



Planbureau voor de Leefomgeving

Discussion Paper

Regionale bevolkings- en huishoudensprognose: methodiek prognoseintervallen

De standpunten in dit document zijn die van de auteur(s) en komen niet noodzakelijk overeen met het beleid van Centraal Bureau voor de Statistiek

2017 | 21

**Coen van Duin (CBS)
Andries de Jong (PBL)**

Content

1. Inleiding	4
2. Methoden om onzekerheid in beeld te brengen	6
3. Methodiek ter bepaling van prognose-intervallen in de CBS nationale prognose	9
4. Methode en veronderstellingen voor schatten regionale prognose-intervallen	12
4.1 Algemene methode	12
4.2 Veronderstellingen voor de demografische varianten	15
4.3 Veronderstellingen voor de huishoudensdynamiek varianten	17
4.4 Veronderstellingen voor de binnenlandse migratie-varianten	18
5. Uitkomsten	21
6. Vergelijking met de WLO scenario's van 2015	26
7. Conclusies en evaluatie	30
Literatuur	31
Appendix 1: Schatten van de onzekerheidsmarges uit de uitkomsten van de varianten	32
Appendix 2: Afleiden inputveronderstellingen voor de nationale demografische varianten	35
Appendix 3: Toevoegen extra regionale onzekerheid aan de nationale varianten	39
Appendix 4: Schatten van de onzekerheid in de binnenlandse migratie	42

Summary

Bij de regionale prognose van 2016 zijn voor het eerst onzekerheidsintervallen rond de prognosecijfers gepubliceerd. Om deze te kunnen schatten is een nieuwe methode ontwikkeld, die in dit artikel wordt beschreven. Rekening houdend met deze intervallen is voor 70 procent van de Nederlandse gemeenten niet met zekerheid te stellen of het inwonertal tussen 2015 en 2040 zal groeien of krimpen. Voor 20 procent is groei waarschijnlijk, voor 10 procent krimp. Bij andere prognose-uitkomsten is er meer stelligheid. Zo geldt een afname van het aantal 20- tot 65-jarigen voor 60 procent van de gemeenten als waarschijnlijk en zal 85 procent van de gemeenten in 2040 waarschijnlijk meer huishoudens tellen

Keywords

Prognoses, Bevolking, Onzekerheidsmarges, Regio

Bij de regionale prognose van 2016 zijn voor het eerst onzekerheidsintervallen rond de prognosecijfers gepubliceerd. Om deze te kunnen schatten is een nieuwe methode ontwikkeld, die in dit artikel wordt beschreven. Rekening houdend met deze intervallen is voor 70 procent van de Nederlandse gemeenten niet met zekerheid te stellen of het inwonertal tussen 2015 en 2040 zal groeien of krimpen. Voor 20 procent is groei waarschijnlijk, voor 10 procent krimp. Bij andere prognose-uitkomsten is er meer stelligheid. Zo geldt een afname van het aantal 20- tot 65-jarigen voor 60 procent van de gemeenten als waarschijnlijk en zal 85 procent van de gemeenten in 2040 waarschijnlijk meer huishoudens tellen.

1. Inleiding

In oktober 2016 publiceerden het PBL en het CBS de zesde editie van de regionale prognose (Kooiman, de Jong, Huisman, van Duin en Stoeldraijer, 2016). Deze prognose beschrijft de verwachte toekomstige ontwikkeling van het aantal inwoners naar leeftijd en geslacht en het aantal huishoudens naar type voor de Nederlandse gemeenten. De prognose kijkt ongeveer een kwart eeuw vooruit, tot 2040. In eerdere edities van de regionale prognose werden enkel puntschattingen gepubliceerd, zonder indicatie van de onzekerheid hieromheen. Om toch een beeld van de onzekerheid te geven, werden (in Huisman, de Jong, van Duin en Stoeldraijer 2013) twee alternatieve toekomstbeelden gepresenteerd, gebaseerd op de Ruimtelijke Verkenning 2011 (PBL 2011). In de prognose van 2016 wordt de onzekerheid echter expliciet gemaakt, door tevens prognose-intervallen te publiceren.

Prognoses beogen de meest waarschijnlijke toekomstige ontwikkelingen te beschrijven, maar de werkelijke ontwikkelingen zullen er altijd in meer of mindere mate van afwijken. Intervallen rond de prognoses geven een schatting van hoe groot die afwijking gemiddeld zal zijn. Door dit soort intervallen te publiceren, informeren het PBL en het CBS gebruikers over de mate van onzekerheid in de cijfers en geven daarmee ook een indicatie van de kans dat in een volgende editie van de prognoses die cijfers anders zullen zijn. Uiteraard zijn deze intervallen, net zo goed als de prognose zelf, schattingen, die in latere edities van de prognose, als meer informatie beschikbaar is, zullen veranderen.

Dit artikel beschrijft de methode die is gebruikt om de intervallen te berekenen en bespreekt de resultaten. In de volgende paragraaf wordt eerst ingegaan op manieren om onzekerheid in beeld te brengen en welke keuzen zijn gemaakt bij het ontwikkelen van de methodiek. In paragraaf 3 wordt de stochastische methodiek van de nationale prognose in meer detail besproken, omdat de aanpak voor de regionale prognose hierop is geïnspireerd. Paragraaf 4 beschrijft hoe de regionale onzekerheidsintervallen worden geschat. Paragraaf 5 bespreekt de resultaten. Door de prognose-uitkomsten en de intervallen te vergelijken wordt in kaart gebracht welke regionale demografische ontwikkelingen waarschijnlijk en welke

onwaarschijnlijk zijn. In paragraaf 6 wordt gekeken hoe de prognose-intervallen zich verhouden tot de uitkomsten van de 4 scenario's uit de Welvaart een Leefomgeving studies. In de slotparagraaf wordt kort ingegaan op mogelijkheden om de methodiek verder te ontwikkelen. Voor wiskundige achtergronden wordt verwezen naar de appendices.

2. Methoden om onzekerheid in beeld te brengen

Prognoses hebben tot doel de meest waarschijnlijke toekomstige ontwikkeling te beschrijven. Ze geven een beeld van één mogelijke toekomst. Als de prognose perfect is opgesteld, is de kans dat het inwonertal in de toekomst hoger is dan de prognose 50 procent en de kans dat het lager ligt ook. Dit geldt niet alleen voor het inwonertal, maar voor alle uitkomsten van de prognose. Prognoses doen dus uitspraken als: ‘het inwonertal van Amsterdam in 2040 ligt met gelijke kans hoger of lager dan 1,1 miljoen’. Wat hun toepasbaarheid beperkt, is dat ze geen uitspraak doen over hoe dichtbij die 1,1 miljoen bij het werkelijke inwonertal zal liggen. Om daarover iets te zeggen, moet een beeld van de onzekerheid van de prognose-uitkomsten worden gegeven.

Er zijn traditioneel drie manieren om de onzekerheid van toekomstige ontwikkelingen in kaart te brengen: de scenariobenadering, de variantenbenadering en de stochastische benadering. Al deze benaderingen hebben gemeen dat de input van de prognose wordt gevarieerd om te kijken hoe de uitkomsten hierdoor veranderen.

In de scenariobenadering worden een klein aantal verhaallijnen opgesteld voor de toekomstige ontwikkeling; deze verhaallijnen kijken breder dan alleen de demografische ontwikkeling. Vervolgens worden voor elk toekomstbeeld kwantitatieve veronderstellingen afgeleid die consistent zijn met de verhaallijnen, en die gebruikt kunnen worden als input voor de doorrekening van het simulatie model. Veronderstellingen voor demografische kerncijfers zoals het toekomstige kindertal, de levensverwachting en het migratiesaldo worden in samenhang gevarieerd. Een voorbeeld van deze aanpak zijn de welvaart- en leefomgeving-scenario's (WLO) die door de Nederlandse overheid worden gebruikt (Manders en Kool, 2015). Een ander voorbeeld zijn de Shared Socioeconomic Pathways (SSP) scenario's (Lutz, Butz & KC, 2014), die worden gebruikt bij het doorrekenen van effecten van klimaatverandering. Waar prognoses trachten de meest waarschijnlijke ontwikkeling te beschrijven, zetten de scenario's een aantal mogelijke ontwikkelingen naast elkaar. In het algemeen kan je niet aangeven hoe waarschijnlijk ieder scenario is. De mate waarin de uitkomsten variëren tussen verschillende scenario's geeft echter wel een indicatie van hun onzekerheid.

