

GESLACHT EN LEEFTIJD IN REGIONAAL ECONOMISCH ONDERZOEK
NAAR ARBEIDSPARTICIPATIE

J. Paul Elhorst*

Samenvatting

De doelstelling van dit artikel is om meer klaarheid te brengen in de relatie tussen geslacht, leeftijd en arbeidsparticipatie indien de laatste wordt geschat op basis van regionale data. Geslacht en leeftijd kunnen zowel aan de kant van de verklarende variabelen als aan de kant van de te verklaren variabele een rol spelen. Het eerste doet zich voor als geslacht en leeftijd worden gebruikt ter beschrijving van de samenstelling van de bevolking, een belangrijke verklarende variabele van arbeidsparticipatie. Aangezien vaak meerdere variabelen benodigd zijn om deze grootheid te beschrijven en deze variabelen in de regel sterk correleren, is het precieze effect van de samenstelling van de bevolking naar geslacht en leeftijd echter moeilijk vast te stellen. Het artikel gaat nader op dit probleem in en geeft een oplossing hiervoor. Het tweede doet zich voor als wordt overwogen niet de arbeidsparticipatie van de totale beroepsbevolking te verklaren maar van verschillende bevolkingsgroepen uitgesplitst naar geslacht en leeftijd. De vraag of dit zinvol is blijkt onder meer af te hangen van het doel dat aan het onderzoek wordt gesteld.

* J. Paul Elhorst
Faculteit der Economische Wetenschappen
Rijksuniversiteit Groningen
Postbus 800
9700 AV Groningen
tel. 050-633893

1. Inleiding

De statistieken laten er geen twijfel over bestaan. De participatiegraad van de potentiële beroepsbevolking uitgesplitst naar geslacht en leeftijd loopt sterk uiteen. In het algemeen is de participatiegraad van mannen in de leeftijdscategorie van 15-24 jaar laag, neemt de participatiegraad sterk toe bij de overgang naar de leeftijdscategorie van 25-34 jaar, bereikt deze zijn hoogtepunt in de leeftijdscategorie van 35-44 jaar en neemt de participatiegraad daarna vanaf 45 jaar eerst licht en vanaf 55 jaar vervolgens sterk af. De participatiegraad van vrouwen in de leeftijdscategorie van 15-24 jaar is aanvankelijk vrijwel gelijk aan die van mannen. Daarna doet zich bij vrouwen, bij de overgang naar de volgende leeftijdscategorie van 25-34 jaar, echter nauwelijks nog een stijging voor en zet zich al weer een geleidelijke daling in bij de overgang naar de leeftijdscategorie van 35-44 jaar.

In tegenstelling tot deze statistieken blijkt er meer twijfel te bestaan over de modelmatige relatie tussen geslacht, leeftijd en arbeidsparticipatie, getuige de verschillende aanpakken die in de literatuur gevolgd zijn. Deze verschillen zullen eerst in het kort worden beschreven. Verschillen met betrekking tot het geslacht doen zich in die zin voor dat er enerzijds auteurs (Steinle, 1983; Baumann et al., 1988; Nord, 1989; Gallaway et al., 1991) zijn te onderscheiden die de participatiegraad van de totale potentiële beroepsbevolking hebben onderzocht en anderzijds auteurs (Bowen en Finegan, 1969; Fleisher en Rhodes, 1976; Van der Veen, 1982; Siegers, 1983; Dickmann et al., 1989) die daarbij een uitsplitsing hebben gemaakt naar mannen en vrouwen. Voorts is er een omvangrijke groep van auteurs (Siegers en Zandanell, 1981; Van der Veen en Evers, 1983 en 1984; Molho, 1983; Molho en Elias, 1984; Op de Beke, 1988; Van den Bout et al., 1989) die alleen de participatiegraad van vrouwen heeft onderzocht, vaak met als argument dat de participatiegraad van mannen in de tijd en over regio's nauwelijks varieert.

