
Notitie Replicatie verdeelmodel Participatiewet 2015

datum	28 oktober 2015
aan	SCP/SZW
van	SEO
Rapportnummer	2015-77

Inleiding

Het ministerie van SZW heeft behoefte aan een replicatie van het multiniveauverdeelmodel dat is gebruikt om de middelen voor de uitkeringen in het kader van de Participatiewet voor 2015 over gemeenten te verdelen. Doel van deze replicatie is om te achterhalen of de toegepaste verdeling exact nagebootst kan worden. Het SCP heeft SEO Economisch Onderzoek gevraagd deze replicatie uit te voeren. Het gaat hier niet om contraexpertise door een derde partij, maar een onderzoek met als doel om op een transparante manier aspecten te verduidelijken die het SCP en/of anderen hebben geconstateerd.

Tijdens deze replicatie is een aantal verbeter-suggesties naar voren gekomen. De verdeelmodellen voor zowel 2016 als 2015 zijn naar aanleiding van deze suggesties verbeterd. Met deze verbeter-suggesties zijn beide verdeelmodellen nu exact te repliceren. SEO is betrokken bij de berekeningen van het model. Deze notitie bespreekt deze replicatie en verbeter-suggesties.

Het multiniveauverdeelmodel bestaat uit drie onderdelen:

1. Aanmaken van het analysebestand
2. Schatten van het verklaringsmodel
3. Toepassen van het verdeelmodel

Deze notitie gaat in op de eerste twee onderdelen: het maken van het analysebestand en het schatten van het verklaringsmodel.

Aanmaken van het analysebestand

De analysebestanden voor de verdeling van de middelen voor 2015 en 2016 zijn gebaseerd op drie jaargangen van de Enquête Beroepsbevolking (EBB). Voor 2015 is gebruik gemaakt van de jaren 2010, 2011 en 2012 (voor 2016: 2011, 2012 en 2013). Aan deze bestanden worden allerlei aanvullende gegevens gekoppeld zoals uitkeringsafhankelijkheid, gemeentekennmerken, buurtkennmerken, huishoudinkomens, enzovoorts. Er wordt eerst per jaar een analysebestand gemaakt en vervolgens worden deze samengevoegd om tot één analysebestand te komen. Uit de replicatie komen de volgende dingen naar voren:

- Repliceerbaarheid van het analysebestand
De huishoudkenmerken in het analysebestand zijn vrijwel exact te repliceren. De persoonskenmerken kunnen niet exact gerepliceerd worden. Dit komt doordat de gegevens op verschillende plaatsen gesorteerd worden, waarbij meerdere records soms dezelfde waarden hebben op sorteerkennmerken. Bijvoorbeeld als op persoon en enquêtedag wordt gesorteerd en een zelfde persoon meerdere keren op die dag is bevraagd, of als op huishoudnummer wordt gesorteerd en een huishouden uit meerdere personen bestaat. De statistische programmatuur (STATA) past dan een random sortering toe, wat betekent dat de ene keer de ene persoon uit het huishouden bovenaan staat en de andere keer de andere. Als vervolgens de eerste persoon uit dat huishouden of op die enquêtedag wordt geselecteerd, kan dit steeds een andere selectie opleveren. Deze random selectie heeft vooral gevolgen op persoonsniveau, niet op huishoudniveau: de huishoudkenmerken zijn immers voor iedere persoon in het huishouden gelijk. Omdat het model op huishoudniveau wordt geschat is het effect van deze random selectie op de modeluitkomsten verwaarloosbaar.¹ Toch is het voor de repliceerbaarheid van belang dat steeds exact hetzelfde analysebestand wordt aangemaakt. Om deze reden is de programmatuur aangepast.
- Inconsistenties in samenstelling van het huishouden
In de EBB wordt de huishoudsamenstelling geregistreerd. In de databron zit voor dit kenmerk een aantal inconsistenties als ook gekeken wordt naar de andere respondenten in het huishouden. Er zijn bijvoorbeeld alleenstaanden met kinderen die als alleenstaanden met overigen genoteerd staan, of mensen die als paar genoteerd staan, maar bij wie geen gegevens over een ander huishoudlid aanwezig zijn (alle huishoudleden zitten in principe in de EBB). Hiernaast kon in de vorige modeltoepassing een eenouderhuishouden ook een alleenstaande ouder met een ander zijn, bijvoorbeeld oma of opa. Deze inwonende andere kon de schattingsresultaten enigszins verstoren, bijvoorbeeld als deze persoon hoogopgeleid was en de alleenstaande ouder niet. Een deel van deze inconsistenties zijn in de vorige modeltoepassing in een laat stadium (bij toepassing van onderdeel 3) gecorrigeerd. Er is nu voor gekozen om deze inconsistenties in een vroeg stadium te identificeren en buiten de analyse te laten.
Bij het uitsplitsen van eenouderhuishoudens naar geslacht, is een fout gemaakt waardoor soms het geslacht van het kind is overgenomen. Een alleenstaande moeder met twee zoons kon daardoor onterecht aan de groep eenouder-vader worden toegekend. Ook dit is gecorrigeerd.
- Coderen ontbrekende waarden van kenmerken
Er is bij de vorige toepassing van het verdeelmodel voor gekozen om ontbrekende waarden toe te kennen aan een categorie. Zo zijn bijvoorbeeld de huishoudens met een ontbrekende waarde bij het hebben van een koop/huurwoning toegekend aan de groep huurwoning en bij etniciteit zijn de onbekende etniciteiten tot de groep niet-westerse allochtonen gerekend. Het gaat hier om kleine aantallen, toch is het beter deze ontbrekende waarden niet in te vullen. Het is immers niet bekend of het een koop/huurwoning of (niet-westerse) allochtoon of autochtoon betreft.
- Aanmaken te verklaren variabele (bijstand)
Om het model te kunnen toepassen moet een kenmerk bijstandsontvanger (ja/nee) worden aangemaakt. Dit kenmerk bestaat niet alleen uit de huishoudens die op dit moment bijstand gebruiken, maar ook uit personen die door invoering van de Participatiewet niet meer in de

