
Arbeidsmarktstatistieken in Nederland.

De opzet van het sociaal statistisch bestand

Enquêtes op het terrein van de arbeidsmarkt hebben hun beperkingen om in de snel veranderende informatiebehoefte te voorzien. In dit artikel wordt het alternatief gepresenteerd van het Sociaal Statistisch Bestand (SSB) van het Nederlandse Centraal Bureau voor de Statistiek. In het SSB worden gegevens uit registers en enquêtes op het niveau van de individuele personen gecombineerd tot een geïntegreerde data-infrastructuur. De informatie in het SSB beperkt zich niet tot de arbeidsmarktstatistieken, maar beslaat een aantal verschillende statistische thema's.

Veranderde informatiebehoefte over de arbeidsmarkt

De arbeidsmarkt is niet alleen zelf sterk in beweging, ook de vraag naar statistische informatie over de arbeidsmarkt verandert. Zo is er een groeiende behoefte aan informatie waarin thema's die vroeger tot verschillende wetenschappelijke disciplines behoorden, met elkaar in verband worden gebracht. Naast het klassieke interdisciplinaire onderwerp waarin de relatie tussen onderwijs en arbeidsmarkt wordt onderzocht, bestaat er nu bijvoorbeeld ruime belangstelling voor de gezondheid van werknemers. Dit houdt verband met de vergrijzing van de bevolking. Hierdoor is het noodzakelijk dat ouderen langer doorwerken en er zoveel mogelijk gebruik gemaakt wordt van het arbeidspotentieel van arbeidsgehandicapten. Ook moet kostbaar ziekteverzuim en uitreding uit het arbeidsproces door arbeidsongeschiktheid worden voorkomen.

Een andere ontwikkeling die we in de vraag naar gegevens constateren, is die naar stromen op de arbeidsmarkt. Deze informatie is onmisbaar voor een goed inzicht in het functioneren van deze markt en het nagaan van de effectiviteit van werkgelegenheidsbeleid. De informatiebehoefte is de afgelopen jaren veranderd door de toegenomen differentiatie gedurende het leven van mensen. Voorbeelden hiervan zijn de nog steeds in beweging zijnde rolverdeling tussen mannen en vrouwen, de toename

van het aantal mensen dat werk en zorg combineert, de flexibilisering en toegenomen dynamiek van de arbeidsmarkt en de herziening van de stelsels van sociale zekerheid (Manting, 2003; Schmid & Gazier, 2002).

Informatie over de stromen in en uit uitkeringssituaties zijn ook belangrijk voor de evaluatie van het arbeidsmarktbeleid. Wijzigingen in socialezekerheidsarrangementen zijn bij uitstek geschikt om processen op de arbeidsmarkt bij te sturen. Daarvoor is onder meer informatie nodig over wie gebruik maakt van verschillende uitkeringen, welke categorieën er in slagen om weer een baan te vinden en met welke snelheid dit gebeurt.

Een andere belangrijke ontwikkeling in de informatiebehoefte is de toegenomen vraag naar de relatie tussen de kenmerken van bedrijven en het functioneren van de arbeidsmarkt. Is de loonvor-

ming bij op winst gerichte bedrijven anders dan bij bedrijven in de non-profit sector? Wie vloeit af als een bedrijf zijn personeelsbestand afslankt? Zijn er verschillen tussen bedrijfstukken in dit opzicht? Waar komen degenen die zijn ontslagen vervolgens terecht? Welke invloed hebben investeringen in research and development?

Voor lokaal arbeidsmarktbeleid is informatie over het functioneren van de regionale arbeidsmarkt van groot belang. Ook worden we geconfronteerd met veel vragen naar informatie over kleine groepen in de samenleving. Dit laatste heeft onder meer te maken met de sterke differentiatie van de positie van allochtonen op de arbeidsmarkt. Sommige allochtone groepen doen het goed in het onderwijs en op de arbeidsmarkt, andere zijn juist minder succesvol. Niet alleen is het van belang te weten hoe het met de niet-westerse allochtonen in het algemeen op de arbeidsmarkt gaat. Ook is de vraag hoe het met bijvoorbeeld de laag opgeleide allochtonen van Antilliaanse afkomst tussen 12 en 25 jaar in de gemeente Delfzijl gaat.