De variantenbenadering lijkt op de scenariobenadering, maar dan zonder verhaallijnen. In deze benadering wordt meer systematisch verkend welke invloed een andere keuze voor elk van de demografische kerncijfers op de gesimuleerde bevolkingsontwikkeling heeft. Een voorbeeld zijn de Bevölkerungsvorausberechnungen van het Duitse federale statistische bureau (Destatis, 2015). Die bestaan uit 8 varianten waarin het kindertal, de netto buitenlandse migratie en de toekomstige levensverwachting afwisselend hoog of laag

worden ingesteld. De spreiding van de uitkomsten over deze varianten geeft een beeld van hun onzekerheid.

Zowel de scenario- als de variantenbenadering geven weliswaar een beeld van de gevoeligheid van uitkomsten voor de gebruikte veronderstellingen, maar zijn niet bruikbaar om onzekerheidsintervallen rond prognose-uitkomsten te schatten. Eén probleem is dat in scenario's wordt uitgegaan van volledige samenhang in de tijd tussen de afwijkingen ten opzichte van de prognose in de verschillende componenten, bijvoorbeeld voor alle jaren hogere vruchtbaarheid én hogere levensverwachting én hogere internationale migratie. In werkelijkheid zijn de onzekerheden in die componenten grotendeels onafhankelijk van elkaar, waardoor hun effecten elkaar deels zullen compenseren. Hierdoor zal de mate van onzekerheid minder snel oplopen dan scenario's suggereren (De Beer en Alders, 1999).

Een ander probleem is dat, afhankelijk van de doelvariabele waarvan men de onzekerheid wil beschrijven, andere scenario's moeten worden geformuleerd. De eerder genoemde hoog/laag-scenario's met hoge respectievelijk lage vruchtbaarheid, levensverwachting en netto migratie beschrijven twee toekomstbeelden met een duidelijk verschil in de ontwikkeling van het inwonertal, maar met mogelijk slechts een klein verschil voor de vergrijzing. Immers: een hogere levensverwachting leidt tot een hoger aandeel ouderen, maar een hoger vruchtbaarheidscijfer en migratiesaldo geven juist een lager aandeel. Wie toekomst met hoge en lage vergrijzing wil beschrijven zou dus naar andere scenario's moeten kijken, waarin lage vruchtbaarheid en migratie met een hoge levensverwachting wordt gecombineerd of hoge vruchtbaarheid en migratie met een lage levensverwachting. In de variantenaanpak wordt dit probleem deels ondervangen. Er wordt niet voor één combinatie van hoge/lage veronderstellingen voor de verschillende componenten gekozen, maar de resultaten van alle mogelijke combinaties worden naast elkaar getoond. Er wordt echter geen schatting van de gecombineerde onzekerheid ten gevolge van de verschillende componenten gegeven. Bovendien produceren de varianten typisch een te snelle toename van de onzekerheid in de loop van de prognose, omdat er geen rekening mee wordt gehouden dat fluctuaties in de componenten zichzelf over meerdere jaren kunnen uitmiddelen. Dit probleem speelt overigens ook bij scenario's.

Om deze problemen te verhelpen zijn in de jaren '90 van de vorige eeuw stochastische prognoses ontwikkeld (Alho en Spencer, 1985; Pflaumer, 1988). Deze methode wordt gebruikt om de onzekerheidsintervallen van de nationale CBS prognoses te schatten (De Beer en Alders, 1999a). Bij een stochastische prognose wordt een tijdreeksmodel gespecificeerd voor elke inputcomponent, waarbij de onzekerheid voor elke component in beeld wordt gebracht via een kansverdeling. Op grond van deze input wordt een zeer groot aantal varianten doorgerekend. De verdeling van de uitkomsten over de varianten wordt vervolgens geïnterpreteerd als een schatting van de kansverdeling voor de toekomstige ontwikkelingen. Uit deze kansverdeling kunnen prognose-intervallen worden bepaald.

Idealiter zouden de onzekerheidsintervallen voor de regionale prognose ook stochastisch worden bepaald. Dit zou echter gepaard gaan met een zeer lange doorlooptijd. Dit komt door het grote aantal inputveronderstellingen dat gevarieerd zou moeten worden: voor rond vierhonderd gemeenten zijn alle combinaties van vruchtbaarheidscijfers, sterftcijfers, buiten en binnenlandse migratiecijfers, woningbouwcijfers en overgangskansen tussen (zeven) huishoudensposities nodig. Door het grote aantal vrijheidsgraden zou een zeer groot aantal varianten gesimuleerd moeten worden om een voldoende omvangrijke steekproef van de mogelijke combinaties van inputveronderstellingen te krijgen. Een extra complicatie hierbij is dat de rekentijd per variant van het regionale prognosemodel PEARL veel hoger ligt dan die van het nationale prognosemodel.

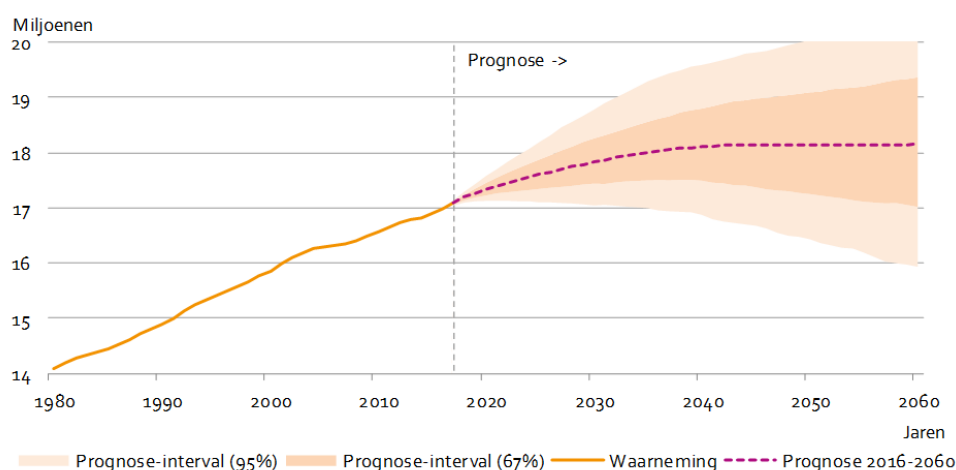
Gezien de bovenstaande problematiek is gezocht naar een alternatieve methode. Er is een uitbreiding van de klassieke variantenaanpak ontwikkeld waarmee wel onzekerheidsintervallen kunnen worden geschat. Per gemeente worden 16 varianten berekend, waarin steeds één inputcomponent wordt gevarieerd. Per component worden twee varianten opgesteld: één waarin de veronderstellingen hoger en één waarin ze lager dan volgens de prognose worden gekozen. De componenten waarvoor varianten worden berekend zijn: vruchtbaarheid, sterfte, internationale migratie, binnenlands vestigingssaldo, binnenlandse mobiliteit en 3 soorten overgangen tussen huishoudensposities. Deze methodiek verschilt in twee opzichten van de klassieke variantenaanpak. Ten eerste wordt het tijdsverloop van de inputveronderstellingen, net als bij een stochastische prognose, afgeleid door een stochastisch tijdreeksmodel voor de component te specificeren. Ten tweede worden de uitkomsten van de 16 varianten gecombineerd tot één onzekerheidsinterval per variabele. Op nationaal niveau geeft deze procedure onzekerheidsmarges die de intervallen van de stochastische nationale prognose goed benaderen. Door deze methodiek ook regionaal toe te passen wordt consistentie met de CBS-intervallen op nationaal niveau verkregen.

In paragraaf 4 wordt de methodiek voor het schatten van prognose-intervallen rond de regionale prognose in meer detail beschreven. Eerst wordt echter nader ingegaan op de methodiek van de nationale stochastische prognose, die als inspiratie dient voor de aanpak bij de regionale prognose.

