



## **Centraal Bureau voor de Statistiek**

Divisie Macro-economische statistieken en publicaties  
Sector Ontwikkeling en ondersteuning

*Postbus 4000  
2270 JM Voorburg*

---

# **Longitudinale micro-integratie van de SSB- banenbestanden voor een consistente tijdreeks van de baandynamiek**

**Robin Milot en Peter Kee**

Kennisgeving:

De in dit rapport weergegeven opvattingen zijn die van de auteurs en komen niet noodzakelijk overeen met het beleid van het Centraal Bureau voor de Statistiek.

---

*Projectnummer:*

*MOO-203875*

*BPA-nummer:*

*2005-40-MOO*

*Datum:*

*28 april 2005*

# LONGITUDINALE MICRO-INTEGRATIE VAN DE SSB- BANENBESTANDEN VOOR EEN CONSISTENTE TIJDREEKS VAN DE BAANDYNAMIEK

*Samenvatting: Op basis van de SSB-banenbestanden proberen wij de continue bruto baandynamiek van werknemersbanen te bepalen. Het gaat ons hierbij om de totale creatie en vernietiging van banen in een jaar. Wanneer dit alleen gebeurt op basis van de aanvangsdatum en einddatum van de werknemersbanen dan leidt dit tot inconsistente tijdreeksen. Door het toepassen van voorwaartse en achterwaartse longitudinale micro-integratie hebben wij de inconsistentie in de tijdreeks tot een acceptabel niveau weten terug te brengen.*

*Trefwoorden: Baandynamiek, arbeidsmarktdynamiek, stromen, AR, SSB.*

## 1. Inleiding

Tot op heden publiceert het CBS in de Arbeidsrekeningen (AR) alleen (sub-)populatie totalen over de werkgelegenheid. Hieruit zijn alleen netto veranderingen in de werkgelegenheid af te leiden. De netto veranderingen zijn slecht het topje van de ijsberg van de totale arbeidsmarktdynamiek. De totstandkoming van het Sociaal Statistisch Bestand (SSB) maakt het mogelijk om de arbeidsmarktdynamiek continu te meten. Het CBS heeft zich het doel gesteld om de Arbeidsrekeningen de komende jaren uit te breiden met gegevens over de baandynamiek en personendynamiek op de arbeidsmarkt. Met een baan bedoelen wij een arbeidsovereenkomst, en niet een arbeidsplaats.

Dit rapport laat zien op welke manier continue stromen van de baancreatie en baanvernietiging van werknemersbanen in een jaar zijn te bepalen uit de SSB-banenbestanden van werknemers. Uit onze analyse blijkt dat directe bepaling op basis van de aanvangsdatum en einddatum van de banen niet direct in een consistente tijdsreeks resulteert. Met behulp van voorwaartse en achterwaartse longitudinale micro-integratie is de inconsistentie in de tijdreeks terug te brengen tot een acceptabel niveau.

Wij zullen nu eerst in paragraaf 2 de gebruikte methoden en data bespreken. Vervolgens zullen wij in paragraaf 3 voor de verschillende methoden de resultaten presenteren en de verschillen bediscussiëren. Wij sluiten af met de conclusie in paragraaf 4.

## 2. Methoden en data.

Wij bespreken in deze paragraaf de gebruikte methoden en data. In paragraaf 2.1 beschrijven wij welke definitie wij volgen voor een werknemersbaan, en hoe een integraal SSB banenbestand met werknemersbanen binnen het SSB is samengesteld uit verschillende administratieve en statistische bronnen. Vervolgens definiëren wij in paragraaf 2.2 verschillende bruto baanstromen, en laten wij zien hoe de baanstromen van afzonderlijke jaren onderling consistent moeten zijn. In paragraaf 2.3 komen dan tot slot drie methoden naar voren om baanstromen uit de verschillende banenbestanden te bepalen.