Verschillen met betrekking tot leeftijd doen zich op vergelijkbare wijze voor. Molho (1983) en Molho en Elias (1984), die veruit de meeste aandacht besteden aan de relatie tussen arbeidsparticipatie en leeftijd, rangschikken het leeftijdsaspect onder de verklarende variabelen. Voor een groot aantal leeftijdsgroepen worden variabelen geconstrueerd die het aandeel van deze leeftijdsgroepen in de totale bevolking dienen aan te geven. Vervolgens wordt één vergelijking geschat ter verklaring van de participatiegraad. Van der Veen (1982) daarentegen verwerkt het leeftijdsaspect in de te verklaren variabele in die zin dat voor elke leeftijdsgroep (15-24, 24-49 en 50-64 jarigen) een aparte vergelijking wordt geschat. In aansluiting hierop is het denkbaar dat een combinatie wordt gebruikt van deze twee. Een bekend voorbeeld daarvan is de omvangrijke empirische studie van Bowen en Finegan (1969). Naast de tweedeling op grond van geslacht onderscheiden zij zowel bij de te verklaren

variabele als bij de verklarende variabelen verschillende leeftijdsgroepen. Tenslotte mag ook de studie van Op de Beke (1988) niet onvermeld blijven. Net als Van der Veen (1982) verwerkt hij het leeftijdsaspect in de te verklaren variabele, in dit geval voor 10 leeftijdsgroepen. Vervolgens schat hij echter niet één aparte vergelijking voor elke leeftijdsgroep, maar veronderstelt hij dat alle leeftijdsgroepen hetzelfde participatiegedrag vertonen. Op grond hiervan kan met slechts één vergelijking worden volstaan, die dan bovendien geschat kan worden op basis van een groter aantal waarnemingen, te weten het oorspronkelijke aantal maal het aantal onderscheiden leeftijdsgroepen.

Aangezien tot op heden niet duidelijk is welke modelmatige aanpak onder welke omstandigheden de voorkeur verdient, stelt het onderhavige onderzoek zich ten doel meer klaarheid te brengen in deze materie. Daartoe is het onderzoek als volgt opgezet. Eerst wordt de theoretische achtergrond van het modelleringsvraagstuk besproken, waarbij ervan wordt uitgegaan dat de participatiegraad wordt verklaard met behulp van regressieanalyse. Vervolgens zullen de bevindingen van deze theoretische analyse worden geïllustreerd aan de hand van een empirisch onderzoek, een onderzoek dat is gebaseerd op arbeidsparticipatiegegevens die betrekking hebben op de regio's van de EG over de periode 1983-1989. Tenslotte wordt een aantal conclusies getrokken.

2. Theoretische achtergrond

Een onderscheid naar geslacht en leeftijd roept vragen op zowel rechts als links van het gelijkheidsteken in de regressievergelijking, respectievelijk aan de kant van de verklarende variabelen en aan de kant van de te verklaren variabele. Een onderscheid aan de rechterkant van het gelijkheidsteken roept de vraag op van multicollineariteit. Een aantal van de genoemde onderzoeken in de inleiding heeft overduidelijk aangetoond dat de samenstelling van de bevolking naar geslacht en leeftijd invloed uitoefent op de mate van arbeidsparticipatie. Het punt is echter dat de samenstelling van de bevolking in beginsel niet in één kengetal is te vangen. In feite is de samenstelling van de bevolking een complex geheel waarin diverse variabelen een rol spelen. In de literatuur wordt meestal gebruik gemaakt van variabelen die de omvang of het aandeel van verscheidene bevolkingsgroepen in het totaal van de bevolking aangeven. Zo onderscheidden Molho en Elias (1984) maar liefst 7 van deze variabelen ter beschrijving van de samenstelling van de bevolking. Vanzelfsprekend kunnen al deze variabelen tegelijkertijd in de regressievergelijking worden opgenomen, maar dit leidt onvermijdelijk tot een grote mate van correlatie tussen de verklarende variabelen, beter bekend als multicollineariteit.

Een onderscheid naar geslacht en leeftijd aan de linkerkant van het gelijkheidsteken roept de vraag op onder welke omstandigheden dit zinvol is. Het waarom van deze vraag zal worden geïllustreerd aan de hand van een

voorbeeld zonder dat dit gepaard gaat met verlies aan algemeenheid. Stel dat een keuze gemaakt kan worden uit het verklaren van de participatiegraad van de beroepsbevolking in zijn geheel of die van mannen en vrouwen ieder afzonderlijk. In wezen heeft men dan de keuze uit twee modellen. Een eerste model waarin zowel de participatiegraad van mannen als van vrouwen wordt verklaard

$$Y_m = WP_m / LF_m = a_m' X_m + e_m, \quad (1a)$$

$$Y_v = WP_v / LF_v = a_v' X_v + e_v, \quad (1b)$$

en een tweede model waarin de participatiegraad van de totale beroepsbevolking wordt verklaard