3. Methodiek ter bepaling van prognose-intervallen in de CBS nationale prognose

Een prognose is een door de prognosemaker gemaakte, met feiten onderbouwde, inschatting. Ook de prognose-intervallen zijn een inschatting, maar in dit geval van het bereik rondom de prognose uitkomsten. Sinds 1998 bevat de nationale bevolkingsprognose van het CBS prognose-intervallen. Grafiek 3.1. toont het 67% en 95% interval van de bevolking, volgens de (kern)prognose 2016-2060. Wanneer verder in de toekomst wordt gekeken, neemt de onzekerheid van de prognosecijfers toe.

3.1 Inwonertal, prognose 2016-2060



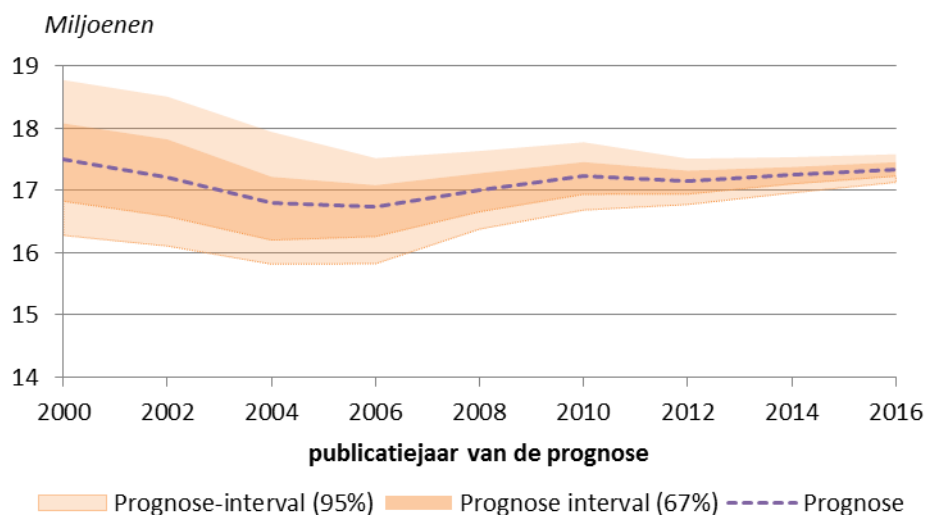
Bron: Centraal Bureau voor de Statistiek

Omgekeerd geldt dat de onzekerheid steeds kleiner wordt, naarmate het voorspelde kalenderjaar dichterbij het publicatiejaar van de prognose ligt. Grafiek 3.2 geeft als voorbeeld hiervan het verwachte inwonertal van Nederland in 2020 volgens opeenvolgende prognoses sinds 2000, met de bijbehorende prognose-intervallen. Naarmate 2020 dichterbij komt, neemt het vertrouwen in de voorspellingen toe en worden de prognose-intervallen smaller. In 2000 meldde het CBS dat het inwonertal in 2020 waarschijnlijk tussen de 16,8 en 18,1 miljoen zou liggen; in de prognose die 16 jaar later uitkwam was dit interval fors gekrompen.

De bevolkingsprognose wordt berekend aan de hand van veronderstellingen over het aantal immigranten, de vruchtbaarheidscijfers en de sterfte- en emigratiekansen per prognosejaar. Op grond van deze veronderstellingen en de waargenomen bevolking aan het begin van de prognoseperiode wordt de bevolking in latere jaren uitgerekend. Voor de stochastische bevolkingsprognose worden veronderstellingen gedaan over kansverdelingen voor de vruchtbaarheidscijfers, immigratie-aantallen en emigratie-

en sterftetekansen. Op basis van deze kansverdelingen wordt een groot aantal tijdspaden berekend voor de vruchtbaarheid, migratie en sterfte.

3.2 Inwonertal in 2020 volgens CBS-prognoses sinds 2000



De kansverdelingen voor de inputcomponenten worden vastgelegd door tijdreeksmodellen te specificeren. Twee vragen zijn daarbij van belang: wat is het interval waarbinnen de component waarschijnlijk aan het eind van de prognoseperiode zal liggen en hoe sterk middelen fluctuaties in de component zichzelf in loop van de tijd uit? De eerste vraag bepaalt de onzekerheid op de lange termijn, de tweede de wijze waarop de mate van onzekerheid gedurende de prognoseperiode toeneemt. Veronderstellingen voor beide zijn bepaald door een mix van methoden: analyse van fouten in oude prognoses, tijdreeksextrapolatie en expert-inschattingen (De Beer en Alders, 1999; Alders en de Beer, 2005, Carolina en Van Duin, 2010).

Voor het kindertal per vrouw en de levensverwachting wordt aangenomen dat de onzekerheden zichzelf slechts in beperkte mate uitmiddelen. De bandbreedte rond de veronderstellingen voor deze grootheden blijft daardoor ook verder in de toekomst toenemen, al gaat dat wel steeds trager. Bij internationale migratie is vaak sprake van gebeurtenissen waardoor de migratie tijdelijk sterk stijgt om vervolgens weer te dalen. Voor deze component wordt daarom een tijdreeksmodel gebruikt waarin een toename van het verschil met de prognoseveronderstelling in het ene jaar een afname van dat verschil in het volgende jaar waarschijnlijker maakt. In zo'n tijdreeksmodel middelen de afwijkingen zich sterker uit. De bandbreedte rond de veronderstellingen voor het migratiesaldo stijgt snel in de eerste jaren van de prognose, maar neemt op de langere termijn nauwelijks verder toe. In Appendix 2 wordt de wiskundige beschrijving van de tijdreeksmodellen gegeven.

Met behulp van de tijdreeksmodellen worden willekeurige tijdspaden voor vruchtbaarheid, sterfte en migratie gegenereerd. Deze worden vervolgens gebruikt om duizend varianten van de prognose te berekenen, elk met een andere combinatie

van veronderstellingen voor de componenten. Iedere variant staat voor een mogelijke toekomstige ontwikkeling van de Nederlandse bevolking. Zo wordt een “steekproef van toekomsten” gegenereerd. Voor elke gewenste grootte kunnen vervolgens onzekerheidsintervallen worden berekend uit de verdeling van de waarden over deze steekproef (De Beer, 1999).

Voor de stochastische huishoudensprognose wordt met een eenvoudiger model gewerkt dan voor de deterministische (niet-stochastische) huishoudensprognose. In het deterministische model worden overgangen tussen huishoudensposities gegenereerd. Het beschrijft daardoor niet alleen hoe de huishoudenssamenstelling van de bevolking van jaar op jaar verandert, maar ook door welke gebeurtenissen dat komt: zijn er meer alleenstaanden omdat er minder stellen gaan samenwonen, of omdat er meer gaan scheiden? De stochastische huishoudensprognose werkt met een zogenaamd “statisch” model, waarbij de bevolking met een leeftijds- en geslachtsafhankelijke verdelingsfunctie over huishoudensposities wordt verdeeld. Voor de stochastische huishoudensprognose worden veronderstellingen gemaakt over een kansverdeling voor deze verdelingsfunctie. Door deze kansverdeling te combineren met de uitkomsten van de duizend varianten van de stochastische bevolkingsprognose wordt een groot aantal huishoudensprognose-varianten gegenereerd. Deze worden gebruikt om onzekerheidsmarges rond de uitkomsten van de huishoudensprognose te schatten (Alders, 1999).

