

Werkloosheidsduren op basis van de Enquête beroepsbevolking, 2002–2007

09

*Marcel Kerkhofs,
Organisatie voor Strategisch Arbeidsmarktonderzoek
Harry Bierings, Robert de Vries,
Centraal Bureau voor de Statistiek*

Met dank aan, Henk-Jan Dirven, Léander Kuijvenhoven, Joop Schippers en Johan van der Valk voor hun waardevolle commentaar op eerdere versies van dit rapport. De inhoud van dit rapport stemt niet noodzakelijkerwijs overeen met het beleid van het Centraal Bureau voor de Statistiek.

Publicatiedatum CBS-website: 19 mei 2009



Verklaring van tekens

.	= gegevens ontbreken
*	= voorlopig cijfer
x	= geheim
–	= nihil
–	= (indien voorkomend tussen twee getallen) tot en met
0 (0,0)	= het getal is kleiner dan de helft van de gekozen eenheid
niets (blank)	= een cijfer kan op logische gronden niet voorkomen
2007–2008	= 2007 tot en met 2008
2007/2008	= het gemiddelde over de jaren 2007 tot en met 2008
2007/'08	= oogstjaar, boekjaar, schooljaar enz., beginnend in 2007 en eindigend in 2008
2005/'06–2007/'08	= oogstjaar, boekjaar enz., 2005/'06 tot en met 2007/'08

In geval van afronding kan het voorkomen dat het weergegeven totaal niet overeenstemt met de som van de getallen.

Colofon

Uitgever

Centraal Bureau voor de Statistiek
Henri Faasdreef 312
2492 JP Den Haag

Prepress

Centraal Bureau voor de Statistiek - Grafimedia

Omslag

TelDesign, Rotterdam

Inlichtingen

Tel. (088) 570 70 70
Fax (070) 337 59 94
Via contactformulier: www.cbs.nl/infoservice

Bestellingen

E-mail: verkoop@cbs.nl
Fax (045) 570 62 68

Internet

www.cbs.nl

Samenvatting

In dit onderzoeksverslag gaat het vooral om de vraag of statistieken van werkloosheidsduren ontwikkeld kunnen worden op basis van duurmodellen. Hiervoor is de kansverdeling van werkloosheidsduren bepaald met behulp van een *competing risks* duurmodel, waarbij informatie is gebruikt van de vijf opeenvolgende peilingen van personen uit de Enquête Beroepsbevolking (EBB). De analyses borduren voort op eerdere fasen van het onderzoeksproject waarin de werkloosheidsduur van werklozen nog bepaald is op basis van uitsluitend de eerste peiling van de EBB. Door de vervolgpeilingen van de EBB te gebruiken, kan het probleem van *length biased sampling* (oververtegenwoordiging van lange duren in de steekproef) op een adequate manier worden opgelost. Op basis van alle vervolgpeilingen uit de EBB 2002–2007 is informatie over het vinden van een baan toegepast om de kans op werk voor werklozen te bepalen. Voor elke persoon met gegeven persoonskenmerken kan dan de verwachte werkloosheidsduur bepaald worden voor als deze nu werkloos zou worden. Ook is het mogelijk om op basis van het duurmodel de verwachte resterende werkloosheidsduur te bepalen en daarmee nog lopende werkloosheidsduren te voltooiën of (in vakjargon) te imputeren voor personen die al een bepaalde tijd werkloos zijn geweest.

Hoewel het in dit onderzoeksverslag hoofdzakelijk gaat om de vraag of en welke statistieken van werkloosheidsduren ontwikkeld kunnen worden op basis van duurmodellen is ook aandacht voor alternatieve statistieken die gebaseerd zijn op directe tellingen. Het gaat dan bijvoorbeeld over de werkloosheidsduur tot peilmoment. Daarnaast kan echter ook de invalshoek van stromen en transities worden gekozen om het dynamische aspect van werkloosheid te beschrijven. Een voorbeeld van een dergelijke statistiek is het percentage van de mensen die een jaar geleden werkloos waren dat sindsdien een baan heeft gevonden. Maar ook van het aantal aaneengesloten maanden in het afgelopen jaar dat die mensen daarna nog werkloos bleven, kan een statistiek worden gemaakt. Dergelijke statistieken kunnen gebaseerd worden op dezelfde informatie als gebruikt is voor bovengenoemde duuranalyse en zijn een relevante aanvulling op de statistieken over duren. In dit rapport wordt dan ook aanbevolen om zowel over duren als stromen van werkloosheid statistieken samen te stellen en deze als pakket te presenteren.

Inhoudsopgave

1. Inleiding	7
2. Analyse van duren	9
2.1 Duurmodellen	9
2.2 Het schatten van duurmodellen	11
2.3 <i>Competing risks</i> modellen	15
2.4 De gebruikte schattingsmethode	16
3. Data	19
3.1 Inleiding	19
3.2 Definitie werkloosheid en werkloosheidsduur	19
3.3 Enkele beschrijvende statistieken	20
4. Schattingen	23
4.1 Schattingen van het duurmodel	23
4.2 Schattingen van het <i>competing risks</i> model	26
5. Statistieken van werkloosheidsduren en -stromen	31
5.1 Inleiding	31
5.2 Het berekenen van de statistieken	33
5.3 Statistieken op basis van het aanvullen van lopende duren	36
5.4 Statistieken op basis van verwachte werkloosheidsduren van instromers	39
5.5 Andere statistieken op basis van <i>competing risks</i> schattingen	42
5.6 Statistieken op basis van transities uit werkloosheid	44
6. Conclusies	47
Referenties	49
Bijlage 1: Gesimuleerde duren	51
Bijlage 2: Gebruikte variabelen	55
Bijlage 3: Suggesties voor implementatie	57

1. Inleiding

Werkloosheid is één van de hoofdthema's van sociaaleconomisch beleid. Daarom is het van belang over goede informatie te beschikken over de ontwikkeling van de werkloosheid en de samenstelling daarvan. De statistieken van het Centraal Bureau voor de Statistiek (CBS) vormen daarvoor een belangrijke bron. Daarbij is van belang dat werkloosheid pas goed begrepen kan worden als niet alleen de omvang maar ook de in- en uitstroom uit het werkloosheidsbestand en de samenstelling goed in beeld komt. Eenzelfde werkloosheidspercentage kan horen bij een arbeidsmarkt met een grote groep mensen die zo nu en dan voor korte tijd werkloos zijn, maar ook bij een arbeidsmarkt met een kleine groep langdurig werklozen. Beleidsmatig vragen deze arbeidsmarktsituaties om een zeer verschillend optreden. De eerste situatie vraagt om het verbeteren van het proces waarmee werkzoekende en werkgevers elkaar kunnen vinden, de tweede situatie vraagt om een meer op de problemen van de werkzoekende gericht activeringsbeleid met scholing, werkgewenning en eventueel sancties. Voor dit laatste soort beleid is het nodig om ook goed zicht te hebben op de kenmerken van de langdurig werklozen. Bovendien is het voor het beoordelen van de effectiviteit van dergelijk beleid van belang te weten hoe de werkloosheidsduur zich ontwikkelt.

Voor statistische analyse zijn duren echter conceptueel complex en statistieken van werkloosheidsduren zijn daarom niet eenvoudig te construeren. Dit heeft te maken met de praktische problemen bij het meten van duren zoals:

- het feit dat de werkloosheidsperiode van huidige werklozen nog niet is afgesloten en werkloosheidsduren dus pas gedeeltelijk bekend zijn,
- soms het precieze aanvangsmoment van werkloosheid onbekend is,
- dat in een bestand van mensen die op een bepaald moment werkloos zijn lange duren oververtegenwoordigd zijn en korte duren ondervertegenwoordigd zijn in vergelijking met de kansverdeling van duren bezien vanaf het moment waarop de werkloosheid aanvangt.

Om de werkloosheidsduur goed te kunnen bestuderen is het daarom nodig gebruik te maken van statistische methoden (duurmodellen) die het mogelijk maken om de kansverdeling van de werkloosheidsduur te bepalen rekening houdend met de eigenschappen en problemen van de beschikbare gegevens.

Deze complicaties, gevoegd bij het feit dat werkloosheidsduren zich niet eenduidig laten koppelen aan een tijdstip of periode, zijn de reden dat de beschikbare statistieken doorgaans de werkloosheid als voorraadgrootte behandelen in de zin dat ze aangeven welk deel van de bevolking of van een specifieke deelgroep daarvan geen betaald werk heeft, maar daar wel naar op zoek is. Met dit soort statistieken kan in beeld worden gebracht hoe groot de onderbenutting van het arbeidsaanbod op een gegeven moment is en eventueel welk deel daarvan bestaat uit mensen die al langer dan een jaar (of bijvoorbeeld langer dan drie jaar) zonder werk zitten, maar niet hoe lang iemand die nu werkloos is nog moet zoeken alvorens een nieuwe baan te vinden.

Het onderzoek in deze rapportage maakt onderdeel uit van een langer lopend project van het CBS, dat wordt uitgevoerd in het kader van het CBS-speerpunt Sociale Dynamiek en Uitbreiding Arbeidsrekeningen. Hierin wordt de mogelijkheid onderzocht om met behulp van duuranalyses statistieken van de werkloosheidsduur te construeren. Hiermee zou het mogelijk moeten worden om de verdeling van de werkloosheidsduur en de ontwikkeling daarvan in beeld te brengen. In twee eerdere fasen zijn deze mogelijkheden al verkend. In de eerste fase is nagegaan welke gegevens zich het best lenen voor het vaststellen van het moment waarop iemand werkloos wordt en het moment waarop een werkloosheidsperiode wordt beëindigd. De uitkomst hiervan was dat de Enquête Beroepsbevolking (EBB) daarvoor de beste mogelijkheden biedt. Op basis hiervan is de werkloosheidsduur van werklozen op het eerste peilmoment bepaald. In de tweede fase zijn duurmodellen gebruikt om de werkloosheidsduur van werklozen te voltooien. Dit heeft geleid tot een gezamenlijk rapport van onderzoekers van de Stichting voor Economisch Onderzoek (SEO) en het CBS (Berkhout, Van Klaveren, Van der Valk en De Vries, 2007). Dit onderzoek heeft aanleiding gegeven tot een derde fase van onderzoek

waarin een aantal vervolgvragen centraal stonden. Deze hebben betrekking op de mogelijkheid om (1) de vervolgpellingen van de EBB te gebruiken in de duuranalyse en (2) om bij het analyseren van werkloosheidsduren rekening te houden met de mogelijkheid dat iemand op een gegeven moment ontmoedigd is om verder te zoeken naar werk. De persoon stopt met zoeken naar werk en de werkloosheidsduur eindigt zonder dat een nieuwe baan is gevonden. Doel van de eerste vervolgvraag is na te gaan welke mogelijkheden de gegevens uit de telefonische enquêterondes van de EBB bieden voor de analyse van werkloosheidsduren. In de tweede fase van dit project zijn bij het schatten van het duurmodel alleen de gegevens uit het eerste *face-to-face* interview (CAPI) van de deelnemers gebruikt. Deze vinden gespreid over het jaar plaats met een aselechte steekproef van nieuwe respondenten, elke maand ongeveer 9 000 personen. Vervolgens worden deze respondenten om het kwartaal nog vier maal benaderd voor een telefonische vervolgenquête (CATI). Door van deze gegevens gebruik te maken, kan het aantal waarnemingen worden vergroot en mogelijk ook de nauwkeurigheid van de meting van de werkloosheidsperiodes. Lopende de derde fase is gebleken dat de vervolgpellingen een nog veel belangrijkere toegevoegde waarde hebben doordat zij een natuurlijke oplossing bieden om vertekening door *length biased sampling* in de steekproefopzet te voorkomen.

Breed gesteld is het doel van de tweede vervolgvraag na te gaan wat de gevolgen zijn van het gebruik van minder restrictieve modelveronderstellingen, in het bijzonder de veronderstelling dat een periode van werkloosheid uitsluitend beëindigd kan worden met een transitie naar een betaalde werkkring. Afhankelijk van het doel van de analyse kan het daarvoor nodig zijn een ander soort duurmodel te gebruiken, waarin onderscheid wordt gemaakt tussen verschillende manieren waarop een werkloosheidsduur kan worden beëindigd. Als de werkloosheidsduur om verschillende redenen kan worden beëindigd, maar daar in statistieken geen onderscheid naar hoeft te worden gemaakt, is het strikt genomen niet nodig om dat expliciet in het model mee te nemen en kan de werkloosheidsduur met een gewoon duurmodel worden geanalyseerd. De uitstroombintensiteit (*hazard rate of exit rate*) kan dan namelijk betrekking hebben op de verschillende soorten transities vanuit werkloosheid tegelijk. Als het soort transitie voor de te ontwikkelen statistiek wél uitmaakt, moet gebruik worden gemaakt van een zogenaamd *competing risk* model. Dit is bijvoorbeeld het geval als de statistiek niet zozeer de werkloosheidsduur in kaart moet brengen, maar de tijd die het iemand die werkloos wordt kost om een nieuwe betaalde werkkring te vinden. *Competing risks* modellen zijn bij uitstek geschikt voor dit soort toepassingen. Het *competing risks* model is een generalisatie van het gewone duurmodel waarbij voor elk soort transitie vanuit werkloosheid een afzonderlijke *hazard rate* geldt en die transitiespecifieke *exit rates* bepalen gezamenlijk wat de kans is dat de duur in een bepaald tijdsinterval wordt beëindigd. Ook als het voor de statistieken niet vereist is om de manier van beëindigen te onderscheiden, biedt dit model toch een flexibelere specificatie, waarin de manier van beëindigen bij het schatten als additionele informatie wordt meegenomen. Bij het beantwoorden van de tweede vervolgvraag worden ook andere modelveronderstellingen tegen het licht gehouden, in het bijzonder de gebruikte functionele vorm voor duurzaamheidsduur en de rol van tijdvariërende variabelen.

Deze rapportage beschrijft het onderzoek voor de derde fase van het 'werkloosheidsduren' project en is als volgt opgebouwd. In het volgende hoofdstuk wordt kort ingegaan op duurmodellen en het schatten daarvan. Vervolgens worden in hoofdstuk 3 de gebruikte data beschreven. Hoofdstuk 4 bespreekt de schattingen van het duurmodel en het *competing risks* model. In hoofdstuk 5 wordt ingegaan op de stap van modelschattingen naar statistieken van de verwachte werkloosheidsduur alsmede op andere statistieken gebaseerd op directe tellingen van stromen. Hoofdstuk 6 sluit af met de belangrijkste conclusies en aanbevelingen.

2. Analyse van duren

2.1 Duurmodellen

Zoals in de inleiding al werd aangegeven zijn duren conceptueel niet eenvoudig te analyseren. Werkloosheidsstatistieken zijn daarom veelal gebaseerd op ‘voorraad-grootheden’ zoals het aantal werklozen of het percentage werklozen in de potentiële beroepsbevolking of in de bruto arbeidsdeelname. Weliswaar zijn dit pure momentopnames, maar het duuraspect van werkloosheid kan in dergelijke statistieken bijvoorbeeld worden geïntroduceerd door binnen het werklozenbestand groepen te onderscheiden op basis van het aantal jaren dat ze al werkloos zijn.

In het klassieke onderscheid tussen voorraadgrootheden, stroomgrootheden en duren kan voor het beschrijven van het dynamische aspect van werkloosheid ook de invalshoek van stroomgrootheden worden gekozen, bijvoorbeeld: het aantal werkloosheidsdagen in een periode (kwartaal of jaar) of van de mensen die 1 jaar geleden werkloos waren het percentage dat sindsdien een baan heeft gevonden of het aantal aaneengesloten maanden in het afgelopen jaar dat die mensen daarna nog werkloos bleven (maximaal 12). Dergelijke statistieken beschrijven slechts enkele aspecten van de werkloosheidsduur, maar hebben als voordeel dat ze duidelijk te plaatsen zijn in de tijd doordat de datum of de periode waarop zij betrekking hebben ondubbelzinnig vast ligt. Dat ligt bij de werkloosheidsduren zelf moeilijker. Daarom is de kernvraag van dit onderzoeksproject – nagaan of het mogelijk is om statistieken te ontwikkelen van de werkloosheidsduur zelf – uiterst relevant. In Berkhout *et al.* (2007) wordt hiervoor als hoofdreden genoemd dat het van belang is de dynamiek op de arbeidsmarkt goed in kaart te brengen, vooral waar standcijfers daarin tekort schieten doordat (1) standcijfers achterlopen bij de economische ontwikkelingen, (2) op grond van werkloosheidsduren de dynamiek van de arbeidsmarkt voor subgroepen beter in beeld kan worden gebracht en (3) het meten van individuele werkloosheidsduren noodzakelijk is “om de vanuit beleids oogpunt uiterst belangrijke groep langdurige werklozen te kunnen identificeren”.

De intrinsieke complexiteit van duren blijkt al bij het meten van duren en in het bijzonder bij het meten van de verdeling – kansverdeling, frequentieverdeling of populatieverdeling – van duren. Zoals in de inleiding al kort werd aangestipt zijn waargenomen duren vaak incompleet, bijvoorbeeld omdat niet bekend is hoelang iemand vóór een bepaald tijdstip al werkloos was, of omdat niet bekend is hoelang iemand ná een bepaald tijdstip nog werkloos is gebleven. Een tweede probleem staat bekend als *length biased sampling*, hoewel het niet alleen steekproeven, maar ook populaties betreft. Normaliter wordt onder de verdeling van duren de kansverdeling verstaan van een werkloosheidsduur die op een bepaald moment begint of – daaraan gerelateerd – de populatieverdeling van werkloosheidsduren die op een bepaald tijdstip aanvangen. Dergelijke duren worden wel *flow sampled* genoemd. Daar staan de zogenaamde *stock sampled* gegevens tegenover. Dat zijn werkloosheidsduren die op een bepaald tijdstip ‘actief’ zijn. Gegevens over de werkloosheidsduur van mensen die op een peildatum werkloos zijn vormen dus een *stock sample* en zijn vertekend in de zin dat lange duren daarin oververtegenwoordigd zijn en korte duren ondervertegenwoordigd ten opzichte van de *flow sampled* verdeling van duren die op een bepaald moment zijn aangevangen. Deze laatste verdeling is te prefereren voor het doen van uitspraken over de verdeling van werkloosheidsduren, maar vaak worden de daarvoor benodigde gegevens slechts waargenomen in de vorm van een *stock sample*, waardoor de werkloosheidsduur systematisch wordt overschat. Bovendien leidt de oververtegenwoordiging van langere duren daarbij tot een oververtegenwoordiging van mensen met kenmerken die samenhangen met relatief lange duren. Voor een deel kan dat bovendien gaan om niet waargenomen kenmerken.

Om deze complicaties bij het meten van werkloosheidsduren het hoofd te kunnen bieden hebben aan het einde van de jaren '70 de duurmodellen hun intrede gedaan in de arbeidseconomie (Lancaster, 1979; Lancaster, 1990).¹⁾ Deze modellen – later ook toege-

¹⁾ Zie Theeuwes, Kerkhofs en Lindeboom (1988) voor een eerste toepassing op Nederlandse gegevens.

past bij het analyseren van andere duren, zoals baanduren of vacatureduren – werden overigens al langer toegepast op onderzoekerreinen als de epidemiologie en de demografie. Een werkloosheidsduur wordt hierin beschreven als de uitkomst van een in de tijd voortschrijdend proces, waarin iemand die zich in een bepaalde situatie bevindt constant bloot staat aan de mogelijkheid een transitie te maken naar een andere toestand en daarmee de lopende duur in de huidige situatie te beëindigen. De kans dat zich op een bepaald moment zo'n gebeurtenis voordoet²⁾ wordt de uitstroomintensiteit of *hazard rate* genoemd. Deze kan variëren over de tijd (bijvoorbeeld door de invloed van de conjunctuur op de kans om weer werk te vinden), tussen individuen (op grond van meer of minder aantrekkelijke kenmerken of verschillen in zoekgedrag) en naar gelang de duur die iemand op dat moment al werkloos is (als gevolg van ontmoediging, stigmatisering of aflopen van uitkeringsrechten).

Door gebruik te maken van duurmodellen is het mogelijk om op statistisch verantwoorde manier rekening te houden met het soort gegevens dat wordt gebruikt en met het probleem van onvolledig waargenomen duren. Daarnaast bieden deze modellen ook de mogelijkheid om rekening te houden met verschillen tussen mensen op grond van kenmerken die niet zijn waargenomen (wat econometristen zijn gaan aanduiden met niet waargenomen heterogeniteit, maar in de epidemiologie bekend staat onder de noemer *frailty*). Dit is van belang omdat na 1 maand werkloosheid de samenstelling van de groep werklozen al anders is dan aan het begin: de kansrijksten hebben namelijk al weer werk gevonden. Hierdoor zal de samenstelling van de groep werklozen veranderen naarmate deze langer werkloos zijn.³⁾

De bouwsteen voor dergelijke modellen is de *hazard rate* of uitstroomintensiteit, die in modellen van duren aangeeft met welke waarschijnlijkheid zich binnen een korte periode een gebeurtenis voordoet waardoor de huidige duur wordt beëindigd. Bij het analyseren van werkloosheidsduren geeft de *hazard rate* dus de waarschijnlijkheid dat iemand die momenteel werkloos is in de komende tijd werk vindt (of om andere redenen ophoudt naar werk te zoeken). Dergelijke processen kunnen als discrete of als continue processen worden voorgesteld. Bijvoorbeeld bij maandabonnementsen op een tijdschrift is de tijd dat iemand abonnee is per definitie een geheel aantal maanden, bij het vervallen van atomen en andere kleine deeltjes gaat het daarentegen om een continu verlopend proces en kan een duur elk positief reëel getal zijn. Veelal wordt van een continu model uitgegaan als de bijbehorende tijdseenheid – de natuurlijke tijdseenheid voor definitie en waarneming – klein is t.o.v. de mogelijke duren. In het geval van werkloosheidsduren is de duur strikt genomen een discreet aantal dagen, maar is een substantieel deel van de duren langer dan een maand of langer dan een jaar. In dat geval ligt het voor de hand om in het model uit te gaan van een proces dat zich continu in de tijd afspeelt, ook al zijn de waargenomen duren niet nauwkeuriger gedefinieerd of waarneembaar dan in termen van een geheel aantal dagen. Gedefinieerd in een continu tijdsverloop hangt de kans dat een gebeurtenis zich in een gegeven periode voordoet mogelijk af van de tot dan toe verstreken duur en van andere relevante achtergrondkenmerken, kortweg aangeduid met x , maar uiteraard vooral van de lengte van de periode. De hazard is daarom voor een continu proces als volgt gedefinieerd:

$$\theta(d, x_t) \equiv \lim_{\Delta d \downarrow 0} \frac{\text{Kans op beëindiging van de huidige duur tussen } d \text{ en } d + \Delta d}{\Delta d}$$

en geeft aan hoeveel de beëindigingskans toeneemt met elke extra fractie van de tijd. Deze uitstroomintensiteit bepaalt volledig de kansverdeling van de duren. De kans dat een duur minstens d duurt, de *survival* functie, is dan:

$$S(d|\{x_t\}) \stackrel{\text{def}}{=} \Pr(\text{Duur} > d|\{x_t\}) = \exp\left(-\int_0^d \theta(s, x_{t_0+s}) ds\right)$$

²⁾ Vandaar ook de nauwe verwantschap tussen duurmodellen en methodes voor event history analyse (Singer en Willett, 2003).

³⁾ Overigens is een probleem met dit soort duurmodellen dat het niet goed mogelijk is om deze veranderende samenstelling gedurende de werkloosheid te onderscheiden van de manier waarop de uitstroomintensiteit verandert naarmate de werkloosheid langer duurt (duurafhankelijkheid).

Op vergelijkbare manier kunnen van allerlei kansen en kansdichtheden worden berekend die nodig zijn voor het schatten van de onbekende parameters van de *hazard* met behulp van de *maximum likelihood* methode (zie paragraaf 2.2).

De notatie hierboven laat een paar bestanddelen zien die bij het specificeren van een duurmodel van belang zijn:

- de duurzaamheidsduur: hoe hangt de uitstroomintensiteit af van de reeds volbrachte duur. [Bij werkloosheidsduren: hoe beïnvloedt de tijd die iemand al naar werk zoekt de kans op het spoedig vinden van een baan, bijvoorbeeld door stigmatisering, door ontmoediging, door verval van het *human capital*]
- heterogeniteit: hoe hangt de uitstroomintensiteit af van andere waargenomen of niet waargenomen kenmerken. [Bij werkloosheidsduren: hoe wordt de kans op het vinden van een baan beïnvloed door het opleidingsniveau, de leeftijd, de regio, of door iets moeilijk meetbaars als doorzettingsvermogen of fysieke aantrekkelijkheid.]
- omdat de duren tot stand komen in de loop van de tijd, kunnen de omstandigheden en relevante kenmerken in x veranderen, waardoor het mogelijk nodig is om onderdelen van x als tijdvariërende variabelen in het model op te nemen (soms aangeduid als niet-stationariteit). [Bij werkloosheidsduren bijvoorbeeld door veranderlijke kenmerken als gezinssamenstelling en leeftijd, maar mogelijk ook conjunctuurindicatoren of seizoenspatronen]

2.2 Het schatten van duurmodellen

Het specificeren van de uitstroomintensiteit als functie van duur d en tijdspatronen van achtergrondkenmerken $\{x_i\}$ ligt aan de basis van een modelmatige analyse van duren, maar daarnaast is er een aantal andere zaken dat het analyseren van duren compliceert.

Achtereenvolgens:

- (1) onvolledig waargenomen duren
- (2) *length-biased sampling*
- (3) niet waargenomen heterogeniteit

Onvolledig waargenomen duren

Duren ontstaan in de loop van de tijd en kunnen, afhankelijk van de manier waarop zij worden gemeten, gedeeltelijk buiten het *sampling frame* vallen of anderszins onvolledig worden waargenomen. Van de mensen die nu werkloos zijn, kun je te weten komen hoe lang ze al op zoek zijn naar werk, maar niet hoelang ze dat nog blijven. De totale duur tot het vinden van werk is dus onvolledig waargenomen, visueel voorgesteld is de duur aan de rechterkant afgekapt of, zoals het ook wel wordt aangeduid, 'rechts gecensureerd'.