$$Y_t = \frac{WP_m + WP_v}{LF_m + LF_v} = a_t' X_t + e_t. \quad (2)$$

Merk op dat $Y_t = W_m Y_m + W_v Y_v$ met $W_m = LF_m / (LF_m + LF_v)$ en $W_v = LF_v / (LF_m + LF_v)$. Hierbij is Y gedefinieerd als de participatiegraad, X als de verklarende variabelen, e als de storingsterm, WP als de beroepsbevolking - de werkenden en de werklozen - en LF als de totale potentiële beroepsbevolking, de laatste twee beide met betrekking tot de leeftijdscategorie van 15-64 jaar. Aangenomen is dat de verklarende variabelen in deze drie vergelijkingen niet dezelfde behoeven te zijn. Elke vergelijking kan één of meer verklarende variabelen bevatten die niet in de andere twee vergelijkingen voorkomen. De reden om uit te gaan van verschillende X is dat een onderscheid naar verschillende groepen vaak wordt toegepast op grond van de mogelijkheid om andere verklarende variabelen te kunnen kiezen.

Op grond van het bovenstaande voorbeeld kan worden aangetoond dat het antwoord op de vraag of een uitsplitsing naar mannen en vrouwen bij de analyse van arbeidsparticipatie zinvol is in de eerste plaats afhangt van de doelstelling van het onderzoek. Een denkbare doelstelling is dat men het arbeidsaanbod van verschillende groepen wil verklaren. Het Nederlandse overheidsbeleid richt zich op het bevorderen van de arbeidsparticipatie (zie WRR, 1990), met name van die groepen die een zwakke positie op de arbeidsmarkt innemen zoals vrouwen, jongeren en ouderen. In het licht van deze doelstelling is een uitsplitsing van het arbeidsaanbod naar verschillende geslacht- en leeftijdsgroepen bij voorbaat aan te bevelen, omdat op deze wijze onderzocht kan worden of individuen behorende tot deze groepen al of niet ander gedrag vertonen. Mocht in dat geval blijken dat mannen en vrouwen inderdaad anders reageren op bepaalde verklarende variabelen, dan kan de overheid haar beleid daarop afstemmen door bijvoorbeeld enerzijds alleen voor die beleidsmaatregelen te opteren die de grootste bijdrage leveren aan de arbeidsparticipatie van vrouwen en

anderzijds die beleidsmaatregelen achterwege te laten die onbedoeld de arbeidsparticipatie van mannen bevorderen. Mocht daarentegen blijken dat mannen en vrouwen hetzelfde gedrag vertonen, dan kunnen zij net zo goed onder één noemer worden gebracht. Uitgaande van het voorbeeld in vergelijking (1a) en (1b) en onder de aanname dat de X variabelen in beide vergelijkingen dezelfde zijn kan middels de hypothese $H_0: a_m = a_v$ getoetst worden of twee groepen al dan niet hetzelfde gedrag vertonen. Deze toets kan worden uitgevoerd door de vergelijkingen (1a) en (1b) en ieder afzonderlijk te schatten op basis van n waarnemingen en gezamenlijk op basis van $2n$ waarnemingen. Het aantal verklarende variabelen is in alle gevallen gelijk aan k . Zij $\hat{e}_m'\hat{e}_m$ de residuele kwadratensom van de 1^e vergelijking, $\hat{e}_v'\hat{e}_v$ de residuele kwadratensom van de 2^e vergelijking en $\hat{e}_t'\hat{e}_t$ de residuele kwadratensom van de gezamenlijke vergelijking, dan is de toetsgrootte

$$\frac{(\hat{e}_t'\hat{e}_t - \hat{e}_m'\hat{e}_m - \hat{e}_v'\hat{e}_v)/k}{(\hat{e}_m'\hat{e}_m + \hat{e}_v'\hat{e}_v)/2n} \quad (3)$$

F-verdeeld met $(k, 2n)$ vrijheidsgraden.