4. Methode en veronderstellingen voor schatten regionale prognose-intervallen

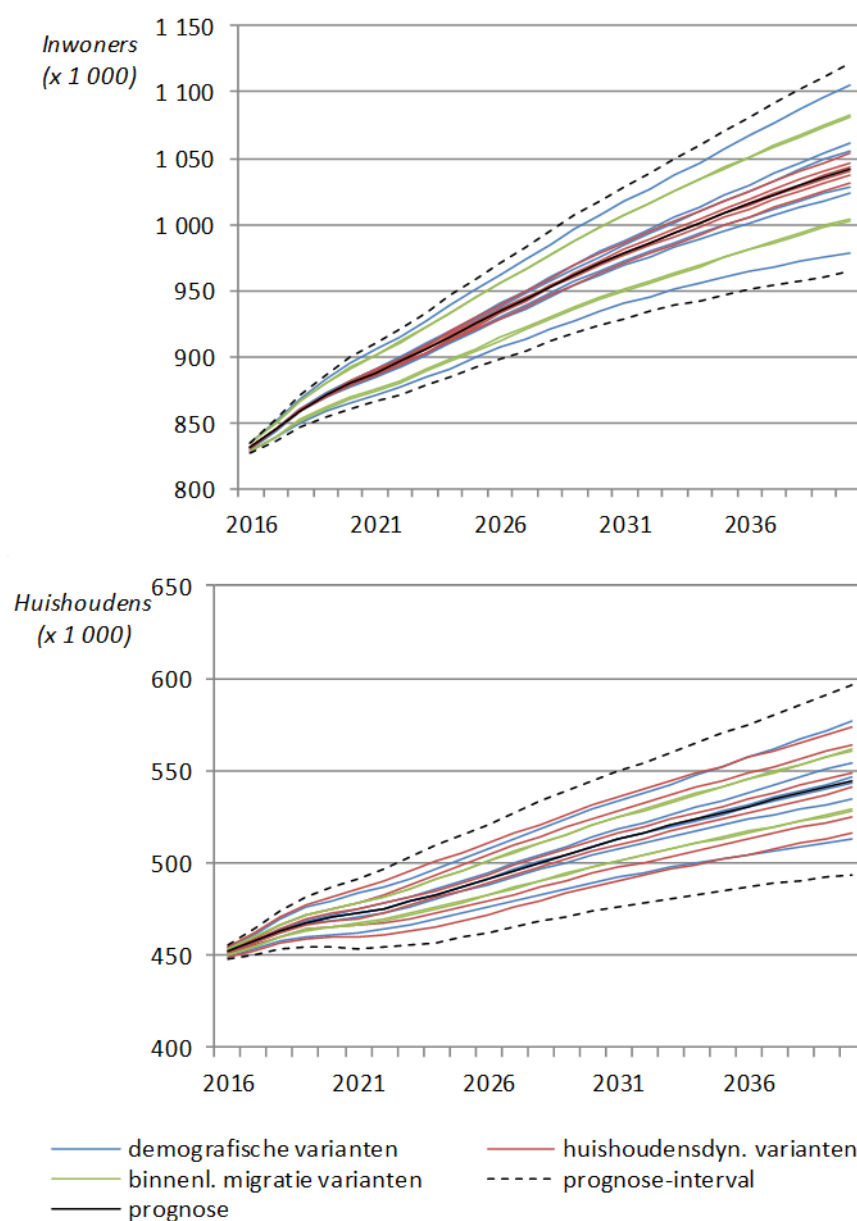
4.1 Algemene methode

De regionale prognose wordt berekend met het model PEARL (De Jong et al., 2005). Als input voor het model dienen vruchtbaarheidscijfers, die de kans beschrijven dat een vrouw van een gegeven leeftijd een kind krijgt, sterfrequenties, die de kans op overlijden beschrijven, immigratieaantallen en emigratieaantallen, frequenties voor overgangen per jaar tussen verschillende huishoudensposities, verhuisfrequenties en ten slotte parameters die bepalen in welke bestemmingsgemeenten verhuizende personen terecht komen. Op basis van deze input simuleert het model de toekomstige bevolkingsontwikkeling.

Om de onzekerheidsmarges te berekenen is voor elk soort inputparameter bepaald wat het effect is van een “typische” fout daarin op de uitkomsten van de prognose. Dit wordt gedaan door een prognose-variant te berekenen waarin die input is aangepast. Hoe deze aangepaste input is gekozen wordt toegelicht in paragraaf 4.1-4.3. Er zijn 3 paar demografische varianten doorgerekend, waarin respectievelijk de input voor het kindertal per vrouw, de levensverwachting en het saldo buitenlandse migratie zijn gewijzigd. Ook zijn 3 paar huishoudensdynamiekvarianten berekend, waarin de overgangsfrequenties tussen verschillende huishoudensposities zijn gevarieerd. Per gemeente zijn daarnaast nog 4 varianten berekend waarin de veronderstellingen voor binnenlandse migratie (verhuizingen tussen gemeenten) zijn aangepast. De binnenlandse migratievarianten zijn voor elke gemeente afzonderlijk berekend met een vereenvoudigd prognosemodel (paragraaf 4.4). De 8 demografische en huishoudensdynamiekvarianten zijn doorgerekend in PEARL, waarbij voor alle gemeenten tegelijk de hoge dan wel lage veronderstellingen werden gebruikt. Uit de verschillen tussen de uitkomsten van de 16 varianten en de prognose worden onzekerheidsmarges afgeleid waarin het effect van de onzekerheid in elk van de inputparameters is meegenomen. Dit gebeurt in twee stappen. Eerst worden uit de uitkomsten van de 16 varianten cijfers voor “combinatievarianten” geschat. Dit zijn varianten waarin, net als in een echte stochastische prognose, alle componenten tegelijk worden gevarieerd. Acht componenten die elk hoger of lager dan volgens de prognose worden ingesteld, geeft in totaal $2^8=256$ combinatievarianten. Vervolgens worden prognose-intervallen bepaald door te kijken naar de gemiddelde kwadratische prognosefout over deze combinatievarianten. De berekeningen zijn beschreven in Appendix 1.

Ten opzichte van een echte stochastische prognose heeft deze methode een aantal beperkingen. Ten eerste worden alleen onzekerheidsintervallen berekend en wordt geen volledige kansverdeling van de prognose-uitkomsten gegeven. Ten tweede is de methode niet geschikt voor het schatten van intervallen rond stroomcijfers, zoals het jaarlijkse aantal geboorten of verhuizende personen. De methode is toegesneden op het schatten van intervallen rond standcijfers, zoals het aantal inwoners of huishoudens. Ook kunnen er intervallen voor gecumuleerde stroomcijfers vanaf het startjaar van de prognose mee worden geschat (bijvoorbeeld voor het totale aantal geboorten vanaf het startjaar tot en met 2040).

4.1.1 Uitkomsten van de varianten voor Amsterdam (inwonertal en aantal huishoudens) en de resulterende prognose-intervallen



Grafiek 4.1.1 toont als voorbeeld voor de gemeente Amsterdam de uitkomsten van de varianten voor het inwonertal en voor het aantal huishoudens en de daaruit afgeleide onzekerheidsmarges. Dit demonstreert dat de onzekerheid voor verschillende uitkomsten door een andere mix van componenten wordt gedomineerd. De onzekerheid in het inwonertal van Amsterdam heeft vooral te maken met de buitenlandse en binnenlandse migratie. Voor de onzekerheid in het aantal huishoudens spelen ook huishoudensdynamiek-componenten een belangrijke rol.

4.2 Veronderstellingen voor de demografische varianten

Voor elke gemeente zijn er 6 demografische varianten doorgerekend. In de GebHoog en GebLaag variant wordt uitgegaan van een hoger en lager gemiddeld kindertal per vrouw dan in de prognose. In de SterfteHoog en SterfteLaag variant wordt een lagere/hogere levensverwachting dan in de prognose verondersteld. In de MigHoog en MigLaag variant wordt van een hoger en lager migratiesaldo uitgegaan. Bij het afleiden van de hoge en lage veronderstellingen voor deze varianten is in twee stappen gewerkt. Eerst zijn voor Nederland als geheel, dus zonder onderscheid naar regio, hoge en lage varianten geconstrueerd waarin het vruchtbaarheidscijfer, de sterftetekansen of de netto migratie aangepast zijn ten opzichte van de prognose. Deze varianten zijn zo gekozen dat de standcijfers die uit de uitkomsten van de varianten kunnen worden afgeleid, de 67% boven- en ondermarge volgens de nationale CBS stochastische prognose zo goed mogelijk benaderen. In een tweede stap zijn de veronderstellingen voor de hoge en lage variant per gemeente aangepast om mee te nemen dat demografische ontwikkeling op gemeentelijke niveau in het algemeen minder zeker zijn dan op nationaal niveau.