### 2.1 Banen

#### 2.1.1 Definitie

Het Europees Systeem van Rekeningen (ESR 1995) definieert een baan als volgt:

*“Een baan is een expliciete of impliciete overeenkomst tussen een persoon en een ingezeten institutionele eenheid om gedurende een bepaalde periode of tot nader order tegen beloning werk te verrichten.”*

Onder deze definitie vallen zowel banen van werknemers als banen van zelfstandigen. In dit rapport beperken wij ons tot **werknemersbanen**, waarvan sprake is wanneer de betrokkene niet tot dezelfde institutionele eenheid behoort als de werkgever.

#### 2.1.2 Banenbestanden van werknemers in het SSB

De SSB-banenbestanden van werknemers zijn volumebestanden, dat wil zeggen dat alle waargenomen werknemersbanen over het gehele jaar zijn opgenomen. Arts en Hoogteijling (2002) geven een uitvoerige beschrijving van de opbouw van het banenbestand van werknemers. Wij volstaan hier met een korte samenvatting. Het banenbestand van werknemers is samengesteld uit vier bronnen:

1. de verzekerdenadministratie werknemers (VZA)
2. de enquête werkgelegenheid en lonen (EWL)
3. het bestand met voorheffingen voor de loonbelasting (Fibase)
4. de gekoppelde gemeentelijke basis administraties (GBA)

De eerste drie bronnen bevatten dienstverbanden tussen personen en bedrijven. De EWL bevat de statistische bedrijfseenheid. De administratieve eenheden van bedrijven in de Fibase (loonbelastingnummer) en VZA (bedrijfsverenigings-aansluitnummer) worden met een aantal bewerkingen omgezet naar de statistische bedrijfseenheid. Door het correct verbinden van overeenkomstige records in de eerste drie bronnen worden statistische banenrecords afgeleid.

De eerste drie bronnen bevatten ook allen gegevens over de aanvangsdatum en einddatum van het dienstverband. Aan de hand van de aanvangsdata en einddata uit

de records afkomstig van de drie bronnen en andere informatie uit de verschillende bronnen worden een aanvangsdatum en einddatum afgeleid die beter aansluiten bij het betalingsconcept. Alleen indien banenrecords koppelen met banenrecords uit andere jaarbestanden dan wordt informatie uit eerdere jaren over de aanvangsdatum meegenomen in de vaststelling van de aanvangsdatum. Op dit moment zijn er banenbestanden beschikbaar voor de jaren 1999 tot en met 2002. Voor de bepaling van de baandynamiek hebben wij gebruik gemaakt van de tweede versie van de banenbestanden van werknemers. De banenrecords met nabetalingen hebben wij in de analyse weggelaten.

## 2.2 Baanstromen

Met de banenbestanden uit het SSB hebben wij voor elk jaar een integraal kader met werknemersbanen. Met behulp van de begin- en einddatum van de banen kunnen wij de baanstromen afleiden. Wij maken daarbij onderscheid tussen eerste en tweede orde baanstromen.

### 2.2.1 Eerste orde baanstromen: creatie en vernietiging van banen

Bij de eerste orde baanstromen classificeren wij de baanstromen onafhankelijk van elkaar naar de creatie of vernietiging van de baan:

- De **bruto instroom (gross inflow, GI)** van banen in de populatie van banen tijdens het referentiejaar wordt veroorzaakt door **creatie** van banen tijdens het referentiejaar. Dit zijn de banen met een aanvangsdatum binnen het referentiejaar.
- De **bruto uitstroom (gross outflow, GO)** van banen in de populatie van banen tijdens het referentiejaar wordt veroorzaakt door **vernietiging** van banen tijdens het referentiejaar. Dit zijn de banen met een einddatum binnen het referentiejaar.

Het absolute aantal banen in de eerste orde bruto instroom en uitstroom zullen wij aanduiden met de symbolen  $F_{GI}$  en  $F_{GO}$ .