Anders wordt het als een onderzoek zich ten doel stelt slechts het arbeidsaanbod van de totale beroepsbevolking te verklaren. Onder deze omstandigheid is het nut van een uitsplitsing naar mannen en vrouwen niet langer gegarandeerd. Het staat namelijk niet op voorhand vast dat als de verklaringskracht van de afzonderlijke vergelijkingen voor mannen en vrouwen toeneemt ook de verklaringskracht van de per waarneming gewogen som van deze twee vergelijkingen toeneemt. De schattingen van $\hat{a}_m$ en $\hat{a}_v$ worden verkregen door de residuele kwadratensom van vgl (1a) en vgl (1b) te minimaliseren, respectievelijk

$$\Sigma (Y_m - a_m'X_m)^2 \text{ en } \Sigma (Y_v - a_v'X_v)^2, \quad (4)$$

en de schatting van $\hat{a}_t$ door de residuele kwadratensom van vgl (2) te minimaliseren

$$\Sigma (Y_t - a_t'X_t)^2. \quad (5)$$

De vraag is dan langs welke weg de hoogste verklaringsgraad van het arbeidsaanbod van de beroepsbevolking in zijn geheel wordt gehaald. Langs directe weg door de schatting van vgl (2) of langs indirecte weg door per waarneming vgl (1a) en vgl (1b) bij elkaar op te tellen rekening houdend met de bijbehorende gewichten W_m en W_v . Het eerste resulteert in de residuele kwadratensom

$$\Sigma (Y_t - \hat{a}_t' X_t)^2 = \Sigma \hat{e}_t^2, \quad (6)$$

en het tweede in de residuele kwadratensom

$$\Sigma [W_m(Y_m - \hat{a}_m' X_m) + W_v(Y_v - \hat{a}_v' X_v)]^2 = \Sigma [W_m \hat{e}_m + W_v \hat{e}_v]^2. \quad (7)$$

Op grond van deze vergelijkingen is niet op voorhand aan te geven welke aanpak de beste resultaten zal geven. Per waarneming kunnen de residuen $\hat{e}_m$ en $\hat{e}_v$ vanwege de grotere verklaringskracht van de afzonderlijke vergelijkingen weliswaar kleiner zijn, maar omdat ze vervolgens bij elkaar worden opgeteld is het niet uitgesloten dat de som van deze twee het residu van de vergelijking in zijn geheel overtreft. Of dat werkelijk het geval is en ook geldt voor de residuele kwadratensom over alle waarnemingen zal van geval tot geval verschillen en dient nader onderzocht te worden.

3. Een empirische analyse

Om de theorie die hierboven is besproken aan de hand van een praktijkvoorbeeld te illustreren is een onderzoek gedaan naar de regionale participatiegraad in de EG. Met behulp van gegevens van Eurostat is een databestand geconstrueerd bestaande uit 835 waarnemingen die betrekking hebben op 146 verschillende regio's in de 12 landen van de EG over de periode 1983-1989.

Uitgangspunt bij de empirische verklaring van arbeidsparticipatie vormen de volgende verklarende variabelen:

- het regionale werkloosheidscijfer
- het netto uurloon van arbeid
- de sociale zekerheidsuitgaven aan arbeidsongeschiktheid per hoofd van de potentiële beroepsbevolking
- aantal jaren dat voor het bereiken van de 65-jarige leeftijd via een daartoe getroffen regeling uit het arbeidsproces kan worden getreden
- het aandeel van de werkgelegenheid in de landbouw
- het aandeel van de werkgelegenheid in de industrie
- het aandeel mannen werkzaam in deeltijd
- de scholingsgraad van de potentiële beroepsbevolking
- de samenstelling van de bevolking naar leeftijd en geslacht gemeten met behulp van twee principale componenten, zie de tekst voor een nadere toelichting

De keuze van deze verklarende variabelen is gebaseerd op eerder onderzoek dat op dit terrein is verricht (zie Elhorst, 1993).

Met betrekking tot geslacht en leeftijd biedt de beschikbare dataset de volgende mogelijkheden. Aan de linkerkant van het gelijkheidsteken kan een

uitsplitsing worden gemaakt naar maximaal 10 verschillende groepen in de beroepsbevolking (mannen en vrouwen in 5 verschillende leeftijdsklassen: 15-24, 25-34, 35-44, 45-54 en 55-64) en aan de rechterkant van het gelijkheidsteken kunnen maximaal 14 verschillende variabelen worden onderscheiden ter beschrijving van de samenstelling van de bevolking (mannen en vrouwen in 7 verschillende leeftijdsklassen: 0-14, 15-24, 25-34, 35-44, 45-54, 55-64 en 65+).