¹ In de praktijk blijkt dat in de EBB binnen een huishouden soms het adres van een huishoudlid ontbreekt. Als precies dat huishoudlid wordt geselecteerd, kunnen voor dat huishouden geen aanvullende gemeente-, wijk- en buurtgegevens gekoppeld worden.

Wsw of Wajong kunnen stromen en daardoor een beroep moeten doen op bijstand, de zogenaamde nieuwe doelgroep. Bij het aanmaken van het kenmerk bijstandsontvanger zijn de mensen met een IOAZ-uitkering onterecht niet meegeteld. Het gaat hier om zeer kleine aantallen. Bij de nieuwe doelgroep zijn een tweetal groepen onterecht wel meegeteld: in 2012 zijn per abuis de volledig arbeidsongeschikte Wajongers ook meegeteld en bij de Wsw is niet gekeken naar de maand waarin de Wsw is gestart. Iedereen die in 2008 in de Wsw is gestroomd, telt daardoor in 2012 mee als “afgelopen vier jaar ingestroomd”. Als iemand in januari 2008 in de Wsw is gestroomd en in november 2012 geënquêteerd, is dit langer dan vier jaar geleden. Het kenmerk bijstandsontvanger is opnieuw berekend, waarbij bovenstaande punten zijn gecorrigeerd.

- **Keuze ten aanzien van gebruik arbeidshandicap**
Ervaren gezondheid, blijkend uit het al dan niet hebben van een arbeidshandicap, lijkt een belangrijke factor bij het risico op bijstand. Dit kenmerk is gebaseerd op waarnemingen uit de EBB. Er is echter slechts voor een beperkt deel (ongeveer 20% per jaargang) van de respondenten bekend of er sprake is van een arbeidshandicap. Daarom is de waarneming van een persoon op een bepaald tijdstip ingevuld voor alle meetmomenten van die persoon. Stel dat een persoon in 2012 aangeeft een arbeidshandicap te ervaren, dan wordt verondersteld dat deze persoon op eerdere en latere meetmomenten ook een arbeidshandicap had. De vraag rijst of het verstandig is een dergelijk kenmerk, met zoveel ontbrekende waarden, op te nemen. Vanwege de sterke verklaringskracht en het gebrek aan een alternatief is ervoor gekozen dit kenmerk te handhaven.
- **Kleine onvolkomenheden**
Bij deze replicatie kwamen ook een aantal kleine onvolkomenheden aan het licht, die gevolgen hadden voor een beperkt aantal waarnemingen. Het gaat hier bijvoorbeeld om het gebruik van een verkeerd jaar in het adressenbestand, de keuze om enquêtedag op de 15^e te zetten terwijl de werkelijke dag bekend is en het invullen van ontbrekende gemeentecodes met codes uit 2010 die in 2011 of 2012 niet meer voorkwamen vanwege herindelingen. Deze zijn vervolgens aangepast.

