenten en VUT-ters is het begin van de arbeidsmarktcarrière verlaat resp. het eind vervroegd; voor de invaliden is niet zeker of zij ooit nog uit de groep 'geen keuze' zullen ontsnappen. Bij dienstplichtigen (die wel geacht worden te participeren) kan men er zeker van zijn dat zij slechts tijdelijk 'geen keuze' hebben.

Op het eerste gezicht lijkt het misschien redelijk om ook de groep 'moeders met jonge kinderen' in te delen bij de groep 'geen keuze'; dit zou echter niet in overeenstemming zijn met de praktijk, omdat van de ondervraagde moeders ongeveer 20% werkt of werk zoekt. In hoeverre dit verband houdt met de aanwezigheid van kinderopvang is helaas niet na te gaan.

In de analyse-fase van het onderzoek wordt de groep 'geen keuze' buiten beschouwing gelaten, omdat hun categorie vastligt en dus niet kan worden beïnvloed door allerlei persoonsgebonden variabelen. Dit betekent echter niet dat wij deze groep ook in de simulatie-fase kunnen weglaten.

Wanneer wij de groep 'geen keuze' buiten beschouwing laten, luidt onze

definitie van de categorieën als volgt:

- A) participeert: deze groep bestaat uit alle personen die in de periode 1-1-1980 tot 1-1-1985 minder dan drie maanden zonder werk zijn geweest, ongeacht de arbeidstijd per week;
- C) participeert niet: alle personen die van 1-1-1980 tot 1-1-1985 minder dan drie maanden betaalde arbeid hebben verricht;
- B) wisselt: alle overigen.

Het bestaan van de categorie 'geen keuze' noodzaakt tot de volgende correcties:

- personen die op het moment van ondervraging invalide, dagstudent, dienstplichtig, VUT-ter of gepensioneerd waren, vallen in de categorie 'geen keuze' ongeacht hun arbeidsmarktgedrag in de voorafgaande periode;
- personen die tussen 1-1-1980 en 1-1-1985 enige tijd dagstudent of dienstplichtig zijn geweest, worden ingedeeld volgens een aparte procedure, waarbij de tijd dat zij al of niet werkten wordt gerelateerd aan de periode dat zij 'wel keuze' hadden.

Bij de indeling in de drie groepen A, B en C wordt geen rekening gehouden met de aard van het dienstverband of met de arbeidstijd per week. De groep B, die afwisselend werkt en niet werkt, heeft dus geen betrekking op deeltijdarbeid, oproepcontracten of andere vormen van flexibele arbeid; wij zien alleen naar de mate waarin het individu in de afgelopen vijf jaar werk heeft gehad, ongeacht het karakter van dat werk. Bij deze keuze heeft mede een rol gespeeld dat er aanwijzingen zijn dat de vraag naar de aard van het dienstverband in de enquête niet altijd goed is beantwoord, zeker bij de reconstructie van de arbeidsmarktgeschiedenis over de laatste vijf jaar. Ook is bij de indeling in de drie groepen alleen aandacht besteed aan de vraag of men werkte of niet, zonder nader onderscheid naar de aard van het niet werken. De categorie C bevat iedereen die in de afgelopen vijf jaar niet of nauwelijks heeft gewerkt, en dat zijn niet alleen (zeer) langdurig werklozen maar ook, bij de vrouwen, moeders met kleine kinderen. De oorzaak van niet participeren kan dus heel sterk verschillen, en de interpretatie

verschilt vooral naar het geslacht. Vanuit dat gezichtspunt is het goed dat de verdeling over de drie groepen A, B en C voor mannen en vrouwen gescheiden wordt geanalyseerd.

### 3.3 Beperkingen en veronderstellingen

Binnen de beperkte opzet van dit onderzoek, die door de korte beschikbare tijd werd opgelegd, is niet gestreefd naar nieuwe technische of theoretische vondsten, en is in hoofdzaak gebruik gemaakt van reeds beschikbare programmatuur. Dit heeft verschillende consequenties. Ten eerste zijn een aantal verklarende variabelen niet zo mooi als wij ze eigenlijk zouden wensen. Hiervoor verwijzen wij naar Hoofdstuk 4.2, waar de keuze van de verklarende variabelen wordt besproken.

Een belangrijker gevolg is de aard van het onderzoek; wij doen geen enkele poging om de onderlinge beïnvloeding van huisgenoten te verklaren. De analyse is uitsluitend beschrijvend. In een verklarend model zou men de wijze waarop het beslissingsproces binnen een huishouden plaatsvindt, expliciet moeten formuleren\*. Het onderhavige onderzoek is slechts een inventarisatie, als eerste stap in de richting van een verklarend model.

Een andere praktische beperking wordt ons opgelegd door de probleemstelling. De indeling in categorieën is gebaseerd op het arbeidsmarktgedrag in de periode 1-1-1980 tot en met 1-1-1985. Impliciet is dus verondersteld dat het structurele arbeidsmarktgedrag over een vijfjarige periode niet of niet vaak zal wijzigen. Dit is een realistische veronderstelling voor ouderen, maar in mindere mate voor personen die aan het begin van hun arbeidsmarktcarrière staan.

Dan zijn er nog beperkingen die worden veroorzaakt doordat wij ons baseren op een enquête die voor een geheel ander doel is gehouden dan het onderhavige onderzoek. Het zou wat irreëel zijn om te verwachten dat alle vragen en definities uit het OSA-survey precies bij onze probleemstelling

---

\* zie voor aanzetten in deze richting de dissertatie van P. Kooreman (1986)

aansluiten. Een voorbeeld is de wijze van steekproeftrekking. De doelgroep van de enquête bestond uit alle personen die tussen 15 en 60 jaar oud waren, behalve dienstplichtig militairen en dagstudenten; personen die kostwinner waren, werden altijd ondervraagd. Het probleem waar dit toe leidt kan het best geïllustreerd worden aan de hand van de dienstplichtig militairen: het databestand bevat niemand die op het tijdstip van ondervraging in dienst zat, maar wel 52 personen die tijdens de voorafgaande vijf jaar hun volledige dienstplicht hebben vervuld. Van deze personen is het arbeidsmarktgedrag in de periode 1980-1985 wel bekend, en zij kunnen dus na een correctie voor hun diensttijd in één van de onderscheiden categorieën worden ingedeeld. Het gevolg van deze correctie is dat een groep waarnemingen systematisch ondervertegenwoordigd is, namelijk degenen die op het moment van ondervraging in dienst zaten. Precies hetzelfde geldt uiteraard voor dagstudenten en voor werkende niet-kostwinners ouder dan 60, al is de laatste groep waarschijnlijk niet groot.

Een correcte, maar nogal ingrijpende manier om voor deze selectie te corrigeren is om de analyse te beperken tot personen die tijdens de onderzoeksperiode ouder dan 25 waren, zodat alle - potentiële - dagstudenten en dienstplichtigen uit de steekproef zijn verwijderd. Het nadeel van deze methode is, dat daarmee tegelijk een zeer interessante groepering, de thuiswonende werkende jongeren, van het onderzoek wordt uitgesloten. Daarom zullen wij dagstudenten en dienstplichtigen in de groep 'geen keuze' indelen, hoewel dit strikt genomen niet juist is. De fout die hiermee wordt gemaakt is waarschijnlijk niet groot.

Een volgende beperking die moet worden genoemd is dat de verklarende variabelen meestal slechts bekend zijn op het moment van ondervraging. Dit moment ligt tussen april en juni 1985. Wij kunnen dus niet waarnemen welke waarden zij in de voorafgaande jaren hebben aangenomen. Dit geldt zowel voor de periode 1980-1985 als voor de maanden tussen 1-1-1985 en het ogenblik van enquêtering. Dit is jammer. Wij hebben de onderzoeksperiode vastgesteld op de vijf jaar tussen 1-1-1980 en 1-1-1985, omdat de variabelen die de arbeidsmarktgeschiedenis beschrijven over deze periode voor alle respondenten bekend zijn. De einddatum van de opgegeven arbeids-

marktgeschiedenissen varieert tussen respondenten; februari en augustus 1985 komen beiden voor. Overigens moeten wij uiteraard maar op het geheugen van de respondenten vertrouwen.

Ten slotte moet erop worden gewezen dat ons model het werkelijke arbeidsmarktgedrag verklaart uit aanbodfactoren. Beperkingen van de vraagzijde van de arbeidsmarkt laten wij buiten beschouwing.



## 4 METHODE

4.1 Het logitmodel

Veronderstel dat er  $N$  personen zijn die  $K$  typen van arbeidsmarktgedrag kunnen vertonen. De kans dat persoon  $i$  gedrag type  $j$  heeft,  $P_{ij}$ , is afhankelijk van een aantal kenmerken van  $i$ . Deze kenmerken geven wij aan als een vector  $x_i$ .

Volgens het logitmodel is de kans  $P_{ij}$  als volgt bepaald:

$$P_{ij} = \frac{\exp x_i' \beta_j}{\sum_{k=1}^K \exp x_i' \beta_k} \quad (4/1)$$

Hierin is  $\beta_j$  een vector met dezelfde dimensie als  $x_i$ .  $\beta_j$  bevat de te schatten coëfficiënten. Het is duidelijk dat de kansen  $P_{ij}$  tot één sommeren.

De vector van kenmerken  $x_i$  bevat zulke variabelen als leeftijd, opleidingsniveau, etcetera. Wij verwachten dat deze variabelen invloed hebben op het type arbeidsmarktgedrag. De richting van deze invloed op de kans op arbeidsmarktgedrag  $j$  komt tot uiting in het teken van de desbetreffende coëfficiënt. Het aantal coëfficiënten  $\beta_j$  neemt rechtstreeks toe met het aantal typen arbeidsmarktgedrag dat wordt onderscheiden.

Het logitmodel heeft een aantal belangrijke eigenschappen. De eerste eigenschap is, dat het model kan worden afgeleid uit een nutsmaximeringsprobleem. Veronderstel dat voor individu  $i$  het arbeidsmarktgedrag type  $j$  een nut  $u_{ij}$  oplevert, en dat

$$u_{ij} = x_i' \beta_j + \varepsilon_{ij} \quad (4/2)$$



met  $\epsilon_{ij}$  een stochastische storingsterm. Bij bepaalde veronderstellingen over de verdeling van  $\epsilon_{ij}$  leidt maximering van dit nut bij de keuze van het arbeidsmarktgedrag voor individu  $i$  tot Formule (4/1) voor de kans  $P_{ij}$ .

Een tweede eigenschap van het logitmodel wordt in de literatuur wel aangeduid als "Independence from Irrelevant Alternatives" (IIA). Dit betekent dat de zogenaamde 'odds ratio'  $P_{ij}/P_{ik}$ , dus de verhouding van de kansen voor twee alternatieven, onafhankelijk is van het aantal en het karakter van de andere alternatieven. Dit valt eenvoudig in te zien:

$$\frac{P_{ij}}{P_{ik}} = \frac{\exp x_i' \beta_j}{\exp x_i' \beta_k} \quad (4/3)$$

In deze formule komen alleen de coëfficiënten van de alternatieven  $j$  en  $k$  voor, de andere zijn afwezig; het doet er zelfs niet toe welke alternatieven dit zijn. Vergelijking (4/3) zal dus ook gelden als wij naast  $j$  en  $k$  een nieuwe indeling van de andere alternatieven invoeren.

Uit (4/3) kunnen wij een derde eigenschap afleiden:

$$\log \frac{P_{ij}}{P_{ik}} = x_i' (\beta_j - \beta_k) \quad (4/4)$$

Uit (4/4) is duidelijk dat het model in deze vorm niet geïdentificeerd is. Wij kunnen slechts het verschil tussen de  $\beta_j$ 's schatten. Dit probleem wordt opgelost door de introductie van een referentiegroep, waarvoor de coëfficiënten op nul worden gesteld. Indien wij categorie 1 als zodanig

kiezen, luidt het model:

$$P_{i1} = \frac{1}{1 + \sum_{k=2}^K \exp x_i' \beta_k} \quad (4/5)$$

$$P_{ij} = \frac{\exp x_i' \beta_j}{1 + \sum_{k=2}^K \exp x_i' \beta_k}, \quad j = 2 \dots K \quad (4/6)$$

Vanzelfsprekend worden de uitkomsten van het model niet beïnvloed door de keuze van de referentiegroep. Wel dient men bij de interpretatie van de schattingen rekening te houden met het feit dat er slechts verschillen tussen coëfficiënten worden geschat. Om de interpretatie te vereenvoudigen presenteren wij in Hoofdstuk 5 een aantal eenvoudiger grootheden die uit de coëfficiënten kunnen worden afgeleid.