Omgekeerd kan het voorkomen dat wél bekend is hoelang iemand vanaf een bepaalde datum werkloos is gebleven, maar niet hoelang hij of zij het daarvoor al was. Duren die onvolledig worden waargenomen door het ontbreken van deze 'beginduren' worden wel aangemerkt als 'links gecensureerd'.

Daarnaast kan het voorkomen dat duren niet exact kunnen worden gemeten, maar dat slechts op discrete momenten in de tijd wordt waargenomen of een duur nog voortduurt of tussen het vorige en het huidige peilmoment is beëindigd. Dit type gegevens wordt aangeduid als *grouped of interval censored durations* (Lancaster, 1990; Jenkins, 1995).

Laat vooropgesteld zijn dat volledig waargenomen duren het mooist zijn, maar ook met onvolledige duren zijn informatieve en correcte analyses van het onderliggende duurproces mogelijk.⁴⁾ Zaak is wel om daarbij de analysemogelijkheden zo goed als mogelijk te benutten, bijvoorbeeld door in het empirisch model de aard van de informatiebeper-

⁴⁾ Afhankelijk van de beschikbare data – volledige duren of gecensureerde duren, uit een *stock sample* of een *flow sample* – kunnen de parameters uit het onderliggende duurmodel worden geschat met de *maximum likelihood* methode. Dezelfde parameters kunnen dus op verschillende manieren en op basis van verschillende soorten informatie worden geschat door voor de waargenomen informatie de correcte *likelihood*-bijdragen (kansen en kansdichtheden) te gebruiken.

kingen expliciet te modelleren en niet de vertekend of onvolledig waargenomen duren als benadering van de echte duren te behandelen. Bij rechtsgecensureerde duren vormt de waargenomen duur een ondergrens van de werkelijke duur en deze informatie kan als *likelihood*-bijdrage worden gebruikt bij het schatten van de parameters van de onderliggende duurverdeling met de *maximum likelihood* methode.⁵⁾ In dat geval is de *likelihood* van een volledig waargenomen duur gelijk aan de kansdichtheid van die duur en voor rechts gecensureerde duren is de *likelihood*-bijdrage gelijk aan de kans op een duur van die lengte of langer. Deze kunnen als volgt afgeleid worden uit de uitstroombintensiteit θ . De verdelingsfunctie, de kans op een duur korter dan d , is:

$$F(d|\{x_t\}) \stackrel{\text{def}}{=} \Pr(\text{Duur} \leq d|\{x_t\}) = 1 - \exp\left(-\int_0^d \theta(s, x_{t_0+s}) ds\right)$$

en hieruit kan de kansdichtheid van de duur worden afgeleid als:

$$f(d|\{x_t\}) = \theta(d, x_{t_0+d}) \cdot \exp\left(-\int_0^d \theta(s, x_{t_0+s}) ds\right)$$

Bij linksgecensureerde duren is het informatietekort in het algemeen ernstiger. Als er duurzaamheidsduur is, is op geen enkel moment bekend hoe lang de duur al loopt en dus ook niet wat de waarschijnlijkheid is dat een duur al zo lang loopt. Hierdoor is het nodig om vanuit het theoretische model van het duurproces een *limited information* model af te leiden dat beschrijft hoe de duurverdeling er uitziet, van duren vanaf een bepaald peilmoment ongeacht de voorafgaande duur. Deze aanpak is zeer complex en vergt veel veronderstellingen over transitieprocessen in en uit werkloosheid dan nodig zijn als de 'beginduur' wel bekend is.⁶⁾

Bij intervalgecensureerde duren – zoals bijvoorbeeld bij het gebruik van de vervolgmetingen van de EBB – is de beste aanpak de kans op het waargenomene, namelijk dat de duur tussen een bekende onder- en bovengrens ligt, uit het onderliggende duurmodel af te leiden en als *likelihood*-bijdrage te gebruiken:

$$\Pr(d_0 \leq D \leq d_1|\{x_t\}) = \exp\left(-\int_{d_0}^{d_1} \theta(s, x_{t_0+s}) ds\right)$$

Een alternatieve aanpak is om bijvoorbeeld bij een duur waarvan bekend is dat deze minstens twee en hoogstens vijf maanden heeft geduurd, in het midden te gaan zitten en de duur op drieëneenhalve maand te schatten. Door deze duren als volledig waargenomen duren te analyseren, wordt bij het berekenen van de nauwkeurigheid van de geschatte parameters de informatie in de data te rooskleurig voorgesteld. Daarnaast is informatie toegevoegd op een manier die niet consistent is met het onderliggende model. Hoewel bij relatief korte intervallen de vertekeningen niet groot zullen zijn, verdient het toch de voorkeur om de eerste aanpak te volgen en de *likelihood* op maat te bepalen voor de beschikbare informatie.

Length-biased sampling

In het algemeen wordt de kansverdeling van duren opgevat als de kansverdeling van duren die op een bepaald moment in de tijd beginnen en waarvan het verloop vanaf dat moment wordt gevolgd tot voltooiing. Het daarmee overeenkomende steekproefkader is dat van een instroomsteekproef (of *flow sample*) waarbij personen willekeurig worden getrokken uit een bestand van mensen die op een bepaald moment (net) werkloos zijn geworden, of waarbij een willekeurige getrokken groep mensen wordt gevolgd totdat zij werkloos worden en vanaf dat instroommoment hun duur als werkloze wordt geregistreerd. Kenmerk van deze steekproeven is dat de trekkingskans niet door de lengte van

⁵⁾ Door het maximaliseren van de likelihood worden de onbekende parameter zodanig geschat dat de waarschijnlijkheid van het waargenomene zo groot mogelijk is. Bovendien geeft de theorie van de maximum likelihood methode standaard uitdrukkingen voor de nauwkeurigheid van de parameterschattingen (asymptotisch geldige standaardfouten en betrouwbaarheidsintervallen).

⁶⁾ Heckman en Singer (1984) geven een overzicht van dit soort econometrische problemen in duurmodellen en mogelijke oplossingen. Niet zonder betekenis bleken juist de formules voor stock sampled duren in dit overzicht onjuist (Heckman en Singer, 1985).

de werkloosheidsperiode wordt beïnvloed aangezien zij al in de steekproef zitten als die duur na instroom in werkloosheid nog moet gaan ontstaan. Als de steekproef hier niet aan voldoet en de trekkingskans wel door de lengte van de al lopende duur wordt beïnvloed, zal de kansverdeling er geheel anders uitzien. Het meest voorkomende steekproefdesign waarvoor dit geldt, is de zogenaamde *stock sample*, die in het geval van werkloosheidsduren tot stand komt door een willekeurige steekproef te trekken uit de mensen die op een bepaald tijdstip werkloos zijn. Personen met lange werkloosheidsduren hebben een grotere kans om in dit bestand te zitten dan werklozen met korte duren. En van verschillende trekkingen uit dezelfde duurverdeling hebben lange duren een grotere trekkingskans dan korte. Kortom: in een *stock sample* zijn lange duren oververtegenwoordigd en korte duren ondervertegenwoordigd en zijn mensen met kenmerken die gemiddeld genomen horen bij langere werkloosheidsduren oververtegenwoordigd en mensen met kenmerken die negatief correleren met de werkloosheidsduur ondervertegenwoordigd. Deze vertekening van de duurverdeling door selectiviteit van de steekproef op basis van de lengte van de duren wordt wel aangeduid als *length-based* of, minder neutraal, *length-biased sampling*.⁷⁾

De vertekening van de duurverdeling heeft in eerste instantie betrekking op de duur (de duur op het peilmoment en daarmee ook de volledige duur) en indirect op de samenstelling van de steekproef naar achtergrondkenmerken. Daarnaast wordt ook de plaatsing van de duren in de tijd vertekend waargenomen. Bij een steekproef in 2002 zijn duren die in 1998 zijn begonnen per definitie minstens 3 jaar lang. Als niet expliciet met de opbouw van de steekproef wordt rekening gehouden, zou daaruit ten onrechte kunnen worden geconcludeerd dat mensen die in 1998 werkloos worden onevenredig lang werkloos zijn gebleven en een extreem lage uitstroomintensiteit hebben.

Overigens is ook de duurverdeling in een *stock sample* een op zichzelf goed gedefinieerde duurverdeling en kan deze met behulp van duurmodellen worden geanalyseerd. Alleen moet daarbij worden opgepast met de interpretatie van deze duurverdeling. Zo kan de geschatte duurverdeling op basis van een *stock sample* niet zonder meer worden gebruikt om te voorspellen hoelang iemand die vandaag werkloos wordt werkloos zal blijven – gemiddelde, mediaan, kans dat het binnen drie maanden lukt om een baan te vinden – of om in een *stock sample* gevonden onvoltooid (rechtsgecensureerde) duren aan te vullen tot volledige duren. Beide toepassingen zijn relevant voor het ontwikkelen van statistieken op basis van schattingen van een duurmodel. De duurverdeling in een *stock sample* is overigens volledig bepaald door de duurverdeling van het onderliggende duurmodel voor *flow data*. Het is daarom in principe mogelijk om de ene verdeling om te rekenen naar de andere, zij het dat dit in het algemeen zeer ingewikkeld is en vergaande aanvullende veronderstellingen vergt over het proces dat de stromen van en naar werkloosheid bepaalt. Daarom is een alternatieve aanpak veruit te prefereren als er verdere informatie beschikbaar is waardoor het niet nodig is om het waargenomen proces terugwaarts te analyseren. De mogelijkheid daarvoor bestaat als de personen in de *stock sample* na de trekking van de steekproef nog een tijd worden gevolgd (al dan niet met tussenpozen). Het probleem van *length-biased sampling* is dat de steekproef selectief is met betrekking tot de duur die een persoon al werkloos was, niet met betrekking tot de duur die de persoon vanaf dat moment nog werkloos blijft. Deze laatste duur (de rest-duur) kan daarom wél direct worden gerelateerd aan de onderliggende duurverdeling van instromers. De analyse wordt dan een analyse van de werkloosheidsduur conditioneel op de duur die de werkloze op het moment van steekproefselectie al werkloos was. In die zin wordt de resterende duur de echte te verklaren variabele (endogene) van de analyse en de duur dat de persoon voor selectie al werkloos was, wordt een verklarende variabele (exogene). De aanpak waarin de duren op een dergelijke manier worden beschreven wordt wel aangeduid als *delayed entry* methode. De duurverdeling kan correct worden geschat door de *likelihood*-bijdrage te bepalen voor de beschikbare informatie over de duur na dat eerste peilmoment, bijvoorbeeld de volledige resterende duur, een rechtsgecensureerde vervolgduur of een of meer metingen op latere peildata. Dit kan worden

⁷⁾ In Bijlage B.1 wordt aan de hand van gesimuleerde duurgegevens geïllustreerd wat het effect van het steekproefontwerp op de waargenomen duurverdeling is.

gedaan door de *survival*-functie, verdelingsfunctie en kansdichtheid conditioneel op de op de eerste peildatum al verstreken duur d_0 te bepalen, respectievelijk:

$$S(d|\{x_t\}, D > d_0) = \exp\left(-\int_{d_0}^d \theta(s, x_{t_0+s}) ds\right)$$

$$F(d|\{x_t\}, D > d_0) = 1 - \exp\left(-\int_{d_0}^d \theta(s, x_{t_0+s}) ds\right)$$

$$f(d|\{x_t\}, D > d_0) = \theta(d, x_{t_0+d}) \cdot \exp\left(-\int_{d_0}^d \theta(s, x_{t_0+s}) ds\right)$$

De *likelihood*-bijdrage van een voltooide duur is dan de kansdichtheid $f(d|\{x_t\}, D > d_0)$, van een rechtsgecensureerde duur de survival-kans $S(d|\{x_t\}, D > d_0)$ en voor een *interval censored* duur zijn de *likelihood*-bijdragen:

$$\Pr(D \leq d_1|\{x_t\}, D > d_0) = \exp\left(-\int_{d_0}^{d_1} \theta(s, x_{t_0+s}) ds\right)$$

$$\Pr(D > d_1|\{x_t\}, D > d_0) = 1 - \exp\left(-\int_{d_0}^{d_1} \theta(s, x_{t_0+s}) ds\right)$$

voor duren die in het interval worden beëindigd en duren die in het interval nog niet worden beëindigd, respectievelijk.

De meest gebruikte specificatie van dit duurmodel is de *proportional hazard* specificatie:

$$\theta(d, x_t) = \psi(x_t) \cdot \lambda(d)$$

waarbij ψ aangeeft hoe de uitstroomintensiteit afhangt van de achtergrondkenmerken x en λ de duurzaamheidskans weergeeft. Een veelgebruikte specificatie voor de duurzaamheidskans is de Weibull-specificatie:

$$\theta(d, x_t) = \psi(x_t) \cdot \alpha \cdot d^{\alpha-1}, \text{ waarbij doorgaans } \psi(x_t) = \exp(x_t' \beta)$$

De parameter α moet positief zijn. Is α kleiner dan 1, dan daalt de uitstroomintensiteit naarmate de duur vordert (negatieve duurzaamheidskans), als α groter is dan 1, dan neemt de uitstroomintensiteit gedurende het proces toe (positieve duurzaamheidskans). Deze specificatie werd ook in het Berkhout *et al.* (2007) gebruikt, zij het dat daarin de achtergrondvariabelen x constant waren over de tijd, impliciet stationariteit veronderstellend. Daarnaast worden in de literatuur ook andere specificaties gebruikt. Als alternatief voor de Weibull-specificatie gebruiken we in dit onderzoek ook een specificatie van de duurzaamheidskans als stap-functie.

Niet waargenomen heterogeniteit

Een reden om parametrische duurmodellen te schatten, is om na te gaan hoe verschillen tussen personen zich vertalen in verschillende duurverdelingen. Dat kan in een empirische analyse uiteraard alleen voor kenmerken die worden waargenomen. Dat geldt voor alle regressietechnieken en daarvoor hebben dergelijke modellen een 'storingsterm' die dient als vergaarbak voor alle niet waargenomen kenmerken waarin waarnemingenseenheden kunnen verschillen. Zolang die storingstermen dan voldoen – of verondersteld worden te voldoen – aan de eis dat ze de uitkomsten van de analyses niet verstoren, is daarmee de kous af. Bij duurmodellen is dat niet zo en dat heeft alles te maken met het feit dat duren pas in de loop van de tijd ontstaan. Van de mensen die op een bepaald moment werkloos worden, vindt een deel weer snel een nieuwe baan. Deze groep is echter niet aselekt samengesteld en bestaat voor het overgrote deel uit mensen met 'gunstige' waargenomen of niet waargenomen kenmerken. Dat wil zeggen kenmerken die horen bij gemiddeld kortere werkloosheidsduren. Het gevolg is dat gaandeweg het duurproces de samenstelling van het gevolgde groep werklozen verandert – minder gunstig wordt – en dat als gevolg daarvan de uitstroomintensiteit kleiner wordt. Dit lijkt op de duurzaamheidskans uit paragraaf 2.1 maar is iets heel anders. Vooral voor beleid maakt het nogal wat uit of de afname van de uitstroomintensiteit uit werkloosheid, als oorzaak heeft dat de duur die iemand al werkloos is zelf de kansen op werk kleiner

maakt (de echte duurzaamheidskans) of dat de gemiddelde samenstelling van de overblijvers steeds slechter wordt. In het eerste geval moet je zorgen dat mensen zo snel mogelijk weer uit het werklozenbestand komen, bij de tweede verklaring kun je de succesvollen zichzelf laten redden en alle aandacht richten op de 'minder gunstige' groepen. Vandaar dat niet waargenomen heterogeniteit een belangrijk onderwerp is bij dit soort analyses. Bij de waargenomen heterogeniteit kun je met de veranderende samenstelling van een groep werklozen expliciet rekening houden, bij de niet waargenomen heterogeniteit uiteraard niet. Ten principale zijn duurzaamheidskans en compositie-effecten door niet waargenomen heterogeniteit niet te onderscheiden, maar dat kan bereikt worden door extra modelveronderstellingen te maken (bijvoorbeeld de veel gebruikte *proportional hazard* veronderstelling of veronderstellingen over de kansverdeling van de storings-term die de niet waargenomen heterogeniteit representeert).

Het is echter maar de vraag of dit identificatieprobleem ook echt wel een probleem is als het gaat om het ontwikkelen van statistieken. Bij de statistieken gaat het per slot van rekening niet om het beschrijven van het duurproces van een gegeven persoon (waarvan gedurende zijn werkloosheidsduur de hoogstvoltooide opleiding niet zal dalen), maar om het beschrijven van het duurproces van de groep werklozen (waarvan gedurende de werkloosheid het gemiddelde niveau van de hoogstvoltooide opleiding wel kan dalen en de gemiddelde leeftijd nog harder kan toenemen dan in het leven van alledag). Om die reden is het voor het maken van statistieken niet belangrijk de samenstellingseffecten door niet waargenomen heterogeniteit te onderscheiden van de duurzaamheidskans. De relevante uitstroomintensiteit van de mensen die drie maanden werkloos zijn is de gemiddelde uitstroomintensiteit van de groep mensen die na drie maanden nog geen werk heeft gevonden. De duurzaamheidskans hiervan mag zowel individuele duurzaamheidskans als compositie-effecten omvatten.

2.3 Competing risks modellen

In een *competing risks* model wordt een duurproces gemodelleerd waarin een lopende duur beëindigd kan worden door verschillende soorten gebeurtenissen, of als uitkomst van transities in verschillende richtingen. In het geval van werkloosheidsduren, kan het zoekproces succesvol worden afgesloten met het vinden van een nieuwe baan, maar kan de duur ook om andere redenen worden beëindigd. Dat kan het resultaat zijn van een bewuste keuze, bijvoorbeeld als iemand zich volledig gaat toeleggen op het huishouden of een opleiding gaat volgen. Ook overlijden of andere vormen van steekproefuitval kunnen als een alternatieve *exit route* worden gemodelleerd. In vergelijking met het *single risk* duurmodel analyseren de *competing risks* modellen dus niet alleen de waargenomen duren, maar ook de manier waarop deze worden afgesloten. Hierdoor wordt het mogelijk om een gedetailleerdere beschrijving te geven van het duurproces, waarbij de uitstroomintensiteit voor elke *exit route* afzonderlijk kan afhangen van de reeds verstreken duur, tijdvariërende factoren en persoonskenmerken. Dit maakt het bijvoorbeeld mogelijk om onderscheid te maken tussen de invloed van een conjuncturele verslechtering enerzijds op de kans dat een werkloze binnen afzienbare tijd een baan vindt, anderzijds op de kans dat deze de arbeidsmarkt verlaat zonder verder te zoeken. De waargenomen werkloosheidsduren geven daarvan slechts het netto-effect.

Bezien vanuit de *event history* benadering van het duurproces kunnen in een competing risks model verschillende *events* plaatsvinden, elk met een eigen onafhankelijke uitstroomintensiteit. De som van deze exit-specifieke uitstroomintensiteiten bepaalt de totale uitstroomintensiteit en daarmee de duurverdeling, de verhouding tussen de afzonderlijke uitstroomintensiteiten bepaalt de kans dat de duur met een transitie volgens de betreffende *exit route* wordt afgesloten. De uitstroomintensiteit voor *exit route j* is dan gedefinieerd als:

$$\theta_j(d, x_t) = \lim_{\Delta d \downarrow 0} \frac{\text{Kans op transitie naar } j \text{ tussen } d \text{ en } d + \Delta d}{\Delta d}$$

en de bijbehorende kansdichtheid van een duur van lengte d die wordt beëindigd met een transitie j is:

$$f(d, \text{exit} = j | \{x_t\}) = \frac{\theta_j(d, x_{t_0+d})}{\sum_i \theta_i(d, x_{t_0+d})} \cdot \exp\left(-\int_0^d \sum_i \theta_i(s, x_{t_0+s}) ds\right)$$

Twee verschillen daardoor: (1) de analyse gebruikt ook de informatie over hoe een duur is beëindigd (bijvoorbeeld: wel of niet door het vinden van een nieuwe baan) en (2) in het duurmodel wordt in plaats van één hazard rate voor de beëindiging van de duur nu voor elke beëindigingswijze een eigen hazard gedefinieerd. In feite heeft iedere beëindigingswijze zijn eigen virtuele event en de eerste gebeurtenis daarvan bepaalt wanneer de lopende duur wordt beëindigd en op welke wijze.

2.4 De gebruikte schattingsmethode

Uit de bespreking in paragraaf 2.2 volgt dat in de analyses in dit onderzoek niet expliciet rekening hoeft te worden gehouden met de niet waargenomen heterogeniteit, maar dat het wel belangrijk is dat de gebruikte schattingsmethode afgestemd is op de aard van de gebruikte informatie over werkloosheidsduren. De duurverdeling in paragraaf 2.1 wordt volledig bepaald door de uitstroomintensiteit θ en de daarbij behorende kansdichtheid en verdelingsfunctie. Die kunnen direct worden gebruikt om het duurmodel te schatten op grond van informatie uit een zogenaamde *flow sample*, ook als de duren *right censored* of *interval censored* worden waargenomen. Als de gegevens anders zijn, bijvoorbeeld afkomstig uit een *stock sample*, dan moeten *likelihood*-bijdrages worden gebruikt van de verdeling van de duren in een dergelijk steekproefontwerp. Dat geeft weliswaar een andere schattingsmethode, maar uiteindelijk worden daarmee de parameters van dezelfde duurverdeling geschat. Kortom: het geschatte model is hetzelfde, maar de gebruikte schattingsmethode dient afgestemd te zijn op de manier waarop de steekproef tot stand is gekomen en zal dus anders zijn voor een *flow sample* dan voor een *stock sample*. Voor de analyse van de lopende duren uit de eerste peilingen van de EBB zou daarom een analysetechniek moeten worden gebruikt die rekening houdt met het specifieke steekproefontwerp van een 'stock sample', wat zeer complex is en vergaande vereenvoudigende modelveronderstellingen vereist. De vervolopeilingen van de EBB maken het mogelijk het hele probleem van *length-biased sampling* te vermijden door de waargenomen informatie te analyseren als *delayed entry* duren. De lopende, nog niet voltooide duren uit de eerste EBB-peilingen (de CAPI interviews) worden als gegeven genomen (als exogene) en schattingen van de onbekende parameters van de duurverdeling ('hét duurmodel') worden gebaseerd op de informatie uit de daaropvolgende CATI metingen. Dit gebeurt in feite door de kans dat een lopende werkloosheidsduur voor de eerste peildatum wordt beëindigd uit het duurmodel af te leiden en de parameters daaruit zo te kiezen dat deze het best overeenstemmen met de feitelijk waargenomen beëindigingen en voortzettingen. Op deze manier wordt in de schattingsmethode expliciet rekening gehouden met het feit dat de duren in de beschikbare informatie *interval censored* zijn, d.w.z. dat niet de precieze duur vanaf de eerste peiling kennen, maar weten of de duur wel of niet voor de eerstvolgende peiling is beëindigd. Van de duren die op de tweede peildatum nog niet zijn beëindigd wordt vervolgens op dezelfde manier naar de periode tussen de tweede en de derde peiling bekeken en dit procedé herhaalt zich tot de vijfde peiling of zoveel eerder als de werkloosheidsduur wordt beëindigd. Deze aanpak maakt het ook mogelijk om nieuwe werkloosheidsduren, d.w.z. duren die na de eerste peiling en voor de vierde peiling ontstaan op dezelfde wijze te behandelen. Het tijdstip halverwege het interval waarin de duur begint, is daarbij als startmoment gedefinieerd.

In dit onderzoek wordt dus de *delayed entry*, *interval censored* methode als schattingsmethode gebruikt om de onbekende parameters in het onderliggende duurmodel met de beschikbare gegevens correct te schatten. Op basis van de parameterschattingen kan voor elke persoon met gegeven persoonskenmerken de verwachte werkloosheidsduur worden bepaald als deze nu werkloos zou worden, maar ook de verwachte resterende werkloosheidsduur als deze al een bepaalde tijd werkloos is geweest.

De gevolgde *delayed entry* aanpak heeft als bijkomend voordeel dat de invloed van de retrospectief vastgestelde duur van de werkloosheid op de eerste (CAPI) peiling maar

beperkte invloed heeft op de schattingen. Meetfouten die het gevolg kunnen zijn van het feit dat deze duren op herinnering zijn gebaseerd, hebben daardoor nauwelijks invloed⁸⁾ en ook reikt de veronderstelling dat het duurmodel stabiel is hierdoor niet oncontroleerbaar ver terug in de tijd. Bovendien is het met de gekozen aanpak ook niet nodig om een steekproef van 'baanvinders' (mensen met recent beëindigde duren) aan de steekproef van lopende duren toe te voegen.⁹⁾

Bij de *competings risks* analyse wordt in feite een generalisatie van bovenstaand duurmodel geschat en daarbij gelden dezelfde overwegingen voor het kiezen van een geschikte schattingsmethode. Voor het schatten van het duurmodel en het *competing risks* model met het soort gegevens dat hier wordt gebruikt is geen standaard software beschikbaar. Daarom zijn in dit project routines in Stata geprogrammeerd waarmee dit mogelijk is. Deze routines zijn uitgebreid getest met gesimuleerde data (zie Bijlage B1).

⁸⁾ Anders dan in Berkhout *et al.* (2007) zijn om deze reden de 'beginduren' hier niet gladgestreken (van pieken en afrondingen ontdaan).

