

Amsterdam, september 2007

De werkloosheidsduur in Nederland (2002-2005)

dr. P.H.G. Berkhout (SEO)
drs C. van Klaveren (SEO)
drs J. van der Valk (CBS)
drs. R. de Vries (CBS)



Centraal Bureau voor de Statistiek



seo economisch onderzoek

SEO Economisch Onderzoek doet onafhankelijk onderzoek in opdracht van overheid en bedrijfsleven. Ons onderzoek helpt onze opdrachtgevers bij het nemen van beslissingen. SEO Economisch Onderzoek is gelieerd aan de Universiteit van Amsterdam. Dat geeft ons zicht op de nieuwste wetenschappelijke methoden. We hebben geen winstoogmerk en investeren continu in het intellectueel kapitaal van de medewerkers via promotietrajecten, het uitbrengen van wetenschappelijke publicaties, kennisnetwerken en congresbezoek.

SEO-rapport nr. 2007-51

ISBN 978-90-6733-410-5

Inhoudsopgave

SAMENVATTING.....	I
1 INLEIDING	1
2 RAAMWERK EN OPERATIONALISERING	3
2.1 Raamwerk	3
2.2 Operationalisering werkloosheidsduur.....	5
3 BEREKENING VAN DE WERKLOOSHEIDSDUUR	7
3.1 Data imperfecties.....	7
3.2 Modelspecificatie	10
3.3 Plausibiliteitscontrole.....	13
4 ONTWERP VAN EEN STATISTIEK VOOR WERKLOOSHEIDSDUREN	15
4.1 Selectie en weging.....	15
4.2 Duurstatistieken.....	18
5 DE WERKLOOSHEIDSDUUR IN 2002-2005	21
5.1 Frequentieverdeling, mediaan en gemiddelde	21
5.2 Trendcijfers 2002-2005	22
5.3 Werkloosheidsduur en achtergrondkenmerken.....	24
6 CONCLUSIES EN AANBEVELINGEN.....	27
BIJLAGE 1: BEPALING WERKLOOSHEIDSDUUR	29
Operationalisering	29
De onvoltooide werkloosheidsduur.....	29
De voltooide werkloosheidsduur.....	33
BIJLAGE 2: DUURMODEL.....	39
Duurmodellen en de uittredevoet.....	39
Het Weibull model	40
Duurafhankelijkheid, niet-waargenomen heterogeniteit.....	41
Onvoltooide duren voltooien.....	42
Imputatie bij ontbrekende informatie.....	42
Effenen van onwaarschijnlijke pieken.....	43
BIJLAGE 3: STATA VERWERKINGSTAPPEN.....	45

Samenvatting

In dit rapport wordt een onderzoek beschreven naar de mogelijkheden voor ontwikkeling van een statistiek die de duur van werkloosheid in beeld brengt op basis van gegevens uit de Enquête Beroepsbevolking. Het onderzoek vond plaats in het kader van het CBS-speerpunt Sociale dynamiek en uitbreiding Arbeidsmarktreceningen. Het onderzoek werd uitgevoerd door SEO Economisch Onderzoek in samenwerking met het CBS.

De centrale doelstelling is het formuleren en produceren van een betrouwbare statistiek van de werkloosheidsduur, waarbij rekening wordt gehouden met de complicaties die meting van duren met zich meebrengt. Die complicaties zijn 1) het bestaan van onvoltooide duren; en 2) selectiviteit van de steekproef. De steekproef van een dwarsdoorsnede meting is selectief omdat langdurige perioden van werkloosheid een grotere kans hebben om te worden waargenomen dan korte perioden van werkloosheid.

De werkloosheidsduur wordt gedefinieerd als een ononderbroken tijdsperiode van 'niet-werken' voorafgaand aan een positie waarin wel wordt gewerkt. Een werkloosheidsduur ontstaat tijdens een transitie op de arbeidsmarkt naar werk. De toestand vóór de transitie kan elke van de volgende drie arbeidsmarktposities zijn: werkend, studierend en niet-participerend. In welke mate een individu in transitie zich de gehele periode beschikbaar stelt voor werk en actief hiernaar zoekt, blijft buiten beschouwing. Met andere woorden; er wordt verondersteld dat een persoon in transitie beschikbaar is voor werk van 12 uur of meer per week en hier ook actief naar of zoek is. Bovendien wordt verondersteld dat mensen in transitie in het geheel niet werken, ook niet in deeltijd. Ook kan men in transitie niet (voltijd) studeren of naar school gaan.

De werkloosheidsduur wordt geoperationaliseerd door twee in maanden gemeten tijdstippen van elkaar af te trekken. Voor de bepaling van de startdatum van werkloosheid is gebruik gemaakt van de einddatum van de laatste baan, de begindatum van zoeken naar werk en datum van schoolverlaten. Hiervan is de meest recente datum genomen. De einddatum van de (onvoltooide) werkloosheidsduur is voor mensen die werkloos zijn op het moment van enquêteren gelijk aan de enquêtemaand. Voor mensen die werken is de einddatum gelijk aan de begindatum van de huidige baan. Verondersteld wordt dat mensen in een nieuwe baan beginnen op de eerste dag van de maand en dat een baan eindigt op de laatste dag van de maand.

Een regressie-duurmodel wordt gebruikt om de duur van mensen die werkloos zijn op het moment van enquêteren te voltooien. Dit model wordt tevens ingezet om enkele imperfecties weg te werken in de beschikbare EBB-data die het opstellen van een betrouwbare statistiek voor werkloosheidsduren bemoeilijken. Het gaat dan om ontbrekende informatie over de periode van inactiviteit van personen en pieken in de duren als gevolg van afronding door respondenten bij retrospectieve vragen.

Voordat een statistiek gemaakt kan worden dienen enkele keuzes te worden gemaakt. De eerste keuze betreft de soort statistiek. Hoewel het gemiddelde van de werkloosheidsduur wellicht voor de hand lijkt te liggen, verdient de mediaan de voorkeur. Deze statistiek is namelijk veel minder

gevoelig voor extreem lange waarnemingen dan het gemiddelde. Tevens blijft het effect van eventuele overschattingen in de modelvoorspellingen daardoor beperkt. De mediaan is niet ongebruikelijk. Bijvoorbeeld bij inkomensstatistieken wordt deze ook sinds jaar en dag toegepast.

De tweede keuze betreft de selectie van waarnemingen. Een belangrijke functie van een door het CBS te publiceren statistiek voor werkloosheidsduren is dat ermee een beeld wordt gegeven van de toestand op de arbeidsmarkt in een bepaalde periode. Een dergelijk cijfer moet snel na afsluiting van de periode kunnen worden gepubliceerd. De keuze van de lengte van de periode heeft invloed op de statistiek. Hoe korter de periode, hoe meer de nadruk komt te liggen op mensen die werkloos worden waargenomen op het moment van enquêteren. Hoe langer de periode, hoe meer de nadruk ligt op mensen die in de betreffende periode een baan vonden. Het gevolg hiervan is dat bij een lange periode van beschouwing de mediaan lager ligt dan bij een korte periode. Dit komt voort uit het feit dat mensen met een lange werkloosheidsduur een grotere kans hebben om als werkloze te worden waargenomen op het moment van enquêteren, dan mensen met een korte werkloosheidsduur. Ten aanzien van de rapportageperiode wordt voorgesteld om te kiezen voor het kwartaal. De statistiek doet een uitspraak over de leden van de beroepsbevolking waarvan de werkloosheidsduur (voor een deel) in het betreffende kwartaal viel. Het betreft dus enerzijds mensen die in het betreffende kwartaal een baan vonden, en anderzijds mensen die aan het einde van het kwartaal nog werkloos waren ongeacht wanneer de werkloosheidsduur begon.

Bij de selectie van waarnemingen moet ook een keuze worden gemaakt met betrekking tot de baanwisselaars. Dit zijn mensen die niet later dan drie maanden na de vorige baan in de huidige baan begonnen. Weglaten van deze mensen houdt in dat een relatief grote groep van korte werkloosheidsduren buiten beschouwing blijft, met uiteraard gevolgen voor de uitkomst van de statistiek. Mensen die van baan veranderen zonder daarbij werkloos te zijn geweest, horen niet thuis in een statistiek die onvrijwillige werkloosheid moet verbeelden. Aan de andere kant is de periode van drie maanden in de definitie van baanwisselaars behoorlijk lang. Hierdoor worden bijvoorbeeld mensen van wie het contract afliep maar binnen enkele maanden weer een nieuwe baan vonden, ten onrechte aangezien voor baanwisselaars. Gekozen is om een middenweg te bewandelen. Alleen de baanwisselaars met een werkloosheidsduur van minder dan één maand worden niet meegenomen bij het berekenen van de werkloosheidsduur.

De hierboven beschreven methode, waarbij gebruik wordt gemaakt van kwartaalcijfers, levert de volgende resultaten op voor de jaren 2002-2005. In 2005 is de mediaan van de werkloosheidsduur 15,3 maanden. De mediaan was vooral in 2004 en 2005 hoog. In 2002 en 2003 vonden werklozen relatief snel een baan en lag de mediaan van de werkloosheidsduur een stuk lager (respectievelijk 6,2 maanden en 10,8 maanden). Deze cijfers gelden voor de populatie als geheel; er bestaan echter grote verschillen in werkloosheidsduur naar herkomst, leeftijd en opleidingsniveau. Vooral jongeren, autochtonen en hoger opgeleiden vinden relatief snel een baan. Voor 15-24 jarigen is de mediaan van de werkloosheidsduur 7,7 maanden. Vervolgens loopt deze op met de stijging van de leeftijd tot maar liefst 35,3 maanden voor 55-64 jarigen. Voorts vinden autochtonen in 12,7 maanden een baan in 2005, terwijl dit voor niet-westerse allochtonen 26,1 maanden is. Wat betreft het opleidingsniveau geldt dat vooral personen met alléén basisonderwijs er lang over doen om een baan te vinden. Voor hen is de mediaan van de werkloosheidsduur 38,2 maanden in 2005. Ter vergelijking; personen met een hbo-diploma vinden doorgaans in 8,2

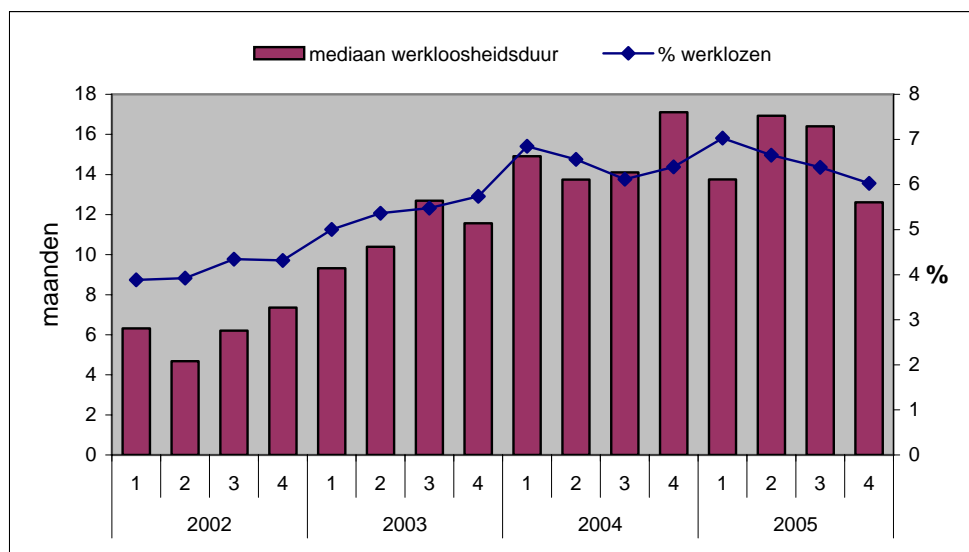
maanden een baan. De hier vermelde resultaten betreffen voorlopige cijfers. Zoals aangegeven wordt in dit rapport een onderzoek beschreven naar de mogelijkheden voor de ontwikkeling van een statistiek die de duur van werkloosheid in beeld brengt. De cijfers in dit rapport hebben dan ook een voorlopig karakter en kunnen eventueel in een later stadium worden bijgesteld.

Tabel S1. De voltooide werkloosheidsduur 2002-2005 op basis van kwartaalcijfers, mediaan en in klassen van 5 maanden

	Mediaan (maanden)	0-5 maanden %	6 - 11 maanden %	12 - 17 maanden %	18 - 23 maanden %	>= 24 maanden %
Jaar						
2002	6,2	49	14	8	5	24
2003	10,8	37	16	11	6	30
2004	14,7	30	14	11	6	39
2005	15,3	31	13	11	6	38

Bron: CBS (EBB 2002-2005)

Figuur S1. Mediaan werkloosheidsduur en percentage werklozen per kwartaal, 2002-2005



Bron: CBS (EBB 2002-2005)

Het voltooiën van de werkloosheidsduur van werklozen heeft op twee manieren invloed op de uiteindelijke resultaten. Ten eerste is door voltooiing van de werkloosheidsduur van werklozen de mediaan van de werkloosheidsduur (in bijvoorbeeld 2005) toegenomen van 8,0 maanden tot 15,3 maanden. Ten tweede zijn de verhoudingen binnen groepen veranderd na voltooiing van de dueren. De verschillen naar leeftijd, herkomst en opleidingsniveau die er al waren in de onvoltooide dueren, zijn groter geworden. Zo zijn de verschillen in werkloosheidsduur tussen jongeren en ouderen, autochtonen en niet-westerse allochtonen en laagopgeleiden en hoog

opgeleiden toegenomen. Aldus ontstaat een wat somberder beeld voor ouderen, niet-westerse allochtonen en de laagst opgeleiden van de kans om op korte termijn een baan te vinden.

De gebruikte informatie uit de EBB is nog niet optimaal. Zo is een zeer ongewenste omissie dat van de baanvinders niet bekend is wanneer zij begonnen met zoeken naar hun huidige baan. Hierdoor is de meting van de werkloosheidsduur van mensen die na een periode van inactiviteit weer zijn gaan zoeken van mindere kwaliteit. Dit betreft met name herintreders en arbeidsongeschikten. Meting van het zoekmoment onder baanvinders maakt de werkloosheidsduurstatistiek veel betrouwbaarder voor deze vanuit beleidsoogpunt zeer belangrijke arbeidsmarktgroepen. Voorts is de duur gemeten in maanden. Omdat echter niet bekend is op welke dag iemand begint met werken en op welke dag in de maand de vorige activiteit is beëindigd, is het lastig om deze maanden exact te bepalen. Daarom is aangenomen dat de vorige activiteit aan het einde van de maand plaatsvond en de baan aan het begin van de maand. Onderzocht zou moeten worden of deze aanname gegrond is. Aanbevolen wordt om deze informatie te verzamelen. Dit zou eventueel kunnen met behulp van registraties. Tenslotte is het van belang om bij baanwisselaars die niet direct van de ene baan op de andere overstappen, te onderzoeken of er een periode van werkloosheid tussen de ene en de andere baan zat. Hierover zou in de EBB informatie kunnen worden verzameld.

Het onderzoek draagt een aantal onderwerpen voor vervolgonderzoek voor. Zo is tijdens dit project alleen nog maar gebruik gemaakt van de informatie uit de eerste, *face-to-face*, peiling van de EBB. De vervolgpeilingen per telefoon zijn niet gebruikt voor het opstellen van de duurstatistiek, wel voor een plausibiliteitscontrole (zie paragraaf 3.3). Gebruik van de telefonische data zal de betrouwbaarheid van de statistiek verhogen, door een toename van het aantal waarnemingen. Een tweede punt dat onderzocht zou moeten worden is de situatie dat iemand zich na een periode van werkloosheid terugtrekt van de arbeidsmarkt. Vooralsnog is aangenomen dat elke persoon die op het moment van ondervragen werkloos is, blijft zoeken naar een baan. Bij oudere werknemers en ontmoedigden is dit wellicht niet correct. Hierdoor ontstaan mogelijk heel lange werkloosheidsduren, terwijl de betreffende respondenten helemaal niet meer zoeken. Om dit te ondervangen kan een ander duurmodel worden toegepast (*competing risks* model), zodat het mogelijk wordt om vanuit de transitie over te gaan naar twee mogelijke toestanden: werk en inactiviteit. Ten derde is binnen dit project niet meegenomen dat informatie uit registers, in het bijzonder ingeschreven staan bij het CWI of een uitkering hebben, gebruikt kan worden om de werkloosheidsduur beter af te bakenen. Dit zal ook onderzocht moeten worden.

1 Inleiding

Dit rapport is samengesteld in het kader van een project van het CBS-speerpunt Sociale dynamiek en uitbreiding Arbeidsrekeningen. Doelstelling van het speerpunt is een verbetering van de statistische informatievoorziening over de sociaaleconomische levensloop van mensen. Het speerpunt Sociale Dynamiek omvat de beschrijving van veranderingen in demografische kenmerken, onderwijs (*human capital*), arbeidsmarktpositie, materiële welvaart en maatschappelijke participatie en hun onderlinge samenhang. Over demografische transitie, schoolloopbanen en inkomensdynamiek worden door het CBS al geruime tijd regulier gegevens samengesteld. Dat geldt veel minder voor de dynamiek op de arbeidsmarkt en de onderlinge samenhang van de genoemde dimensies. In 2007 staat de dynamiek op de arbeidsmarkt centraal en de samenhang daarvan met *human capital*. Om inzicht te verkrijgen in de voortdurende beweging op de arbeidsmarkt, is het van belang de dynamiek op de arbeidsmarkt goed in kaart te brengen. In overeenstemming met de Lissabonstrategie is het arbeidsmarktbeleid gericht op de bevordering van de dynamiek op de arbeidsmarkt. Een dynamische arbeidsmarkt past niet alleen in het streven naar volledige werkgelegenheid, maar is ook een antwoord op de huidige, meer dynamische mondiale economie. Het maken van statistieken van de werkloosheidsduur, naast de standcijfers van werkloosheid, is van belang om een drietal redenen. Ten eerste lopen standcijfers van de werkloosheid altijd achter bij economische ontwikkelingen. Ten tweede kan de dynamiek van de arbeidsmarkt voor allerlei subgroepen er beter mee in beeld worden gebracht. Ten derde is meting van de individuele werkloosheidsduur noodzakelijk om de vanuit beleidsoogpunt uiterst belangrijke groep langdurige werklozen te kunnen identificeren.

De centrale doelstelling in dit onderzoek is het ontwikkelen van een betrouwbare statistiek van de werkloosheidsduur, waarbij rekening wordt gehouden met de complicaties die meting van duren met zich meebrengt. Die complicaties zijn in beginsel: 1) het bestaan van onvoltooide duren; en 2) selectiviteit van de steekproef. De steekproef is selectief omdat langdurig werklozen een grotere kans hebben om als werkloze te worden waargenomen dan kortdurig werklozen. Het probleem van onvoltooide duren doet zich voor wanneer respondenten op het moment van enquêteren werkzoekend zijn. Van hen is niet bekend wat de uiteindelijke werkloosheidsduur zal worden. Er is in dat geval sprake van onvolledige informatie: we weten niet precies hoe lang de duur was, maar alleen hoe lang zij ten minste was. Het gevolg hiervan is dat statistieken die men doorgaans gebruikt om een populatie te beschrijven (zoals het gemiddelde, de standaard deviatie en de mediaan) niet meer zonder grote onzuiverheid kunnen worden berekend. In Berkhout (2002) wordt beschreven hoe men hiervoor kan corrigeren. De gedachte achter die methode is om onvoltooide duren te completeren op basis van modelberekeningen. In het voorliggende rapport is onderzocht of het met een dergelijke methode mogelijk is om betere schattingen te maken van de werkloosheidsduur. Hiervoor is data gebruikt van de Enquête Beroepsbevolking 2002-2005.

2 Raamwerk en operationalisering

In dit hoofdstuk wordt in paragraaf 2.1 ingegaan op de gehanteerde definitie van de werkloosheidsduur. Paragraaf 2.2 beschrijft hoe die definitie in de praktijk is geoperationaliseerd.