De schattingsmethode kan het best worden geïllustreerd aan de hand van een eenvoudig geval met een verdeling tussen slechts twee typen arbeidsmarktgedrag die afhangt van één verklarende variabele. Wij zouden bijvoorbeeld kunnen definiëren:

- j = 1: participeert wel
- j = 2: participeert niet

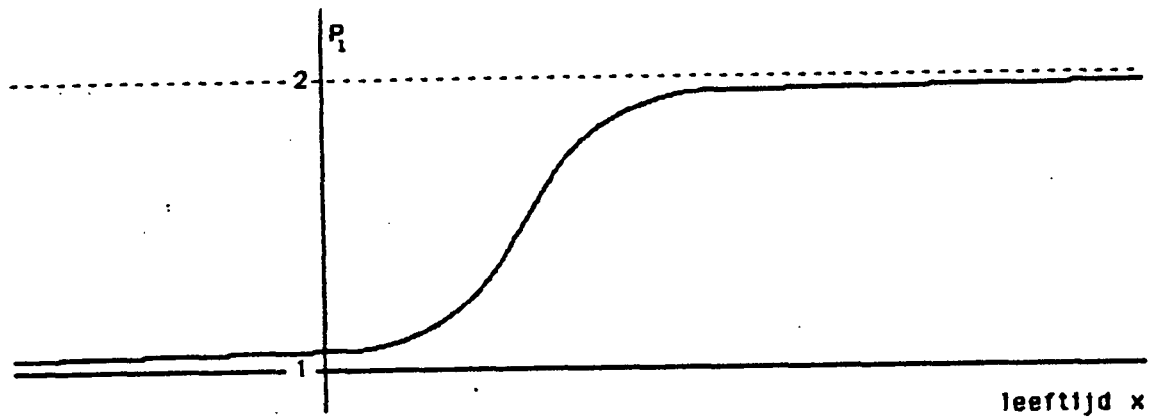
en als enige verklarende variabele x de leeftijd kiezen. In plaats van (4/5) en (4/6) vinden wij nu:

$$P_{i1} = \frac{1}{1 + \exp (\beta_0 + \beta_1 x_i)} \quad (4/7)$$

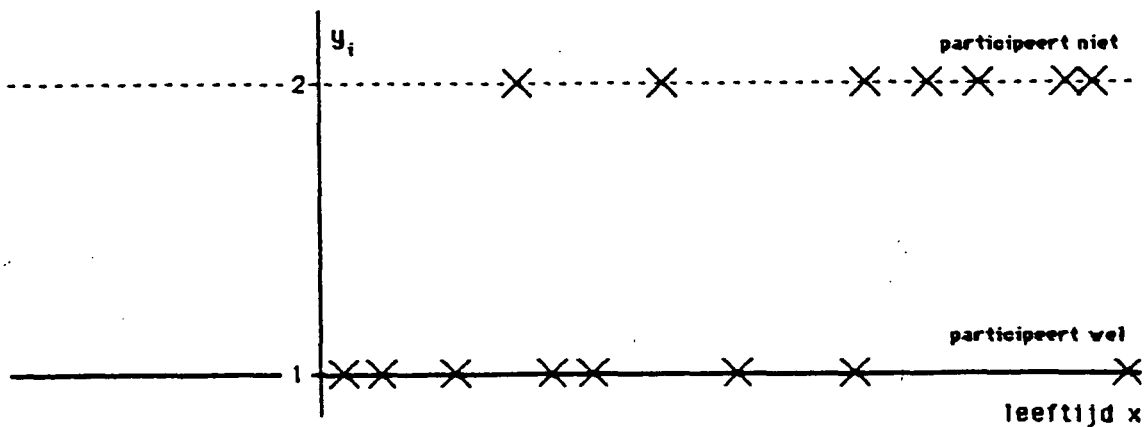
$$P_{i2} = \frac{\exp (\beta_0 + \beta_1 x_i)}{1 + \exp (\beta_0 + \beta_1 x_i)}. \quad (4/8)$$

Dankzij de sterke vereenvoudigingen die wij hebben aangebracht kunnen wij (4/8) in een figuur weergeven. In Figuur 4/1 is (4/8) gegeven, met  $\beta_1 > 0$ : de kans dat iemand niet participeert neemt toe met de leeftijd, en wel volgens een S-vormige curve. In Figuur 4/2 zijn steekproefwaarnemingen getekend waarmee dit model wordt geconfronteerd. Het zijn individuele waarnemingen, telkens van één persoon, die ofwel participeert (1) ofwel niet (2).

Figuur 4/1 Het verloop van (4/8) met  $\beta_1 > 0$



Figuur 4/2 Waarnemingen van 15 individuen met leeftijd  $x_i$



Het zal nu duidelijk zijn dat men door een goede keuze van  $\beta_0$  en  $\beta_1$  in (4/8) er voor kan zorgen dat de kansen van het model zo goed mogelijk overeenkomen met de frequenties die zich in de praktijk voordoen. De waarden die het beste resultaat opleveren aanvaarden wij als schatting van de coëfficiënten. Aangezien het logitmodel een ingewikkelde vorm heeft gebeurt de bepaling van de schatters in de praktijk door een systematisch proces van "trial and error". Beschikt men over een goede computer en een behoorlijk budget voor de berekeningen dan levert dit geen moeilijkheden op.

Aangezien de logitformule (4/1) kansen of waarschijnlijkheden definieert hebben wij hier te maken met een stochastisch model. De schattingen die wij berekenen kunnen dan ook op de gebruikelijke wijze volgens de statistische theorie worden geïnterpreteerd. Dit betekent dat zij, net zoals regressie-coëfficiënten, standaardfouten hebben die wij kunnen berekenen, en dat ook kan worden nagegaan of zij al dan niet significant van nul verschillen. Het zou echter te ver voeren om de theorie die hieraan ten grondslag ligt uiteen te zetten.

#### 4.2 Te verklaren en verklarende variabelen

De te verklaren variabele in het model is de kansverdeling voor een individu over drie categorieën van arbeidsmarktgedrag. Deze kansverdeling beschrijft keuzegedrag, en de keuze tussen de drie categorieën wordt verklaard uit persoonskenmerken. Personen die niet kunnen kiezen worden opgenomen in een vierde categorie, de groep 'geen keuze'. Deze personen worden in de analysefase buiten beschouwing gelaten. De precieze definitie van deze categorieën is gegeven in Paragraaf 3.2.

Om de indeling van personen naar de vier onderscheiden categorieën vast te stellen, moeten wij gebruik maken van de oorspronkelijke gezinsformulieren van het OSA-survey. Niet alle leden van de benaderde huishoudens zijn namelijk ondervraagd. Gelukkig behoren alle niet-ondervraagde personen tot de categorie 'geen keuze' die in de analyse-fase niet wordt meegenomen. De

verdeling van onze steekproef over de categorieën staat vermeld in Tabel 4.2/1. Het totaal aantal ondervraagde personen telt niet op tot 4020, omdat alle gezinnen waarvan tenminste één lid zijn inkomen niet kon of wilde opgeven, uit het bestand zijn verwijderd. Bovendien zijn die gezinsleden niet meegenomen, van wie het uurloon niet bekend was. De groep "niet van toepassing" bevat personen (meestal nog jonge kinderen) die niet zijn ondervraagd. De bejaarden in deze groep worden in de categorie 'geen keuze' ondergebracht.

Tabel 4.2/1 Verdeling van de steekproef over vier categorieën van arbeidsmarktgedrag

categorie	ondervraagd	niet ondervraagd	totaal
participeert	1683	---	1683
wisselt	538	---	538
participeert niet	751	---	751
geen keuze	241	1381	1622
niet van toepassing	0	601	601
totaal	3213	1982	5195

Bij de keuze van de verklarende variabelen hebben wij ons laten leiden door de resultaten van reeds door anderen verricht onderzoek. Wij gebruiken de volgende variabelen:

1. GSL geslacht
2. LFT leeftijd op 1-1-85
3. SLF LFT kwadraat
4. VKD verzorgt kind, ja = 1, nee = 0
5. NIV niveau van de hoogst gevolgde opleiding
6. INA niet-arbeidsinkomen van individu
7. ULN uurloon (feitelijk of verwacht)
8. BSP woont samen met partner, ja = 1, nee = 0
9. BSS gescheiden, ja = 1, nee = 0
10. RWL werkloosheid in provincie
11. UST woont in stad, ja = 1, nee = 0

De volgende variabelen verdienen nadere toelichting.

GSL Omdat wij verwachten dat voor een aantal variabelen de coëfficiënten van het geslacht zullen afhangen, gebruiken wij deze variabele om twee aparte deelbestanden aan te maken. De overige coëfficiënten worden dan voor mannen en vrouwen apart geschat. Geslacht is dus wel een belangrijke variabele, maar de invloed wordt niet door een coëfficiënt gemeten.

LFT en SLF Het effect van de leeftijd op het arbeidsmarktgedrag is tweeledig: enerzijds neemt met de leeftijd de ervaring toe, waardoor de participatiekans stijgt; aan de andere kant wordt bij een hoge leeftijd de gezondheidstoestand minder, zodat de participatiekans afneemt. Daarom veronderstellen wij een kwadratisch verband tussen leeftijd en arbeidsmarktgedrag.

VKD Deze dummy-variabele neemt de waarde 1 aan wanneer er in het huishouden iemand van 17 jaar of jonger aanwezig is, die niet werkt. Bovendien moet de persoon zelf 18 jaar of ouder zijn, en niet de zoon of dochter van de kostwinner zijn, dit laatste om te voorkomen dat oudere broers of zusters als 'verzorgers' worden aangemerkt. Wij verwachten dat de aldus gemeten aanwezigheid van kinderen een positief effect heeft op de participatie van mannen, en een negatief effect op die van vrouwen.

NIV Voor het opleidingsniveau nemen wij het niveau van de hoogst gevolgde opleiding volgens de Standaard Onderwijs Indeling 1978 van het CBS. Wij onderscheiden 5 niveaus, waarbij het laagste de waarde 2 aanneemt en het hoogste de waarde 6. Wij verwachten een positief effect van opleiding op participatie. Vanwege de vrij grote non-respons was het niet mogelijk om opleidingen ook naar richting te onderscheiden. Voor de precieze definitie van NIV verwijzen wij naar Appendix A.

INA Het niet-arbeidsinkomen is een lastige variabele, omdat het afhangt van het arbeidsmarktgedrag van huisgenoten. Het arbeidsinkomen van één huisgenoot is immers voor alle andere huisgenoten niet-arbeidsinkomen.

Terwille van de volgende fase van het onderzoek, waarin de onderlinge beïnvloeding tussen huisgenoten aan de orde komt, zouden wij nu geen verklarende variabelen willen opnemen waarin het arbeidsmarktgedrag van anderen verwerkt is. Voor het inkomenseffect van het delen in het gezamenlijke inkomen maken wij een uitzondering.

Het niet-arbeidsinkomen is berekend door het totale werkelijke inkomen van alle gezinsleden bij elkaar op te tellen, daar het eigen arbeidsinkomen van de betreffende persoon van af te trekken, en vervolgens te delen door het aantal gezinsleden. Dit betekent dat inkomen uit uitkering ook is opgenomen in het niet-arbeidsinkomen. In eerste instantie zou men hiertegen bezwaar kunnen maken, omdat inkomen uit uitkeringen een bepaald arbeidsmarktgedrag kan opleggen. Dit geldt echter uitsluitend voor het gedrag op een bepaald moment; voor zover een uitkering tijdelijk is wordt het structurele arbeidsmarktgedrag niet beïnvloed.

ULN Het uurloon is het quotiënt van twee variabelen, te weten het arbeidsinkomen en het aantal arbeidsuren. De constructie van deze variabelen wordt beschreven in Appendix A. Vanzelfsprekend zijn arbeidsinkomen en arbeidsuren alleen bekend voor personen die betaalde arbeid verrichten. Wanneer wij het uurloon als verklarende variabele in het model willen opnemen, moeten wij dus een kunstgreep toepassen om het - potentiële - uurloon te schatten van personen die niet werken. Gelukkig is in het OSA-survey een vraag opgenomen waarbij niet-werkers moeten opgeven hoeveel zij denken te verdienen, wanneer zij zouden werken (vraag 63.....welk nettoloon denkt u dat u op dit moment in een volledige baan zou kunnen verdienen?). Deze vraag kende een vrij hoge non-respons. Hiervoor corrigeren wij door voor degenen die wel antwoordden, een regressievergelijking te schatten met het verwachte loon als functie van leeftijd, geslacht en opleiding. Uit de gevonden relatie schatten wij dan vervolgens het verwachte arbeidsinkomen van degenen die dat niet hebben opgegeven. Omdat in vraag 63 expliciet over een volledige baan wordt gesproken, kunnen wij hieruit het verwachte uurloon afleiden.

BSP Deze dummy-variabele neemt de waarde 1 aan als de respondent samenwoont met een partner of gehuwd is. Beide gevallen worden over één kam geschoren.

RWL De werkloosheid in de betreffende provincie op 1-1-1985 wordt opgenomen om het zogenaamde 'discouraged worker-effect' te kunnen meten. Het zou mogelijk zijn om het gemiddelde over de periode 1980-1985 te berekenen, maar in verband met de onderlinge consistentie van de variabelen kiezen wij het percentage per 1-1-1985. Omdat de werkloosheid van mannen en van vrouwen per provincie aanmerkelijk kan verschillen, wordt in de vergelijkingen van beide groepen het eigen werkloosheidspercentage opgenomen. Dit is de enige verklarende variabele die niet afkomstig is uit het OSA-survey, maar uit CBS-publikaties.

UST In eerste instantie hadden wij 6 dummy's opgenomen die 6 categorieën van urbanisatiegraad aangaven. Deze variabelen hadden zó weinig verklarende kracht, dat wij ze hebben samengevoegd tot één dummy, die de waarde 1 aanneemt als de persoon in een stad woont (urbanisatiegraad C1 t/m C6) en 0 anders.