⁹⁾ In Berkhout *et al.* (2007) zijn deze toegevoegd omdat het model niet geschat kon worden op een steekproef van alleen maar duren die nog niet zijn afgesloten. Het is echter zeer moeilijk om deze additionele steekproef correct te analyseren aangezien de trekkingskans van deze duren zowel van de duur afhangen als van het feit dat ze bijtijds (voor de peildatum) zijn afgesloten.

3. Data

3.1 Inleiding

De gebruikte data zijn afkomstig uit de EBB voor de periode 2002–2007 en bevatten ook de gegevens uit de telefonische vervolgpeilingen. De gebruikte steekproef bestaat uit alle respondenten die op de eerste peildatum, het CAPI interview, tussen de 15 en 64 jaar oud waren en waarvan de eerste peiling in de periode 2002–2007 plaatsvond en bevat de gegevens uit de eerste peiling en de vervolgpeilingen voor zover die in deze periode vielen.

3.2 Definitie werkloosheid en werkloosheidsduur

Net als in Berkhout *et al.* (2007) is de arbeidsmarktstatus bepaald op basis van de variabele beroepsbevolking in de EBB (BB_NED), die voor elke peildatum aangeeft of iemand werkt (BB_NED=1), werkloos is (BB_NED=2) of niet participeert (BB_NED=3). In het Berkhout *et al.* (2007) werden de scholieren (EDUCSTA=1 of 3) buiten beschouwing gelaten. Door het gebruik van de vervolgpeilingen kan de scholierstatus (EDUCSTA) nu echter van peiling tot peiling verschillen. Daarom zou voor een vergelijkbare aanpak in deze analyse de scholierstatus geïntegreerd moeten worden met de arbeidsmarktstatus BB_NED (1=werk, 2=werkzoekend werkloos, 3=overig niet werkend), door deze variabele aan te vullen met toestand '4=scholier'. Dit impliceert echter dat iemand niet gelijktijdig werkloos en scholier kan zijn, wat niet spoot met de definitie in de werkloosheidsstatistiek van het CBS. Hierin bevinden zich onder de werklozen namelijk ook scholieren, reden waarom in dit onderzoek scholieren in de analyses worden meegenomen.

In samenhang daarmee is de definitie van de werkloosheidsduur aangepast. In de eerdere fases van dit project is deze gedefinieerd als het minimum van 3 duren: (1) de duur sinds het verlaten van de dagopleiding, (2) de duur sinds het verlaten van de laatste baan en (3) de duur die de persoon volgens eigen opgave al op zoek is naar betaald werk. Nu ook scholieren werkloos kunnen zijn, vervalt de eerste van de drie samenstellende duren. De resulterende definitie – als minimum van de duur die de persoon al zoekt en de duur sinds het verlaten van de laatste baan – komt ook overeen met de internationaal bij statistische bureaus en Eurostat gehanteerde definitie van de werkloosheidsduur. Aangezien de op de eerste peildatum al verstreken werkloosheidsduren slechts als afhankelijke variabele worden gemodelleerd, zijn de retrospectief gemeten duren uit de EBB direct gebruikt en, anders dan in Berkhout *et al.* (2007), niet eerst in verband met 'pieken' gladgestreken.

Verder zijn voor het schatten van het duurmodel en het competing risks model alleen de personen gebruikt die op de eerste peildatum van de EBB, het CAPI interview, werkloos waren of op de tweede, derde of vierde peildatum werkloos blijken te zijn geworden.¹⁰⁾ Deze personen worden in de daaropvolgende peilingen gevolgd om na te gaan of zij daarin nog steeds werkloos zijn. Als een duur in het interval tussen twee peilmomenten is beëindigd, is overigens niet bekend op welk datum dat precies gebeurde. In de toegepaste *interval censored* schattingsmethode wordt daarom alleen de informatie gebruikt of de duur wel of niet in het betreffende tijdsinterval werd afgesloten. Wel wordt impliciet verondersteld dat personen die op twee opeenvolgende peilingen werkloos zijn de hele periode tussen die twee peildata werkloos zijn geweest. Korte onderbrekingen van de werkloosheidsduur tussen twee peildata worden aldus buiten beschouwing gelaten.

¹⁰⁾ Van respondenten die 1 of meer peilingen hebben overgeslagen wordt alleen de informatie tot de eerste onderbreking gebruikt, d.w.z.: geen aanvulling van de ontbrekende patronen. Na een onderbreking aangevangen nieuwe werkloosheidsduren zijn wel weer meegenomen.

3.3 Enkele beschrijvende statistieken

Tabel 1 geeft een overzicht van het aantal waarnemingen per peiling, uitgesplitst naar het jaar waarin de eerste meting plaatsvond. De eerste kolom geeft het totale aantal respondenten, de tweede kolom het aantal mensen dat op de peildatum werkloos is. De derde kolom geeft het aantal werkloze respondenten als daaruit de scholieren worden weggelaten.

Tabel 1
Aantal personen in analysebestand, per meting en jaar van eerste meting

Jaar van eerste meting	Meting	Allen	Werkloos	Werkloos & geen scholier
2002	1	78 494	1 970	1 637
	2	57 140	1 287	1 030
	3	53 011	1 242	1 030
	4	49 935	1 230	1 035
	5	48 310	1 222	1 032
2003	1	80 370	2 590	2 180
	2	61 416	1 842	1 542
	3	58 203	1 838	1 551
	4	55 583	1 682	1 428
	5	52 802	1 688	1 425
2004	1	90 489	3 499	2 947
	2	70 471	2 435	2 005
	3	65 814	2 231	1 849
	4	61 939	2 099	1 741
	5	59 256	1 874	1 571
2005	1	86 589	3 476	2 931
	2	72 192	2 335	1 944
	3	67 862	2 046	1 720
	4	64 396	1 853	1 531
	5	62 131	1 603	1 313
2006	1	78 249	2 869	2 344
	2	66 954	1 837	1 500
	3	62 610	1 563	1 241
	4	59 184	1 378	1 094
	5	56 592	1 269	1 026
2007	1	76 780	2 228	1 778
	2	48 812	1 084	849
	3	29 845	579	463
	4	13 936	257	217
	5			
Totaal	1	490 971	16 632	13 817
	2	376 985	10 820	8 870
	3	337 345	9 499	7 854
	4	304 973	8 499	7 046
	5	279 091	7 656	6 367

De aantallen voor de eerste peiling laten een duidelijk conjunctureel patroon zien voor het aantal werklozen, al is dat op het oog iets sterker door een gelijk opgaande ontwikkeling van de totale steekproefomvang. Als percentage van de totale steekproef bedraagt de werkloosheid ongeveer 2,5 procent in 2002, 3,2 in 2003, 3,9 procent in 2004, 4,0 procent in 2005, 3,7 procent in 2006 en 2,9 procent in 2007. Dit patroon laat zich goed door-trekken in de vervolgpelingen. Dit patroon onderstreept het belang van het gebruik van conjuncturele indicatoren of jaardummy's in het te schatten duurmodel.

De vergelijking van de aantallen tussen de opeenvolgende peilingen geeft een indruk van de paneluitval. In de analyses zullen wij uitgaan van de veronderstelling dat de uitval aselekt is. Van respondenten waarvan de eerste peiling in 2007 plaatsvond, heeft een deel van de vervolgpelingen pas na 2007 plaatsgevonden, waardoor die informatie in het analysebestand ontbreekt. De sterke daling van het aantal waarnemingen van dit cohort in opeenvolgende metingen is daarom hoofdzakelijk het gevolg van het steekproefkader en maar voor een deel van steekproefuitval.

Tabel 2 geeft de longitudinale tegenhanger van de transversale gegevens in Tabel 1. Hiervoor is het analysebestand getransformeerd naar een persoon-periode bestand waarin de waarnemingseenheden de tijdsintervallen zijn die in de analyse worden gebruikt. Per respondent zijn dat maximaal vier intervallen tussen twee opeenvolgende peildata. Dit longitudinale bestand bevat in totaal 1 276 498 periodes, waarvan er 37 848 beginnen in een situatie van werkloosheid.

Voor deze periodes wordt in Tabel 3 nagegaan in hoeveel daarvan een transitie vanuit werkloosheid naar een andere arbeidsmarkttoestand plaatsvindt, in het bovenste paneel in aantallen, in het onderste paneel als percentages. Ook hier zien we weer een sterk conjunctuurpatroon. Gemiddeld genomen wordt in ruim de helft van de periodes (van gemiddeld 3 maanden) een transitie van werkloosheid naar een andere toestand gemaakt. Iets vaker naar niet-participatie, maar toch in 25,7 procent van de gevallen naar een nieuwe baan.

Tabel 2
Aantal tijdsintervallen (persoon x periode) gedefinieerd als periode tussen twee opeenvolgende peilingen, per meting en jaar van eerste meting

Jaar van eerste meting	Metingen	In beide metingen	In beide metingen en werkloos in de eerste
2002	1-2	57 140	1 436
	2-3	50 977	1 111
	3-4	47 970	1 095
	4-5	46 504	1 126
	<i>totaal</i>	<i>202 591</i>	<i>4 768</i>
2003	1-2	61 416	1 968
	2-3	55 981	1 631
	3-4	53 758	1 666
	4-5	51 590	1 554
	<i>totaal</i>	<i>222 745</i>	<i>6 819</i>
2004	1-2	70 471	2 668
	2-3	64 886	2 206
	3-4	61 397	2 032
	4-5	58 723	1 937
	<i>totaal</i>	<i>255 477</i>	<i>8 843</i>
2005	1-2	72 192	2 799
	2-3	66 695	2 077
	3-4	63 118	1 862
	4-5	61 003	1 726
	<i>totaal</i>	<i>263 008</i>	<i>8 464</i>
2006	1-2	66 954	2 362
	2-3	61 279	1 628
	3-4	57 936	1 404
	4-5	55 464	1 280
	<i>totaal</i>	<i>241 633</i>	<i>6 674</i>
2007	1-2	48 812	1 397
	2-3	28 793	626
	3-4	13 439	257
	4-5		
	<i>totaal</i>	<i>91 044</i>	<i>2 280</i>
Totaal	1-2	376 985	12 630
	2-3	328 611	9 279
	3-4	297 618	8 316
	4-5	273 284	7 623
	<i>totaal</i>	<i>1 276 498</i>	<i>37 848</i>

Tabel 3
Transities vanuit werkloosheid in het persoon-periode bestand: aantal personen in EBB die aan begin van periode werkloos zijn, uitgesplitst naar situatie aan het eind van de periode ¹⁾

Jaar van begin van interval	Arbeidsmarktpositie (aan eind van interval)			
	Werkend	Werkloos	Niet participierend	Totaal
<i>aantallen</i>				
2002	873	1 160	975	3 008
2003	1 459	2 774	1 683	5 916
2004	1 897	4 252	2 321	8 470
2005	1 991	4 181	2 483	8 655
2006	1 873	3 266	2 291	7 430
2007	1 269	1 664	1 436	4 369
<i>Totaal</i>	<i>9 362</i>	<i>17 297</i>	<i>11 189</i>	<i>37 848</i>
<i>rij-percentages</i>				
2002	29,00	38,60	32,40	100,00
2003	24,70	46,90	28,40	100,00
2004	22,40	50,20	27,40	100,00
2005	23,00	48,30	28,70	100,00
2006	25,20	44,00	30,80	100,00
2007	29,00	38,10	32,90	100,00
<i>Totaal</i>	<i>24,70</i>	<i>45,70</i>	<i>29,60</i>	<i>100,00</i>

¹⁾ Periodes afgebakend door twee opeenvolgende metingen van een persoon in EBB.

In Tabel 4 wordt de vergelijkbare informatie gegeven als de intervallen worden uitgesplitst naar het nummer van de peiling. Daaruit blijkt dat tussen de eerste en tweede peiling vaker een beëindiging van de werkloosheid plaatsvindt, dan tussen latere peilingen. Dit patroon blijkt zich in elk van de jaren in de steekproef voor te doen en verdient nadere bestudering.

Tabel 4
Transities vanuit werkloosheid in het persoon-periode bestand, 2002–2007; aantal personen in de EBB die aan begin van de periode werkloos zijn, uitgesplitst naar de situatie aan het eind van de periode ¹⁾

Interval	Arbeidsmarktpositie (aan eind van interval)			Totaal
	werkend	werkloos	niet participierend	
<i>aantallen</i>				
1–2	3 256	5 294	4 080	12 630
2–3	2 348	4 227	2 704	9 279
3–4	1 926	4 067	2 323	8 316
4–5	1 832	3 709	2 082	7 623
<i>Totaal</i>	<i>9 362</i>	<i>17 297</i>	<i>11 189</i>	<i>37 848</i>
<i>rij-percentages</i>				
1–2	25,8	41,9	32,3	100
2–3	25,3	45,6	29,1	100
3–4	23,2	48,9	27,9	100
4–5	24,0	48,7	27,3	100
<i>Totaal</i>	<i>24,7</i>	<i>45,7</i>	<i>29,6</i>	<i>100</i>

¹⁾ Periode= interval afgebakend door twee opeenvolgende metingen van een persoon in EBB.

Bij de methode die in dit onderzoek wordt gebruikt om de onbekende parameters van het duurmodel te schatten, zijn de waarnemingseenheden eigenlijk de intervallen uit Tabel 3 en Tabel 4, d.w.z. de periodes tussen twee opeenvolgende EBB-peilingen voor zover de betreffende persoon aan het begin werkloos was en de situatie aan het eind van het interval bekend is. Voor dit doel is op basis van het oorspronkelijke analysebestand een *long* bestand gemaakt, met daarin informatie over alle 37 848 van dergelijke periodes. Bijlage B.2 geeft een overzicht van de in de analyses gebruikte variabelen in dit bestand.

4. Schattingen

4.1 Schattingen van het duurmodel

Tabel 5 geeft de parameterschattingen van het duurmodel, zoals beschreven in hoofdstuk 2. In de schattingen is gebruik gemaakt van de variabelen uit Tabel B2.1 in Bijlage 2. Deze variabelen betreffen kenmerken als leeftijd, geslacht, opleiding en gezinssamenstelling. Daarnaast zijn ook indicatoren opgenomen voor de maand en het jaar waarin de duur loopt. Deze laatste variabelen hebben op de aangegeven momenten invloed op de uitstroomintensiteit. Zonder hiervoor variabelen in het model op te nemen, wordt impliciet uitgegaan van stationaire omstandigheden, wat gezien de conjunctuurgebonden ontwikkelingen op de arbeidsmarkt geen acceptabele veronderstelling is. Aan deze kritiek kan daarom eenvoudig tegemoet worden gekomen door:

- dummy-variabelen voor de maand en
- dummy-variabelen voor het jaar.

aan de basisspecificatie toe te voegen. Daarnaast kan door het opnemen van dummy-variabelen voor duur-categorieën een alternatieve vorm van duurzaamheidsafhankelijkheid in het model worden geïntroduceerd, naast of in plaats van de Weibull duurzaamheidsafhankelijkheid. In feite wordt de *hazard* op die manier ook in de vorm van een stapfunctie afhankelijk van de reeds verstreken duur. In Tabel 5 worden eerst de schattingen zonder duurzaamheidsafhankelijkheid gegeven. Daarna de schattingen met Weibull duurzaamheidsafhankelijkheid en daarna met een stapfunctie voor de duurzaamheidsafhankelijkheid daaraan toegevoegd.

De parameters voor variabelen als leeftijd, geslacht, herkomstgroepering en regio hebben de verwachte tekens; opleiding vertoont, enigszins verrassend, geen duidelijk systematisch verband. Het verband tussen de uitstroomintensiteit en leeftijd is overigens niet monotoon dalend, maar bereikt zijn minimum bij de leeftijdscategorie van 50 tot en met 54 jaar. Bij hogere leeftijden is ook de uitstroomintensiteit weer hoger, wat wijst op een selectieve samenstelling van de groep mensen die na hun 55-ste op zoek is naar werk. Werklozen met een partner hebben een hogere uitstroomintensiteit, ongeacht of de partner een baan heeft. Verder blijkt, net als in Berkhout *et al.* (2007), dat de arbeidsmarktpositie voorafgaand aan de werkloosheid – op basis van het onderscheid tussen baanwisselaars, herintreders, schoolverlaters en overigen – een sterk, zij het niet excessief effect heeft. Mensen die vanuit een baan werkloos zijn geworden, hebben een significant lagere uitstroomintensiteit. Ook het zoekgedrag heeft een sterke significante samenhang met de uitstroomintensiteit, maar met tegendraadse tekens. Hoe meer gezocht is, hoe kleiner de kans op beëindiging. Dat geldt zowel voor de variabele die aangeeft of er recent is gezocht, als voor de ontplooiende zoekspanningen. Dit wijst op een vorm van selectiviteit aangezien de zoekvariabelen betrekking hebben op de periode van vier weken voorafgaand aan de beschreven periode en de personen in de schattingssteekproef aan het begin van de periode werkloos zijn. Als je actief hebt gezocht en desondanks nog werkloos bent, is de kans groot dat je andere voor de arbeidsmarkt minder aantrekkelijke kenmerken hebt, wat wellicht ook de reden was om actiever te zoeken. De duurzaamheidsafhankelijkheid is, zoals verwacht mocht worden, keurig negatief.

Tabel 5
Parameterschattingen van duurmodel, met verschillende specificaties van duurzaamheidsafhankelijkheid

Specificatie duurzaamheidsafhankelijkheid:	Geen		Weibull		Weibull en stapfunctie	
	coëfficiënt	t-waarde	coëfficiënt	t-waarde	coëfficiënt	t-waarde
<i>Geslacht (ref.: man)</i>						
vrouw	0,112	6,20	0,113	6,21	0,113	6,22
<i>Leeftijd (ref.: 35–39)</i>						
15–19	0,710	16,09	0,599	13,42	0,581	13,00
20–24	0,582	15,96	0,494	13,42	0,477	12,92
25–29	0,362	10,76	0,311	9,20	0,302	8,91
30–34	0,154	4,85	0,136	4,29	0,134	4,23
40–44	–0,084	–2,77	–0,076	–2,49	–0,078	–2,54
45–49	–0,164	–4,96	–0,140	–4,24	–0,141	–4,25
50–54	–0,270	–7,56	–0,234	–6,55	–0,240	–6,70
55–59	–0,232	–6,04	–0,176	–4,55	–0,186	–4,80
60–64	0,093	1,68	0,153	2,74	0,116	2,08

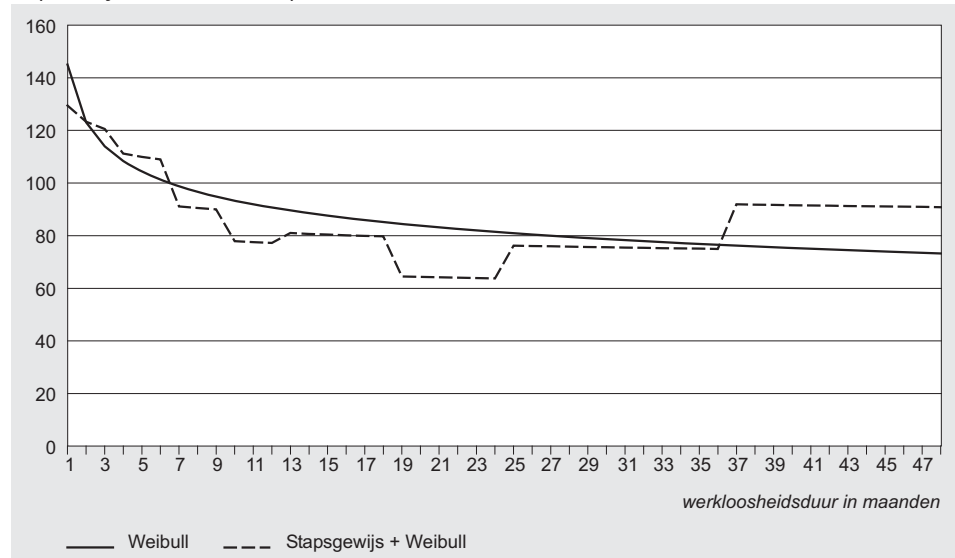
Tabel 5 (slot)
Parameterschattingen van duurmodel, met verschillende specificaties van duurzaamheidskarakteristiek

Specificatie duurzaamheidskarakteristiek:	Geen		Weibull		Weibull en stapfunctie	
	coëfficiënt	t-waarde	coëfficiënt	t-waarde	coëfficiënt	t-waarde
<i>Opleiding (ref.: mbo, havo, vwo)</i>						
basisonderwijs	-0,053	-1,71	-0,031	-0,99	-0,034	-1,10
vmbo	-0,044	-2,30	-0,034	-1,75	-0,037	-1,93
hbo	0,006	0,23	-0,002	-0,07	-0,005	-0,19
wo	-0,067	-2,10	-0,072	-2,24	-0,073	-2,29
<i>Beroepsrichting (ref.: economisch)</i>						
techniek	0,090	3,51	0,086	3,33	0,088	3,40
verzorgend/sociaal-cultureel	0,120	5,12	0,116	4,94	0,113	4,80
overige	0,128	5,28	0,121	4,98	0,121	4,95
<i>Herkomstgroepering (ref.: autochtoon)</i>						
westers allochtoon	-0,037	-1,37	-0,034	-1,26	-0,034	-1,27
niet-westers allochtoon	-0,140	-5,51	-0,125	-4,93	-0,115	-4,54
<i>Gezinsituatie(ref.: alleen)</i>						
paar	0,191	3,84	0,169	3,39	0,179	3,57
ander lid	0,052	1,05	0,051	1,02	0,054	1,09
ln(# personen)	0,086	1,53	0,079	1,41	0,076	1,35
ln(1+# kinderen 0- 5)	-0,123	-2,93	-0,120	-2,84	-0,120	-2,86
ln(1+# kinderen 6-11)	-0,018	-0,46	-0,006	-0,15	-0,009	-0,23
ln(1+# kinderen 12-17)	0,086	2,29	0,087	2,32	0,090	2,39
ln(1+# kinderen 18+)	0,026	0,64	0,026	0,64	0,024	0,60
Partner werkt	-0,007	-0,26	-0,006	-0,22	-0,009	-0,37
<i>Opleiding partner (ref.: midden)</i>						
laag	-0,024	-0,97	-0,019	-0,76	-0,019	-0,78
hoog	0,017	0,67	0,012	0,48	0,016	0,61
<i>Woonregio (ref.: Midden-Nederland)</i>						
Groningen	-0,145	-3,39	-0,129	-3,00	-0,126	-2,95
Friesland	-0,096	-2,19	-0,075	-1,70	-0,071	-1,60
Drenthe	-0,097	-2,12	-0,084	-1,85	-0,088	-1,93
IJssel-Vecht/Twente	-0,105	-2,65	-0,091	-2,30	-0,091	-2,30
IJssel-Veluwe	-0,043	-0,95	-0,036	-0,78	-0,032	-0,69
Arnhem/O-Gelderl./Nijm./Riv.l.	-0,088	-2,30	-0,080	-2,07	-0,078	-2,03
Flevoland	-0,197	-3,81	-0,179	-3,45	-0,168	-3,23
Noord-Holland Noord	0,133	2,76	0,135	2,81	0,136	2,82
Zuidelijk Noord-Holland	-0,046	-1,18	-0,037	-0,93	-0,037	-0,93
Rijnstreek	-0,020	-0,41	-0,026	-0,55	-0,028	-0,60
Haaglanden	-0,023	-0,52	-0,021	-0,49	-0,027	-0,62
Rijnmond	-0,090	-2,48	-0,079	-2,19	-0,078	-2,16
Zeeland	-0,081	-1,43	-0,065	-1,13	-0,071	-1,24
Midden en West Brabant	-0,040	-0,99	-0,033	-0,82	-0,034	-0,84
Noordoost Brabant	0,064	1,34	0,065	1,38	0,068	1,43
Zuidoost Brabant	-0,081	-1,87	-0,073	-1,67	-0,072	-1,65
Limburg	-0,138	-3,79	-0,122	-3,34	-0,120	-3,28
<i>Soort werkloze (ref: voorheen werkzaam)</i>						
herintreder	0,195	11,12	0,155	8,73	0,118	6,56
voorheen scholier	0,074	2,35	0,143	4,49	0,158	4,94
overig	0,166	4,46	0,292	7,68	0,249	6,48
<i>Zoeken naar werk</i>						
In afgelopen 4 weken gezocht	-0,188	-6,70	-0,219	-7,76	-0,202	-7,13
Zoekactiviteiten factor 1	-0,243	-24,44	-0,235	-23,55	-0,232	-23,21
Zoekactiviteiten factor 2	-0,025	-3,20	-0,025	-3,19	-0,024	-3,02
<i>Jaardummies (ref.: 2005)</i>						
2002	0,221	6,78	0,189	5,80	0,182	5,57
2003	-0,005	-0,19	-0,029	-1,13	-0,028	-1,08
2004	-0,120	-5,19	-0,129	-5,58	-0,123	-5,29
2006	0,102	4,44	0,106	4,60	0,101	4,39
2007	0,205	7,94	0,200	7,72	0,187	7,20
<i>Maandummies (ref.: september)</i>						
januari	-0,226	-6,16	-0,235	-6,39	-0,243	-6,61
februari	-0,071	-1,79	-0,076	-1,90	-0,080	-2,01
maart	-0,030	-0,86	-0,036	-1,04	-0,041	-1,19
april	-0,013	-0,37	-0,017	-0,47	-0,019	-0,55
mei	-0,029	-0,79	-0,035	-0,96	-0,042	-1,17
juni	-0,001	-0,04	-0,001	-0,03	-0,003	-0,07
juli	-0,042	-1,19	-0,037	-1,05	-0,034	-0,96
augustus	-0,051	-1,46	-0,055	-1,54	-0,053	-1,50
oktober	-0,017	-0,48	-0,017	-0,48	-0,021	-0,60
november	-0,101	-2,58	-0,102	-2,60	-0,098	-2,48
december	-0,188	-5,22	-0,194	-5,38	-0,196	-5,44
<i>Stapfunctie duurzaamheidskarakteristiek (ref.: 3-6 maanden)</i>						
0- 3 maanden					0,066	2,92
6- 9 maanden					-0,172	-5,25
9-12 maanden					-0,310	-6,43
12-18 maanden					-0,260	-4,93
18-24 maanden					-0,471	-6,74
24-36 maanden					-0,291	-3,95
24 maanden en langer					-0,086	-0,94
Constante	-1,631	-28,89	-1,127	-17,7	-1,335	-14,43
<i>Weibull duurzaamheidskarakteristiek</i>						
α			0,85	98,07	0,956	31,29
$\alpha-1$			-0,15	-17,3	-0,044	-1,45
N (# waarnemingen)	36 595		36 595		36 595	
AIC	45 918,90		45 617,60		45 454,40	
BIC	46 505,90		46 213,10		46 109,50	
k (# parameters)	69		70		77	
Log likelihood	-22 890,50		-22 738,80		-22 650,20	
χ^2	4 680,80		3 896,10		4 080,40	

Noot: schaduwkleur geeft aan dat de betreffende coëfficiënt op 5%-niveau significant van nul verschilt. De χ^2 -waarde heeft betrekking op de statistische toets van de hypothese dat alle parameters behalve de constante term nul zijn. Dit is een formele op de verklaringskracht van het model, waarbij de hypothese met 95% significantie wordt verworpen bij kritieke waarden 88,3, 89,4 en 97,4 bij drie schattingen in de tabel, respectievelijk.