In eerste aanleg is onderzocht in hoeverre multicollineariteit onder de geslacht- en leeftijdsvariabelen aan de rechterkant van het gelijkheidsteken een potentieel probleem vormt. Dit is gedaan door het percentage kruislingse correlatiecoëfficiënten tussen de 14 geslacht- en leeftijdsvariabelen te bepalen dat groter is dan 0.5. Hieruit bleek dat maar liefst 33% (30 van de 91) groter is dan 0.5. Evenzo bleek dat 3.3% groter is dan 0.9. Dit duidt erop dat multicollineariteit inderdaad een potentieel probleem vormt en dat een regressievergelijking waarin meerdere geslacht- en leeftijdsvariabelen worden opgenomen tot moeilijk interpreteerbare resultaten zal leiden. Om die reden is besloten het aantal geslacht- en leeftijdsvariabelen te reduceren volgens de methode van principale componentenanalyse. Een lastig probleem hierbij vormt de bepaling van het aantal componenten. Ofschoon statistische toetsgrootheden bij de oplossing van dit probleem behulpzaam kunnen zijn, bestaan er geen objectieve of algemeen geaccepteerde criteria om het aantal componenten te bepalen (Jackson, 1991: 41-51). Om deze reden en omdat de te bepalen componenten zullen worden gebruikt als verklarende variabelen van de participatiegraad, is de nadruk gelegd op het op een zinvolle wijze kunnen interpreteren van deze componenten. Om het aantal verklarende variabelen van de participatiegraad en daarmee het aantal componenten binnen de perken te houden is het laatste aantal voorts als voldoende aangemerkt als het verklaarde deel van de variantie binnen de volledige verzameling van geslacht- en leeftijdsvariabelen tenminste groter is dan het onverklaarde deel. Op grond van deze criteria bleek het aantal van 2 principale componenten optimaal. Beide staan nader omschreven in tabel 1 en verklaren tezamen 65.5% van de totale variantie.

De eerste van deze twee componenten kan worden geïnterpreteerd als de ontgroening en vergrijzing van de bevolking, het feit dat het aantal ouderen alsmaar toeneemt, terwijl het aantal jongeren dat daar tegenover staat alsmaar afneemt. Deze ontwikkeling komt duidelijk tot uitdrukking in de coëfficiënten. Het teken van de jongere mannelijke en vrouwelijke leeftijdsgroepen is negatief en van de oudere mannelijke en vrouwelijke leeftijdsgroepen positief. De tweede component kan worden geïnterpreteerd als dat deel van de bevolking dat volledig beschikbaar is voor de arbeidsmarkt. Alleen de coëfficiënten van de mannelijke en vrouwelijke leeftijdsgroepen in de middengroepen zijn positief. Omgekeerd zijn de coëfficiënten van de mannelijke en vrouwelijke leeftijdsgroepen tot 24 jaar en vanaf 55 jaar negatief vanwege het feit dat zij niet volledig voor de

Tabel 1 Principale componenten analyse
toegepast op de variabelen ter
beschrijving van de samenstel-
ling van de bevolking naar
leeftijd en geslacht

Geslacht/ leeftijd	Gemiddelde waarde	Eerste component	Tweede component
M 0-14	8.95%	-.0947	-.0275
V 0-14	8.55%	-.0991	-.0278
M 15-24	8.40%	-.1384	-.1590
V 15-24	8.46%	-.1814	-.1390
M 25-34	7.22%	.0012	.3016
V 25-34	7.24%	-.0198	.3450
M 35-44	6.75%	.1039	.4044
V 35-44	6.70%	.1043	.3727
M 45-54	6.13%	.1381	.0531
V 45-54	6.17%	.1478	.0004
M 55-64	5.37%	.1415	-.2043
V 55-64	5.99%	.1522	-.1416
M 65+	5.59%	.0830	-.1156
V 65+	8.46%	.0794	-.0271

arbeidsmarkt beschikbaar zijn. Jongeren zijn niet volledig beschikbaar, omdat ze in steeds grotere getalen een opleiding volgen, en ouderen zijn niet volledig beschikbaar, omdat ze in steeds grotere getalen vroegtijdig uit het arbeidsproces worden gestoten. Per waarneming zijn de waarden van deze twee componenten berekend, waarna ze vervolgens zijn gebruikt als verklarende variabelen.

In tweede aanleg is aan de hand van de toetsgrootheid in vergelijking (3) onderzocht of de verschillende geslacht- en leeftijdsgroepen al dan niet ander gedrag vertonen. Hierbij zijn de betreffende participatiegraden geschreven als een lineaire functie van de bovenomschreven verklarende variabelen en vervolgens zowel geschat met behulp van OLS als met 2SLS. De reden om ook 2SLS toe te passen houdt verband met de "simultaneity bias" die kan ontstaan als het werkloosheidscijfer ten onrechte wordt beschouwd als een exogene grootheid. Volgens Fleisher en Rhodes (1976) worden de participatiegraad en het werkloosheidscijfer in regionale arbeidsmarkten simultaan bepaald en leidt de toepassing van OLS op een regressievergelijking waarin de participatiegraad wordt geschat als functie van het werkloosheidscijfer om deze reden tot onzuivere resultaten. De eenvoudigste wijze om hiervoor te corrigeren is om te schatten met 2SLS, een methode die in de literatuur veelvuldig is toegepast (Fleisher en Rhodes, 1976; Van der

Veen, 1982; Siegers, 1983; Gallaway et al., 1991)¹.

