

Het Sociaal Statistisch Bestand 1998 en 1999

C.H. Arts en E.M.J. Hoogteijling¹⁾

Het Sociaal Statistisch Bestand (SSB) bevat gegevens over alle personen, banen en uitkeringen in Nederland die op micro-niveau aan elkaar relateerbaar zijn. Het SSB biedt een aantal nieuwe mogelijkheden voor statistische output, zoals samenloop van banen en uitkeringen, pendel, diep-regionale uitsplitsingen, beschrijvingen van groepen allochtonen en stroominformatie.

1. Inleiding

Eind jaren negentig is het CBS begonnen met het ontwikkelen van het Sociaal Statistisch Bestand (SSB). Hiervoor worden gegevens over personen uit registraties en enquêtes op micro-niveau gekoppeld en waar nodig consistent gemaakt. Een belangrijke doelstelling van het SSB is om een gegevensbestand samen te stellen dat een samenhangende, consistente beschrijving geeft van een aantal aspecten van de bevolking van Nederland. Voor het verslagjaar 1995 is een prototype ontwikkeld, waarbij de nadruk lag op het koppelen van de verschillende bronnen (Arts en Van Lith, 1999; Arts, Bakker en Van Lith, 2000).

Op dit moment zijn voor de verslagjaren 1998 en 1999 Sociaal Statistische Bestanden beschikbaar. Deze bestanden bieden nieuwe mogelijkheden voor statistische output in aanvulling op het bestaande pakket. In deze aflevering van de Sociaal-Economische Maandstatistiek worden enkele mogelijkheden van het gebruik van het SSB gepresenteerd. De nadruk ligt daarbij op de statistische uitkomsten.

Dit artikel gaat in op de gebruikte bronnen en de wijze van samenstellen van de verschillende deelbestanden van het SSB. In paragraaf 2 worden de gebruikte bronnen kort toegelicht en worden enkele kerncijfers op basis van het SSB gepresenteerd. Ook komt kort de inhoud van de hiernavolgende artikelen aan de orde. In paragraaf 3 wordt ingegaan op de wijze van samenstellen van de integrale deelbestanden van het SSB en het toevoegen van enquêtegegevens. In paragraaf 4 komt de methode van herhaald wegen aan bod, die wordt toegepast bij het gebruik van gegevens uit enquêtes.

2. Bronnen en kerncijfers van het SSB

2.1 Bronnen

Staat 1 geeft een overzicht van de gegevensbronnen die voor het SSB worden gebruikt. Vrijwel alle bronbestanden van het SSB zijn zogenaamde volumebestanden, dat wil zeggen dat gegevens over het gehele jaar zijn opgenomen. Het bestand van de enquête werkgelegenheid en lonen (EWL) vormt hierop een uitzondering: hierin zijn alleen gegevens opgenomen over banen van werknemers die eind december bestonden. In totaal zijn voor het SSB van 1999 ongeveer 65 miljoen records met gegevens uit registraties gebruikt.

Staat 1

Gegevensbronnen van het SSB naar aantal records, herkomst, eenheden en verslagjaar

Bron	Aantal records		Afkomstig van	Belangrijke variabelen
	1998	1999		
	<i>x 1 000</i>			
Gemeentelijke basisadministratie vanaf 1995 (GBA)	25 599	28 716	Gemeenten	geslacht, geboortedatum, adres, burgerlijke staat, nationaliteit, geboorteland, gezinsrelaties
Voorheffing loonbelasting (Fibase)	18 548	18 793	Belastingdienst	soort inkomen, fiscaal loon, aanvang en einde, loonbelastingnummer bedrijf/instell.
Verzekerenadministratie werknemers (VZA)	10 432	10 795	Uitvoeringsinstellingen werknemersverzekeringen	blsv, sociale verzekeringsdagen, soort dienstverband, aanvang en einde, aansluitingsnummer bedrijf/instell.
Enquête werkgelegenheid en lonen (EWL)	2 770	2 990	CBS, enquête onder bedrijven	soort dienstverband, looncomponenten, arbeidsduur, aanvang en einde, economische activiteit
Administratie arbeidsongeschiktheidswet (AO)	1 015	1 037	Uitvoeringsinstellingen werknemersverzekeringen	soort uitkering, mate en aard van arbeidsongeschiktheid, aanvang en einde, aanvang en einde, reden van beëindiging, WSW-indicatie, bedragen
Definitieve aanslaggegevens fiscale ondernemingswinst natuurlijke personen (Winst)	787	767	Belastingdienst	winstbedrag
Administratie werkloosheidswet (WW)	740	573	Uitvoeringsinstellingen werknemersverzekeringen	aanvang en einde, reden beëindiging, bedragen
Administratie studiefinanciering (WSF)	654	654	Informatie Beheer Groep	soort studiefinanciering, bedragen
Administratie algemene bijstandswet (ABW)	606	540	Gemeenten	soort uitkering, leefvorm, aanvang en einde, bedragen
Administratie wet inkomensvoorziening oudere en gedeeltelijk arbeidsongeschikte werknemers (IOAW)/arbeidsongeschikte gewezen zelfstandigen (IOAZ)	31	32	Gemeenten	soort uitkering, leefvorm, aanvang en einde, bedragen

Bron: CBS, Sociaal Statistisch Bestand.

De basis van het SSB is een bestand met persoonsgegevens, waarin jaarbestanden van de *Gemeentelijke Basisadministratie* (GBA) vanaf 1995 zijn opgenomen (gecumuleerd GBA). Het bestand bevat gegevens zoals geslacht, geboortedatum, geboorteland, burgerlijke staat, woonplaats en herkomstgroepering van alle personen die vanaf 1995 in Nederland woonachtig zijn of zijn geweest. In het bestand zijn (demografische) wijzigingen opgenomen. Wanneer bijvoorbeeld iemand is verhuisd, dan komen gegevens van deze persoon twee keer in het bestand voor: één keer met het oude adres en één keer met het nieuwe adres, met de bijbehorende aanvangs- en beëindigingsdatum van de betreffende situatie. Hiermee is het mogelijk voor ieder persoon op elk moment in de tijd de demografische kenmerken te bepalen.

Het *Fibase*-bestand van de Belastingdienst bevat gegevens over dienstverbanden en uitkeringen, die vallen onder de voorheffing loonbelasting. De gegevens zijn afkomstig van bedrijven en instellingen. Naast gegevens over uitkeringen van werknemersverzekeringen, zoals in het kader van de werkloosheidswet (WW) en arbeidsongeschiktheidswet (AO), zijn gegevens over uitkeringen van volksverzekeringen opgenomen. Voorbeelden hiervan zijn uitkeringen van de algemene ouderdomswet (AOW) en algemene nabestaandenwet (ANW). Ook bevat het bestand gegevens over pensioenuitkeringen en uitkeringen van particuliere verzekeringen, zoals lijfrenten. Het bestand bevat gegevens over aanvang en beëindiging van dienstbetrekkingen en uitkeringen, fiscaal loon, loonbelasting en het bedrijf of instelling waarvan de gegevens afkomstig zijn.

In de *verzekerdensadministratie werknemers* (VZA) zijn gegevens over jaaropgaven van dienstverbanden van werknemers opgenomen, wanneer deze vallen onder de werknemersverzekeringen. Het gaat onder andere om gegevens over het bruto jaarloon sociale verzekeringen (BLSV), het aantal sociale verzekeringsdagen, betaalde werknemerspremies, type werknemer (flexibel en niet-flexibel) en het bedrijf of instelling waarvan de gegevens afkomstig zijn.

De *enquête werkgelegenheid en lonen* (EWL) wordt door het CBS uitgevoerd en bevat gegevens over loon en arbeidsduur van banen van werknemers, zoals waargenomen in bedrijfsadministraties. Voorbeelden hiervan zijn pensioenpremies, spaarloon, bijzondere beloningen, wekelijkse arbeidsduur en vakantie-, verlof- en ziekte-dagen. De enquête is een combinatie van integrale waarneming en aanvullend steekproefonderzoek (De Vries, 1998; CBS, 2000a).

De *uitkeringsadministraties* van de uitvoeringsinstellingen werknemersverzekeringen bevatten op maandbasis gegevens over betalingen van arbeidsongeschiktheids- en werkloosheidsuitkeringen (AO en WW). Hierin zijn kenmerken opgenomen als reden van beëindiging, de mate en aard van arbeidsongeschiktheid.