Schatten van het verklaringsmodel

Het multiniveaumodel 2015 is een zogenaamd “random intercept model”. Dit betekent dat het effect van de verklarende variabelen op de kans op bijstand voor iedereen hetzelfde is, maar dat het gemiddelde bijstandsniveau kan variëren tussen de verschillende niveaus. In de ene gemeente (buurt, wijk, etc.) kan de kans op bijstand dus structureel hoger liggen dan in de andere gemeente, maar het effect van bijvoorbeeld etniciteit op de kans op bijstand is overal gelijk. Voordat het verklaringsmodel wordt geschat, worden de EBB-analysebestanden verder bewerkt tot een definitief analysebestand. Uit de replicatie komen de volgende dingen naar voren:

- **Repliceerbaarheid van de modelschattingen**
Voordat het verklaringsmodel geschat wordt, wordt een selectie getrokken uit de nieuwe doelgroep. Hiermee wordt een inschatting gemaakt van de nieuwe instroom in de bijstand als gevolg van de nieuwe doelgroep. Het is de bedoeling om een random selectie uit deze groep mensen te trekken, het is immers nu niet bekend wie daadwerkelijk onder de Participatiewet zou komen te vallen. Doordat deze selectie iedere keer volledig ‘at random’ wordt gedaan, wordt er steeds

een andere set van mensen geselecteerd als nieuwe doelgroep. Dit resulteert in iets andere modelschattingen. Omdat het de voorkeur heeft dat het model iedere keer exact dezelfde modelschattingen oplevert, is ervoor gekozen om deze random selectie te handhaven, maar ervoor te zorgen dat in iedere toepassing van het model dezelfde selectie gedaan wordt. Er wordt nu dus maar eenmaal een random selectie gedaan. Het gaat om kleine aantallen: de effecten op de modelschattingen zijn minimaal.

- Aansluiting tussen niveaus in het model en in de data
Het multiniveaumodel bestaat uit vier verschillende niveaus, namelijk huishoudniveau, wijkniveau, intergemeentelijke sociale dienst (ISD)-niveau en COROP-niveau. De niveaus in het model zijn daardoor niet in alle gevallen gelijk aan de niveaus van de verklarende variabelen (wijk en ISD versus buurt en gemeente). Het lijkt logischer om de niveaus op elkaar te laten aansluiten. Dit zijn immers de niveaus waarvoor een andere bijstandskans verwacht wordt. De afwijking is ontstaan door keuzes die in de ontwikkelingsfase van het model gemaakt zijn. Zo is er voor wijk- in plaats van buurtniveau gekozen omdat het verdeelmodel (onderdeel 3) anders niet convergeerde. De keuze voor ISD is op advies van een inhoudelijk expert gedaan, omdat het beleid van een gemeente op ISD-niveau wordt bepaald. Herschatting van het model met andere niveaus (buurt en gemeente) geeft soortgelijke modelschattingen als het huidige model.
- Inconsistenties in gebruikte indelingen in de data
De gebruikte gegevens voor de schattingen zijn in indelingen in een bepaald jaar. Gegevens over leefbaarheid zijn in de buurtindeling van 2011 en gemeente- en COROP-gegevens in de indeling van 2012. Dit resulteert in onjuistheden als er sprake is geweest van herindelingen. In 2012 is de gemeente Hollands Kroon bijvoorbeeld gevormd uit Anna Paulowna, Niedorp, Wieringen en Wieringermeer. Aan een persoon die in 2010 in Wieringermeer woonde, zijn nu gegevens over de kansen voor hoogopgeleiden in 2010 gehangen voor de gemeente Hollands Kroon (dus al deze vier gemeenten tezamen), terwijl er op dat moment nog sprake was van individuele gemeenten. Bij de COROP-gegevens is dit geen probleem omdat er geen verschuivingen in de COROP-indeling zijn geweest in deze jaren. Het effect hiervan is naar verwachting niet groot, omdat er sprake is van een beperkt aantal gemeentelijke en buurtherindelingen. Bovendien sluiten de gegevens in de heringedeelde gemeenten (of buurten) naar verwachting aan bij de niet-heringedeelde gemeenten (of buurten). De gebruikte brongegevens zijn op dit moment niet beschikbaar in andere indelingen, het is daarom niet mogelijk om de gegevens aan te passen aan de indeling van het betreffende jaar.
- Onjuiste codering van niet-westerse allochtonen
In het model is bij etniciteit onderscheid gemaakt naar kernleden van een huishouden en niet-kernleden. De kernleden van een huishouden zijn bijvoorbeeld de alleenstaande, de alleenstaande ouder, de leden van een paar en inwonende kinderen. Niet-kernleden zijn inwonende grootouders, kleinkinderen, schoonfamilieleden of inwonende mensen die geen familie zijn. Bij mensen die geen kernlid zijn is de etniciteit op ontbrekend gezet. Door een fout in de codering zijn deze ontbrekende waarden vervolgens bij het aanmaken van de indicator niet-westerse allochtoon ja/nee als allochtoon meegerekend. Deze fout is nu hersteld.
- Inconsistenties in de lijst met Intergemeentelijke sociale diensten
Er is voor gekozen om in het model het ISD-niveau te gebruiken. De ISD-indeling is gebaseerd op een lijst aangeleverd door COELO. Vergelijking van deze lijst met de regioatlas van het ministerie van BZK en gegevens op de internetsites van de verschillende gemeenten bracht een