Enquêtes en registers

De veranderde informatiebehoefte noopt het CBS tot het opnieuw inrichten van de arbeidsmarktstatistiek. Belangrijk daarbij is de constatering dat de vraag naar arbeidsmarktinformatie verandert en dat zal blijven doen. Dat vraagt om een data-infrastructuur waarmee gemakkelijk een groot aantal verschillende vragen kan worden beantwoord, ook vragen die nu nog niet voorzien zijn.

De meest aangewezen enquête om vragen over de arbeidsmarkt te beantwoorden is de Enquête beroepsbevolking, de Nederlandse variant van de Europese Labour Force Survey. Enquêtes kunnen echter in beperkte mate aan de hierboven omschreven nieuwe vraag voldoen. Immers, als te veel thema's in een enquête worden opgenomen dan wordt deze veel te lang. Als stromen op de arbeidsmarkt moeten worden waargenomen, krijgt de enquête een paneldesign of moeten retrospectieve vragen opgenomen worden. Bij een paneldesign is paneluitval een serieus probleem, terwijl bij retrospectieve vragen geheugeneffecten verstorend werken. Kenmerken van bedrijven worden waarschijnlijk in enquêtes vrij slecht waargeno-

men, omdat de respondenten hiervoor in het algemeen de kennis missen. Verder is een enquête altijd op steekproefbasis waardoor het moeilijk is om betrouwbare informatie over kleine regio's of groepen samen te stellen (zie ook Tielens, 2004, blz. 93).

Door alle beschikbare registerinformatie te koppelen, zoals in de kern van het SSB wordt gedaan, kan een belangrijk deel van de benodigde informatie worden geleverd. Toch blijven enquêtes noodzakelijk voor het samenstellen van arbeidsmarktstatistiek. In de eerste plaats omdat een deel van de informatie die nodig is niet in registers aanwezig is. In de tweede plaats omdat de registers soms een behoorlijke vertraging hebben, waardoor de informatie niet actueel genoeg is. Enquêtes kunnen voor de belangrijkste arbeidsmarktindicatoren een snel cijfer leveren.

De kern van het Sociaal Statistisch Bestand

De ruggengraat van het SSB bestaat uit de Gemeentelijke Basisadministratie personen (GBA). Dit bestand bevat demografische gegevens zoals geslacht, geboortedatum, burgerlijke staat, geboorteland en woonplaats van alle personen die vanaf 1995 enig moment tot de bevolking van Nederland behoorden. Alle wijzigingen hierin, zoals verhuizingen en echtscheidingen, zijn in het bestand opgenomen. Alle overige bronnen van het SSB zijn gekoppeld aan de GBA. Daaronder zijn integrale registraties van de belastingdienst van werknemers, zelfstandigen en uitkeringsgerechtigden, registraties van het Uitvoeringsinstituut Werknemersverzekeringen (UWV) die naast registraties van werknemers ook de werkloosheids- en arbeidsongeschiktheidsuitkeringen verzorgt, en registraties van de Informatie Beheergroep (IBG) die verantwoordelijk is voor een belangrijk deel van de registraties in het onderwijsveld. Daarnaast is een (grotendeels elektronische) bedrijfsenquête opgenomen waarin iets meer dan de helft van het aantal werknemers is opgenomen: de Enquête Werkgelegenheid en Lonen.

Eerst koppelen we zoveel mogelijk records op basis van het sociaal-fiscaal nummer (sofi-nummer), het in Nederland veel gebruikte persoonlijk identi-

ficatienummer voor fiscale doeleinden en uitvoering van sociale verzekeringswetten. Koppelingen op het sofi-nummer leveren een rendement van bijna 100% op. Records die geen sofi-nummer bevatten, koppelen we op de kenmerken geslacht, geboortedatum en adres. Het koppelingsrendement op deze combinatie hangt af van de kwaliteit van de koppelingsgegevens, maar ligt tussen de 95 en 100%.¹

Hoewel de registers meestal van een goede kwaliteit en zeker niet zwaar vervuild zijn, is het niet zo dat alle bronnen direct bruikbaar zijn voor de statistiek. Het proces waarin de registers worden verwerkt tot statistische gegevens wordt micro-integratie genoemd (zie ook Van der Laan, 2000; Bakker, 2002; Arts en Hoogetijling, 2002; Bakker en Arts, 2003).

