

Een Empirisch Onderzoek naar de Determinanten van de Werkloosheidsduur

J.M.G. Frijns^{★)}

J.G.F. Sibbing

1. Inleiding
2. Literatuurverkenning en Probleemstelling
3. Te Gebruiken Statistische Methoden
 - 3.1. Logit-Analyse
 - 3.2. Regressie met een Afgeknotte Afhankelijke Variabele
4. Resultaten en Conclusies
 - 4.1. Data, Selectiviteit en Meetfouten
 - 4.2. Resultaten van de Logit-Analyse
 - 4.3. Resultaten van de Afgeknotte Afhankelijke Variabele Methode
 - 4.4. Enige Voorzichtige Conclusies

★) C.P.B., Den Haag en K.H.T., Tilburg. Ten tijde van dit onderzoek waren beide auteurs aan de Katholieke Hogeschool Tilburg verbonden.

1. Inleiding.

In dit paper wordt verslag gedaan van een onderzoek naar de herintrede-kansen van werklozen. Via statistisch onderzoek wordt nagegaan welke variabelen deze herintrede-kansen en daarmee de totale duur van de werkloosheid beïnvloeden. Als verklarende variabelen voor deze vertrekken en daarmee voor de totale individuele werkloosheidsduur komen in aanmerking persoonsgebonden kenmerken, zoals opleiding, leeftijd, voorts vraag- en aanbodverhoudingen op de betrokken arbeidsmarkt en tenslotte variabelen die van invloed zijn op het zoekgedrag van een werkloze zoals verwachte loon versus hoogte van de werkloosheidsuitkering.

Dit onderzoek kan worden toegespitst op een tweetal vragen. In de eerste plaats kunnen we onderzoeken of er homogene sociaal-economische subgroepen te onderscheiden zijn met duidelijk verschillende herintrede-kansen. Indien dit het geval is zou dit een reden zijn om een gericht, selectief werkgelegenheidsbeleid te ontwikkelen. De andere vraag luidt in hoeverre algemeen economische variabelen als vraag- en aanbodverhoudingen en verschillen in beloning van belang zijn voor de verklaring van de kansen op herintrede. Indien vooral deze variabelen van belang zijn lijkt een globaal werkgelegenheidsbeleid meer gewenst.

De te gebruiken statistische methodes, logit-analyse en regressie met een afgeknotte afhankelijke variabele, zijn in de algemene econometrie minder bekend. Volledigheidshalve zullen we in Paragraaf 3 deze methodes in het kort uiteen zetten. De gebruikte data zijn afkomstig van een I.V.A.-steekproef onder werklozen in mei en november 1974. Daar onze onderzoeksopzet verschilde van die waarvoor de steekproef oorspronkelijk getrokken is was aanpassing en hercodering van het steekproefmateriaal noodzakelijk. De optredende selectiviteit en de meetfouten ten gevolge van deze noodzakelijke aanpassingen nopen tot grote behoedzaamheid bij het interpreteren van de gevonden resultaten. Deze studie is slechts een eerste verkenning.

De indeling van dit paper is als volgt. In Paragraaf 2 wordt een kort overzicht gegeven van recente literatuur en wordt de probleemstelling geformuleerd, die als basis voor het empirisch onderzoek zal dienen. In Paragraaf 3 worden de te gebruiken statistische methoden kort besproken. In Paragraaf 4 worden de resultaten van de schattingen gepresenteerd en worden enige enige voorzichtige conclusies getrokken.

2. Literatuurverkenning en Probleemstelling.

De literatuur over de determinanten van de werkloosheidsduur is zeer omvangrijk. Aan het ene uiteinde bevinden zich de "new micro-economists" zoals Phelps [11] en Mortensen [7] die de nadruk op het zoekgedrag van de werkloze leggen en dan vooral aandacht besteden aan de rol van beloningsvariabelen, zoals de verhouding tussen verwacht loon en huidig uitkeringsniveau. Aan de andere extreme zijde bevinden zich meer sociologisch georiënteerde onderzoekers die de nadruk leggen op sociale en psychologische factoren en die wijzen op de oververtegenwoordiging van maatschappelijke randgroepen binnen de totale groep werklozen. Tenslotte is er een grote groep economen die de nadruk legt op een tekort schietende (effectieve) vraag naar arbeid. Het is hier niet de plaats om uitgebreid op de (theoretische) merites en beperkingen van de verschillende theorieën in te gaan. Interessanter voor ons zijn een aantal empirische studies van recente datum van o.a. McGregor [6], Nickell [8], [9] en Lancaster [4]. In deze studies wordt getracht m.b.v. gegevens die betrekking hebben op individuele werklozen en gebruikmakend van min of meer geavanceerde statistische methoden de relevantie van de verschillende theorieën voor de verklaring van de herintrede-kans en de totale werkloosheidsduur na te gaan. Een korte bespreking van deze studies lijkt op zijn plaats.

McGregor onderzoekt een steekproef van 786 mannelijke werklozen, getrokken uit het kaartenbestand van een Engels arbeidsbureau. Zijn belangrijkste conclusie is, dat de kans op herintrede daalt naarmate de werkloosheidsperiode langer duurt. Voorts blijkt de invloed van het opleidingsniveau, de woonplaats en de bedrijfstakvariabelen aantoonbaar. De invloed van beloningsvariabelen kon niet onderzocht worden, omdat gegevens hierover ontbraken.

Nickell heeft de beschikking over een steekproef van 426 mannelijke werklozen, afkomstig van de Engelse General Household Survey van 1972. Zijn conclusie is, dat de replacement-ratio, gedefinieerd als de verhouding tussen het niveau van de werkloosheidsuitkering en het gemiddeld loon in de beroepsgroep van de betrokken werkloze, de werkloosheidsduur in belangrijke mate beïnvloedt. Voorts blijken ook de lokale vraag naar arbeid en persoonlijke karakteristieken van belang, terwijl de kans op herintrede afneemt naarmate de persoon langer werkloos is. Tenslotte past Nickell een verfijning toe, waarbij de invloed van de replacement-ratio mag verschillen voor kortstondig en langdurig werklozen. Het blijkt dat de invloed van deze ratio sterk afneemt bij

personen die langdurig werkloos zijn.