aantal onjuistheden aan het licht. Bij herberekening van de verdeling voor 2015 en 2016 is deze lijst geactualiseerd en aangepast.

- **Coderen ontbrekende waarden van kenmerken**
Ook in dit onderdeel is er bij het aanmaken van kenmerken voor gekozen om ontbrekende waarden toe te kennen aan een categorie, bijvoorbeeld bij leefbaarheid in de buurt. Er is in deze verbeterslag voor gekozen om deze ontbrekende waarden niet in te vullen.
- **Kleine onvolkomenheden**
Ook in de programmatuur voor het verklaringsmodel zaten een aantal kleine onvolkomenheden, die vervolgens zijn aangepast. Het gaat hier bijvoorbeeld om een fout in de hercodering van één gemeente, het feit dat de nieuwe doelgroep variabele op persoons- en niet op huishoudniveau is berekend en het toevoegen van de nieuwe doelgroep niet alleen voor de eerste EBB-waarneming gebeurt.

Conclusie

De huishoudkenmerken in het analysebestand waarop het model geschat wordt, zijn vrijwel exact te repliceren. De modelschattingen verschillen bij iedere toepassing van het model enigszins omdat een random selectie uit de nieuwe doelgroep getrokken wordt. Dit is geen fout: het is immers de bedoeling om deze groep per toeval te kiezen. Het heeft echter de voorkeur om iedere keer op exact dezelfde modelschattingen uit te komen en dus deze random selectie slechts eenmaal uit te voeren en vervolgens te handhaven.

Tijdens deze replicatie is een aantal verbetersuggesties naar voren gekomen. Het verdeelmodel 2015 en het verdeelmodel 2016 zijn naar aanleiding van deze suggesties verbeterd. Beide verdeelmodellen zijn met deze verbeteringen nu exact te repliceren. Het uiteindelijke effect van deze verbeteringen is beperkt. Voor 2015 gaat het om een verschuiving van ruim € 26 miljoen (0,5 % van het beschikbare macrobudget).²

² Deze verschuiving in de verdeling wijkt af van de verschuiving genoemd in de brief van staatssecretaris van Sociale Zaken en Werkgelegenheid Jetta Klijnsma aan de Tweede Kamer (referentie 2015-0000278952) aangezien het hier gaat om het effect van de verbetersuggesties die naar voren zijn gekomen uit deze replicatie. De twee correcties die eerder dit jaar zijn toegezegd door de staatssecretaris zijn hier buiten beschouwing gelaten.