2.1 Raamwerk

In het dynamische model van de ‘transitionele’ arbeidsmarkt laveren individuen in de tijd tussen een beperkt aantal afgebakende ‘toestanden’. Dat wil zeggen: zij kunnen enige tijd werken, vervolgens terug gaan naar de schoolbanken voor een aanvullende studie, zich een periode terugtrekken van de arbeidsmarkt om zorgtaken te verrichten, tijdelijk arbeidsongeschikt zijn, etc. Voor de definitie van de werkloosheidsduur gaan we ervan uit dat mensen in één van de volgende drie toestanden verkeren: 1) werkend; 2) studerend; 3) niet-participerend. De laatste genoemde categorie is een restgroep met individuen die op dat moment om allerlei redenen niet werken of studeren, maar later weer kunnen herintreden. Bijvoorbeeld mensen die arbeidsongeschikt zijn, maar ook mensen die thuis zorgtaken verrichten. De keuze voor deze drie toestanden impliceert dat werkloosheid zelf geen toestand is in ons model. Wij beschouwen werkloosheid als een periode van ‘in-transitie-zijn’ tussen twee toestanden; als een frictie die gepaard gaat aan de overgang van de ene toestand naar de andere.

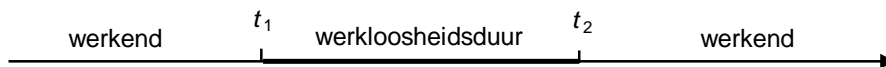
De *werkloosheidsduur* wordt gedefinieerd als een ononderbroken tijdsperiode van ‘niet-werken’ voorafgaand aan een positie waarin wel wordt gewerkt. Een werkloosheidsduur ontstaat dus alleen tijdens een transitie op de arbeidsmarkt naar werk. De toestand vóór de transitie kan elke mogelijke arbeidsmarktpositie zijn: werkend, studerend en inactief. In welke mate een individu in transitie zich beschikbaar stelt voor werk en actieve zoekactiviteiten onderneemt, blijft buiten beschouwing. Deze begrippen zijn in de praktijk uiterst moeilijk meetbaar en bovendien variëren ze per individu in de tijd. Om complicaties te voorkomen veronderstellen we dat mensen in transitie in het geheel niet werken, ook niet in deeltijd. Ook kan men in transitie niet (voltijd) studeren of naar school gaan.

Omdat we drie arbeidsmarkttoestanden onderscheiden, zijn er drie typen werkloosheidsduur. Figuur 2.1 geeft ze grafisch weer. De werkloosheidsduur wordt afgebakend door twee tijdstippen: t_1 en t_2 . Bij type A (baanwisselaars/ontslagwerklozen) wordt t_1 gemarkeerd door het einde van de vorige baan. Bij type B (schoolverlaters) begint de werkloosheidsduur op het moment van schoolverlaten, al dan niet met diploma. Voor herintreders begint de werkloosheidsduur op het moment dat zij zich weer beschikbaar stellen voor de arbeidsmarkt en zoekactiviteiten ondernemen. Op dat moment eindigt hun periode van inactiviteit. Voor typen A, B en C wordt t_2 gegeven door het moment waarop de huidige baan aanvangt of de toekomstige baan zal aanvangen. Het tijdstip t_2 wordt waargenomen indien het voorafgaat aan het moment van enquête. We spreken dan van een *voltooid* werkloosheidsduur. Ook is mogelijk dat individuen in transitie worden waargenomen, dat wil zeggen: op het moment van enquêteren nog geen werk hebben gevonden. Het tijdstip t_2 is dan nog onbekend. We spreken in die gevallen van een *onvoltooid* werkloosheidsduur. Deze waarnemingen leveren incomplete informatie, omdat ze

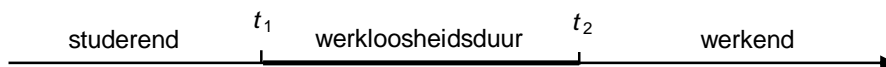
slechts informatie bevatten over hoe lang de werkloosheid ten minste heeft geduurd. Onvoltooiden kunnen zonder problemen worden gecompleteerd. Met behulp van wat bekend is van de persoon kan voor iedereen de ontbrekende informatie worden geschat aan de hand van een duurmodel.

Figuur 2.1 Drie typen werkloosheidsduur

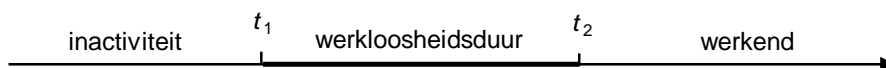
A. Baanwisselaars/ Ontslagwerklozen



B. Schoolverlaters



C. Herintreders



We merken op dat in de gekozen definitie de onvrijwilligheid van de werkloosheid geen rol speelt. De werkloosheidsduur meet heel algemeen de lengte van een periode van niet-werken voorafgaand aan het vinden van een (nieuwe) baan. Dit betekent dat vrijwillige baanwisselaars en ontslagwerklozen identiek worden behandeld. De vrijwillige baanwisselaar die tussen de twee banen een maand niet werkt, heeft een werkloosheidsduur van één maand. De ontslagwerkloze die na zijn gedwongen ontslag een maand later weer begint in een nieuwe baan heeft ook een werkloosheidsduur van één maand.

De gekozen definitie staat toe dat de werkloosheidsduur nul is. Dat doet zich voor indien men tijdens de transitie niet zonder werk zit, met andere woorden: de uitgangssituatie gaat naadloos over in een toestand van betaald werk verrichten. Mensen die van baan wisselen en daarbij van de ene op de andere dag bij een andere baas werken, hebben volgens onze definitie een werkloosheidsduur van nul bij hun transitie. De herintreder die besluit weer te gaan werken nadat hem een baan is aangeboden, heeft eveneens een werkloosheidsduur van nul maanden.

De hier gepresenteerde definitie van de werkloosheidsduur wijkt af van gangbare definities met betrekking tot werkloosheid. De definitie staat beschikbaar voor de arbeidsmarkt, zoekintensiteit en de vrijwilligheid van de werkloosheid in elke hoedanigheid toe. Zelfs een werkloosheidsduur van nul, waaraan in wezen geen werkloosheid te pas komt, is in de definitie mogelijk. Wij kiezen voor een zo ruime definitie omdat we bij het operationaliseren van het concept werkloosheidsduur dan niet of nauwelijks tot het maken van ad-hoc keuzes worden gedwongen. Voor aansluiting met reeds bestaande statistieken en definities vormt onze ruime definitie geen probleem, omdat zij strengere definities omvat en men altijd van elke gewenste subpopulatie de werkloosheidsduur kan beschouwen door selecties te maken van de juiste eenheden.

2.2 Operationalisering werkloosheidsduur

De gegevens die in dit onderzoek worden gebruikt, zijn afkomstig van de Enquête Beroepsbevolking (EBB), jaargangen 2002-2005. Hiervan zijn baanvinders en werklozen in de leeftijd 15-64 jaar genomen. Baanvinders zijn personen die op het moment van ondervragen minder dan een jaar geleden in hun huidige baan van 12 uur of meer per week begonnen zijn. Werklozen zijn personen die op het moment van ondervragen tot de werkloze beroepsbevolking behoren. Scholieren en studenten worden niet in de analyse opgenomen. Tot de scholieren/studenten wordt iedereen gerekend die op het moment van ondervragen een reguliere voltijd, deeltijd of duale opleiding volgt. In tegenstelling tot bijvoorbeeld schoolverlaters zijn scholieren/studenten niet in-transitie tussen twee toestanden. Zij worden hier beschouwd als personen met als belangrijkste bezigheid het volgen van onderwijs. In Bijlage 1 van dit rapport wordt uitvoerig ingegaan op de operationalisering van de werkloosheidsduur voor baanvinders en werklozen.

Voor baanvinders is de voltooide werkloosheidsduur bepaald. Zoals in de vorige paragraaf is aangegeven wordt de werkloosheidsduur afgebakend door twee tijdstippen t_1 en t_2 (zie Figuur 2.1). De voltooide werkloosheidsduur is de duur tussen de startdatum van werkloosheid t_1 tot en met de datum van het vinden van een baan t_2 . Voor de bepaling van de startdatum van werkloosheid van baanvinders is gebruik gemaakt van de einddatum van de laatste baan, de begindatum van zoeken naar werk en datum van schoolverlaten. Hiervan is de meest recente datum genomen als startdatum van werkloosheid. De begindatum van zoeken naar werk is voor herintreders onder de baanvinders niet bekend. Dat moment wordt gesimuleerd aan de hand van model-schattingen (zie paragraaf 3.1.2).

Voor werklozen is het tijdstip van het vinden van een baan t_2 onbekend. Voor werklozen is derhalve de onvoltooide werkloosheidsduur bepaald. De onvoltooide werkloosheidsduur is de duur tussen de startdatum van werkloosheid t_1 tot en met de datum van ondervraging. Voor de bepaling van de startdatum van werkloosheid van werklozen is eveneens gebruik gemaakt van de datum van de laatste baan, de begindatum van zoeken naar werk en datum van schoolverlaten. Van deze drie data wordt de meest recente genomen als startdatum van werkloosheid.

De startdatum van werkloosheid, de datum van het vinden van de baan en de datum van ondervraging hebben betrekking op de maand waarin dit heeft plaatsgevonden. De werkloosheidsduur wordt dus berekend in maanden. Vooral voor personen die kort werkloos zijn, kan dit enigszins een vertekening van de werkelijkheid geven. Het kan zijn dat iemand die op 31 januari is gestopt met werken op 1 februari wordt geënquêteerd. Deze persoon heeft dan een werkloosheidsduur van 1 maand, terwijl dit in werkelijkheid slechts 1 dag is. Aan de andere kant kan iemand aan het begin van de maand werkloos zijn geworden en aan het eind van de maand worden geënquêteerd. Voor deze persoon is een werkloosheidsduur van 0 maanden berekend, terwijl dit in werkelijkheid eerder 1 maand is. Wij hebben ervoor gekozen om iedereen met een werkloosheidsduur van 0 of 1 maanden in de modelberekeningen een werkloosheidsduur van een halve maand te geven.

3 Berekening van de werkloosheidsduur

De Enquête Beroepsbevolking (EBB) heeft tot doel de Nederlandse beroepsbevolking in kaart te brengen. Dat het mogelijk is om met EBB-gegevens individuele werkloosheidsduren vast te stellen, komt doordat de EBB uitgebreid ingaat op het arbeidsmarkt- en schoolverleden van respondenten. In dat verleden worden veel momenten geregistreerd op basis waarvan men inzicht kan krijgen in de werkloosheidsduur van mensen. In hoofdstuk 2 (paragraaf 2.2) en meer in detail in Bijlage 1 is beschreven hoe de werkloosheidsduur is geoperationaliseerd. Daarin komt naar voren dat de EBB-gegevens een aantal imperfecties vertonen, die het maken van een statistiek voor werkloosheidsduren bemoeilijken. Wij gebruiken in dit onderzoek een regressiemodel om die imperfecties te ondervangen, zodat de te ontwerpen duurstatistiek minder gevoelig zal zijn voor anomalieën in de data. In dit hoofdstuk beschrijven wij op welke momenten en op welke wijze het regressiemodel wordt ingezet. Voor een technische beschrijving van het model verwijzen we naar Bijlage 2. Drie vormen van data-imperfecties doen zich voor: onvoltooide duren; ontbrekende informatie en onwaarschijnlijke frequentiepieken. We bespreken ze in afzonderlijke secties in paragraaf 3.1. In paragraaf 3.2 beschrijven we de keuzes die gemaakt zijn bij de modelspecificatie. Het hoofdstuk sluit af met een paragraaf over de plausibiliteit van de modelberekeningen.

3.1 Data imperfecties

3.1.1 Onvoltooide duren voltooien

Een werkloosheidsduur is een zich in de tijd afspelende grootte en dat betekent dat het onvermijdelijk is dat we bij de momentopname van een enquête duren waarnemen die nog op hun voltooiing wachten. Dit heeft ernstige gevolgen voor de berekening van bijvoorbeeld het populatiegemiddelde. Het zou namelijk onjuist zijn om voltooide en onvoltooide duren op te tellen en ook het weglaten van de onvoltooide duren biedt geen soelaas omdat daarmee een selectieve groep van waarnemingen wordt weggelaten. Met andere woorden: het bestaan van onvoltooide duren maakt het onmogelijk om simpele beschrijvende statistieken te berekenen van de werkloosheidsduur.

Een werkloosheidsduur is nog niet voltooid als de EBB-respondent op het moment van enquête nog 'in transitie' (werkloos) is. We voltooien alle onvoltooide werkloosheidsduren op basis van wat bekend is van de individu met behulp van de modelschattingen. Voor elke respondent met een onvoltooide werkloosheidsduur wordt het (nog onbekende) restant geschat in het duurmodel. Wanneer het resultaat van deze schatting voorbij de wettelijke pensioenleeftijd ligt, dan wordt de werkloosheidsduur bij de pensioengerechtigde leeftijd afgekap. Het eindpunt van de kunstmatig voltooide werkloosheidsduur ligt uiteraard in de nog onbekende toekomst. Het is niet waarschijnlijk dat de betreffende respondenten in werkelijkheid precies de werkloosheidsduur zullen krijgen die wij met behulp van het model simuleren. Echter, voor het presenteren van statistieken voor grote groepen is dat geen probleem: waar de ene respondent een kortere duur heeft gekregen, krijgt de andere een langere. Het gemiddelde en de mediaan van de groep

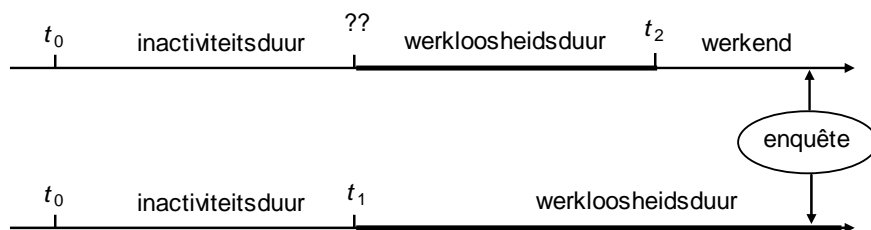
ondervinden daarvan geen last. Voor een formele beschrijving van de schatting van voltooide werkloosheidsduren verwijzen we naar Bijlage 2.

3.1.2 Imputatie bij ontbrekende informatie

Ten tweede doet zich het probleem voor dat de EBB niet voor alle respondenten dezelfde informatie verzamelt, waardoor soms informatie ontbreekt om het beginpunt van de werkloosheidsduur vast te stellen. Wij benaderen die informatie voor enkele speciale groepen op basis van modelvoorspellingen.

In het bijzonder het moment waarop mensen zijn begonnen met zoeken naar werk wordt niet aan alle respondenten gevraagd. Dit blijkt een probleem in het geval van herintreders. Bij hen wordt de werkloosheidsduur voorafgegaan door een lange periode van inactiviteit, dat wil zeggen door een periode van niet-participeren op de arbeidsmarkt. Indien de herintreder werkzaam is op het moment van enquête dan ontbreekt die informatie, is hij werkloos dan is die informatie wel voor handen. Wij simuleren het aanvangsmoment van zoeken naar werk door de inactiviteitsduur van werklozen voorafgaand aan de werkloosheidsduur te analyseren met behulp van het duurmodel. Aan de hand van het geschatte model en de persoonskenmerken van de herintreder wordt vervolgens de lengte van zijn inactiviteitsduur gesimuleerd, met als gevolg dat daarmee impliciet het begin van de werkloosheidsduur wordt vastgesteld.

Figuur 3.1 Inactiviteitsduren



In Figuur 3.1 wordt grafisch weergegeven hoe de inactiviteitsduur wordt gebruikt bij het bepalen van de werkloosheidsduur. In de bovenste tijdlijn is de herintreder werkzaam op het moment van enquêteren. Van hem zijn tijdstippen t_0 en t_2 bekend: we weten wanneer de herintreder inactief werd en wanneer hij begon in zijn huidige baan. Onbekend is het moment waarop de respondent weer begon met zoeken naar een baan. In de figuur wordt dat aangegeven met twee vraagtekens. In de onderste tijdlijn wordt de situatie afgebeeld voor herintreders die werkloos zijn op het moment van enquêteren. Ook nu zijn twee tijdstippen bekend: t_0 en t_1 . We weten hoe lang de herintreder inactief was alvorens hij zich weer aanbood op de arbeidsmarkt en zijn huidige werkloosheidsduur begon. We gebruiken gecompleteerde inactiviteitsduren van werkloze herintreders om die van werkende respondenten te schatten, zodat van hen het beginpunt van de werkloosheidsduur wordt vastgesteld.

De achterliggende gedachte is dat de duur van inactiviteit van baanvinders kan worden voorspeld op basis van waargenomen persoonskenmerken en dat die duur nooit langer voortduurt dan tot tijdstip t_2 . Indien de voorspelde inactiviteitsduur langer is dan $t_2 - t_0$ dan wordt de daaruit

volgende werkloosheidsduur automatisch gelijk aan nul en dat betekent dat aan de baan geen zoekperiode vooraf ging. Het betreft dan een herintreder die weer besluit te gaan werken omdat hem een baan werd aangeboden.

Wij merken op dat wordt verondersteld dat inactiviteitsduren van baanvinders en werklozen – met verder dezelfde achtergrondkenmerken - niet structureel van elkaar verschillen. Dat is een sterke aanname. Men kan zich voorstellen dat mensen die lang inactief zijn geweest, veel moeilijker een baan zullen vinden dan zij die kort inactief waren. Aangezien mensen met een lange werkloosheidsduur met grotere kans als werkloos worden waargenomen in de EBB, mag dus worden verwacht dat daaraan een relatief lange inactiviteitsduur voorafging. De inactiviteitsduur van de baanvinders zal derhalve worden overschat, als ze wordt bepaald met behulp van een model geschat op inactiviteitsduren van werklozen. De werkloosheidsduur wordt daardoor onderschat. We benadrukken dat dit probleem zich voordoet voor een relatief kleine groep in de beroepsbevolking: mensen die na een periode van inactiviteit weer toetreden tot de beroepsbevolking. Er zijn voor het aanvullen van deze ontbrekende informatie echter geen alternatieven voor handen. In de toekomst kan dit probleem eenvoudig worden voorkomen door in de EBB ook aan baanvinders te vragen wanneer zij zijn begonnen met zoeken naar hun huidige baan.

Een andere vorm van ontbrekende informatie is dat de werkloosheidsduur op basis van EBB-informatie nooit preciezer dan in maanden kan worden berekend. Voor de berekening van het populatiegemiddelde is dat geen probleem, maar wanneer men bijvoorbeeld de mediaan van de werkloosheidsduur wil presenteren, dan kan de meeteenheid van maanden te grof zijn.

Voor een formele beschrijving van imputatie van ontbrekende informatie verwijzen we naar Bijlage 2.

3.1.3 Modelleren van onwaarschijnlijke frequentiepieken

Wanneer aan werklozen wordt gevraagd wanneer zij in het verleden zijn begonnen met zoeken naar werk, dan vertonen ruwe frequentietabellen van de verstreken tijd tot aan het moment van enquêteren (onwaarschijnlijke) pieken bij duren van 6, 12, 18, 24, 36 (etc) maanden. Dit doet vermoeden dat respondenten bij het opgeven van retrospectieve informatie afronden op halve jaren in het nabije verleden en op hele jaren in het verre verleden. In sommige gevallen wordt de werkloosheidsduur bepaald door middel van dit opgegeven zoekmoment. Hierdoor ontstaan onwaarschijnlijke pieken in de werkloosheidsduur bij veelvouden van 6 maanden. Deze pieken kunnen worden gladgestreken wanneer we bereid zijn te veronderstellen dat respondenten bij het afronden niet structureel in een bepaalde richting afronden (bijvoorbeeld altijd naar beneden), maar dat doen volgens de regels van het afronden.

Het gladstrijken van frequentiepieken gaat als volgt. We stellen vast in welk interval het door de respondent gegeven antwoord moet liggen. Stel, we nemen te veel zessen waar en vermoeden dat een deel van die zessen afrondingen zijn van getallen in het interval $\{3; 9\}$. Op basis van het aantal waarnemingen in dat interval en het aantal waargenomen zessen, wordt een proportioneel deel van de 6-waarnemingen volgens de individueel geschatte verdelingsfunctie willekeurig gesimuleerd in het interval $\{3; 9\}$. Met andere woorden: een deel van de 6-waarnemingen wordt

verdeeld over het interval. Voor een formele beschrijving van het modelleren van onwaarschijnlijke pieken verwijzen we naar Bijlage 2.