### 2.2.2 Tweede orde baanstromen: combinatie van creatie en vernietiging van banen

Bij de tweede orde baanstromen combineren wij de classificaties voor de creatie en vernietiging. Dit leidt tot vier mogelijke baanstromen:

1. **Zonder stroom (without flow, WF):** Geen creatie en geen vernietiging van de baan.
2. **Alleen instroom (single inflow, SI):** *Creatie*, maar geen vernietiging van de baan.
3. **Alleen uitstroom (single outflow, SO):** *Vernietiging*, maar geen creatie van de baan.
4. **Dubbele stroom (double flow, DF):** Zowel *creatie* als wel *vernietiging* van de baan.

Het absolute aantal banen in de tweede orde bruto baanstromen worden aangeduid met de symbolen  $F_{WF}$ ,  $F_{SI}$ ,  $F_{SO}$ , en  $F_{DF}$ .

### 2.2.3 Relaties tussen stromen in een jaar

Tussen de eerste en tweede orde baanstromen gelden rekenkundige verbanden. De eerste orde bruto instroom is gelijk aan de banen met alleen instroom plus de banen met dubbele stroom:

$$F_{GI} = F_{SI} + F_{DF} .$$

De eerste orde bruto uitstroom is dan dus als volgt:

$$F_{GO} = F_{SO} + F_{DF} .$$

De totale absolute hoeveelheid banen  $J$  in een jaar is daarnaast ook altijd gelijk aan de som van alle tweede orde baanstromen:

$$J = F_{WF} + F_{SI} + F_{SO} + F_{DF} .$$

De grootte van de **netto stroom**  $\Delta F$  kan zowel berekend worden uit de eerste orde bruto stromen als uit de tweede orde enkelvoudige stromen door

$$\Delta F = F_{GI} - F_{GO} = F_{SI} - F_{SO} ,$$

Zodat een positieve waarde van  $\Delta F$  betekent dat er een netto instroom van banen is, en een negatieve waarde van  $\Delta F$  dat er een netto uitstroom van banen is gedurende het referentiejaar. De grootte van de dubbele stroom heeft dus geen enkel effect op de netto stroom voor het referentiejaar.

### 2.2.4 Donatie en ontvangst van banen tussen jaren

Arbeidscontracten tussen werkgevers en werknemers kunnen meerdere jaren doorlopen, zodat de banen van werknemers bij een bedrijf in meerdere jaarbestanden teruggevonden moeten kunnen worden. De banen die doorlopen in het volgende jaar worden als het ware gedoneerd door jaar  $T$  en ontvangen in jaar  $T+1$ . De totale donatie  $D(T)$  van banen uit een jaar moet gelijk zijn aan de totale ontvangst  $R(T+1)$  in het volgende jaar. De donatie en ontvangst kunnen berekend worden uit de tweede orde stromen. De totale donatie is gelijk aan alle banen die *niet* uitstromen tijdens het jaar:

$$D = F_{WF} + F_{SI} .$$

De totale ontvangst is dan gelijk aan alle banen die *niet* instromen tijdens het jaar:

$$R = F_{WF} + F_{SO} .$$

Bij het vergelijken van de stroomgegevens van de banenbestanden voor de verschillende jaren kwamen wij er achter dat de donatie en ontvangst op macroniveau niet consistent zijn. De inconsistentie  $I$  in jaar  $T$  berekenen we door:

$$I(T) = D(T-1) - R(T) .$$

Een positieve  $I(T)$  duidt dus op een donatie-overschot, en een negatieve  $I(T)$  op een donatietekort.

## 2.3 Methoden voor instroom en uitstroom toekenning

De inconsistenties tussen donatie en ontvangst zijn door longitudinale micro-integratie te reduceren. Om de effecten van de correcties te tonen zullen wij drie methoden beschrijven om de instroom en uitstroom te bepalen uit de banenbestanden.