Lancaster gebruikt een in 1973 getrokken steekproef van 479 ongeschoolde werklozen. Hij concludeert dat de leeftijd, het lokale werkloosheidspercentage en de replacement-ratio een negatieve invloed hebben op de herintrede-kans. De invloed van de reeds verstreken werkloosheidsduur blijkt negatief.

De resultaten van deze onderzoeken zijn niet opzienbarend. Zowel betalingsverhoudingen, vraag- en aanbodverhoudingen als persoonlijke kenmerken blijken de werkloosheidsduur te beïnvloeden. Vooraf kunnen we dus geen van deze variabelen uitsluiten. Bij de verklaring van de kans op herintrede is tenslotte nog van belang deze kans te beschouwen als functie van de reeds verstreken werkloosheidsduur. Uit de eerder besproken studies blijkt een negatief verband, d.w.z. dat van twee individuen met gelijke scores t.a.v. de eerder genoemde kenmerken de herintrede-kans het grootst is voor het individu met de kortste reeds verstreken werkloosheidsduur. Dit kan zeer wel te wijten zijn aan een statistisch artefact. Stel dat de herintrede-kans ook afhangt van andere niet expliciet opgenomen variabelen, bijv. motivatie, gezondheid, uiterlijk, dan kan een lange reeds verstreken werkloosheidsduur duiden op een voor herintrede ongunstige score op deze kenmerken.

In deze studie wordt een tweetal onderzoeksmethoden gebruikt, nl. logit-analyse voor het verklaren van de herintrede-kans en een afgeknotte afhankelijke variabele methode voor het verklaren van de totale werkloosheidsduur. In Par. 3 zullen deze methoden in het kort uiteengezet worden.

3. Te Gebruiken Statistische Methoden.

3.1. Logit-Analyse.

Een lastig probleem in de econometrie is het modelleren van problemen waarbij de afhankelijke variabele dichotoom is, bv. wel of niet verhuizen, van werk veranderen, aanschaffen van een auto etc. Een mogelijkheid is het invoeren van een dummy-variabele die de waarde een of nul aanneemt. Zo zou als model voor het verklaren van het aanschaffen van een auto door individu j met inkomen x_j kunnen gelden

$$y_j = \beta_0 + \beta x_j + \epsilon_j$$

waarin $y_j = 1$ bij aankoop en $y_j = 0$ elders. Onaantrekkelijk in deze specificatie zijn de weinig sierlijke eigenschappen van de storingsterm ϵ_j en het

feit dat de voorspelde waarde $\hat{y}_j$ alle waarden op interval $(-\infty, +\infty)$ kan aannemen terwijl y_j beperkt is tot 0 of 1.

Een aantrekkelijker aanpak is het werken met conditionele kansen, bv. de conditionele kans dat individu j een auto koopt gegeven zijn inkomen x_j . Deze conditionele kansen vormen het systematische deel van het keuzegegedrag. Het feitelijk gedrag komt tot stand doordat individu j met kans p_j mogelijkheid 1 en met kans $(1-p_j)$ mogelijkheid 2 kiest en is zodoende van het toeval afhankelijk.

Een in de praktijk veel gebruikte modellering is de zgn. logit-specificatie die stelt dat de logaritme van de verhouding der conditionele kansen p_j en $(1-p_j)$ afhangt van een rij van k verklarende variabelen x_{ij} :

$$e^{\log \frac{(1-p_j)}{p_j}} = \sum_i x_{ij} \beta_i$$

Voor de conditionele kans p_j van individu j volgt dan

$$p_j = \frac{1}{1 + e^{\sum_i x_{ij} \beta_i}}, \quad (1-p_j) = \frac{e^{\sum_i x_{ij} \beta_i}}{1 + e^{\sum_i x_{ij} \beta_i}}$$

De ML-schatter voor de parametervector β is gebaseerd op de likelihood-functie

$$L(\beta) = \prod_{j=1}^n p_j^{y_j} (1-p_j)^{(1-y_j)}$$

waarbij n het aantal waarnemingen (individueen in de steekproef) is en y_j een dummy-variabele die de waarde 1 aanneemt bij keuze van individu j voor mogelijkheid 1 en nul elders. Substitutie van de specificatie voor p_j en het overgaan op de log-likelihood-functie geeft

$$L = e^{\log L(\beta)} = \prod_{j=1}^n [(1-y_j) (\sum_i x_{ij} \beta_i) - e^{\log (1 + e^{\sum_i x_{ij} \beta_i})}]$$

De ML-schatter voor β kunnen we m.b.v. het Newton-Raphson-algoritme berekenen door het oplossen van de eerste-orde voorwaarden voor een maximum

$$\frac{\partial L}{\partial \beta_i} = 0, \quad i = 1, \dots, k.$$

Over de eigenschappen van deze ML-schatter zijn slechts asymptotische resultaten bekend. In McFadden [5] wordt gedetailleerd op deze eigenschappen ingegaan. Wij volstaan met het vermelden dat onder algemene regulariteitsvoorwaarden de schatter $\hat{\beta} = (\hat{\beta}_1, \dots, \hat{\beta}_k)$ asymptotisch normaal verdeeld is met verwachtingswaarde β en als covariantie-matrix de inverse informatie-

matrix. Toetsen t.a.v. β kunnen rechtstreeks gebaseerd worden op deze asymptotische normale verdeling; een alternatief is een expliciete likelihood-ratio toets.