## 5 ARBEIDSMARKTGEDRAG VAN INDIVIDUEN

5.1 Algemeen

In dit hoofdstuk brengen wij verslag uit van de analysefase, waarin het arbeidsmarktgedrag verklaard wordt uit individuele kenmerken. Het individuele arbeidsmarktgedrag is voor mannen en vrouwen apart geanalyseerd. Bij een aantal verklarende variabelen is het namelijk mogelijk dat de invloed geheel anders is voor mannen dan voor vrouwen. Een duidelijk geval is VKD, de dummy die de aanwezigheid van een kind beschrijft. We verwachten dat de aanwezigheid van een kind bij vrouwen de kans op niet-participatie zal vergroten, terwijl bij mannen dit effect minder sterk, of misschien zelfs omgekeerd zal zijn. Het in Hoofdstuk 4 beschreven logitmodel is dus geschat op twee bestanden, één van vrouwen en één van mannen, waarvan alle benodigde variabelen beschikbaar zijn en die niet in de categorie "geen keuze" vallen. In deze categorie ligt het arbeidsmarktgedrag immers vast. Slechts voor twee variabelen is non-respons opgetreden: het uurloon (ULN) en het niet-arbeidsinkomen (INA). In deze fase van het onderzoek zijn alleen die individuen verwijderd voor wie deze variabelen niet bekend zijn. In de simulatiefase (zie Hoofdstuk 6) moeten ook hun huisgenoten buiten de analyse worden gehouden. Zoals uit Tabel 5.1/1 blijkt blijven er na de correctie voor nonrespons meer vrouwen dan mannen over. De verdeling van vrouwen en mannen over de drie categorieën van arbeidsmarktgedrag verschilt aanzienlijk.

Tabel 5.1/1 Verdeling van vrouwen en mannen over drie categorieën

categorie	vrouwen	mannen
A: participeert	35%	82%
B: wisselt	21%	15%
C: participeert niet	44%	4%
totaal aantal	1586	1386

Bij de interpretatie van deze cijfers moet men zich goed realiseren dat zij betrekking hebben op de arbeidsmarktpositie gedurende vijf jaar. In dat licht is het percentage vrouwen in categorie A, dat derhalve vrijwel onafgebroken heeft gewerkt, met 35% hoger dan men zou verwachten. Vervolgens valt het percentage in B lager uit dan voor de hand ligt, en in C hoger: nog altijd 44% van de vrouwen heeft in de afgelopen vijf jaar niet noemenswaard gewerkt. Dit kan lang niet allemaal aan de aanwezigheid van (jonge) kinderen worden toegeschreven. Bij de mannen valt op, dat 4% de hele vijf jaar van de arbeidsmarkt is uitgesloten. Dit cijfer is exclusief arbeidsongeschikten, vervroegd gepensioneerden en dagstudenten. Deze 4% vormt dus een harde kern van langdurig werklozen. De bedoeling van dit hoofdstuk is, te achterhalen welke eigenschappen van een persoon bepalen tot welke categorie zij of hij behoort.

De afhankelijke variabele in het model bestaat uit de waargenomen keuze: A, B of C. Zoals in Hoofdstuk 4.1 uiteengezet, is het logitmodel zeer geschikt voor de analyse van individuele keuze-processen. Het model is geschat met de methode van maximum likelihood, zodat de parameterschattingen verschillende gewenste eigenschappen hebben; zij zijn raak, asymptotisch efficiënt, en asymptotisch normaal verdeeld. Gezien het grote aantal waarnemingen kunnen wij deze asymptotische eigenschappen met enig vertrouwen vermelden.

De interpretatie van de geschatte coëfficiënten wordt vereenvoudigd als we denken in termen van "nut". Een individu moet een keuze doen uit één van de drie categorieën, en zal de meest aantrekkelijke categorie kiezen (de categorie die het hoogste nut oplevert). Omdat we alleen vergelijkenderwijs kunnen spreken (bijvoorbeeld de keuze A is "gunstiger" voor een bepaald individu dan de keuze C), kiezen we een referentiegroep. De keuze van de referentiegroep maakt voor de vergelijking tussen groepen geen verschil, en beïnvloedt onze conclusies niet. Dit kan worden geïllustreerd aan de hand van een eenvoudig voorbeeld. Stel we hebben in ons model alleen VKD (de dummy die de aanwezigheid van een kind beschrijft) als verklarende

variabele, en we vinden bij referentiegroep A de volgende coëfficiënten van deze variabele:

A: PARTICIPEERT	0	(referentiegroep)
B: WISSELT	2.4	
C: PARTICIPEERT NIET	3.0	

Voor een vrouw met een kind (VKD = 1) heeft dus B een hoger nut dan A, maar C weer hoger dan B. Nemen we B als referentiegroep dan verschuiven de coëfficiënten als volgt:

A: PARTICIPEERT	-2.4	
B: WISSELT	0	(referentiegroep)
C: PARTICIPEERT NIET	0.6	

Uit de geschatte coëfficiënten kunnen de kansen dat een individu één van deze drie categorieën kiest afgeleid worden. Er is dus een verschil tussen wat we waargenomen hebben en wat we voorspellen: waargenomen is de keuze uit de drie categorieën, en dat kan er maar één zijn, voorspeld wordt de kans op A, B en C voor ieder individu.

Zoals gezegd zijn de kansen een functie van de verklarende variabelen. Deze functie is kromlijng, zoals blijkt uit Figuur 4/1. Een goed beeld van de invloed van de verklarende variabelen op de kansen wordt gegeven door de pseudo-elasticiteiten van de kansen naar de verklarende variabelen. Deze grootheid is gedefinieerd als:

$$\frac{\partial P}{\partial \log x} = x \frac{\partial P}{\partial x} \quad (5/1)$$

of, in woorden: het aantal procentpunten dat de kans toeneemt wanneer de verklarende variabele met 1% stijgt. Is de pseudo-elasticiteit negatief, dan daalt de kans dat de betreffende categorie wordt gekozen bij een kleine toename van de verklarende variabele.

Natuurlijk is het onmogelijk dat door een kleine verandering in een verklarende variabele, de kansen op alle categorieën zouden stijgen. Wanneer de kans op één categorie stijgt, moet de kans op minstens één

andere categorie dalen. Met andere woorden: omdat de kansen tot één sommeren, sommeren de pseudo-elasticiteiten tot 0.

Omdat het verband tussen kansen en verklarende variabelen kromlijng is, hangen de pseudo-elasticiteiten af van de gekozen waarden van de verklarende variabelen, en daarmee ook van de waarde die P bij die waarden aanneemt. Wij zijn bij de berekening van de pseudo-elasticiteiten steeds van het steekproefgemiddelde uitgegaan.

## 5.2 Schattingsresultaten

De pseudo-elasticiteiten zijn gegeven in Tabel 5.2/1. Zij zijn berekend uit de geschatte coëfficiënten die staan vermeld in Bijlage B. De pseudo-elasticiteiten van het uurloon bedragen bij vrouwen voor de categorieën A, B en C respectievelijk 0.137, -0.087, en -0.049. Dit betekent dat wanneer voor alle vrouwen de uurlonen met 1% stijgen de gemiddelde participatie stijgt met 0.137 procentpunt, en het aandeel van de categorieën B en C daalt met respectievelijk 0.087 en 0.049 procentpunt. Dit kan worden verduidelijkt aan de hand van het voorbeeld van een fictieve vrouw waarvan het uurloon met 20% wordt verhoogd. Het effect van zo'n loonsverhoging kan als volgt worden weergegeven (de uitgangssituatie is willekeurig gekozen).

Schema 5.2/1 Geschatte effect van 20% loonsverhoging voor fictieve vrouw

categorie	kansen vóór loonsverhoging	verandering door loonsverhoging	kansen ná loonsverhoging
PARTICIPEERT	0.40	0.0274	0.4274
WISSELT	0.20	-0.0174	0.1826
PARTICIPEERT NIET	0.40	-0.0098	0.3902
totaal*	1.00	0.00	1.00

\* verschillen door afronding

Tabel 5.2/1 Pseudo-elasticiteiten van de kansen naar de verklarende variabelen, berekend in het steekproefgemiddelde

	VROUWEN			MANNEN		
	PARTIC.	WISSELT	PART. NIET	PARTIC.	WISSELT	PART. NIET
LFT	-0,165** (2,91)	-0,381** (7,94)	0,547** (8,85)	0,140** (2,79)	-0,120** (2,70)	-0,020 (0,90)
VKD	-0,289** (15,01)	0,072** (4,35)	0,217** (9,91)	-0,050** (3,02)	0,030* (2,08)	0,020* (2,45)
NIV	0,168* (2,58)	0,311** (5,43)	-0,479** (6,78)	0,091 (1,59)	-0,066 (1,30)	-0,025 (0,94)
INA	-0,200** (6,17)	0,040 (1,52)	0,160** (4,48)	-0,121** (9,75)	0,095** (8,69)	0,027** (5,24)
ULN	0,137** (3,14)	-0,087 (1,79)	-0,049 (0,99)	0,039 (0,77)	-0,045 (0,99)	0,006 (0,28)
RWL	-0,146* (2,24)	0,019 (0,36)	0,128 (1,87)	-0,278** (5,02)	0,205** (4,22)	0,072** (2,68)
BSP	0,015 (0,36)	-0,015 (0,45)	0,000 (0,00)	0,073** (2,67)	-0,034 (1,44)	-0,038** (3,23)
BSS	-0,004 (1,27)	0,000 (0,18)	0,004 (1,02)	-0,001 (0,76)	0,001 (0,65)	0,000 (0,47)
UST	0,015 (1,00)	0,025* (2,02)	-0,040* (2,49)	-0,025 (1,83)	0,015 (1,28)	0,010 (1,48)

absolute t-waarden tussen haakjes

\* significant op 5% niveau

\*\* significant op 1% niveau

De betekenis van de verklarende variabelen is als volgt:

LFT leeftijd in jaren  
 VKD dummy voor aanwezigheid van een kind in het huishouden  
 NIV opleidingsniveau (van 2 (laag) tot 6 (hoog))  
 INA niet-arbeidsinkomen in duizenden gulden per jaar  
 ULN uurloon in gulden  
 RWL regionale werkloosheidspercentage  
 BSP dummy voor wel/geen partner  
 BSS dummy voor wel/niet gescheiden  
 UST dummy voor wel/niet in stad wonen

De meeste effecten in Tabel 5.2/1 komen overeen met wat men zou verwachten. Het is opvallend dat de coëfficiënten van de mannen in de meeste gevallen

in absolute waarde kleiner zijn dan de coëfficiënten van de vrouwen. Kennelijk vertoont het arbeidsmarktgedrag van mannen minder variatie, en is het minder gevoelig voor veranderingen in hun omstandigheden.

Bij de coëfficiënten van de leeftijd is een toelichting vereist; zoals in Hoofdstuk 4 is vermeld hebben wij een kwadratisch verband verondersteld tussen leeftijd en arbeidsmarktgedrag. De pseudo-elasticiteit bij LFT is dan ook berekend uit twee coëfficiënten, die van de leeftijd zelf en die van de leeftijd kwadraat. Dit betekent dat men voor deze variabele zich goed moet realiseren dat de gegeven pseudo-elasticiteiten gelden voor het steekproefgemiddelde en elders heel andere waarden kunnen aannemen. Voor de andere variabelen geldt dit weliswaar ook, maar in mindere mate. Omdat het effect van de leeftijd zo complex is bespreken wij eerst de andere coëfficiënten, en komen dan op de leeftijd terug.

Zeer significant is de aanwezigheid van een kind bij vrouwen (VKD=1). De kans om te participeren neemt dan sterk af en de kans op geen participatie neemt sterk toe. Bij mannen zouden we het tegengestelde verwachten, maar dit is niet het geval, hoewel de invloed op de kansen klein is.

Een hogere opleiding leidt bij mannen tot een wat hogere kans op participatie, vooral ten koste van de categorie 'WISSELT'. Bij vrouwen is er een veel sterker effect, en wel een daling van de kans op 'PARTICIPEERT NIET', vooral ten bate van 'WISSELT'. Men kan dus stellen dat een hogere opleiding de kans op participatie in het algemeen vergroot, maar de precieze uitwerking verschilt naar geslacht.

Het uurloon blijkt bij mannen geen significante invloed te hebben op het arbeidsmarktgedrag. Bij vrouwen daarentegen heeft een hoog uurloon een positief effect op de participatie.

Mensen met een partner kiezen vaker voor participatie dan mensen zonder partner, terwijl gescheiden personen vaker in de categorieën 'WISSELT' of 'PARTICIPEERT NIET' vallen.

In vergelijking met de andere variabelen is het effect van de burgerlijke staat op het arbeidsmarktgedrag van vrouwen verwaarloosbaar klein. Dit resultaat wekt in eerste instantie verwondering omdat het in de rede ligt dat bijvoorbeeld gehuwde vrouwen een ander arbeidsmarktgedrag hebben dan ongehuwde. Blijkbaar wordt dit verschil in gedrag geheel verklaard door de andere variabelen die de gezinssituatie beschrijven: VKD (aanwezigheid

kind) en INA (niet-arbeidsinkomen). De burgerlijke staat op zichzelf heeft dus geen grote invloed op het arbeidsmarktgedrag; voor zover gehuwde vrouwen een ander gedrag vertonen dan ongehuwde wordt dit veroorzaakt door financiële factoren of door de aanwezigheid van jonge kinderen.

Het werkloosheidspercentage is als verklarende variabele opgenomen om het zgn. 'discouraged worker effect' te kunnen meten. Dit is het verschijnsel dat bij hoge werkloosheid mensen die eigenlijk zouden willen participeren, dit niet doen omdat zij verwachten toch geen werk te zullen vinden. In regio's met hoge werkloosheid blijkt de participatie inderdaad te dalen; ontmoedigde vrouwen stromen vooral naar de categorie 'PARTICIPEERT NIET' en mannen naar 'WISSELT'.

Ten slotte blijken mannen die in een stad wonen minder vaak te participeren, en vrouwen juist meer.