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. Dikwijls worden de statistische variabelen die nodig zijn voor publicaties afgeleid uit een aantal variabelen die in verschillende registers zijn opgenomen, om daarmee te voldoen aan de definities die in de statistiek gebruikelijk zijn. Een voorbeeld daarvan is het inkomen, dat wordt afgeleid vanuit bijvoorbeeld loon en uitkeringsbedragen die in verschillende registers voorkomen.

Ten tweede bevatten de registers niet altijd de gehele populatie van eenheden die onderwerp zijn van de statistiek. Een populatie van eenheden is bijvoorbeeld de verzameling van banen die bij de bedrijven in Nederland bestaan. Door alle registers die betrekking hebben op soortgelijke eenheden samen te voegen, krijgen we een completer beeld van de populatie van eenheden. Als gegevens over de banen uit verschillende registers worden samengevoegd, wordt de populatie van banen compleet, omdat in ieder van de registers banen voorkomen die niet in de andere registers voorkomen.

Ten derde bevatten registers en enquêtes fouten van allerlei aard. Het is daarom gewenst de verschillende bronnen met elkaar te vergelijken. De fouten vallen dan op en kunnen worden verbeterd. Als twee bronnen dezelfde variabele bevatten, dan horen die voor dezelfde persoon ook dezelfde

waarde op te leveren. Het kan echter ook zijn dat er een logische relatie is tussen variabelen, bijvoorbeeld het uurloon vermenigvuldigd met het aantal gewerkte uren per maand levert het maandloon op. Door gebruik te maken van deze systematische relaties, worden fouten in de bronnen opgespoord. Alleen als de fouten zo omvangrijk zijn dat ze van invloed zijn op de te publiceren uitkomsten, wordt naar een manier gezocht om deze te corrigeren.

Op dit moment zijn gegevens voor de verslagjaren 1999-2002 beschikbaar. In juni 2005 volgt het verslagjaar 2003. Uit de kern van het SSB wordt regelmatig gepubliceerd op de elektronische outputdatabase van het CBS: StatLine (CBS StatLine, 2005) en in tal van artikelen.

De satellieten van het Sociaal Statistisch Bestand

De kern van het SSB bevat ook verschillende koppelv variabelen, zodat andere informatie voor specifiek onderzoek daaraan kan worden gekoppeld. In de eerste plaats kunnen gegevens van personen worden gekoppeld. Zo zijn ook persoons-enquêtes in het SSB opgenomen. De Enquête Beroepsbevolking zorgt bijvoorbeeld voor aanvullende informatie over het opleidingsniveau en het beroep dat mensen hebben, zodat onderzoek naar loonverschillen gecorrigeerd voor deze kenmerken mogelijk is (Van Cruchten, et al., 2002; Schulte Nordholt, 2003). In de tweede plaats kan informatie over adressen worden gekoppeld. Zo kan bijvoorbeeld de relatie worden gelegd tussen kenmerken van mensen die in een huur- of in een koopwoning wonen. In de derde plaats kan de informatie van ouders aan die van de kinderen worden gekoppeld indien deze jonger zijn dan 25 jaar. Daarmee kan bijvoorbeeld worden beoordeeld of kinderen afkomstig uit een gebroken gezin vaker een job hebben of juist niet. En last but not least kunnen gegevens van bedrijven aan de kern van het SSB worden gekoppeld.

De koppelv variabelen worden ook gebruikt om voor bepaalde onderwerpen een zogenaamde satelliet van het SSB te maken. De informatie over een bepaald onderwerp die in registers en enquêtes voorkomt, wordt aan de boven beschreven kern van het SSB gekoppeld. Er zijn inmiddels sa-

tellieten voor de beschrijving van de integratie van allochtonen (Van Rijn, *et al.*, 2004), criminaliteit (Blom, *et al.*, 2005), arbeidsmobiliteit en de virtuele volkstelling (CBS, 2003; Gircour, Hartgers & Schulte Nordholt, 2004). Voor iedere satelliet wordt ook weer het proces van micro-integratie doorlopen, zij het dat de gegevens die in de kern van het SSB voorkomen, overgenomen worden. Dit laatste is noodzakelijk voor het bewaken van de consistentie tussen de satellieten en de kern van het SSB.