3.2. Regressie met een Afgeknotte Afhankelijke Variabele.

Het afgeknotte afhankelijke variabele probleem treedt op als er van de afhankelijke variabele niet genoeg waarnemingen beschikbaar zijn. In ons geval geldt dat de werkloosheidsduur slechts geobserveerd is tot het moment van de laatste enquête, zeg tijdstip t . Stel dat a_j de op tijdstip t waargenomen werkloosheidsduur is en y_j de werkelijke totale werkloosheidsduur. Dan geldt:

$$y_j = a_j \quad \text{indien persoon } j \text{ niet meer werkloos is op tijdstip } t$$

$$y_j > a_j \quad \text{indien persoon } j \text{ nog werkloos is op tijdstip } t$$

Laat verder voor de eerste n_1 steekproeftrekkingen gelden dat de totale duur observeerbaar is en laat voor de resterende $(n - n_1)$ trekkingen gelden dat slechts de ondergrens a_j observeerbaar is. Verder veronderstellen we dat de werkloosheidsduur y_j gamma verdeeld is met parameters $r = 2$ en een individu-afhankelijke parameter λ_j .¹⁾ Ergo

$$f_j(y_j) = \lambda_j^2 y_j e^{-\lambda_j y_j}, \quad y_j \geq 0, \quad \lambda_j \geq 0$$

met verdelingsfunctie $F_j(y_j)$ en

$$E[y_j] = 2/\lambda_j, \quad \text{Var}[y_j] = 2/\lambda_j^2$$

-
- 1) De keuze voor een gamma verdeling berust op gronden van eenvoud. Beschouwen we de verdeling van de werkloosheidsduur als wachttijdverdeling met tijds-afhankelijke uittrede kansen dan komt de negatief exponentiële verdeling die wordt gekenmerkt door constante uittrede kansen niet in aanmerking. De gamma verdeling met $r = 2$ en $\lambda > 0$ is een rechts-scheve verdeling met top bij $1/\lambda$ en verwachtingswaarde $2/\lambda$.

De keuze voor $r = 2$ berust op de overweging dat voor deze waarde de gamma verdeling een maximale scheefheid bezit (m.u.v. de niet in aanmerking komende negatief exponentiële verdeling waarvoor $r = 1$). Dit sluit aan bij de geobserveerde forse scheefheid van de verdeling van de totale werkloosheidsduur.

Het zou voor de hand liggen te experimenteren met waarden voor r groter dan 2. Dit viel echter buiten het bestek van dit onderzoek en blijft dus liggen voor een vervolgonderzoek.

De parameter λ_j is individu-afhankelijk. Veronderstellen we dat de gemiddelde duur afhangt van een k-tal afhankelijke variabelen x_{ij} via een eenvoudige lineaire functie dan kunnen we schrijven

$$\frac{2}{\lambda_j} = E[y_j] = \sum_i x_{ij} \beta_i$$

ofwel

$$\lambda_j = \frac{1}{\frac{1}{2} \sum_i x_{ij} \beta_i}$$

De likelihood-functie van de steekproef heeft een hybride vorm.

Veronderstellen we onderling onafhankelijke steekproeftrekkingen dan is deze functie het product van n_1 dichtheidsfuncties $f_j(y_j)$ en $(n-n_1)$ discrete kansen $[1-F_j(a_j)]$, nl.

$$L = \prod_{j=1}^{n_1} f_j(y_j) \prod_{j=n_1+1}^n [1-F_j(a_j)]$$

Dit model is oorspronkelijk voorgesteld door Tobin [13].

Het gebruik van de ML-schatter lijkt voor de hand te liggen.²⁾ Een gedegen analyse van de (asymptotische) statistische eigenschappen van deze ML-schatter als y_j normaal verdeeld is, is te vinden in Amemiya [1]. Onder algemene voorwaarden³⁾ blijkt de ML-schatter voor β asymptotisch raak en asymptotisch normaal verdeeld met als covariantiematrix de inverse informatiematrix.

Hoewel deze resultaten afgeleid zijn voor een normale verdeling lijkt ons vertrouwen op de robuustheid van deze resultaten voor het geval van de gamma verdeling gerechtvaardigd.

- 2) Het gebruik van een OLS-schatter in het lineaire regressiemodel

$$y_j = \sum_i x_{ij} \beta_i + u_j$$

leidt in het geval dat y_j een afgeknotte variabele is tot inconsistente schatters voor β_i , zowel wanneer we alleen die waarnemingen gebruiken waarvoor we de werkelijke waarde y_j kennen als wanneer we de nog onbekende waarden van y_j vervangen door de waargenomen onder grens a_j . In beide gevallen zal namelijk de verwachtingswaarde van y_j ongelijk aan nul zijn en afhangen van $\sum_i x_{ij} \beta_i$.

R. Olsen [10] geeft voor het geval dat y_j normaal verdeeld is correctiefactoren voor de OLS-schattingen. Blijkens zijn simulatie-resultaten leidt dit tot zeer acceptabele resultaten.

- 3) Voor aangename asymptotische eigenschappen is o.a. vereist dat een uniek maximum bestaat en dat de log-likelihoodfunctie strict concaaf is in de omgeving van dit maximum. Voor snelle numerieke convergentie naar dit maximum zou globale concaviteit van de log-likelihood-functie zeer plezierig zijn. Hierop mogen we echter niet rekenen. Het is dan de kunst een startvector voor de te gebruiken numerieke ML-procedure te vinden die in de buurt van maximum ligt.