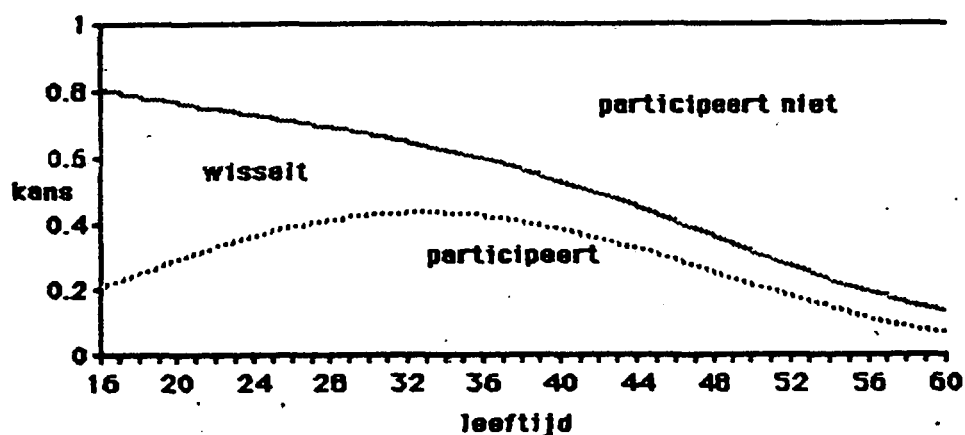
De Figuren 5.2/1 en 5.2/2 illustreren het verloop van de kansen bij respectievelijk vrouwen en mannen wanneer de leeftijd varieert van 16 tot 60 jaar. Overeenkomstig de verwachting is voor beide geslachten de kans op participatie relatief klein voor de zeer lage en de zeer hoge leeftijdsgroepen. Opvallend is dat meer dan de helft van de jonge vrouwen wisselend gedrag vertoont, tegen ongeveer een kwart van de jonge mannen. Voor vrouwen is de participatiekans het hoogst op 33-jarige leeftijd, voor mannen ligt de top bij 42 jaar. Dit verklaart waarom in Tabel 5.2/1 de pseudo-elasticiteiten voor vrouwen en mannen van teken verschillen. De gemiddelde leeftijd van vrouwen in de steekproef is 35,6 jaar. Dit is meer dan 33, de leeftijd waar de kans op categorie A maximaal is, en dus geldt in het steekproefgemiddelde dat de kans op A daalt als de leeftijd stijgt. Voor mannen geldt het omgekeerde; hun gemiddelde leeftijd is 37,05.

Het is verleidelijk, maar onjuist, om de Figuren 5.2/1 en 5.2/2 te interpreteren als een soort "arbeidsmarktlevensgeschiedenissen van de gemiddelde Nederlander". Ten eerste zou men door dat te doen, leeftijds- en cohorteffecten verwarren: het is absoluut niet zeker dat degenen die nu 25 jaar oud zijn, zich op 55-jarige leeftijd net zo zullen gedragen als de ouderen van nu. Een andere reden is dat bij het ouder worden in werkelijkheid niet alleen de leeftijd verandert. Ook de andere variabelen maken een ontwikkeling door. Bij de Figuren 5.2/1 en 5.2/2 is alleen de

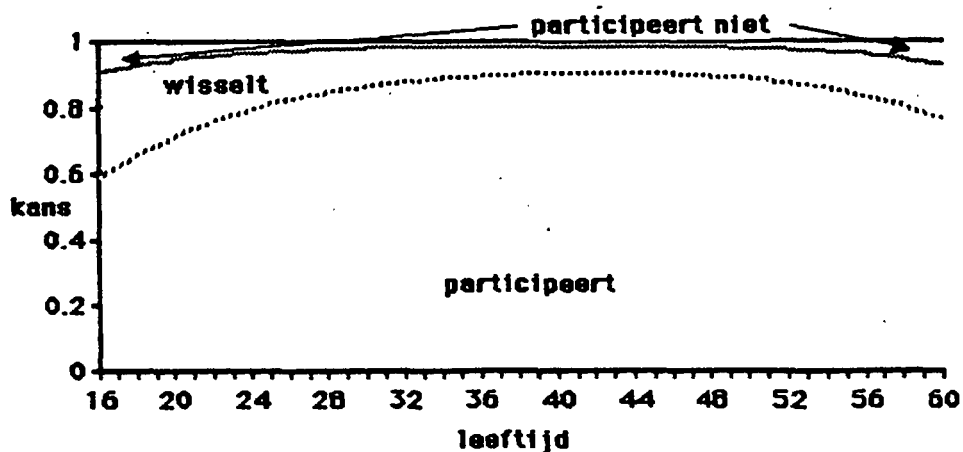


leeftijd gevarieerd, en de andere variabelen zijn vastgeprikt op het steekproefgemiddelde. Om werkelijk een realistische levensgeschiedenis te construeren moet men veronderstellingen maken over het verloop van de burgerlijke staat, aantal kinderen, inkomen en dergelijke van een persoon over een lange periode. Wij hebben hiervan afgezien.

Figuur 5.2/1 Verloop van de kansen als functie van de leeftijd bij vrouwen



Figuur 5.2/2 Verloop van de kansen als functie van de leeftijd bij mannen



Om inzicht te krijgen in het gecumuleerde effect van verschillende variabelen geven wij in Tabel 5.2/2 voor een aantal "standaardpersonen" de berekende kansen op elk van de drie toestanden. Een normindividu is een individu dat voor alle variabelen precies het steekproefgemiddelde heeft, zoals aangegeven in Schema 5.2/2.

Schema 5.2/2 De waarden van de verklarende variabelen voor de norm-individue

variabele	vrouw	man
LFT	35,600	37,050
VKD	0,512	0,499
NIV	3,734	4,041
INA	8,921	4,271
ULN	10,444	12,780
RWL	18,080	15,570
BSP	0,820	0,822
BSS	0,039	0,018
UST	0,548	0,564

Merk op dat een normindividu in werkelijkheid niet kan bestaan: ten eerste omdat de dummyvariabelen voor het normindividu ongelijk zijn aan 0 of 1, ten tweede omdat voor het normindividu de variabele "leeftijd kwadraat" niet het steekproefgemiddelde aanneemt (immers, het kwadraat van het gemiddelde is kleiner dan het gemiddelde van het kwadraat).

Tabel 5.2/2 Geschatte kansen op de toestanden A, B en C voor verschillende standaardpersonen

kans op	VROUWEN			MANNEN		
	A	B	C	A	B	C
norm individu	0,424	0,178	0,398	0,900	0,087	0,013
zonder kind	0,725	0,094	0,181	0,924	0,069	0,007
met kind	0,179	0,230	0,591	0,867	0,110	0,023
NIV= 2 (laagst)	0,312	0,083	0,605	0,869	0,112	0,018
NIV= 6 (hoogst)	0,470	0,361	0,170	0,923	0,068	0,009
INA - 2	0,473	0,167	0,360	0,929	0,062	0,009
INA + 2	0,377	0,188	0,435	0,861	0,121	0,019
ULN - 5	0,356	0,220	0,425	0,889	0,099	0,012
ULN + 5	0,495	0,141	0,364	0,910	0,077	0,014
extreme types:						
A	0,821	0,138	0,041	0,974	0,024	0,001
C	0,031	0,168	0,801	0,449	0,406	0,145

Tabel 5.2/2 geeft eerst de kansen voor het normindividu\*, en vervolgens de kansen voor een aantal individuen die slechts in één opzicht van het normindividu afwijken. De effecten komen overeen met wat we verwachten op grond van de pseudo-elasticiteiten uit Tabel 5.2/1. Zo is bijvoorbeeld het gevolg van de aanwezigheid van een kind bij vrouwen veel groter dan bij mannen. Ook hier blijkt dat het arbeidsmarktgedrag van mannen minder flexibel is dan dat van vrouwen.

In het tweede deel van Tabel 5.2/2 onderzoeken wij het gecumuleerde effect van de verklarende variabelen. Daartoe hebben wij, zowel voor vrouwen als voor mannen, een "extreem A-type" en een "extreem C-type" geconstrueerd. Dat zijn individuen met zodanige waarden van de verklarende variabelen, dat de kans op A respectievelijk C zo groot mogelijk is. De kenmerken van de extreme types staan weergegeven in Schema 5.2/3.

Schema 5.2/3 Kenmerken van de extreme types

variabele	extreem A-type		extreem C-type	
	vrouw	man	vrouw	man
LFT	33	42	25	35
VKD	0	0	1	0
NIV	6	6	2	2
INA	10	5	16	8
ULN	15	15	9	9
RWL	10	10	20	20
BSP	1	1	0	0
BSS	0	0	1	1
UST	1	0	0	1

Overigens is er bij de constructie van de extreme types wel enigszins rekening gehouden met de praktijk. Zo had bijvoorbeeld bij het mannelijke extreme C-type de combinatie "geen partner" en "wel een kind" een nog grotere kans op categorie C opgeleverd; omdat deze combinatie van kenmerken

\* de afwijking ten opzichte van de fracties in Tabel 5.1/1 wordt voor het overgrote deel veroorzaakt door het kwadratische verband in de leeftijd.

bij Nederlandse mannen zelden voorkomt is er voor gekozen om hem slechts één van deze kenmerken toe te kennen. Natuurlijk kan de lezer zelf, met behulp van de coëfficiënten uit Bijlage B en Formules (4/5) en (4/6), voor iedere denkbare combinatie van verklarende variabelen de bijbehorende kansen berekenen.

Ook uit de resultaten van de extreme types blijkt dat het gedrag van mannen meer vastligt dan dat van vrouwen; zelfs een man met typische C-kenmerken heeft nog een kans van bijna 45% om in categorie A te vallen.

Een analogon van  $R^2$  bij regressie, dus een eenvoudig getal dat aangeeft of de verklarende kracht van het model goed of slecht is, bestaat niet voor het logitmodel. Een aanwijzing voor de kwaliteit van de schattingen kunnen wij krijgen door de "waargenomen" en de "voorspelde" toestand van ieder individu te vergelijken. Het zou natuurlijk het mooiste zijn als voor elk individu de waargenomen en voorspelde toestand overeenkomen. In de Tabellen 5.2/3 en 5.2/4 wordt dit onderzocht voor respectievelijk vrouwen en mannen.

Tabel 5.2/3 Waargenomen versus voorspelde keuze MANNEN

voorspelde keuze	waargenomen keuze			TOTAAL
	A	B	C	
A: PARTICIPEERT	1105	168	36	1309
B: WISSELT	27	35	13	75
C: PARTICIPEERT NIET	0	1	1	2
TOTAAL	1132	204	50	1386

Tabel 5.2/4 Waargenomen versus voorspelde keuze VROUWEN

voorspelde keuze	waargenomen keuze			TOTAAL
	A	B	C	
A: PARTICIPEERT	355	102	85	542
B: WISSELT	27	82	41	150
C: PARTICIPEERT NIET	169	150	575	894
TOTAAL	551	334	701	1586

Wanneer het model perfect zou voorspellen, zouden alle waarnemingen geconcentreerd zijn in de diagonalen van de tabellen. Als "voorspelde toestand" hebben wij genomen de toestand waarvan de berekende kans het hoogst was. Deze kans is berekend door de geschatte coëfficiënten in te vullen in Vergelijking (4/1).

Uit deze vergelijkingen komt een eigenschap naar voren die het logitmodel wel vaker vertoont: voor toestanden die niet vaak voorkomen voorspelt het model niet echt goed op individueel niveau. Dit is vooral duidelijk bij de niet-participerende mannen, waarvan voor slechts één de waargenomen toestand juist wordt voorspeld.

Het feit dat het model op individueel niveau de zeldzame toestanden niet goed voorspelt, vormt geen probleem. Wij zijn immers geïnteresseerd in het voorspellen van frequenties op een grote populatie, en niet van het gedrag van een specifieke persoon. Met behulp van het logitmodel kunnen wij uitspraken doen in de trant van "een gehuwde man met leeftijd X heeft een kans  $P_A$  om in categorie A te vallen,  $P_B$  om in categorie B te vallen en  $P_C$  om in categorie C te vallen". In uitspraken als "meneer X participeert" zijn wij niet geïnteresseerd.

### 5.3 Conclusie

Overzien wij deze resultaten, dan zou men haast vergeten dat de analyse niet, zoals gebruikelijk, betrekking heeft op een momentopname, maar op de arbeidsmarktpositie over een veel langere termijn. De verdeling over de drie categorieën verschilt niet zo heel veel van het traditionele beeld, met veel vrouwen en weinig mannen die blijvend en volledig buiten de arbeidsmarkt staan (zie Tabel 5.1/1). Bij de vrouwen is dit lang niet allemaal het gevolg van werkloosheid, of van de aanwezigheid van (jonge) kinderen, maar vermoedelijk toch ook een kwestie van voorkeur. Bij de mannen is werken nog steeds de norm, maar toch is er een duidelijke kern van langdurig inactieven aanwezig.

Bij een nadere analyse van de oorzaken van de arbeidsmarktpositie op de lange termijn vinden wij dezelfde variabelen en dezelfde effecten; zij het misschien niet met precies dezelfde coëfficiënten. Eigenlijk zou men dit nog eens precies willen vaststellen door een numerieke vergelijking van de elasticiteiten in deze twee gevallen, en de verleiding is groot de hier gevonden waarden te vergelijken met elasticiteiten uit ander onderzoek. Die werkwijze roept echter als regel allerlei vragen van definitie en methode op; wij laten zo'n vergelijking daarom achterwege. Wat wèl een goed antwoord op de gestelde vraag zou geven is een herhaling van de huidige analyse, met dezelfde verklarende variabelen, maar met een indeling van personen op grond van hun huidige arbeidsmarktpositie en niet die van de afgelopen vijf jaar. Alleen is nog niet duidelijk hoe wij categorie B ('wisselt') dan moeten definiëren, ja zelfs of er voor die categorie wel plaats is.