In het bestand met definitieve aanslaggegevens *fiscale ondernemingswinst natuurlijke personen*, afkomstig van de Belastingdienst, is de hoogte van de winst opgenomen. Het bestand over het verslagjaar 1999 is niet geheel compleet, omdat nog niet alle definitieve aanslagen bekend zijn. Naar schatting ontbreken gegevens over enkele tienduizenden zelfstandigen. Het bestand met alle definitieve aanslaggegevens over 1999 is begin 2003 beschikbaar.

De bestanden van de gemeentelijke sociale diensten bevatten gegevens over *bijstandsuitkeringen* en uitkeringen in het kader van de *IOAW/IOAZ*. Het gaat hierbij om maandelijks betalingsgegevens en kenmerken van de uitkering en uitkeringsgerechtigde.

Alle bestanden zijn gekoppeld aan het gecumuleerde GBA-bestand op basis van het sofi-nummer. Voor records die geen sofi-nummer bevatten, is gekoppeld op de kenmerken geslacht, geboortedatum en adres. Na koppeling zijn direct identificerende kenmerken uit de bestanden verwijderd en is het sofi-nummer vervangen door een uniek, willekeurig nummer (Arts en Van Lith, 1999; Al en Altena, 2000).

Op basis van de gekoppelde bronnen is een groot aantal gegevens over personen, banen en uitkeringen integraal beschikbaar voor statistische doeleinden. Het is echter niet zo dat alle bronnen direct bruikbaar zijn voor de statistiek. Ten eerste zijn in administratieve bronnen soms gegevens opgenomen die niet relevant zijn voor statistische doeleinden of een andere definitie kennen dan in de statistiek. Ten tweede bestrijken niet alle bronnen de gehele populatie van eenheden, die onderwerp zijn van de statistiek. Het is daarom noodzakelijk de verschillende bronnen te integreren. De hiervoor gebruikte methoden worden beschreven in paragraaf 3.

2.2 Kerncijfers van het SSB 1999

Het SSB van 1999 bestaat uit verschillende deelbestanden, waarin gegevens over personen, banen en uitkeringen over het gehele jaar zijn opgenomen. Er is een bestand met persoonsgegevens, een bestand met gegevens over banen van werknemers, een zelfstandigenbestand en er zijn bestanden met gegevens over uitkeringen. Het personenbestand bevat vooral gegevens afkomstig uit de GBA. Voor het banenbestand werknemers zijn gegevens uit de *Fibase*, de *Verzekerdensadministratie werknemers* en de *EWL* geïntegreerd. Voor de uitkeringenbestanden is gebruik gemaakt van de verschillende uitkeringsadministraties en de *Fibase*. De deelbestanden van het SSB zijn op microniveau gekoppeld, zodat combinaties van gegevens uit verschillende bestanden mogelijk zijn.

In de tweede kolom van staat 2 staan de aantallen van het totale volume van de verschillende bestanden. In de derde kolom staat om hoeveel personen het gaat. In de volgende kolommen staan hoeveel personen daarvan in het buitenland woonachtig waren dan wel tot de bevolking van Nederland behoorden. Tot slot wordt de stand van eind oktober 1999 gegeven ²⁾.

Het SSB van 1999 bevat gegevens over in totaal 16,4 miljoen verschillende personen. Ten eerste gaat het daarbij om 16,1 miljoen personen die in 1999 enige tijd of het gehele jaar in Nederland woonden en als zodanig bij een gemeente ingeschreven zijn geweest. Ten tweede bevat het SSB gegevens over personen die niet in Nederland woonden, maar hier wel werkten of een uitkering hadden (bijna 300 duizend) ³⁾. Personen met pensioen/AOW en personen met een arbeidsongeschiktheidsuitkering wonen relatief vaak in het buitenland. Eind oktober 1999 bestond de bevolking van Nederland uit 15,9 miljoen personen.

Het banenbestand werknemers bevat gegevens over 10,7 miljoen banen van 7,7 miljoen personen die in 1999 als werknemer in Nederland werkzaam zijn geweest. Het gaat daarbij om arbeid geregistreerd in bedrijfsadministraties, in het kader van werknemersverzekeringen en de voorheffing loonbelasting. Het zelfstandigenbestand bevat gegevens over 767 duizend personen die in 1999 fiscaal gezien als zelfstandige hebben gewerkt. In totaal waren er in 1999 volgens het SSB 8,3 miljoen personen die minstens enige tijd in Nederland werkzaam zijn geweest; eind oktober waren van de bevolking van Nederland volgens het SSB 7,4 miljoen mensen werkzaam.

Het SSB bevat gegevens over 2,9 miljoen personen met een pensioenuitkering. Eind oktober waren er onder de bevolking van Nederland bijna 2,7 miljoen personen met pensioen. Daarnaast hadden 1,8 miljoen personen andere uitkeringen, zoals AO-, WW-, en Abw-uitkeringen.

In staat 3 staan diverse grootheden die betrekking hebben op het banenbestand van werknemers. Het *gemiddeld* aantal banen van werknemers in 1999 bedroeg 7,2 miljoen. Het totale *volume* van 10,7 miljoen banen omvat 6,3 miljoen *volledige jaren*, uitgedrukt in eenheden van 260 sociale verzekeringsdagen. Het totale jaarbedrag aan loon (BLSV) was in 1999 rond 151 miljard euro. Wanneer we dit totale bedrag delen door de aantallen verschillende grootheden die in staat 3 staan vermeld dan komen we tot de gemiddelde lonen per grootheid. Zo was het loon per baan in 1999

Staat 2
Kerncijfers van het SSB 1999

Eenheid	Totaal aantal	Aantal personen	w.v.		
			Woonachtig in het buitenland ¹⁾	Behorende tot de bevolking van Nederland	
				Volume	Stand eind oktober
<i>x 1 000</i>					
Personen	16 381	16 381	293	16 088	15 853
Banen	11 424	8 250	117	8 133	7 385
w.v.					
Banen van werknemers	10 657	7 685	114	7 571	6 784
Zelfstandigen	767	767	4	763	761
Pensioen/AOW ²⁾	4 808	2 935	151	2 783	2 651
Uitkeringen	2 926	2 428	41	2 388	1 772
w.v.					
Arbeidsongeschiktheidsuitkeringen	988	972	27	945	850
Werkloosheidsuitkeringen	537	450	2	448	169
Bijstandsuitkeringen	618	588	3	585	455
Uitkeringen IOAW/IOAZ	54	53	0	52	43
Overige uitkeringen ³⁾	728	701	10	691	363

¹⁾ Deze groep bevat ook gegevens van personen die ten onrechte niet met het GBA-bestand konden worden gekoppeld.

²⁾ Deze categorie bevat ook lijfrente-uitkeringen en uitkeringen in het kader van de algemene nabestaandenwet (ANW).

³⁾ Hieronder vallen alle niet eerder genoemde uitkeringen, waaronder uitkeringen in het kader van de ziektewet en de wachtgeldregeling overheid. Studiefinanciering is overigens niet in het overzicht opgenomen.

Bron: CBS, Sociaal Statistisch Bestand.

Staat 3
Kerncijfers van banen van werknemers uit het SSB 1999

	Totaal	w.v. behorende tot de bevolking van Nederland
<i>x 1 000</i>		
Aantal banen	10 657	10 508
Aantal personen	7 685	7 571
Gemiddeld aantal banen	7 186	7 106
Volledige jaren ¹⁾	6 264	6 192
<i>x 1 miljard euro</i>		
Som van het brutoloon sociale verzekeringen (BLSV)	151	149
<i>x 1 000 euro</i>		
BLSV per baan	14,1	14,2
BLSV per persoon	19,6	19,6
BLSV per gemiddelde baan	21,0	20,9
BLSV per volledig jaar	24,1	24,0

¹⁾ In eenheden van 260 sociale verzekeringsdagen.

Bron: CBS, Sociaal Statistisch Bestand.

ruim 14 duizend euro, terwijl dit per persoon bijna 20 duizend euro was. Per volledig jaar was het BLSV 24 duizend euro en per gemiddelde baan bijna 21 duizend euro.