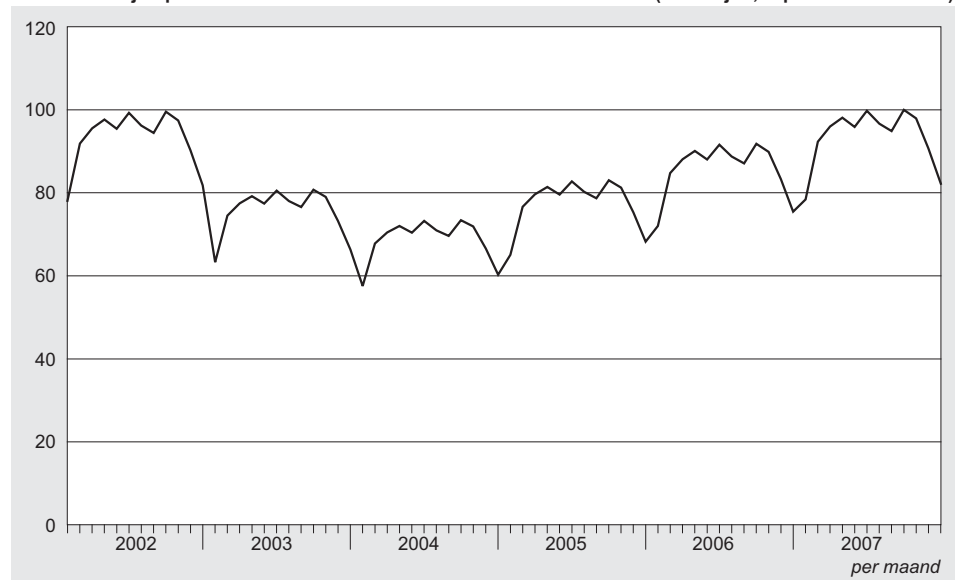
In de derde set schattingen is de Weibull duurzaamheidsindex aangevuld met een stapfunctie met als durercategorieën 0–3 maanden, 3–6 maanden, 6–9 maanden, 9–12 maanden, 12 tot 18, 18 tot 24 maanden, twee tot drie jaar en drie jaar of langer. In deze specificatie blijkt de Weibull-coëfficiënt niet meer significant van 1 te verschillen. Figuur 1 vergelijkt de duurzaamheidsindex van de hazard rate in de specificaties van de tweede en derde kolom en Figuur 2 toont de maand- en jaarpatronen in de laatste schattingen in Tabel 5.

1. Duurzaamheidsindex met Weibull specificatie en met combinatie van Weibull specificatie en stapfunctie (indexcijfer, 6 maanden =100)



De geschatte stapfunctie laat in het eerste jaar een negatieve duurzaamheidsindex zien, maar na pakweg anderhalf à twee jaar bereikt de uitstroomintensiteit het laagste niveau en lijkt de gemiddelde uitstroomintensiteit weer iets op te leven. Voor de dueren tot twee jaar lijkt de Weibull specificatie te volstaan, maar bij langere dueren geeft deze een iets te pessimistisch beeld. De staart van de duurverdeling wordt daarom beter beschreven door de niet-monotone stapfunctie.

2. Maand- en jaarpatronen in de uitstroomintensiteit vanuit werkloosheid (indexcijfer, september 2007 =100)



De tijdspatronen in Figuur 2 laten zowel duidelijke seizoenspatronen als conjuncturele schommelingen zien. De laatste weerspiegelen een scherpe neergang tussen 2002 en 2004, gevolgd door een periode van herstel in de periode van 2005 tot en met 2007. Bij de seizoenspatronen zijn de uitstroomkansen in januari en december aanzienlijk lager dan in de rest van het jaar.

4.2 Schattingen van het competing risks model

Bij het duurmodel uit de vorige paragraaf is geen onderscheid gemaakt tussen de verschillende manieren waarop een werkloosheidsduur kan worden beëindigd. Om onderscheid te maken tussen beëindiging vanwege het vinden van een nieuwe werkkring en beëindiging zonder dat werk is gevonden kan het duurmodel worden gegeneraliseerd tot een *competing risks* model. De duren worden daarbij gemodelleerd als de resultante van twee processen die tot een transitie kunnen leiden. De eerste transitie van deze transities die zich voordoet, beëindigt de lopende werkloosheidsduur en bepaalt de aard van de beëindiging, i.c. met dan wel zonder nieuwe baan.

Elk van de twee hazards kan op vergelijkbare manier worden gespecificeerd als in het gewone duurmodel.¹¹⁾

Doordat de parameters van beide hazards sterk verschillen, wijken de schattingen uiteraard ook af van die voor het *single risk* duurmodel uit de vorige paragraaf. Afgezien van de informatie over de manier waarop een werkloosheidsduur is afgesloten, is bij beide schattingen dezelfde informatie gebruikt. Bij nadere inspectie blijken de schattingen in kwalitatieve zin dan ook goed op elkaar aan te sluiten in de zin dat de schattingen uit de vorige paragraaf het netto-effect geven van de geschatte effecten voor de beide in deze paragraaf onderscheiden uitstroomintensiteiten. De afzonderlijke behandeling van de twee manieren waarop een werkloosheidsduur kan worden beëindigd, biedt op een aantal punten een nadere duiding van de eerder gevonden verbanden. Allereerst is dat het geval bij de verschillen tussen mannen en vrouwen. Alles bijeen genomen, bleken vrouwen in de vorige paragraaf een hogere uitstroomintensiteit te hebben dan mannen.

De schattingen in Tabel 6 laten zien dat de uitstroomintensiteit naar een nieuwe baan bij vrouwen juist lager is, maar dat de hogere uitstroomintensiteit voor beëindiging zonder nieuwe baan de verklaring is voor de gemiddeld genomen kortere werkloosheidsduren bij vrouwen. Ook het leeftijds patroon krijgt in de *competing risks* schattingen een genuanceerdere duiding. De relatie tussen leeftijd en de uitstroom naar een nieuwe baan blijkt sterk en monotoon dalend. Dat de uitstroomintensiteit in het duurmodel na de leeftijd van 54 jaar weer toenam, blijkt volledig voor rekening te komen van de uitstroom naar niet-participatie. De relatie tussen werkloosheidsduur en opleidingsniveau blijkt eveneens genuanceerder te liggen dan de schattingen van duurmodel konden lieten zien. Er is een duidelijk positief verband tussen het opleidingsniveau en de uitstroomintensiteit naar een nieuwe baan, maar daar staat een tegengesteld verband tegenover voor het verband met de uitstroomintensiteit naar niet-participatie. De geschatte parameters voor de beroepsrichting laten zien dat de hogere uitstroomintensiteit voor verzorgende beroepen in vergelijking met economische beroepen volledig voor rekening komt voor de uitstroom naar niet-participatie.

Voor de technische beroepen heeft de hogere uitstroomintensiteit wel degelijk te maken met een grotere kans op het vinden van een nieuwe baan. Daarnaast blijkt dat de lagere uitstroomintensiteit voor niet-westerse allochtonen in het duurmodel geflatteerd werd door een hogere uitstroomintensiteit in de richting van niet-participatie. Een vergelijkbare nuancering bieden de coëfficiënten voor het soort werkloosheid. Werklozen die voorheen een baan hadden, hebben een significant lagere uitstroomintensiteit in het gewone duurmodel, maar als onderscheid wordt gemaakt tussen de manieren waarop de werkloosheidsduur kan worden beëindigd, blijken zij vooral minder snel uit te stromen naar niet-participatie. De uitstroomintensiteit naar een nieuwe baan is even hoog als voor schoolverlaters en significant hoger dan voor herintreders.

¹¹⁾ Eventueel kunnen ook andere risks worden onderscheiden, zoals het gaan volgen van een dagopleiding, paneluitval, overlijden of om andere redenen stoppen met zoeken. Voor het huidige project lijkt het onderscheid tussen het wel of niet gevonden hebben van een nieuwe baan het belangrijkste.

Tabel 6
Parameterschattingen van competing risks model

	Exit: baan		Exit: geen baan	
	coëfficiënt	t-waarde	coëfficiënt	t-waarde
<i>Geslacht (ref.: man)</i>				
vrouw	-0,065	-2,46	0,248	10,00
<i>Leeftijd (ref.: 35-39)</i>				
15-19	0,496	7,65	0,750	12,01
20-24	0,613	12,18	0,359	6,65
25-29	0,459	10,02	0,037	0,72
30-34	0,192	4,31	0,049	1,10
40-44	-0,175	-3,88	0,005	0,12
45-49	-0,415	-8,20	0,089	2,02
50-54	-0,717	-12,54	0,142	3,05
55-59	-1,053	-15,10	0,401	8,17
60-64	-1,444	-10,03	0,830	12,51
<i>Opleiding (ref.: mbo, havo, vwo)</i>				
basisonderwijs	-0,252	-4,72	0,032	0,83
vmbo	-0,145	-4,89	0,030	1,18
hbo	0,167	4,71	-0,205	-5,58
wo	0,145	3,32	-0,347	-7,30
<i>Beroepsrichting (ref.: economisch)</i>				
techniek	0,163	4,59	0,023	0,62
verzorgend/sociaal-cultureel	0,094	2,77	0,167	5,13
overige	0,036	0,98	0,226	6,88
<i>Herkomstgroepering (ref.: autochtoon)</i>				
westers allochtoon	-0,104	-2,60	0,048	1,35
niet-westers allochtoon	-0,406	-9,94	0,134	4,05
<i>Gezinssituatie(ref.: alleen)</i>				
paar	0,075	0,98	0,190	2,92
ander lid	-0,141	-1,89	0,144	2,10
ln(# personen)	0,214	2,52	0,029	0,39
ln(1+# kinderen 0- 5)	-0,334	-5,25	0,091	1,64
ln(1+# kinderen 6-11)	-0,122	-2,07	0,078	1,55
ln(1+# kinderen 12-17)	-0,018	-0,32	0,131	2,64
ln(1+# kinderen 18+)	0,072	1,17	-0,030	-0,57
Partner werkt	0,043	1,09	-0,024	-0,71
<i>Opleiding partner (ref.: midden)</i>				
laag	-0,044	-1,16	0,000	0,01
hoog	0,045	1,19	-0,006	-0,19
<i>Woonregio (ref.: Midden-Nederland)</i>				
Groningen	-0,325	-5,07	0,029	0,51
Friesland	-0,163	-2,55	-0,019	-0,32
Drenthe	-0,243	-3,52	0,017	0,29
IJssel-Vecht/Twente	-0,198	-3,43	0,000	0,00
IJssel-Veluwe	-0,098	-1,46	0,023	0,36
Arnhem/O-Gelderl./Nijm./Rivierenl.	-0,171	-3,06	-0,013	-0,25
Flevoland	-0,201	-2,64	-0,147	-2,13
Noord-Holland Noord	0,101	1,43	0,175	2,74
Zuidelijk Noord-Holland	-0,060	-1,07	-0,020	-0,36
Rijnstreek	-0,107	-1,57	0,012	0,18
Haaglanden	-0,064	-1,01	-0,026	-0,43
Rijnmond	-0,143	-2,74	-0,028	-0,58
Zeeland	-0,147	-1,72	-0,007	-0,09
Midden en West Brabant	-0,084	-1,47	0,001	0,01
Noordoost Brabant	-0,068	-1,00	0,158	2,51
Zuidoost Brabant	-0,171	-2,67	-0,001	-0,01
Limburg	-0,245	-4,59	-0,027	-0,55
<i>Soort werkloze (ref. voorheen werkzaam)</i>				
herintreder	-0,451	-15,83	0,602	23,24
voorheen scholier	-0,023	-0,51	0,403	8,94
overig	0,000	0,00	0,504	10,44
<i>Zoeken naar werk</i>				
ln afgelopen 4 weken gezocht	-1,251	-27,73	0,620	15,12
Zoekactiviteiten factor 1	0,082	4,90	-0,434	-32,48
Zoekactiviteiten factor 2	0,170	15,70	-0,196	-16,56
<i>Jaardummies (ref.: 2005)</i>				
2002	0,156	3,22	0,172	3,94
2003	-0,122	-3,23	0,015	0,43
2004	-0,196	-5,70	-0,080	-2,58
2006	0,155	4,51	0,062	2,02
2007	0,344	9,07	0,082	2,38
<i>Maandummies (ref.: september)</i>				
januari	-0,390	-7,35	-0,136	-2,73
februari	-0,143	-2,51	-0,055	-0,99
maart	-0,059	-1,21	-0,038	-0,80
april	-0,079	-1,58	0,019	0,40
mei	-0,181	-3,48	0,072	1,46
juni	-0,169	-3,14	0,135	2,76
juli	-0,261	-4,96	0,136	2,92
augustus	-0,248	-4,75	0,106	2,31
oktober	-0,063	-1,26	0,016	0,32
november	-0,244	-4,23	0,017	0,32
december	-0,353	-6,71	-0,073	-1,50
<i>Stapfunctie duurzaamheid, ref.: 3-6 maanden)</i>				
0- 3 maanden	0,114	3,29	0,019	0,64
6- 9 maanden	-0,229	-4,48	-0,125	-2,81
9-12 maanden	-0,405	-5,28	-0,214	-3,17
12-18 maanden	-0,530	-5,82	-0,026	-0,36
18-24 maanden	-0,867	-7,09	-0,116	-1,23
24-36 maanden	-0,753	-5,64	0,030	0,29
24 maanden en langer	-0,484	-2,94	0,180	1,35
Constante	-0,607	-3,67	-3,448	-21,03

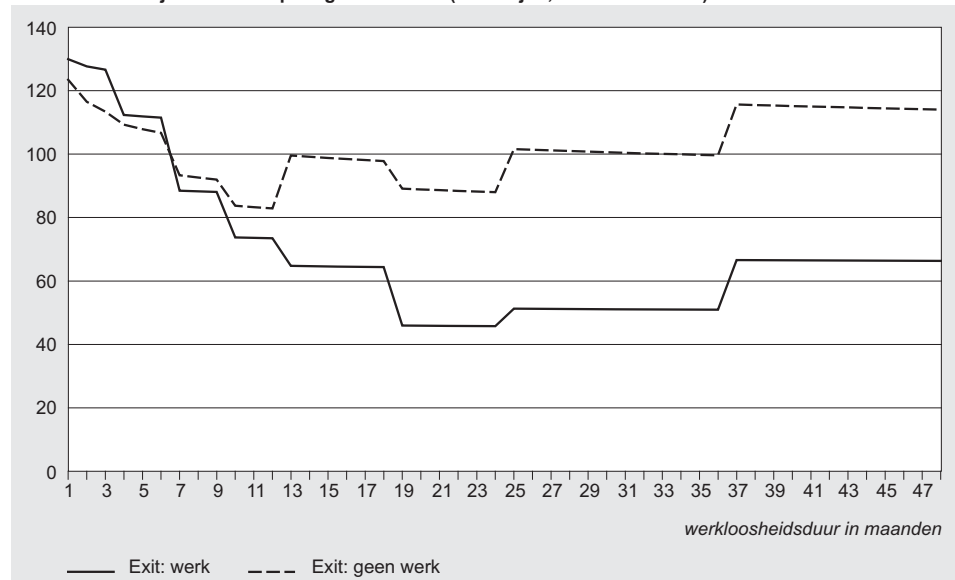
Tabel 6 (slot)
Parameterschattingen van competing risks model

	Exit: baan		Exit: geen baan	
	coëfficiënt	t-waarde	coëfficiënt	t-waarde
Weibull duurzaamheid				
α	0,984	15,84	0,947	19,61
$\alpha-1$	-0,016	-0,26	-0,053	-1,09
N (# waarnemingen)		36 595,00		
AIC		66 762,60		
BIC		68 072,80		
k (# parameters)		154,00		
Log likelihood		-33 227,30		
χ^2		3 674,90		

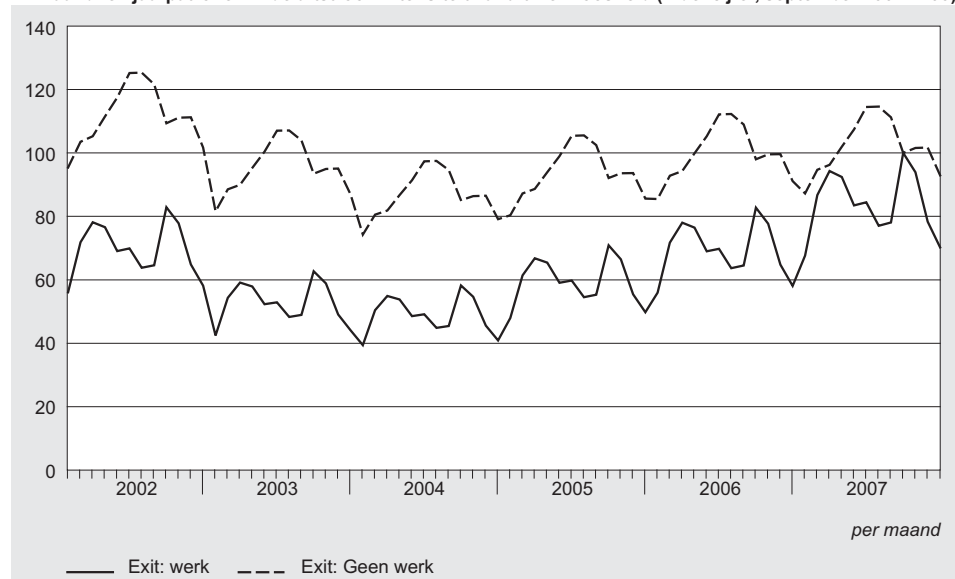
Noot: schaduwkleur geeft aan dat de betreffende coëfficiënt op 5%-niveau significant van nul verschilt.
De χ^2 -waarde heeft betrekking op de statistische toets van de hypothese dat alle parameters behalve de constante term nul zijn. Dit is een formele toets op de verklaringskracht van het model, waarbij de hypothese met 95% significantie wordt verworpen bij kritieke waarde 182,9.

Figuur 3 geeft de geschatte duurzaamheid van de beide uitstroomintensiteiten, Figuur 4 de bijbehorende maand- en jaarpatronen.

3. Duurafhankelijkheid in competing risks model (indexcijfer, 6 maanden =100)



4. Maand- en jaarpatronen in de uitstroomintensiteit vanuit werkloosheid (indexcijfer, september 2007 =100)



Beide vertonen een dalende duurzaamheidsgraad in de eerste maanden, waarna de uitstroomintensiteit naar participatie weer toeneemt, terwijl de uitstroomintensiteit naar het vinden van een nieuwe werkkring verder afneemt. De invloed van de werkloosheidsduur op de kans dat iemand zich van de arbeidsmarkt terugtrekt, is daarom slechts beperkt, terwijl de kans op het vinden van een baan aanmerkelijk sterker afhangt van het aantal maanden dat mensen al op zoek zijn naar werk.

Het seizoenspatroon in de werkloosheidsbeëindigingen gaf in de schattingen van het gewone duurmodel in de vorige paragraaf vooral een dip te zien in de maanden januari en december. Door de beëindigingen met en zonder nieuwe baan te onderscheiden, kunnen ook de bijbehorende seizoenspatronen in beeld worden gebracht. Deze maken duidelijk dat de periodes maart/april en september/oktober de grootste kans hebben op het vinden van een baan. In de zomerperiode wordt de terugval gecamoufleerd door een groter aantal werklozen dat (tijdelijk) stopt met zoeken.

5. Statistieken van werkloosheidsduren en -stromen

5.1 Inleiding

Op basis van de geschatte duurmodellen kunnen statistieken van (verwachte) werkloosheidsduren worden afgeleid. In de algemene aanpak hiervoor kunnen twee stappen worden onderscheiden:

- I. bepaal correcte schattingen van het duurmodel;
- II. bereken daarmee voor een gegeven steekproef of populatie de gewenste statistiek van de gekozen duur.

In de eerste stap wordt de kansverdeling van de werkloosheidsduren correct geschat. Voor de meeste toepassingen zal dat een instroomverdeling zijn, ook als de gebruikte gegevens zelf anders zijn georganiseerd (bijvoorbeeld als *stock sample*) of onvolledig zijn waargenomen. Deze stap is in de voorgaande hoofdstukken uitvoerig besproken. Het geschatte model beschrijft hoe de kansverdeling van de werkloosheidsduren afhangt van kenmerken die tussen personen of tijdstippen kunnen verschillen. Hiermee kan in de tweede stap de voor het afleiden van statistieken beschikbare informatie worden aangevuld.

Voor deze tweede stap moet nog een aantal fundamentele keuzes worden gemaakt.

- Van welke duren wordt een statistiek bepaald? Van de reeds waargenomen duren (in de eerste peiling EBB) aangevuld met de verwachte restduur, van de verwachte restduur zelf of van de verwachte duren van mensen die ‘nu’, op een bepaald moment of in een gegeven periode werkloos worden of zijn geworden.
- En daaraan gerelateerd: voor welke populatie? De werklozen op een bepaald peilmoment, de instromende werklozen in een bepaalde periode, de gemiddelde Nederlander?
- Onder wat voor veronderstellingen over de arbeidsmarkt worden verwachte duren berekend en/of aangevuld? Onder gemiddelde omstandigheden in de observatieperiode, onder de huidige omstandigheden of onder ‘verwachte’ omstandigheden?
- Over welk aspect of kenmerk van de duurverdeling rapporteert de statistiek? Het gemiddelde, de mediaan, of bijvoorbeeld het percentage korter dan 3 maanden of langer dan 1 jaar?
- Terugkomend op de gebruikte duren, kan ook de vraag worden gesteld of het om werkloosheidsduren gaat of om de duur tot het vinden van een nieuwe werkkring (de baanvindduur)? Niet iedere werkloosheidsduur wordt namelijk succesvol afgesloten. Door in het competing risks model onderscheid te maken tussen de manieren waarop een werkloosheidsduur kan worden afgesloten – met of zonder het vinden van een nieuwe baan – is het nu ook mogelijk om te berekenen hoe lang iemand met bepaalde kenmerken zal moeten zoeken tot een nieuwe baan wordt gevonden, of anders gezegd: wat de verwachte werkloosheidsduur is als deze pas bij het vinden van een nieuwe betaalde baan wordt beëindigd en niet om andere redenen, zoals pensionering, tijdelijk terugtrekken van de arbeidsmarkt of sterfte.
- En terugkomend op stap I: met welke van de modelschattingen?

Duurmodellen maken het mogelijk om op verantwoorde wijze de verdeling van werkloosheidsduren te analyseren. Het doel van dit project is echter na te gaan of op basis van de uitkomsten van de analyse goede statistieken van de verwachte werkloosheidsduur zijn te maken. Dit stelt verdergaande eisen. In ieder geval moeten de statistieken verantwoord, betrouwbaar en inzichtelijk zijn.

Met ‘verantwoord’ wordt hier bedoeld dat de statistieken op een correcte – wetenschappelijk verantwoorde – wijze zijn bepaald en bijvoorbeeld op geldige veronderstellingen over de arbeidsmarkt zijn gebaseerd. Voor ‘betrouwbaar’ is nodig dat de statistieken niet gevoelig zijn voor arbitraire veronderstellingen in de empirische specificatie van het duurmodel. En de ‘inzichtelijkheid’ vereist dat de statistieken intuïtief duidelijk zijn en de gewenste informatie verschaffen. Met deze criteria in het achterhoofd wordt in dit hoofdstuk nagegaan op welke manier de statistische analyse van werkloosheidsduren uit het vorige hoofdstuk kan worden ingezet voor het ontwikkelen van statistieken van verwachte werkloosheidsduren die bestaande statistieken op dit terrein kunnen aanvullen.

Veilig kan worden gesteld dat de beoogde statistieken een antwoord moeten geven op de vraag hoe lang een werkloze op een gegeven moment gemiddeld zoekt voordat hij of zij een baan vindt en hoe die zoekduur tussen deelpopulaties verschilt en zich over de tijd ontwikkelt. Statistieken op basis van duurmodellen zouden daaraan kunnen bijdragen door de verdeling van de verwachte voltooide werkloosheidsduren in beeld te brengen. Bovengenoemde keuzemogelijkheden bieden de nodige vrijheidsgraden om goed aan te sluiten bij het doel van de statistiek of om op verschillende manieren verschillende facetten van werkloosheidsduren te belichten.