Op zichzelf suggereert de overeenstemming in de determinanten van de structurele arbeidsmarktpositie en de analyses van momentopnamen dat de lange termijn arbeidsmarktpositie niet wezenlijk verschilt van die op korte termijn. Natuurlijk is het onjuist - vooral in het geval van vrouwen - om het behoren tot categorie C (PARTICIPEERT NIET) gelijk te stellen met onvrijwillige werkloosheid. Ongetwijfeld zal een gedeelte van de personen in categorie C uit eigen vrije wil verkiezen om niet te werken. Dat is echter niet in tegenspraak met de conclusie dat langdurige werkloosheid of voortdurende participatie geen andere oorzaken en geen ander karakter hebben dan korte termijn verschijnselen. Hetzelfde proces beheerst beide categorieën. Of er daarbinnen nog een zelfversterkend effect van werkloosheidsduur optreedt laat onze analyse echter in het midden; daarvoor is een andere aanpak, zoals in het onderhanden onderzoek van Theeuwes gevolgd, noodzakelijk. Onze conclusie betekent dat langdurige werkloosheid wel bijzondere aandacht van het beleid verdient wegens zijn ongewenste sociale gevolgen, maar dat er geen aanleiding is voor een specifieke keuze van beleidsinstrumenten: lange werkloosheid kan bestreden worden met dezelfde middelen als kortstondige werkloosheid, want zij heeft dezelfde oorzaken, en dezelfde doelgroep wordt er door getroffen.



## 6 ARBEIDSMARKTGEDRAG VAN GEZINNEN

### 6.1 Vraag en methode

In dit hoofdstuk besteden wij aandacht aan de tweede onderzoeksvraag. Deze is of er op het punt van arbeidsmarktgedrag onderlinge beïnvloeding optreedt tussen de individuen die deel uitmaken van één huishouden, afgezien van het voor de hand liggende inkomenseffect. Dit laatste houdt in dat de prikkel om door werk een inkomen te verwerven des te groter is naarmate het individu minder middelen aan het overige gezinsinkomen kan ontlenen\*. Wat wij nu willen onderzoeken is of er, in de individuele arbeidsmarktpositie op de langere termijn, daarenboven nog uit anderen hoofde sprake is van onderlinge beïnvloeding. Deze kan het karakter dragen van onderlinge navolging (voorbeeldwerking), of, daaraan tegengesteld, van afschrikking: wat één individu doet, doet een ander in hetzelfde huishouden juist niet (compensatie). Daarbij moet men bedenken dat het niet alleen gaat om individueel gedrag, met een zekere keuzevrijheid, zoals de terminologie suggereert, maar dat er ook een invloed kan uitgaan van opgelegde keuzes, zoals in het geval van invaliditeit.

De vraag of er sprake is van een dergelijke onderlinge beïnvloeding en, zo ja, in welke richting, kan worden beantwoord aan de hand van statistische toetsen die in Appendix C nader worden beschreven. Een van deze toetsen kan op verschillende niveau's worden toegepast, en aldus worden gebruikt om door herhaalde toetsing van steeds preciezer vragen voor steeds specifiekere deelgroepen van de steekproef de onderlinge beïnvloeding (zo die bestaat) te localiseren.

Voor een beter begrip van onze werkwijze moeten wij eerst een korte schets geven van de statistische techniek, waarbij wij voor de details naar

---

\* Wij gebruiken de termen 'gezin' en 'huishouden' door elkaar.



Appendix C verwijzen. Als uitgangspunt dient dat wij op grond van onze eerdere onderzoeking voor ieder individu de kans kunnen uitrekenen dat zij of hij tot categorie A, B of C behoort, gegeven haar of zijn individuele omstandigheden - met inbegrip van het gezinsinkomen waarin zij of hij zonder te werken kan delen. Op grond van deze kansen kan nu ook worden berekend hoe groot de kans voor ieder huishouden van bekende samenstelling is dat de gezinsleden naar arbeidsmarktpositie een bepaald patroon vertonen indien geen onderlinge beïnvloeding optreedt. Dit is het geval van onafhankelijkheid van de individuele arbeidsmarktposities. Wij zullen nagaan of de waargenomen samenstelling van huishoudens hiermee in overeenstemming is of niet. Is dit het geval, dan is er geen beïnvloeding, en valt er niets meer te onderzoeken.

Is dit niet het geval, dan zullen wij vervolgens de kans nagaan dat de huishoudens aan een bepaalde voorwaarde voldoen. Zo'n voorwaarde kan bijvoorbeeld luiden:

"Voorwaarde X: Alle volwassen leden van het huishouden hebben dezelfde arbeidsmarktpositie, hetzij A, B of C."

Men kan ook menen dat de "tussenpositie" B niet scherp genoeg is gedefinieerd om in een definitie van homogeen gedrag te worden gebruikt. Dit leidt tot

"Voorwaarde Y: Alle volwassen leden van het huishouden hebben hetzij arbeidsmarktpositie A, hetzij arbeidsmarktpositie C."

Ook zijn ingewikkelder voorwaarden denkbaar, zoals bijvoorbeeld

"Voorwaarde Z: Gegeven dat een huishouden uit twee personen van verschillend geslacht bestaat behoort de man tot categorie A en de vrouw niet."

Zoals men ziet gaat het steeds om een enkelvoudige voorwaarde, die voor ieder huishouden maar twee mogelijkheden open laat: het voldoet aan de voorwaarde of niet.

Wij vergelijken nu de kans, die bij onafhankelijkheid uit de kenmerken van de individuen volgt, met de feitelijke frequentie in de steekproef van huishoudens die aan de gestelde voorwaarde voldoen. In het geval van onafhankelijkheid komen kans en frequentie overeen. Dit geval zal niet vaak optreden, want als er inderdaad geen beïnvloeding is zouden wij de analyse al na de eerdere toetsing hebben gestaakt. Treedt er wèl een afwijking tussen berekende kans en waargenomen frequentie op, dan is er onderlinge beïnvloeding. Wij kunnen nu ook vaststellen in welke richting deze werkt, en zijn aard vervolgens nader preciseren.

Als voorbeeld nemen wij de zojuist geformuleerde voorwaarde X (alle gezinsleden dezelfde arbeidsmarktpositie). Als wij vinden dat er méér gezinnen aan voorwaarde X voldoen dan te verwachten was, is er blijkbaar positieve onderlinge beïnvloeding of voorbeeldwerking. Men kan nu menen dat dit vooral geldt voor A en C, maar niet voor de "tussenpositie" B die niet scherp genoeg is gedefinieerd om tot homogeen gedrag te concluderen als meerdere gezinsleden tot B behoren. Dit leidt tot een onderzoek van voorwaarde Y, zoals hierboven gegeven. Zo kunnen wij, als daar aanleiding toe bestaat, voortgaan met steeds specifiekere hypothesen over afwijkingen van het geval van onafhankelijkheid te toetsen\*.

Een en ander kan het best worden verduidelijkt door het toe te passen. Wij doen dit achtereenvolgens voor de huishoudens die, behalve kinderen (of bejaarden), juist twee of drie volwassen personen bevatten waarvan de arbeidsmarktpositie kan worden vastgesteld. Huishoudens met één volwassene zijn voor de analyse van onderlinge beïnvloeding niet bruikbaar, en van huishoudens met vier of meer volwassenen zijn er niet genoeg in de steekproef.

---

\* Strikt genomen zijn opeenvolgende toetsen niet onderling onafhankelijk, zodat men met de interpretatie moet oppassen.

## 6.2 Huishoudens met twee (volwassen) personen

In de OSA steekproef zijn 1057 huishoudens met (behalve kinderen of bejaarden) precies twee personen tussen de 16 en 60 jaar waarvan de arbeidsmarktpositie bekend is, en waarvoor wij over voldoende gegevens beschikken om de geschatte kansen volgens het model van Hoofdstuk 5 te berekenen. Van deze  $2 \times 1057 = 2114$  personen zijn er 185, of bijna 9%, die geen keuze hebben; zij zijn veroordeeld tot de activiteit van categorie A omdat zij in militaire dienst zijn, of zij zijn niet actief (categorie C) als scholier, invalide of VUT-er.

Nu kan de gedwongen arbeidsmarktpositie van het éne gezinslid de keuze van de ander beïnvloeden, maar het omgekeerde is uiteraard niet het geval. Voor een symmetrische analyse, waarbij geen nader onderscheid tussen de twee betrokken personen wordt gemaakt, valt er wat voor te zeggen de gezinnen met één (of meer) personen zonder keuze weg te laten. Dit scheelt 162 gezinnen, hoofdzakelijk door de aanwezigheid van gedwongen non-actieven van categorie C.

Zonder nader onderscheid van de twee personen naar hun rol of positie in het gezin kan een huishouden zes verschillende combinaties van de twee arbeidsmarktposities vertonen. In Schema 6.2/1 vindt men de verdeling over deze mogelijkheden telkens drie maal afgebeeld, namelijk de waargenomen verdeling, de berekende verdeling (in het geval van onafhankelijkheid), en de verschillen. De som van deze verschillen over de zes cellen is, als gevolg van de toegepaste methode, altijd precies nul; wat er van één combinatie te veel is, is er van een andere te weinig. Wij geven deze opstelling in aantallen en in procenten.

Het eerste dat opvalt is dat de afwijkingen betrekkelijk gering zijn. Voor de aanpassingstoets (zie Appendix C) vinden wij

$$\chi^2 = 5,52.$$

Dit is, zelfs bij correctie van het aantal vrijheidsgraden tot 3, niet significant. Er is, met andere woorden, geen aanwijzing dat er onderlinge beïnvloeding in arbeidsmarktgedrag optreedt tussen de leden van één gezin, afgezien dan van de directe inkomenseffecten.

Schema 6.2/1 Verdeling van gezinnen met twee volwassen personen over de mogelijke combinaties

---



---

ALLEEN PERSONEN MET VRIJE KEUZE

---



---

AANTALLEN

	waargenomen			berekend			verschil		
	A	B	C	A	B	C	A	B	C
A	244	210	355	223	224	366	+21	-14	-11
B		33	43		27	46		+6	-3
C			10			9			+1

---

PERCENTAGES VAN 895

---

	waargenomen			berekend			verschil		
	A	B	C	A	B	C	A	B	C
A	27,3	23,5	39,7	24,9	25,0	40,9	+2,3	-1,6	-1,2
B		3,7	4,8		3,0	5,1		+0,6	-0,3
C			1,1			1,0			+0,1

---



---

ALLE PERSONEN

---



---

AANTALLEN

	waargenomen			berekend			verschil		
	A	B	C	A	B	C	A	B	C
A	245	210	419	224	224	438	+21	-14	-19
B		33	65		27	64		+6	+1
C			85			80			+5

---

PERCENTAGES VAN 1057

---

	waargenomen			berekend			verschil		
	A	B	C	A	B	C	A	B	C
A	23,2	19,9	39,6	21,2	21,2	41,4	+2,0	-1,3	-1,8
B		3,1	6,1		2,6	6,1		+0,6	+0,1
C			8,0			7,6			+0,5

Voor zover er afwijkingen zijn, wijzen die naar een positieve samenhang: van de homogene gezinnen van de types AA, BB en CC vinden wij er meer dan verwacht, van de andere minder. Bezien wij de indelingstoets voor de voorwaarde

"V1: alle leden van het huishouden hebben  
dezelfde positie, hetzij A, B of C"

dan komt er een toetsgrootte uit met de waarde 2,27\*. Dit bevestigt de indruk dat er een tendens is naar homogeniteit of voorbeeldwerking. Heel erg sterk is het effect echter niet, zoals blijkt wanneer wij het verder trachten te localiseren bij één van de drie categorieën A, B of C. Wij vinden dan lagere waarden.

Betrekken wij ook huishoudens met personen zonder keuze in de beschouwing dan vinden wij het beeld van de tweede helft van Schema 6/1. Zoals men ziet behoren deze personen overwegend tot de niet-actieven van categorie C: van de 85 huishoudens van het type CC zijn er bijvoorbeeld maar 10 waarin beide individuen een keuze hebben kunnen maken. De resultaten van de toetsen verschillen echter nauwelijks. Voor de aanpassingstoets vinden wij

$$\chi^2 = 5,22.$$

weer met 3 vrijheidsgraden - opnieuw geen aanwijzing dat er beïnvloeding optreedt. Voor de indelingstoets voor V1 vinden wij een waarde van 2,41. Er is dus een lichte, positieve samenhang.

Dit blijkt ook als wij andere, speciale gevallen onderzoeken. De opvatting "één harde werker is wel genoeg" komt overeen met de voorwaarde

"V2: één lid van het gezin behoort tot  
categorie A en het andere lid niet"

---

\* Onder de nulhypothese heeft deze toetsgrootte een standaardnormale verdeling. Zie Appendix C.

Dit komt minder voor dan uit onafhankelijkheid zou volgen (verwacht aantal gevallen 662, waargenomen 629, toetsgroottheid -2,33). Hetzelfde geldt voor de specifieke norm "de man is kostwinner", weergegeven in de voorwaarde

"V3: de man behoort tot categorie A  
en de vrouw niet."