Uit de langs deze weg verkregen resultaten bleek dat de H_0 -hypothese of twee bevolkingsgroepen hetzelfde arbeidsaanbodgedrag vertonen zelfs bij een hoge betrouwbaarheidsdrempel van 99% telkenmale moest worden verworpen, ongeacht welke twee bevolkingsgroepen werden vergeleken en ongeacht de toegepaste schattingsmethode. Aan deze resultaten zijn dan ook twee belangrijke conclusies verbonden. In de eerste plaats kan worden geconcludeerd dat als men een beleid wil ontwikkelen ten einde de participatiegraad van specifieke bevolkingsgroepen te verhogen, zoals vrouwen, jongeren of ouderen, het zinvol is een analyse te maken per deelgroep, omdat elke deelgroep zich anders blijkt te gedragen. In de tweede plaats kan worden geconcludeerd dat Op de Beke (1988) in zijn onderzoek naar arbeidsparticipatie wel zeer stringente voorwaarden oplegt aan de parameters door te veronderstellen dat alle leeftijdsgroepen op dezelfde wijze op de verklarende variabelen reageren. Alleen de constante kan verschillende waarden aannemen doordat voor elke leeftijdsgroep een aparte dummy is ingevoerd. Voorts dient te worden opgemerkt dat onder de verklarende variabelen ook het kindertal is opgenomen - met name relevant voor vrouwen in de middelste leeftijdsklassen -, terwijl deze in dit onderzoek ontbreekt. Het uitsluiten van de mogelijkheid dat de reactieparameters van de verklarende variabelen over de verschillende leeftijdsgroepen kunnen variëren, lijkt te zijn ingegeven door het feit dat langs deze weg een groter aantal waarnemingen wordt verkregen. In plaats van 7 waarnemingen wordt door 10 verschillende leeftijdsgroepen te onderscheiden een totaal van 70 waarnemingen verkregen. Op de Beke (1988: 126) verdedigt deze handelswijze met de opmerking dat "*as a result of the larger number of data points more degrees of freedom are available, which enables the parameters to be estimated more efficiently*". In werkelijkheid is deze efficiency-winst echter schijn doordat heterogene groepen ten onrechte als zijnde homogeen worden behandeld. Met andere woorden: de efficiency van de parameters neemt wel toe, maar de representativiteit neemt af. Daardoor blijven ze ongeschikt om er beleidsmaatregelen aan te ontleen.

In derde aanleg is onderzocht of een afzonderlijke analyse van de participatiegraad van verschillende geslacht- en leeftijdsgroepen ook winst oplevert als men uitsluitend is geïnteresseerd in het geheel van deze groepen. In de vorige paragraaf is uiteengezet dat het antwoord op deze vraag van geval tot geval zal verschillen. Om te toetsen hoe het in dit specifieke geval uitpakt is in beginsel uitgegaan van een kwadratische vergelijking, zij het dat alleen die interactietermen zijn geselecteerd die significant van nul verschillend zijn op 5%-niveau

¹Een andere methode is om tevens een vergelijking te ontwikkelen waarin het werkloosheidscijfer wordt verklaard en het dan ontstane stelsel te schatten met 3SLS of FIML (zie Van der Veen en Evers, 1983; en Nord, 1989).

$$Y = a_0 + a_1'X + 1/2 X'AX,$$

(8)

