

Reactie op “Verdeelmodel Participatiewet”, versie juni 2016 door SEO.

Door Bas van der Klaauw (Vrije Universiteit)

De afgelopen periode heb ik in de expertgroep gezeten die met SEO heeft meegekeken bij de doorontwikkeling van het verdeelmodel voor de bijstand. Ik ben door het Ministerie van Sociale Zaken gevraagd een schriftelijke reactie te geven op het model zoals dat er begin juni 2016 lag.

Het verdeelmodel dat SEO gemaakt heeft is een verbetering ten opzichte van het verdeelmodel dat eerder door SCP gemaakt was. Het SCP verdeelmodel had een aantal tekortkoming (zie mijn eerder notitie over het SCP model voor een uitleg):

1. Het gebruik van enquêtegegevens uit de EBB.
2. Implementatie van het multilevel logit model.
3. Keuze van verklarende variabelen in het verdeelmodel.

In het nieuwe verdeelmodel van SEO is hierin een verbeteringsslag gemaakt. Er wordt nu gebruik gemaakt van integrale data waardoor non-respons bij de EBB geen probleem meer is en het omzetten van geschatte bijstandskansen naar gemeentebudgetten logisch volgt uit het model. Bij deze laatste stap worden ook geen vereenvoudigende veronderstellingen gemaakt omtrent de random effecten die in het SCP model op verschillende niveaus toegevoegd waren. Deze random effecten zitten nu niet meer in het model waardoor ook het schatten van het model veel makkelijker is. Ten slotte, is nu een procedure gevolgd waarbij veel zorgvuldiger gekeken is naar welke verklarende variabelen opgenomen zijn in het verdeelmodel.

Vanwege de bovenstaande aanpassing ben ik van mening dat het verdeelmodel van SEO een grote verbetering is ten opzichte van het eerdere model van SCP en dat dit model ook voor komend jaar al bruikbaar is. Maar bij het rapport zoals het er nu ligt, wil ik drie opmerkingen maken, waar naar moet worden gekeken bij de verdere ontwikkeling voor latere jaren:

1. Evalueer het model op meerdere peildata dan alleen 5 januari van het jaar.
2. Aanpassing van de prestatie maatstaf waarmee het model geëvalueerd wordt.
3. Bekijk verschillen tussen gemeenten nader.

Keuze voor Logit model

In het rapport is de afweging gemaakt tussen een Logit model en een lineair model. Een lineair model heeft als voordeel dat de geschatte coëfficiënten eenvoudig te interpreteren zijn, maar heeft als nadeel dat geschatte kansen negatief kunnen worden. Bij het Logit model kan dat niet, maar moet een peildatum gekozen worden waarop bepaald wordt of een huishouden wel (1) of niet (0) in de bijstand zit. Bij een lineair model kan part-time bijstand of tijdelijke bijstand eenvoudiger meegenomen worden. SEO heeft gekozen voor een Logit model, omdat er bij een lineair model (te) vaak een negatieve voorspelde bijstandskans is. Dat is een valide argument.

Voor het Logit model is gekozen voor een peildatum van 5 januari. Op dat moment ontvangen gemiddeld per gemeente 986 huishoudens bijstand. Over een heel jaar ligt dat met gemiddeld 1013 iets hoger. Een mogelijke verbetering is om met meerdere peildata te werken, bijvoorbeeld de 5^e van elke maand. Op die manier kan het model omgaan met seizoensfluctuaties in het aantal bijstandsontvangers. Het nadeel is dat het aantal datapunten dat geanalyseerd moet worden (veel) groter wordt. Een alternatief zou kunnen zijn om voor elke huishouden willekeurig een peildatum te kiezen, of om voor elke kalendermaand een apart model te schatten. Ook op die

manier benadert het model het jaargemiddelde beter en wordt rekening gehouden met seizoensfluctuaties.

Prestatiemaat

In het rapport wordt het Logit model beoordeeld met behulp van de GGAA, wat kijkt in hoeverre het model voor elke gemeente de bijstandspopulatie op 5 januari goed kan voorspellen. Terwijl het lineaire model beoordeeld wordt op verklaring van het aantal bijstandsjaren (gedurende het jaar) in de gemeente. In het algemeen is het geen goed idee om bij modelselectie een verschillende maatstaf te gebruiken voor verschillende modellen. De GGAA is als prestatie maat te beperkt. Het doel van het model is om per gemeente de bijstandsuitgaven gedurende het gehele jaar te voorspellen. Dus modellen moeten worden beoordeeld of ze het aantal bijstandsjaren per gemeente goed kunnen voorspellen.¹ Bij de verdere ontwikkeling lijkt het mij nuttig om dit als maatstaf te gebruiken in plaats van GGAA.

In de gepresenteerde prestatie maten krijgt elke gemeente (meestal) een gelijk gewicht. Dat is goed te rechtvaardigen als naar absolute verschillen in bedragen wordt gekeken. Maar in dit geval wordt gekeken naar absolute verschillen in percentages. Dat betekent dat huishoudens in de kleine gemeenten een groter gewicht krijgen. De variantie van het percentage bijstandsgerechtigden is nu eenmaal groter in kleine gemeenten dan in grote gemeenten. Dit wordt erkend in het rapport en valt ook te zien in de uitsplitsing naar grootte van de gemeente.

Verschillen tussen gemeenten

Om een beter beeld te krijgen van de prestatie van de modellen is meer informatie over de prestatie nuttig. Wat is bijvoorbeeld de maximale afwijking die optreedt? Het rapport laat ook voor een aantal gemeentekennmerken zien dat er geen structurele afwijkingen zijn. De uitzondering is de grootte van de gemeente. De bijstandskans in gemeenten met 50 tot 100 duizend inwoners wordt overschat, terwijl bij de vier grote steden het aantal bijstandsgerechtigden onderschat wordt. Uit de figuren in het rapport kan niet goed worden gezien hoe groot deze overschatting en onderschatting is. Er is tijdens het proces van het ontwikkelen van het model zorgvuldig gekeken naar de selectie van verklarende variabelen. Maar als uit de exacte herverdeel effecten per gemeente blijkt dat de verschillen groot zijn, dan zal nog eens gekeken moeten worden of er misschien toch nog variabelen zijn die ontbreken die structureel onderschatten of overschatten van bepaalde soorten gemeenten kunnen verklaren. Hierbij moet wel worden opgemerkt dat de selectie van variabelen op gemeenteniveau of regioniveau het meest lastige is omdat deze vaak samenhangen met gemeentebeleid. Maar dit kan bij verdere ontwikkeling nader onderzocht worden.

Conclusie

SEO heeft het door SCP ontwikkelde verdeelmodel op een aantal punten verbeterd en mijn aanbeveling is dan ook zo snel mogelijk over te stappen naar het verdeelmodel van SEO. Op een paar punten is het model van SEO misschien nog verder te verbeteren, maar dat kan meegenomen worden in de verdere ontwikkeling.

¹ Extra berekeningen van SEO laten zien dat het Logit model het hier ook beter doet dan het lineaire model, zelfs als dat alleen geschat wordt op bijstandsontvangst op 5 januari. Het is nuttig om dit te documenteren in het rapport in plaats van de GGAA.