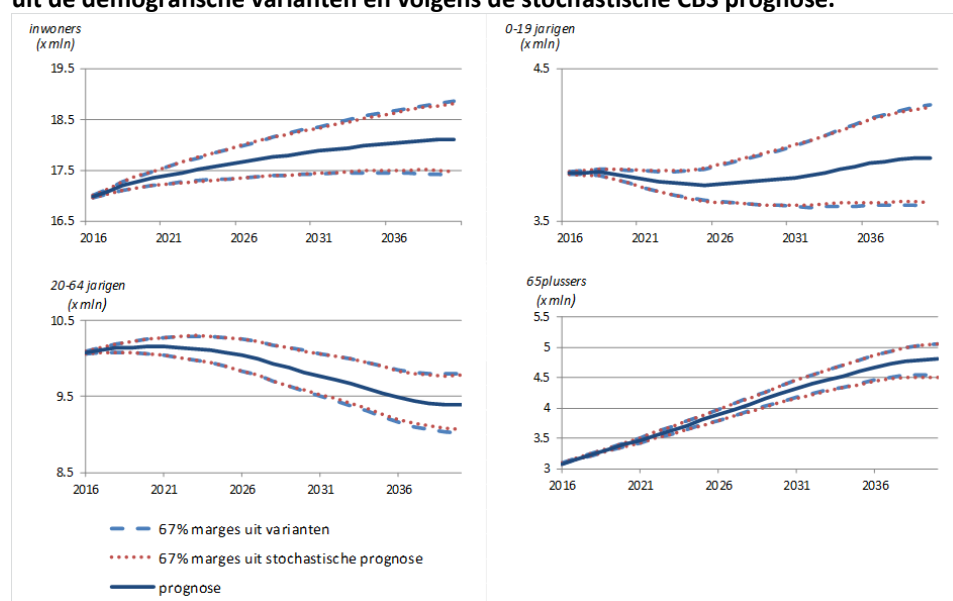
De input voor de nationale Geb-, Mig- en Sterftevarianten gaat in eerste instantie uit van de veronderstellingen voor de onzekerheid van de vruchtbaarheidscijfers, de buitenlandse migratie-aantallen en de sterftetekansen uit de nationale stochastische prognose (paragraaf 3). Vervolgens wordt de 67% bandbreedte in de waarde voor de inputvariabele voor de varianten gereduceerd met een tijdsafhankelijke factor. De versmallingsfactor corrigeert ervoor dat een deel van de onzekerheid in de inputvariabele veroorzaakt wordt door fluctuaties die zichzelf na een aantal jaar uitmiddelen en daardoor niet bijdragen aan de onzekerheid in de standcijfers. De versmallingsfactor wordt geschat uit de mate waarin de fluctuaties in opeenvolgende jaren elkaar uitmiddelen. Hoe sterker dit effect, hoe meer versmalling (Appendix 2). Tabel 4.2.1. geeft een indruk van de marge tussen de lage en hoge variant wat betreft de geboorte, sterfte en netto buitenlandse migratie (immigratie minus emigratie). Het verschil in het aantal gebeurtenissen tussen de hoge en lage variant is het grootst voor de Mig-varianten. Voor de periode tot 2040 wordt dus aangenomen dat internationale migratie van de drie demografische componenten de grootste onzekerheid heeft.

Grafiek 4.2.2. laat de 67% marge zien voor bevolking van Nederland volgens de nationale CBS stochastische prognose en de geschatte 67% marge op basis van de 6 demografische varianten. Zowel voor het inwonertal als voor het aantal jongeren, ouderen en de potentiële beroepsbevolking (20-64-jarigen) lijken de resultaten sterk op elkaar. Ook voor verhoudingscijfers zoals de grijze en groene druk is de overeenkomst goed. Voor meer informatie over de methodiek op basis van 6 demografische varianten wordt verwezen naar Van Duin (2017).

4.2.1 Gemiddeld aantal gebeurtenissen per 10 duizend inwoners per jaar in de periode 2015-2039, volgens de prognose en de 12 demografische en huishoudensdynamiek varianten

		Prognose	Variante Laag	Variante Hoog
Naam varianten	Input die is gevarieerd	<i>geboorten</i>		
Geb(Laag/Hoog)	Kindertal per vrouw	107	101	113
		<i>sterfgevallen</i>		
Sterfte(Laag/Hoog)	Levensverwachting	94	88	101
		<i>migranten (netto)</i>		
Mg(Laag/Hoog)	Internationale migratie	15	5	25
		<i>kinderen die uit huis gaan (netto)</i>		
Thuis(Laag/Hoog)	Leeftijd bij uit huis gaan	101	106	94
		<i>personen die gaan samenwonen</i>		
Paar(Laag/Hoog)	Samenwonen- en scheidingskans	228	213	240
		<i>personen die stoppen met samenwonen</i>		
		187	198	176
		<i>personen die in instelling gaan wonen (netto)</i>		
Inst(Laag/Hoog)	Kans om naar instelling te gaan	16	9	23

4.2.2 Prognose uitkomsten voor inwonertal naar leeftijd, met prognose-intervallen uit de demografische varianten en volgens de stochastische CBS prognose.



Op het niveau van gemeenten is de onzekerheid in de demografische kerncijfers over het algemeen groter dan op nationaal niveau, omdat tegengestelde ontwikkelingen in verschillende regio's op nationaal niveau uit kunnen middelen. Om dit te modelleren worden de gemeentelijke boven- en ondergrenzen in de varianten relatief ruimer ingesteld dan de nationale. Er is geschat hoeveel extra onzekerheid voor elke gemeente moest worden toegevoegd door de tijdreeksen voor het kindertal, het sterftecijfer en de internationale migratie voor individuele gemeenten te vergelijken met de tijdreeksen op nationaal niveau (appendix 3). Voor kleinere

gemeenten blijkt de onzekerheid in vruchtbaarheid, migratie en sterfte groter te zijn dan voor grotere. Voor gemeenten met een hoog vruchtbaarheidscijfer is het kindertal stabiel en dus beter te voorspellen dan voor gemeenten met lage vruchtbaarheid. Voor gemeenten die verhoudingsgewijs veel immigranten ontvangen ten opzichte van het inwonertal zijn de migratiestromen onzekerder dan voor gemeenten die weinig immigratie kennen.

4.3 Veronderstellingen voor de huishoudensdynamiek varianten

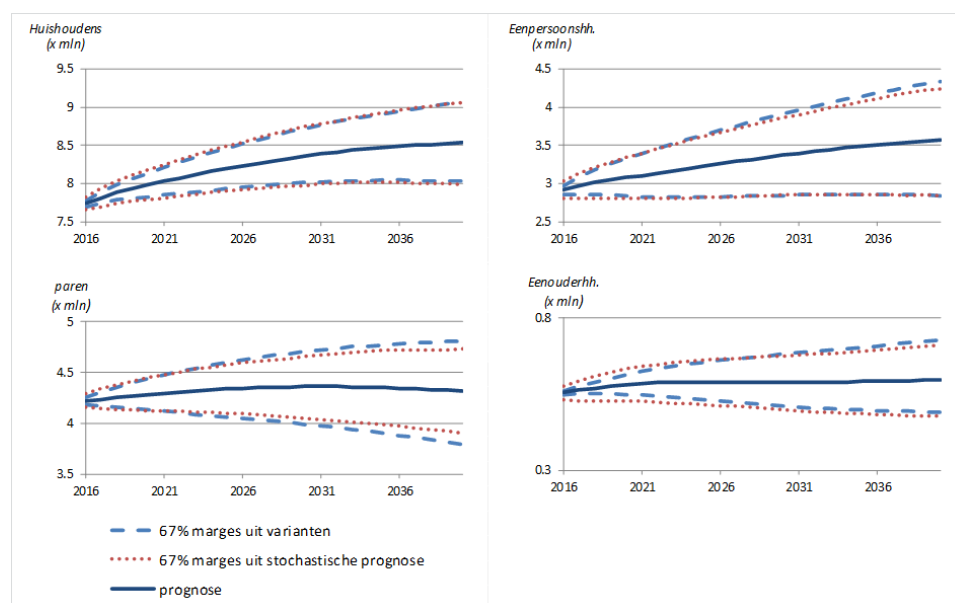
In de huishoudensdynamiek varianten worden de frequenties van overgangen tussen huishoudensposities gevarieerd ten opzicht van de veronderstellingen die in de prognose zijn gebruikt. Daarbij gaat het om kinderen die het ouderlijk huis verlaten, stellen die gaan samenwonen of scheiden, en personen (vooral ouderen) die in een instelling gaan wonen. Ook hier worden eerst nationale varianten gemaakt. Die worden zo gekozen dat de daaruit afgeleide onzekerheidsintervallen voor het aantal inwoners naar huishoudenspositie en het aantal huishoudens naar type die van de nationale stochastische huishoudensprognose benaderen.