met A een symmetrische matrix en X de verzameling van verklarende variabelen. De filosofie achter deze opzet is de volgende. Eén van de voordelen van het afzonderlijk analyseren van het arbeidsaanbodgedrag van verschillende geslacht- en leeftijdsgroepen is dat door de inbreng van specifieke groepseigenschappen een grotere verklaringskracht kan worden verkregen. Nu voert het in het kader van dit artikel te ver om een diepgaand onderzoek naar deze specifieke groepseigenschappen in te stellen. Bovendien biedt de beschikbare dataset daartoe onvoldoende mogelijkheden. Als alternatief is daarom voor een kwadratische functievorm gekozen. Een kwadratische functievorm behoort tot de klasse van flexibele functievormen; flexibel in de zin dat het beschouwd kan worden als een 2^e orde benadering voor de werkelijke, doch onbekende arbeidsaanbodfunctie die wij zouden willen bepalen. Daar komt bij dat de verklaringskracht van een kwadratische vergelijking in termen van significantie hoger is dan een lineaire vergelijking, omdat in een 2^e orde benadering rekening wordt gehouden met interactie tussen de verklarende variabelen. Schattingen van de kwadratische functievorm hebben echter tevens duidelijk gemaakt dat maar een beperkt deel van de interactietermen significant is, alsook dat dit deel per onderzochte bevolkingsgroep verschillend is. Door nu per bevolkingsgroep alleen die interactietermen te selecteren die significant van nul verschillend zijn, wordt een verantwoorde vergelijking tussen het directe model in de vorm van vgl(2) en het indirecte model samengesteld uit vgl(1a) en vgl(1b) mogelijk. In de eerste plaats wordt door voor elke vergelijking uit te gaan van een flexibele functievorm een zo hoog mogelijke verklaringskracht gewaarborgd en in de tweede plaats wordt door alleen significante interactietermen toe te laten elke vergelijking gekenmerkt door een andere verzameling van verklarende variabelen. Zo wordt op bescheiden wijze toch het specifieke karakter van elk van de bevolkingsgroepen tot uitdrukking gebracht. De op deze wijze verkregen residuele kwadratensommen staan vermeld in tabel 2.

Uit deze tabel blijkt dat de samengestelde residuele kwadratensom van de mannelijke en vrouwelijke beroepsbevolking ten opzichte van de residuele kwadratensom van de gehele beroepsbevolking kleiner is, alsook de samengestelde residuele kwadratensom van de 5 mannelijke en 5 vrouwelijke leeftijdsgroepen ten opzichte van de residuele kwadratensom van de totale mannelijke respectievelijk vrouwelijke beroepsbevolking. Bovendien gelden deze uitkomsten wederom ongeacht de toegepaste schattingsmethode. Met andere woorden: zelfs als dit onderzoek zich uitsluitend ten doel zou stellen de participatiegraad van de totale beroepsbevolking te verklaren, dan nog is een uitsplitsing naar verschillende geslacht- en leeftijdsgroepen aan te bevelen, omdat langs deze weg de residuele kwadratensom wordt gereduceerd.

Tabel 2 Vergelijking van residuele kwadratensommen

Aard vergelijking	Schattingsmethode	
	OLS	2SLS
totale beroepsbevolking/ sommatie over mannen en vrouwen	4711/3636	4940/4016
mannen/sommatie over leeftijdsgroepen	2144/1985	2397/2131
vrouwen/sommatie over leeftijdsgroepen	10357/10046	10904/10575

4. Conclusies

De doelstelling van dit onderzoek was om meer klaarheid te brengen in de modelmatige relatie tussen geslacht, leeftijd en arbeidsparticipatie. Geslacht en leeftijd kunnen bij de verklaring van arbeidsparticipatie zowel aan de rechter- als aan de linkerkant van het gelijkheidsteken van de regressievergelijking een rol spelen. Dat geslacht en leeftijd aan de rechterkant van het gelijkheidsteken een rol spelen leidt eigenlijk geen twijfel. Verscheidene onderzoekers hebben aangetoond dat de samenstelling van de bevolking naar geslacht en leeftijd invloed uitoefent op de arbeidsparticipatie. Het probleem is echter dat vaak meerdere variabelen benodigd zijn om de samenstelling van de bevolking in een regressievergelijking tot uitdrukking te brengen, variabelen die onderling sterk blijken te correleren. Dit leidt tot multicollineariteit met alle problemen vandien. In dit onderzoek hebben wij laten zien hoe dit probleem kan worden opgelost door de toepassing van principale componenten-analyse.

Of men als onderzoeker geslacht en leeftijd ook aan de linkerkant van het gelijkheidsteken een rol moet laten spelen - dit door de participatiegraad van verschillende bevolkingsgroepen te verklaren - hangt in de eerste plaats af van datgene waarin de onderzoeker is geïnteresseerd. Als een onderzoeker is geïnteresseerd in het arbeidsaanbodgedrag van verschillende bevolkingsgroepen, dan is een uitsplitsing van de beroepsbevolking naar leeftijd en geslacht bij voorbaat aan te bevelen. Alleen als zou blijken dat het arbeidsaanbod van twee bevolkingsgroepen niet significant van elkaar verschilt, ontbreekt de aanleiding deze nog afzonderlijk te behandelen. In het empirische gedeelte van dit onderzoek is echter gebleken dat deze situatie zich nimmer voordoet.