### 2.3.1 Zonder longitudinale micro-integratie

Bij de methode zonder longitudinale micro-integratie nemen wij aan dat de aanvangsdata en einddata in de afzonderlijke banenbestanden zonder meer waar zijn. Een baan wordt als instroom gezien indien aanvangsdatum binnen de jaargrenzen ligt, en als uitstroom indien de einddatum in het jaar valt. Alle banen met een onbekende einddatum worden in dit geval als doorlopend en *nooit* als uitstromend beschouwd.

### 2.3.2 Voorwaartse longitudinale micro-integratie

Administraties kunnen vertraging hebben in de verwerking, waardoor banen nog een onbekende einddatum hebben terwijl deze feitelijk al beëindigd zijn. Door voorwaartse longitudinale micro-integratie kunnen wij deze banen alsnog afsluiten (op 31 december) en als uitstroom classificeren.

Aan de banen (met persoonsnummer en bedrijfseenheidnummer) van jaar T worden daarvoor de overeenkomstige banen (met persoonsnummer en volgtijdelijk vergelijkbare bedrijfseenheidnummer) van jaar T+1 op microniveau gekoppeld. Indien een baan van jaar T geen koppeling heeft gelegd met een baan in jaar T+1, dan kan aangenomen worden dat de baan niet meer bestaat. De niet-gekoppelde banen uit jaar T met een onbekende einddatum worden dan alsnog beëindigd in het jaar T en als uitstroom geëvalueerd.

### 2.3.3 Achterwaartse longitudinale micro-integratie

Administraties kunnen ook vertraging hebben in de verwerking van nieuwe banen. Zo kunnen banen met een aanvangsdatum in jaar T-1 pas voor het eerst voorkomen in de bestanden van jaar T. Deze banen zijn dan niet waargenomen als een enkelvoudige instroom in jaar T-1 en tellen hierdoor ook niet mee in de donatie. In jaar T worden deze banen echter wel waargenomen en gerekend tot de ontvangst. Door achterwaartse longitudinale micro-integratie kunnen wij deze banen alsnog kunstmatig laten instromen in jaar T, waardoor het donatietekort wordt verminderd.

Aan de banen (met persoonsnummer en volgtijdelijk vergelijkbare bedrijfseenheidnummer) van jaar T worden daarvoor de overeenkomstige banen (met persoonsnummer en bedrijfseenheidnummer) van jaar T-1 op microniveau gekoppeld. Indien een baan van jaar T geen koppeling heeft gelegd met een baan in jaar T-1, dan wordt de baan in jaar T als instroom geëvalueerd.

### 3. Resultaten en discussie

Om de effecten van de longitudinale micro-integratie op de uitkomsten van de stromen te laten zien, zullen wij voor de drie methodes de uitkomsten laten zien. Wij laten eerst de uitkomsten zien zonder longitudinale micro-integratie, vervolgens met voorwaartse longitudinale micro-integratie, en als laatste met zowel voorwaartse als achterwaartse micro-integratie.

#### 3.1 Zonder longitudinale micro-integratie

In Tabel 1 staan de tweede orde stroomgegevens zonder longitudinale micro-integratie, dus alleen gebaseerd op de aanvangsdatum en einddatum van de banen in de banenbestanden voor de afzonderlijke jaren. Volgens deze methode neemt het aantal banen zonder stroom  $F_{WF}$  elk jaar toe. Het aantal banen met dubbele stroom  $F_{DF}$  en alleen instroom  $F_{SI}$  neemt voor deze tijdsperiode af. Het aantal banen met alleen uitstroom  $F_{SO}$  neemt in 2000 nog toe en daarna ook af, waardoor de netto stroom  $\Delta F$  in 2000 sterk afneemt en vervolgens ongeveer gelijk blijft.