2.3 Nieuwe statistische output

Door het onderlinge verband tussen gegevens over verschillende statistische thema's in het SSB, ontstaan nieuwe mogelijkheden voor statistische output. Zo is het mogelijk de relatie tussen uitkeringen en banen te laten zien. De integrale bestanden bieden tevens de mogelijkheid om veel details weer te geven, bijvoorbeeld op regionaal niveau of voor kleine bevolkingsgroepen. Bovendien hebben veel bestanden een volumekarakter, zodat longitudinale analyses tot de mogelijkheden behoren. Vooralsnog zijn deze analyses beperkt tot het volumebestand van één verslagjaar. De hierna volgende artikelen illustreren de mogelijkheden van het SSB.

In Das, Van Toor en Arts (2002) wordt een overzicht gepresenteerd van de verschillende inkomensbronnen van de bevolking van 15-64 jaar. Enkele groepen personen die naast een baan ook een uitkering hebben worden nader onder de loep genomen. Melser en Van Toor (2002) en Boerdam (2002) gaan in op de arbeidsmarktsituatie van allochtonen en autochtonen. Linder (2002) laat enkele mogelijkheden zien van de regionale component in het SSB.

Door het toevoegen van variabelen uit andere bronnen is het mogelijk nieuwe verbanden te leggen. Zo belichten Corpeleijn en Heerschop (2002) de pendel tussen regio's en steden, waarvoor de vestigingsplaats van het bedrijf aan het banenbestand werknemers van 1998 is toegevoegd. In Florquin (2002) is door koppeling van het studentenbestand hoger onderwijs informatie voorhanden over de arbeidsmarktpositie van deze groep.

Door in de volumebestanden van het SSB twee peilmomenten te kiezen (ultimo april en oktober) is het mogelijk op transitiebasis overgangen in de arbeidsmarktpositie van mensen te laten zien. Dit komt in Arts, Van Toor en De Vries (2002) aan de orde.

Door koppeling met enquêtes is het mogelijk statistische informatie, die (nog) niet in registraties voorkomt, toe te voegen. In Van Cruchten, Hartgers en Schulte-Nordholt (2002) wordt op basis van de Enquête Beroepsbevolking (EBB) bijvoorbeeld ingegaan op de relatie tussen loon en opleidingsniveau en Weidum en Linder (2002) belichten het zoekgedrag op de arbeidsmarkt van uitkeringsgerechtigden.

3. Opzet van het Sociaal Statistisch Bestand

3.1 Het betalingsconcept in het SSB

Bij het bepalen van aanvangs- en beëindigungsdatums van banen en uitkeringen is binnen het SSB gekozen voor het *betalingsconcept*. Een baan of uitkering telt in deze optiek mee als er daadwerkelijk betalingen hebben plaatsgevonden. Met name binnen de sociale zekerheidsstatistiek is het tot nog toe gebruikelijk om uitkeringen te presenteren op *registratiebasis* of op *transactiebasis*.

Een presentatie op registratiebasis komt neer op het weergeven van de uitkering (of baan) vanaf het moment van invoering in de administratie tot aan het moment van afvoeren. Bij het weergeven op transactiebasis wordt uitgegaan van uitkeringsrechten. Dat wil zeggen dat uitkeringen meetellen op het moment dat belanghebbende recht had op de uitkering. Dit betekent dat bij het samenstellen van transactiecijfers een correctie nodig is voor administratieve vertraging. Immers, de feitelijke toekenning van een recht op een uitkering kan met terugwerkende kracht plaatsvinden als gevolg van administratieve, bezwaar- of beroepsprocedures. Doordat het lang duurt voordat de meeste vertraagde gevallen zijn verwerkt, is een nadeel van deze wijze van presenteren dat definitieve cijfers pas laat beschikbaar zijn.

Bij zowel de presentatie op registratie- als op transactiebasis komen uitkeringen voor waarbij geen betalingen zijn gedaan. Deze zogenaamde 'nuluitkeringen' kunnen worden veroorzaakt door korting op de uitkering door inkomsten uit arbeid, sancties, schorsingen, terugvordering en verrekening. Daarnaast kan een recht op bijvoorbeeld een arbeidsongeschiktheidsuitkering bij werkherleving blijven bestaan. Wanneer blijkt dat terugkeer op de arbeidsmarkt niet mogelijk is, wordt het 'slapende' recht weer geëffectueerd. Ook bij banen, die op registratiebasis worden waargenomen, komen nulbetalingen voor. Het gaat daarbij bijvoorbeeld om personen die bij een werkgever staan geregistreerd, maar een bepaalde tijd geen werk hebben verricht. Daarnaast kan administratieve vervuiling de oorzaak zijn van nulbetalingen in de bestanden.

Bij het weergeven van een uitkering of baan op betalingsbasis wordt uitgegaan van de betalingen die volgens de registratie zijn gedaan. Er wordt een onderscheid gemaakt tussen reguliere betalingen en nabetalingen. Reguliere betalingen zijn betalingen die plaats vinden vóór het moment dat het recht op een uitkering verloopt of het dienstverband is beëindigd. Nabetalingen zijn betalingen die hierna plaats vinden. Uitkeringen- of banenrecords in de registraties zonder betalingen behoren volgens het betalingsconcept niet tot de populatie uitkeringen of banen.

Het SSB biedt de mogelijkheid cijfers over *samenloop* van banen en uitkeringen samen te stellen. Het bepalen van samenloop op registratie- of transactiebasis geeft een overschatting van het werkelijk samengaan van een baan met een uitkering. Immers, bij een 'slapend' recht of bij volledige korting van de uitkering is feitelijk geen sprake van een uitkering. Ook is bij een dienstverband, waarbij geen betalingen zijn gedaan, geen sprake van een baan. Het betalingsconcept sluit op deze perceptie aan. Alleen wanneer iemand loon ontvangt en daarnaast nog een uitkeringsbedrag, wordt dit beschouwd als samenloop.

3.2 Banenbestanden

3.2.1 Werknemers en zelfstandigen

Het SSB bevat twee bestanden met gegevens over banen: het banenbestand *werknemers* en het banenbestand *zelfstandigen*. Het laatste bestand is gebaseerd op winstgegevens van de Belastingdienst (zie paragraaf 2.1). Er is verondersteld dat iedere persoon die hierin voorkomt in 1999 als zelfstandige heeft gewerkt en als zodanig één baan als zelfstandige heeft gehad⁴⁾. Aangezien er geen gegevens over aanvang en beëindiging in het bestand zijn opgenomen, is bovendien verondersteld dat de zelfstandige het gehele jaar actief is geweest. Dit leidt tot een overschatting van

het aantal zelfstandigen op één peilmoment. Het banenbestand zelfstandigen voor 1999 is niet geheel compleet, omdat pas begin 2003 alle definitieve aanslagen van 1999 bekend zijn. Er zijn op dit moment nog geen gegevens beschikbaar over de soort economische activiteit waarin de zelfstandigen actief zijn.

3.2.2 Toekennen van bedrijfseenheden en verbinden van bronnen

Bij het samenstellen van het banenbestand werknemers is gebruik gemaakt van vier bronnen:

1. de verzekerdenadministratie werknemers (VZA),
2. de enquête naar werkgelegenheid en lonen (EWL),
3. bestand met voorheffingen loonbelasting (Fibase),
4. de gekoppelde gemeentelijke basis administraties van 1995–1999 (GBA).

De eerste drie bronnen zijn gebruikt voor het vaststellen van de banen van werknemers, de GBA is gebruikt als hulpmiddel bij het afbakenen van perioden en het bepalen van grensarbeid.

Een *baan* van een werknemer is gedefinieerd als een relatie tussen een persoon en een bedrijf of instelling in statistische zin, de zogenaamde *bedrijfseenheid*, zoals geoperationaliseerd in het Algemeen Bedrijfsregister (ABR) van het CBS (CBS, 2000b). We spreken in dit artikel over een *dienstverband* als het gaat om de relatie tussen een persoon en een bedrijf of instelling zoals vastgelegd in de verschillende registraties. We hebben het dan in feite over de records in de registratiebestanden.

Een aantal bewerkingen heeft betrekking op het bepalen van de statistische bedrijfseenheid van een baan en het toekennen van het bijbehorende bedrijfseenheididentificatienummer. Hiervoor is het noodzakelijk dat administratieve eenheden van bedrijven zoals opgenomen in de Fibase (loonbelastingnummer) en VZA (bedrijfsvereniging aansluitingsnummer) worden vertaald naar de statistische eenheid. Met het toekennen van de statistische bedrijfseenheid is tevens informatie beschikbaar over de economische activiteit waarin een persoon werkzaam is.