Het opstellen van de varianten verloopt technisch iets anders dan bij de demografische varianten. De onzekerheidsmarges van de nationale CBS stochastische huishoudensprognose worden berekend in een model waarin geen overgangen tussen huishoudensposities worden gesimuleerd (paragraaf 3). Omdat het model geen huishoudenovergangen beschrijft zijn er geen veronderstellingen over de onzekerheid van die overgangen beschikbaar, om de inputveronderstellingen voor de huishoudensdynamiek varianten op te baseren. Om deze reden wordt er gefit aan de uitkomsten van de stochastische huishoudensprognose. De input voor de 6 huishoudensdynamiekvarianten is zo gekozen dat de resulterende marges voor het aantal huishoudens naar type en de bevolking naar huishoudenspositie aansluiten bij de marges volgens de stochastische huishoudensprognose. Tabel 4.2.1 toont voor de drie paar huishoudensdynamiekvarianten het gemiddeld aantal gebeurtenissen per jaar voor de huishoudenovergang die in de betreffende variant is aangepast. Grafiek 4.3.1 vergelijkt voor een aantal prognose-uitkomsten de marges die uit deze 6 varianten worden geschat met de marges volgens de stochastische huishoudensprognose. Er is een goede overeenkomst, al is de aansluiting wel iets slechter als bij de demografische onzekerheidsmarges. De onzekerheidsmarges voor het aantal paren zijn op de lange termijn ruimer dan volgens de stochastische huishoudensprognose en de marges voor het aantal eenouderhuishoudens op de korte termijn smaller. De marges voor het totale aantal huishoudens en het aantal eenpersoonshuishoudens sluiten over de hele periode goed aan.

Net als bij de demografische stromen valt te verwachten dat de onzekerheid op gemeentelijk niveau groter is dan op nationaal niveau. Omdat geen data beschikbaar was om dit verschil voor de huishoudenovergangen te onderzoeken is gewerkt met de veronderstelling dat het verschil in onzekerheid tussen grote en kleine gemeenten bij de huishoudenovergangen hetzelfde patroon volgt als bij het

vruchtbaarheidscijfer (Appendix 3). Op grond van deze aanname is in de gemeentelijke cijfers extra bandbreedte tussen de hoge en lage varianten toegevoegd.

4.3.1 Prognose uitkomsten voor huishoudens naar type, met prognose-intervallen uit de demografische varianten en volgens de stochastische CBS prognose.



4.4 Veronderstellingen voor de binnenlandse migratie-varianten

In het prognosemodel PEARL wordt de binnenlandse migratie in een aantal stappen gesimuleerd. Verschillende soorten inputparameters spelen hierbij een rol: verhuiskansen, het aandeel van de verhuizende personen dat de gemeente verlaat, het aandeel mensen binnen die groep dat zich verder of juist minder ver dan 35 km van de herkomstgemeente vestigen, bestemmingspatronen voor de lange afstandsmigranten, inputparameters die bepalen hoe de bestemmingskeuze van de korteaftandsmigranten samenhangt met de afstand tot de vertrekgemeente en met het inwonertal van de bestemmingsgemeente (De Jong et al., 2006). Ook worden de verwachtingen voor de (netto) woningbouw per gemeente als input gebruikt bij het vaststellen van de bestemmingen van korteaftandsmigranten. Er is voor gekozen om ten behoeve van de berekening van de onzekerheidsmarges niet te werken met varianten van al deze inputparameters. Methodologisch is het zeer complex om uit deze met elkaar samenhangende inputparameters onzekerheidsmarges af te leiden. Bovendien zou doorrekenen in PEARL gepaard gaan met een zeer lange doorlooptijd voor de berekeningen. In plaats daarvan is er daarom voor gekozen om direct het binnenlands vertrek- en vestigingscijfer van elke gemeente te variëren. De binnenlandse migratie-varianten zijn doorgerekend in een cohort-component model

waarin geboorte, sterfte, binnen- en buitenlandse migratie werden gesimuleerd voor één gemeente. De input voor het model, zoals vruchtbaarheidscijfers, sterftcijfers en leeftijds patronen van vestigers en vertrekkers, zijn ontleend aan de output van PEARL. Om het effect van de vier binnenlandse migratievarianten op het aantal huishoudens te bepalen, is de bevolking via headship rates uitgesplitst naar huishoudenstypes.

Hieronder wordt in algemene termen beschreven hoe de inputveronderstellingen voor de varianten voor de binnenlandse migratie zijn afgeleid. Voor een meer wiskundige behandeling wordt verwezen naar appendix 4.

Als kernindicatoren voor de binnenlandse migratie worden in het model bruto vestigings- en vertrekcijfers gebruikt: de vestiging in een regio en het vertrek uit een regio, gedeeld door het inwonertal van de regio. In principe zou (net als bij de geboorte en sterfte) zowel voor vestiging als vertrek een lage en hoge variant kunnen worden opgesteld. De tijdreeksen voor de vestiging en vertrek zijn echter sterk positief gecorreleerd. In jaren dat veel mensen uit een gemeente vertrekken, vestigen zich hier ook veel mensen. Dit ligt voor de hand, aangezien gemeenten een redelijk stabiele opvangcapaciteit in de woningvoorraad hebben, die slechts geleidelijk verandert door toevoegingen en onttrekkingen aan de woningvoorraad of veranderingen in de woningbezetting. Binnenlandse migratie is daardoor vooral een proces van doorstroming, hetgeen een sterke correlatie tussen het aantal mensen dat vertrekt en zich hier vestigt 'afdwingt'. Dit betekent dat in de binnenlandse migratie-varianten van vestiging en vertrek niet onafhankelijk van elkaar gevarieerd mogen worden, omdat dat te wijde bandbreedtes in de uitkomsten zou geven. Om deze reden is met twee andere indicatoren voor de binnenlandse migratie gewerkt, waarvoor de tijdreeksen vrijwel ongecorrleerd zijn. De eerste indicator is het geometrisch gemiddelde van het vestigings- en vertrekcijfer. Dit zegt iets over de mobiliteit: hoeveel mensen stromen totaal in en uit de gemeente? De tweede indicator is de verhouding tussen het bruto vestigings- en vertrekcijfer. Dit zegt iets over het binnenlands migratiesaldo. Aan de hand van twee indicatoren kunnen vervolgens per gemeente vier varianten worden berekend door een hoge of lage waarde voor de eerste indicator te combineren met een hoge of lage waarde voor de tweede.

Voor beide indicatoren dient de onzekerheid te worden bepaald, ten behoeve van het opstellen van de lage en hoge variant. Om dit te schatten is gebruik gemaakt van een eenvoudig model om voorspelfouten te genereren. Hiertoe is voor de jaren 2006-2014 een projectie van vestiging en vertrek gemaakt door het toepassen van constante bruto cijfers voor deze twee componenten. Vervolgens kunnen tijdreeksen van voorspelfouten worden bepaald door per kalenderjaar de voorspelde cijfers te vergelijken met de waargenomen cijfers. Zowel de tijdreeksen van de afwijkingen voor mobiliteit als migratiesaldo kent enige autocorrelatie, hoewel deze minder sterk is dan voor de internationale migratie. Dit betekent dat bij de binnenlandse migratie meer sprake van tijdelijke fluctuaties die voor een belangrijk deel uitmiddelen dan bij buitenlandse migratie.