*Tabel 1: Stromen zonder longitudinale micro-integratie*

---

	$F_{WF}$	$F_{SI}$	$F_{SO}$	$F_{DF}$	$\Delta F$
1999	5.067.081	1.930.154	1.176.090	2.416.316	754.064
2000	5.301.578	1.837.059	1.378.811	2.332.598	458.248
2001	5.491.398	1.801.806	1.340.808	2.198.147	460.998
2002	5.737.396	1.724.774	1.283.854	1.933.748	440.920

---

In Tabel 2 staan de ontvangst, donatie en inconsistentie voor de methode zonder longitudinale micro-integratie. Hierin is een inconsistentie  $I(T)$  van rond de 300 duizend banen tussen de donatie  $D(T-I)$  en ontvangst  $R(T)$  te zien. Dit houdt in dat er meer banen worden gedoneerd in T-I, dan er in het jaar T worden ontvangen. Meest voor de hand liggende verklaring hiervoor is dat er te weinig banen administratief worden afgesloten. Gerelateerd aan de ontvangst is dat een relatieve inconsistentie van ongeveer 5 procent.

*Tabel 2: Ontvangst en donatie zonder longitudinale micro-integratie*

---

	$R$	$D$	$I$
1999	6.243.171	6.997.235	
2000	6.680.389	7.138.637	316.846
2001	6.832.206	7.293.204	306.431
2002	7.021.250	7.462.170	271.954

---

### 3.2 Met voorwaartse longitudinale micro-integratie

Om het aantal ten onrechte niet beëindigde banen in het jaar terug te dringen hebben wij vervolgens voorwaartse longitudinale micro-integratie toegepast. In Tabel 3 staan hiervoor de resultaten voor de tweede orde stroomgegevens. Door de voorwaartse integratie gaat het laatste jaar van de tijdreeks verloren.

*Tabel 3: Stromen met voorwaartse micro-integratie*

	$F_{WF}$	$F_{SI}$	$F_{SO}$	$F_{DF}$	$\Delta F$
1999	4.890.262	1.714.412	1.352.909	2.632.058	361.503
2000	5.101.007	1.644.795	1.579.382	2.524.862	65.413
2001	5.313.393	1.618.118	1.518.813	2.381.835	99.305
2002	.	.	.	.	.

Ten opzichte van de methode zonder micro-integratie is het aantal banen zonder stroom  $F_{WF}$  en met alleen instroom  $F_{SI}$  afgenomen, en het aantal banen met alleen uitstroom  $F_{SO}$  en dubbele stroom  $F_{DF}$  toegenomen. Het extra beëindigen van banen leidt namelijk tot een verplaatsing van banen zonder stroom naar banen met alleen uitstroom en van banen met alleen instroom naar banen met een dubbele stroom. Door deze verplaatsingen is de netto stroom  $\Delta F$  afgenomen in vergelijking met de methode zonder voorwaartse micro-integratie.

Uit Tabel 4 blijkt dat de positieve inconsistentie  $I$  is omgezet naar een negatieve inconsistentie. De negatieve inconsistentie is in relatieve zin wel terug gebracht tot bijna 1 procent van de ontvangst-stroom.

*Tabel 4: Ontvangst en donatie met voorwaartse micro-integratie*

	$R$	$D$	$I$
1999	6.243.171	6.604.674	
2000	6.680.389	6.745.802	-75.715
2001	6.832.206	6.931.511	-86.404
2002	7.021.250	.	-89.739

Hoewel het aantal banen zonder stroom en met alleen uitstroom niet te bepalen is voor 2002, is de som hiervan (de ontvangst  $R$ ) wel te bepalen. Dit komt doordat bij de voorwaartse micro-integratie alleen banen tussen deze beide stroomgrootheden worden verplaatst. In vergelijking met de methode zonder micro-integratie wordt dus alleen de donatie gecorrigeerd.



### 3.3 Met voorwaartse en achterwaartse longitudinale micro-integratie

Om de negatieve inconsistentie te beperken hebben wij vervolgens boven op de voorwaartse longitudinale micro-integratie ook nog achterwaartse longitudinale micro-integratie toegepast. In Tabel 5 staan hiervan de resultaten voor de tweede orde stromen. Zowel het eerste als het laatste jaar van de tijdreeks is door toepassing van beide methoden onbruikbaar geworden om de stromen voor dat jaar te bepalen.