Het is daarom van belang om goed voor ogen te houden wat de mogelijke informatie-behoefte zijn waarin de nieuwe statistieken zouden kunnen voorzien. Is het doel de actuele situatie op de arbeidsmarkt te karakteriseren? Is het doel de werkloosheidsduren van de huidige werklozen te karakteriseren? Of is het doel om de verwachte werkloosheidsduur zo goed mogelijk te voorspellen voor de mensen die nu werkloos worden of de verwachte resterende werkloosheidsduur voor de mensen die nu al werkloos zijn? In het eerste geval zal men vooral willen weten hoe lang een werkloze momenteel gemiddeld zoekt voordat hij of zij onder de huidige omstandigheden een baan vindt. Dat kan gaan over alle huidige werklozen, maar ook over mensen die net werkloos zijn geworden. Bij het tweede doel gaat het om statistieken die per definitie achter de feiten aanlopen, namelijk door te boekstaven hoelang de huidige werklozen in de voorafgaande tijd al vruchteloos naar werk hebben gezocht. Voor het derde doel moeten de uitkomsten van de duuranalyse worden verrijkt met veronderstellingen over hoe de arbeidsmarkt zich naar alle waarschijnlijkheid gaat ontwikkelen, waardoor de statistiek aan betrouwbaarheid verliest, maar wellicht aan relevantie wint. Het eerste doel sluit het best aan bij de taak van het CBS en de ambitie om met de additionele statistieken van de werkloosheidsduren niet te zeer het verleden te archiveren.

Met deze overwegingen in het achterhoofd is besloten twee verschillende statistieken op basis van de duurmodellen te onderzoeken

1. statistieken van verwachte werkloosheidsduren van de mensen die in een bepaalde peilperiode werkloos zijn (aanvulling van lopende werkloosheidsduren),
2. statistieken van werkloosheidsduren van mensen die net werkloos zijn geworden (instroomduren).

Daarnaast wordt ook gekeken naar een statistiek waarin de duur tot het vinden van een nieuwe baan centraal staat:

3. statistieken van de duur tot iemand een nieuwe baan heeft gevonden (baanvindduren), wat zowel kan voor het bestand van werklozen als voor de nieuwe instroom in werkloosheid.

Bij de eerste van deze statistieken wordt uitgegaan van de groep respondenten die in eerste peilingen (de CAPI-interviews) van de EBB werkloos zijn. Voor de tweede soort statistieken worden de respondenten uit zowel de eerste peiling als de vier telefonische vervolopeilingen gebruikt mits zij op dat moment korter dan een kwartaal werkloos waren.¹²⁾ De derde soort statistieken kan zowel voor bestaande als instromende werklozen worden bepaald. Om de complete duur als 'baanvindduur' te kunnen opvatten, wordt hier voor de instromende werklozen gekozen.¹³⁾

Het eerste type statistieken heeft als voordeel dat het dicht bij de waargenomen duren blijft. Het blijft beperkt tot het aanvullen van wat op de peildatum nog moest komen en hierbij wordt uitgegaan van de huidige (economische) omstandigheden. Deze aanpak sluit bovendien het best aan bij de in de vorige onderzoeksfase (Berkhout *et al.*, 2007) ingeslagen weg. Daarnaast ligt de onderzoekspopulatie bij deze aanpak vast, namelijk de personen die op een bepaald peilmoment werkloos zijn.

¹²⁾ Als variant hierop wordt ook een periode van een jaar als tijdraam gebruikt, zij het dat eventuele dubbelellingen van dezelfde duur daarbinnen worden weggegooid. Met een periode van een jaar krijgen we meer waarnemingen dan met een kwartaal, maar sluipt het karakter van een stock sample binnen, waardoor dit hier slechts als *benchmark* is opgenomen.

¹³⁾ Voor elk van de drie statistieken wordt dus een bijpassende steekproef uit de EBB gebruikt. Indien het voor de representativiteit nodig is om de EBB-gegevens te wegen, moet dit ook bij deze toepassingen worden gedaan.

Door te vertrekken van de tot op een peildatum waargenomen lopende werkloosheidsduren heeft deze statistiek meer dan de andere twee statistieken de blik naar het verleden gericht, maar maakt de aanvulling met de verwachte resterende duur deze cijfers relevant voor de verdere ontwikkeling van het huidige werklozenbestand. De op de peildatum waargenomen lopende durven vormen een *stock sample*, met verhoudingsgewijs veel langere en weinig kortere durven. Hierdoor geeft de statistiek weliswaar geen goed beeld van de durven die een nieuw instromende werkloze mag verwachten, maar correspondeert de verdeling volledig met het huidige werklozenbestand op de peildatum en is de verwachte resterende duur (anders dan in Berkhout *et al.*, 2007) berekend zonder vertekening door het steekproefdesign, d.w.z. als ‘verlate instroom’ conditioneel op de op de peildatum al verstreken duur. Daarmee is de definitie van de statistiek consistent en intuïtief helder, enerzijds doordat in de statistiek onderscheid kan worden gemaakt tussen de verwachte resterende duur en de verwachte totale duur en anderzijds doordat de statistieken betrekking hebben op een relevante groep mensen: de mensen die op het betreffende moment werkloos zijn.

Het tweede type statistieken geeft het meest een beeld van de actuele situatie op de arbeidsmarkt, kijkt niet achterom, maar gaat uit van de huidige situatie en pretendeert ook niet om toekomstige veranderingen of ontwikkelingen correct te voorspellen. Het brengt voor de groep van nieuwe werklozen de verwachte werkloosheidsduur onder huidige omstandigheden in kaart. De tot de peiling waargenomen werkloosheidsduren worden in deze statistiek alleen gebruikt om de kortelings ingestroomde werklozen in de eerste EBB-peiling te identificeren. De verwachte werkloosheidsduur wordt direct uit de geschatte verdeling berekend.

Deze beide statistieken vertonen een analogie met demografische ‘life tables’ die aangeven wat de verwachte levensduur van een jonggeborene is, maar ook hoe lang de gemiddelde Nederlander nog voor de boeg heeft. Het derde type statistieken benut de meerwaarde van het *competing risks* model en geeft het best antwoord op de vraag hoe lang iemand die nu werkloos wordt er op de huidige arbeidsmarkt naar verwachting nodig heeft om een nieuwe baan te vinden.

In de volgende paragraaf wordt nader ingegaan op de manier waarop durven aangevuld of geïmputeerd kunnen worden op basis van een statistische analyse van duurmodellen. Vervolgens komen in paragrafen 5.3, 5.4 en 5.5 de drie hierboven genoemde statistieken op basis van duurmodellen aan de orde. De gegevens die zijn gebruikt bij het schatten van de duurmodellen kunnen ook op een directere manier in statistieken worden omgezet, namelijk door op basis van de vervolgpelingen van de EBB te turven hoeveel procent van de werklozen in de daaropvolgende peiling (1) nog steeds werkloos is, (2) een baan heeft gevonden of (3) om andere redenen niet langer werkloos is. Deze statistieken hebben als voordeel dat ze geen parametrisch model nodig hebben en automatisch de *competing risks* in kaart brengen. Daar staat tegenover dat het natuurlijk geen statistieken van (verwachte) werkloosheidsduren zijn. Bovendien is dit soort statistieken kwetsbaar voor selectieve uitval in het roterende panel.¹⁴⁾ In paragraaf 5.6 worden de mogelijkheden van dit soort statistieken nader toegelicht.

5.2 Het berekenen van de statistieken

Voor het berekenen van het gemiddelde van een duurverdeling of andere kenmerken van de geschatte kansverdelingen is de Monte Carlo methode doorgaans de eenvoudigste aanpak. Voor gegeven kenmerken x en parameters α en β is de volledige kansverdeling van de durven bekend en kan een aantal willekeurige trekkingen uit die verdeling worden gedaan. Daarvan kan het gemiddelde worden berekend als schatting van het gemiddelde van de verdeling en kunnen op vergelijkbare manier schattingen worden berekend van de mediaan of van de procentuele verdeling over een aantal vooraf gedefi-

¹⁴⁾ Mogelijk kan hiervoor worden gecorrigeerd door middel van een weegschema op basis van waargenomen kenmerken, waarmee specifiek de uitval in de opeenvolgende golven van het roterende panel kan worden gecompenseerd.

nieerde duurcategorieën. De trekkingen uit de duurverdeling kunnen worden berekend door een willekeurig reëel getal u uit de uniforme verdeling over het interval $[0,1]$ te trekken en hiervan de waarde van de inverse van de verdelingsfunctie $F(d)$ te bepalen. In het geval van een proportioneel *hazard* model met Weibull duurzaamheidsafhankelijkheid geeft dit:

$$d = \left[-\frac{\ln(u)}{\exp(x' \beta)} \right]^{1/\alpha}$$

Deze aanpak kan ook worden gevolgd om een resterende duur te trekken uit de kansverdeling van duren waarvan al een beginduur d_0 is verstreken. In dat geval wordt de uitdrukking:

$$r = \left[-\frac{\ln(u)}{\exp(x' \beta)} + d_0^\alpha \right]^{1/\alpha} - d_0$$

Om deze formules te kunnen gebruiken is wel van belang dat x constant is (of wordt verondersteld) en dat de duurzaamheidsafhankelijkheid van de Weibull-vorm is. Om die reden is voor de statistieken in dit hoofdstuk gebruik gemaakt van de schattingen met alleen Weibull duurzaamheidsafhankelijkheid en zijn de exogenen constant gehouden. Als hiervan wordt afgeweken kan eenzelfde aanpak worden gevolgd, maar het is in dat geval minder eenvoudig om de inverse van de verdelingsfunctie te berekenen.

De nauwkeurigheid van de Monte Carlo schattingen is uiteraard groter naarmate deze op een groter aantal trekkingen worden gebaseerd. De in dit hoofdstuk gerapporteerde statistieken zijn berekend op basis van een gegeven steekproef van enkele duizenden personen waarbij voor elke waarneming 100 willekeurige trekkingen uit de bijbehorende duurverdeling zijn genomen.

Voor enkele statistieken kunnen ook gesloten uitdrukkingen worden afgeleid. De procentuele verdeling van de duren over gegeven duurklassen, kan eenvoudig worden berekend met behulp van de verdelingsfuncties $F(d)$.

Voor een Weibull duurmodel met – voor het bepalen van de verwachte duren – constant gehouden exogenen geldt:

$$E(d) = \int_0^1 \left(-\frac{\ln(u)}{\exp(x' \beta)} \right)^{1/\alpha} du = \exp\left(-\frac{x' \beta}{\alpha}\right) \cdot \Gamma\left(1 + \frac{1}{\alpha}\right)$$

Deze uitdrukking kan dus worden gebruikt voor het berekenen van het gemiddelde, de mediaan of andere kenmerken van de instroomduur, d.w.z. de werkloosheidsduur van mensen die op een bepaald moment werkloos worden.

Deze formule kan niet direct worden gebruikt om het gemiddelde van de verwachte resterende duur van reeds lopende duren te bepalen. Voor dergelijke *delayed entry* duren, d.w.z. duren die als *stock sample* zijn verzameld en waarvan het moment van de peiling al een duur d_0 was verstreken, geldt:¹⁵⁾

$$\begin{aligned} E(d|d_0) &= \int_0^1 \left(-\frac{\ln(u)}{\exp(x' \beta)} + d_0^\alpha \right)^{1/\alpha} du \\ &= \exp\left(-\frac{x' \beta}{\alpha} + \exp(x' \beta) \cdot d_0^\alpha\right) \cdot \Gamma\left(1 + \frac{1}{\alpha}; \exp(x' \beta) \cdot d_0^\alpha\right) \end{aligned}$$

Voor de duren in een *competing risks* model zijn dit soort analytische uitdrukkingen van de gemiddelde duur of frequentieverdelingen niet beschikbaar en is daarom uitsluitend gebruik gemaakt van de Monte Carlo methode. Dit is gedaan door voor elke manier waarop de duur kan worden beëindigd een duur te trekken uit de duurverdeling voor de betreffende *hazard rate* en de kleinste van deze virtuele duren te berekenen. In dit hoofdstuk wordt bij de werkloosheidsduren slechts onderscheid gemaakt tussen twee *exit routes*: met en zonder nieuwe baan. Met de boven beschreven aanpak kunnen virtuele duren worden berekend voor het baan vinden en voor het stoppen met zoeken om andere redenen. De kleinste van de twee is een trekking uit de duurverdeling van het

¹⁵⁾ Waarbij $\Gamma(x,y)$ de niet geschaalde *upper incomplete gamma* functie is (Abramowitz en Stegun, 1972).

competing risks model, maar de eerste virtuele duur is als baanvindduur op zichzelf ook interessant. Deze drukt uit hoe lang het naar verwachting duurt om een baan te vinden als het zoekproces niet voortijdig om een andere reden wordt afgebroken. Ook hiervan zullen zoals hiervoor werd aangegeven statistieken worden bepaald.

Bij de aanvulling van lopende werkloosheidsduren is gebruik gemaakt van de steekproef van respondenten die op de eerste EBB-peiling (CAPI) werkloos waren. Voor statistieken over de duurverdeling van personen die in werkloosheid instromen, wordt gebruik gemaakt van de personen die recent, in het afgelopen kwartaal of in het afgelopen jaar, werkloos zijn geworden. Daarbij is het goed om op te merken dat dit gaat om de bepaling van de groep personen waarvoor de statistieken van de verwachte duren worden bepaald. De daarbij gebruikte parameterschattingen zijn in deze rapportage gebaseerd op alle gegevens uit de jaren van 2002 tot en met 2007.¹⁶⁾ De statistieken zijn berekend op basis van de schattingen van het duurmodel en het *competing risks* model met alleen Weibull duuraafhankelijkheid. De aantallen personen waarop de statistieken zijn gebaseerd staan in Tabel 7.

Tabel 7
Aantallen waarnemingen in de schattingssteekproef die zijn gebruikt voor het berekenen van de statistieken

Kwartaal	Lopende werkloosheidsduren en instroom ¹⁾	Instroom in afgelopen kwartaal	Instroom in afgelopen jaar
2002.1	569	259	415
2002.2	436	373	501
2002.3	482	625	763
2002.4	483	676	822
2003.1	648	1 014	1 208
2003.2	574	961	1 144
2003.3	674	1 054	1 255
2003.4	694	1 047	1 297
2004.1	973	1 360	1 684
2004.2	849	1 270	1 572
2004.3	791	1 157	1 445
2004.4	886	1 272	1 602
2005.1	866	1 427	1 725
2005.2	875	1 199	1 484
2005.3	841	1 247	1 524
2005.4	894	1 177	1 471
2006.1	795	1 239	1 484
2006.2	690	1 045	1 244
2006.3	767	1 025	1 232
2006.4	617	1 047	1 214
2007.1	593	1 118	1 296
2007.2	566	863	1 035
2007.3	584	898	1 046
2007.4	485	825	964
Totaal	16 632	24 178	29 427

¹⁾ Lopende duren uit de eerste peiling en nieuw ingestroomde werklozen in de eerste peiling na instroom.

Voor de statistieken voor de verwachte duren van de personen die op een bepaald peilmoment werkloos zijn, worden alleen de waarnemingen uit de eerste peilingen van de EBB gebruikt. De samenstelling van deze groep is daarom niet beïnvloed door paneluitval in de telefonische vervolgpeilingen. In totaal geeft dit tussen 2002 en 2005 16 632 waarnemingen, waarvan de eerste kolom van de tabel de verdeling geeft over de kwartalen. Deze aantallen reflecteren het conjunctureel verloop van de werkloosheid in deze periode. Om een goed bestand samen te stellen als basis voor de statistieken voor 'instromers' nemen we van deze waarnemingen alleen degenen die op de peildatum pas korter dan drie maanden werkloos zijn – dat zijn er 5 513 – en voegen daar de personen aan toe die tussen twee opeenvolgende peilingen, gemiddeld met tussentijd van een kwartaal, werkloos zijn geworden. Dat zijn er 5 651 tussen de eerste en de tweede peiling, 5 021 tussen de tweede en de derde peiling, 4 205 tussen de derde en de vierde

¹⁶⁾ Bij het produceren van statistieken kunnen de parameterschattingen op grond van nieuwe informatie worden aangepast. Dat kan door de schattingssteekproef te laten groeien of door deze met een vaste lengte van de tijdsperiode te laten 'voortschrijven'.

peiling en 3 788 tussen de vierde en de vijfde peiling. Dat geeft in totaal 24 178 waarnemingen als aselechte steekproef uit de personen die ‘zoiust’ zijn ingestroomd. Ter vergelijking is op dezelfde manier ook de steekproef samengesteld van personen die op een peildatum korter dan een jaar werkloos waren. Van elk van deze duren is alleen de eerste waarneming meegenomen om dubbeltellingen te voorkomen. In totaal levert dit 29 427 waarnemingen op, waarvan de duren op de eerste peildatum niet meer dan een jaar mogen bedragen, maar toch al een lichte length bias kunnen bevatten. Om die reden wordt voor de statistieken van de verwachte werkloosheidsduren van ‘instromers’ de voorkeur gegeven aan de aantallen waarnemingen in de tweede kolom van Tabel 7 boven die in de derde. Wel worden de laatste ook gebruikt om na te gaan hoe gevoelig de statistieken zijn voor deze definitie.

5.3 Statistieken op basis van het aanvullen van lopende duren

Tabel 8 geeft een overzicht van gemiddelde (verwachte) duren uitgesplitst naar het jaar van de meting waarop deze zijn gebaseerd. In de kolommen staan:

- de op het moment van de peiling reeds verstreken duur;
- de verwachte resterende duur (conditioneel op de reeds volbrachte duur);
- complete duur = beginduur + restduur.

De statistieken van de laatste twee duren zijn bovendien zowel met het duurmodel als met het *competing risks* model berekend. Wat direct opvalt is het verschil in grootte tussen de reeds verstreken duur en de verwachte resterende duur. Dit verschil weerspiegelt het feit dat de verstreken duren een *stock sample* vormen, waarin lange duren en mensen met kenmerken die samenhangen met gemiddeld lange duren oververtegenwoordigd zijn. Verder valt op dat de schattingen van het duurmodel en van het *competing risks* model nagenoeg dezelfde statistieken opleveren. Dit is opmerkelijk omdat beide modellen op het oog nogal verschillende parameterschattingen geven. Kennelijk bieden de extra vrijheidsgraden – 77 extra parameters in het *competing risks* model – en de additioneel gebruikte gegevens – informatie over de manier waarop de werkloosheidsduur is beëindigd – geen extra handvatten om de beschrijving van het duurproces nog substantieel te verbeteren.

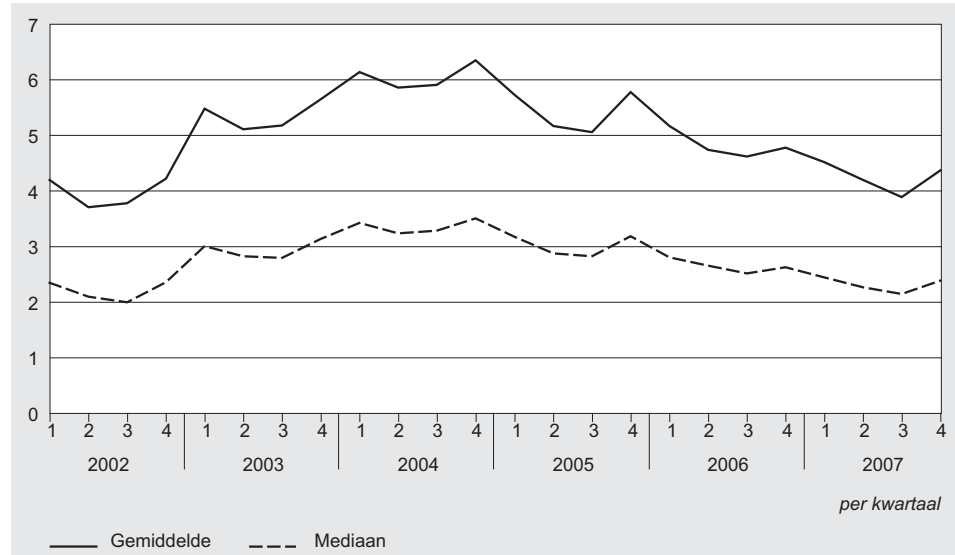
De mediane duren zijn lager dan de gemiddelde duren, wat wijst op een scheve verdeling met lange duren. De verwachte resterende duur volgt een duidelijk anticyclisch conjunctureel bepaald patroon. Dat geldt echter niet voor de ‘beginduren’. Een verklaring hiervoor kan zijn dat in goede tijden de harde kern van de werklozen met gemiddeld lange duren overblijft, terwijl daar in een recessie veel nieuwe werklozen bijkomen, die de gemiddelde (en mediane) duur omlaag trekken.

Tabel 8
Gemiddelde en mediaan van verwachte werkloosheidsduren in maanden, 2002–2007

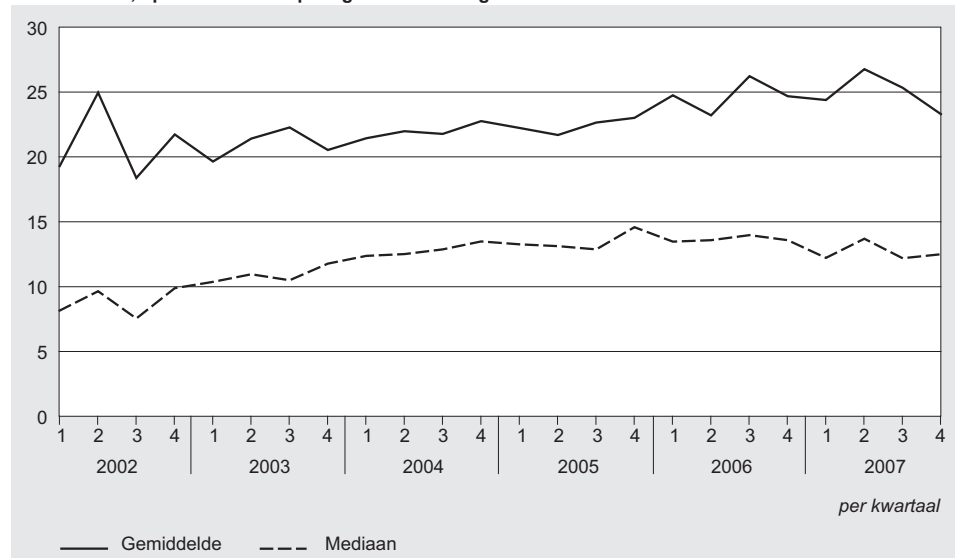
Jaar	Reeds verstreken duur	Duurmodel		Competing risks model	
		Verwachte resterende duur	Verwachte complete duur	Verwachte resterende duur	Verwachte complete duur
<i>gemiddelde duren</i>					
2002	16,9	3,9	20,8	4,0	20,9
2003	15,5	5,3	20,8	5,4	20,9
2004	15,8	6,0	21,9	6,1	22,0
2005	17,0	5,4	22,3	5,4	22,4
2006	20,0	4,8	24,7	4,8	24,7
2007	20,7	4,2	24,9	4,2	25,0
2002–2007	17,5	5,1	22,6	5,1	22,7
<i>mediane duren</i>					
2002	5,0	2,2	8,6	2,2	8,7
2003	6,0	3,0	10,7	2,9	10,9
2004	6,0	3,4	12,7	3,4	12,8
2005	8,0	3,1	13,4	3,0	13,5
2006	9,0	2,7	13,6	2,7	13,6
2007	8,0	2,4	12,6	2,3	12,6
2002–2007	6,0	2,8	12,3	2,8	12,4

Figuur 5 geeft het gemiddelde en de mediaan van de verwachte resterende werkloosheidsduur per kwartaal. Hierbij zijn de *competing risks* schattingen gebruikt. De eerst stijgende en dan dalende trend reflecteert de conjuncturele ontwikkelingen. Figuur 6 geeft de overeenkomstige grafiek voor de volledige verwachte dueren, dat wil zeggen de waargenomen beginduur aangevuld met de verwachte resterende werkloosheidsduur.

5. Gemiddelde en mediaan van de verwachte resterende werkloosheidsduur in maanden, per kwartaal, 2002–2007, op basis van competing risks schattingen



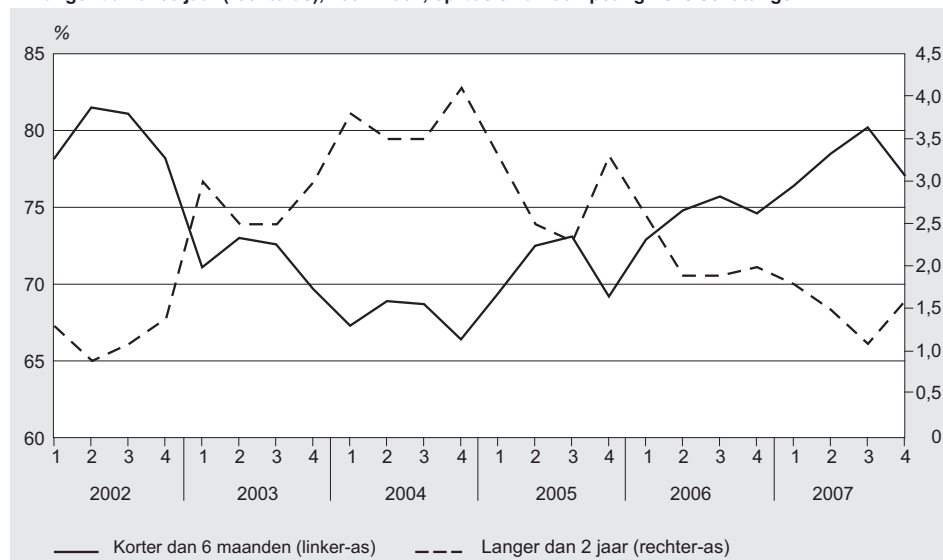
6. Gemiddelde en mediaan van de verwachte volledige werkloosheidsduur in maanden, per kwartaal, 2002–2007, op basis van competing risks schattingen



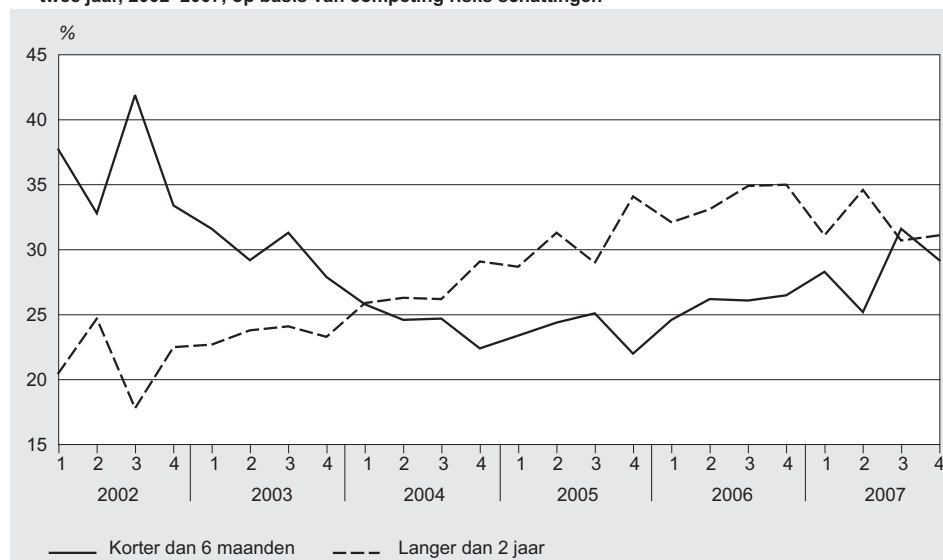
Twee alternatieve mogelijkheden om de verdeling van de verwachte resterende en totale werkloosheidsduren in statistieken weer te geven, is door middel het percentage korte en lange dueren. Figuur 7 en Figuur 8 geven beide, respectievelijk voor de verwachte resterende werkloosheidsduren en voor de verwachte complete werkloosheidsduren.