De hier genoemde satellieten worden allemaal nog verder ontwikkeld. Bovendien worden enkele nieuwe satellieten voorzien voor de komende jaren (sociale en ruimtelijke mobiliteit; schoolloopbanen, gezondheid). Daarbij wordt steeds beoordeeld of met de inhoud van de satellieten de belangrijkste vragen van statistiekgebruikers kunnen worden beantwoord. Om deze gebruikersbehoefte goed in kaart te brengen, wordt ook dikwijls samengewerkt met andere instellingen die op het betreffende terrein over veel expertise beschikken. Tevens wordt regelmatig overlegd met belangrijke gebruikers zoals ministeries en planbureaus.

Gebruik van enquêtes

Voor de satellieten van het SSB is het gebruik van enquêtes onontbeerlijk om aan de informatiebehoefte te voldoen. Door de enquêtes eerst op microniveau te koppelen aan de kern van het SSB en vervolgens met behulp van de methode van 'consistent en herhaald wegen' tabellen te schatten, worden inconsistenties voorkomen. Belangrijkste wijziging van deze weegmethode ten opzichte van bestaande weegmethodes is dat niet langer één gewicht wordt toegekend aan ieder record van een bronbestand, maar dat per tabel weegfactoren worden bepaald. Nadat initiële gewichten per record zijn bepaald om te corrigeren voor verschillende insluitkansen en selectieve non-respons, worden deze gewichten per tabel zodanig aangepast dat de uitkomsten consistent zijn met eerder gepubliceerde of gelijktijdig te publiceren gegevens. Er wordt bovendien steeds herwogen naar het meest betrouwbare gegeven, uit welke bron(nen) dit ook afkomstig is. Dit is dikwijls een registervariabele die integraal beschikbaar is, maar kan ook een variabele zijn die gecoördineerd wordt waargenomen in meerdere surveys. Als alle tabellen op deze wijze

worden geschat, zijn de uitkomsten onderling consistent. De methode van het herhaald wegen is eerder uitvoerig beschreven, bijvoorbeeld door Kroese, Renssen & Trijssenaar (2000) en Houbiers (2004).

De registers kunnen ook gebruikt worden om een efficiëntere steekproef te trekken, omdat deze kan worden gestratificeerd naar kenmerken die in de gekoppelde registers voorkomen. Een deel van de voordelen van het SSB vallen overigens weg bij het gebruik van enquêtegegevens (betrouwbare informatie over kleine groepen, veel regionaal detail, longitudinale analyses).

Ten slotte

Wat betekent het SSB nu concreet voor statistieken op het terrein van de arbeidsmarkt? In de groeiende behoefte aan informatie waarin de samenhang van een groot aantal verschillende verschijnselen wordt beschreven, wordt grotendeels voorzien doordat alle beschikbare bronnen onderling zijn gekoppeld en consistent zijn gemaakt. Er kunnen ook longitudinale analyses worden uitgevoerd, omdat over meerdere jaren gekoppelde en consistente gegevens beschikbaar zijn. Bovendien hebben veel bestanden een volumekarakter, zodat ook duuranalyses tot de mogelijkheden behoren. Doordat gebruik wordt gemaakt van integrale registerinformatie, kan meer en betrouwbaarder informatie worden samengesteld over kleine regionale eenheden en kleine groepen in de samenleving.

Een nadeel van het SSB is de geringe actualiteit van een deel van de gegevens. Hoewel de meeste vraagstellingen in wetenschappelijk onderzoek meestal 'tijdloos' zijn, wordt voor beleidsdoelinden de actualiteit belangrijk gevonden. De werkwijze die wordt gevolgd bij de samenstelling van de kern van het SSB heeft tot gevolg dat de langzaamste bron de mate van actualiteit bepaalt. Op dit moment komt de kern van het SSB ongeveer anderhalf jaar na afloop van het verslagjaar beschikbaar. Een van de langzamere bronnen betreft werknemersgegevens van het UWV. Deze bron wordt vanaf 1 januari 2006 vervangen door actuele, maandelijks leveringen door de belastingdienst en UWV in de zogenaamde polisadministratie. De langzaamste bron betreft echter de gegevens voor

zelfstandigen van de belastingdienst. Als ook hier voor een oplossing wordt gevonden, mag worden verwacht dat de kern van het SSB in de toekomst sneller kan worden samengesteld.