*Tabel 5: Stromen met voorwaartse en achterwaartse micro-integratie*

	$F_{WF}$	$F_{SI}$	$F_{SO}$	$F_{DF}$	$\Delta F$
1999	.	.	.	.	.
2000	5.070.505	1.675.297	1.549.554	2.554.690	125.743
2001	5.273.267	1.658.244	1.485.873	2.414.775	172.371
2002	.	.	.	.	.

In vergelijking met alleen voorwaartse micro-integratie is de netto stroom  $\Delta F$  weer iets toegenomen door de achterwaartse longitudinale micro-integratie. Dit wordt veroorzaakt door het verplaatsen van banen zonder stroom naar banen met alleen instroom, en van banen met alleen uitstroom naar banen met dubbele stroom.

Hoewel de tweede orde stromen niet bepaald kunnen worden voor 1999 en 2002, kunnen wel de donatie  $D$  voor 1999 en de ontvangst  $R$  van 2002 worden berekend. Uit Tabel 6 blijkt dat door toepassing van zowel voorwaartse als achterwaartse longitudinale micro-integratie de inconsistentie  $I$  tussen jaren nog niet volledig is opgeheven. Een negatieve inconsistentie van gemiddeld 17 duizend blijft nog onopgelost. Gerelateerd aan de ontvangst is de inconsistentie nu afgenomen tot 0,3 procent. De relatieve fout is dus met een factor 16 teruggebracht ten opzichte van de methode zonder longitudinale micro-integratie. In verhouding tot de enkelvoudige stromen ligt de inconsistentie voor de jaren 2000 en 2001 in de buurt van de 1 procent. Om de omvang en de verandering van de baandynamiek te bepalen, zijn de voorwaartse en achterwaartse longitudinale micro-integratie nauwkeurig genoeg. Wel dient bij analyses op mesoniveau met de inconsistentie rekening gehouden te worden, omdat administratieve fouten geconcentreerd kunnen zijn in enkele bedrijfstakken.

*Tabel 6: Ontvangst en donatie met voorwaartse en achterwaartse micro-integratie*

	$R$	$D$	$I$
1999	.	6.604.674	.
2000	6.620.059	6.745.802	-15.385
2001	6.759.140	6.931.511	-13.338
2002	6.954.410	.	-22.899

Absoluut gezien heeft de inconsistentie nog wel enige vertekening tot gevolg op de netto instroom. Dit komt doordat een negatieve inconsistentie duidt op een tekort aan instroom. Op de netto stroom vindt er een vertekening plaats door de kunstmatige instroom in het jaar T, terwijl volgens de aanvangsdata instroom in het jaar T-1 plaats had moeten vinden. Voor de absolute waarde van de instroom over de jaren geeft het echter weinig vertekening.

In principe is het mogelijk om op basis van de gegevens uit het banenbestand van jaar T extra banen met instroom toe te voegen aan het banenbestand van jaar T-1. Wij doen dit op dit moment echter niet, omdat er dan inconsistenties gaan ontstaan voor het gemiddeld aantal banen en het aantal banen op peildata tussen de originele banenbestanden en onze nieuwe longitudinaal geïntegreerde banenbestanden. Het is eventueel wel mogelijk dat dit in de toekomst een onderdeel kan gaan worden van het revisieproces van de SSB-banenbestanden.

#### **4. Conclusie**

Door de methoden van voorwaartse en achterwaartse longitudinale micro-integratie van de SSB-banenbestanden is het mogelijk redelijk consistente tijdreeksen van de baandynamiek te maken.

#### **Referenties**

Arts, C.H. en E.M.J. Hoogteijling (2002), Het sociaal statistisch bestand 1998 en 1999, *Sociaal-economische maandstatistiek*, 2002/12, Centraal Bureau voor de Statistiek, Voorburg/Heerlen, 13-21.

ESR 1995, *Europees systeem van rekeningen 1995*, EUROSTAT.