Dit hebben wij getoetst voor de huishoudens die uit een man en een vrouw (afgezien van eventuele jonge kinderen) bestaan, en waarin geen personen zonder keuze voorkomen. Dit zijn er 885. Daarvan voldoen er 519 aan V3 tegen een verwacht aantal van 540. Er is derhalve geen extra belemmering van de arbeidsmarktparticipatie van de vrouw gelegen in de arbeid van haar partner. Daarbij moet men natuurlijk wèl bedenken dat in het individuele kansmodel een duidelijke invloed bleek van geslacht en kinderverzorging, en dat deze effecten in de onderhavige analyse als gegeven worden aangenomen. Op de arbeidsmarkt is de vrouw er slecht aan toe, zeker als ze voor kinderen moet zorgen, maar de aanwezigheid van een werkende man maakt het niet erger.

### 6.3 Huishoudens met drie volwassen personen

De steekproef bevat 163 huishoudens met drie volwassen personen met bekende arbeidsmarktpositie en volgens ons eerder model berekenbare kansen. Van deze  $163 \times 3 = 489$  personen hebben er echter 164, of een derde van het totaal, geen keuze. Dit is veel meer dan bij de huishoudens met twee volwassenen. Het betreft overwegend thuiswonende scholieren en studenten. Laten wij de huishoudens waarin ten minste één persoon geen keuze heeft, weg, dan blijven er nog maar 48 over. Dit is te weinig voor een verantwoorde statistische analyse. De gegevens zijn dus óf vertekend, óf te gering in aantal om een goed antwoord te geven op de vraag naar onderlinge beïnvloeding. Wij geven de uitkomsten dan ook hoofdzakelijk uit de behoefte aan volledige rapportage.

De waargenomen en berekende aantallen vindt men in de Schema's 6.3/1 en 6.3/2. De opstelling is minder overzichtelijk dan bij twee personen, toch kan men wel zien dat de afwijkingen klein zijn, en geen duidelijk patroon vertonen.

Schema 6.3/1 Verdeling van gezinnen met drie volwassen personen over alle mogelijke combinaties, alleen personen met vrije keuze

WAARGENOMEN		
AAA 6	AAB 5 AAC 15	ABC 10
BBB 1	ABB 2 BBC 0	
CCC 0	ACC 9 BCC 0	
BEREKEND		
AAA 6	AAB 7 AAC 14	ABC 11
BBB 0	ABB 2 BBC 2	
CCC 0	ACC 4 BCC 1	
YERSCHIL		
AAA 0	AAB -2 AAC 1	ABC -1
BBB 1	ABB 0 BBC -2	
CCC 0	ACC 5 BCC -1	

Schema 6.3/2 Verdeling van gezinnen met drie volwassen personen over alle mogelijke combinaties, alle personen

WAARGENOMEN		
AAA 8	AAB 6 AAC 41	ABC 19
BBB 1	ABB 2 BBC 1	
CCC 9	ACC 64 BCC 12	
BEREKEND		
AAA 9	AAB 8 AAC 44	ABC 23
BBB 0	ABB 2 BBC 3	
CCC 10	ACC 50 BCC 13	
VERSCHIL		
AAA -1	AAB -2 AAC -3	ABC -4
BBB 1	ABB 0 BBC -2	
CCC -1	ACC 14 BCC -1	

De toetsen bevestigen dat er geen duidelijke afwijking van onafhankelijkheid optreedt. Wij geven de waarden eerst voor 48 huishoudens, dan (tussen haakjes) voor de 163 huishoudens waaronder met mensen zonder keuze.



Voor de aanpassingstoets vinden wij

$$\chi^2 = 15,25 (9,60).$$

met 9 vrijheidsgraden. Het eerste cijfer is bijna significant. Het is echter de vraag of de gebruikte benaderingen bij slechts 48 waarnemingen nog wel opgaan. Het tweede geval, voor 163 gezinnen, is helemaal niet significant.

Voor de indelingstoetsen hebben wij gekeken naar

"V4: alle leden van het huishouden behoren tot dezelfde categorie, hetzij A, B of C."

Hiervoor vinden wij de toetsgrootheid 0,15 (-0,40), hetgeen geheel overeenkomt met afwezigheid van onderlinge beïnvloeding. Dit blijft zo als wij het criterium voor homogeniteit wat afzwakken tot

"V5: ten minste 2 van de 3 personen behoren tot dezelfde categorie."

Dit levert de waarde 0,34 (0,97).

De conclusie luidt dat er geen noemenswaardige onderlinge beïnvloeding kan worden aangetoond.

#### 6.4 Conclusie

De conclusie van dit hoofdstuk is dat er geen sterke aanwijzingen zijn voor onderlinge beïnvloeding van het arbeidsmarktgedrag van leden van één gezin, na inachtneming van de determinanten van dat gedrag die in Hoofdstuk 5 zijn genoemd. Als er al onderlinge beïnvloeding is dan is die positief (navolging, voorbeeldwerking), maar heel zwak. Men kan dit effect verwaarlozen.

Aan de ene kant is het teleurstellend dat er geen aanwijzingen zijn voor het interessante verschijnsel dat wij dachten te vinden. Aan de andere kant bevestigt het resultaat dat het model van Hoofdstuk 5 ter verklaring van lange termijn arbeidsmarktgedrag althans op het punt van de persoonlijke omgeving volledig is, en de individuele keuze afdoende beschrijft. Met deze validatie moeten wij tevreden zijn.



## LITERATUUR

Bekkering, J.M., Y.K. Grift en J.J. Siegers (1986), Belasting- en premieheffing en de arbeidsmarktparticipatie door gehuwde vrouwen, een econometrische analyse, Ministerie van Sociale Zaken en Werkgelegenheid, Den Haag

Cramer, J.S. (1986), Econometric Applications of Maximum Likelihood Methods, Cambridge

CBS, Standaard Onderwijs Indeling SOI 1978

Doornik, J.A. (1985), LOGITJD, een computerprogramma voor het schatten van onafhankelijke logitmodellen, Interfaculteit der Actuariële Wetenschappen en Econometrie der Universiteit van Amsterdam, AE Note N5/85

Hartog, J. en J.J.M. Theeuwes (1983), "De onstuitbare opkomst van de werkende gehuwde vrouw", ESB, pp 1152-1157

Kapteyn, A. en I. Woittiez (1986), Arbeidsaanbod en Voorkeursvorming, een verkennende studie naar de dynamiek in het gedrag van aanbieders van arbeid, Organisatie voor Strategisch Arbeidsmarktonderzoek, werkdokument W23, Den Haag

Kooreman, P. (1986), Essays on Micro-econometric Analysis of Household Behaviour, Tilburg

Opstal, R van en J. Theeuwes (1985), Jeugdige werklozen en hun kans op een baan, OSA-werkdocument nr. 14, Den Haag

Siegers, J.J. (1983), Arbeidsduurverkorting en het arbeidsaanbod door gehuwde vrouwen, Ministerie van Sociale Zaken en Werkgelegenheid, Den Haag

Siegers, J.J. (1985), Arbeidsaanbod en Kindertal, een micro-economische analyse, Groningen

Vissers, A.M.C., A.M. de Vries en Th. Schepens (1986), Arbeidsmarktgedrag ten tijde van massale werkloosheid, Den Haag, OSA-voorstudie nr. 12

Woittiez, I. (1986), Een verslag over OSA-data 1985, intern rapport, Katholieke Universiteit Brabant, Tilburg

## APPENDIX A HET OSA-SURVEY

In het OSA-rapport 'Arbeidsmarktgedrag ten tijde van massale werkloosheid' worden enige gegevens verstrekt over het OSA-survey. Het survey is gebaseerd op een steekproef uit het PTT-afgiftepuntenbestand. De doelgroep van het onderzoek was de 'potentiële beroepsbevolking van Nederland', gedefinieerd als 'personen in de leeftijd van 16 tot en met 60 jaar, voorzover zij geen dagstudent of dienstplichtig militair zijn'. Indien in een huishouding één of meer personen in aanmerking kwamen voor ondervraging, maar de hoofdkostwinner daarbuiten viel (b.v. door leeftijd) werd de hoofdkostwinner toch ondervraagd. Bij nonrespons werd via de random-route methode een vervangend adres gezocht. Toch is er kennelijk nog wat nonrespons opgetreden, want op de 2387 benaderde afgiftepunten werden 2132 gesprekken met gezinnen gerealiseerd, die in totaal 6348 personen bevatten. 4020 van hen zijn ondervraagd; dat waren dus, volgens het voorgaande, blijkbaar 2132 hoofdkostwinners en nog eens 1888 gezinsleden tussen de 16 en 60, geen scholier of militair.

In bijlage 1 van het OSA-rapport wordt de representativiteit van de steekproef onderzocht. De onderzoekers concluderen dat de onderzoeksgroep en een aantal interessante subgroepen daarvan naar een groot aantal kenmerken niet significant verschillen van de populatie (resp. deelpopulatie). Dit geldt ten aanzien van geslacht, huishoudensgrootte, regionale spreiding, duur van de werkloosheid, arbeidstijd en type ABW-uitkering. Een aantal deelgroepen zijn echter in de steekproef systematisch ondervertegenwoordigd: jongeren, RWW-ers, werklozen en zelfstandigen. Deze groepen zijn minder responsgeneigd.

Het OSA-survey bevat een grote hoeveelheid informatie over de 'levenshouding' en de werksatisfactie van de ondervraagde personen. Deze informatie is in het onderhavige onderzoek niet gebruikt. Wij beperken ons tot feitelijke informatie over de respondenten.

Drie variabelen, te weten het onderwijsniveau, het inkomen en het aantal arbeidsuren, zijn door ons geconstrueerd uit een vrij grote hoeveelheid andere gegevens.

### Onderwijsniveau

Het OSA-survey bevat 56 variabelen die informatie verschaffen over de opleiding van de respondent. Wij hebben hieruit twee variabelen geconstrueerd, die het opleidingsniveau beschrijven:

- niv1: niveau van de hoogst gevolgde opleiding
- niv2: niveau van de hoogst voltooide opleiding

Hierbij is uitgegaan van de Standaard Onderwijs Indeling (SOI) van het CBS. Volgens deze indeling wordt het niveau bepaald door de nominale duur van de opleiding.

De variabele niv1 is berekend door voor elk individu van alle opleidingen die hij/zij gevolgd heeft, de opleiding met het hoogste niveau te kiezen. Voor niv2 is dezelfde procedure gevolgd, nu met de restrictie dat uitsluitend de met een diploma, akte of getuigschrift afgeronde opleidingen werden bekeken.

In beide gevallen bleven er een aantal individuen over van wie geen enkel onderwijsniveau bekend was. Indien deze individuen de lagere school hadden doorlopen, werden zij ingedeeld op het laagste niveau. De restgroep werd ingedeeld op basis van het eindjaar van de laatst gevolgde (voltooide) opleiding. Indien dit eindjaar onwaarschijnlijk hoog was werd het individu 2e niveau, 2e trap toegekend (13-15 jaar nominale opleidingsduur).

De resulterende verdeling over niveau's is in Tabel A/1 vermeld, waarbij de kolom 'niveau bekend' de respondenten bevat waarvan het opleidingsniveau bekend was. In de kolom 'toegekend' staan die respondenten waarvoor wij het niveau uit andere gegevens (zoals het eindjaar van opleiding) berekend hebben.

Tabel A/1 Verdeling steekproef over opleidingsniveau's

niveau	niv1		niv2	
	hoogst bekend	gevolgde niveau toegekend	hoogst bekend	voltooid niveau toegekend
1e niveau	13	309	8	749
2e niveau, 1e trap	987	1	936	0
2e niveau, 2e trap	1770	6	1577	71
3e niveau, 1e trap	728	0	537	0
3e niveau, 2e trap	206	0	142	0
totaal	3704	316	3200	820

Om de representativiteit van de steekproef te onderzoeken hebben wij een vergelijking gemaakt met de Arbeidskrachtentelling 1983, waarvan de gegevens beschikbaar zijn. Bij deze vergelijking kunnen enkele kanttekeningen gemaakt worden.

- de selectiecriteria van de AKT komen grotendeels, maar niet helemaal overeen met de selectiecriteria van de OSA-enquête:
- er zijn alleen gegevens beschikbaar van de AKT '83 en deze worden vergeleken met een enquête uit 1985. Het is echter onaannemelijk dat er zich grote sprongen in het opleidingsniveau van de beroepsbevolking hebben voorgedaan.
- de gegevens van de AKT '83 kunnen alleen goed vergeleken worden met niv2 (hoogst afgeronde opleiding), omdat het behalen van een diploma, akte of getuigschrift bij de AKT een voorwaarde is om bij een bepaald niveau ingedeeld te worden.

Tabel A/2 Vergelijking van de OSA-enquête met de AKT.

	AKT '83	- OSA - enquête -	
		"niv2" (voltooid)	"niv1" (gevolgd)
1e niveau	24.9	18.8	8.0
2e niveau, 1e trap	30.9	23.3	24.6
2e niveau, 2e trap	32.1	41.0	44.2
3e niveau, 1e trap	9.0	13.4	18.1
3e niveau, 2e trap	3.0	3.5	5.1

(cijfers zijn procenten van de onderzoeksgroep)

Zoals men ziet zijn er aanzienlijke verschillen tussen het onderwijsniveau van de AKT '83 en niv2 uit de OSA-enquête. Deels kan dat het gevolg zijn van de arbitraire methode waarmee voor ongeveer 20% van de individuen (namelijk indirect aan de hand van het aantal jaren onderwijs) de variabele niv2 berekend werd. Daar staat tegenover dat voor de variabele niv1 het niveau van de hoogst gevolgde opleidingen voor een zeer groot gedeelte direkt uit het datamateriaal kon worden gefilterd. Men mag aannemen dat een variabele die het hoogst gevolgde niveau aangeeft gemiddeld een hogere inschatting geeft van het niveau dan de hoogst afgeronde opleiding. De verschillen met de AKT'83 zijn echter zo groot dat het vermoeden bestaat dat in de onderzoeksgroep van de OSA individuen met een hoge opleiding aanzienlijk oververtegenwoordigd zijn.