De relatie tussen administratieve eenheden van bedrijf en de statistische eenheid is lang niet altijd één op één. Zowel administratieve eenheden die bestaan uit meerdere statistische eenheden (1-op-n-relatie) als relaties waarbij m administratieve eenheden behoren bij n statistische eenheden (m-op-n-relatie) vormen een probleem. Bij beide typen is het namelijk noodzakelijk de banen behorende bij een administratieve eenheid van bedrijf te verdelen over de bijbehorende statistische bedrijfseenheden. Dit doet zich voornamelijk voor bij grotere bedrijven en heeft dus betrekking op relatief grote aantallen banen. Bij situaties, waarbij één statistische eenheid bestaat uit meerdere administratieve eenheden is het soms noodzakelijk verschillende dienstverbanden van dezelfde persoon bij verschillende administratieve eenheden van bedrijf samen te voegen tot één baan. De records in het EWL-bestand zijn al voorzien van statistische bedrijfseenheden. De daar gebruikte verdeelsleutels bij 1-op-n en m-op-n-relaties worden gebruikt voor verdeling van banen die niet tot de ultimostand behoren. Banenrecords van bedrijven die niet in de EWL worden waargenomen, worden via zogenaamde relatiebestanden voorzien van een statistische bedrijfseenheid. Deze relatiebestanden bevatten gegevens over de relaties tussen administratieve en statistische eenheden van bedrijf en zijn gebaseerd op het ABR.

Tegelijkertijd moeten records uit verschillende bronnen correct met elkaar worden verbonden. Zo kunnen personen in de loop van het jaar meerdere banen na elkaar hebben gehad, maar ook kunnen zij tegelijkertijd meerdere banen hebben. Gegevens over al deze banen zijn in de bestanden opgenomen. Vaak zijn gegevens over hetzelfde dienstverband in meerdere bronnen opgenomen, maar het kan ook voorkomen dat de ene baan alleen voorkomt in de ene bron, terwijl de andere baan slechts in een andere bron voorkomt. Het correct verbinden van de records is essentieel voor een juiste schatting van het aantal banen: het ten onrechte koppelen van records uit verschillende bronnen ('miskoppeling') leidt tot een onderschatting van het aantal banen, terwijl het ten onrechte niet koppelen ('gemiste koppeling') leidt tot

een overschatting. In het eerste geval ontstaat één banenrecord terwijl sprake is twee banen; in het tweede geval ontstaan twee banenrecords, terwijl sprake is van één en dezelfde baan. Voor het bepalen van het aantal als werknemer werkzame *personen* heeft dit overigens geen effect.

Om banen uit verschillende bronnen aan elkaar te koppelen wordt gebruik gemaakt van een uniek nummer van de persoon, zoals opgenomen in de bestanden. Daarnaast wordt gebruik gemaakt van relatietabellen tussen administratieve en statistische bedrijfs-eenheden en de hoogte van het (jaar)loon van de verschillende bronnen.

3.2.3 Selectie van banen en toekennen van datums

Wijziging van aanvangs- en beëindigingsdatums en het selecteren van bepaalde banenrecords hebben tot doel beter aan te sluiten op het betalingsconcept (zie paragraaf 3.1). Er wordt daarbij aangenomen dat het betaalde loon in de vastgestelde periode is verdiend.

In administraties komen records voor die niet onder de definitie 'baan' vallen. Records uit de Fibase worden als werknemersbanen beschouwd als het vermelde fiscale loon groter is dan nul, de loonbelastingtabel aangeeft dat het om inkomen uit tegenwoordige arbeid gaat en het soort inkomen aangeeft dat het geen uitkering is. Voor het VZA-bestand gelden soortgelijke afbakening: het BLSV moet groter zijn dan nul, tenzij dezelfde baan voorkomt in het loonbelastingbestand met een fiscaal loon groter dan nul. Als een record van een persoon in het VZA-bestand uitsluitend als uitkeringsrecord voorkomt in het fiscale bestand, waarvan het fiscale loon overeenstemt met het BLSV uit het VZA-bestand, dan komt het VZA-record als baan te vervallen.

Veel dienstverbandrecords hebben in de Fibase en de VZA als aanvangsdatum 1 januari 1999 en als beëindigingsdatum 31 december 1999. De reden hiervoor is dat voor veel dienstverbandenrecords 1 januari 1999 en 31 december 1999 een soort standaarddatums zijn. Voor de *anciënniteit* is nagegaan of de persoon met dezelfde baan al eerder voorkwam in banenbestanden van 1995–1998, die gebaseerd zijn op VZA- en EWL-bestanden van die jaren. Wanneer dat het geval was, is de vroegste aanvangsdatum daaruit overgenomen, rekening houdend met de leeftijd van de persoon. Wanneer verschillende bronnen verschillende aanvangs- en beëindigingsdatums aangeven, dan worden keuzen gemaakt. Wanneer een baan in één bron een aanvangsdatum van 1 januari 1999 heeft en een andere bron een andere datum geeft, dan wordt de datum uit de laatste bron overgenomen. Dit geldt ook voor banen met een beëindigingsdatum van 31 december 1999. Voor banen die mede gebaseerd zijn op informatie uit de VZA en de EWL geldt bovendien als voorwaarde dat het aantal sociale verzekeringsdagen past binnen een eventueel nieuw vastgestelde kalenderperiode, dat is de periode tussen aanvang en einde van de baan.

Wanneer een persoon met een dienstverband in 1999 volgens het GBA-bestand is gestorven, dan geldt de sterftedatum als beëindigingsdatum van de baan. Wanneer de sterftedatum vóór 1999 ligt, dan is geen sprake van een baan en wordt het dienstverband niet in het banenbestand opgenomen. Het gaat hierbij om enkele honderden gevallen.

Bij veel kleine banen is het niet waarschijnlijk dat de vermelde periode betrekking heeft op de periode waarin feitelijk is gewerkt. Het gaat daarbij bijvoorbeeld om uitzendkrachten die het hele jaar ingeschreven zijn geweest bij een uitzendbureau, maar slechts een korte periode hebben gewerkt, gezien het aantal sociale verzekeringsdagen. Het meenemen van deze kleine banen op basis van de aanvangs- en beëindigingsdatum zou tot een overschatting van het aantal banen leiden. Daarom zijn bij een aantal van deze kleine banen nieuwe datums geïmputeerd. Er is gekozen voor banen met minder dan 42 sociale verzekeringsdagen omdat vanaf 42 sociale verzekeringsdagen sprake kan zijn van een deel-

tijdbaan van één dag per week. Hierbij is het dus mogelijk dat het gehele jaar is gewerkt. Bij de imputatie van de datums wordt rekening gehouden met het seizoenspatroon van banen in dezelfde economische activiteit. Als laatste vinden correcties op datums van banen plaats op basis van de confrontatie met uitkeringen (zie paragraaf 3.4).

3.2.4 Loongegevens en arbeidsduur

Uitgangspunt bij de integratie van banen is dat voor alle banen een aantal basisgegevens beschikbaar is, terwijl andere gegevens alleen beschikbaar zijn voor records die in de EWL zijn opgenomen. Bij dit laatste gaat het bijvoorbeeld om gegevens over diverse looncomponenten en de arbeidsduur. Het BLSV is in zowel de VZA als de EWL beschikbaar. In de confrontatie van de VZA met de EWL is voor het vaststellen van het BLSV uitgegaan van de VZA. De EWL neemt het BLSV op maand-, kwartaal of jaarbasis waar. Hierdoor is het mogelijk dat bij bepaalde bedrijven de waarneming van het BLSV van werknemers onvolledig is. Het totale BLSV moet in dergelijke gevallen bij de EWL dus worden bijgeschat. Bij de VZA gaat het om jaaropgaven en is in principe volledigheid van het BLSV over het jaar gegarandeerd. Wanneer het BLSV tussen VZA en EWL afwijkt, worden de verschillende looncomponenten van de EWL dan ook aangepast aan het bedrag van de VZA. Wanneer de verschillen onverklaarbaar groot zijn, worden de EWL-gegevens voor de betreffende baan niet meege-nomen.

3.2.5 Opbouw van het banenbestand uit de bronnen

De bewerkingen die hierboven staan beschreven leiden tot grote wijzigingen ten aanzien van de brongegevens (staat 4). Van het VZA werknemersbestand zijn bijna 1,3 miljoen records verwijderd. Voor een deel gaat het daarbij om records die niet aan de gestelde definitie voldeden. Een ander deel is door samenvoeging met andere VZA-records verdwenen. Records worden samengevoegd wanneer zij betrekking hebben op dezelfde persoon en sprake is van dezelfde statistische bedrijfseenheid. Meestal gaat het daarbij om records met verschillende administratieve bedrijfseenheden die tot dezelfde statistische bedrijfseenheid blijken te horen (zie ook paragraaf 3.2.2).