3.2 Modelspecificatie

In de analyse van de werkloosheidsduur wordt een aantal achtergrondkenmerken van respondenten meegenomen waarvan wordt aangenomen dat deze van invloed zijn op de werkloosheidsduur. Het betreft kenmerken van personen die over het algemeen worden gebruikt in onderzoek naar de kans op werk. Natuurlijk zijn er nog meer variabelen te bedenken die de werkloosheidsduur van personen bepalen, maar deze zijn niet altijd in de EBB voor handen. We denken met de door ons geselecteerde variabelen de werkloosheidsduur goed te kunnen schatten. Het gaat om de volgende kenmerken: geslacht, leeftijd, herkomst, opleidingsniveau, opleidingssector, leeftijd jongste kind, huishoudsituatie, werkloosheidstype, rba-gebied waarin een persoon woont, het aantal uren dat een persoon wil werken, werk partner en opleidingsniveau van de partner.

Het geslacht maakt onderscheid tussen mannen en vrouwen. De leeftijd is gemeten in jaren. Wat betreft herkomst wordt een onderscheid gemaakt tussen autochtonen, westerse allochtonen en niet-westerse allochtonen. Herkomst wordt bepaald aan de hand van het geboorteland van de persoon of dat van de moeder, tenzij de moeder in Nederland is geboren. In dat geval is de persoon ingedeeld naar het geboorteland van de vader. Autochtonen zijn personen van wie beide ouders in Nederland zijn geboren, ongeacht het land waar ze zelf zijn geboren. Personen worden gerekend tot de allochtonen als ten minste één ouder in het buitenland is geboren. Tot de categorie 'niet-westers' behoren allochtonen uit Turkije, Afrika, Latijns-Amerika en Azië met uitzondering van Indonesië en Japan.

Het opleidingsniveau gaat over de hoogst behaalde opleiding van personen. Deze is als volgt ingedeeld: basisonderwijs, vmbo, havo/vwo/mbo, hbo en wo. Ook de opleidingssector gaat over het hoogst behaalde opleidingsniveau. Deze is ingedeeld in vier sectoren: economie, techniek, zorg/sociaal-cultureel en overig. De leeftijd van het jongste kind geeft de leeftijd van het jongste kind in het huishouden. De huishoudsituatie geeft aan of de persoon samenwoont met een partner of niet, met de categorieën (1) partner (2) geen partner.

Werkloosheidstype kent de volgende categorieën: (1) baanwisselaar, (2) schoolverlaters, (3) korter dan één jaar geleden gestopt in de laatste baan, (4) één jaar tot twee jaar geleden gestopt in de laatste baan en (5) langer dan twee jaar geleden gestopt in de laatste baan. Baanwisselaar betreft personen die korter dan een jaar in de huidige werkkring werkzaam zijn en drie maanden of korter voor de aanvang van de huidige baan hun vorige werkkring beëindigden. Schoolverlaters zijn personen die korter dan een jaar geleden een reguliere opleiding hebben verlaten. De categorieën 3, 4 en 5 hebben betrekking op het moment dat een persoon gestopt is in de laatste baan van 12 uur of meer waarin langer dan één jaar werd gewerkt.

Om regionale verschillen in Nederland te laten zien, wordt gebruikt gemaakt van het rba-gebied (regionaal bestuur voor de arbeidsvoorziening). Deze heeft betrekking op het rba-gebied waarin een persoon woont. Er wordt onderscheid gemaakt tussen 18 verschillende rba-gebieden: Groningen, Friesland, Drenthe, IJssel/Vecht/Twenthe, IJssel/veluwe, Arnhem-oost/Gelderland/Rivland, Flevoland, Midden-nederland, Noord-Holland-noord, Rijnmond, zuidelijk Noord-

Holland, Rijnstreek, Haaglanden, Zeeland, Midden- en West-Brabant, Noord-Oost Brabant, Zuid-Oost Brabant en Limburg.

Daarnaast is er een variabele opgenomen in de analyse die aangeeft hoeveel uren een persoon wil werken, met de volgende categorieën: (1) persoon wil 0-11 uren werken (2) persoon wil 12-24 uren werken (3) persoon wil 25-32 uren werken (4) persoon wil 33 of meer uren werken. Tot slot zijn er partnergegevens toegevoegd aan de analyse. Er is een variabele opgenomen die aangeeft of de partner werkzaam is. Hierbij is de volgende indeling gemaakt. (1) partner werkt (12 uur of meer) (2) partner werkt niet (3) onbekend. Het hoogst behaalde opleidingsniveau van de partner is als volgt ingedeeld: laag, middelbaar en hoog. Tabel 3.1 t/m Tabel 3.2 geven een beschrijving van deze variabelen.

Tabel 3.1 Statistische beschrijving van de gebruikte modelvariabelen

	Percentage	N
Man	52	4288
Vrouw	48	3956
Leeftijd (enquêtemoment, gemiddelde)	35	8244
Autochtoon	80	6533
Westerse allochtoon	9	730
Niet-westerse allochtoon	12	981
Basisonderwijs	7	552
Vmbo	22	1846
Mbo/havo/vwo	43	3549
Hbo	17	1377
Wo	11	885
Opleidingssector		
Techniek	26	2117
Economie	21	1700
Zorg/sociaal-cultureel	26	2127
Overig	28	2300
Arbeidsmarktgroep		
Baanwisselaar	28	2160
Schoolverlater	12	889
> 1 jaar geleden gewerkt	18	1363
>=1 jaar geleden gewerkt<2 jaar	9	683
Overig	21	1587
Leeftijd jongste kind (gemiddelde)	2,5	8244
Huishoudsituatie		
Paar	74	6113
Geen paar	26	2131

Bron: CBS (EBB, 2005)

Tabel 3.2 (Vervolg) Statistische beschrijving van de gebruikte modelvariabelen

	Percentage	N
Woonplaats (Rba-gebied)		
Groningen	4	305
Friesland	4	297
Drenthe	3	266
IJssel/vecht/twenthe	7	540
IJssel/veluwe	4	375
Arnhem-Oost/gelderland/rivland	7	576
Flevoland	3	243
Midden-nederland	9	730
Noord-holland-noord	4	299
Rijnmond	8	701
Zuidelijk noord-holland	5	387
Rijnstreek	5	390
Haaglanden	11	913
Zeeland	2	194
Midden- en west-brabant	7	579
Noord-oost brabant	3	260
Zuid-oost brabant	6	478
Limburg	9	711
Aantal uren willen werken		
0-11 uur	4	346
12-24 uur	24	1783
25-32 uur	15	1110
>= 33 uur	61	4613
Partner werkzaam		
Partner werkzaam	45	3689
Partner niet werkzaam	13	1076
Geen partner/onbekend	42	3479
Opleidingsniveau partner		
Laag	15	1204
Middelbaar	25	2094
Hoog	17	1437
Geen partner/onbekend	43	3509
Baanvinders		
Baanvinders	60	4942
Werklozen	40	3302

Bron: CBS (EBB, 2005)

3.3 Plausibiliteitscontrole

De EBB kent meerdere peilingen. Dit betekent dat iedere persoon meerdere keren wordt ondervraagd. De eerste peiling is *face-to-face*. Daarna is er elke drie maanden een vervolgpeiling die telefonisch plaatsvindt. Voor iedere respondent is er dus informatie uit de eerste peiling (moment t) en uit vier vervolgpeilingen ($t+3$, $t+6$, $t+9$ en $t+12$). De modelschattingen van de werkloosheidsduur in dit rapport zijn gebaseerd op informatie uit de eerste peiling. De vervolgpeilingen bevatten vergelijkbare informatie, maar dan op een later tijdstip. Deze informatie is vooralsnog niet meegenomen in de modelschattingen. Dit zal waarschijnlijk wel in de toekomst gebeuren, waarbij eerst de kwaliteit van deze data wordt getoetst.

In deze paragraaf wordt de informatie uit de vervolgpeilingen gebruikt voor een plausibiliteitscontrole van de modelberekeningen. Dit is als volgt gedaan. Aan personen die in de eerste peiling werkloos zijn, wordt in de vervolgpeilingen gevraagd of ze werkzaam zijn in een baan van minimaal 12 uur per week. Indien dit het geval is, hebben deze personen een baan gevonden en is de werkloosheidsduur voltooid. Dit kan voor alle vier vervolgpeilingen worden bekeken. Op deze manier kan worden bepaald of werklozen drie, zes, negen of twaalf maanden na de eerste peiling een baan hebben gevonden, en welk deel van de werklozen meer dan 12 maanden nodig hebben om een baan te vinden. Dit kan worden vergeleken met de voorspellingen van het regressiemodel. Tabel 3.13 laat deze vergelijking zien. Voor de EBB vervolgpeiling geldt dat niet iedereen aan elke vervolgpeiling heeft meegedaan. Hier is als volgt mee omgegaan; wanneer voor iemand in een bepaalde vervolgpeiling een waarde ontbreekt, is de waarde van de vorige peiling genomen. Respons op de vervolgpeilingen is mogelijk selectief. Indien de welwillende en gemotiveerde mensen met grotere kans responderen, dan geven de EBB-vervolgpeilingen een te rooskleurig beeld van de uitstroom uit werkloosheid.

Tabel 3.3 **Vergelijking regressiemodel en EEB vervolgpeiling: aantal maanden dat werklozen na de eerste peiling nodig hebben om een baan te vinden**

	model %	EBB vervolgpeilingen %
Aantal maanden na eerste peiling		
0 -3 maanden	13,3	22,9
4 -6 maanden	12,4	11,3
7 -9 maanden	7,8	7,5
9 -11 maanden	6,6	4,9
>=12 maanden	59,9	53,4

Bron: EBB 2005; N = 1.980

Uit Tabel 3.13 blijkt dat de modelvoorspellingen slechts in beperkte mate verschillen van de EBB vervolgpeilingen. Vooral in de eerste drie maanden na de eerste peiling is er een duidelijk verschil. Het model voorspelt dat 13,3% van de werklozen binnen drie maanden een baan vindt. Bij de EBB vervolgpeiling ligt dit percentage met 22,9% beduidend hoger. Voor het overige geldt

dat de verschillen tussen het duurmodel en de vervolgpeilingen beperkt zijn. Een onderschatting van de uitstroom uit werkloosheid door het model mag overigens op voorhand worden verwacht. Wanneer in het Weibull-duurmodel niet alle relevante determinanten van de baanvindkans zijn opgenomen, vertaalt zich dat in een overschatting van de duurzaamheidsparameter, met als gevolg een onderschatting van de voorspelde uitstroom uit werkloosheid.

Geconcludeerd wordt dat het duurmodel in het eerste jaar na de eerste peiling een iets lagere uitstroom uit werkloosheid voorspelt (40,1%) in vergelijking met de resultaten uit de vervolgpeilingen van de EBB (46,6%). De waarheid ligt waarschijnlijk in het midden: het model onderschat mogelijk de uitstroom, de vervolgpeilingen geven mogelijk een te rooskleurig beeld van de uitstroom door selectieve respons.

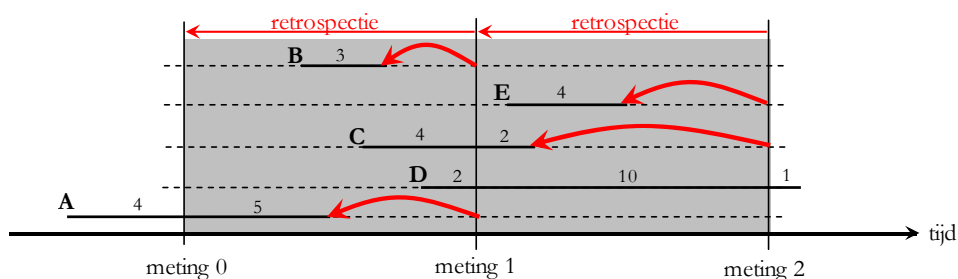
4 Ontwerp van een statistiek voor werkloosheidsduren

In dit hoofdstuk worden enkele duurstatistieken over werkloosheid ontworpen. Voordat we dat doen, gaan we eerst in paragraaf 4.1 in op de selectie en weging van waarnemingen die nodig zijn om over een representatieve steekproef van werkloosheidsduren te beschikken.

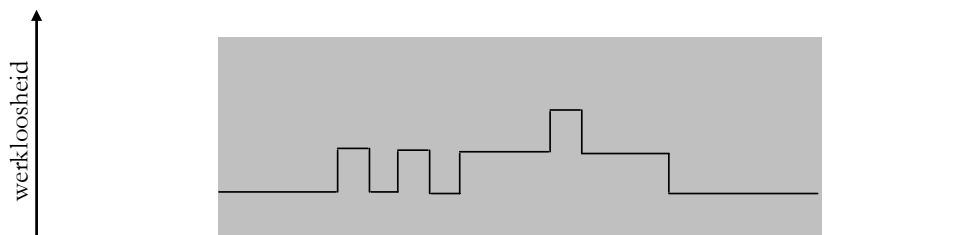
4.1 Selectie en weging

De EBB-gegevens kunnen worden gezien als een reeks van achtereenvolgende dwarsdoorsnedemetingen van de beroepsbevolking. De problemen die zich voordoen bij het meten van werkloosheidsduren door middel van zo'n reeks dwarsdoorsnedemetingen verduidelijken wij aan de hand van een simpel voorbeeld dat is weergegeven in Figuur 4.1. De periode van beschouwing is de grijs gearceerde periode. In die periode worden van een denkbeeldige populatie 5 individuen werkloos, aangeduid met de letters A t/m E. De 5 mensen doorlopen hun pad in de tijd langs een horizontale stippellijn van links naar rechts. Hun werkloosheidsduur doet zich voor als de stippellijn een ononderbroken lijn wordt. Drie keer wordt een (representatieve) dwarsdoorsnede meting verricht (metingen 0, 1 en 2), waarin informatie over het moment van meten alsmede retrospectieve informatie wordt vergaard. Zo wordt in meting 1 vastgesteld dat individuen C en D werkloos zijn op het moment van meten, en wordt van individuen A en B vastgesteld dat hun werkloosheidsduur in de periode tussen meting 0 en 1 werd beëindigd. De cijfers zijn lengte eenheden. Individu C is 6 tijdseenheden werkloos. Op het moment van meting 1 heeft hij 4 eenheden daarvan achter de rug; de resterende 2 zijn op dat moment nog ongewis. Pas bij meting 2 kan van C de voltooide duur worden waargenomen.

Figuur 4.1 Meting van werkloosheidsduren door middel van dwarsdoorsnede enquêtes



Figuur 4.2 De voorraad werklozen in de tijd



Figuur 4.2 geeft voor dit hypothetische voorbeeld grafisch de ontwikkeling van de werkloosheid in de tijd weer. De werkloosheid bedraagt 1 bij aanvang van de periode, daarna schommelt de werkloosheid enige tijd tussen 1 en 2. Kort na meting 1 bereikt de werkloosheid een hoogtepunt van 3. Daarna neemt zij weer af. Uit Figuur 4.2 blijkt dat uit een representatieve steekproef van werkloosheidsduren kan worden afgeleid hoe het aantal werklozen zich ontwikkelt in de tijd.

Bij gebruik van duurgegevens uit herhaalde dwarsdoorsnede metingen ligt het gevaar van selectiviteit op de loer. Anders gezegd: we beschouwen niet een representatieve steekproef van duren, maar een selectieve groep van (in dit geval) lange duren. Dat zou uiteraard leiden tot onzuivere statistieken en dat is niet wenselijk. We verduidelijken het selectiemechanisme aan de hand van het voorbeeld in Figuur 4.1. Stel, men zou op de drie meetmomenten de volledige populatie ondervragen. Duur A zou dan onvoltooid worden waargenomen in meting 0 en voltooid in meting 2. Voor duur C geldt het zelfde, maar dan telkens een meting later. Duren B en E worden precies één keer waargenomen: voltooid in respectievelijk meting 1 en 2. Duur D wordt twee keer onvoltooid waargenomen. Veronderstel dat we de onvoltooiden duren met behulp van een model perfect kunnen voorspellen. Als we nu van alle drie metingen de voltooide en onvoltooiden duren tot één steekproef voor de periode van beschouwing samenvoegen, dan worden duren A, C en D dubbel geteld. Niet toevallig zijn dat ook nog eens relatief lange duren, omdat juist die duren met grotere kans op een willekeurig moment als onvoltooid worden waargenomen. Om deze dubbeltelling te voorkomen moeten we alle onvoltooiden duren buiten beschouwing laten die in metingen tijdens de periode van beschouwing worden waargenomen. Alleen de onvoltooiden duren van de laatste meting dienen in de analyse te worden betrokken. Dit inzicht is van wezenlijk belang.

Kortom: voor een representatieve selectie van duren voor een periode waarin herhaaldelijk dwarsdoorsnede enquêtes zijn gehouden en waarin retrospectief informatie is verzameld, moet als volgt worden geselecteerd:

- 1) Selecteer alle duren die in de periode van beschouwing retrospectief en voltooid werden waargenomen, in Figuur 4.1 A en B in meting 1 en C en E in meting 2;
- 2) en tevens alle duren die onvoltooid werden waargenomen in de afsluitende meting: D in meting 2;
- 3) Laat buiten beschouwing de onvoltooiden duren in de metingen gedurende de periode, bijvoorbeeld in meting 1 wordt C onvoltooid waargenomen, maar deze duur wordt alléén meegenomen als voltooide duur in meting 2.

Wij werken dit principe uit voor het maken van jaarcijfers en kwartaalcijfers aan de hand van EBB-gegevens.

Jaarcijfers

Wanneer we aan de hand van EBB-gegevens een uitspraak willen doen over de werkloosheidsduur in een kalenderjaar, dan moet een representatieve selectie van waarnemingen van een enquêtejaar in beschouwing worden genomen. Tevens zal daarbij moeten worden herwogen. De selectie wordt als volgt gemaakt. We selecteren alle respondenten die op het moment van enquêteren werken in een baan die aanving in het jaar van beschouwing. Dit zijn baanvinders met de extra beperking dat zij in het huidige kalenderjaar met de baan begonnen. Het standaardgewicht dat aan EBB-respondenten wordt toegekend, wordt voor deze baanvinders aangepast door vermenigvuldiging met een factor. Er wordt herwogen omdat de eerste maand van het jaar veel vaker retrospectief wordt waargenomen, dan de laatste maand van het jaar. Die factor bedraagt $1/12$ voor respondenten die in januari met hun baan begonnen, $1/11$ als zij in februari begonnen, $1/10$ als dat in maart was, etc. Voor baanvinders in de maand december bedraagt de factor 1. Naast de geselecteerde baanvinders worden ook alle respondenten die in december worden waargenomen en op het moment van enquête tot de werkloze beroepsbevolking worden gerekend, eveneens met factor 1.

Kwartaalcijfers

Met de EBB-gegevens is ook mogelijk om meerdere malen per jaar uitspraken te doen over ontwikkelingen in de werkloosheidsduur. Zo kan bijvoorbeeld de werkloosheidsduur per kwartaal worden gepubliceerd. Van belang is daarbij dat de statistiek zo snel mogelijk na afsluiting van het kwartaal wordt gepubliceerd. De kwartaalstatistiek wordt gebaseerd op de EBB-maanden van het kwartaal. Voor het eerste kwartaal worden de respondenten geselecteerd die in de maanden januari, februari en maart een baan vonden. Het gewicht van de januari-baanvinders wordt gecorrigeerd met een factor $1/3$. Voor de februari-baanvinders geldt een factor van $1/2$ en voor die van maart een factor van 1. Naast de baanvinders worden de werklozen van de enquêtemaand maart in de analyse betrokken, eveneens met een gewichtsfactor van 1. In de overige kwartalen wordt analoog geselecteerd en herwogen. Tabel 4.1 geeft een overzicht.