4. Resultaten en Conclusies.

4.1. Data, Selectiviteit en Meetfouten.

De gegevens van dit onderzoek zijn afkomstig van een IVA-onderzoek in het kader van het project Bestrijding Excessieve Werkloosheid, zie hiertoe Koopmans [3]. Daartoe werd een steekproef van 580 werklozen geënquêteerd in mei 1974 met een vervolgenquête in november 1974. De door ons gestelde eisen t.a.v. volledigheid van informatie per persoon leidden ertoe dat voor onze steekproef slechts 158 personen resteerden. Deze selectie op volledigheid van informatie kan echter gemakkelijk leiden tot selectiviteit, nl. als ontbreken van gegevens gecorreleerd is met de waarde van de endogene variabele. Als b.v. van kortstondig werklozen vaker gegevens ontbreken leidt dit tot selectiviteit. Over de gevolgen van deze selectiviteit kunnen wij weinig zinnigs zeggen, daarvoor weten we te weinig over de aard en omvang van de optredende selectiviteit.⁴⁾

Een tweede belangrijke bron van systematische schattingsfouten zijn meetfouten aan de verklarende variabelen. In ons geval dienen we vooral beducht te zijn voor meetfouten bij de berekening van de inverse replacement ratio. Problemen ontstonden vooral doordat geen voldoende fijne indeling in beroepsgroepen mogelijk bleek, zodat we ons moesten behelpen met een ruwe schatting van het gemiddelde te verdienen loon in de betrokken bedrijfstak. Veronderstellen we dat de gevolgen van deze meetfouten in de door ons gebruikte statistische methoden overeenkomen met die in het standaard lineaire regressiemodel dan geldt het volgende. De directe invloed van meetfouten aan een verklarende variabele is een onderschatting van de absolute waarde van de bijbehorende regressie-coëfficiënt; zie bv. Theil [12, Par. 12.2]. De indirecte invloed van meetfouten aan een verklarende variabele op de schattingen van de andere coëfficiënten hangt af van de samenhang tussen de betrokken verklarende variabelen in de steekproef. Bij geringe samenhang, wat in ons geval een gerechtvaardigde assumptie lijkt, is deze invloed ook gering.

Een laatste oorzaak van systematische schattingsfouten is het "omitted variables" probleem. Dit kan zeer nare gevolgen hebben, in het bijzonder als

4) In een eenvoudig lineair regressie-model zouden wij deze selectiviteit kunnen definiëren als:

$$E[y_j | x_j, \text{Sample Selection Rule}] = x_j \beta + E[u_j | \text{SSR}]$$

Vertekening t.g.v. selectiviteit treedt op als $E[u_j | \text{SSR}]$ ongelijk nul is en afhangt van j . OLS-schatters zijn in dit geval niet meer zuiver, en inconsistent. Zie bv. J. Heckman [2].

een van de wel opgenomen regressoren (ongewild) tevens als proxy voor een of meerdere omitted variables functioneert. In het voorgaande zijn we reeds dieper op het probleem van de omitted variables ingegaan, zie Par. 2.

4.2. Resultaten van de Logit-Analyse.

Het in Par. 3.1. beschreven logitmodel behoeft voor de praktische toepassing ter verklaring van de herintredkans nog een definitie van een te gebruiken maat van aanpassing, een specificatie van de lengte van het interval $(t, t+s)$ waarbinnen werkloze al dan niet werk vindt en een specificatie van de op te nemen verklarende variabelen. Als de herintredkans geheel onafhankelijk zou zijn van de specifieke kenmerken van individu j en dus voor iedereen gelijk, dan zou $\bar{p} = \Sigma y_j / n$ met

$$y_j = \begin{cases} 1 & \text{bij herintrede van individu } j \\ 0 & \text{elders} \end{cases}$$

en n het aantal individuen een zuivere schatter van deze kans zijn. Bij de beoordeling van de mate van verklaring die het logitmodel verschaft hebben we dan ook willen nagaan in hoeverre de schatter

$$\hat{p}_j = 1 / [1 + \exp(\sum_i x_{ij} \hat{\beta}_i)]$$

die uit het logitmodel volgt, betere resultaten oplevert dan $\bar{p}$. Daartoe hebben we de volgende grootheid gedefinieerd⁵⁾:

$$e = 1 - \left[\frac{\sum_{j=1}^n (y_j - \hat{p}_j)^2}{\sum_{j=1}^n (y_j - \bar{p})^2} \right]$$

De aanpassing (verklaring) m.b.v. het logitmodel is beter naarmate e dichter bij 1 ligt.

5) Een andere mogelijkheid is gebaseerd op de loglikelihood functies; de toets

$$H_0: p_j = \bar{p} \quad \text{t.o.v.} \quad H_1: p_j = 1 / [1 + \exp(\sum_i x_{ij} \beta_i)] \quad j = 1, \dots, n$$

levert als toetsingsgrootheid $2(L_1 - L_0)$, die een χ^2 verdeling volgt met k df (k = aantal regressoren excl. constante term) en waarin

$$L_1 - L_0 = \sum_{j=1}^n \left[y_j \ln \frac{\hat{p}_j}{\bar{p}} + (1 - y_j) \ln \frac{1 - \hat{p}_j}{1 - \bar{p}} \right]$$

De intervallengte s staat voor de periode vanaf 1 juni; we onderzoeken of individu j in deze periode (her)intreedt. Achtereenvolgens is geëxperimenteerd met 1, 2, 3 en 4 maanden. Het aantal intreders was respectievelijk 18, 43, 57 en 68 van de 158 in de steekproef opgenomen personen. Het aantal intreders binnen een periode van 1 maand is erg gering; bij dit kleine aantal spelen toevalsfactoren een te grote rol. Het lijkt derhalve raadzaam deze periode niet in de analyse te betrekken. Als aanvullend criterium bij de keuze tussen 2, 3 en 4 maanden is de waarde van de aanpassingsmaatstaf e gebruikt. Deze nam toe van 0.04 bij 2 maanden, via 0.25 bij 3 maanden tot 0.44 bij 4 maanden. Omdat de geschatte waarden van de coëfficiënten β_1 nauwelijks verschilden voor de analyses bij 3 resp. 4 maanden zullen we gelet op de hogere waarde van e bij 4 maanden ons in het vervolg beperken tot deze laatste periode.