Een groot deel van de verwijderde records uit de Fibase zijn uitkeringenrecords, maar ook worden rond 750 duizend dienstverbandrecords door samenvoeging verwijderd. Rond 70 duizend Fibase-records hebben een fiscaal loon van nul en zijn verwijderd. Een klein deel van de EWL-records is verwijderd als gevolg van te grote verschillen met het BLSV van de VZA en de confrontatie van beëindigingsdatums met de andere bronnen.

Van de records uit de Fibase die in het banenbestand zijn opgenomen is van 70% de aanvangsdatum gewijzigd. Voor een groot gedeelte gaat het daarbij om banen die al bestonden voor 1999. Deze waren oorspronkelijk in de Fibase vastgelegd met een aanvangsdatum van 1 januari 1999. Bijna een kwart van de beëindigingsdatums in de Fibase is gewijzigd, door confrontatie met andere bronnen.

Het fiscaal loon van bijna 3% van de Fibase-records is gewijzigd als gevolg van het samenvoegen met andere Fibase-records. Ook voor de BLSV-wijzigingen van VZA-records is de voornaamste oorzaak het samenvoegen van meerdere records. Bij de EWL wordt het BLSV bij bijna zes van de tien records gewijzigd. Bij een deel gaat het daarbij om afrondingsverschillen die in de EWL zijn ontstaan door optelling van bedragen van meerdere betalingsperiodes tot een jaarloon. Bij een ander deel gaat het om het ontbreken in de EWL van één of meerdere betalingsperiodes. Omdat het bruto jaarloon mede is gebaseerd op het BLSV, wijzigt dit ook vaak.

In schema 1 wordt de samenstelling van het banenbestand werknemers van 1999 gepresenteerd. Het grootste gedeelte bestaat uit banenrecords die zowel in de VZA als Fibase voorkomen, al dan niet in combinatie met gegevens uit de EWL. Het loonbelastingbestand vormt een behoorlijke aanvulling op VZA en EWL (9,5

Staat 4
Enkele wijzigingen op bronbestanden SSB banen van werknemers, 1999

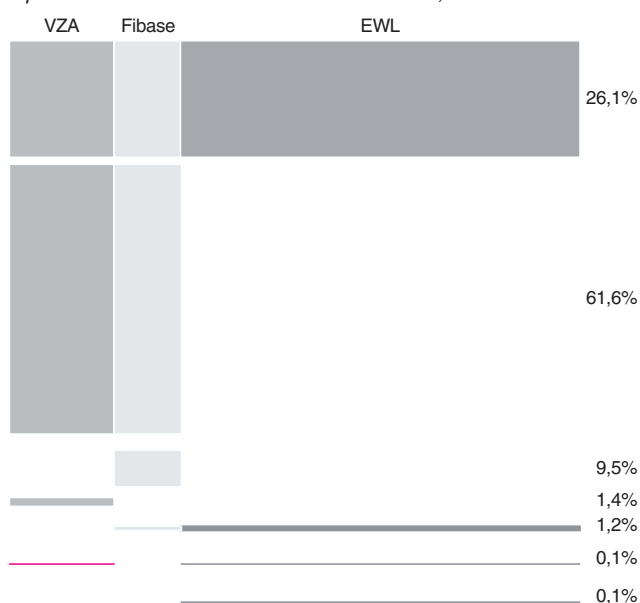
	Fibase	VZA	EWL
	x 1 000		
Totaal	18 793	10 795	2 990
w.v.			
verwijderd door selectie of samenvoeging	8 305 ¹⁾	1 287	64
resterend in SSB banenbestand werknemers	10 488	9 508	2 926
	%		
w.o. gewijzigd			
aanvangsdatum	70	20	8
beëindigingsdatum	24	15	10
fiscaal loon	3	—	—
BLSV	—	3	58
bruto jaarloon	—	—	71

¹⁾ Inclusief uitkeringen.

Bron: CBS, Sociaal Statistisch Bestand.

procent van het totaal aantal banen). Onder banenrecords die uitsluitend gebaseerd zijn op het loonbelastingbestand komen banen voor van directeur-groootaandeelhouders (DGA's), die niet onder de werknemersverzekeringen vallen. Voor een deel neemt de EWL deze banen overigens ook waar. Ook zitten in de aanvulling uit de loonbelasting banen die niet in bedrijfsadministraties voorkomen. Daarbij gaat het bijvoorbeeld om banen van kranten- en folderbezorgers (CBS, 1998).

Schema 1
Opbouw van het banenbestand SSB naar bron, 1999



De Arbeidsrekeningen (AR) is een op macro-niveau geïntegreerd stelsel van arbeidsmarktgegevens (Leunis en Verhage, 1996). Het totaal aantal personen dat in 1999 werkzaam was als werknemer of zelfstandige bedraagt volgens de Arbeidsrekeningen 7,9 miljoen. Het aantal werkzame personen volgens het SSB is 7,4 miljoen ⁵⁾. Het verschil wordt onder andere veroorzaakt door een bijrschatting in de AR van 270 duizend zelfstandigen die voor particuliere huishoudens werken (CBS, 2001a). Bovendien zijn in de Arbeidsrekeningen gegevens opgenomen over meewerkende gezinsleden. Aangezien er geen microgegevens over beide groepen werkenden beschikbaar zijn, ontbreken zij in het SSB. Tot slot ontbreken in het SSB ook gegevens over enkele tienduizenden zelfstandigen, vanwege incompleteitheid van de bron (zie ook paragraaf 3.2.1).

3.3 Uitkeringen

3.3.1 AO-, WW-, Abw-, en IOAW/IOAZ-uitkeringen

Voor de uitkeringen in het kader van de AO, WW, Abw en IOAW/IOAZ zijn betalingsgegevens per maand beschikbaar en is het mogelijk uitkeringen zuiver op betalingsbasis af te bakenen. Voor de zogenaamde 'nuluitkeringen', die niet op betalingsbasis meetellen, wordt nagegaan of er in het loonbelastingbestand positieve betalingsgegevens opgenomen zijn. Wanneer dat niet het geval is, worden zij verwijderd.

Er wordt een onderscheid gemaakt tussen *reguliere betalingen* en *nabetalingen*. Bepalend hiervoor is het einde van het *uitkeringsrecht*. Alle betalingen die hebben plaatsgevonden voordat het uitkeringsrecht verstreken is maken deel uit van het reguliere betalingstraject. Is het uitkeringsrecht verstreken en vinden er nog betalingen plaats, bijvoorbeeld als gevolg van het uitkeren van vakantiegeld, verrekeningen en dergelijke, dan valt dat onder het *nabetalingstraject* ⁶⁾. Wanneer sprake is van het samengaan van een reguliere betaling van een uitkering met een baan, spreken we van *samenloop*. Het gelijktijdig voorkomen van een baan en nabetalingen van uitkeringen wordt binnen het SSB niet als samenloop beschouwd.

Volgens het principe van het betalingsconcept zou een uitkering geen maand met een nulbetaling mogen bevatten. Dat impliceert dat tussentijdse perioden met een nulbedrag tot opdeling leiden van de uitkering in meerdere uitkeringen met nieuwe aanvangs- en beëindigingsdatums. Voorlopig is er voor gekozen deze strikte toepassing niet te gebruiken, omdat in de uitkeringsadministratie sprake kan zijn van administratieve vertraging van de betalingsinformatie. Dit komt bijvoorbeeld tot uiting bij een uitkering met een regelmatig patroon van maandbedragen, onderbroken door een nulbedrag en vervolgd met een maand met een dubbel maandbedrag.

Doordat betalingsperioden met nulbetalingen voorkomen en het SSB zich toelegt op één verslagjaar ontstaat waarschijnlijk een overschatting van de in- en uitstroom. Zo wordt een uitkering met in december 1998 een betaling, in januari 1999 een nulbetaling en daarna weer reguliere betalingen, in feite ten onrechte als instroom gezien. Hetzelfde geldt voor een uitkering met voor november 1999 reguliere betalingen, in december 1999 een nulbetaling en in januari 2000 weer reguliere betalingen: deze wordt ten onrechte als uitstroom gezien. Bovendien tellen beide soorten uitkeringen ten onrechte niet mee aan het begin en aan het einde van het jaar. Hiervoor is te corrigeren door meerdere verslagjaren in de afbakening van het betalingsconcept mee te nemen. Voor de beschreven uitkeringenbestanden van 1999 is dit nog niet gebeurd.