Anders wordt het als een onderzoeker uitsluitend is geïnteresseerd in de

verklaring van participatiegraad van de beroepsbevolking in zijn geheel. In het licht van deze doelstelling heeft een indirecte analyse uitgesplitst naar verschillende bevolkingsgroepen alleen zin als de gewogen som van de regressieschattingen per bevolkingsgroep een nauwkeuriger verklaring geeft dan een directe regressieschatting voor het geheel van deze bevolkingsgroepen. In dit onderzoek is aangetoond dat op theoretische gronden niet is te zeggen welke van de twee benaderingen het beste resultaat geeft. Op empirische gronden is echter gebleken dat een uitsplitsing naar bevolkingsgroepen is te prefereren. Men moet er echter voor waken dit niet te verheffen tot een algemeen resultaat voor onderzoek naar arbeidsparticipatie, omdat zich in een ander onderzoek met andere data of een andere groepsindeling wel eens het omgekeerde zou kunnen voordoen.

Literatuur

- Baumann, J., Fischer, M.M. en U. Schubert (1988), 'A choice-theoretical labour-market model: empirical tests at the meso-level'. *Environment and Planning A* 20: 1085-1102.
- Bout, E.R., Laan, L. van der en G.A. van der Knaap (1989), *Het deelnemingsgedrag van vrouwen op de Nederlandse arbeidsmarkt; een regionale analyse*. Rotterdam, Economisch Geografisch Instituut.
- Bowen, W.G. en T.A. Finegan (1969), *The economics of labor force participation*. Princeton, University Press.
- Dickmann, A., Laan, L. van der en G.A. van der Knaap (1989), *Regionale ontwikkelingen van de arbeidsmarkt participatie in de Europese Gemeenschap (1979-1985)*. Rotterdam, Economisch Geografisch Instituut.
- Elhorst, J.P. (1993), *Regional economic research on labour force participation*. Groningen, RuG.
- Fleisher, B.M. en G. Rhodes (1976), 'Unemployment and the labor force participation of married man and women: a simultaneous model'. *The Review of Economics and Statistics* 58: 398-406.
- Gallaway, L., Vedder, R. en R. Lawson (1991), 'Why people work: an examination of interstate variations in labor force participation'. *Journal of Labor Research* 12: 47-59.
- Jackson, J.E. (1991), *A user's guide to principal components*. New York, John Wiley & Sons.
- Molho, I.I. (1983), 'A regional analysis of the distribution of married women's labour force participation rates in the UK'. *Regional Studies* 17: 125-134.
- Molho, I. en P. Elias (1984), 'A study of regional trends in the labour force participation of married women in the UK, 1968-1977'. *Applied Economics* 16: 163-173.
- Nord, S. (1989), 'The relationships among labor-force participation, service-sector employment, and underemployment'. *Journal of Regional Science* 29: 407-421.

- Op de Beke, J.M.J. (1988), 'Endogenous extrapolation of labor force participation rates of married women'. *De Economist* 136: 118-135.
- Siegers, J.J. en R. Zandanel (1981), 'A simultaneous analysis of the labour force participation of married women and the presence of young children in the family'. *De Economist* 129: 382-393.
- Siegers, J.J. (1983), 'An economic-demographic ten-equation model'. *De Economist* 131: 400-443.
- Steinle, W.J. (1983), Regional labour markets: trends, structure and relevance from a European perspective'. *Papers of the Regional Science Association* 52: 3-21.
- Veen, A. van der (1982), 'Het aanbod van arbeid en de spanning op de arbeidsmarkt'. *Maandschrift Economie* 46: 3-17.
- Veen, A. van der en G.H.M. Evers (1983), 'A simultaneous model for regional labor supply, incorporating labor force participation, commuting and migration'. *Socio Economic Planning Sciences* 17: 239-250.
- Veen, A. van der en G.H.M. Evers (1984), 'A labour supply function for females in the Netherlands'. *De Economist* 132: 367-376.
- WRR (1990), *Een werkend perspectief; arbeidsparticipatie in de jaren '90*. Den Haag, SDU/Uitgeverij.

Ontvangen: 24-1-1994

Geaccepteerd: 18-4-1994