Actuelere arbeidsmarktstatistieken worden nu al vanuit de afzonderlijke registraties en de Enquête Beroepsbevolking samengesteld. Met deze gegevenssets kan snel ingespeeld worden op de actualiteit; voor het beschrijven van de samenhang zijn zij echter beperkt. Door de breedheid aan onderwerpen en het integrale karakter is het SSB daarvoor geschikter.

Bart Bakker

Koos Arts

Leo van Toor

Centraal Bureau voor de statistiek

Noot

1. Koppelen op de combinatie van geslacht, geboortedatum en adres wordt toegepast voor de enquêtes. Deze maken geen deel uit van de kern van het SSB, maar wel van de satellieten die in paragraaf 5 worden besproken.

Bibliografie

- Arts, C.H. & Hoogteijling, E.M.J. (2002). Het Sociaal Statistisch Bestand 1998 en 1999. *Sociaal-economische maandstatistiek*, december 2002, pp. 13-21.
- Bakker, B.F.M. (2002). Statistics Netherlands' Approach to Social Statistics: The Social Statistical Dataset. *OECD Statistics Newsletter*, nr. 11, pp. 4-6.
- Bakker, B.F.M., & Arts, C.H. (2003). Dynamiek op de arbeidsmarkt; gegevens over stromen uit het Sociaal Statistisch Bestand. In B.F.M. Bakker & L. Putman (Reds.), *De virtuele volkstelling en SSB* (pp. 59-70). Amsterdam: SISWO/CBS.
- Blom, M., J.; Oudhof, R.V. Bijl & Bakker, B.F.M. (Reds.) (2005). *Verdacht van criminaliteit. Allochtonen en*

- autochtonen nader bekeken. WODC-cabier*, 2. Den Haag/Voorburg: WODC/CBS.
- CBS (2003). www.cbs.nl/nl/publicaties/publicaties/algemeen/volkstelling-2001/volkstelling-2001.htm
- CBS StatLine (2005), <http://statline.cbs.nl/statweb/table.asp?pa=70899ned&D1=0-4>
- Cruchten, J.M.J. van, Hartgers, M.I. & Schulte Nordholt, E. (2002). Loon naar opleidings- en beroepsniveau: het Loonstrucuuronderzoek 1998. *Sociaal-economische maandstatistiek*, december 2002, pp. 72-81.
- Gircour, R., M. Hartgers & Schulte Nordholt E. (2004). *The Dutch virtual census of 2001. Analysis and methodology*. Voorburg/Heerlen: CBS.
- Houbiers, M. (2004). Towards a Social Statistical Database and unified estimates at Statistics Netherlands. *Journal of Official Statistics*, 20(1) pp. 55-75.
- Kroese, A.H., Renssen, R.H. & Trijssenaar, M. (2000). Weighting or imputation: constructing a consistent set of estimates based on data from different sources, In P. Al en B.F.M. Bakker (Reds.), *Re-engineering social statistics by micro-integration of different sources. Themanummer Netherlands Official Statistics*, 15, nr. summer (pp. 23-31).
- Laan, P. van der (2000). Integrating administrative registers and household surveys. In P. Al en B.F.M. Bakker (red.), *Re-engineering social statistics by micro-integration of different sources. Themanummer Netherlands Official Statistics*, 15, nr. summer (pp. 7-15).
- Manting, D., (2003). Levensloopintegratie: een utopie?, In J. Nobel, S. Algera, M. Biemans & P. van der Laan (red.), *Gedacht en gemeten*. Voorburg/Heerlen: CBS.
- Rijn, A.S. van, A. Zorlu, R.V. Bijl & B.F.M. Bakker (2004). *De ontwikkeling van een Integratiekaart. WODC-cabier 2004-9*. Den Haag/Voorburg: WODC/CBS.
- Schmid, G. & B. Gazier (red.) (2002). *The dynamics of full employment: social integration through transitional labour markets*. Brookfield: Edward Elgar.
- Schulte Nordholt, E. (2003). Loonstrucuuronderzoek – verrijkte data door integratie. In B.F.M. Bakker & Lisa Putman (Reds.), *De virtuele volkstelling en SSB* (pp. 73-80) Amsterdam: SISWO/CBS.
- Tielens, M. (2004). Administratieve bestanden of enquêtes. De sterktes en zwaktes op een rijtje. *Over.Werk*, 14(3) pp. 92-97.