Bij de Abw-uitkeringen zijn uitsluitend gegevens over periodiek algemene uitkeringen meegenomen. Uitkeringen in het kader van de bijzondere bijstand zijn buiten beschouwing gelaten. Abw- en

IOAW/IOAZ-uitkeringen worden uitgekeerd aan huishoudens. Zowel aanvragers als partners van aanvragers van een uitkering komen in de bronbestanden voor. In het SSB zijn uitkeringsrecords waarbij sprake is van een partner gesplitst in een aanvrager- en een partnerrecord, met ieder de helft van de bedragen.

3.3.2 Overige uitkeringen

Het Fibase-bestand vormt de enige bron voor andere uitkeringen dan genoemd in paragraaf 3.3.1. Het gaat daarbij bijvoorbeeld om uitkeringen in het kader van de Algemene Ouderdomswet (AOW), al dan niet gecombineerd met pensioenuitkeringen, de Algemene Nabestaanden Wet (ANW), Ziektewet (ZW) en de Wachtgeldregeling Overheid. Evenals bij de banen is het niet mogelijk te bepalen of het gaat om reguliere betalingen of dat uitsluitend van nabetalings sprake is.

3.4 Confrontatie van banen en uitkeringen

De confrontatie van banen en uitkeringen is bedoeld om de werkelijke samenloop van banen en uitkeringen vast te stellen en samenloop als gevolg van bijvoorbeeld administratieve redenen te verwijderen. Werkelijke samenloop is het gelijktijdig voorkomen van reguliere betalingen uit een baan en een uitkering bij één persoon. Er zijn diverse oorzaken waardoor 'onwerkelijke' samenloop ontstaat. Zo kan in de administratie van werknemers geen beëindigingsdatum van een baan zijn opgenomen. Wanneer deze werknemer na beëindiging van het dienstverband bijvoorbeeld een WW-uitkering ontvangt, ontstaat een administratieve overlap. Ook kan een baan administratief later worden afgesloten als gevolg van nabetalings.

De bronnen van de banen (EWL, VZA-werknemers en Fibase) bevatten uitsluitend gesommeerde betalingsgegevens over het gehele jaar met daarbij een administratieve aanvangs- en beëindigingsdatum⁷⁾. De reguliere betalingsperiodes van de uitkeringsbestanden AO, Abw, IOAW/IOAZ en WW zijn bepaald aan de hand van gerapporteerde maandbetalingen. Hierdoor sluiten de aanvangs- en beëindigingsdatums van de uitkeringsbestanden aanzienlijk beter aan op het betalingsconcept. In de confrontatie met de banen is dan ook als uitgangspunt gekozen dat de aanvangs- en beëindigingsdatums van de genoemde uitkeringen niet worden gecorrigeerd. Hierop zijn twee uitzonderingen. Wanneer het wijzigen van de dag van aanvang of beëindiging van de uitkering leidt tot het verdwijnen van overlap met de baan, dan wordt dit toegepast. De aanname is dan dat de overlap tussen een baan en een uitkering op administratieve afhandeling berust. De tweede uitzondering wordt gemaakt voor nuluitkeringen met een positief bedrag in het loonbelastingbestand. Aangezien niet duidelijk is in welke periode betalingen hebben plaatsgevonden, kan er met de aanvangs- en beëindigingsdatum worden geschoven.

Bij de banen is niet altijd bekend in welke periode de betalingen precies hebben plaatsgevonden. Voor een groot aantal banen gaat het om maandelijkse, reguliere betalingen, in de aangegeven gelegen administratieve registratieperiode (de in de bestanden aanwezige datums). Bij een ander deel is echter sprake van onregelmatig werk, waarbij de registratieperiode niet de feitelijke betalingsperiode weergeeft. En juist deze banen, bijvoorbeeld in de uitzendbranche, komen vaker voor in combinatie met uitkeringen. Er is minimaal één beperking aan wijziging van aanvang en beëindiging, namelijk dat het aantal sociale verzekeringsdagen van de baan in de betreffende kalenderperiode moet blijven 'passen'. Bij de confrontatie van banen en uitkeringen wordt dus geprobeerd zoveel mogelijk oneigenlijke samenloop te elimineren door te schuiven met de aanvangs- en beëindigingsdatum van de baan, zodanig dat het aantal sociale verzekeringsdagen past in de nieuw bepaalde betalingsperiode van de baan. De consequentie van deze benadering is dat deeltijdbanen die werkelijk samenloop vertonen met een uitkering een grotere kans hebben in looptijd te worden gekort dan voltijdbanen. Deeltijdbanen hebben immers minder sociale verzekeringsdagen dan voltijdbanen.

De confrontatie gaat voorbij aan tussentijdse nulbetalingen in de betalingsperiode van de uitkeringen. Het is mogelijk dat een persoon een baan heeft gehad tijdens de reguliere looptijd van de uitkering en dat de betaling van de uitkering tijdelijk is stopgezet. In feite is er dan geen sprake van samenloop. Tussentijdse nulbetalingen zijn in de huidige opzet opgenomen in de looptijd van de reguliere betalingen. Dit betekent dat in dit geval ten onrechte samenloop wordt geconstateerd. Hierbij zijn enkele relativerende opmerkingen op zijn plaats. Ten eerste vindt administratieve verrekening van inkomsten uit loon met het uitkeringsbedrag niet altijd direct plaats in de maand(en) dat arbeid wordt verricht. Hierdoor blijft er toch een zekere mate van 'onterechte' samenloop bestaan tussen betalingen van de uitkering en het verrichten van werk. Ten tweede is zowel de omvang van het aantal banen als de looptijd van de banen die bij een dergelijke samenloop horen, relatief gering. Dit betekent dat de beschreven onterechte samenloop gering zal zijn.

3.5 Het personenbestand

De basis voor het personenbestand is het gecumuleerde GBA-bestand. Hieruit worden alle records van personen geselecteerd, die in 1999 tot de bevolking van Nederland behoorden. Er vindt hierop een aantal aanvullingen plaats. Ten eerste worden gegevens van personen toegevoegd, die in 1999 niet tot de bevolking behoorden, maar wel in Nederland werkzaam waren of hier een uitkering hadden³⁾. Ten tweede wordt voor personen, waarvan al via het GBA-bestand gegevens in het personenbestand zijn opgenomen, een extra record toegevoegd wanneer blijkt dat een baan- of uitkeringsperiode niet correspondeert met de periode dat een persoon tot de bevolking van Nederland behoorde. Wanneer iemand bijvoorbeeld in augustus van het jaar is geëmigreerd, maar het banenrecord had een beëindiging in november, dan wordt voor deze persoon een extra record toegevoegd met kenmerken over de periode tussen de datum van emigratie en de beëindiging van de baan in november. Toevoegen van extra persoonsrecords in het personenbestand zorgt voor consistentie met banen- en uitkeringsbestanden. De toegevoegde records zijn herkenbaar gemaakt, zodat voor ieder peilmoment bekend is of iemand al of niet tot de bevolking van Nederland behoorde.

3.6 Het SSB van 1998

De werkwijze voor het SSB van 1999 wijkt af van de werkwijze van 1998. De definitie van werknemers sluit voor 1998 zoveel mogelijk aan bij de door de EWL gehanteerde definitie: 'banen van werknemers in bedrijfsadministraties' (CBS, 2000c). Dit betekent dat banen van werknemers die niet in bedrijfsadministraties zitten, in 1998 zijn uitgesloten. Ook worden banen van werknemers in dienst van huishoudens en internationale organen volgens de EWL-definitie niet meegenomen. Wel is besloten om de EWL-populatie te completeren met werknemers in dienst van bedrijven die in het Algemeen Bedrijfsregister (ABR) getypeerd waren als 'bedrijven zonder personeel'. Gegevens hierover zijn opgenomen in het VZA-werknemersbestand. De Fibase is vervolgens gebruikt om twee groepen banen bij te schatten, namelijk banen van DGA's en overige banen van werknemers die door VZA en EWL worden 'gemist'. Voor de eerste groep zijn records met de specificatie 'loon/salaris directeuren NV/BV verzekerd voor de werknemersverzekeringen' uit de Fibase geselecteerd. Voorzover deze records niet koppelden met banenrecords uit VZA en EWL en een statistische bedrijfseenheid kon worden toegekend, zijn deze opgenomen in het banenbestand van 1998. Voor de tweede groep zijn records met loon/salaris uit de Fibase geselecteerd, waarvan de persoon in de GBA voorkomt, maar niet met een banenrecord in het (voorlopige) banenbestand. Daarbij is het noodzakelijk een onderscheid te maken tussen banen die niet tot de doelpopulatie behoren, zoals van kranten- en folderbezorgers of stagiaires, en fiscaal loon of salaris dat ligt in de sfeer van vergoedingen (vrijwilligers). Dit is geoperationaliseerd door alleen die

records uit de Fibase te selecteren met een fiscaal loon van 2 500 gulden (1 135 euro) of meer op jaarbasis en waaraan een statistische bedrijfseenheid kon worden toegekend.