De variabele NIV die wij in de analyse van Hoofdstuk 5 gebruiken correspondeert met niv1.

### Inkomens

Uit de vele variabelen die het inkomen beschrijven hebben wij acht soorten inkomen geconstrueerd, zoals aangegeven in Tabel A/3. De naam is de naam van de SPSS-variabele die wij aan de betreffende inkomenssoort verbinden. 'Nonrespons' slaat op personen die de inkomensvraag niet wilden of konden beantwoorden.

Uit Tabel A/3 blijkt dat de nonrespons op de vraag naar bijkomende financiële voordelen van werken bijzonder hoog is (1429 van de 2502 werkers). Bovendien werden er soms wel erg onwaarschijnlijke bedragen opgegeven bij deze variabele. Daarom laten wij deze inkomenscomponent in het vervolg buiten beschouwing.

Ook de individuele huursubsidie laten wij buiten beschouwing, omdat dit een inkomensafhankelijke regeling is en omdat er soms evident foute bedragen zijn opgegeven.



Tabel A/3 Enige gegevens over inkomens in het OSA-survey

naam	non-respons	groep	betaling (periode)	opmerkingen
INKLOON	137	werknemers	netto (per maand)	voor zelfstandigen opgenomen bij INKZELF
INKZELF	38	zelfstandigen	verschilt	uit verschillende variabelen geconstrueerd
INKZWRT	0	allen	per maand	zwart inkomen
INKFINV	1429	werkers	netto (verschilt)	bijkomende financiële voordelen van baan (soms vreemde antwoorden)
INKPENS	180	allen	netto (per maand)	VUT, AOW, pensioen
INKSTUD		allen	bruto (per jaar)	beurs, bijdrage ouders
INKUITK		allen	netto (per maand)	WAO, WW, WWV, RWW, ABW, AWW. kinderbijslag
INKEXTR		allen	bruto (per jaar)	alimentatie, vermogen, verhuur

Het inkomen van zelfstandigen is berekend aan de hand van de antwoorden op verschillende vragen; indien het inkomen van 1985 niet bekend was is het inkomen uit het laatst bekende jaar opgenomen, met een correctie voor het prijspeil.

Aangezien sommige bedragen alleen bruto bekend waren moest er een correctie voor belasting en sociale premies worden toegepast; op grond van gegevens uit de Nationale Rekeningen is de gemiddelde netto/bruto verhouding in 1983, 1984 en 1985 op 0,64 gesteld. De netto/bruto correctie is, waar nodig toegepast bij de inkomenscomponenten INKZELF en bij INKEXTR, maar is achterwege gelaten bij het studie-inkomen omdat hierover in het algemeen geen belasting wordt betaald.

In de analysefase van het onderzoek worden de zeven inkomenscomponenten die hier worden onderscheiden, geaggregeerd tot twee variabelen: arbeidsinkomen en niet-arbeidsinkomen. De totale nonrespons op het inkomen bedraagt dan 339. Deze individuen en al hun gezinsleden moesten uit het bestand worden verwijderd. Voor de anderen is het arbeidsinkomen gebruikt voor de constructie van het uurloon ULN, en het arbeidsinkomen en niet-arbeidsinkomen van alle gezinsleden voor de constructie van het individuele niet-arbeidsinkomen INA (zie Paragraaf 4.2).

### Arbeidsuren

Het aantal arbeidsuren is noodzakelijk om het uurloon van werkenden te berekenen. In het vragenformulier van de OSA-enquête komen op verschillende plaatsen vragen voor over het aantal arbeidsuren, die telkens een nuance verschillen. Wanneer wij ons beperken tot de vragen over de situatie op het tijdstip van ondervraging dan leidt dit tot het volgende lijstje:

- 13d hoeveel uren werkt u gemiddeld per week (reisuren niet meetellen);
- 13n verricht u ... regelmatig thuiswerk ... zo ja, hoeveel uren per week werkt u hier gemiddeld aan;
- 14b hoeveel uren gemiddeld per week verricht u nu in totaal betaald werk;
- 25 voor hoeveel uren per week bent u thans in dienst ... het aantal uren dat u met uw werkgever overeengekomen bent;
- 26a hoeveel uren werkt u gemiddeld per week ... als u eventuele overuren ook meerekent;
- 26b hoeveel betaalde overuren zitten daar gemiddeld dan in;
- 26c hoeveel onbetaalde overuren zitten daar gemiddeld in;
- 26d (geconstrueerde variabele: v26a-v26b-v26c, de enquêteur moet controleren of dit aantal overeenkomt met v25);
- 83b (i.v.m. zwart werk) om hoeveel uren gaat het gemiddeld per maand.

Het onderlinge verband tussen deze vragen is weergegeven in Schema A/1. Stippellijntjes betekenen dat uit de formulering niet duidelijk blijkt of

een variabele wel of niet moet worden meegerekend.

Schema A/1 Het onderlinge verband tussen de vragen over uren in de OSA-enquête

v26a {	onbetaalde overuren	v26c	} v14b
	betaalde overuren	v26b	
	uren in hoofdbaan excl. overuren	v25 $\stackrel{?}{=}$ v26d	
	uren in extra baan	(restterm)	
	uren thuiswerk	v13n	
	uren zwart werk	v83b	

} v13d

Voor ons doel, het berekenen van het uurloon, is de variabele v26d het meest geschikt. Het arbeidsinkomen is namelijk gegeven exclusief overwerk, bijbaan enzovoort. Ook inkomen uit zwart werk is buiten beschouwing gebleven.

Door Woittiez (1985) is er op gewezen dat een aantal respondenten de verschillende vragen over uren op tegenstrijdige wijze hebben beantwoord. Zulke individuen moeten uit de steekproef verwijderd worden. In Schema A/2 geven wij aan welke tegenstrijdigheden er zoal voorkomen volgens Woittiez. De aantallen verschillen hier en daar van die van Woittiez; de oorzaak is waarschijnlijk een verschil in de "missing values". Dat is de codering van antwoordcategorieën zoals "niet van toepassing", "wil niet zeggen" enzovoort. De normale procedure in het OSA-survey is om "missing values" de codes 96 t/m 99 te geven, maar in de urenvariabelen komen ook een aantal keren waarden tussen 50 en 95 voor. Volgens het IVA berust dit waarschijnlijk op codeerfouten. Veel van deze waarnemingen verdwijnen door het verwijderen van de tegenstrijdigheden, maar niet alle. Daarom hebben wij, behalve degenen die tegenstrijdige antwoorden gaven, ook die respondenten verwijderd die in loondienst waren maar bij vraag 26d

aangaven dat hun baan, exclusief overuren, 50 of meer arbeidsuren per week in beslag nam. Voor zelfstandigen hebben wij aangenomen dat zulke lange werkweken wel reëel zijn.

Schema A/2 Tegenstrijdigheden in antwoorden op vragen over arbeidsuren in het OSA-survey

Tegenstrijdigheid	aantal waarnemingen
1. v25 ≠ v26d	4
2. v25 > v26a	2
3. v26a ≠ v26b + v26c + v26d	3
4. v14b < v26d	17
5. v14b > v26d geen betaalde overuren } geen extra baan } geen thuiswerk }	100
6. v13d > v26a geen extra baan } geen thuiswerk }	65
7. v13d > v14b geen onbetaalde overuren }	54
8. v13d < v14b geen extra baan } geen thuiswerk }	29

Het totaal aantal personen dat uit de steekproef moet worden verwijderd vanwege de variabele "arbeidsuren" is 272. Een aantal van hen zou toch al verwijderd zijn wegens nonrespons op het inkomen. In Tabel A/4 geven wij de frequentieverdeling van het aantal arbeidsuren na uitsluiting van de 272 individuen. Zoals men ziet werken verreweg de meeste mensen nog gewoon 40 uur per week.

In de analysefase van het onderzoek worden alleen individuen verwijderd, maar in de simulatiefase leidt dat tot verwijdering van de desbetreffende huishoudens. Wij kunnen immers alleen huishoudens gebruiken waarvoor de gegevens van alle leden bekend zijn. Tabel A/5 laat het effect zien van de (individuele) nonrespons op inkomen en arbeidsuren. Er blijven voor de simulatiefase slechts 1630 gave huishoudens over. Eén daarvan moest alsnog verwijderd worden omdat niet alle relevante leden waren ondervraagd. Van de 1629 huishoudens die overblijven zijn er 345 niet gebruikt omdat zij slechts één persoon bevatten (afgezien van bejaarden en kinderen), en 64

niet omdat zij uit meer dan drie personen bestaan.

Tabel A/4 Verdeling aantal arbeidsuren per week, exclusief nonrespons

uren per week	aantal	uren per week	aantal	uren per week	aantal
1	1	21	4	42	9
2	2	22	9	43	3
3	4	23	4	44	2
4	12	24	15	45	18
5	12	25	21	46	2
6	15	26	4	48	4
7	6	27	3	50	7
8	16	28	8	52	1
9	16	29	9	53	1
10	23	30	19	55	3
11	7	31	4	56	1
12	26	32	84	57	1
13	7	34	5	60	7
14	5	35	16	62	1
15	23	36	39	68	1
16	9	37	22	70	5
17	7	38	293	76	1
18	8	39	18	80	5
19	3	40	2491	85	2
20	104	41	4		

Tabel A/5 Verdeling huishoudens over 2 categorieën nonrespons

arbeidsuren	inkomen		totaal
	nonrespons	geen nonrespons	
nonrespons	52	204	256
geen nonrespons	246	1630	1876
totaal	298	1834	2132

## APPENDIX B HET ECONOMETRISCHE MODEL

In deze appendix geven wij enkele formules en schattingsresultaten voor het in dit rapport behandelde multinominale logitmodel.

Wij gaan uit van  $N$  individuen en  $K$  typen arbeidsmarktgedrag. De vector van verklarende variabelen geven wij aan met  $x_i$ , de parametervector met  $\beta_j$ . Hierin duidt  $i$  een individu aan en  $j$  een toestand.

Wij definiëren:

$$y_{ij} = \begin{cases} 1 & \text{als individu } i \text{ gedrag type } j \text{ vertoont,} \\ 0 & \text{anders} \end{cases} \quad (\text{B/1})$$

Het logitmodel luidt:

$$\left\{ \begin{array}{l} P_{ij} = P\{y_{ij} = 1\} = \frac{\exp(x_i' \beta_j)}{\sum_{k=1}^K \exp(x_i' \beta_k)} \\ \beta_1 = 0 \end{array} \right. \quad (\text{B/2})$$

De loglikelihood functie is:

$$\log L(\beta) = \sum_{i=1}^N \sum_{j=1}^K y_{ij} \log P_{ij}(\beta) \quad (\text{B/3})$$

De elementen van  $\beta$  zijn bepaald door deze functie te maximeren.

Een vergelijking met het baseline-model, dat is een versie van het model met alleen een constante term als verklarende variabele, geeft inzicht in de verklarende kracht van het model. Het maximum van de loglikelihood van het baseline-model is

$$\log L_B = \sum_{j=1}^K N_j \log f_j \quad (B/4)$$

waarin:

$N_j$  = aantal waarnemingen in categorie  $j$   
 $f_j$  = fractie waarnemingen in categorie  $j$

De afgeleiden van de kansen naar de regressoren zijn

$$\frac{\partial P_j}{\partial x_h} = P_j \left\{ \beta_{jh} - \sum_{k=1}^K \beta_{kh} \frac{P_k}{P_j} \right\} \quad (B/5)$$

De log-odds ratio is tenslotte

$$\log \frac{P_{ij}}{P_{ik}} = x_i' (\beta_j - \beta_k) \quad (B/6)$$

In Tabel B/1 geven wij de schattingen van de coëfficiënten van het model, voor de regressoren van Hoofdstuk 4.2, met als referentiecategorie "A: PARTICIPEERT". Tabel B/2 geeft de vergelijkbare coëfficiënten bij de referentiecategorie "C: PARTICIPEERT NIET". De pseudo-elasticiteiten in Tabel 5.2/1 zijn berekend met de coëfficiënten uit Tabel B/1. De variabele SLF is het kwadraat van LFT (leeftijd).