Wij merken op dat in een kwartaalcijfer meer nadruk ligt op werklozen, dan in een jaarcijfer gebaseerd op een geheel enquêtejaar. Dit mechanisme is inherent aan het feit dat op een willekeurig moment van meten een lange werkloosheidsduur een grotere kans heeft te worden waargenomen, dan een korte werkloosheidsduur. Hoe korter de beschouwde periode, des te groter de nadruk op de werklozen en des te hoger de mediaan van de werkloosheidsduur. Maakt men de beschouwde periode zo extreem kort, bijvoorbeeld een dag, dan selecteren we in de statistiek in wezen alleen werklozen en geen enkele baanvinder. Een jaarcijfer gebaseerd op een geheel enquêtejaar resulteert dus onvermijdelijk in een ander cijfer dan wanneer men het gemiddelde neemt van de afzonderlijke kwartaalcijfers. De vraag is welke de voorkeur verdient. Kwartaalcijfers geven een recent beeld en leggen meer de nadruk op werklozen, terwijl in jaarcijfers meer gewicht wordt toegekend aan mensen die in dat jaar een baan vonden. We geven de voorkeur aan kwartaalcijfers. Aan de hand hiervan kan dan ook een jaarcijfer worden gepresenteerd door de vier kwartalen van het betreffende jaar te middelen.

Tabel 4.1 Kwartaalcijfers: weging en selectie van duren

	EBB-maanden	selectie van waarnemingen	factor
Kwartaal 1	januari	baanvinders in januari	1/3
	t/m	baanvinders in februari	1/2
	maart	baanvinders in maart	1
		werklozen op moment van enquêteren in maart	1
Kwartaal 2	april	baanvinders in april	1/3
	t/m	baanvinders in mei	1/2
	juni	baanvinders in juni	1
		werklozen op moment van enquêteren in juni	1
Kwartaal 3	juli	baanvinders in juli	1/3
	t/m	baanvinders in augustus	1/2
	september	baanvinders in september	1
		werklozen op moment van enquêteren in september	1
Kwartaal 4	oktober	baanvinders in oktober	1/3
	t/m	baanvinders in november	1/2
	december	baanvinders in december	1
		werklozen op moment van enquêteren in december	1

Baanwisselaars

Baanwisselaars zijn mensen die van de ene baan overgaan naar de andere. Het CBS definieert de baanwisselaar als volgt. Indien de huidige baan niet later dan drie maanden na het beëindigen van de vorige baan aanving, dan wordt de betreffende respondent gezien als een baanwisselaar. In onze definitie van de werkloosheidsduur wordt verondersteld dat banen altijd beginnen op de eerste dag van de maand en eindigen op de laatste dag van de maand. Baanwisselaars die in opeenvolgende maanden een baan beëindigen en met een baan beginnen, krijgen een werkloosheidsduur van nul. Deze baanwisselaars blijven buiten de beschouwing in de duurstatistiek, omdat hun transitie niet gepaard gaat met onvrijwillige werkloosheid. Ligt er een hele maand tussen de twee banen (bijvoorbeeld vorige eindigt in januari, nieuwe baan vangt aan in maart) dan heeft de baanwisselaar een werkloosheidsduur van één maand en wordt hij wel meegeteld in de duurstatistiek.

4.2 Duurstatistieken

Mogelijke statistieken

Centraal in dit onderzoek staat de ontwikkeling van duurstatistieken over werkloosheid in Nederland. We formuleren in deze paragraaf enkele alternatieven die aan de hand van het voorbeeld in Figuur 4.1 concreet worden gemaakt. In de figuur staan 5 werkloosheidsduren met een lengte van resp. 9, 3, 6, 13 en 4.

- Alternatief 1: Hoe lang waren de mensen die in de gearceerde periode werkloos waren, werkloos? We kijken dan naar alle mensen van wie een deel van de werkloosheidsduur in de beschouwde periode viel. Aan het feit dat een deel van hun werkloosheid mogelijk

buiten de beschouwde periode ligt (zowel er voor als er na), wordt voorbij gegaan. De gemiddelde werkloosheidsduur in het voorbeeld van Figuur 4.1 bedraagt dan 7 (=35:5).

- Alternatief 2: Hoe lang waren de mensen die in de gearceerde periode een baan vonden, werkloos? We kijken naar de duren die beëindigd werden in de beschouwde periode: A, B, C en E. De gemiddelde werkloosheidsduur bedraagt dan 5,5 (=22:4).
- Alternatief 3: Hoe lang waren de mensen die in de gearceerde periode werkloos waren, werkloos tot aan het einde van de beschouwde periode? We kijken naar dezelfde groep mensen als bij alternatief 1, maar de tijd die zij werkloos waren ná de periode waarover wordt gerapporteerd, blijft buiten beschouwing. De gemiddelde werkloosheidsduur in het voorbeeld van Figuur 4.1 bedraagt dan 6,8 (=34:5).
- Alternatief 4: Hoe lang waren de mensen die in de gearceerde periode werkloos waren, in die beschouwde periode werkloos? We kijken naar dezelfde groep mensen, maar de tijd die zij werkloos waren voor en na de periode waarover wordt gerapporteerd, blijft buiten beschouwing. De gemiddelde werkloosheidsduur in het voorbeeld bedraagt dan 6 (=30:5). Dit alternatief is interessant wanneer men een lange periode in beschouwing neemt.
- Alternatief 5: De cohortbenadering: hoe lang waren de mensen die in de gearceerde periode werkloos zijn geworden, werkloos? We volgen dan de mensen die tussen meting 0 en meting 2 werkloos zijn geworden tot ze een baan hebben gevonden. In figuur 4.1 zijn dat duren B, C, D en E. De gemiddelde werkloosheidsduur in het voorbeeld bedraagt dan 6,5 (=26:4).

Ook over langdurig werklozen kunnen statistieken worden ontwikkeld door alle duren boven een zekere grens (bijvoorbeeld 7 tijdseenheden in het voorbeeld van Figuur 4.1) te beschouwen. Wij denken aan de volgende statistieken:

- Alternatief 6: Hoe lang waren de langdurig werklozen die in de gearceerde periode werkloos waren, werkloos? In het voorbeeld van Figuur 4.1 zijn dat duren A en D. Het gemiddelde daarvan bedraagt 11.
- Alternatief 7: Welk percentage van de werklozen in de beschouwde periode was langdurig werkloos? In het voorbeeld van Figuur 4.1 bedraagt dat percentage 40%.

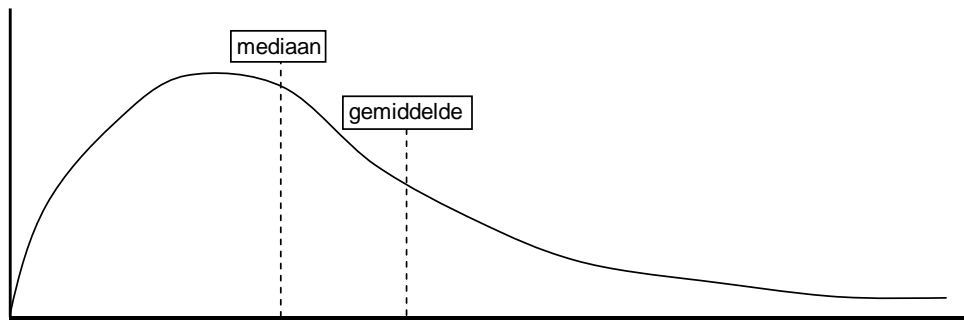
Wij concentreren ons in dit onderzoek op uitwerking van alternatief 1. Voor de berekening van Alternatief 1 is het noodzakelijk dat een schatting wordt gemaakt van het restant van de duur van waarneming D. Merk voorts op dat de keuze van alternatief bepalend is voor de uitkomst van de statistiek. Wellicht kunnen de andere alternatieven in een vervolgonderzoek aan bod komen.

De mediaan versus het gemiddelde

Duurgrootheden volgen in de praktijk vaak een scheve verdeling met een lange rechterstaart voor de extreem lange duren. Men vindt in dergelijke verdelingen veel meer massa onder het gemiddelde dan erboven. Dit doet zich bijvoorbeeld ook voor bij statistieken over beloning (uur- en maandloon) en vermogen: extreem hoge waarden trekken daar het gemiddelde omhoog, waardoor de meerderheid van de mensen een waarde onder het gemiddelde hebben. In statistieken over beloning en vermogen wordt dan ook veelal de mediaan in plaats van het gemiddelde gebruikt.

In Figuur 4.3 worden de mediaan en het gemiddelde van een scheve verdeling grafisch weergegeven. De mediaan verdeelt de massa onder grafiek in twee gelijke delen van 50%. Bij het gemiddelde (de verwachting) worden alle mogelijke uitkomsten gewogen met de kans dat zij voorkomen. De mediaan ligt bij dergelijke scheve verdelingen onder het gemiddelde.

Figuur 4.3 De mediaan en het gemiddelde van een scheve verdeling



Voor de werkloosheidsduur is de mediaan om de volgende redenen de meest geschikte statistiek om te publiceren. Ten eerste is deze statistiek minder gevoelig voor extreem lange waarnemingen dan het gemiddelde. Tevens blijft het effect van eventuele overschattingen in de modelvoorspellingen daardoor beperkt. Ten tweede blijkt de mediaan in de praktijk meer aan te sluiten bij wat mensen als een plausibel gemiddelde van de werkloosheidsduur ervaart. Een mogelijke verklaring hiervoor is dat men het gemiddelde als een symmetrisch concept beschouwt, waarin een bovengemiddelde waarneming wordt gecompenseerd door een waarneming precies even ver onder het gemiddelde. Bij de scheve verdeling van de werkloosheidsduur met afkapping bij nul, kan die compensatie van lange duren alleen maar plaatsvinden door massaverhouding onder en boven het gemiddelde.

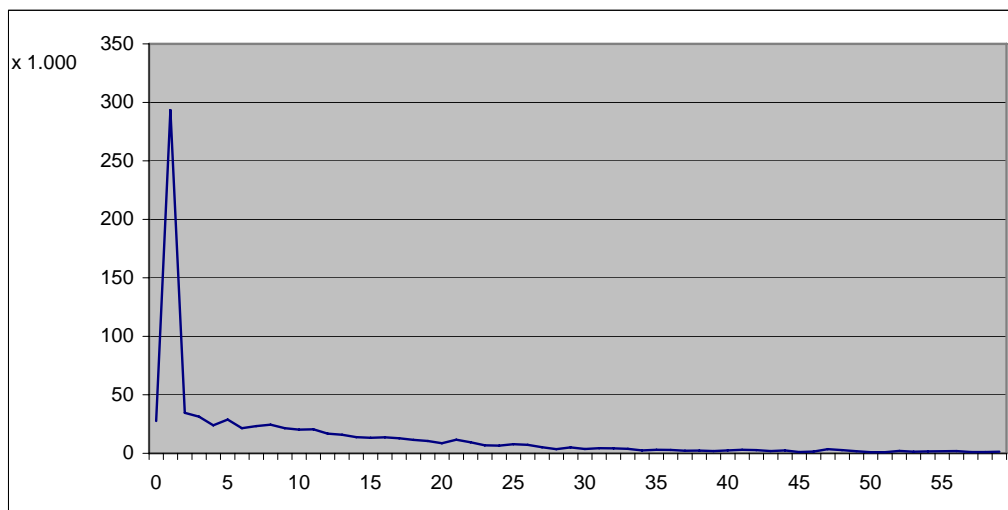
5 De werkloosheidsduur in 2002-2005

In deze paragraaf worden resultaten gepresenteerd van de voltooide werkloosheidsduur. In paragraaf 5.1 wordt de frequentieverdeling van de voltooide werkloosheidsduur voor het jaar 2005 beschreven. Vervolgens wordt ingegaan op trendcijfers van de werkloosheidsduur voor de periode 2002-2005. De derde paragraaf laat een uitsplitsing zien naar de achtergrondkenmerken geslacht, leeftijd, etniciteit en opleiding. Zoals eerder aangegeven, hebben de cijfers in dit rapport een voorlopig karakter en deze kunnen in een later stadium worden bijgesteld.

5.1 Frequentieverdeling, mediaan en gemiddelde

In Figuur 5.1 wordt de frequentieverdeling van de voltooide werkloosheidsduur van baanvinders en werklozen grafisch weergegeven voor zover de duur korter is dan 60 maanden. Bijna 18% (180 duizend mensen in 2005) hadden een werkloosheidsduur van langer dan 60 maanden. Uit de figuur blijkt duidelijk de scheefheid van de frequentieverdeling van de werkloosheidsduur. De meeste personen (bijna 30% van het totaal) hebben een werkloosheidsduur van rond de één maand. De overige ruim 70% heeft een werkloosheidsduur van 2 of meer maanden. Het gemiddelde ligt ongeveer bij 40 maanden, de mediaan bij 8 maanden. We stellen voor om in het vervolg gebruik te maken van de mediaan bij het maken van statistieken over de werkloosheidsduur van personen in Nederland (zie ook uitleg bij Figuur 4.3). Het gemiddelde van 40 maanden is voor het gevoel erg hoog en dat komt door een relatief klein aantal extreem lange duren. Door de mediaan te kiezen als statistiek, blijven deze extremen tellen als massa maar heeft de waarde van de extremen geen effect op de statistiek.

Figuur 5.1 Voltooide werkloosheidsduur van baanvinders en werklozen in maanden, 2005



Bron: CBS (EBB 2005)

5.2 Trendcijfers 2002-2005

In hoofdstuk 4 werd een aantal alternatieven besproken van statistieken over de werkloosheidsduur. In deze paragraaf beschrijven we jaar- en kwartaalcijfers voor de jaren 2002-2005. Tabel 5.1 toont jaarcijfers van de werkloosheidsduur voor de periode 2002-2005. Ook is aangegeven welk deel van de personen 0-5, 6-11, 12-17, 18-23 en 24 maanden of meer werkloos is. Hieruit komt naar voren dat in de jaren 2002 de mediaan van de werkloosheidsduur relatief laag is. Daarna neemt deze geleidelijk toe tot 2004 en 2005. Hetzelfde beeld komt naar voren uit het aandeel personen dat binnen 12 maanden een baan heeft gevonden. In 2004 en 2005 heeft om en nabij 57% van de personen een werkloosheidsduur van minder dan 12 maanden. In 2002 en 2003 heeft respectievelijk 73% en 68% van de personen een baan gevonden binnen 12 maanden. Bovenstaande cijfers komen overeen met gegevens over het percentage werklozen in Nederland dat vanaf 2002 toeneemt.

Tabel 5.1 De voltooide werkloosheidsduur 2002-2005, jaarcijfers

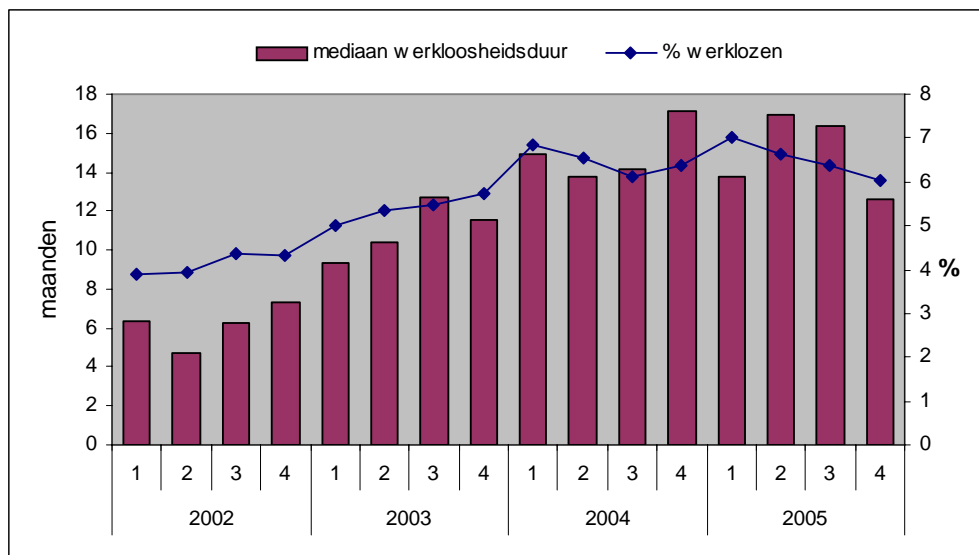
Jaar	Mediaan (maanden)	0-5 maanden %	6 - 11 maanden %	12 - 17 maanden %	18 - 23 maanden %	>= 24 maanden %
2002	1,7	62	11	6	3	18
2003	5,0	53	13	8	5	21
2004	8,2	45	12	9	5	28
2005	8,2	45	13	8	6	28

Bron: CBS (EBB 2002-2005)

Zoals in hoofdstuk 4 is vermeld kan de werkloosheidsduur ook per kwartaal worden berekend. Hiervoor zijn personen die in de eerste, tweede en derde maand van het kwartaal een baan hebben gevonden samen genomen met personen die in de derde maand van het kwartaal werkloos waren. In Figuur 5.2 is de mediaan van de werkloosheidsduur per kwartaal weergegeven voor de periode 2002-2005. Ook hieruit blijkt dat met name in de jaren 2004 en 2005 de werkloosheidsduur relatief hoog is. Bovendien laat deze figuur zien dat de werkloosheidsduur sterk is afgenomen in het laatste kwartaal van 2005. Die daling is vergelijkbaar met de daling in de kwartaalcijfers van het percentage werklozen.

Voorts laat deze figuur hier en daar een grillig verloop zien. Nadere inspectie van de figuur leert dat binnen de jaren geen eenduidig seizoenpatroon in het oog springt. Seizoensinvloeden als mogelijke verklaring hiervoor zijn dus niet waarschijnlijk. Nader onderzoek moet uitwijzen in hoeverre het feit dat de kwartaalcijfers zijn gebaseerd op vrij kleine aantallen (ongeveer 500 personen per kwartaal) hieraan ten grondslag ligt. Het is derhalve van belang om bij de bepaling van de werkloosheidsduur gebruik te maken van telefonische data uit de EBB. Hierdoor kan het aantal personen per kwartaal aanzienlijk worden vergroot, waardoor eventuele ruis in de data minder invloed heeft op de uiteindelijke resultaten.

Figuur 5.2 Mediaan werkloosheidsduur en percentage werklozen per kwartaal, 2002-2005



Bron: CBS (EBB 2002-2005)

Tabel 5.2 De voltioede werkloosheidsduur 2002-2005 op basis van kwartaalcijfers, mediaan en in klassen van 5 maanden

Jaar	Mediaan (maanden)	0-5	6 - 11	12 - 17	18 - 23	>= 24
		maanden %	maanden %	maanden %	maanden %	maanden %
2002	6,2	49	14	8	5	24
2003	10,8	37	16	11	6	30
2004	14,7	30	14	11	6	39
2005	15,3	31	13	11	6	38

Bron: CBS (EBB 2002-2005)

Met behulp van de kwartaalcijfers kan ook een jaargemiddelde worden bepaald door de cijfers van de kwartalen van een jaar samen te nemen (zie Tabel 5.2). Het jaarcijfer gebaseerd op de vier kwartalen is hoger dan de jaarcijfers in Tabel 5.1. Dit komt omdat bij de bepaling van het jaarcijfer in Figuur 5.2 vier maal gebruik wordt gemaakt van gegevens van werklozen, namelijk werklozen die aan het einde van elk kwartaal werkloos zijn. Bij de bepaling van het jaarcijfer in Tabel 5.1 is dat slechts éénmaal: de werklozen die in december van het jaar werkloos zijn. De werkloosheidsduur met behulp van de kwartaalcijfers ligt dan hoger omdat de mediaan van de werkloosheidsduur van werklozen doorgaans hoger ligt dan die van de baanvinders. Merk op dat dit mechanisme inherent is aan het feit dat op een willekeurig moment van meten een lange werkloosheidsduur een grotere kans heeft te worden waargenomen, dan een korte werkloosheidsduur. Hoe korter de beschouwde periode, des hoger de mediaan van de werkloosheidsduur van de mensen die in die periode werkloos waren. Kortom: kwartaalcijfers geven een recent beelden en leggen meer de nadruk op werklozen, terwijl in jaarcijfers meer gewicht wordt toegekend aan mensen die in dat jaar een baan vonden. Een jaarcijfer gebaseerd op een geheel enquêtejaar resulteert dus per definitie in een ander cijfer dan wanneer men het gemiddelde

neemt van de afzonderlijke kwartaalcijfers. Zoals in hoofdstuk 4 is aangeven, verdient het jaarcijfer gebaseerd op kwartaalcijfers onze voorkeur.