Voor het banenbestand van 1999 is de Fibase integraal gebruikt. Dit betekent dat ook alle andere records met loon of salaris uit de Fibase, anders dan hierboven genoemd, in het banenbestand van 1999 zijn opgenomen. De uitkomsten van de banenbestanden van 1998 en 1999 zijn dan ook niet vergelijkbaar. Bovendien zijn alle overige uitkeringen, zoals waargenomen in de Fibase voor het verslagjaar 1999 in een apart bestand opgenomen (zie subparagraaf 3.3.2), terwijl dat voor 1998 niet is gedaan.

Het banenbestand van 1998 is gebruikt voor een aantal vervolgtrajecten: het toekennen van de werkgemeente (Corpeleijn en Heerschoop, 2002), de koppeling met het studentenbestand hoger onderwijs (Florquin, 2002) en de koppeling met enquêtes (Van Cruchten, Hartgers en Schulte-Nordholt, 2002; Weidum en Linder, 2002).

4. Toevoegen van enquêtevariabelen

Zoals hierboven is beschreven is het SSB voor een belangrijk deel opgebouwd uit registergegevens. Er bestaan echter niet voor alle relevante variabelen registergegevens. Voor gegevens over variabelen als opleidingsniveau, beroepsniveau en zoekgedrag op de arbeidsmarkt zijn enquêtes onder personen en huishoudens nodig. Deze enquêtes worden eerst op microniveau gekoppeld aan de integrale bestanden van het SSB, hierna genaamd *registers*. Vervolgens worden met behulp van de methode van het *herhaald wegen* tabellen geschat die consistent zijn met de uitkomsten uit de registers.

Bij het herhaald wegen wordt elke tabel geschat op basis van zoveel mogelijk gegevens. Daarbij wordt rekening gehouden met alle gemeenschappelijke (sub)populatiетotalen van deze tabel en eerder berekende tabellen. Een nieuw te schatten tabel vereist een nieuw weegmodel, dat aangeeft met welke (sub)populatiетotalen van eerder berekende tabellen rekening moet worden gehouden. Het herwegen naar deze gegeven (sub)populatiетotalen wordt *kalibreren* genoemd. Als alle tabellen op deze wijze worden geschat, zijn de uitkomsten onderling consistent.

4.1 Koppelen op peilmoment of op enquêtedatum

Bij het koppelen van de enquêtes aan de registers zijn verschillen in de spreiding van de waarneming in de tijd van belang. Het SSB heeft de vorm van een volumebestand waaruit over iedere willekeurige datum een standgegeven kan worden afgeleid. Enquêtes worden meestal verspreid over een bepaalde periode of op een bepaalde peildatum afgenomen. Om de variabelen uit enquêtes goed te kunnen ophogen naar de (sub)populatiетotalen uit de registers, moeten alle bronnen in principe betrekking hebben op dezelfde periode of peildatum.

Doorlopende enquêtes worden bij voorkeur op enquêtedatum gekoppeld aan volumeregisters en kunnen vervolgens worden opgehoogd naar de bijbehorende jaargemiddelden. Het koppelen op enquêtedatum is met name een eis als de variabelen sterk situationeel bepaald zijn. Dit geldt bijvoorbeeld voor variabelen als de arbeidsmarktpositie en het zoeken naar werk. De tabellen over de arbeidsmarktpositie van personen met een uitkering zijn geschat door de variabelen uit de EBB op enquêtedatum te koppelen aan de registers en deze daarna op te hogen naar registertotalen, die door middeling van kwartaalstanden tot stand zijn gekomen.

Niet alle gegevens zijn echter op enquêtedatum te koppelen. In sommige gevallen moeten tabellen geschat worden over de situatie op een bepaalde datum, in andere gevallen wordt een bepaald gedeelte van de variabelen op een vaste datum waargenomen en

een ander gedeelte verspreid door het jaar. Dat laatste komt voor bij het Loonstrucuuronderzoek (LSO). Een gedeelte van de variabelen voor het LSO komt uit de EWL en een ander gedeelte uit de EBB.

De EWL-variabelen, zoals loon en arbeidsduur, hebben betrekking op de situatie in de laatste week van december. De EBB-variabelen, zoals opleiding en beroep, worden verspreid over het jaar waargenomen. Omdat deze EBB-variabelen minder frequent van waarde veranderen dan variabelen als arbeidsmarktpositie, hoeven deze niet per se op enquêtedatum te worden gekoppeld. De belangrijkste reden om het gehele jaar te koppelen is dat er weinig EBB-records met een enquêtedatum eind december 1998 zijn. Bij het schatten van de LSO-tabellen over 1998 zijn daarom de opleidings- en beroepsgegevens uit de EBB 1998 en 1999 gekoppeld aan de EWL-gegevens over de laatste week van 1998 en de registerinformatie over diezelfde week. Op deze wijze wordt de massa aanzienlijk vergroot.

4.2 Herhaald wegen

De methode van het herhaald wegen zorgt ervoor dat de te publiceren tabellen consistent worden geschat. Elke tabel die uit de enquêtes geschat wordt, is daarmee consistent met alle relevante (sub)populatiетotalen die uit de registers geteld kunnen worden, ook als deze niet in het oorspronkelijke weegschema van de enquête zijn opgenomen. Door telkens te kalibreren op de registerstellingen of op eerder geschatte (sub)populatiетotalen, worden de uitkomsten onafhankelijk van de gebruikte bron. Zo zijn bijvoorbeeld de (sub)populatiетotalen van de tabel *maandloon naar economische activiteit en opleidingsniveau* die uit de doorsnede van de EWL en de EBB geschat worden, gelijk aan de (sub)populatiетotalen uit de tabellen *maandloon naar economische activiteit* (EWL) en de tabel *opleidingsniveau* (EBB).

De methode van het herhaald wegen is eerder uitvoerig beschreven (Kooiman, Kroese en Renssen (2000), Kroese, Renssen en Trijssenaar (2000) en Renssen, Kroese en Willeboordse (2001)). Voor een uitgebreidere beschrijving van de toepassing van het herhaald wegen bij het schatten van SSB-tabellen wordt verwezen naar CBS (2001b) en naar CBS (2002).

Om elke tabel zo nauwkeurig mogelijk te kunnen schatten, is elke schatting gebaseerd op een zo groot mogelijk aantal waarnemingen. Daarom worden bij het herhaald wegen telkens de gegevens uit de grootst mogelijke bron of grootst mogelijke combinatie van bronnen gebruikt. Tabellen die uitsluitend registervariabelen bevatten, worden geteld uit de registers. Tabellen die een combinatie van variabelen uit één of meer steekproeven bevatten, worden geschat uit de grootst mogelijke combinatie van steekproeven. Deze combinaties heten *rechthoekige datablokken*.