Ten behoeve van het opstellen van een lage en hoge variant wordt vervolgens de standaarddeviatie van de afwijkingen van de twee tijdreeksen via een regressiemodel gerelateerd aan kenmerken van de betrokken gemeenten. Het blijkt dat de standaarddeviatie van de ruistermen in beide tijdreeksen afnemen met het inwonertal van de gemeente. De relatieve onzekerheid in de mobiliteit en het vestigingsratio is derhalve kleiner voor grotere gemeenten. Voor de vier grote gemeenten is een dummy in het regressiemodel opgenomen dat de extra grote onzekerheid voor deze gemeenten weerspiegelt. Ook voor kleine gemeenten met minder dan 50 duizend inwoners is een dummy opgenomen, aangezien hier de onzekerheid bij het vestigingssaldo groter is en bij de mobiliteit juist kleiner.

Tabel 4.4.1 toont als voorbeeld worden de inputwaarden voor vijf gemeenten met een verschillend inwonertal. Voor kleinere gemeenten wordt met deze methode een relatief grotere onzekerheid in het bruto vestigingssaldo gevonden. Voor Renswoude (5 duizend inwoners in 2015) is het verschil tussen de hoogste en laagste waarde van het gemiddelde vestigingssaldo 8 procent van de totale verhuisstroom over de gemeentegrens (vestiging + vertrek). Voor Venlo (100 duizend inwoners) en Utrecht (ruim 300 duizend) ligt het op 3,5 procent. Omdat Utrecht per inwoner grotere vertrek- en vestigingsstromen heeft, is de absolute bandbreedte in het bruto vestigingssaldo groter dan bij Venlo.

4.4.1 Gemiddeld aantal binnenlandse migranten per jaar 2015-2039 volgens de prognose en de 4 binnenlandse migratie varianten voor geselecteerde gemeenten

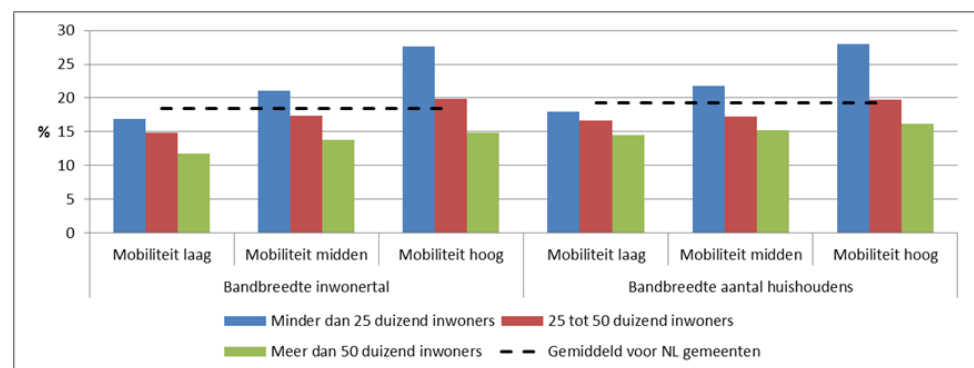
gemeente	Prognose	Mobiliteit hoog		Mobiliteit laag	
		Saldo hoog	Saldo laag	Saldo hoog	Saldo laag
<i>aantal binnenlandse migranten per 10 duizend inwoners</i>					
Renswoude (5 duizend inw.)					
vestiging	356	387	361	352	327
vertrek	355	358	386	328	352
saldo	1	29	-25	24	-25
Boekel (10 duizend inw.)					
vestiging	332	357	335	330	309
vertrek	337	339	362	314	334
saldo	-4	18	-26	15	-25
Kampen (51 duizend inw.)					
vestiging	274	288	278	270	261
vertrek	284	287	298	270	280
saldo	-10	0	-20	0	-20
Venlo (101 duizend inw.)					
vestiging	253	264	255	251	244
vertrek	291	294	303	280	289
saldo	-38	-30	-48	-29	-45
Utrecht (334 duizend inw.)					
vestiging	602	623	603	601	582
vertrek	596	596	618	576	596
saldo	6	27	-15	26	-14

5. Uitkomsten

Het blijkt dat voor het inwonertal van gemeenten de 67%-bandbreedte in 2040 gemiddeld op 18 procent geschat wordt. Dat komt neer op een onzekerheid rond de prognosewaarde van plus of min 9 procent. Kleinere gemeenten hebben relatief een grotere geschatte bandbreedte en daarmee een grotere onzekerheid rond de prognose, dan grote gemeenten. Ook blijkt van belang hoeveel mobiliteit er in de gemeentelijke bevolking is, dus hoeveel mensen zich er vestigen of vertrekken gerelateerd aan het inwonertal. Hoe hoger de mobiliteit, hoe onzekerder de prognose. Voor de periode 2015-2019 zijn Vlagtwedde, Vlieland en Wageningen de gemeenten met de hoogste mobiliteit in de prognose. Urk, Edam-Volendam en Rijssen-Holten hebben de laagste mobiliteit. Grafiek 5.1 toont de bandbreedte voor gemeenten naar het inwonertal en de mobiliteit. Voor grote gemeenten met een lage mobiliteit (zoals Apeldoorn of Emmen) is de relatieve bandbreedte gemiddeld bijna drie keer kleiner dan voor kleine gemeenten met een hoge mobiliteit (b.v. Ameland of Westervoort).

De relatieve bandbreedte voor het aantal huishoudens in 2040 komt gemiddeld op 19 procent, dus vergelijkbaar met die voor het inwonertal. Het verschil in bandbreedte tussen grotere en kleinere gemeenten en gemeenten met meer en minder mobiliteit is kleiner dan voor het inwonertal.

5.1. Gemiddelde relatieve bandbreedte van het inwonertal en het aantal huishoudens voor gemeenten naar gemeentegrootte (2015) en mobiliteit (2015-2019)



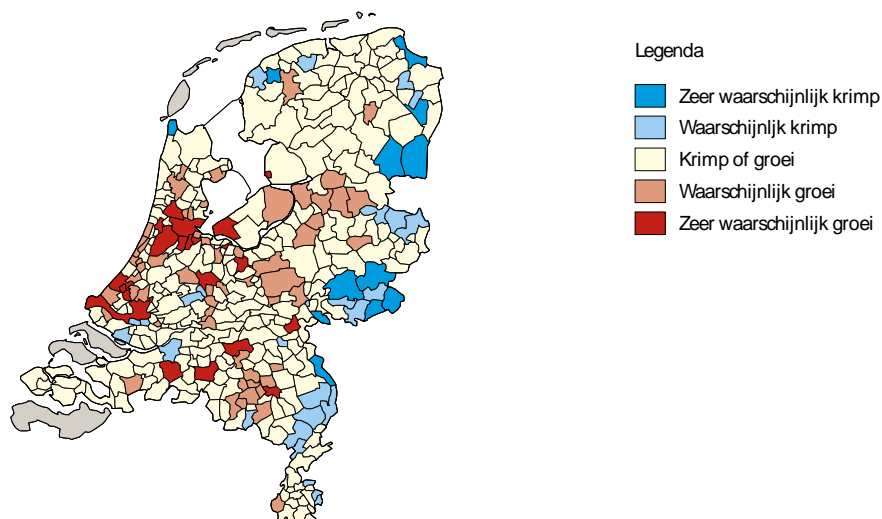
Er zijn onzekerheidsmarges geschat voor het inwonertal naar leeftijdsgroep (totaal, 0 tot 20 jaar, 20 tot 65 jaar en 65 jaar of ouder) en voor het aantal huishoudens naar type (totaal, eenpersoonshuishoudens, paren, eenouderhuishoudens). Deze informatie is beschikbaar op de websites van het CBS en het PBL (CBS, 2016; PBL, 2016). De uitkomsten bevatten ook onzekerheidsmarges voor provincies en COROP-regio's.

Met behulp van de onzekerheidsintervallen kan worden afgeleid of het waarschijnlijk is dat een gemeente de komende jaren gaat groeien of krimpen. Als de

benedengrens voor het inwonertal in 2040 hoger ligt dan het inwonertal in 2015, dan is het waarschijnlijk dat de gemeente over die periode groeit. Ligt de benedengrens van het interval onder de waarde in 2015 maar de bovengrens erboven, dan is groei voor die gemeente niet veel meer of minder waarschijnlijk dan krimp. Als de bovengrens in 2040 lager ligt dan de startwaarde zal de gemeente in 2040 waarschijnlijk kleiner zijn dan in 2015.