5.3 Werkloosheidsduur en achtergrondkenmerken

Tot slot worden in deze paragraaf voor een aantal achtergrondkenmerken cijfers gepresenteerd voor het jaar 2005. Beschreven wordt de werkloosheidsduur naar geslacht, herkomst, leeftijd en opleidingsniveau. De cijfers voor 2005 zijn gebaseerd op het kwartaalcijfer. In Tabel 5.3 is de mediaan van de werkloosheidsduur weergegeven naar verschillende achtergrondkenmerken. Ook is het aandeel personen vermeld dat 0-5, 6-11, 12-17, 18-23 en 24 maanden of meer werkloos is. Nota bene: de werkloosheidsduur voor het jaar 2005 is relatief hoog in vergelijking met de jaren 2002 en 2003. Waarschijnlijk bevindt de werkloosheidsduur zich in 2005 op het hoogste niveau en gaat deze vanaf 2006 weer dalen.

Tabel 5.3 Statistieken voltooide werkloosheidsduur naar geslacht, herkomst, leeftijd en opleidingsniveau, mediaan en in klassen van 5 maanden

	Mediaan (maanden)	0-5 maanden %	6 - 11 maanden %	12 - 17 maanden %	18 - 23 maanden %	>= 24 maanden %
Geslacht						
Man	14,1	33	14	11	7	35
Vrouw	16,3	29	13	11	6	42
Etniciteit						
Autochtoon	12,7	37	12	11	7	34
Westerse allochtoon	16,4	22	18	14	8	38
Niet-westerse allochtoon	26,1	17	15	8	5	54
Leeftijd						
15-24 jaar	7,7	43	21	11	4	21
25-34 jaar	9,9	39	15	10	6	30
35-44 jaar	17,4	28	12	12	7	41
45-54 jaar	25,6	20	7	12	8	53
55-64 jaar	35,3	9	9	7	11	64
Opleiding						
Basisonderwijs	38,2	16	7	10	7	60
VMBO	16,5	27	15	10	6	42
MBO/HAVO/VWO	14,8	31	14	11	7	37
HBO	8,1	46	12	12	4	26
WO	10,2	38	16	11	7	28
Totaal	15,3	31	13	11	6	38

Bron: CBS (EBB 2005)

Uit Tabel 5.3 blijkt dat mannen en vrouwen iets verschillen wat betreft de werkloosheidsduur. Mannen doen er ongeveer 14 maanden over om een baan te vinden; vrouwen ongeveer 16 maanden. Daarbij geldt dat vrouwen iets vaker dan mannen 24 maanden of meer werkloos zijn

geweest. Grote verschillen vinden we echter bij herkomst, opleidingsniveau en vooral leeftijd. Jongeren vinden relatief snel een baan, terwijl ouderen hier lang over doen. De mediaan van de werkloosheidsduur loopt op van 7,7 maanden bij 15-24 jarigen tot 35,3 maanden bij 55-64 jarigen. Verder komt naar voren dat vooral personen met alléén basisonderwijs lang nodig hebben om een baan te vinden. Hierin speelt de leeftijd echter ook een grote rol. Vooral veel ouderen hebben als hoogst behaalde opleidingsniveau het basisonderwijs; onder jongeren komt dit vrij weinig voor. Natuurlijk is het ook mogelijk om leeftijd en opleidingsniveau met elkaar te kruisen. Op deze manier wordt dan duidelijk dat vooral veel jongeren met een hoog opleidingsniveau (hbo en wo) relatief snel een baan hebben gevonden. Tot slot geldt dat voor niet-westerse allochtonen de mediaan van de werkloosheidsduur hoog is (26,1 maanden voor niet westerse allochtonen; tegen 12,7 maanden voor autochtonen). De uitsplitsing van de werkloosheidsduur naar de diverse achtergrondkenmerken levert een vertrouwd beeld op. Vrouwen zoeken langer dan mannen, allochtonen langer dan autochtonen, ouderen langer dan jongeren en laag opgeleiden langer dan hoog opgeleiden.

De uitkomsten van dit hoofdstuk laten zien dat het voltooiën van de werkloosheidsduren op twee manieren van invloed is geweest op de uiteindelijke resultaten. Ten eerste is de werkloosheidsduur van werklozen toegenomen na het voltooiën van de duren. Wanneer wordt gekeken naar de onvoltooide werkloosheidsduur van werklozen dan is de mediaan 8,0 maanden in 2005. Zoals we hebben gezien, is de voltooide werkloosheidsduur 15,3 maanden in 2005 (op basis van kwartaalcijfers). Ten tweede zijn na het voltooiën van de duren de verhoudingen binnen groepen enigszins veranderd. Verschillen in werkloosheidsduur die we vonden bij de onvoltooide werkloosheidsduur tussen jongeren en ouderen, autochtonen en niet-westerse allochtonen en de laagst opgeleiden en hoger opgeleiden zijn na voltooiing van de duren toegenomen. Zo bezien geeft de voltooiing van de duren een wat somberder beeld van de werkgelegenheidskansen van ouderen, niet-westerse allochtonen en de laagst opgeleiden.

6 Conclusies en aanbevelingen

De resultaten van het onderzoek laten zien dat het mogelijk is om een statistiek voor werkloosheidsduren te ontwikkelen op basis van gegevens uit de Enquête Beroepsbevolking (EBB). De informatie in de EBB is een goede basis om een model te schatten op basis van de baanvinders en deze toe te passen op de werklozen om hun werkloosheidsduur te voltooien. Vervolgens is het mogelijk om hiermee statistieken te maken over werkloosheidsduren. Aan de hand daarvan kan een schatting worden gemaakt van het aantal langdurig werklozen. Tevens is het mogelijk om een beeld te geven van de tijd die een werkzoekende nodig heeft om een baan te vinden. Vanwege de aanwezigheid van grote extreme waarden is het verstandig om de mediaan daarvoor te gebruiken in plaats van het gemiddelde.

Een belangrijke functie van een door het CBS te publiceren duurstatistiek is dat ermee een beeld wordt gegeven van de toestand op de arbeidsmarkt in een bepaalde periode. Een dergelijk cijfer moet snel na afsluiting van de periode kunnen worden gepubliceerd. De keuze van de lengte van de periode heeft invloed op de statistiek. Hoe korter de periode, hoe meer de nadruk komt te liggen op mensen die werkloos worden waargenomen op het moment van enquêteren. Hoe langer de periode, hoe meer de nadruk ligt op mensen die in de betreffende periode een baan vonden. Het gevolg hiervan is dat bij een lange periode van beschouwing de mediaan lager ligt dan bij een korte periode. Dit komt voort uit het feit dat mensen met een lange werkloosheidsduur een grotere kans hebben om als werkloze te worden waargenomen op het moment van enquêteren, dan mensen met een korte werkloosheidsduur. Ten aanzien van de rapportageperiode wordt voorgesteld om te kiezen voor het kwartaal. De statistiek doet een uitspraak over de leden van de beroepsbevolking waarvan de werkloosheidsduur (voor een deel) in het betreffende kwartaal viel. Het betreft dus enerzijds mensen die in het betreffende kwartaal een baan vonden, en anderzijds mensen die aan het einde van het kwartaal nog werkloos waren ongeacht wanneer de werkloosheidsduur begon. Kortom: we geven de voorkeur aan kwartaalcijfers; een jaarcijfer kan vervolgens worden gepresenteerd door de vier kwartalen van het betreffende jaar te middelen.

Bij de selectie van waarnemingen moet ook een keuze worden gemaakt met betrekking tot de baanwisselaars. Dit zijn mensen die niet later dan drie maanden na de vorige baan in de huidige baan begonnen. Weglaten van deze mensen houdt in dat een relatief grote groep van korte werkloosheidsduren buiten beschouwing blijft, met uiteraard gevolgen voor de uitkomst van de statistiek. Mensen die van baan veranderen zonder daarbij werkloos te zijn geweest horen niet thuis in een statistiek die onvrijwillige werkloosheid wil meten. Aan de andere kant is de periode van drie maanden in de definitie van baanwisselaars behoorlijk lang. Hierdoor worden bijvoorbeeld mensen van wie het contract afliep maar binnen enkele maanden weer een nieuwe baan vonden, ten onrechte aangezien voor baanwisselaars. Gekozen is om een middenweg te bewandelen. Alleen de baanwisselaars met een werkloosheidsduur van minder dan één maand worden niet meegenomen bij het berekenen van de werkloosheidsduur.

De statistieken die nu geproduceerd kunnen worden, laten een ander beeld zien van de verschillen in werkloosheidsduur dan zichtbaar was op basis van onvoltooide werkloosheidsduren van werklozen. Natuurlijk daalt na het voltooien van werkloosheidsduren het aandeel

personen dat minder dan 12 maanden werkloos is (geweest). Daarbij zijn de verschillen die er al waren naar leeftijd en opleidingsniveau nog iets toegenomen; jongeren vinden snel een baan, terwijl ouderen hier relatief lang over doen. Hetzelfde geldt voor hoog en laag opgeleide personen.

De gebruikte informatie uit de EBB is nog niet optimaal. Zo is een zeer ongewenste omissie dat van de baanvinders niet bekend is wanneer zij begonnen met zoeken naar hun huidige baan. Hierdoor is de meting van de werkloosheidsduur van mensen die na een periode van inactiviteit weer zijn gaan zoeken van mindere kwaliteit. Dit betreft met name herintreders en arbeidsongeschikten. Meting van het zoekmoment onder baanvinders maakt de werkloosheidsduurstatistiek veel betrouwbaarder voor deze vanuit beleidsoogpunt zeer belangrijke arbeidsmarktgroepen. Voorts is de duur gemeten in maanden. Omdat echter niet bekend is op welke dag iemand begint met werken en op welke dag in de maand de vorige activiteit is beëindigd, is het lastig om deze maanden exact te bepalen. Daarom is aangenomen dat de vorige activiteit aan het einde van de maand plaatsvond en de baan aan het begin van de maand. Onderzocht zou moeten worden of deze aanname gegrond is. Aanbevolen wordt om deze informatie te verzamelen. Dit zou eventueel kunnen met behulp van registraties. Tenslotte is het van belang om bij baanwisselaars die niet direct van de ene baan op de andere overstappen, te onderzoeken of er een periode van werkloosheid tussen de ene en de andere baan zat. Hierover zou in de EBB informatie kunnen worden verzameld.

Het onderzoek draagt een aantal onderwerpen voor vervolgonderzoek voor. Zo is tijdens dit project alleen nog maar gebruik gemaakt van de informatie uit de eerste, *face-to-face*, peiling van de EBB. De vervolgpeilingen per telefoon zijn niet gebruikt voor het opstellen van de duurstatistiek, wel voor een plausibiliteitscontrole (zie paragraaf 3.3). Gebruik van de telefonische data zal de betrouwbaarheid van de statistiek verhogen, door een toename van het aantal waarnemingen. Een tweede punt dat onderzocht zou moeten worden is de situatie dat iemand zich na een periode van werkloosheid terugtrekt van de arbeidsmarkt. Vooralsnog is aangenomen dat elke persoon die op het moment van ondervragen werkloos is, blijft zoeken naar een baan. Bij oudere werknemers en ontmoedigden is dit wellicht niet correct. Hierdoor ontstaan mogelijk heel lange werkloosheidsduren, terwijl de betreffende respondenten helemaal niet meer zoeken. Om dit te ondervangen kan een ander duurmodel worden toegepast (*competing risks* model), zodat het mogelijk wordt om vanuit de transitie over te gaan naar twee mogelijke toestanden: werk en inactiviteit. Ten derde is binnen dit project niet meegenomen dat informatie uit registers, in het bijzonder ingeschreven staan bij het CWI of een uitkering hebben, gebruikt kan worden om de werkloosheidsduur beter af te bakenen. Dit zal ook onderzocht moeten worden.

Bijlage 1: Bepaling werkloosheidsduur

Operationalisering

In deze bijlage worden de variabelen beschreven die gebruikt zijn bij de bepaling van de werkloosheidsduur van werklozen en baanvinders. Eerst komen drie variabelen aan bod die hierbij centraal staan; datum van de laatste baan, datum van zoeken naar werk en datum van schoolverlaten. De gegevens die worden gebruikt zijn afkomstig van de Enquête Beroepsbevolking (EBB), jaargangen 2002-2005. De EBB betreft een onderzoek die maandelijks wordt gehouden onder een steekproef van personen van 15 jaar of ouder in Nederland, met uitzondering van personen in inrichtingen, instellingen en tehuizen. De steekproef heeft een omvang van ongeveer 1 procent van de Nederlandse bevolking. De respondenten worden gevraagd naar verschillende aspecten van het werk, arbeidsverleden, het zoeken naar werk, het gevolgde onderwijs. Bovendien wordt naar achtergrondkenmerken als leeftijd, herkomst, huishoudsituatie en dergelijke gevraagd. Uit de EBB zijn personen geselecteerd in de leeftijd 15-64 jaar.

Werklozen betreffen personen die op het moment van de enquête behoren tot de werkloze beroepsbevolking. De werkloze beroepsbevolking bestaat uit in Nederland wonende mensen van 15-64 jaar zonder werk (of met werk van minder dan 12 uur per week) die actief zoeken naar een baan van 12 uur of meer per week en daarvoor ook beschikbaar zijn.

Baanvinders zijn personen die deel uitmaken van de werkzame beroepsbevolking (15-64 jaar) en minder dan een jaar geleden in hun huidige baan van 12 uur of meer per week begonnen zijn.

Scholieren/studenten worden niet opgenomen in de uiteindelijke analyse. Scholieren/studenten worden beschouwd als personen met als belangrijkste bezigheid het volgen van onderwijs. Zo bezien zijn scholieren/studenten - in tegenstelling tot bijvoorbeeld schoolverlaters die zoeken naar werk - niet 'in-transitie' tussen twee toestanden. Bovendien gaat het bij het overgrote deel van de scholieren/studenten die werk zoeken om een vakantie- of bijbaantje.

De onvoltooide werkloosheidsduur

Voor werklozen is eerst de datum bepaald wanneer iemand werkloos is geworden (de startdatum van werkloosheid). Door hieraan de datum van de enquête te koppelen kan voor elke persoon de onvoltooide werkloosheidsduur worden berekend. Deze onvoltooide werkloosheidsduur is het aantal maanden tussen de startdatum van werkloosheid tot en met de datum van de enquête. Uit de EBB kunnen drie mogelijke data worden gereconstrueerd voor de bepaling van deze startdatum van werkloosheid; stopdatum, zoekdatum en schoolverlatersdatum (svdatum). Hieronder worden deze drie data eerst kort beschreven. Daarna wordt de constructie van de onvoltooide werkloosheidsduur van werklozen en een aantal resultaten voor het jaar 2005 besproken.

Stopdatum

Stopdatum betreft de datum (jaar en maand) wanneer personen zijn gestopt in de laatste baan van 12 uur of meer, waarin ze langer dan één jaar hebben gewerkt. Deze is als volgt gereconstrueerd. Aan iedereen van 18 jaar of ouder die op het moment van de enquête aangeeft geen werk te hebben, of werk te hebben van minder dan 12 uur per week én aan iedereen van 18 jaar¹ of ouder die korter dan 1 jaar geleden is begonnen in de huidige werkkring wordt de volgende vraag gesteld: heeft u ooit een betaalde werkkring gehad van 12 uur of meer per week, voor langer dan 1 jaar? Aan degenen die hier “ja” op hebben geantwoord is vervolgens gevraagd in welk jaar en in welke maand ze gestopt zijn met werken of minder dan 12 uur per week zijn gaan werken. Hiermee kan het jaar en de maand worden bepaald van de laatste baan (van 12 uur per week, voor langer dan 1 jaar). Aan personen die langer dan drie jaar geleden zijn gestopt in de laatste baan wordt *de maand* waarin men is gestopt niet meer gevraagd. Van deze personen is alléén het jaar bekend. Van 78% van de werklozen is de stopdatum bekend. De onbekenden betreft vooral jongeren die nog op school zitten en nog niet eerder een baan van 12 uur of meer per week hebben gehad.

Zoekdatum

De zoekdatum betreft de datum (jaar en maand) wanneer men is gaan zoeken naar werk voor tenminste 12 uur per week of meer. In de EBB is aan iedereen die de afgelopen vier weken heeft gezocht naar werk gevraagd hoe lang men op zoek is naar werk. Dit kan worden aangegeven in jaren en in maanden. Aan de hand hiervan kan het moment worden gereconstrueerd wanneer iemand op zoek is gegaan naar werk.

Opmerkelijk is echter dat het merendeel van de personen bij de beantwoording van de vraag “hoe lang ze op zoek zijn naar werk” afronden op halve en hele jaren. Dit betekent bijvoorbeeld dat veel personen die in werkelijkheid vijf maanden hebben gezocht, zeggen zes maanden op zoek te zijn naar werk en personen die elf of dertien maanden zoeken naar werk geven aan twaalf maanden op zoek te zijn. De datum van zoeken naar werk geeft dus een vrij ruwe schatting van de werkelijkheid. Van 87% van de werklozen is de zoekdatum bekend.

Schoolverlatersdatum

De datum van schoolverlaten (svdatum) is als volgt bepaald. Aan iedereen is gevraagd of ze na de basisschool of lagere school nog een opleiding of cursus hebben gevolgd van 6 maanden of meer. Vervolgens is van deze opleiding onder meer de naam, soort, richting, duur, gemiddeld aantal lessen en begin- en einddatum (jaar en maand) gevraagd. Aan de hand hiervan kan worden achterhaald of de opleiding een reguliere of niet-reguliere opleiding is. Dit is gedaan voor maximaal zes opleidingen die men na de basisschool of lagere school kan hebben gevolgd. Een reguliere opleiding is een voltijd- of deeltijdopleiding in het vmbo, havo, vwo, mbo (bol en bbl), hbo of wo. Wanneer de naam van de gevolgde opleiding onbekend is worden opleidingen die langer duren dan 1 jaar gerekend tot reguliere opleidingen.

¹ Deze vraag is dus niet gesteld aan personen jonger dan 18 jaar. Voor de bepaling van de startmaand van werkloosheid levert dit waarschijnlijk weinig problemen op. Het overgrote deel van de 15-,16- en 17-jarigen volgt nog een opleiding (95%) en slechts een klein deel is werkloos (3%).

Voor de bepaling van de datum van schoolverlaten worden alléén de reguliere opleidingen meegenomen. De datum van schoolverlaten betreft dan de einddatum van de meest recent gevolgde reguliere opleiding. Van personen die langer dan drie jaar geleden de opleiding hebben verlaten is de maand van het verlaten van de opleiding niet meer gevraagd. Alléén het jaar is hiervan bekend. De maand is op augustus gezet omdat dit de meest voorkomende maand is waarin leerlingen/studenten een opleiding verlaten. Vanzelfsprekend hebben personen die op het moment van de enquête een reguliere opleiding volgen geen schoolverlatersdatum. Van 85% van de werklozen is de svdatum bekend.

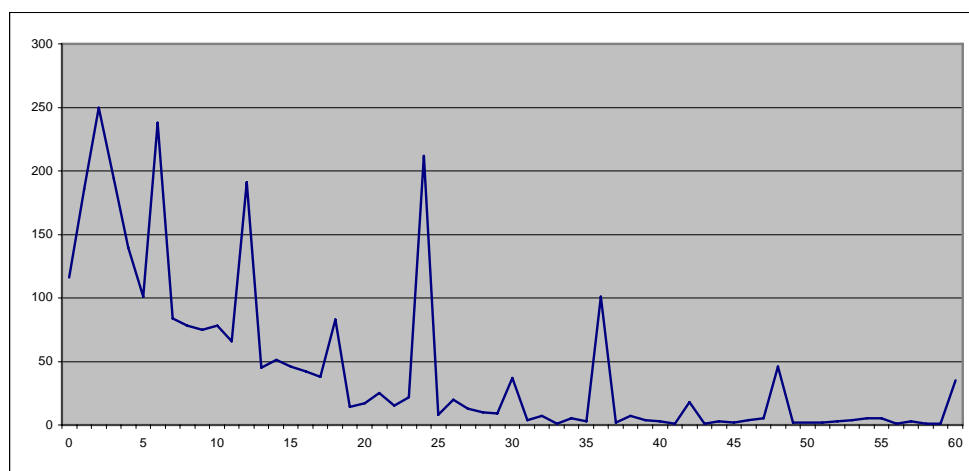
Bepaling van de onvoltooide werkloosheidsduur

Als startdatum van werkloosheid wordt de meest recente datum van de bovengenoemde drie data genomen. Wanneer deze startdatum van werkloosheid is bepaald, wordt deze gekoppeld aan de datum van de enquête en kan de onvoltooide werkloosheidsduur (in maanden) worden berekend. In totaal is er in de periode 2002-2005 van 97% van de werklozen de onvoltooide werkloosheidsduur bekend. Dit zijn in totaal 11.535 werklozen.