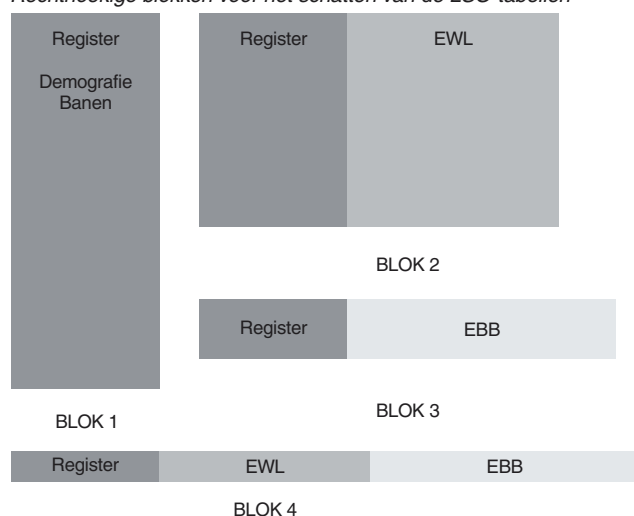
4.3 Rechthoekige datablokken

Het vormen van volledig gevulde rechthoekige datablokken, die worden gebruikt bij het schatten van de verschillende tabellen, is daarom een belangrijke stap in het proces van het herhaald wegen. Een voorbeeld kan het gebruik van rechthoekige blokken verduidelijken. Schema 2 laat de bestanden zien die worden gebruikt bij het schatten van de LSO-tabellen. Door de combinatie van registers en steekproeven ontstaan verschillende blokken, waaruit de LSO-tabellen worden geschat. In dit voorbeeld zijn er vier rechthoekige datablokken. Deze blokken bevatten steeds een selectie van variabelen en alle records waarvoor deze variabelen waargenomen zijn.

Blok 1 beschrijft alle registervariabelen zoals geslacht, leeftijd en economische activiteit. Omdat deze variabelen voor alle banen bekend zijn, bevat blok 1 alle records uit het banenregister. Blok 2 beschrijft variabelen zoals het jaarloon en bestaat uit alle EWL-records, waaraan de registervariabelen uit blok 1 zijn toege-

voegd. Blok 3 beschrijft opleiding en beroep en bestaat uit de EBB-records en de registervariabelen uit blok 1. Blok 4 tenslotte beschrijft de samenhang tussen lonen, opleiding, beroep en de registervariabelen en bestaat uit de doorsnede van de records uit de EWL en de EBB, waaraan de registervariabelen zijn toegevoegd.

Schema 2
Rechthoekige blokken voor het schatten van de LSO-tabellen



Elke gewenste tabel kan nu worden geschat (of geteld) uit het grootst mogelijke blok. Het aantal banen naar economische activiteit en banen naar geslacht en leeftijd komt uit blok 1. De tabellen over lonen naar geslacht en leeftijd of naar economische activiteit worden geschat uit blok 2. Blok 3 levert de tabellen over banen naar opleiding en beroep, terwijl de tabellen over lonen naar opleiding en beroep worden geschat uit blok 4. Herhaald wegen leidt ertoe dat dat een (sub)populatietotaal dat berekend is uit het ene blok gelijk is aan hetzelfde (sub)populatietotaal dat berekend is uit een ander blok.

Literatuur

- Al, P. en J.W. Altena (2000), 'Data security, privacy and the SSB', *Netherlands Official Statistics*, **15**, pp. 47–50.
- Arts, C.H., en F.J. van Lith (1999), 'Een nieuwe weg voor statistieken over personen: het Sociaal Statistisch Bestand (SSB)', *Sociaal-economische maandstatistiek*, september 1999, pp. 22–29.
- Arts, C.H., Bakker, B.F.M. en F.J. van Lith (1999), 'Linking administrative registers and household surveys', *Netherlands Official Statistics*, **15**, pp. 16–21.
- Arts, C.H., Van Toor, L. en S. de Vries (2002), 'Stromen op de arbeidsmarkt, april–oktober 1999', *Sociaal-economische maandstatistiek*, december 2002, pp. 66–71.
- Boerdam, A.A. (2002), 'Hoeveel allochtonen ontvangen een uitkering?', *Sociaal-economische maandstatistiek*, december 2002, pp. 35–41.
- Centraal Bureau voor de Statistiek (1992), *Standaard bedrijfsindeling (SBI) overzicht en schakelschema's*, CBS, Voorburg/Heerlen.
- Centraal Bureau voor de Statistiek (1998), *Arbeidsrekeningen 1996–1997*, CBS, Voorburg/Heerlen.
- Centraal Bureau voor de Statistiek (2000a), *Bestandsindeling en recordbeschrijving CBS-uitvraag salarisadministraties met ingang van 2001*, CBS, Voorburg/Heerlen.
- Centraal Bureau voor de Statistiek (2000b), *Bedrijven in Nederland 2000*, CBS, Voorburg/Heerlen.
- Centraal Bureau voor de Statistiek (2000c), 'Banen, lonen en arbeidsduur van werknemers, 1999–2000', *Sociaal-economische maandstatistiek*, december 2000, pp. 30–43.
- Centraal Bureau voor de Statistiek (2001a), *De Nederlandse economie 2000*. CBS, Voorburg/Heerlen.

Centraal Bureau voor de Statistiek (2001b), *Weegproces SSB, Vol. 1 Handleiding voor het schatten van tabellen met herhaald wegen*, CBS, Voorburg.

Centraal Bureau voor de Statistiek (2002; nog te verschijnen), *Opname van enquêtevariabelen in het SSB; methodebeschrijving en uitkomsten voor het loonstrucuuronderzoek en de arbeidsmarktpositie van personen met een uitkering, 1998*, CBS, Voorburg.

Cruchten, J.M.J. van, Hartgers, M.I. en E. Schulte-Nordholt (2002), 'Loon naar opleidings- en beroepsniveau: het Loonstrucuuronderzoek 1998.', *Sociaal-economische maandstatistiek*, december 2002, pp. 72–81.

Corpeleijn, A.W.F. en M.J. Heerschop (2002), 'Inkomende en uitgaande pendel, 1998', *Sociaal-economische maandstatistiek*, december 2002, pp. 53–58.

Das, M., Van Toor, L. en C.H. Arts, (2002), 'Samenloop van banen en uitkeringen', *Sociaal-economische maandstatistiek*, december 2002, pp. 22–28.

Florquin, F.F. (2002), 'Banen van studenten, april 1998', *Sociaal-economische maandstatistiek*, december 2002, pp. 59–65.

Kooiman, P., Kroese A.H. en R.H. Renssen (2000), *Official Statistics an Estimation Strategy for the IT-Era*, Research Paper, CBS, Voorburg.

Kroese, A.H., Renssen R.H. en M. Trijssenaar (2000), 'Weighting or imputation: constructing a consistent set of estimates based on data from different sources.', *Netherlands Official Statistics*, **15**, pp. 23–31.

Leunis, W.P. en K.G. Verhage (1996), *Arbeidsrekeningen. Kern van het statistisch systeem over arbeid*, Academisch Proefschrift, CBS, Voorburg.

Linder, F.S. (2002), 'Bronnen van inkomen in de regio', *Sociaal-economische maandstatistiek*, december 2002, pp. 42–52.

Melser, C. en L. van Toor (2002), 'Hoeveel allochtonen zijn werknemer?', *Sociaal-economische maandstatistiek*, december 2002, pp. 29–34.

Renssen, R.H., Kroese A.H. en A.J. Willeboordse (2001), *Aligning Estimates by Repeated Weighting*, CBS, Heerlen.

De Vries, Sylvia (1998) 'Elektronische dataverzameling van arbeid- en loongegevens', *Sociaal-economische maandstatistiek*, november 1998, pp. 25–28.

Weidum, J. en F.S. Linder (2002) 'Arbeidspositie en opleidingsniveau van personen met een uitkering, 1998', *Sociaal-economische maandstatistiek*, december 2002, pp. 82–92.

Noten in de tekst

- 1) Paragraaf 4.
- 2) Er is gekozen voor de maand oktober, omdat dit een relatief stabiele periode is.
- 3) Deze groep bevat ook gegevens van personen die ten onrechte niet met het GBA-bestand konden worden gekoppeld, de zogenaamde gemiste koppelingen.
- 4) Kleine groepen personen met winst zijn niet als zelfstandige actief geweest. Deze zijn niet als zodanig in het bestand herkenbaar.
- 5) Bij de Arbeidsrekeningen gaat het om het jaargemiddelde van 1999, bij het SSB om de stand van ultimo oktober 1999 van personen die in Nederland woonden. In het AR-cijfer is de inkomende pendel vanuit het buitenland opgenomen, in het SSB-cijfer niet.
- 6) Het kan ook gaan om negatieve bedragen (in het geval van terugbetalingen).
- 7) De brongegevens van de EWL bevatten veelal betalingsgegevens per maand of kwartaal; voor het SSB is echter een bewerkt bestand gebruikt dat uitsluitend gegevens bevat over banen ultimo van het jaar.