De in onze analyse betrokken verklarende variabelen hebben betrekking op persoonlijke kenmerken, zoals leeftijd, geslacht, opleiding etc., de beloningsverhouding, bv. in de vorm van de replacement ratio en vraag- en aanbodfactoren. Tenslotte is nog als aparte variabele opgenomen de tijd die een persoon op 31 mei al werkloos was. Een eerste schatting is uitgevoerd m.b.v. de als volgt gedefinieerde regressoren:

- het geslacht van individu j is opgenomen als een dummy-variabele. Deze variabele heeft de waarde 1 als de betrokkene een vrouw is en nul elders.
- de leeftijd van de betrokkene is gemeten in jaren.
- de burgerlijke staat is opgenomen als een dummy-variabele. Deze neemt de waarde 1 aan als betrokkene ongehuwd is en nul elders.
- de dummy-variabele voor de nationaliteit is 0 als de persoon Nederlander is en 1 elders.
- het beroepsniveau, zoals "gemeten" door arbeidsbemiddelaar is opgenomen op (0,1)-schaal, lopend van eenvoudige arbeid tot zeer ingewikkelde arbeid. Het beroepsniveau fungeert als een belangrijke factor in het vraag/aanbod proces op de arbeidsmarkt.⁶⁾
- de inverse replacement ratio is gedefinieerd als gemiddeld bruto weekloon in de bedrijfstak van betrokkene gedeeld door zijn netto inkomen per week tijdens de huidige werkloosheid.
- de werkloosheidsduur tot 31 mei is gemeten in maanden.

6) De achterliggende idee hierbij is dat in de zeventiger jaren een vraagoverschot bestond voor eenvoudige productie-arbeid en een aanbodoverschot voor hogere niet-productiegebonden functies.

De resultaten van deze schatting staan in Tabel 1; de e-waarde bedraagt 0.418.

regressor	geslacht	leeftijd	burg. staat	nationa- liteit	beroeps- niveau	inverse replace- ment ratio	werk- loos- heids- duur
geschatte coëfficiënt	-0.215	0.0166	0.768	1.799	0.561	-0.508	0.0414
geschatte standaardaf- wijking*	0.640	0.0113	0.468	1.173	0.725	0.204	0.026
gemiddelde waarde regressor over							
- de hele groep	0.09	37.57	0.37	0.04	0.51	2.39	8.13
- de herintreders	0.07	34.34	0.34	0.01	0.49	2.45	6.02

* Als standaardafwijking is gebruikt de geschatte standaardafwijking van de asymptotische verdeling.

Tabel 1, Logit Analyse, Het Basismodel.

Bij de interpretatie van deze resultaten dient allereerst opgemerkt te worden dat verhoging van $\sum_i x_{ij} \beta_i$ een verlaging van de herintrede kans betekent want

$p_j = 1/(1 + e^{\sum_i x_{ij} \beta_i})$. Er geldt dus $\partial p_j / \partial x_{ij} = -\beta_i p_j (1 - p_j)$; in de steekproef schommelt de geschatte $\hat{p}_j (1 - \hat{p}_j)$ tussen 0.085 en 0.25 met een gemiddelde van 0.221. Factoren die de kans op herintrede verlagen zijn dan hogere leeftijd, langere werkloosheidsduur, ongehuwd zijn, het niet bezitten van de Nederlandse nationaliteit en een hoger beroepsniveau (zij het dat deze laatste variabele een lage t-waarde⁷⁾ heeft). Een hogere waarde voor de inverse replacement ratio vergroot de kans op herintrede. Ook vrouwen lijken meer kans op herintrede te hebben, zij het dat hier de t-waarde wel erg dicht bij nul ligt. De gevonden resultaten voor de variabelen leeftijd, werkloosheidsduur, nationaliteit, beroepsniveau en inverse replacement ratio zijn in (kwalitatieve) overeenstemming met de theorie. Een indruk van het kwantitatieve belang van de afzonderlijke regressoren krijgen we door $\partial p / \partial x_i$, geëvalueerd in het steekproefgemiddelde $\bar{x}$, te vermenigvuldigen met $\Delta x_i = \bar{x}_i^h - \bar{x}_i$ waarbij $\bar{x}_i^h$ het gemiddelde van x_i over de herintreders is. We vinden dan bv. een stijging van de geschatte p met 0.0035 voor de

7) De t-waarde is gedefinieerd als geschatte coëfficiënt gedeeld door de geschatte standaardafwijking.

variabele beroepsniveau en met 0.0075 voor de inverse replacement ratio.

Op dit basismodel zijn een aantal amenderingen aangebracht. Allereerst wordt een opleidingsvariabele opgenomen die de waarde 1 heeft als de persoon ongeschoold is en de waarde 0 elders. Verder splitsen we de steekproef op in 3 leeftijdsgroepen via de introductie van 2 dummy variabelen: bij een leeftijd lager dan 25 neemt één dummy variabele de waarde 1 aan en bij een leeftijd hoger dan 45 neemt de andere dummy de waarde 1 aan. De idee achter deze opsplitsing is dat niet de leeftijd zelf een belemmerende invloed heeft op de herintrede, maar veeleer het overschrijden van bepaalde leeftijdsgrenzen die we hier op resp. 25 en 45 jaar hebben gesteld. De resultaten die aldus verkregen zijn worden weergegeven in Tabel 2. Bij de schatting van het geamendeerde model bleek de t-waarde van de geslachts-dummy bijna nul te zijn zodat deze variabele uit de lijst van regressors verwijderd is. De e-waarde van het geschatte model bedraagt 0,436.

regressor \	lage of geen opl.	leeftijd < 25 > 45		burg. staat	natio- naliteit	ber. niveau	inverse replace- ment ratio	werkl. duur
geschatte coëfficiënt	0,676	-0,677	0,536	0,866	1,996	1,176	-0,489	0,0335
geschatte standaard- afwijking*	0,385	0,537	0,442	0,455	1,171	0,706	0,204	0,026
gemiddelde hele groep	0,34	0,15	0,28	0,37	0,04	0,51	2,39	8,13
gemiddelde bij herin- trede	0,28	0,21	0,18	0,34	0,01	0,49	2,45	6,01

* zie opm. bij Tabel 1.