Voorts kunnen de gemiddelde of mediane dueren ook worden gerapporteerd voor deelgroepen. Tabel 9 geeft voor 2002 tot 2007 de gemiddelde werkloosheidsduren uitgesplitst naar enkele achtergrondkenmerken.

7. Percentage verwachte resterende werkloosheidsduren korter dan 6 maanden (linkeras) respectievelijk langer dan twee jaar (rechteras), 2002–2007, op basis van competing risks schattingen



8. Percentage verwachte complete werkloosheidsduren korter dan 6 maanden respectievelijk langer dan twee jaar, 2002–2007, op basis van competing risks schattingen



Tabel 9
Gemiddelde verwachte volledige werkloosheidsduren in maanden, naar deelgroepen

	2002	2003	2004	2005	2006	2007
Man	18,6	20,0	22,2	23,0	25,5	24,4
Vrouw	22,6	21,8	21,8	21,9	24,2	25,4
Totaal	20,9	20,9	22,0	22,4	24,7	25,0
15 t/m 24 jaar	6,5	9,0	9,5	9,9	8,9	8,5
25 t/m 34 jaar	14,0	13,7	17,2	17,6	17,4	17,6
35 t/m 44 jaar	24,8	23,0	23,8	24,8	25,1	24,9
45 t/m 54 jaar	31,7	34,7	33,6	30,1	33,9	33,9
55 t/m 64 jaar	37,6	35,3	34,1	34,4	40,3	38,6
Totaal	20,9	20,9	22,0	22,4	24,7	25,0
Basisonderwijs	32,0	24,3	26,4	26,6	26,8	33,1
Vmbo	19,5	22,9	22,2	21,2	23,9	23,1
Mbo/havo/vwo	18,9	19,9	20,8	22,0	24,5	25,3
Hbo	21,0	17,6	23,2	21,7	24,6	23,3
Wo	18,6	19,4	19,7	23,9	26,2	23,2
Totaal	20,9	20,9	22,0	22,4	24,7	25,0

Tabel 9l (slot)
Gemiddelde verwachte volledige werkloosheidsduren in maanden, naar deelgroepen

	2002	2003	2004	2005	2006	2007
Groningen	18,2	24,4	21,8	34,3	23,3	29,9
Friesland	18,6	19,1	25,2	22,5	28,0	27,9
Drenthe	22,7	30,0	18,7	21,9	24,6	26,9
IJssel-Vecht/Twente	18,2	18,2	23,0	24,2	23,5	23,2
IJssel-Veluwe	19,7	25,2	22,4	20,1	20,5	24,6
Arnhem/O.Gelderl. e.v.	20,6	23,5	23,9	20,9	25,3	26,4
Flevoland	24,4	19,1	26,3	20,6	32,2	22,1
Midden-Nederland	21,5	17,5	18,2	20,7	22,1	25,6
Noord-Holland Noord	23,7	14,5	25,6	15,4	23,5	20,0
Zuidelijk N.-Holland	14,9	20,5	19,4	22,6	24,0	23,3
Rijnstreek	20,5	19,7	17,7	18,7	22,5	29,3
Haaglanden	16,7	21,8	17,6	21,1	21,0	22,2
Rijnmond	26,0	20,0	22,0	21,8	21,8	25,8
Zeeland	29,8	29,4	28,4	24,5	40,6	35,1
Midden en W. Brabant	20,4	17,8	24,3	21,4	23,0	23,9
Noordoost Brabant	19,6	18,4	19,8	19,4	21,5	22,3
Zuidoost Brabant	21,8	25,5	21,5	25,5	31,8	19,0
Limburg	21,5	20,4	24,3	23,7	31,7	24,0
Totaal	20,9	20,9	22,0	22,4	24,7	25,0
Autochtoon	20,8	21,1	21,5	22,4	24,9	25,5
Westers allochtoon	23,0	20,6	24,1	22,7	27,1	26,0
Niet-westers allochtoon	19,8	20,4	22,9	22,2	22,8	22,0
Totaal	20,9	20,9	22,0	22,4	24,7	25,0

De uitsplitsing naar herkomstgroepering laat overigens ook zien dat statistieken van aangevulde duren soms moeilijk te interpreteren kunnen zijn: de verwachte resterende duur geven weliswaar het verwachte beeld – gemiddeld het langst voor niet-westerse allochtonen en het kortst voor autochtonen – maar als gevolg van de steekproefopbouw (*stock sample*) geven de reeds verstreken duren en de op basis daarvan berekende complete duren een ander beeld.¹⁷⁾ Merk daarbij overigens wel op dat de beginduren direct waargenomen zijn en dus niet van de modelschattingen afhangen.

5.4 Statistieken op basis van verwachte werkloosheidsduren van instromers

Tabel 10 geeft het gemiddelde en de mediaan van de instroomduren. Deze zijn beduidend lager dan de voltooide duren uit Tabel 8. Dit komt omdat de daarbij gebruikte beginduren een *stock sample* vormen, in de steekproef van lopende werkloosheidsduren minder gunstige kenmerken vaker voorkomen dan onder de nieuw ingestroomde werklozen en de negatieve duurzaamheidsduur ervoor zorgt dat de verwachte resterende duren gemiddeld ook langer duren dan de instroomduren. Opnieuw liggen de statistieken op basis van de schattingen van het duurmodel dicht bij de statistieken op basis van het *competing risks* model.

Verder valt op dat de instromers in het afgelopen jaar net iets hogere duren hebben dan de instromers in het afgelopen kwartaal. Het verschil is echter klein. Wel laat dit zien dat bij een tijdraam van een jaar in plaats van een kwartaal de invloed van *length bias sampling* al merkbaar is, ook al zijn de waargenomen duren zelf uitsluitend gebruikt om af te bakenen welke werklozen kort geleden zijn ingestroomd en zijn de reeds verstreken duren zelf niet gebruikt als beginduur waar een verwachte resterende duur bij wordt opgeteld, zoals bij de statistieken in de vorige paragraaf. Bij de statistiek in deze paragraaf worden de verwachte duren namelijk uitgerekend alsof de persoon pas op de peildatum werkloos is geworden. De steekproef wordt uitsluitend gebruikt om de kenmerken van de werkloosheidsinstroom goed in kaart te brengen. Dat de verwachte duren bij een tijdraam van een jaar hoger zijn, komt dus alleen doordat onder de personen die op de eerste peildatum al langer dan drie maanden, maar korter dan een jaar werkloos zijn, kenmerken

¹⁷⁾ Een mogelijke verklaring hiervoor is dat de groep autochtonen heterogener is samengesteld – bijvoorbeeld naar kenmerken als opleiding en werkervaring – dan de groepen allochtonen. Daardoor kan *length biased sampling* bij de autochtonen leiden tot een sterkere selectiviteit m.b.t. kenmerken die met gemiddeld langere werkloosheidsduren samenhangen.

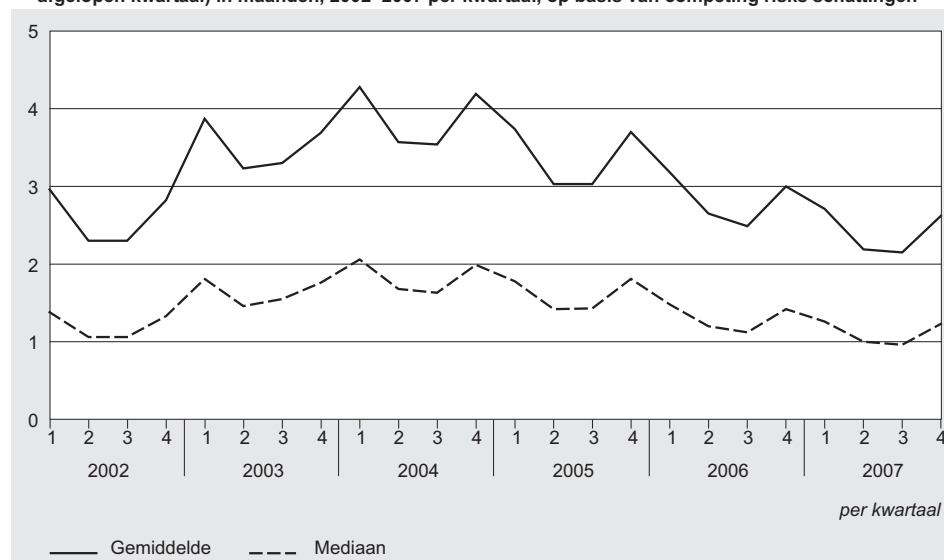
die gunstig zijn voor een snelle werkhervatting al minder voorkomen dan in de groep die pas korter dan drie maanden werkloos is.

Tabel 10
Gemiddelde en mediaan van verwachte werkloosheidsduren in maanden, op basis van kenmerken van instromers in afgelopen kwartaal resp. afgelopen jaar, 2002–2007

Jaar	Duurmodel		Competing risks model	
	Instromers in afgelopen		Instromers in afgelopen	
	kwartaal	jaar	kwartaal	jaar
<i>gemiddelde duren</i>				
2002	2,6	2,8	2,6	2,7
2003	3,6	3,7	3,5	3,7
2004	3,9	4,1	3,9	4,1
2005	3,5	3,6	3,4	3,5
2006	2,9	3,0	2,8	2,9
2007	2,5	2,6	2,4	2,5
2002–2007	3,3	3,4	3,2	3,3
<i>mediane duren</i>				
2002	1,3	1,4	1,2	1,3
2003	1,8	1,8	1,6	1,7
2004	1,9	2,0	1,8	1,9
2005	1,7	1,8	1,6	1,7
2006	1,4	1,5	1,3	1,4
2007	1,3	1,3	1,1	1,2
2002–2007	1,6	1,7	1,5	1,5

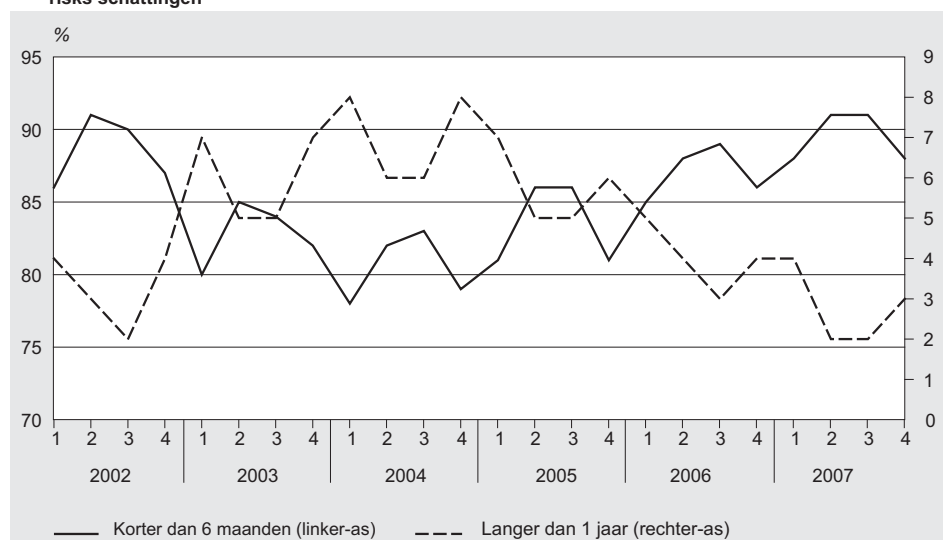
Figuur 9 toont mediaan en gemiddelde van de instroomduur per kwartaal. Hierbij zijn de schattingen van het *competing risks* model gebruikt met als basis de in het afgelopen kwartaal ingestroomde werklozen.

9. Gemiddelde en mediaan van de verwachte werkloosheidsduur van nieuwe werklozen (instromers in afgelopen kwartaal) in maanden, 2002–2007 per kwartaal, op basis van competing risks schattingen



In Tabel 11 zijn de statistieken van de verwachte instroomduren uitgesplitst naar een aantal kenmerken.

10. Percentage verwachte complete werkloosheidsduren van nieuwe werklozen (instromers in afgelopen kwartaal) korter dan 6 maanden respectievelijk langer dan één jaar, 2002–2007, op basis van competing risks schattingen



Anders dan bij de statistieken in de vorige paragraaf geeft de uitsplitsing naar herkomstgroepering hier het verwachte beeld aangezien de statistiek berekend wordt voor de groep van instromende werklozen.

Tabel 11
Gemiddelde verwachte werkloosheidsduren van nieuwe werklozen (instromers in afgelopen kwartaal) in maanden, op basis van competing risks schattingen

	2002	2003	2004	2005	2006	2007
Man	2,8	3,9	4,3	3,6	3,1	2,5
Vrouw	2,4	3,3	3,6	3,2	2,7	2,4
Totaal	2,6	3,5	3,9	3,4	2,8	2,4
15 t/m 24 jaar	1,4	1,9	2,2	1,8	1,5	1,3
25 t/m 34 jaar	2,6	3,3	3,8	3,2	2,6	2,1
35 t/m 44 jaar	3,1	4,2	4,7	3,9	3,3	2,8
45 t/m 54 jaar	3,8	5,2	5,8	4,9	4,2	3,6
55 t/m 64 jaar	3,8	5,0	5,6	5,2	4,6	3,7
Totaal	2,6	3,5	3,9	3,4	2,8	2,4
Basisonderwijs	2,7	3,6	4,1	3,5	2,9	2,6
Vmbo	2,5	3,3	3,7	3,1	2,6	2,2
Mbo/havo/vwo	2,5	3,5	3,9	3,4	2,9	2,5
Hbo	2,6	3,7	4,0	3,4	3,0	2,3
Wo	2,9	4,2	4,4	4,1	3,2	2,7
Totaal	2,6	3,5	3,9	3,4	2,8	2,4
Groningen	2,6	3,8	4,4	3,7	3,3	2,5
Friesland	2,5	3,4	4,3	3,3	3,1	2,5
Drenthe	3,0	3,4	3,9	3,7	3,1	2,4
IJssel-Vecht/Twente	2,9	3,7	4,0	3,5	3,1	2,5
IJssel-Veluwe	2,5	3,5	3,7	3,4	2,6	2,3
Arnhem/O.Gelderl. e.v.	2,6	3,7	4,1	3,4	2,8	2,4
Flevoland	2,7	4,0	4,2	3,6	3,0	2,7
Midden-Nederland	2,2	3,4	3,5	3,1	2,6	2,1
Noord-Holland Noord	1,8	2,5	2,8	2,5	2,0	1,9
Zuidelijk N.-Holland	2,8	3,6	3,9	3,3	2,7	2,4
Rijnstreek	2,5	3,1	3,5	3,1	2,9	2,4
Haaglanden	2,6	3,5	3,8	3,3	2,5	2,4
Rijnmond	2,5	3,5	4,0	3,6	2,9	2,6
Zeeland	2,1	3,9	3,8	3,0	2,9	2,6
Midden en W. Brabant	2,5	3,4	3,7	3,5	2,9	2,5
Noordoost Brabant	2,4	3,0	3,3	2,7	2,5	2,0
Zuidoost Brabant	3,0	3,6	4,2	3,5	3,1	2,4
Limburg	2,6	3,9	4,4	3,8	3,1	2,6
Totaal	2,6	3,5	3,9	3,4	2,8	2,4
Autochtoon	2,5	3,4	3,8	3,3	2,8	2,3
Westers allochtoon	2,8	4,0	4,3	3,8	3,2	2,7
Niet-westers allochtoon	3,1	4,3	4,6	3,8	3,2	2,9
Totaal	2,6	3,5	3,9	3,4	2,8	2,4

5.5 Andere statistieken op basis van competing risks schattingen

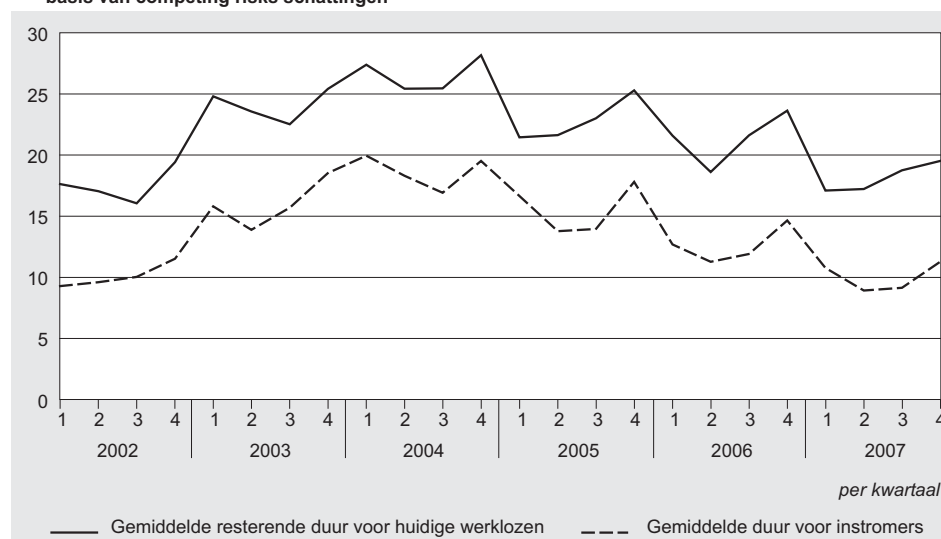
In de twee voorafgaande paragrafen zijn statistieken bepaald waarvoor zowel het gewone duurmodel als het *competing risks* model kunnen worden gebruikt. Beide blijken nagenoeg identieke resultaten te geven. In deze paragraaf wordt tot slot nog ingegaan op een voorbeeld van statistieken die met de schattingen van het *competing risks* model mogelijk worden. Tabel 12 geeft de gemiddelde verwachte duren die zich zouden realiseren als werkloosheidsduren alleen nog maar beëindigd kunnen worden door het vinden van een nieuwe baan. Ter onderscheid van werkloosheidsduren, die immers ook kunnen worden afgesloten zonder dat een nieuwe baan is gevonden, kunnen deze duren worden aangeduid als verwachte baanvindduren, gedefinieerd als de verwachte duur die zich zou realiseren als het zoekproces niet eerder om andere redenen dan het vinden van een nieuwe baan wordt beëindigd.

Tabel 12
Gemiddelde en mediaan van baanvindduur in maanden, voor lopende duren en voor nieuw instromende werklozen, 2002–2007, op basis van schattingen van het competing risks model

Jaar	Aanvullen lopende duren		Verwachte duur van nieuwe werklozen	
	verwachte resterende duur	verwachte volledige duur	in afgelopen kwartaal	in afgelopen jaar
<i>gemiddelde duur (maanden)</i>				
2002	17,6	34,5	10,4	10,7
2003	24,1	39,7	16,0	16,1
2004	26,7	42,6	18,7	18,5
2005	22,9	39,8	15,6	15,3
2006	21,3	41,2	12,6	12,8
2007	18,1	38,8	10,1	10,2
2002–2007	22,3	39,9	14,5	14,5
<i>mediane duur (maanden)</i>				
2002	5,9	14,1	2,4	2,6
2003	8,1	17,4	3,7	3,9
2004	9,3	20,4	4,4	4,5
2005	8,2	20,9	3,7	3,8
2006	7,7	21,6	2,9	3,0
2007	6,3	19,2	2,3	2,4
2002–2007	7,7	19,2	3,3	3,4

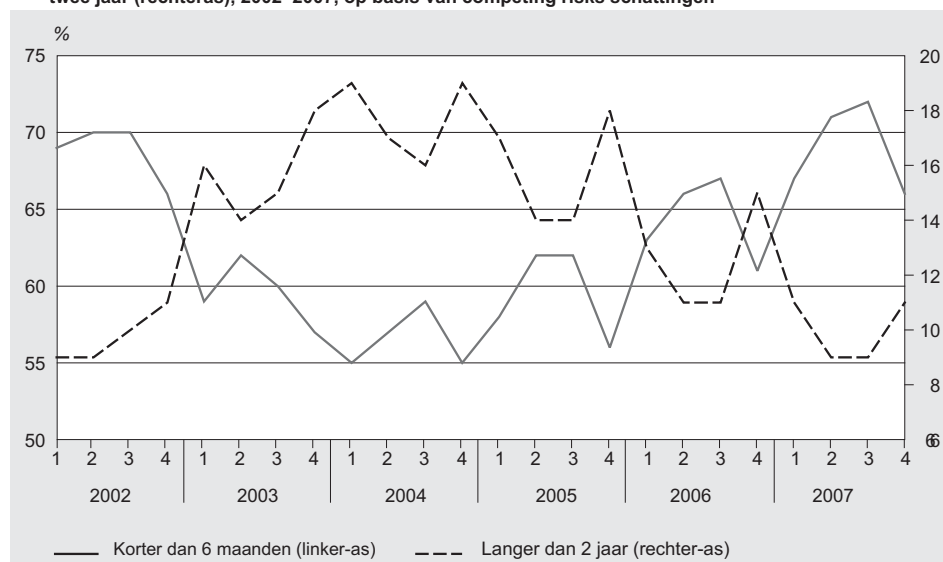
Voor de instroomduur maakt het opnieuw niet zoveel uit of de instromers uit het afgelopen jaar of uit het afgelopen kwartaal worden gebruikt. Deze statistiek is het makkelijkst te interpreteren voor instromers, aangezien bij de aanvulling de *counterfactual* (geen exit langs een andere route dan door het vinden van een baan) daarbij wel op de aanvulling, maar niet op de beginduur betrekking heeft. Bij de instroomduren betreft dit de hele verwachte duur.

11. Gemiddelde duur tot vinden baan (geen andere beëindiging mogelijk) voor lopende werkloosheidsduren en voor nieuwe werklozen (instromers in afgelopen kwartaal) in maanden, 2002–2007 per kwartaal, op basis van competing risks schattingen



Net als in de vorige paragrafen kunnen ook statistieken worden gemaakt van het percentage korte en het percentage lange duren om de betreffende duurverdeling op een andere manier in beeld te brengen. Figuur 12 geeft het percentage van de duren dat korter dan 6 maanden is en het percentage dat langer dan 2 jaar duurt voor de baanvindduren. Tot slot geeft Tabel 13 de uitsplitsing van de gemiddelde baanvindduren naar een aantal achtergrondkenmerken.