Figuur 6.1 laat de onvoltooide werkloosheidsduur van werklozen zien voor het jaar 2005. Vermeld is het aantal werklozen dat 1,2, 3 maanden et. cetera werkloos is, tot een maximum van 60 maanden. Uit de figuur blijkt dat relatief veel werklozen kort werkloos zijn (1 à 2 maanden). Daarna daalt het aantal werklozen snel met de toename van de werkloosheidsduur. Wat verder opvalt aan deze figuur zijn de pieken bij een werkloosheidsduur van 6, 12, 18 maanden et. cetera. De onvoltooide werkloosheidsduur van werklozen geeft zodoende enigszins een ruwe weergave van de werkelijkheid. Deze pieken kunnen met behulp van modelberekeningen worden verwijderd (zie ook paragraaf 3.1.3). Figuur 6.2 laat de onvoltooide werkloosheidsduur van werklozen zien na het gladstrijken van de pieken.

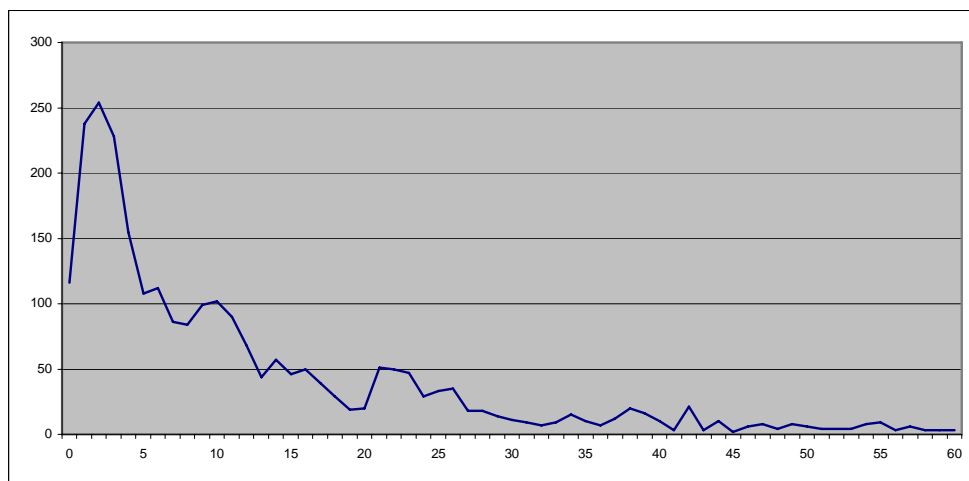
In Tabel 6.1 is de onvoltooide werkloosheidsduur in klassen van 5 maanden vermeld, uitgesplitst naar een aantal achtergrondkenmerken.

Figuur 6.1 Onvoltooide werkloosheidsduur van werklozen in maanden



Bron: CBS (EBB, 2005)

Figuur 6.2 Onvoltooide werkloosheidsduur van werklozen in maanden na gladstrijken



Bron: CBS (EBB, 2005)

Tabel 6.1 Onvoltooide werkloosheidsduur van werklozen, in klassen van 5 maanden

	0 - 5 maanden	6 - 11 maanden	12 - 17 maanden	18 - 23 maanden	> 24 maanden	Totaal
Geslacht						
Man	35	20	14	7	24	100
Vrouw	38	22	13	5	23	100
Etniciteit						
Autochtoon	38	21	13	6	22	100
Westerse allochtoon	32	21	16	7	24	100
Niet-westerse allochtoon	32	20	15	6	26	100
Leeftijd						
15-24 jaar	56	22	11	3	7	100
25-34 jaar	42	21	13	5	18	100
35-44 jaar	33	21	15	6	24	100
45-54 jaar	24	20	15	7	35	100
55-64 jaar	21	18	14	9	38	100
Opleiding						
Basisonderwijs	31	20	15	5	30	100
VMBO	38	20	14	5	22	100
MBO/HAVO/VWO	37	21	14	6	22	100
HBO	39	22	11	7	21	100
WO	34	19	14	7	26	100
Totaal	37	21	14	6	23	100

Bron: CBS (EBB, 2005)

De startdatum van werkloosheid, de datum van het vinden van de baan en de datum van ondervraging hebben betrekking op de maand waarin dit plaatsvond. De werkloosheidsduur is dan ook berekend in maanden. Vooral voor personen die kort werkloos zijn, kan dit enigszins een vertekening van de werkelijkheid geven. Het kan zijn dat iemand die op 31 januari is gestopt met werken op 1 februari wordt geënquêteerd. Deze persoon heeft dan een werkloosheidsduur van 1 maand, terwijl dit in werkelijkheid slechts 1 dag is. Aan de andere kant kan iemand aan het begin van de maand werkloos zijn geworden en aan het eind van de maand worden geënquêteerd. Voor deze persoon is een werkloosheidsduur van 0 maanden berekend, terwijl dit in werkelijkheid eerder 1 maand is. Uiteindelijk is er daarom voor gekozen om iedereen met een (onvoltoide of voltooide) werkloosheidsduur van 0 of 1 maanden een werkloosheidsduur van een halve maand te geven.

De voltooide werkloosheidsduur

Voor baanvinders is ook de startdatum van werkloosheid achterhaald. Daarnaast is van de baanvinders de datum bekend van het vinden van de baan. Aan de hand van beide data kan voor baanvinders de voltooide werkloosheidsduur worden gereconstrueerd. Dit is de duur in maanden van de startdatum van werkloosheid tot en met de datum van het vinden van de baan.

Voor baanvinders zijn - in tegenstelling tot werklozen - twee data beschikbaar als startdatum van werkloosheid, namelijk: de datum van de laatste baan (stopdatum) en de schoolverlatersdatum (svdatum). De datum wanneer een persoon is gaan zoeken naar werk (zoekdatum) ontbreekt voor baanvinders.

Stopdatum

Van 31% van de baanvinders ontbreekt stopdatum. Het gaat hier vooral om jongeren en herintreders. Voor beide groepen geldt dat een groot deel van hen nooit eerder een baan heeft gehad (en daarmee ook geen datum waarop is gestopt in de laatste baan). Voor jongeren levert dit in principe geen problemen op omdat voor hen de datum van schoolverlaten vooral relevant is. Deze svdatum kan voor jongeren worden gebruikt als startdatum van werkloosheid. Bovendien geldt dat veel jongeren nog onderwijs volgen. Wanneer scholieren en studenten niet zouden worden meegenomen, blijkt ook dat het aantal onbekenden voor stopdatum afneemt.

Svdatum

Van 11% van de baanvinders ontbreekt de svdatum. Vooral van veel ouderen is deze datum niet aanwezig. Het relatief grote aantal onbekenden bij ouderen is hier echter een minder groot probleem. Voor ouderen is vooral de stopdatum relevant. Veel ouderen hebben - in tegenstelling tot jongeren - al een baan gehad na het schoolverlaten. Voor veel ouderen is dan ook de datum van de laatste baan het de startdatum van werkloosheid.

Toch is geprobeerd om te bekijken of voor ouderen svdatum alsnog kan worden bepaald. Het groot aantal onbekenden bij de svdatum van ouderen hebben te maken met een verandering die in 2004 is ingevoerd in de vragenlijst van de EBB. Alle personen kunnen in de EBB-vragenlijst voor maximaal zes opleidingen die ze na het basisonderwijs hebben gevolgd een aantal kenmerken opgeven. Aan personen van 36 jaar en ouder is vanaf 2004 echter niet meer gevraagd

wanneer ze met de betreffende opleiding(en) zijn begonnen; nog wel wanneer ze de betreffende opleiding(en) hebben beëindigd. Hierdoor kan voor een deel van de personen van 36 jaar en ouder de duur van de opleidingen niet meer worden achterhaald, en daarmee ook niet de schoolverlatersdatum.

Deze informatie kan alsnog worden aangevuld door als uitgangspunt te nemen dat het einde van de laatste opleiding tevens het begin is van de volgende opleiding. Dit is gedaan voor personen van 36 jaar en ouder waarvan de schoolverlatersdatum ontbreekt. Voor deze personen is er een reeks van einddata bekend van maximaal zes opleidingen. Hiervan is dan de meest recente genomen als schoolverlatersdatum. Er is echter één uitzondering gemaakt. De opleiding moet zijn afgerond voor de leeftijd van 36 jaar. Er is gebleken dat bij de reeks van opleidingen die een persoon kan hebben gevolgd soms grote verschillen tussen de betreffende data bestaan. Bijvoorbeeld een persoon heeft in 1978 een eerste opleiding afgerond, de tweede in 1982 en de derde in 2004. De laatste opleiding is dan waarschijnlijk een cursus of bedrijfsopleiding. Door als selectie op te nemen dat de opleiding voor de leeftijd van 36 jaar moet zijn afgerond wordt dit vermeden. Bovendien mag er van worden uitgegaan dat de meeste opleidingen die na de leeftijd van 35 jaar worden gevolgd cursussen of bedrijfsopleidingen zijn. Na deze toevoeging is het percentage baanvinders waarvan de svdatum ontbreekt gedaald tot 5%.

Verder geldt dat baanvinders die al een baan hebben gevonden voor het verlaten van de opleiding een werkloosheidsduur van 0 maanden hebben gekregen. Dit komt relatief veel voor omdat een deel van de leerlingen/studenten al is begonnen met zoeken naar werk - en een baan heeft gevonden - voordat ze de opleiding verlaten. Daarnaast zal een deel van de scholieren een leer-werk traject of een stage continueren en daardoor al werk hebben op het moment van schoolverlaten.

Zoekdatum

Zoals gezegd ontbreekt voor baanvinders de datum wanneer iemand is gaan zoeken naar werk (zoekdatum). Dit laatste vormt vooral een probleem voor personen die lang geleden gestopt zijn met werken, daarna vele jaren inactief zijn, en recent op zoek zijn gegaan naar werk. Hier gaat het om herintreders. Vooral voor deze personen is het van belang te achterhalen wanneer ze actief zijn gaan zoeken naar werk.

Aangezien de gegevens hierover ontbreken, worden geprobeerd deze zoekdatum zo goed mogelijk te schatten. Dit gebeurt met behulp van gegevens van personen die op het moment van de enquête werkloos zijn. Van deze personen is namelijk wel de stopdatum, zoekdatum en svdatum bekend. Aan de hand hiervan kan voor werkloze personen de "inactiviteitsduur" worden bepaald. Met inactiviteitsduur wordt hier bedoeld: de duur in maanden vanaf het moment van de laatste baan/moment van schoolverlaten tot het moment van zoeken naar werk. Vervolgens kan met behulp van deze informatie voor baanvinders de inactiviteitsduur worden gesimuleerd. Hiermee kan vervolgens de zoekdatum voor baanvinders worden geschat.

Deze simulatie van de zoekdatum vindt alleen plaats onder personen waarvoor dit relevant is. Het betreft personen die langer dan twee jaar geleden zijn gestopt in de laatste baan, die langer dan een jaar geleden een reguliere opleiding hebben verlaten en die geen scholier of student zijn.

In hoofdstuk 3 van dit rapport wordt verder ingegaan op de simulatie van de zoekdatum van baanvinders.

Bepaling van de werkloosheidsduur

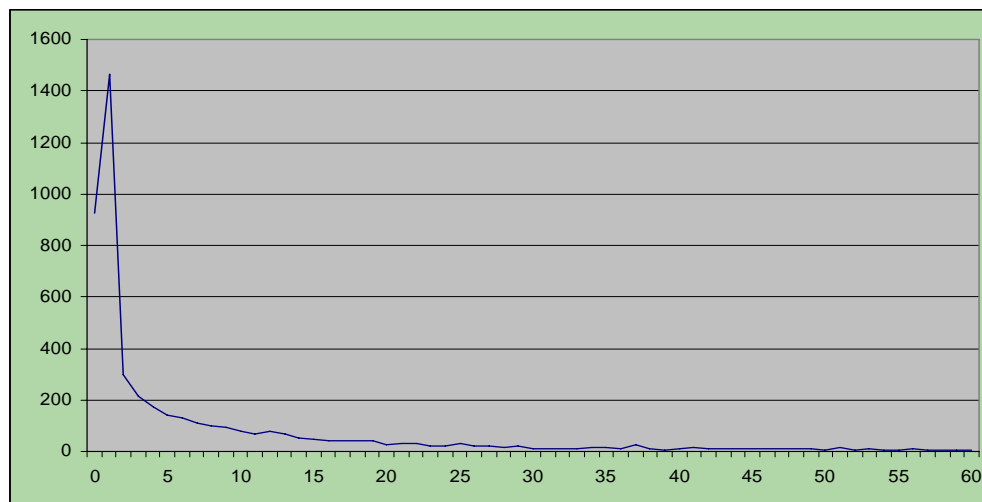
Nadat de zoekdatum is geïmplementeerd kan de startdatum van werkloosheid voor baanvinders worden gereconstrueerd. Uitgangspunt is dat van de stopdatum en svdatum de meest recente wordt genomen (en voor personen waarbij de zoekdatum is geschat, geldt dit voor stopdatum, svdatum en zoekdatum). Vervolgens kan de werkloosheidsduur (in maanden) worden berekend voor baanvinders door de startdatum van werkloosheid te koppelen aan de datum van het vinden van een baan.

In totaal is er in de periode 2002-2005 van 83% van de baanvinders de werkloosheidsduur bekend (26.169 baanvinders). De baanvinders waarvan de werkloosheidsduur onbekend is, zijn voor het merendeel scholieren/studenten. Wanneer deze groep buiten beschouwing wordt gelaten ontbreekt nog maar van 7% van de baanvinders de werkloosheidsduur. Zoals eerder vermeld is ervoor gekozen om de groep scholieren/studenten niet mee te nemen in de uiteindelijke analyse.

Figuur 6.3 vermeldt de werkloosheidsduur van baanvinders (tot een maximum van 60 maanden). De figuur laat zien dat relatief veel baanvinders 0 of 1 maanden werkloos zijn geweest. Dit komt omdat bij de baanvinders ook de baanwisselaars worden meegenomen. Baanwisselaars zijn personen die van de ene naar de andere baan zijn overgestapt (waarbij de duur tussen twee banen minimaal 0 en maximaal 3 maanden bedraagt). Daarna neemt het aantal baanvinders flink af met de stijging van de werkloosheidsduur. In figuur 3 is de werkloosheidsduur van baanvinders nogmaals vermeld, maar dan exclusief de baanwisselaars.

In Tabel 6.2 is de werkloosheidsduur in klassen vermeld, uitgesplitst naar een aantal achtergrondkenmerken. In Tabel 6.3 is dit weergegeven, maar dan exclusief de baanwisselaars.

Figuur 6.3 Voltooide werkloosheidsduur (baanvinders) in maanden, 2005



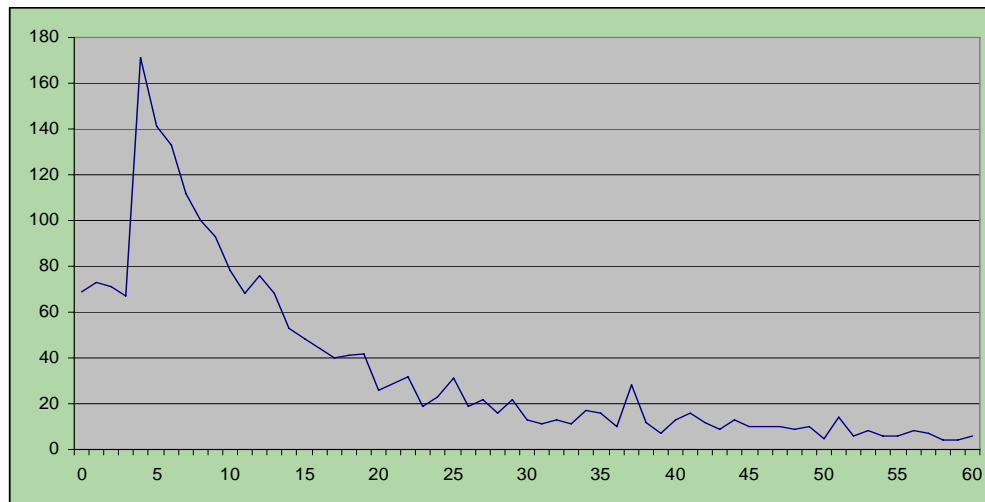
Bron: CBS (EBB, 2005)

Tabel 6.2 Werkloosheidsduur baanvinder, in klassen van 5 maanden, 2005

	0 - 5 maanden	6 - 11 maanden	12 - 17 maanden	18 - 23 maanden	> 24 maanden	Totaal
Geslacht						
Man	69	12	6	4	9	100
Vrouw	63	13	8	3	13	100
Etniciteit						
Autochtoon	68	12	7	3	10	100
Westerse allochtoon	60	15	8	4	13	100
Niet-westerse allochtoon	54	15	8	5	18	100
Leeftijd						
15-24 jaar	62	15	8	4	11	100
25-34 jaar	68	12	6	3	11	100
35-44 jaar	71	11	6	3	9	100
45-54 jaar	66	12	8	3	11	100
55-64 jaar	51	14	10	6	19	100
Opleiding						
Basisonderwijs	58	8	7	5	23	100
VMBO	58	14	8	5	16	100
MBO/HAVO/VWO	66	12	7	3	11	100
HBO	71	13	7	2	7	100
WO	74	11	6	3	6	100
Totaal	66	12	7	3	11	100

Bron: Bron: CBS (EBB, 2005)

Figuur 6.4 Werkloosheidsduur baanvinders in maanden, 2005 (exclusief baanwisselaars)



Bron: CBS (EBB, 2005)

Tabel 6.3 Werkloosheidsduur baanvinder, in klassen van 5 maanden (excl. baanwisselaars)

	0 - 5 maanden	6 - 11 maanden	12 - 17 maanden	18 - 23 maanden	> 24 maanden	Totaal
Geslacht						
Man	26	28	15	9	22	100
Vrouw	29	25	15	6	25	100
Etniciteit						
Autochtoon	29	26	15	7	23	100
Westerse allochtoon	22	30	16	7	25	100
Niet-westerse allochtoon	23	25	13	8	30	100
Leeftijd						
15-24 jaar	42	22	12	7	17	100
25-34 jaar	23	28	15	8	26	100
35-44 jaar	18	30	18	8	27	100
45-54 jaar	15	30	19	9	28	100
55-64 jaar	11	26	18	10	35	100
Opleiding						
Basisonderwijs	19	15	14	9	43	100
VMBO	23	25	14	8	29	100
MBO/HAVO/VWO	27	27	15	8	24	100
HBO	33	30	16	5	17	100
WO	37	28	14	7	14	100
Totaal	28	26	15	7	24	100

Bron: CBS (EBB, 2005)

Bijlage 2: Duurmodel

Duurmodellen en de uittredevoet

De duurvariabelen in dit onderzoek (werkloosheidsduur; inactiviteitsduur) zijn geanalyseerd met behulp van het Weibull-model. Dit is een regressiemodel waarin een duurvariabele wordt gerelateerd aan een set van verklarende variabelen, zoals dat in wezen ook gebeurt in een gewone kleinste kwadraten (OLS) regressie. Het grootste verschil met OLS-regressie is dat in een duurmodel de verklarende variabelen worden opgenomen in een vergelijking die de uittredevoet (ook wel de *hazard rate* genoemd) uitdrukt. De uittredevoet θ is gedefinieerd als de voorwaardelijke kans per oneindig klein tijdsinterval dt om vanuit een zekere toestand op tijdstip t over te gaan naar een andere. In het onderhavige onderzoek beschrijft θ de overgang van werkloosheid naar betaald werk (of in het geval van inactiviteitsduren de overgang van inactiviteit naar aanbod op arbeidsmarkt). Tijdvariabele t meet daarbij de individuele verstreken tijd in de toestand van beschouwing, in dit geval dus werkloosheid. In formele notatie:

$$\theta(t) \equiv \lim_{dt \rightarrow 0} \frac{P(t \leq T < t + dt \mid T \geq t)}{dt} = \lim_{dt \rightarrow 0} \frac{P(t \leq T < t + dt)}{P(T \geq t)} = \frac{f(t)}{1 - F(t)} \quad (\text{B.1})$$

De uittredevoet is in het geval van continue tijd, zoals in bovenstaande vergelijking, een intensiteit en geen kans. Indien de tijd in discrete eenheden wordt behandeld is de uittredevoet per definitie wel een kans:

$$\theta(z_j) \equiv P(T = z_j \mid T \geq z_j) = \frac{f(z_j)}{1 - F(z_j)} \quad j = 1, 2, \dots \quad (\text{B.2})$$

Essentieel in de definitie is de voorwaarde $T \geq t$ (of z_j) die bepaalt dat alleen zij die t of langer werkloos zijn in beschouwing worden genomen; zij die eerder werk vonden doen niet meer mee. Instructief voor begrip van de uittredevoet en het onderscheid dat de voorwaarde aanbrengt, is het te plaatsen in de context van levensduren van mensen. De onvoorwaardelijke kans bij de geboorte om 65 te worden geeft het percentage weer van alle mensen - dus ook zij die dat niet halen - die in het 66-ste levensjaar sterven. De uittredevoet geeft weer welk deel van de 65-jarigen voor hun 66-ste verjaardag sterft. Uit definitie (B.1) volgt dat een één-op-één relatie bestaat tussen de uittredevoet en de verdelingsfunctie $F(t)$ met bijbehorende dichtheidsfunctie $f(t) = \partial F(t) / \partial t$. Met andere woorden: met de keuze van een bepaalde functionele vorm voor $\theta(t)$ wordt impliciet voor een zekere verdelingsfunctie gekozen en andersom. De functie $1 - F(t)$ wordt in de theorie van duurmodellen ook wel de *survival function* genoemd, vaak aangeduid met $S(t)$. Voor meer achtergrond informatie over duurmodellen wordt verwezen naar de standaard werken: Kalbfleish & Prentice (1980) en Lancaster (1990).