Tabel 2, Logit Analyse, Het Geamendeerde Basismodel.

Wat de nieuwe variabelen betreft kan geconcludeerd worden dat een laag opleidingsniveau minder kans op herintrede geeft, evenals het ouder zijn dan 45 jaar, Het jonger zijn dan 25 jaar geeft daarentegen een grotere kans op herintrede. De gevonden e-waarde is hoger dan die in het basismodel en lijkt gezien het feit dat we met individuele gegevens werken redelijk hoog. Voorts hebben alle coëfficiënten het juiste teken, blijken alle t-waarden groter

dan één in absolute waarde. We accepteren daarom dit model als acceptabele beschrijving van de herintredekanalen.

Uitgaande van het bovenstaande model is verder geëxperimenteerd met verschillende specificaties van vraag/aanbod factoren en de relatieve beloningsverhouding. Allereerst zijn specifieke vraag/aanbod variabelen opgenomen, die op een (0,1)-schaal aangeven hoe de arbeidsbemiddelaar de vraag naar resp. het aanbod van personen als betrokkene inschatten.

Deze pogingen leidden echter niet tot een duidelijk betere verklaring: de coëfficiënten hadden weliswaar het juiste teken maar de bijbehorende t-waarden lagen dicht bij nul. Verder hebben we geëxperimenteerd met een variabele die gedefinieerd is als de verhouding tussen de openstaande aanvragen en het aantal werklozen per bedrijfstak. Een probleem hierbij was dat deze gegevens niet voor alle personen in de steekproef beschikbaar waren, zodat we ons moesten beperken tot 148 personen. Schatting van het in Tabel 2 gespecificeerde model aangevuld met deze nieuwe vraag/aanbod variabele leverde weinig schokkends op. De e-waarde, ditmaal gevonden in een steekproef van 148 personen, lag in dezelfde orde van grootte als voorheen; de coëfficiënt van de vraag/aanbod variabele had het goede teken maar de bijbehorende t-waarde lag dicht bij nul.

Tenslotte hebben we onderzocht of een andere definitie resp. een opsplitsing van de relatieve beloningsvariabele betere resultaten opleverde. In Par. 4.1. is al opgemerkt dat we ons moesten behelpen met erg ruwe schattingen van het verwachte loon bij herintrede, nl. gemiddeld verdiende loon per bedrijfstak. Binnen iedere bedrijfstak hebben wij nog enige differentiatie proberen aan te brengen door de inverse replacement ratio te vermenigvuldigen met het beroepsniveau van de betrokken persoon. De aldus gedefinieerde variabele deed het echter beduidend slechter dan de oorspronkelijke. Als proxy voor het te verwachten inkomen is het laatstverdiende inkomen van de betrokken persoon nog geprobeerd, maar ook deze variabele voldeed niet beter dan de oorspronkelijke.

Analoog aan Nickel [9] hebben we de inverse replacement ratio opgesplitst in 2 variabelen; de eerste variabele geeft deze ratio voor personen die korter dan 3 maanden werkloos waren op 31 mei 1974 en is nul elders en de tweede variabele is analoog gedefinieerd voor de personen langer dan 3 maanden werkloos. De resultaten hiervan staan in Tabel 3. De e-waarde bedraagt 0.441.

regressor	lage of geen opl.	leeftijd		burg. staat	natio- nali- teit	beroeps niveau	werkl. duur	inverse replace- ment ratio	
		< 25	> 45					< 3 mnd	> 3 mnd
geschatte coëfficiënt	0.735	-0.629	0.515	0.856	2.033	1.267	0.0152	-0.600	-0.385
geschatte standaard- afwijking*	0.393	0.542	0.445	0.458	1.175	0.719	0.0288	0.228	0.221
gemiddelde to- tale groep	0.34	0.15	0.28	0.37	0.04	0.51	8.13	2.49	2.34
gemiddelde over herin- tredeers	0.28	0.21	0.18	0.34	0.01	0.49	6.01	2.70	2.26

* zie opmerking bij Tabel 1.

Tabel 3, Logit - Analyse, Geamendeerd Basismodel met
Gesplitste Inverse Replacement Ratio

Vergelijking van de resultaten in Tabel 2 met die in Tabel 3 leert ons dat de coëfficiënten van de overige variabelen vrijwel onveranderd zijn gebleven met uitzondering van de coëfficiënt van de werkloosheidsduur voorafgaand aan 31 mei. Deze coëfficiënt daalt vrij fors van 0.0335 naar 0.0152. Het belang van deze variabele, die zoals we eerder zagen o.a. als proxy voor de invloed van "omitted variables" fungeert, vermindert blijkbaar als we de inverse replacement ratio splitsen naar kortstondig en langdurig werklozen. De coëfficiënt van de inverse replacement ratio is bij beide groepen werklozen negatief met een t-waarde van resp. -2.63 en -1.74. Voor de groep langdurig werklozen lijkt dit resultaat niet in overeenstemming met de gevonden gemiddelden over de totale groep en de subgroep van herintreders. Deze gemiddelden resp. 2,34 en 2,26 wijzen namelijk op een negatief verband tussen herintrede kans en hoogte van de inverse replacement ratio voor de groep langdurig werklozen. Dit blijkt moeilijk te verklaren. Vanwege de geringe betrouwbaarheid van de gebruikte inkomensgegevens lijkt het echter ook weinig zinvol dit probleem diepgaand te onderzoeken. Wel wordt de these van Nickell dat de relatieve beloningsverhouding van groter belang is voor kortstondig werklozen dan voor langdurig werklozen door onze resultaten bevestigd.