12. Percentage verwachte baanvindduren korter dan 6 maanden (linkeras) respectievelijk langer dan twee jaar (rechteras), 2002–2007, op basis van competing risks schattingen



Tabel 13
Gemiddelde verwachte baanvindduren van nieuwe werklozen (instromers in afgelopen kwartaal) in maanden, op basis van competing risks schattingen

	2002	2003	2004	2005	2006	2007
Man	9,2	14,3	16,8	13,9	11,9	12,3
Vrouw	11,3	17,4	20,2	16,7	13,1	13,2
Totaal	10,4	16,0	18,7	15,6	12,6	10,1
15 t/m 24 jaar	3,9	6,1	7,1	5,6	4,6	3,4
25 t/m 34 jaar	6,1	9,1	10,8	8,7	6,6	5,0
35 t/m 44 jaar	10,4	15,4	18,3	14,6	11,3	8,8
45 t/m 54 jaar	18,6	27,7	31,6	24,3	20,2	15,2
55 t/m 64 jaar	53,0	65,0	78,7	59,1	46,7	35,8
Totaal	10,4	16,0	18,7	15,6	12,6	10,1
Basisonderwijs	18,6	24,8	33,6	24,6	21,2	14,9
Vmbo	10,5	16,4	18,9	15,7	12,4	10,2
Mbo/havo/vwo	9,1	14,5	16,6	14,8	11,7	9,9
Hbo	8,9	14,2	15,2	11,5	10,8	8,3
Wo	6,8	14,3	15,3	15,8	11,2	8,6
Totaal	10,4	16,0	18,7	15,6	12,6	10,1
Groningen	14,8	20,2	21,8	18,6	15,7	11,1
Friesland	9,7	14,9	19,5	15,9	14,0	9,4
Drenthe	11,8	17,7	20,6	19,6	13,9	10,4
IJssel-Vecht/Twente	11,7	16,9	17,4	15,4	16,7	12,3
IJssel-Veluwe	10,4	12,9	14,4	13,0	11,9	9,2
Arnhem/O.Gelderl. e.v.	10,9	16,8	20,8	16,0	13,0	9,9
Flevoland	11,6	14,1	19,4	15,4	11,2	13,0
Midden-Nederland	7,1	13,8	14,1	13,4	9,5	7,0
Noord-Holland Noord	5,6	12,2	13,2	10,4	9,1	6,2
Zuidelijk N.-Holland	7,9	14,4	16,0	12,9	9,8	10,2
Rijnstreek	10,1	13,0	17,1	15,0	11,9	8,8
Haaglanden	10,3	15,6	18,4	15,5	10,3	9,1
Rijnmond	10,8	18,3	23,4	17,3	13,5	11,3
Zeeland	9,3	15,5	16,6	13,0	12,9	11,5
Midden en W. Brabant	9,2	14,5	16,8	14,7	11,6	10,0
Noordoost Brabant	8,4	14,0	15,8	13,1	11,0	9,3
Zuidoost Brabant	10,5	14,0	20,1	17,2	14,2	10,9
Limburg	12,2	20,5	23,2	18,2	15,9	10,9
Totaal	10,4	16,0	18,7	15,6	12,6	10,1
Autochtoon	9,8	14,8	17,6	14,6	11,8	9,5
Westers allochtoon	11,8	21,2	19,6	17,1	15,1	11,0
Niet-westers allochtoon	13,2	21,3	26,5	20,5	16,2	12,9
Totaal	10,4	16,0	18,7	15,6	12,6	10,1

5.6 Statistieken op basis van transitie uit werkloosheid

De gegevens uit de EBB die gebruikt zijn voor het schatten van de duurmodellen kunnen ook op een directere manier worden gebruikt om statistieken over werkloosheidsduren samen te stellen. Bij het schatten van de duurmodellen is gebruik gemaakt van de informatie over de arbeidsmarktstatus van personen die in de peiling van een kwartaal daarvoor werkloos waren. Dit soort informatie kan ook worden gepresenteerd in de vorm van een transitietabel, zoals Tabel 3 en Tabel 4 in hoofdstuk 3. Dit soort tabellen heeft als voordeel dat ze de waargenomen informatie direct weergeven en eenvoudig te interpreteren zijn. Bovendien geven ze op een natuurlijke wijze het *competing risks* aspect van dit duurproces weer. Naarmate de tijd tussen twee peildata langer is, wordt echter de kans groter dat in de tussentijd meer dan één transitie heeft plaatsgevonden en dat de weergegeven transitie tussen de peildata in een dergelijke tabel niet meer geïnterpreteerd kunnen worden als transitie vanuit de werkloosheidstoestand op de eerste peildatum.¹⁸⁾ Aangezien deze laatste interpretatie relevant is voor statistieken van werkloosheidsduren, is er hier voor gekozen om in de transitietabel aan te geven hoeveel van de werklozen op het eerste peilmoment voor een volgend peilmoment werk heeft gevonden.

Het gaat dan om statistieken die de uitstroom of transitie vanuit werkloosheid tonen. Hierin wordt op basis van de vervolgpeilingen van de EBB bekeken welk deel van de werklozen in de daaropvolgende peiling is uitgestroomd uit werkloosheid. Dit kan zijn omdat ze een baan hebben gevonden ofwel omdat de persoon niet meer participeert op de arbeidsmarkt. Dergelijke statistieken maken geen gebruik van een parametrisch model. Zij zijn eenvoudig samen te stellen en te interpreteren. Aangezien er vier vervolgpeilingen zijn die om het kwartaal plaatsvinden, kunnen de personen die op het eerste peilmoment van de EBB werkloos zijn in principe een jaar lang worden gevolgd. Op deze manier kan worden bepaald welk deel van de werklozen na drie, zes, negen of twaalf maanden uit werkloosheid is gestroomd.

Een voorbeeld van een dergelijke tabel is weergegeven in Tabel 14. Hierin zijn werklozen genomen die in het eerste kwartaal van 2006 hebben deelgenomen aan de eerste *face-to-face* peiling van de EBB. Deze zijn gevolgd tot en met het eerste kwartaal van 2007. De tabel is uitgesplitst naar de op het eerste peilmoment reeds verstreken werkloosheidsduur. Er is vermeld welk deel van de werklozen is uitgestroomd uit werkloosheid omdat ze een baan hebben gevonden van 12 uur of meer per week en het deel dat is uitgestroomd omdat ze niet meer participeren op de arbeidsmarkt. Verder zijn onderaan in tabel 14 beide percentages opgeteld, wat de cumulatieve uitstroomkans uit werkloosheid geeft, dat wil zeggen: het deel van de werklozen dat is uitgestroomd omdat ze een baan hebben gevonden dan wel omdat ze niet meer participeren.

Uit tabel 14 blijkt dat van de personen die op een bepaald moment (het eerste peilmoment) werkloos zijn 23 procent binnen drie maanden een baan vindt. Dit aandeel is na twaalf maanden opgelopen tot 40 procent. Verder wordt duidelijk dat de verstreken werkloosheidsduur van belang is bij het vinden van werk. Vooral veel personen die 0–3 maanden werkloos waren op het eerste peilmoment hebben na twaalf maanden betaald werk (6 van de tien). Onder personen die op het eerste peilmoment al 12 maanden of langer werkloos waren, is dit aandeel een stuk lager (ongeveer een kwart). Wanneer wordt gekeken naar de uitstroom van werklozen naar non-participatie dan geldt juist het omgekeerde. Vooral degenen die al lang werkloos waren, stromen uit naar non-participatie. Dit mag ook worden verwacht. Met name langdurig werklozen zullen op den duur ontmoedigd raken en stoppen met zoeken naar werk. Van degenen die bij de eerste peiling al 12 maanden of meer werkloos waren, is 63 procent een jaar later uitgestroomd naar non-participatie. Bij degenen die 0 tot 3 maanden werkloos waren, is dit 36 procent. Tot slot laten de resultaten van de uitstroom uit werkloosheid, die volgen als we beide transitie samennemen, zien dat het overgrote deel van de werklozen na een jaar is uitgestroomd uit werkloosheid (92 procent). Op basis van het voorgaande kan worden geconcludeerd dat hierbij degenen die kort werkloos waren vooral een baan hebben gevonden, terwijl degenen die langer werkloos waren vaak zijn uitgestroomd naar non-participatie.

¹⁸⁾ Zie Kerkhofs (2006) voor een bespreking hiervan in de context van baan-baanmobiliteit. Overigens is er bij het schatten van de duurmodellen impliciet vanuit gegaan dat een kwartaal voldoende kort is en dat eventuele

Tabel 14
Aandeel werklozen dat na drie, zes, negen of twaalf maanden is uitgestroomd uit werkloosheid, naar verstreken werkloosheidsduur, voor werklozen in het 1e kwartaal van 2006

Werklozen in 1e kwartaal 2006	Uitgestroomd naar werk			
	drie maanden	zes maanden	nege maanden	twaalf maanden
	%			
Verstreken werkloosheidsduur				
0- 3 maanden	39	54	59	60
4- 6 maanden	25	36	40	43
7-11 maanden	21	37	45	45
≥12 maanden	13	19	24	26
Totaal	23	33	39	40
	Uitgestroomd naar non-participatie			
	drie maanden	zes maanden	nege maanden	twaalf maanden
	%			
Verstreken werkloosheidsduur				
0- 3 maanden	32	35	36	36
4- 6 maanden	28	40	46	51
7-11 maanden	25	38	42	45
≥12 maanden	39	55	59	63
Totaal	34	45	49	52
	Uitgestroomd naar werk of non-participatie			
	drie maanden	zes maanden	nege maanden	twaalf maanden
	%			
Verstreken werkloosheidsduur				
0- 3 maanden	71	89	95	97
4- 6 maanden	54	77	86	94
7-11 maanden	46	75	87	91
≥12 maanden	51	73	84	89
Totaal	57	78	88	92

Vergelijkbare tabellen zijn te maken voor alle kwartalen van de periode 2002-2007. Wanneer de tabellen verder worden uitgesplitst naar achtergrondkenmerken wordt duidelijk welke personen in welke tijdsperiode relatief snel zijn uitgestroomd uit werkloosheid. Ook is het mogelijk om de nieuwe instroom van werklozen in het kwartaal als uitgangspunt te nemen van de statistieken. In dat geval worden alléén de personen uit Tabel 14 geselecteerd die 0-3 maanden werkloos waren.

Zoals gezegd is het voordeel van transitietabellen dat ze relatief eenvoudig zijn samen te stellen en te interpreteren. Een nadeel is dat deze statistieken gevoelig zijn voor selectieve paneluitval. De kans bestaat dat personen die moeilijk aan werk komen niet meewerken in de vervolgpelingen van de EBB. Duurmodellen bieden daarentegen de mogelijkheid om in de statistieken hiervoor te corrigeren, voor zover de selectiviteit betrekking heeft op waargenomen kenmerken. Wat betreft de uitval geldt voor het jaar 2006 dat van de werklozen die aan de eerste peiling van de EBB hebben deelgenomen in totaal 14 procent is uitgevallen in de tweede peiling. De uitval neemt in de daaropvolgende peilingen verder toe tot 28 procent in de vijfde peiling. De hier gepresenteerde cijfers zijn gebaseerd op ongewogen cijfers. Er komt binnenkort een weging beschikbaar voor deze tabellen, waarbij rekening wordt gehouden met steekproefuitval. Daarbij wordt zoveel mogelijk aangesloten bij de huidige weging van de EBB. Deze is ook gebaseerd op alle vijf de peilingen van de EBB, en hiermee worden de huidige standcijfers over werkloosheid gemaakt.

Verder geldt voor de verstreken werkloosheidsduur uit Tabel 14 dat deze is gebaseerd op het moment waarop personen zijn gaan zoeken naar werk en de datum van de laatste baan. Bij de beantwoording van deze vragen hebben respondenten vaak de neiging om af te ronden op halve en hele jaren. De vraag is bijvoorbeeld in hoeverre personen die aangeven 6 maanden werkloos te zijn, dit in werkelijkheid zijn? Nadere analyse laat

zien dat personen met een verstreken werkloosheidsduur 6 maanden minder snel werk vinden dan personen met een verstreken werkloosheidsduur van 7, 8 of 9 maanden. Dit is ook de reden waarom in Tabel 14 de verschillen in de uitstroom naar werk tussen degenen met een verstreken werkloosheidsduur van 4–6 maanden en degenen met 7–11 maanden niet groot zijn. Aangezien er een duidelijke relatie is tussen de verstreken werkloosheidsduur en de kans op werk wordt bekeken op welke manier hiermee om moet worden gegaan. Er zijn twee manieren. De eerste is door een zogenoemde 'piekencorrectie' waarbij personen die 6 maanden werkloos zijn opnieuw worden onderverdeeld. De tweede manier is door de categorieën van verstreken werkloosheidsduur anders te kiezen, bijvoorbeeld 0–3 maanden, 4–11 maanden en 12 maanden of meer. Hierbij wordt er dan vanuit gegaan dat personen die aangeven 6 maanden werkloos te zijn binnen de categorie 4–11 maanden werkloos vallen. Voorlopig wordt voor deze tweede optie gekozen bij het samenstellen van transitietabellen.

Een ander nadeel van transitietabellen kan zijn dat de aantallen al snel te klein zijn om betrouwbare cijfers over samen te kunnen stellen. Dit blijkt op het moment dat de tabellen worden uitgesplitst naar achtergrondkenmerken – bijvoorbeeld naar leeftijd en herkomst – en op het moment dat Nederland maar weinig werklozen kent, wat vooral in de meest recente jaren het geval is.

Een mogelijkheid om met deze kleine aantallen om te gaan, is om bij de bepaling van de werklozen het gehele panel te gebruiken. Dit kan worden gedaan voor de instroom in werkloosheid. Het gaat dan om personen die in de eerste peiling 0–3 maanden werkloos zijn, en om de personen die in één van de vervolgpelingen werkloos zijn geworden. Voor deze werklozen kan worden aangegeven welk deel drie maanden later een baan heeft gevonden dan wel is uitgestroomd naar non-participatie. In Tabel 15 wordt dit weergegeven. Voordeel hiervan is dat van het gehele panel van de EBB gebruik wordt gemaakt om de nieuwe instroom in werkloosheid te bepalen. Hierdoor liggen de aantallen veel hoger. Dit maakt het goed mogelijk om bij het samenstellen van statistieken uitsplitsingen te maken naar verschillende achtergrondkenmerken. Nadeel is echter dat alléén maar drie maanden verder kan worden gekeken, aangezien van personen die zich in de vierde peiling bevinden alléén nog de vijfde peiling resteert. Bovendien is een nadeel dat hiermee niet in beeld kan worden gebracht hoe lang langdurig werklozen er over doen om een baan te vinden.

Tabel 15
Aandeel ingestroomde werklozen dat na drie maanden is uitgestroomd uit werkloosheid, voor werklozen in het 1e kwartaal van 2006

Werklozen in 1e kwartaal 2006	Uitgestroomd naar werk	Uitgestroomd naar non-participatie	Totaal uitgestroomd vanuit werkloosheid
	%		
Totaal	21	29	49
Man	23	25	48
Vrouw	19	32	51
Autochtoon	22	30	51
Westerse allochtoon	17	24	41
Niet-westerse allochtoon	15	28	43
Leeftijd 15–24 jaar	32	34	66
Leeftijd 25–34 jaar	32	24	55
Leeftijd 35–44 jaar	18	30	48
Leeftijd 45–54 jaar	13	26	39
Leeftijd 55–64 jaar	8	32	40

6. Conclusies

In dit onderzoek zijn de mogelijkheden onderzocht om met behulp van duuranalyse statistieken te ontwikkelen van verwachte werkloosheidsduren. Daarbij is gebruik gemaakt van de gegevens uit de EBB voor de periode van 2002 tot en met 2007. Directe analyse van de CAPI-informatie over lopende of afgeronde werkloosheidsduren uit de EBB geeft door de steekproefstructuur een vertekend beeld van de werkelijke verdeling van de verwachte werkloosheidsduren. Om die reden is gebruik gemaakt van econometrische analyse van (parametrische) duurmodellen. Voor het correct schatten van de parameters in de duurmodellen bleek bovendien de informatie uit het roterend panel van de EBB essentieel. Naast een 'gewoon' duurmodel is ook gebruik gemaakt van een *competing risks* model, waarin niet alleen de werkloosheidsduur zelf wordt gemodelleerd, maar ook de manier waarop deze duur wordt afgesloten, te weten met dan wel zonder nieuwe baan. De schattingsresultaten voor beide soorten modellen zijn plausibel en robuust. Bovendien blijken de daarmee afgeleide statistieken van de verwachte werkloosheidsduren nagenoeg gelijk. Daarmee bieden de geschatte modellen dus een uitstekende basis voor het afleiden van statistieken over verwachte werkloosheidsduren.

Met behulp van de geschatte duurmodellen is vervolgens een aantal verschillende statistieken van verwachte werkloosheidsduren afgeleid. Allereerst zijn dat statistieken van de werkloosheidsduren die ontstaan door in een bestand van werklozen de al verstreken werkloosheidsduur door imputatie te 'voltoeien'. Daarnaast zijn ook statistieken van verwachte werkloosheidsduren bepaald voor een startcohort van personen die in een bepaalde (korte) periode werkloos zijn geworden. Tot slot zijn, om de extra mogelijkheden van de *competing risks* schattingen te benutten, ook statistieken van baanvindduren afgeleid. Deze geven aan hoelang werklozen erover doen om een nieuwe baan te vinden aangenomen dat zij zich niet voortijdig van de arbeidsmarkt terugtrekken.

Om na te gaan aan welke statistieken de voorkeur moet worden gegeven, zijn ook twee andere statistieken van belang die niet zijn gebaseerd op duurmodellen. Dat zijn statistieken op basis van de waargenomen werkloosheidsduren tot aan de peildatum en statistieken van de transities vanuit werkloosheid op basis van de informatie in het roterend panel van de EBB.

Van statistieken van werkloosheidsduren mag worden verwacht dat zij de informatie verschaffen die een goed beeld geven van werkloosheid en daarmee bij voorkeur informatie geven die voor beleidsmakers relevant is. Wat dat laatste betreft, winnen werkloosheidsstatistieken aan relevantie als zij beter de actuele situatie op de arbeidsmarkt in beeld brengen, in het bijzonder de kansen op het vinden van werk van mensen die nu werkloos zijn. Daarnaast is bij statistieken in het algemeen van belang dat zij betrouwbaar – en dus betrouwbaar te meten – zijn, dat zij dicht bij de verzamelde gegevens blijven en dat zij transparent zijn en eenvoudig te begrijpen. Hoe valt dan het oordeel uit als we met dit wensenpakket in het achterhoofd de verschillende statistieken bekijken?

1. *Werkloosheidsduren op basis van duurmodellen*

Kennis van de verdeling van werkloosheidsduren biedt een veelheid aan mogelijkheden om deze door middel van statistieken te karakteriseren en daarmee beleidsmatig relevante informatie te geven over de werkloosheid en de actuele situatie op de arbeidsmarkt. Naast het boekstaven van de tijd die de werklozen in de voorafgaande maanden en jaren al zonder werk zitten, kunnen hiermee ook de perspectieven voor verschillende groepen werklozen zichtbaar worden gemaakt.

De meest voor de hand liggende insteek is om de verwachte werkloosheidsduur in kaart te brengen voor de personen die in een peilperiode als werkloos worden waargenomen. Dergelijke gegevens geven een goed beeld van het actuele werklozenbestand. Nadeel is dat dit een stock sample is waarin lange duren zijn oververtegenwoordigd. Hierdoor wordt weliswaar de beleidsmatig uiterst relevante groep van langdurig werklozen goed in beeld gebracht, maar kunnen deze cijfers niet direct worden gebruikt om uitspraken te

doen over de werkloosheidsduren van mensen die net werkloos zijn geworden. Daarom is het een interessant alternatief om in statistieken de verwachte duren weer te geven voor de nieuwe instroom van werklozen, bijvoorbeeld de werklozen die in het afgelopen kwartaal werkloos zijn geworden (instroomduren).

Tegenover deze mogelijkheden staat als voornaamste bezwaar dat voor een dergelijke aanpak een statistisch model moet worden gebruikt met de bijbehorende modelveronderstellingen. Tussen de ruwe data en de statistiek staat als het ware een raffinageproces dat de nodige zorg behoeft. De statistische analyse zelf heeft echter ook als voordeel dat het de mogelijkheid biedt om te corrigeren voor selectieve paneluitval in de vervolgpelingen van de EBB voor zover deze betrekking heeft op waargenomen kenmerken.

2. Werkloosheidsduur tot peilmoment

Dit zijn statistieken over de reeds verstreken werkloosheidsduur op peilmoment (onvoltooide werkloosheidsduren). Deze statistieken zijn eenvoudig samen te stellen en te interpreteren. Nadeel is echter dat deze duren incompleet zijn en onbekend is hoelang zij nog doorlopen. Bovendien zijn in een dergelijke stock sample lange duren oververtegenwoordigd, evenals personen met kenmerken die samenhangen met gemiddeld langere werkloosheidsduren. Deze statistieken geven dus terugkijkend in de tijd een beeld van de mensen die op de peildatum (nog) werkloos zijn.

3. Stromen van werkloosheid

Een manier om stromen c.q. transities in kaart te brengen is door directe tellingen op basis van de vervolgpelingen van de EBB. Met de vervolgpelingen kan worden bepaald welk deel van de werklozen na drie, zes, negen of twaalf maanden werk heeft gevonden. Het voordeel hiervan is dat deze statistieken eenvoudig zijn samen te stellen en te interpreteren. Bovendien belichten dergelijke statistieken de actuele situatie op de arbeidsmarkt. Nadeel is echter dat er rekening moet worden gehouden met selectieve paneluitval in de vervolgpelingen van de EBB, een beperking waarvoor in de statistieken op basis van de duurmodellen impliciet is gecorrigeerd door bij het schatten van het duurmodel te conditioneren op de belangrijkste waargenomen kenmerken en bij het berekenen van de statistieken als basis een steekproef uit de EBB te nemen waarop paneluitval niet of maar beperkte invloed heeft. Voor statistieken op basis van de waargenomen transities tussen twee peilmomenten in de EBB is het van belang om na te gaan of voor de invloed van paneluitval kan worden gecorrigeerd door gebruik te maken van weegfactoren.

Als we op deze manier de balans opmaken, kan worden geconstateerd dat elk van de statistieken voor- en nadelen heeft en dat zij bovendien aanvullende informatie geven in de zin dat elk andere aspecten van werkloosheid en verwachte werkloosheidsduren belicht. In dat opzicht is het wellicht raadzaam om niet zozeer één van deze statistieken te kiezen, maar te overwegen de statistieken, eventueel met wisselende frequenties, als pakket verder te ontwikkelen.

Referenties

- Abramowitz, M. En I. A. Stegun, (eds.) (1972), *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. New York: Dover.
- Berkhout, P.H.G., C. van Klaveren, J. van der Valk en R. de Vries (2007), *De werkloosheidsduur in Nederland (2002–2005)*, SEO-rapport nr. 2007-51, Amsterdam: Stichting voor Economisch Onderzoek.
- Blossfeld, H.-P., K. Golsch en G. Rohwer (2007), *Event history analysis with Stata*, New Jersey: Lawrence Erlbaum Associates.
- Kerkhofs, M.J.M. (2006) 'Gaanders en staanders op de arbeidsmarkt: baanmobiliteit in Nederland', in: Ester, P., R. Muffels & J. Schippers (2006) *Dynamiek en Levensloop, de arbeidsmarkt in transitie*, Koninklijke Van Gorcum: Assen, blz. 97–117.
- Heckman, J.J. en B. Singer (1984), 'Econometric Duration Analysis,' *Journal of Econometrics*, vol. 24, pp. 63–132.
- Heckman, J.J. en B. Singer (1985), 'Erratum,' *Journal of Econometrics*, vol. 27, pp. 137–138.
- Jenkins, S.P. (1995), 'Easy Estimation Methods for Discrete-Time Duration Models,' *Oxford Bulletin of Economics & Statistics*, pp. 129–38.
- 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: Cambridge University Press.
- Singer, J.D. en J.B. Willett (2003), *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*, Oxford: Oxford University Press.
- Theeuwes, J.J.M., M.J.M. Kerkhofs en M. Lindeboom (1988), *Toestanden, overgangen en duren op de Nederlandse arbeidsmarkt 1980–1985*, OSA-werkdocument W49, Den Haag: OSA.

Bijlage 1: Gesimuleerde duren

Aangezien voor de analyses in dit onderzoek weliswaar gebruik kan worden gemaakt van het statistisch pakket Stata (Blossfeld, Golsch en Rohwer, 2007), maar daarbinnen eigen programma's moeten worden ontwikkeld, is een databestand gesimuleerd dat is gebruikt om deze zelfgemaakte routines uit te proberen. Qua structuur is dit bestand vergelijkbaar met de gegevens uit de EBB inclusief vervolgpelingen en omdat het datagenerend proces van deze data volledig bekend is (en gevarieerd kan worden) is het hiermee mogelijk om na te gaan of de geprogrammeerde routines en eventuele programma's van derden goed werken. Door de manier waarop dit bestand is gegenereerd, kan het echter ook worden gebruikt om de rol van de in de paragraaf 2.2 genoemde complicaties te belichten.

Het gesimuleerde bestand bestaat uit 100 000 duren. Die duren horen bij gefingeerde personen die alleen qua sexe kunnen verschillen. De sexe is willekeurig getrokken met kans $\frac{1}{2}$ op $\text{sexe}=0$ en kans $\frac{1}{2}$ op $\text{sexe}=1$. Vervolgens zijn de duren willekeurig getrokken uit de duurverdeling bij de uitstroombintensiteit

$$\theta(d, \text{sexe}) = \exp(0 + 0,2 \cdot \text{sexe}) \cdot 0,8 \cdot d^{-0,2}.$$

De uitstroombintensiteit is het product van een stuk dat afhangt van persoonskenmerken en een stuk dat de duurzaamheidsbeschrijving (de *proportional hazard* specificatie). De duurzaamheidsbeschrijving is een zogenaamde Weibull-specificatie en doordat de Weibull parameter kleiner is dan 1 ($\alpha=0,8$) is de duurzaamheidsbeschrijving negatief en wordt de uitstroombintensiteit kleiner naarmate iemand langer naar werk zoekt. Alle heterogeniteit is waargenomen en mannen en vrouwen met $\text{sexe}=1$ hebben een hogere uitstroombintensiteit (en dus kortere werkloosheidsduren) dan die met $\text{sexe}=0$.

Deze collectie duren is dus een willekeurige steekproef uit de verdeling van de duren, en is te zien als een *flow sample*, een steekproef van instromers op een gegeven tijdstip (of in een gegeven periode)

Om een *stock sample* te simuleren, is bij elke duur een willekeurig aanvangsmoment getrokken tussen 1 januari 1958 en 1 januari 2008 (uniform verdeeld) en een willekeurige enquêtedatum tussen 1 januari 2002 en 1 januari 2008. Conform EBB-structuur van de te analyseren echte data zijn vervolgens 4 vervolgpelingen ingeroosterd, gemiddelde na 3, 6, 9 en 12 maanden (elk van deze data willekeurig verdeeld tussen 0,6 maand vroeger en 0,6 maand later)

De complete verzameling van 100 000 duren kan dus gezien worden als *flow sample* en correspondeert met wat je als de echte duurverdeling zou kunnen zien. Daarnaast kunnen twee andere duurverdelingen worden afgeleid: de verdeling van de volledige duren in de *stock sample* van personen die op de eerste peildatum werkloos waren en de verdeling van de duren die de personen in deze steekproef op die peildatum al werkloos waren. Deze drie verdelingen zijn in Tabel B.1 beschreven. De eerste twee verdelingen hebben betrekking op volledige (afgesloten) duren. Het effect van *length-biased sampling* is op een aantal manieren zichtbaar. Allereerst is de gemiddelde duur in de echte verdeling 12 maanden en in de *stock sample* bijna 32 maanden. Ook aan de percentielen is goed te zien dat de kansverdeling van de duren in de *stock sample* sterk naar rechts is verschoven.