Het Weibull model

Het Weibull model ontleent haar naam aan de verdelingsfunctie die resulteert bij de volgende specificatie van de uittredevoet:

$$\theta(t) = \alpha \lambda t^{\alpha-1} \quad (\text{B.3})$$

In (B.3) geeft t de tijd weer en zijn α en λ te schatten parameters, die het verloop van de uittredevoet in de tijd en de hoogte ervan bepalen. In het onderhavige onderzoek kan λ per individu variëren door een relatie te leggen met een set van persoonskenmerken die we aanduiden met x_i . Omdat λ per definitie positief moet zijn, worden de x -variabelen opgenomen in de exponentiële functie:

$$\lambda_i = \exp(x_i' \beta) \quad (\text{B.4})$$

De Weibull verdeling kent de volgende verdelingsfunctie:

$$F(t) = 1 - \exp(-\lambda t^\alpha) \quad (\text{B.5})$$

De survivalfunctie en de inverse daarvan zijn respectievelijk:

$$S(t) = \exp(-\lambda t^\alpha) \quad (\text{B.7})$$

$$S^{-1}(p) = [\ln(p) \cdot \exp(-x_i' \beta)]^{1/\alpha} \quad (\text{B.8})$$

De verwachting $E(t)$ is:

$$E(t) = \exp(-x_i' \beta / \alpha) \cdot \Gamma(1 + \alpha^{-1}) \quad (\text{B.9})$$

waarbij $\Gamma(\cdot)$ de gammafunctie voorstelt. De mediaan $M(t)$ kan eenvoudig worden afgeleid uit (B.7) door gelijk te stellen aan $1/2$ en te inverteren:

$$M(t) = [\ln(2) \cdot \exp(-x_i' \beta)]^{1/\alpha} \quad (\text{B.10})$$

De α -parameter varieert niet per individu, maar wordt geschat voor verschillende arbeidsmarkt-groepen. Daarmee wordt impliciet verondersteld dat het verloop van de uittredevoet in de tijd karakteristiek is voor diverse groepen op de arbeidsmarkt en dat individuen binnen de groep alleen in hoogte van de uittredevoet afwijken van het groepspatroon. Als geldt dat $\alpha < 1$, dan neemt de uittredevoet af in de tijd. De kans op werk neemt dan af naarmate de werkloosheid in de tijd voortduurt. Als geldt dat $\alpha > 1$, dan neemt de uittredevoet juist toe in de tijd en als $\alpha = 1$ dan is de uittredevoet constant in de tijd. Het Weibull model behoort tot de klasse van proportionele hazardmodellen, omdat het effect van persoonlijke karakteristieken tijdens de duur proportioneel onveranderd blijven. Dat wil zeggen: als mannen een 5% hogere uittredevoet hebben dan vrouwen, dan wordt verondersteld dat dat verschil op elk moment geldt: na één maand, na twee maanden of na welk aantal maanden dan ook.

Duurafhankelijkheid, niet-waargenomen heterogeniteit

De aard van tijdafhankelijkheid van de uittredevoet (de α -parameter) is een belangrijk aspect in de analyse van werkloosheidsduren. Bij een *positieve duurafhankelijkheid* wordt kans op een baan groter naarmate de werkloosheid langer voort duurt. Bij *negatieve duurafhankelijkheid* wordt de kans op een baan na verloop van tijd kleiner. Dit laatste verschijnsel wordt in de economie ook wel *hysterese* genoemd en is in verschillende studies aangetroffen bij Nederlandse werklozen in de jaren tachtig (zie Ter Huurne, 1988 en Kerckhoffs & De Neubourg, 1991). In de literatuur is veel aandacht geschonken aan de schijnbare tijdafhankelijkheid van de uittredevoet (zie bijvoorbeeld Lancaster, 1979, Lancaster & Nickell, 1980 en Elbers & Ridder, 1982). Een zwak punt van proportionele hazardmodellen, waarin niet-waargenomen heterogeniteit van de waarnemingen niet wordt ondervangen, is dat negatieve duurafhankelijkheid niet te onderscheiden is van niet-waargenomen heterogeniteit. Met andere woorden: zou men een Weibull-model gebruiken voor duren die in werkelijkheid voortkomen uit een datagenererend proces met constante uittredevoet, maar één of meer verklarende variabelen weglaten, dan wordt ten onrechte negatieve duurafhankelijkheid door het model waargenomen. Dat is in te zien met een voorbeeld. Neem een groep van ten onrechte homogeen veronderstelde waarnemingen. Daarvan hebben degenen met de in werkelijkheid hogere uittredevoet gemiddeld kortere zoekduren dan degenen met een lagere uittredevoet. Indien het model duurafhankelijkheid toestaat dan komt die negatieve correlatie tussen uittredevoet en zoekduur tot uitdrukking in negatieve duurafhankelijkheid. Immers, de uittredevoet lijkt in deze ‘homogene’ groep lager bij de langere duren. Niet-waargenomen heterogeniteit heeft in het Weibull-model volgens Lancaster & Nickell (1980) de volgende effecten: 1) overschatting van de (eventuele) daling van de uittredevoet in de tijd, en 2) onderschatting van het effect van de wel waargenomen variabelen op de uittredevoet en mogelijk onterecht concluderen dat het effect met de tijd afneemt.

In het onderhavige onderzoek zijn schattingen gedaan met een multiplicatieve gamma-verdeelde storingsterm in een Weibull hazardspecificatie. Dit leidt tot duren die de relatief onbekende Burr-verdeling volgen. Voor details over het Burr-model, zie bijvoorbeeld Berkhout (2004). Het Burr-model is te prefereren boven het Weibull-model omdat daarin rekening wordt gehouden met niet-waargenomen heterogeniteit. Het schatten van het Burr-model bracht in de praktijk echter complicaties met zich mee, omdat zich voor enkele subgroepen (numerieke) convergentieproblemen voordeden. Dergelijke complicaties zijn onwenselijk in een proces dat regelmatig moet worden herhaald voor de productie van een kwartaalstatistiek. Daarom is ervoor gekozen het Weibull in alle subgroepen toe te passen. Dat model convergeert altijd zonder problemen. Een nadeel hiervan is wel dat bij de voltooiing van onvoltooide werkloosheidsduren mogelijk een overschatting wordt gemaakt (door niet-waargenomen heterogeniteit). Het effect hiervan blijft echter beperkt om twee redenen. Ten eerste omdat een minderheid van de duren onvoltooid wordt waargenomen. Ten tweede omdat de mediaan als statistiek is gekozen, waardoor niet elke waargenomen onvoltooid duur relevant is voor uitkomst van de mediaan². De geschatte Weibull-coëfficiënten van de verklarende variabelen zijn ook minder betrouwbaar dan die van het Burr-model, omdat een vertekening richting nul kan optreden. Dat betekent dat niet-significante variabelen (volgens de Weibull-schatting) mogelijk wel een effect hebben op de werk-

² Voor een onvoltooid duur (ver) boven de mediaan maakt het niet uit of de voorspelde voltooide duur wordt overschat. Dergelijke waarnemingen tellen ongeacht de modelvoorspelling altijd louter mee als massa boven de mediaan.

loosheidsduur en dat het effect van significante variabelen in werkelijkheid groter is. Men dient de geschatte coëfficiënten met dit in het achterhoofd te interpreteren.

Na de modelschatting zijn voor elke waarneming in beginsel de parameters α en λ bekend op grond van de subgroep waartoe de waarneming behoort en zijn x -variabelen. Dat betekent dat voor elke respondent in de EBB bekend is uit welke verdelingsfunctie de waargenomen duur een trekking is. Daarbij is de vorm van de verdeling (de Weibull) voor iedereen dezelfde, maar is de ligging en spreiding van de verdeling voor elk individu anders (afhankelijk van de parameters α en λ die mede worden bepaald door de individu specifieke x -variabelen). Dit idee gebruiken we bij het oplossen van onvolkomenheden in de EBB-gegevens.

Onvoltooide duren voltooien

Stel T_i is een onvoltooide duur van een respondent met karakteristieken x_i behorend tot een groep op de arbeidsmarkt met geschatte parameters α en β . We veronderstellen dat de voltooide duur nooit langer kan voortduren dan tot het moment waarop de respondent 65 wordt. Die maximale duur duiden we aan met \hat{T}_i . De voltooide duur \tilde{T}_i van respondent i wordt dan als volgt berekend.

$$\pi_1 = S(T_i) = \exp(-\exp(x_i' \beta) \cdot T_i^\alpha) \quad (\text{B.11})$$

$$\tilde{T}_i = \min(\hat{T}_i, [-\ln(u\pi_1) \cdot \exp(-x_i' \beta)]^{1/\alpha}) \quad (\text{B.12})$$

waarbij u een willekeurige trekking is uit de (0,1)-uniforme verdeling³.

Imputatie bij ontbrekende informatie

Stel L_W is de (voltooide) inactiviteitsduur van werkloze respondenten met karakteristieken x_i . Een modelschatting op een steekproef van L_W levert geschatte parameters α en β op. De (niet waargenomen) inactiviteitsduur van baanvinders (eveneens met karakteristieken x_i) duiden we aan met L_B . Voor alle baanvinders die vanuit inactiviteit weer aan het werk gingen simuleren we L_B als volgt.

$$L_B = \min(t_2 - t_0, [-\ln(u) \cdot \exp(-x_i' \beta)]^{1/\alpha}) \quad (\text{B.13})$$

waarbij u een willekeurige trekking is uit de (0,1)-uniforme verdeling.

De werkloosheidsduur wordt waargenomen in gehele aantallen maanden. Om cijfers achter de komma te simuleren gaan we als volgt te werk. Stel $T_i = \{1, 2, 3, 4, \dots\}$ is een zoekduur van een respondent met karakteristieken x_i behorend tot een groep op de arbeidsmarkt met geschatte parameters α en β . Laat \tilde{T}_i de bewerkte duur met informatie achter de komma zijn. De berekeningswijze is als volgt:

³ Met het oog op reproduceerbaarheid van de uitkomsten is de willekeurige trekking u als variabele in de dataset vastgelegd.

$$\pi_1 = S(T_i - 0.5) = \exp(-\exp(x_i' \beta) \cdot (T_i - 0.5)^\alpha) \quad (\text{B.14})$$

$$\pi_2 = S(T_i + 0.5) = \exp(-\exp(x_i' \beta) \cdot (T_i + 0.5)^\alpha) \quad (\text{B.15})$$

$$\tilde{T}_i = [-\ln(u[\pi_1 - \pi_2] + \pi_2) \cdot \exp(-x_i' \beta)]^{1/\alpha} \quad (\text{B.16})$$

waarbij u een willekeurige trekking is uit de (0,1)-uniforme verdeling.

Effenen van onwaarschijnlijke pieken

Stel T_i is een zoekduur van een respondent met karakteristieken x_i behorend tot een groep op de arbeidsmarkt met geschatte parameters α en β . Laat T_i behoren tot een van de waargenomen pieken: $T_i = \{6, 12, 18, 24, 30, 36, 48, 60, \dots\}$. We veronderstellen dat de zoekduur nooit langer kan voortduren dan tot het moment waarop de respondent 65 wordt. Die maximale duur duiden we aan met \hat{T}_i . Een zoekduur van bijvoorbeeld 6 maanden van respondent i wordt dan als volgt opnieuw berekend in het interval $\{3; 9\}$ en wordt genoteerd als \tilde{T}_i :

$$\pi_1 = S(T_i - 3) = \exp(-\exp(x_i' \beta) \cdot (T_i - 3)^\alpha) \quad (\text{B.17})$$

$$\pi_2 = S(T_i + 3) = \exp(-\exp(x_i' \beta) \cdot (T_i + 3)^\alpha) \quad (\text{B.18})$$

$$\tilde{T}_i = \min(\hat{T}_i, [-\ln(u[\pi_1 - \pi_2] + \pi_2) \cdot \exp(-x_i' \beta)]^{1/\alpha}) \quad (\text{B.19})$$

waarbij u een willekeurige trekking is uit de (0,1)-uniforme verdeling.

In bovenstaande formules worden de constanten -3 en +3 aangepast naar gelang de veronderstelde spreiding rond de waargenomen piek anders is. Rond bijvoorbeeld 36 maanden zijn de constanten gelijk aan -6 en +6. Onderstaande schema geeft aan hoe de diverse pieken zij aangepakt.

Piek bij	Interval	Constanten
6 maanden	{3 maanden; 9 maanden}	-3 en +3
12 maanden	{9 maanden; 15 maanden}	-3 en +3
18 maanden	{15 maanden; 21 maanden}	-3 en +3
24 maanden	{21 maanden; 27 maanden}	-3 en +3
30 maanden	{27 maanden; 33 maanden}	-3 en +3
36 maanden	{33 maanden; 42 maanden}	-3 en +6
48 maanden	{42 maanden; 54 maanden}	-3 en +6
60...	{54... maanden; 66... maanden}	-6 en +6

Vergelijkingen (B.17) t/m (B.19) worden echter niet voor alle waarnemingen in een piek berekend, omdat er natuurlijk ook waarnemingen zijn die werkelijk precies 6 of 12 maanden werkloos waren. Voor welke respondenten dat precies geldt, is onbekend. Wij wijzen

(reproduceerbaar willekeurig) een aantal respondenten in de pieken aan, waarvoor een nieuwe duur wordt berekend. Dat aantal wordt vastgesteld op grond van de aanname dat het aantal waarnemingen in de piek na vereffening gelijk is aan het totale aantal waarnemingen in het interval gedeeld door de lengte van het interval. Bijvoorbeeld: we nemen in interval $\{9; 15\}$ met lengte 6 in totaal 600 duren waar, waarvan 150 met een duur van 12 (d.w.z. tussen 11,5 en 12,5 tijdseenheden). Na vereffening moeten er ongeveer $600:6=100$ waarnemingen met een duur van 12 overblijven. Door nu voor 60 van de 150 waarnemingen een nieuwe duur te berekenen in het interval $\{9; 15\}$ wordt het aantal waarnemingen met duur 12 naar schatting $150 - 60 + 60:6=100$. Merk op dat in dit voorbeeld naar verwachting 1 op de 6 herberekeningen toch weer leidt tot een duur van 12; wat tot uitdrukking komt in de term $60:6$.

In formele notatie wordt het aantal te herberekenen duren in een piek als volgt bepaald. Stel N is het aantal waarnemingen in een piekinterval met lengte L en M is het aantal waarnemingen in de betreffende piek. Het aantal willekeurig aan te wijzen respondenten in de piek dat een andere duur krijgt is dan:

$$X = \frac{L}{1-L} \left(\frac{N}{L} - M \right) \quad (\text{B.20})$$

Bijlage 3: STATA verwerkingstappen

In deze bijlage worden de verschillende stappen weergegeven zoals deze in STATA zijn geprogrammeerd om onvoltooide duren te completeren. Ook de imputatie van de inactiviteitsduren en het effenen van de frequentiepieken zijn in deze programmatuur terug te vinden. Waar relevant is bij de verwerkingstappen de output vermeld, zoals de schattingsresultaten van het regressie-duurmodel. Het gaat in deze bijlage om de programmatuur voor het jaar 2005. Deze is vergelijkbaar met die van de andere jaren die zijn geanalyseerd.

VOORAF: DEFINIËREN LOCALS, PROCEDURES EN CONSTANTEN

```
local x1 "vrouw leeftijd leeftidkw ljongst control_ljongst paar walloch nwalloch bo vmbo hbo
wo control_edu groningen friesland drenthe ijsselvechtntwente ijsselveluwe
arnhemostgelderlandrivland flevoland noordhollandnoord rijmond
zuidelijknoordholland rijnstreek haaglanden zeeland middenenwestbrabant noordoostbrabant
zuidoostbrabant limburg arb_baanwiss arb_schoolverl arb_overig wil011 wil1224 wil2532
p_wonb p_werkt psOI_onb psOI_lg psOI_hg sector_economisch sector_verzorgend
sector_overig";
```

```
local x2 "arb_baanwiss arb_schoolverl arb_overig";
```

```
capture program drop duurmodel1
```

```
program define duurmodel1
```

1. version 7.0
2. args lnf lnlambda lnalfa
3. tempvar L alfa
4. local t "\$ML_y1"
5. local v "\$ML_y2"
6. qui gen double `L'=exp(`lnlambda')
7. qui gen double `alfa'=exp(`lnalfa')
8. qui replace `lnf' = `v'*(ln(`alfa'*L*t^(`alfa'-1))-L*t^`alfa')+(1-`v')*(-L*t^`alfa')
9. end

```
capture program drop duurmodel2
```

```
program define duurmodel2
```

1. version 7.0
2. args lnf lnlambda lnalfa
3. tempvar L alfa
4. local t "\$ML_y1"
5. qui gen double `L'=exp(`lnlambda')
6. qui gen double `alfa'=exp(`lnalfa')
7. qui replace `lnf'=ln(`alfa'*L*t^(`alfa'-1))-L*t^`alfa'
8. end

STAP 1: HET AANMAKEN VAN VARIABELENSTAP 2: EFFENEN VAN PIEKEN EN BEPALEN ZOEKDUUR VAN WERKLOZEN

```

replace zoekduur=0.5 if zoekduur==0
zoekduur was int now float
replace zoekduur=. if zoekduur>=120
g zdoud=zoekduur
g zdvolt=1

```

```
ml model lf duurmodel1 (lambda: zoekduur zdvolt =) (alfa:)
```

```
ml maximize
```

```

initial:      log likelihood =    -43000
alternative:  log likelihood = -13628.973
rescale:      log likelihood = -12722.575
rescale eq:   log likelihood = -11995.482
Iteration 0:  log likelihood = -11995.482
Iteration 1:  log likelihood = -11254.228
Iteration 2:  log likelihood = -11243.763
Iteration 3:  log likelihood = -11243.762
Iteration 4:  log likelihood = -11243.762

```

```

Number of obs   =    3105
Wald chi2(0)    =          .
Prob > chi2     =          .