Concluderend kunnen we stellen dat het mogelijk lijkt een redelijke verklaring voor de herintrede kans te vinden. Belangrijke variabelen blijken

persoonsgebonden variabelen als leeftijd, gebrek aan scholing en nationaliteit. Voorts blijkt ook de relatieve beloningsvariabele een zeer duidelijke invloed te hebben op de herintredkans terwijl tenslotte ook aan vraag/aanbod factoren enige, bescheiden, invloed kan worden toegeschreven.

4.3. Resultaten van de Afgeknotte Afhankelijke Variabele Methode.

De in Par. 3.2 besproken afgeknotte afhankelijke variabele methode is gebruikt om de totale werkloosheidsduur te verklaren uit de volgende variabelen⁸⁾:

- de inverse replacement-ratio zoals gedefinieerd in Par. 4.2. In eerste instantie is rekening gehouden met de mogelijkheid van verschillende coëfficiëntwaarden voor deze regressor voor respectievelijk kortstondig en langdurig werklozen,
- de vraag naar werknemers in de beroepsgroep waartoe de betrokken werkloze behoort, door de arbeidsbemiddelaar "gemeten" op (0,1)-schaal,
- de leeftijd van de werkloze in jaren,
- een dummy-variabele die aangeeft of de betrokken werkloze geschoold of ongeschoold is. Deze variabele is één, indien de werkloze ongeschoold is en nul elders.

Deze variabelen worden als afwijkingen van hun steekproefgemiddelde gemeten. Tenslotte wordt een constante term in de regressievergelijking opgenomen.

De resultaten van de eerste schatting, met een uitsplitsing van de inverse replacement-ratio naar kortstondig en langdurig werklozen⁹⁾ zijn weer-gegeven in Tabel 4. De werkloosheidsduur is gemeten in maanden.

8) Voor het numeriek berekenen van de ML-schattingen werd gebruik gemaakt van Newton-achtige methoden, zoals het algoritme van Broyden, Fletcher en Shanno en dat van Fiocco en McCormick. Wij danken Ir. De Beer, Rekencentrum K.H.T., voor zijn steun bij deze fase van het onderzoek.

9) Een persoon is langdurig werkloos, indien zijn totale gemeten werkloosheidsduur groter is dan 7 maanden.

Regress- sor	constante term	inverse replacement ratio		vraag naar betr.	leeftijd in jaren	ongeschoold- geschoold dummy
		kortst. werkl.	langd. werkl.			
geschatte coëfficiënt	19.04	-0.22	3.86	-2.82	0.60	1.82
geschatte standaard afwijking*	1.54	1.48	1.56	4.98	0.12	2.28

* Als standaard-afwijking is gebruikt de geschatte standaard-afwijking van de asymptotische verdeling.

Tabel 4, Resultaten Afgeknotte Regressie voor de Totale Werkloosheidsduur met een Gesplitste Replacement ratio.

Bij de interpretatie van deze resultaten valt allereerst de hoge waarde van de constante term op. Het gevonden steekproefgemiddelde van alle complete en incomplete werkloosheidsperioden bedraagt 12 maanden zodat een verwachte waarde voor de totale werkloosheidsduur van 19 maanden, indien een score nul geldt voor alle andere regressoren, niet excessief hoog hoeft te zijn. Wel dienen we erop te wijzen dat de hoge geschatte waarde van de constante term een gevolg kan zijn van selectiviteit bij de samenstelling van de steekproef, zie hiervoor ook Par. 4.1. Het lijkt aannemelijk dat de gehanteerde selectieregels hebben geleid tot een ondervertegenwoordiging van kortstondig werklozen.

De coëfficiënt van de inverse replacement ratio is negatief voor de groep kortstondig werklozen doch positief voor de groep langdurig werklozen. Debet aan deze laatste, theoretisch onjuiste waarde, is wellicht de opsplitsing van de replacement ratio. Dit resulteert in een dummy-achtig karakter voor de twee aldus gecreëerde variabelen. Het lijkt mogelijk dat de variabele die betrekking heeft op de langdurig werklozen als een proxy functioneert voor de invloed van "omitted variables". Reeds eerder is betoogd dat deze omitted variables vooral van belang zijn bij de groep langdurig werklozen. Anderzijds is de opsplitsing van de replacement ratio die uitgevoerd is aan de hand van de waarde van de endogene variabele vanuit statistisch oogpunt zeer dubieus. Een negatieve waarde voor de coëfficiënt van de eerste variabele en een positieve waarde voor de coëfficiënt van de tweede variabele zijn

dan, gezien het dummy-achtige karakter niet verrassend.

De coëfficiënten van de overige variabelen hebben het juiste teken, significant verschillend van nul lijkt echter alleen de leeftijdsvariabele. Ook de kwantitatieve invloed van deze variabelen lijkt, met uitzondering van de leeftijdsvariabele, gering. Zo leidt ongeschoold zijn tot een verwachte verlenging van de totale werkloosheidsduur van slechts 1.82 maanden t.o.v. de groep geschoolden.

Gelet op bovenstaande discussie lijkt een herschatting van het model, dit maal met een ongesplitste replacement ratio noodzakelijk. De resultaten wijken met uitzondering van de coëfficiënt van de inverse replacement-ratio weinig af van de eerder gevonden schattingen¹⁰⁾ en zijn weergegeven in Tabel 5. De coëfficiënt van de inverse replacement ratio is positief; blijkbaar domineert de groep langdurig werklozen bij de schatting van de coëfficiënt van de opgesplitste variabele. Het hoeft geen betoog dat dit resultaat vanuit theoretisch oogpunt uiterst onbevredigend is.

Regressor	constante term	inverse replacement ratio	vraag naar betr.	leeftijd	ongeschoold-geschoold-dummy
geschatte coëfficiënt	18.48	3.34	-2.40	0.58	1.70
geschatte standaard-afwijking*	1.48	1.35	4.66	0.12	2.32

* Als standaardafwijking is gebruikt de geschatte standaardafwijking van de asymptotische verdeling.