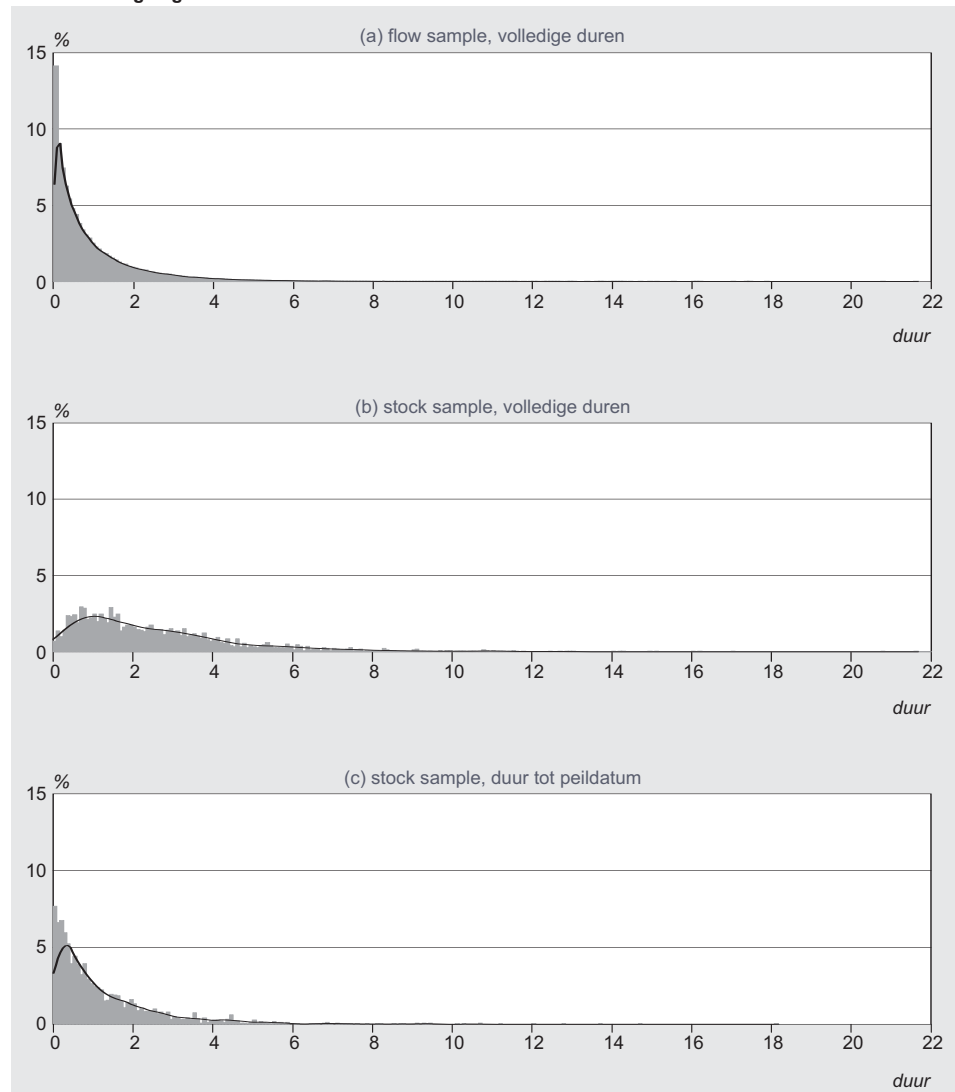
In de derde steekproef is er in vergelijking met de verdeling in de eerste kolom sprake van twee tegengestelde effecten. Allereerst is het een *stock sample* waardoor de gemiddelde duur systematisch wordt overschat. De duren in de steekproef zijn echter rechtsgesensureerde duren, waardoor ze systematisch de totale duur tot het vinden van een nieuwe baan onderschatten. In duurmodellen zonder duurzaamheidsbeschrijving (exponentiële duurmodellen) vallen deze effecten gemiddeld precies tegen elkaar weg. Door de lichte vorm van negatieve duurzaamheidsbeschrijving in de gesimuleerde data gaat dat hier al niet meer op en is de gemiddelde duur in deze *stock sample*, ondanks het feit dat alleen het eerste stuk van de duur wordt meegeteld, bijna 4 maanden langer dan in de achterliggende echte duurverdeling.

Tabel B.1
Verdelingen van de gesimuleerde duren

	Flow sample	Stock sample (werklozen op peildatum ¹⁾)	
	volledige duur	volledige duur	duur tot peildatum
Aantallen			
allen	100 000	2 094	2 094
sexe=0	49 974	1 137	1 137
sexe=1	50 026	957	957
Gemiddelde duur (maanden)			
allen	12,0	31,9	15,9
sexe=0	13,6	35,0	17,3
sexe=1	10,5	28,4	14,3
Percentielen (maanden)			
10%	0,6	6,3	1,4
25%	2,2	12,5	3,7
50%	6,7	24,6	9,5
75%	15,9	42,5	21,0
90%	30,3	66,0	37,2

Aan de aantallen waarnemingen is ook de selectie op – in dit geval waargenomen – kenmerken zichtbaar. Sexe 0 en 1 waren gelijk verdeeld en in de totale steekproef komt *sexe=1* zelfs iets meer voor. Deze groep heeft een hogere uitstroomintensiteit en dus gemiddeld genomen kortere duren. Per saldo komen ze daarom minder voor in de *stock sample*, waarvan ze slechts 46 procent uitmaken.

B1.1 Verdelingen gesimuleerde duren



Figuur B1.1 is een grafische representatie van de drie duurverdelingen. De staven zijn histogrammen, de doorgetrokken lijn een niet-parametrische kernelregressie door de individuele duren. Het mag duidelijk zijn dat de verdelingen sterk kunnen verschillen en dat het recht-toe-recht-aan schatten van een duurverdeling op basis van een *stock sample* ((b) of (c)) in het algemeen een sterk vertekend beeld zal geven van de achterliggende duurverdeling en daardoor niet geschikt is voor het ontwikkelen van statistieken, of deze nu gebaseerd zijn op het aanvullen van reeds gedeeltelijk waargenomen lopende duren of op het berekenen van verwachte duren voor nieuw ingestroomde werklozen.

Zoals gezegd, leiden de tot de peildatum waargenomen duren door de *length biased sampling* tot overschatting, maar door het ontbreken van de resterende duur tot onderschatting van de werkloosheidsduren. Bij negatieve duurzaamheidsduren – de meest aanneemelijke situatie voor werkloosheidsduren – overheerst het eerste effect en door een duurmodel te schatten waarbij deze duren als voltooide duren worden behandeld, wordt een duurverdeling als in figuur B1.1b gevonden, met een oververtegenwoordiging van langere duren. Door op een vergelijkbare manier de nog lopende duren in de vervolgpelingen te analyseren, wordt deze opwaartse vertekening alleen nog maar sterker aangezien het effect van *length biased sampling* dan groter is. Bij de nog lopende duren in de vervolgpelingen wordt namelijk impliciet als selectie criterium gebruikt dat de persoon op de eerste peildatum werkloos is – waardoor in de steekproef langere duren oververtegenwoordigd zijn – en dat de duur minstens tot de betreffende vervolgpeling voortduurt. Tabel B1.2 laat zien hoe deze vorm van selectiviteit de duurverdeling in de vervolgpelingen nog verder naar rechts trekt.

Tabel B1.2
Gemiddelde duur op data vervolgpelingen in gesimuleerde data

	Gemiddelde duur (maanden)		N
	Volledige duur	Rechts gecensureerde duur (tot peildatum)	
Flow sample (alle duren)	12,0		100 000
Stock sample: personen die werkloos zijn op alle peildata tot en met peiling			
1	31,9	15,9	2 094
2	31,8	18,8	2 005
3	36,3	22,4	1 513
4	40,7	26,1	1 174
5	44,9	29,8	917

De vervolgpelingen bieden echter meer toepassingsmogelijkheden. Allereerst maken de vervolgpelingen het mogelijk om in de schattingsmethode expliciet rekening te houden met het feit dat de nog lopende duren onvolledig zijn waargenomen. Met alleen de duren uit de eerste peiling is dat niet mogelijk, omdat er daarvoor in de steekproef ook volledig waargenomen duren moeten zitten. Met de informatie uit vervolgpelingen kunnen de lopende duren uit de eerste peiling worden aangevuld en een deel van de duren zal daarmee worden afgesloten, waardoor voor die waarnemingen de volledige duur is waargenomen. Door expliciet rekening te houden met *right censoring* wordt de *stock sample* verdeling correct geschat, maar hierdoor resteert de overschatting door *length biased sampling* als enige vertekening, wordt een verdeling als in (b) in Figuur B.1.1 geschat, waarmee de achterliggende duurverdeling uiteindelijk nog sterker wordt overschat.

Daarnaast bieden de vervolgpelingen echter ook de mogelijkheid om nieuw instromende werklozen aan de steekproef toe te voegen. Door deze informatie ook te gebruiken wordt de overschatting getemperd, maar in feite wordt in dat geval een nauwelijks traceerbare mengvorm van *stock sample* en *flow sample* verdeling geschat.¹⁹⁾ Als alleen de duren van deze nieuwe instromers worden gebruikt, zou de achterliggende werkloosheidsver-

¹⁹⁾ In feite is in Berkhout *et al.* (2007) iets vergelijkbaars gedaan door aan de steekproef uit de eerste peilingen van de EBB de afgeronde duren toe te voegen van de mensen die in het afgelopen jaar een baan hadden gevonden. De relatie tussen de resulterende steekproef en de achterliggende duurverdeling is daardoor helemaal onduidelijk.

deling wel zonder vertekening worden geschat, maar daarmee kan alleen iets worden gezegd over de verdeling van duren die korter zijn dan de periode tussen de eerste en de laatste peiling. In het geval van de EBB dus alleen van duren korter dan een jaar.

Hoewel dit niet in eerste instantie de motivatie was voor het gebruiken van de vervolgpelingen van de EBB, is het toch de extra informatie uit de vervolgpelingen die het mogelijk maakt om met behulp van de *delayed entry* aanpak de correcte achterliggende duurverdeling te schatten. In essentie komt dit er op neer dat alleen de gebeurtenissen vanaf het eerste peilmoment worden geanalyseerd. De kansprocessen die deze gebeurtenissen bepalen, zijn niet van invloed geweest op de steekproeftrekking en de eigenschappen daarvan kunnen daarom direct uit de achterliggende duurverdeling worden afgeleid. Daardoor kunnen deze gegevens direct worden gebruikt om de onbekende parameters van de achterliggende duurverdeling te schatten. Hierbij wordt de informatie over de periode voor de eerste peiling wel gebruikt, maar uitsluitend als exogene informatie.

Bijlage 2: Gebruikte variabelen

Voor de schatten van de duurmodellen is op basis van de gegevens uit de vijf EBB-peilingen – het eerste *face-to-face* interview en de 4 telefonische vervolgpelingen – een persoon-periode bestand samengesteld, waarin de waarnemingseenheden (cases) de periodes zijn tussen twee opeenvolgende peilingen voor zover de betreffende persoon op de eerste van deze peilingen werkloos was en de arbeidsmarktsituatie op het volgende peilmoment bekend is. Tabel B2.1 geeft een overzicht van de variabelen uit dit bestand die direct in de schattingen zijn gebruikt.

Tabel B2.1
Overzicht van variabelen die gebruikt zijn in de analyses

	Variabele	Aantal waarnemingen	Gemiddelde	Minimum	Maximum
<i>Geslacht (gsl)</i>					
man	(ref.)	37 848	0,438	0	1
vrouw	vrouw	37 848	0,562	0	1
<i>Leeftijd (lft)</i>					
15–19	lft_1519	37 848	0,105	0	1
20–24	lft_2024	37 848	0,112	0	1
25–29	lft_2529	37 848	0,088	0	1
30–34	lft_3034	37 848	0,104	0	1
35–39	(ref.)	37 848	0,133	0	1
40–44	lft_4044	37 848	0,129	0	1
45–49	lft_4549	37 848	0,117	0	1
50–54	lft_5054	37 848	0,104	0	1
55–59	lft_5559	37 848	0,084	0	1
60–64	lft_6064	37 848	0,023	0	1
<i>Opleiding (opl5hb)</i>					
basisonderwijs	opl_bo	37 682	0,094	0	1
vmbo	opl_vmbo	37 682	0,273	0	1
mbo, havo, vwo	(ref.)	37 682	0,425	0	1
hbo	opl_hbo	37 682	0,131	0	1
wo	opl_wo	37 682	0,077	0	1
<i>Beroepsrichting (sector)</i>					
economisch	(ref.)	37 848	0,187	0	1
techniek	ber_tech	37 848	0,205	0	1
verzorgend/sociaal-cultureel	ber_zorg	37 848	0,261	0	1
overige	ber_overig	37 848	0,347	0	1
<i>Herkomstgroepering (autoch)</i>					
autochtoon	(ref.)	37 848	0,786	0	1
westers allochtoon	walloch	37 848	0,094	0	1
niet-westers allochtoon	nwalloch	37 848	0,120	0	1
<i>Gezinssituatie</i>					
alleen	(ref.)	37 848	0,222	0	1
paar	paar	37 848	0,571	0	1
ander lid	anderlid	37 848	0,207	0	1
ln(# personen)	Inaantalpp	37 848	0,974	0	2,565
ln(1+ # kinderen 0–5)	ln1nlf05	37 848	0,143	0	1,609
laag	popl_laag	37 848	0,175	0	1
midden	(ref.)	37 848	0,675	0	1
hoog	popl_hoog	37 848	0,150	0	1
<i>Woonregio (rbawo)</i>					
Groningen	groningen	37 848	0,052	0	1
Friesland	friesland	37 848	0,048	0	1
Drenthe	drenthe	37 848	0,043	0	1
IJssel-Vecht/Twente	ijsselvech--e	37 848	0,070	0	1
IJssel-Veluwe	ijsselveluwe	37 848	0,041	0	1
Arnhem/Oost-Gelderland/	arnhemoost--d	37 848	0,077	0	1
<i>Nijmegen/Rivierenland</i>					
Flevoland	flevoland	37 848	0,030	0	1
Midden-Nederland	(ref.)	37 848	0,079	0	1
Noord-Holland Noord	noordholland	37 848	0,032	0	1
Zuidelijk Noord-Holland	Rijnmond	37 848	0,069	0	1
Rijnstreek	zuidelijkned	37 848	0,036	0	1
Haaglanden	rijnstreek	37 848	0,048	0	1
Rijnmond	haaglanden	37 848	0,103	0	1
Zeeland	zeeland	37 848	0,022	0	1
Midden en West Brabant	middenenwe~t	37 848	0,064	0	1
Noordoost Brabant	noordoostb~t	37 848	0,036	0	1
Zuidoost Brabant	zuidoostbr~t	37 848	0,050	0	1
Limburg	limburg	37 848	0,100	0	1
<i>Soort werkloze (arbgroep)</i>					
voorheen werkzaam	(ref.)	37 395	0,476	0	1
herintreder	vh_np	37 395	0,392	0	1
voorheen scholier	vh_school	37 395	0,087	0	1
overig	vh_overig	37 395	0,045	0	1
<i>Zoeken naar werk</i>					
In afgelopen 4 weken gezocht	gezocht4wkn	37 848	0,869	0	1
Zoekactiviteiten factor 12	zoekact1	37 848	0,000	-2,64300	1,53200
Zoekactiviteiten factor 2	zoekact2	37 848	0,000	-1,36100	4,19800

¹⁾ Waarnemingen zijn hier tijdsintervallen tussen twee opeenvolgende EBB-metingen, waarin de respondent bij aanvang werkloos is en waarvan de situatie aan het eind bekend is.

²⁾ Factoren op basis van factoranalyse van zoekactiviteiten in afgelopen 4 weken.

In de analyses is voor het zoekgedrag van een werkloze een tweetal factoren opgenomen. Deze factoren zijn met behulp van een principale componenten (factor-)analyse geconstrueerd op basis van 9 vragen over recente zoekinspanningen. In elk van deze vragen wordt een activiteit genoemd met de vraag of de persoon die in de voorafgaande vier weken wel of niet heeft ondernomen. De factoranalyse levert twee factoren op met eigenwaarde groter dan 1. Deze twee factoren verklaren vangen ongeveer 47 procent van de totale variantie in de negen vragen. Tabel B2.2 geeft de factorladingen.

Tabel B2.2
Factorladingen voor zoekactiviteiten (na varimax rotatie)

	Factoren	
	zoekact1	zoekact2
Heeft in de afgelopen 4 weken		
personeelsadvertenties nagekeken, niet via internet	0,808	-0,055
gereageerd op personeelsadvertenties, niet via internet	0,756	0,102
sollicitatiegesprek gevoerd of assessment gedaan	0,053	0,714
geinformeerd bij werkgevers	0,457	0,442
geinformeerd bij familie, vrienden of relaties	0,609	0,135
geinformeerd bij CWI	0,652	0,213
geinformeerd of laten inschrijven bij uitzendbureau	0,518	0,272
open sollicitatie geschreven	0,525	0,336
bedrijfsdag bezocht	0,032	0,668

Er tekenen zich duidelijk 2 inhoudelijk verschillende factoren af, waarvan de eerste vooral betrekking heeft op het inwinnen van informatie en de tweede op het contact hebben met een mogelijke nieuwe werkgever. Informatie inwinnen bij de werkgever zit daarom tussen de beide factoren in. De overige acht activiteiten worden duidelijk geassocieerd met één van de twee factoren. In de EBB wordt ook gevraagd of in de afgelopen 4 jaar via internet naar personeeladvertenties is gekeken en of in de afgelopen 4 weken is gereageerd op een personeeladvertentie via internet. Deze vragen zijn in de periode 2004–2006 niet gesteld. Wordt voor de overige jaren een factoranalyse gedaan waarin deze 2 variabelen aan de andere negen zijn toegevoegd, dan resulteren opnieuw twee factoren, met pakweg dezelfde ladingen als in Tabel B2.2 en waarbij de twee extra vragen keurig bij de eerste factor worden ondergebracht. Om die reden is besloten deze twee vragen bij de analyses voor 2002–2007 buiten beschouwing te laten.

Bijlage 3: Suggesties voor implementatie

Bij het schatten van de duurmodellen en het afleiden van statistieken is een aantal keuzes gemaakt waarvan bij praktische toepassing kan worden afgeweken of waarvan de invloed nog kan/moet worden nagegaan. In deze bijlage wordt kort ingegaan op enkele van deze keuzes.

Zoals aangegeven bij de bespreking van de data en de schattingsmethode, is gebruik gemaakt van informatie over werkloosheidsduren van personen die in de eerste EBB-peiling werkloos zijn en van personen die tussen twee peilingen als nieuwe werklozen zijn ingestroomd. In de *delayed entry* en *interval censored* schattingsmethode wordt primair gebruik gemaakt van de informatie of een persoon die in een bepaalde peiling werkloos is dat bij de volgende peildatum nog steeds is of dat hij of zij in de tussentijd een baan heeft gevonden of gestopt is met zoeken naar werk. Bij deze analyses is de werkloosheidsduur op de eerste peildatum dus niet de afhankelijke (te verklaren) variabele, maar speelt deze wel een rol als exogene variabele doordat de *likelihood*-functie een voorwaardelijke kans is conditioneel op deze duur. De duur die iemand op de eerste peildatum werkloos is, wordt gedefinieerd als het minimum van de duur die iemand al zonder werk zit en de duur die iemand al op zoek is naar werk. Als iemand in de daaropvolgende peiling nog steeds werkloos is, wordt de verstreken werkloosheidsduur op dat peilmoment berekend door de tijd tussen de twee peildata bij de eerdere duur op te tellen. Op dezelfde manier worden werkloosheidsduren die nog niet zijn afgesloten, doorgetrokken naar de daaropvolgende peildata. Als iemand tussen twee peildata werkloos wordt, wordt de werkloosheidsduur op de tweede van deze peildata gedefinieerd als de helft van de tijd tussen de twee peildata. Het kan interessant zijn om de gevoeligheid van de statistieken voor deze 'schatting' te onderzoeken, al mag veilig worden aangenomen dat de resultaten nauwelijks door deze veronderstelling worden beïnvloed.

Belangrijker is dat de informatie van de werkloosheidsduur die voor de eerste peiling al is verstreken inhoudelijk anders is dan de in de tijd verder doorgetrokken duur op basis van de vervolgpeilingen. De eerste duur is gebaseerd op retrospectieve informatie die vooral bij langere duren beïnvloed kan zijn door herinneringsfouten en afrondfouten. Doordat de reeds verstreken werkloosheidsduur als onafhankelijke variabele slechts een beperkte rol speelt in de schattingen zullen dergelijke fouten naar verwachting weinig invloed hebben op de schattingen en de statistieken, al kan het interessant zijn om na te gaan wat het verschil is als ontpiekte of ontdiscretiseerde duurgegevens worden gebruikt. Een belangrijker verschil tussen de retrospectieve gegevens uit de eerste peiling en de informatie uit de vervolgpeilingen kan zijn dat op grond van de laatste informatie tijdelijke onderbrekingen vaker als transitie vanuit werkloosheid worden aangemerkt – en dus als einde van de werkloosheidsduur – dan in de retrospectieve informatie. Zo'n onderbreking kan betrekking hebben op een korte tijdelijke dienstbetrekking of een korte periode waarin een persoon niet zoekt of niet per direct beschikbaar is, bijvoorbeeld als gevolg van vakantie of ziekte. Bij retrospectieve bevraging is denkbaar dat dergelijke niet-substantiële onderbrekingen niet altijd als onderbreking van de zoekduur worden aangemerkt. Het zou daarom de moeite waard zijn om te onderzoeken hoeveel de statistieken veranderen als korte onderbrekingen van werkloosheid buiten beschouwing worden gelaten, bijvoorbeeld door bij de definitie van werkloosheid het beschikbaarheidscriterium en/of het zoekcriterium minder stringent toe te passen.

Een aantal andere praktische overwegingen is van belang bij het bepalen van statistieken van verwachte werkloosheidsduren langs de in dit rapport gevolgde lijn. Mogelijk verdienen deze nadere bestudering of uitwerking. Allereerst is bij het bepalen van de statistieken voor de eenvoud uitgegaan van de modelspecificaties en -schattingen met Weibull duurafhankelijkheid en zijn de tijdvariërende variabelen bij de berekening van de statistieken constant gehouden. Dit is rekentechnisch eenvoudiger, maar het is mogelijk de gebruikte programmatuur aan te passen zodat ook de stapsgewijze duurafhankelijkheid kan worden gebruikt – de specificaties die de voorkeur verdienen boven de Weibull specificatie – en eventueel kan bij het bepalen van de verwachte duren ook het maandpatroon en de leeftijd van de respondent variabel worden gemaakt. Dat laatste zal de uit-

komsten naar verwachting nauwelijks beïnvloeden, maar het meenemen van de geschatte maandeffecten als tijdvariërende variabele bij het bepalen van de statistieken kan een scherper beeld opleveren van het seizoenspatroon van werkloosheidsduren. Door voor de duurzaamheid de specificatie op basis van een stapfunctie te gebruiken, zullen er waarschijnlijk verhoudingsgewijs minder van de langste duren voorkomen, aangezien de geschatte stapfuncties in hoofdstuk 4 niet monotoon dalend zijn en na twee jaar weer een lichte toename van de uitstroomintensiteit laten zien.

Zoals de resultaten in hoofdstuk 5 duidelijk laten zien, zijn de berekende statistieken niet gevoelig voor kleine variaties in de opgenomen variabelen of voor het gebruik van een *competing risks* model in plaats van een gewoon duurmodel. Dat laat onverlet dat het van belang is alert te zijn op mogelijke instabiliteit van het model, bijvoorbeeld doordat de jaardummy's onvoldoende ruimte bieden om structurele veranderingen op de arbeidsmarkt te accommoderen, bijvoorbeeld als daardoor de geschatte effecten van een achtergrondvariabele als leeftijd in de loop van de tijd anders worden. Om dit te onderzoeken kan worden nagegaan of de resultaten sterk veranderen als de schattingen voor afzonderlijke jaren (of kwartalen) worden bepaald. Ook kunnen nadere gevoeligheidsanalyses worden gedaan door de schattingen te bepalen voor bepaalde deelsteekproeven, of door alleen de informatie op basis van de eerste twee peilingen te gebruiken, of bijvoorbeeld alleen de informatie waarvoor geen proxy-beantwoording is gebruikt.²⁰⁾

Ook bij de procesmatige aanpak van de statistieken spelen vergelijkbare mogelijkheden een rol. Als de schattingen op gegevens voor een aantal jaren worden bepaald – bijvoorbeeld als blijkt dat het model door de tijd voldoende stabiel is en de schattingen nauwkeuriger zijn naarmate zij op meer waarnemingen zijn gebaseerd – is het toch verstandig om met nieuwe gegevens het model opnieuw te schatten. Dat kan door de nieuwe informatie toe te voegen, al dan niet met het weglaten van oudere gegevens. Met het beschikbaar komen van de gegevens voor 2008 kan het model opnieuw geschat worden voor de periode 2002–2008, maar ook voor 2003–2008. Bij de laatste aanpak wordt dus het aantal jaren waarvan de informatie wordt gebruikt constant gehouden, maar zorgt het verschuiven van het tijdraam ervoor dat de meest recente informatie wordt gebruikt. Ook kan bij het bepalen van statistieken die zijn uitgesplitst naar deelpopulaties gebruik worden gemaakt van afzonderlijke schattingen voor die deelgroepen – bijvoorbeeld apart voor mannen en voor vrouwen – of door meer interacties van deze indelingen met andere variabelen in het model op te nemen.

²⁰⁾ Dit kan ook worden onderzocht door in het model dummy-variabelen op te nemen die aangeven of de betreffende case op proxy-beantwoording is gebaseerd of indicatoren voor de gebruikte peilingen.