```

```
Log likelihood = -11243.762
```

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
lambda					
_cons	-2.365672	.0424884	-55.68	0.000	-2.448948 -2.282397
alfa					
_cons	-.0880627	.0136621	-6.45	0.000	-.1148398 -.0612855

```
predict hulp1, eq(lambda)
```

```
predict hulp2, eq(alfa)
```

```
g zdlambda = exp(hulp1)
```

```
g zdalfa = exp(hulp2)
```

```
drop hulp1 hulp2
```

```
forval i=6(6)30 {
```

```
2.   g zdpiek`i' =(zoekduur>=`i'-3 & zoekduur < `i'+3)
```

```
3. }
```

```
g zdpiek36 =(zoekduur>=33 & zoekduur < 42)
```

```
forval i=48(12)108 {
```

```
2.   g zdpiek`i' =(zoekduur>=`i'-6 & zoekduur < `i'+6)
```

```
3. }
```

```
forval i=6(6)30 {
```

```
2.   bysort zdpiek`i': egen tot`i'=count(zoekduur)
```

```
3. }
```

```

forval i=36(12)108 {
2.   bysort zdpiek `i': egen tot`i'=count(zoekduur)
3. }
bysort zoekduur: egen tot=count(zoekduur)

***pieken proportioneel gladstrijken; alle pieken aflopen
g Smin=.
g Smax=.
g Ssim=.

forval i=6(6)30 {
2. replace Smax=exp(-zdlambda*(zoekduur-3)^zdelta) if zoekduur==`i' & r1>(tot`i'-tot)/(6*tot)
3. replace Smin=exp(-zdlambda*(zoekduur+3)^zdelta) if zoekduur==`i' & r1>(tot`i'-tot)/(6*tot)
4. }

replace Smax=exp(-zdlambda*(zoekduur-3)^zdelta) if zoekduur==36 & r1>(tot36-tot)/(9*tot)
replace Smin=exp(-zdlambda*(zoekduur+6)^zdelta) if zoekduur==36 & r1>(tot36-tot)/(9*tot)

forval i=48(12)108 {
2.replace Smax=exp(-zdlambda*(zoekduur-6)^zdelta) if zoekduur==`i' & r1>(tot`i'-tot)/(12*tot)
3.replace Smin=exp(-zdlambda*(zoekduur+6)^zdelta)if zoekduur==`i' & r1>(tot`i'-tot)/(12*tot)
4. }

replace Ssim = (Smax-Smin)*r2+Smin
replace zoekduur = int((-1*ln(Ssim)/zdlambda)^(1/zdelta)) if Ssim~=.
drop Ssim Smin Smax zdpiek* tot* zdlambda zdelta r1 r2

```

STAP 3: IMPUTEREN ONTBREKENDE INFORMATIE INACTIVITEIT/ZOEKDUUR

```

g inactduur =tijdduur-zoekduur if werkloos==1 & arb_overig == 1
replace inactduur =sv_duur-zoekduur if werkloos==1 & arb_overig == 1 & tijdduur == .
replace inactduur=0.5 if inactduur <= 0
egen MAX = rowmin(tijdduur sv_duur)
replace inactduur=. if werkloos== 1 & arb_overig == 1 & inactduur>MAX

ml model lf duurmodel2 (lambda: inactduur = `x1') (alfa:) if arb_overig == 1
ml maximize
initial:      log likelihood =   -87044.5
alternative:  log likelihood = -11161.473
rescale:      log likelihood = -6797.7464
rescale eq:   log likelihood = -6797.7464
Iteration 0:  log likelihood = -6797.7464
Iteration 1:  log likelihood = -6372.3156
Iteration 2:  log likelihood = -6137.0032
Iteration 3:  log likelihood = -6122.3786
Iteration 4:  log likelihood = -6121.2408
Iteration 5:  log likelihood = -6121.2392
Iteration 6:  log likelihood = -6121.2392

```

Log likelihood = -6121.2392

Number of obs = 1206
Wald chi2(40) = 364.68
Prob > chi2 = 0.0000

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]		
					lambda		
vrouw	-.2470615	.0891298	-2.77	0.006	-.4217527	-.0723703	
leeftijd	-.1445386	.0219011	-6.60	0.000	-.187464	-.1016131	
leeftijdkw	.0016551	.0002692	6.15	0.000	.0011275	.0021826	
ljongst	-.0186331	.0073572	-2.53	0.011	-.033053	-.0042131	
control_lj~t	.0893747	.1192652	0.75	0.454	-.1443807	.3231301	
paar	-.335164	.1272846	-2.63	0.008	-.5846373	-.0856907	
walloch	.1548485	.1058053	1.46	0.143	-.0525262	.3622232	
nwalloch	-.0109472	.089796	-0.12	0.903	-.1869441	.1650497	
bo	-.3258009	.1147981	-2.84	0.005	-.550801	-.1008008	
vmbo	-.277004	.0753959	-3.67	0.000	-.4247773	-.1292307	
hbo	.1130417	.1170266	0.97	0.334	-.1163262	.3424096	
wo	.2418956	.1283201	1.89	0.059	-.0096072	.4933984	
control_edu	-.0374651	.4953833	-0.08	0.940	-1.008398	.9334682	
groningen	.0690088	.1752942	0.39	0.694	-.2745616	.4125792	
friesland	.1664154	.1913499	0.87	0.384	-.2086234	.5414542	
drenthe	-.1737711	.1873803	-0.93	0.354	-.5410297	.1934875	
ijsselvech~e	-.0575018	.155689	-0.37	0.712	-.3626466	.2476429	
ijsselveluwe	.1452271	.1749245	0.83	0.406	-.1976186	.4880727	
arnhemoost~d	.2305046	.1617625	1.42	0.154	-.086544	.5475532	
flevoland	.1907951	.2054166	0.93	0.353	-.211814	.5934043	
noordholla~d	.0881268	.1838106	0.48	0.632	-.2721354	.448389	
rijnmond	-.0243495	.1546183	-0.16	0.875	-.3273957	.2786967	
zuidelijkn~d	.1018749	.2058819	0.49	0.621	-.3016462	.5053961	
rijnstreek	-.0630163	.1817541	-0.35	0.729	-.4192479	.2932152	
haaglanden	-.0644285	.145236	-0.44	0.657	-.3490859	.2202289	
zeeland	-.1882312	.2568764	-0.73	0.464	-.6916997	.3152372	
middenenwe~t	.1228731	.1609627	0.76	0.445	-.192608	.4383543	
noordoostb~t	.2802659	.1968439	1.42	0.155	-.105541	.6660729	
zuidoostbr~t	.2364708	.1678248	1.41	0.159	-.0924598	.5654013	
limburg	.0843368	.1491813	0.57	0.572	-.2080533	.3767268	
wil1224	-.5562501	.0949791	-5.86	0.000	-.7424057	-.3700946	
wil2532	-.2528384	.105619	-2.39	0.017	-.4598479	-.0458288	
p_wonb	-1.359766	.4223318	-3.22	0.001	-2.187521	-.5320112	
p_werkt	-.1867139	.1026497	-1.82	0.069	-.3879037	.0144758	
psoi_onb	.5059958	.3977877	1.27	0.203	-.2736538	1.285645	
psoi_lg	-.25292	.0896037	-2.82	0.005	-.42854	-.0773	
psoi_hg	-.1906906	.1069359	-1.78	0.075	-.4002811	.0188999	
sector_eco~h	.0844314	.1040101	0.81	0.417	-.1194247	.2882874	
sector_ver~d	-.2610196	.0991463	-2.63	0.008	-.4553427	-.0666965	
sector_ove~g	-.0777682	.0944783	-0.82	0.410	-.2629422	.1074059	
_cons	.598946	.5137431	1.17	0.244	-.4079721	1.605864	
alfa	_cons	-.0953337	.0229963	-4.15	0.000	-.1404056	-.0502618

predict hulp1, eq(lambda)

predict hulp2, eq(alfa)

g lambda = exp(hulp1)

g alfa = exp(hulp2)


```

g inactsim=.
g Ssim=r3*exp(-lambda*1^alfa)
replace inactsim = int((-1*ln(Ssim)/lambda)^(1/alfa))
gen hulp3=MAX-inactsim if werkloos == 0 & arb_overig == 1
replace zoekduur=hulp3 if werkloos == 0 & arb_overig == 1
replace zoekduur=0.5 if werkloos == 0 & arb_overig == 1 & zoekduur<=0
drop hulp1 hulp2 hulp3 Ssim lambda alfa MAX inactsim r3

```

STAP 4: BEPALING WERKLOOSHEIDSDUUR

```

drop duur
egen duur=rowmin(zoekduur sv_duur tijdduur)

*** duur begint nooit voor 15e levensjaar
g upperbound= (leeftijd-15)*12
replace duur=upperbound if duur>upperbound & duur<.

*** veronderstelling (baanvinders): begin 1e van de maand; eindig op laatste dag van de maand
replace duur=duur-1 if werkloos==0 & duur==tijdduur & duur>1

*** nullen
replace duur=0.5 if duur==0

*** extremen
*replace duur=. if duur>=120
replace extr=(duur>r(p95) & voltooid==1)

```

STAP 5: MODELSCHATTING

```
ml model lf duurmodel1 (lambda: duur voltooid = `x1' extr) (alfa: `x2') if duur<.
```

```
ml maximize
```

```

initial:      log likelihood = -123811.5
alternative:  log likelihood = -24367.843
rescale:      log likelihood = -17857.484
rescale eq:   log likelihood = -17857.484
Iteration 0:  log likelihood = -17857.484
Iteration 1:  log likelihood = -13429.001
Iteration 2:  log likelihood = -12360.042
Iteration 3:  log likelihood = -11056.358
Iteration 4:  log likelihood = -11010.835
Iteration 5:  log likelihood = -10997.095
Iteration 6:  log likelihood = -10996.785
Iteration 7:  log likelihood = -10996.784

```

```
Log likelihood = -10996.784
```

```

Number of obs   =      7675
Wald chi2(45)   =    3650.44
Prob > chi2     =      0.0000

```

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
-----+-----						
	lambda					
vrouw	-.0559812	.0372256	-1.50	0.133	-.128942	.0169797
leeftijd	-.0210188	.0123221	-1.71	0.088	-.0451696	.003132
leeftijdkw	-.0000891	.0001594	-0.56	0.576	-.0004015	.0002234
ljongst	.0090979	.0039989	2.28	0.023	.0012602	.0169356
control_lj-t	-.0098782	.0503117	-0.20	0.844	-.1084873	.088731
paar	.0562163	.0556571	1.01	0.312	-.0528696	.1653022
walloch	-.1218175	.0548122	-2.22	0.026	-.2292475	-.0143875
nwalloch	-.2531247	.0574374	-4.41	0.000	-.3656999	-.1405496
bo	-.1871811	.0790185	-2.37	0.018	-.3420546	-.0323076
vmbo	-.1610964	.04081	-3.95	0.000	-.2410824	-.0811103
hbo	.0818356	.0432075	1.89	0.058	-.0028497	.1665208
wo	.1510837	.0532225	2.84	0.005	.0467696	.2553978
control_edu	-.3833082	.2990378	-1.28	0.200	-.9694116	.2027951
groningen	-.4052499	.0995287	-4.07	0.000	-.6003225	-.2101772
friesland	-.2576209	.0960595	-2.68	0.007	-.445894	-.0693477
drenthe	-.3023105	.0992526	-3.05	0.002	-.4968421	-.107779
ijsselvech-e	-.1648782	.0750467	-2.20	0.028	-.3119671	-.0177894
ijsselveluwe	-.0682873	.0833698	-0.82	0.413	-.231689	.0951145
arnhemooost-d	-.0852817	.0720007	-1.18	0.236	-.2264005	.055837
flevoland	-.1184497	.1000454	-1.18	0.236	-.3145352	.0776358
noordholla-d	.0729449	.0893509	0.82	0.414	-.1021796	.2480693
rijnmond	-.0262569	.0689359	-0.38	0.703	-.1613688	.1088549
zuidelijkn-d	-.0278146	.0792023	-0.35	0.725	-.1830483	.1274191
rijnstreek	.0405335	.0801869	0.51	0.613	-.11663	.197697
haaglanden	-.1578222	.0651846	-2.42	0.015	-.2855817	-.0300628
zeeland	-.0486552	.1076151	-0.45	0.651	-.2595768	.1622664
middenenwe-t	-.0111131	.0710965	-0.16	0.876	-.1504596	.1282334
noordoostb-t	-.0750084	.0924556	-0.81	0.417	-.256218	.1062012
zuidoostbr-t	-.181845	.0747315	-2.43	0.015	-.328316	-.035374
limburg	-.1803315	.070378	-2.56	0.010	-.3182698	-.0423931
arb_baanwiss	4.382313	.1168343	37.51	0.000	4.153323	4.611304
arb_school-l	2.503228	.1348891	18.56	0.000	2.23885	2.767606
arb_overig	1.41955	.1369293	10.37	0.000	1.151173	1.687926
wil011	.4749789	.2320162	2.05	0.041	.0202355	.9297223
wil1224	-.2081409	.0517747	-4.02	0.000	-.3096174	-.1066643
wil2532	-.0006156	.0452224	-0.01	0.989	-.0892499	.0880188
p_wonb	-.1848152	.2678041	-0.69	0.490	-.7097016	.3400712
p_werkt	.0866197	.0485533	1.78	0.074	-.008543	.1817825
psoi_onb	.1764157	.2614653	0.67	0.500	-.336047	.6888783
psoi_lg	-.0450471	.0495369	-0.91	0.363	-.1421376	.0520435
psoi_hg	.0179809	.0464059	0.39	0.698	-.0729731	.1089349
sector_eco-h	.0150422	.0433032	0.35	0.728	-.0698305	.0999149
sector_ver-d	.04339	.0451134	0.96	0.336	-.0450307	.1318107
sector_ove-g	-.0748224	.0470066	-1.59	0.111	-.1669536	.0173088
extr	-.466704	.0824811	-5.66	0.000	-.6283639	-.3050441
_cons	-3.515154	.2937143	-11.97	0.000	-4.090823	-2.939484
-----+-----						
	alfa					
arb_baanwiss	.4470132	.0298093	15.00	0.000	.3885881	.5054383
arb_school-l	-.4309226	.0388381	-11.10	0.000	-.5070438	-.3548014
arb_overig	-.8664027	.0367389	-23.58	0.000	-.9384096	-.7943958
_cons	.5299828	.0256266	20.68	0.000	.4797556	.58021
-----+-----						

```

predict hulp1, eq(lambda)
predict hulp2, eq(alfa)
g lambda_duur = exp(hulp1)
g alfa_duur = exp(hulp2)
drop hulp1 hulp2

```

STAP 6: AANPASSEN VAN DUREN OP BASIS VAN HET MODEL

```

*** Maximale zoekduur bij 65 jaar
g duurmax = 12*(65-leeftijd)+duur

***** Proportioneel gladstrijken van pieken bij 6, 12, 18, 24, 36 etc en ontdiscretiseren *****
g duuroud=duur

forval i=6(6)18 {
  2.   g piek`i' =(duur>=`i'-3 & duur < `i'+3 & voltooid==0)
  3. }

g piek24 =(duur>=21 & duur < 30 & voltooid==0)

forval i=36(12)108 {
  2.   g piek`i' =(duur>=`i'-6 & duur < `i'+6 & voltooid==0)
  3. }
forval i=6(6)24 {
  2.   bysort voltooid piek`i': egen tot`i'=count(duur)
  3. }
forval i=36(12)108 {
  2.   bysort voltooid piek`i': egen tot`i'=count(duur)
  3. }

bysort voltooid duur: egen tot=count(duur)
g Smin=.
g Smax=.
g Ssim=.

forval i=6(6)18 {
  2. replace Smax=exp(-lambda_duur*(duur-3)^alfa_duur) if voltooid==0 & duur==`i' &
r4>(tot`i'-tot)/(6*tot)
  3. replace Smin=exp(-lambda_duur*(duur+3)^alfa_duur) if voltooid==0 & duur==`i' &
r4>(tot`i'-> tot)/(6*tot)
  4. replace Smax=exp(-lambda_duur*(duur-0.5)^alfa_duur) if voltooid==0 & duur==`i' &
r4<(tot`i'-tot)/(6*tot)
  5. replace Smin=exp(-lambda_duur*(duur+0.5)^alfa_duur) if voltooid==0 & duur==`i' &
r4<(tot`i'-tot)/(6*tot)
  6. }

```

```

replace Smax=exp(-lambda_duur*(duur-3)^alfa_duur) if voltooid==0 & duur==24 &
r4>(tot24-tot)/(9*tot)
replace Smin=exp(-lambda_duur*(duur+6)^alfa_duur) if voltooid==0 & duur==24 &
r4>(tot24-tot)/(9*tot)
replace Smax=exp(-lambda_duur*(duur-0.5)^alfa_duur) if voltooid==0 & duur==24 &
r4<(tot24-tot)/(9*tot)
replace Smin=exp(-lambda_duur*(duur+0.5)^alfa_duur) if voltooid==0 & duur==24 &
r4<(tot24-tot)/(9*tot)

```

```

forval i=36(12)108 {
2. replace Smax=exp(-lambda_duur*(duur-6)^alfa_duur) if voltooid==0 & duur==`i' &
r4>(tot`i'-tot)/(12*tot)
3. replace Smin=exp(-lambda_duur*(duur+6)^alfa_duur) if voltooid==0 & duur==`i' &
r4>(tot`i'-tot)/(12*tot)
4. replace Smax=exp(-lambda_duur*(duur-0.5)^alfa_duur) if voltooid==0 & duur==`i' &
r4<(tot`i'-tot)/(12*tot)
5. replace Smin=exp(-lambda_duur*(duur+0.5)^alfa_duur) if voltooid==0 & duur==`i' &
r4<(tot`i'-tot)/(12*tot)
6. }
(0 real changes made)

```

**** onvoltooide duur 0.5

```

replace Smax = 1 if voltooid==0 & duur==0.5
replace Smin = exp(-lambda_duur*(duur)^alfa_duur) if voltooid==0 & duur==0.5

```

**** alle resterende onvoltooide uren

```

replace Smax = exp(-lambda_duur*(duur-0.5)^alfa_duur) if voltooid==0 & Smax==.
replace Smin = exp(-lambda_duur*(duur+0.5)^alfa_duur) if voltooid==0 & Smin==.

```

**** duur ontdiscretiseren

```

replace Ssim = (Smax-Smin)*r5+Smin if voltooid==0
replace duur = (-1*ln(Ssim)/lambda_duur)^(1/alfa_duur) if voltooid==0 & Ssim~=.
drop Smax Smin Ssim piek* tot* r4 r5

```

***** onvoltooide duur voltooiën: werkloosheidsduur werklozen *****

```

g Ssim = r6*exp(-lambda_duur*duur^alfa_duur) if voltooid==0
g hulp = (-1*ln(Ssim)/lambda_duur)^(1/alfa_duur) if voltooid==0
egen werkloosheidsduur=rowmin(hulp duurmax) if voltooid==0
drop Ssim hulp r6

```

***** ontdiscreteren voltooide duur: werkloosheidsduur baanvinders

```

g Smin = exp(-lambda_duur*(duur+0.5)^alfa_duur) if voltooid==1
g Smax = exp(-lambda_duur*(duur-0.5)^alfa_duur) if voltooid==1
g Ssim = (Smax-Smin)*r7+Smin if voltooid==1
replace werkloosheidsduur = (-1*ln(Ssim)/lambda_duur)^(1/alfa_duur) if voltooid==1
drop Smin Smax Ssim r7

```

Bronnen

Berkhout, P.H.G. (2004), Van bul naar baan, Academisch Proefschrift Universiteit van Amsterdam, Stichting voor Economisch Onderzoek, SEO-rapport 709.

Berkhout, P.H.G. (200X), De zoekduur van EBB-schoolverlaters: definities en berekeningswijze, Stichting voor Economisch Onderzoek der Universiteit van Amsterdam, SEO-rapport XXX.

Elbers, Ch. & G. Ridder (1982), True and spurious duration dependence: the identifiability of the proportional hazard model, *Review of Economic Studies* (1992), pp. 403-409.

Kalbfleish J.D. & R.L. Prentice (1980), *The statistical analysis of failure time data*, John Wiley and Sons, New York.

Kerckhoffs, C. & C. de Neubourg (1991), *Determinanten van individuele werkloosheidsduur*, OSA werkdocument W 84, Den Haag.

Lancaster, T. (1979), *Econometric methods for the duration of unemployment*, *Econometrica* Vol. 47 (4), pp 939-956.

Lancaster, T. (1990), *The econometric analysis of transition data*, Cambridge University press.

Lancaster, T. & S. Nickell (1980), *The analysis of re-employment probabilities for the unemployed*, *J.R.Statist. Soc.* 1980, 143, pp.141-165.

Ter Huurne, A. (1988), *Werkloze jongeren twee jaar later*, OSA Werkdocument w55, Den Haag.