Tabel 5, Resultaten Afgeknotte Regressie voor de Totale Werkloosheidsduur.

Concluderend moeten we stellen dat de resultaten van de afgeknotte regressie methode ter verklaring van de totale werkloosheidsduur tegenvallen.

10) De waarde van de loglikelihoodfunctie daalde van het tweede model t.o.v. het eerste geschatte model slechts met 0.67. Dit lijkt geen krachtige ondersteuning voor de hypothese dat de coëfficiënt van de inverse replacement ratio verschilt voor kortstondig en langdurig werklozen.

Van belang lijken alleen de vraagfactor, de leeftijdsvariabele en een gebrek aan scholing. Daarnaast blijkt de gemiddelde totale werkloosheidsduur verrassend hoog; dit is echter wellicht een gevolg van selectiviteit bij de samenstelling van de steekproef. Betrouwbare schattingen voor de invloed van de beloningsvariabele, in het model opgenomen als de inverse replacement ratio, hebben we echter niet kunnen verkrijgen. We zagen dat dit wellicht te wijten is aan de modelspecificatie die geen onderscheid toelaat in coëfficiëntwaarden voor kortstondig en langdurig werklozen. Een andere verstoringende factor is de invloed van omitted variables.

4.4. Enige Voorzichtige Conclusies.

Alvorens conclusies te trekken n.a.v. de gevonden resultaten lijkt het raadzaam ons eigen oordeel te geven over de gebruikte methoden en suggesties te doen voor verbeteringen. Een belangrijk en voor ons nauwelijks oplosbaar probleem bleek de selectiviteit van de steekproef en de mogelijke aanwezigheid van meetfouten. In toekomstig onderzoek lijken deze problemen echter minder groot indien meer aandacht, moeite en geld gespendeerd kan worden aan het trekken van een (aselecte) steekproef en het verzamelen van aanvullende informatie. Een veel ernstiger probleem vormt het probleem van de weggelaten variabelen vooral bij de verklaring van de totale werkloosheidsduur.

De resultaten van de logit-analyse waren bemoedigend. Het lijkt mogelijk een redelijke verklaring voor de herintrede kans te geven. Belangrijke variabelen blijken persoonsgebonden kenmerken als leeftijd, gebrek aan scholing en nationaliteit. Opvallend is de duidelijke invloed van de inverse replacement ratio; de invloed van vraag/aanbod factoren lijkt daarentegen beperkt. Er werd geen onderscheid tussen mannen en vrouwen gevonden w.b. de herintrede-kans, maar het geringe aantal vrouwen in de steekproef laat geen duidelijke conclusie toe. De resultaten van de afgeknotte regressie ter verklaring van de totale werkloosheidsduur vielen tegen. De resultaten onderstreepten echter de negatieve invloed van de leeftijd op de kans van herintrede. Vraagfactoren en scholingsfactoren hadden wel het juiste teken, maar leken kwantitatief tamelijk onbelangrijk. Tenslotte bleek de invloed van de replacement-ratio tegengesteld aan de theoretisch verwachte invloed. Het is aannemelijk dat het laatste resultaat mede te wijten is aan statistische complicaties, zoals meetfouten en omitted variable bias.

Concluderend moeten we stellen dat de verkregen resultaten geen duidelijke uitspraken toelaten met betrekking tot de actualiteit van de verschil-

lende verklaringen voor de duur van de werkloosheid. Een ondersteuning voor de randgroepen-these, waarbij de werkloosheid geconcentreerd is bij de zwakkere groepen op de arbeidsmarkt, is de aantoonbare invloed van de leeftijd en het ongeschoold zijn.¹¹⁾ De totale geschatte invloed hiervan op de duur van de werkloosheid is echter niet bijzonder groot. Ook aan vraagfactoren kan enige invloed niet worden ontzegd. Tenslotte worden we geconfronteerd met tegenstrijdige informatie over de invloed van de relatieve beloningsvariabele.

LITERATUUR.

- [1] Amemiya, T., Regression Analysis when the Dependent Variable is Truncated Normal, Econometrica, Vol. 41, 1973.
- [2] Heckman, J., The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for such Models, Annals of Economic and Social Measurement, Vol. 5, 1976.
- [3] Koopmans, E. e.a., Onderzoek onder Werklozen, IVA-rapport, Tilburg, 1977.
- [4] Lancaster, T., Econometric Methods for the Duration of Unemployment, Econometrica, Vol. 47, 1979.
- [5] McFadden, D., Conditional Logit Analysis of Quantitative Choice Behavior, in Frontiers in Econometrics, ed. P. Zarembka, Academic Press, 1974.
- [6] McGregor, A., Unemployment Duration and Re-employment Probability, The Economic Journal, Vol. 88, 1978.
- [7] Mortensen, D.T., Unemployment Insurance and Job Search Decisions, Industrial and Labor Relations Review, Vol. 30, 1976.
- [8] Nickell, S.J., The Effect of Unemployment and Related Benefits on the Duration of Unemployment, The Economic Journal, Vol. 89, 1979.

11) In dit verband zij opgemerkt, dat een oververtegenwoordiging van ouderen en ongeschoolden in het werklozenbestand, gelet op de gemiddeld langere werkloosheidsduur van deze groepen, niet impliceert dat zij een grotere kans hebben om werkloos te worden.

- [9] Nickell, S.J., Estimating the Probability of Leaving Unemployment,
Econometrica, Vol. 47, 1979.
- [10] Olsen, R.J. Approximating a Truncated Normal Regression with the Method
of Moments, Econometrica, Vol. 48, 1980.
- [11] Phelps, E. e.a., Microeconomic Foundations of Employment and Inflation
Theory, Norton, 1970.
- [12] Theil, H., Principles of Econometrics, North-Holland, 1971.
- [13] Tobin, J., Estimation of Relationships for Limited Dependent Variables,
Econometrica, Vol. 26, 1958.