Tabel B/1 Schattingscoëfficiënten bij referentiecategorie A: PARTICIPEERT

	VROUWEN		MANNEN	
	WISSELT	PART. NIET	WISSELT	PART. NIET
CONST	1.712 ( 1.61)	0.604 ( 0.61)	0.562 ( 0.50)	0.280 ( 0.15)
LFT	-0.254** ( 4.01)	-0.127* ( 2.37)	-0.221** ( 3.62)	-0.350** ( 3.48)
SLF/100	0.304** ( 3.52)	0.246** ( 3.50)	0.263** ( 3.43)	0.446** ( 3.57)
VKD	2.290** (10.94)	2.579** (13.73)	0.539* ( 2.28)	1.219** ( 2.61)
NIV	0.266** ( 2.62)	-0.420** ( 4.97)	-0.139 ( 1.37)	-0.197 ( 1.04)
INA	0.086** ( 4.34)	0.105** ( 5.84)	0.185** ( 8.98)	0.207** ( 5.95)
ULN	-0.077** ( 2.64)	-0.048* ( 2.48)	-0.028 ( 0.97)	0.009 ( 0.19)
RWL	0.028 ( 1.44)	0.039* ( 2.24)	0.112** ( 4.43)	0.151** ( 3.00)
BSP	-0.136 ( 0.56)	-0.052 ( 0.19)	-0.393 ( 1.67)	-1.403** ( 3.32)
BSS	0.369 ( 0.81)	0.524 ( 1.25)	0.441 ( 0.68)	0.485 ( 0.53)
UST	0.132 ( 0.86)	-0.246 ( 1.81)	0.241 ( 1.40)	0.530 ( 1.58)

tussen haakjes: absolute t-waaden

\* : significant op 5% niveau

\*\* : significant op 1% niveau

INA luidt in duizenden guldens, ULN in guldens



Tabel B/2 Schattingscoëfficiënten bij referentiecategorie B: WISSELT

	VROUWEN PART. NIET	MANNEN PART. NIET
CONST	-1.108 ( 1.01)	-0.282 ( 0.14)
LFT	0.127* ( 1.97)	-0.129 ( 1.25)
SLF/100	-0.058 ( 0.67)	0.184 ( 1.43)
VKD	0.290 ( 1.31)	0.681 ( 1.38)
NIV	-0.686** ( 6.78)	-0.058 ( 0.29)
INA	0.019 ( 0.94)	0.022 ( 0.62)
ULN	0.029 ( 0.96)	0.037 ( 0.71)
RWL	0.011 ( 0.59)	0.039 ( 0.74)
BSP	0.084 ( 0.28)	-1.010* ( 2.31)
BSS	0.156 ( 0.34)	0.045 ( 0.04)
UST	-0.378* ( 2.57)	0.288 ( 0.82)

tussen haakjes: absolute t-waarden

\* : significant op 5% niveau

\*\* : significant op 1% niveau

INA luidt in duizenden guldens, ULN in guldens

## APPENDIX C HET ONDERZOEK NAAR DE ONDERLINGE BEINVLOEDING VAN LEDEN VAN EEN HUISHOUDEN

C1. In de tweede fase van het onderzoek willen wij nagaan of er sprake is van onderlinge beïnvloeding in arbeidsmarktgedrag tussen de leden van één huishouden en, zo ja, welk karakter deze dan heeft. Men kan zich indenken dat er een positieve samenhang is, in die zin dat de aanwezigheid van één actief (of niet-actief) persoon de kans vergroot dat andere leden van hetzelfde huishouden ook actief (of juist niet-actief) zijn. In dat geval is er sprake van voorbeeldwerking of navolging. Het tegendeel kan ook het geval zijn, wanneer de activiteit van de één de kans op activiteit van de ander vermindert, zodat er als het ware binnen het huishouden compensatie optreedt. Men kan zich ook nog andere, ingewikkelder patronen indenken, maar het is natuurlijk ook heel goed mogelijk dat er statistisch gesproken in het geheel geen onderlinge beïnvloeding valt vast te stellen.

In deze bijlage besteden wij aandacht aan de vraag hoe een en ander empirisch kan worden onderzocht, en welke statistische toetsen van toepassing zijn.

C2. Zoals bekend delen wij alle individuen op basis van hun arbeidsmarkt-geschiedenis van de laatste vijf jaar in naar hun structurele arbeidsmarktstatus. Wij onderscheiden daarbij drie categorieën, of, met weder invoering van de individuen die 'geen keuze' hebben, de volgende groepen:

Aa. voortdurend actief, uit voorkeur;

Ab. voortdurend actief, geen keuze (dit geldt voor dienstplichtigen);

B. afwisselend actief en niet-actief;

Ca. niet-actief, uit voorkeur;

Cb. niet-actief, geen keuze (scholier of dagstudent, invalide, zeer oud, enzovoort).

Wij gaan hier voorbij aan de vraag hoe individuen precies zijn ingedeeld, en aan de technische moeilijkheden die zich bij sommige groepen door de werkwijze van de OSA-enquête voordoen. Merk op dat de individuen met 'geen keuze', die bij de analyse van het keuzeproces terecht terzijde werden gesteld, thans weer in het onderzoek kunnen worden betrokken. Ook afgedwongen niet-activiteit van een arbeidsongeschikte kan immers een voorbeeldwerking hebben dan wel compenserend gedrag uitlokken, kortom het gedrag van anderen in het huishouden beïnvloeden. Het omgekeerde geldt echter niet. Daarom voeren wij de analyse tweemaal uit, éénmaal op alle gezinnen en éénmaal op gezinnen zonder personen in de categorie "geen keuze".

- C3. De eerste manier om het effect van de aanwezigheid van anderen in een bepaalde categorie op de keuze te onderzoeken zou zijn dat wij de status van huisgenoten als verklarende variabele voor het individuele gedrag invoeren. Technisch houdt dit in dat er drie variabelen  $Z_A$ ,  $Z_B$  en  $Z_C$  worden ingevoerd, gedefinieerd als de aantallen (andere) personen met arbeidsmarkt-status A, B of C in hetzelfde huishouden. Deze worden als gewone regressoren in het logit-model toegevoegd. Indien wij nu op toestand A normaliseren, dus de coëfficiënten voor toestand A nul stellen, mag men, afhankelijk van de gemaakte veronderstelling, een bepaald patroon voor de coëfficiënten verwachten. Bij voorbeeldwerking is dit patroon als volgt:

	coëfficiënt van		
	$Z_A$	$Z_B$	$Z_C$
toestand B	-	+	-
toestand C	-	-	+

Een voordeel van de methode is dat de feitelijke schattingen ook andere patronen kunnen vertonen, die zich voor interpretatie lenen: de mogelijkheden zijn niet beperkt tot van tevoren bedachte patronen.

Helaas is deze methode echter uit statistisch oogpunt niet toelaatbaar, omdat er (zo wordt nu juist verondersteld) mogelijkwijze beïnvloeding

over en weer optreedt. Dat wil zeggen dat de uitkomst van de keuze van het individu dat wij bestuderen zelf weer effect kan hebben gehad op de keuze van zijn huisgenoten, die wij als verklarende variabele gebruiken. Dit is in strijd met de veronderstellingen, waarop de schattingsmethode berust, en wij zien niet in hoe dit fundamentele gebrek kan worden hersteld.\*

- C4. Een statistische toets op onderlinge beïnvloeding is een toets van het tegendeel, dat is onafhankelijkheid van de kansen op toestand A, B en C van de individuen binnen één gezin, met inachtneming van de regressoren uit Hoofdstuk 5. Daarvoor gebruiken wij twee toetsen, de aanpassingstoets en de indelingstoets.

Voor beide toetsen beschouwen wij huishoudens met een gegeven aantal relevante personen, bijvoorbeeld twee personen, met bekende arbeidsmarktstatus en voldoende gegevens om uit het logit-model de kansen op de drie mogelijke posities te berekenen. Het is niet van belang of er naast deze twee relevante personen nog anderen (kinderen, bejaarden) tot het huishouden behoren, of niet.

- C5. Wij gaan nu als volgt te werk. Wij nummeren de huishoudens met de index  $h=1, 2, \dots, H$ , en de individuen binnen ieder huishouden met index  $j=1, 2, \dots, J_h$ . De combinatie  $hj$  is dus voldoende ieder individu aan te duiden.

Voor ieder individu kennen wij de (geschatte) kans dat zij of hij een van de drie arbeidsmarktstatusen heeft, dat is voor groep A, B, of C

$$\hat{p}_{jh}(A), \quad \hat{p}_{jh}(B), \quad \hat{p}_{jh}(C). \quad (C.1)$$

---

\* Strikt genomen brengt de mogelijkheid van onderlinge beïnvloeding met zich mee dat wij tijdens de analysefase van ieder huishouden slechts één (aselect gekozen) individu in de beschouwing zouden moeten betrekken. Vermoedelijk is het niet erg om van deze eis af te wijken; wij kunnen er echter desgewenst aan voldoen.

Als personen 'zonder keuze' mede in de analyse worden betrokken is voor hen telkens één van de kansen 1 en de andere twee zijn 0. Voor alle anderen volgen de kansen uit het logit-model van Hoofdstuk 5 en de geschatte coëfficiënten, toegepast op de eigenschappen van het betreffende individu.

Voor ieder huishouden  $h$  kan nu op grond van zijn samenstelling uit personen met bepaalde kenmerken en bijgevolg bepaalde kansen op de drie toestanden de kans worden berekend op een bepaalde samenstelling naar arbeidsmarktstatus. Zo is bijvoorbeeld de kans dat een huishouden van twee personen uit de combinatie (A,B) bestaat gelijk aan

$$P_h^*(A,B) = \hat{p}_{1h}(A) \cdot \hat{p}_{2h}(B) + \hat{p}_{1h}(B) \cdot \hat{p}_{2h}(A) \quad (C.2)$$

en net zo voor alle andere combinaties. Het aantal mogelijke combinaties voor gezinnen met twee personen is zes. Soms kunnen combinaties op een zinvolle manier worden samengevoegd tot een voorwaarde. Vanzelfsprekend kunnen wij de kans berekenen dat een gezin aan de voorwaarde voldoet wanneer wij de kansen op de afzonderlijke combinaties kennen. Zo kan de kans dat het huishouden uit twee personen van dezelfde categorie bestaat worden berekend als

$$P_h^*(\text{homogeen}) = P_h^*(A,A) + P_h^*(B,B) + P_h^*(C,C) \quad (C.3)$$

hetgeen door invulling van (C.2) overgaat in

$$P_h^*(\text{homogeen}) = \hat{p}_{1h}(A) \hat{p}_{2h}(A) + \hat{p}_{1h}(B) \hat{p}_{2h}(B) + \hat{p}_{1h}(C) \hat{p}_{2h}(C) \quad (C.4)$$

terwijl

$$Q_h^*(\text{homogeen}) = 1 - P_h^*(\text{homogeen}) \quad (C.5)$$

de kans is op een andere samenstelling. Men kan echter ook heel goed een andere eis dan uniformiteit van de arbeidsmarktstatussen aan een huishouden stellen; met enige kennis van combinatoriek kan de kans op

ieder bepaald geval voor ieder huishouden worden berekend, en als  $P_h^*$  worden behandeld. Soortgelijke formules vindt men voor huishoudens met drie personen.

C6. In de aanpassingstoets gaan wij uit van een tabel waarin de huishoudens verdeeld worden over alle mogelijke combinaties. Op deze tabel passen wij de klassieke  $\chi^2$  goodness-of-fit toets toe:

$$\chi^2 = \sum_k \frac{(n_k - \hat{n}_k)^2}{\hat{n}_k} \quad (C.6)$$

waarin  $\hat{n}_k$  het volgens (C.2) berekende aantal waarnemingen met combinatie  $k$  weergeeft, en  $n_k$  het waargenomen aantal.

Bij gezinnen met twee individuen zijn er in totaal zes combinaties: AA, AB, AC, BB, BC, CC, en bij drie personen zijn het er tien (zie Schema 6.3/1 in de tekst). Het aantal vrijheidsgraden van de toetsgroottheid (C.6) zou dus 5 respectievelijk 9 zijn, ware het niet dat een nadere correctie op zijn plaats is. De kansen waar de  $\hat{n}_k$  op berusten zijn namelijk zó berekend dat de gemiddelde kans op A, B en C overeenkomt met de waargenomen frequentie. Als ze precies geschat waren uit de waarnemingen die in de toets zijn betrokken zou dit restricties opleggen, immers

$$2\hat{n}(AA) + \hat{n}(AB) + \hat{n}(AC) \quad (C.7)$$

moet gelijk zijn aan het aantal personen in A, of

$$2n(AA) + n(AB) + n(AC) \quad (C.8)$$

Er zijn drie categorieën, of drie restricties, maar één is al begrepen in de gelijkheid van totale berekende en waargenomen aantallen. Resten er twee restricties, die in mindering moeten worden gebracht op het aantal vrijheidsgraden.

In feite zijn de kansen niet geschat uit de individuele waarnemingen

van de leden van de tweepersonshuishoudingen, maar het scheelt niet veel. Er is dus o.i. alle reden het aantal vrijheidsgraden van deze toets met 2 te verminderen.

Voor de gezinnen met 3 volwassen personen geldt deze redenering in veel mindere mate, omdat zij maar een klein gedeelte bevatten van de individuele waarnemingen waaruit de kansen zijn geschat, zeker als wij ons beperken tot de gezinnen zonder mensen uit de groep "geen keuze". Hier kan de correctie achterwege blijven.

C7. Voor de indelingstoets beschouwen wij de stochast met waarden 1 of 0:

$$Y_h$$

al naar het huishouden  $h$  wel of niet tot een bepaalde groep behoort. Er geldt dan

$$EY_h = P_h^*, \quad \text{var } Y_h = P_h^* Q_h^*,$$

met  $P^*$  en  $Q^*$  ontleend aan zulke kansen als in (C.4) en (C.5).

Voor  $\hat{n}^*$ , het aantal gezinnen van het gezochte type in de steekproef, geldt bij onafhankelijkheid van het individuele gedrag - want zo is  $P_h^*$  berekend! -

$$E\hat{n}^* = \sum_h P_h^*, \quad \text{var } \hat{n}^* = \sum_h P_h^* \cdot Q_h^*$$

zodat

$$Z = \frac{\hat{n}_h^* - \sum_h P_h^*}{\sqrt{\sum_h P_h^* \cdot Q_h^*}}$$

volgens de Centrale Limiet Stelling bij benadering standaardnormaal verdeeld is. Vergelijking van het feitelijk aantal huishoudens met deze norm leert dan of er significant van een afwijking in de ene of de andere richting sprake is.