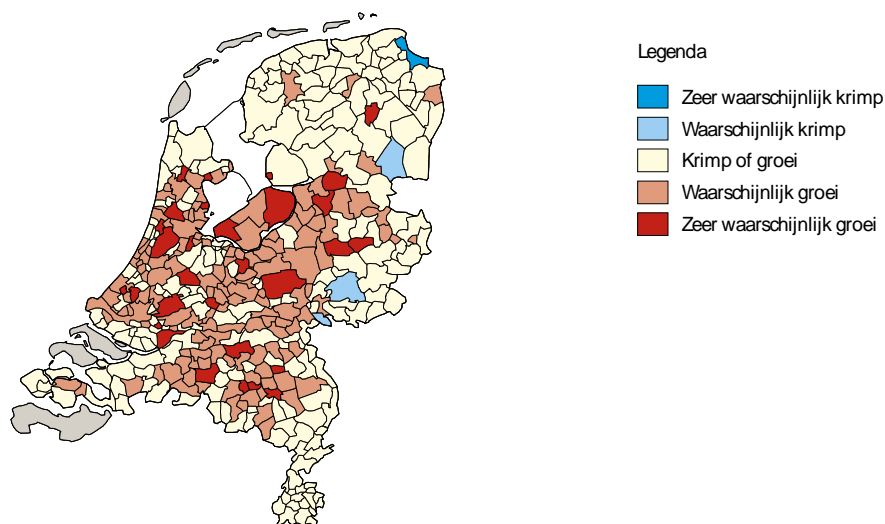
Veel gemeenten in de Randstad zullen volgens de prognose tussen 2015 en 2040 waarschijnlijk groeien in inwonertal, zie kaart 5.2. In de regio's ten oosten, zuiden en noorden zijn groeigemeenten dunner gezaaid. Voor een flink aantal gemeenten aan de oostgrens wordt krimp waarschijnlijk geacht. Voor de meeste gemeenten echter (70 procent) is onzeker of ze zullen groeien of krimpen. In de kaart staat voor sommige gemeenten "zeer waarschijnlijk groei" of "zeer waarschijnlijk krimp" aangegeven. Dit zijn gemeenten waar het verschil tussen het inwonertal in 2040 en 2015 groter is dan twee keer de onzekerheidsmarge op het prognosecijfer voor 2040. Bij de ontwikkeling van het aantal huishoudens is het patroon anders dan bij het inwonertal (kaart 5.3). Hier wordt voor de meeste gemeenten in de "romp" van Nederland tussen 2015 en 2040 een toename verwacht en is voor slechts 4 gemeenten krimp waarschijnlijk. Voor 55 procent van de gemeenten is onzeker of het aantal huishoudens toe of af zal nemen.

5.2 Verwachte ontwikkeling inwonertal per gemeenten 2015-2040



Bron: PBL/CBS regionale bevolkings- en huishoudensprognose 2015-2040

5.3 Verwachte ontwikkeling aantal huishoudens per gemeente 2015-2040

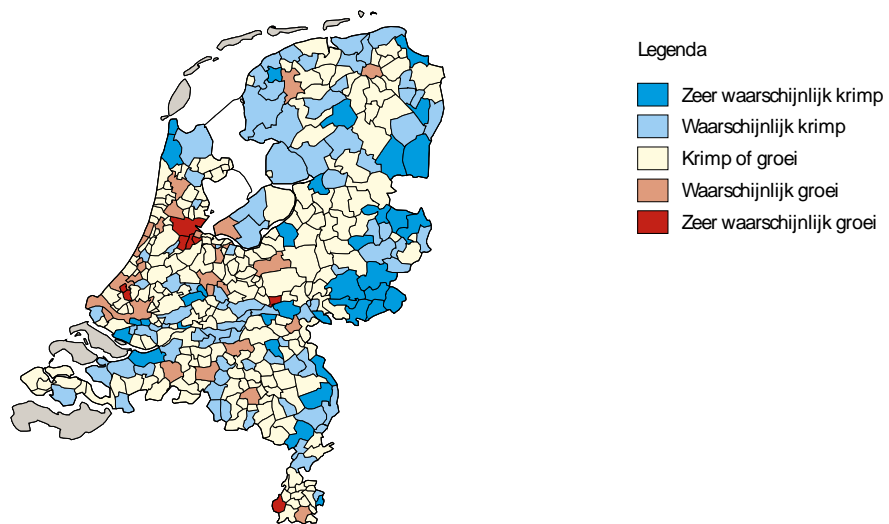


Bron: PBL/CBS regionale bevolkings- en huishoudensprognose 2015-2040

Voor ongeveer 10 procent van de gemeente is het volgens de prognose waarschijnlijk dat het inwonertal in 2040 lager zal zijn dan in 2015. Wordt alleen naar de inwoners in de jongere leeftijden gekeken, dan zijn er veel meer gemeenten waarvoor krimp wordt verwacht. Voor zo'n 30 procent van de gemeenten geldt krimp van het aantal 0-19-jarigen als waarschijnlijk (kaart 5.4). Met name in de kop van Noord Holland, in Friesland, Noord Brabant en de noordelijke Zeeuwse eilanden zijn gemeenten waar de prognose een dalend aantal jongeren voorziet zonder dat duidelijk is of het totale inwonertal ook zal dalen. Voor 60 procent van de gemeenten gaat het aantal 20- tot 65-jarigen waarschijnlijk afnemen (kaart 5.5). Een toenemend aantal 65plussers geldt voor op één na alle gemeenten als waarschijnlijk. Alleen voor Schiermonnikoog ligt een gelijkblijvend aantal ouderen nog nipt binnen de onzekerheidsmarges (al voorziet de prognose ook daar een flinke toename).

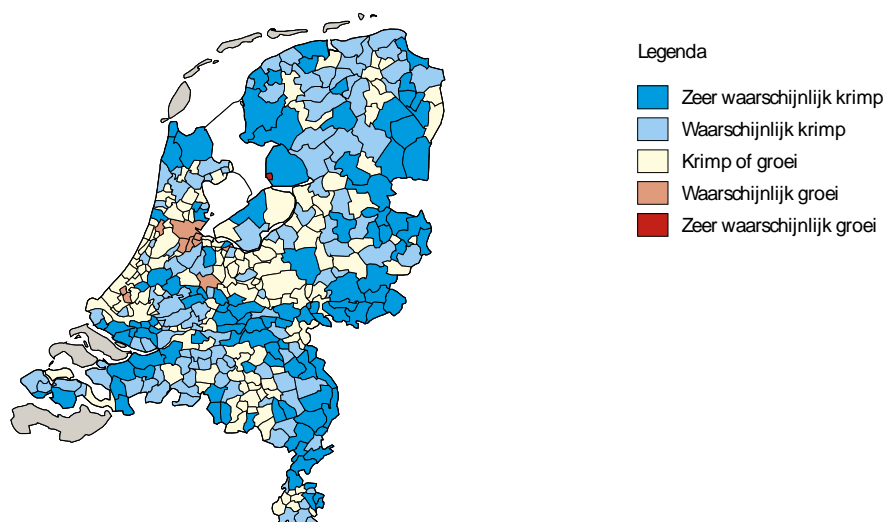
Het aantal gemeenten waar met redelijke zekerheid krimp van het aantal huishoudens wordt verwacht is klein (5 stuks). Dit hangt samen met het feit dat de drijvende kracht achter de bevolkingskrimp, toenemende sterfte door de vergrijzing, in eerste instantie vooral leidt tot een toename van het aantal eenpersoonshuishoudens (als één van de partners in een paar overlijdt) en pas in tweede instantie tot een afname van het aantal huishoudens (als de achtergebleven partner sterft). Wordt niet naar huishoudens in het algemeen gekeken, maar specifiek naar paren, dan zijn er 75 gemeenten waar een afname waarschijnlijk is (kaart 5.6). Slechts voor 27 gemeenten blijkt een toename van het aantal paren waarschijnlijk, vergeleken met 271 gemeenten waarvoor het totale aantal huishoudens waarschijnlijk stijgt.

5.4 Verwachte ontwikkeling aantal 0- tot 20-jarigen per gemeente, 2015-2040



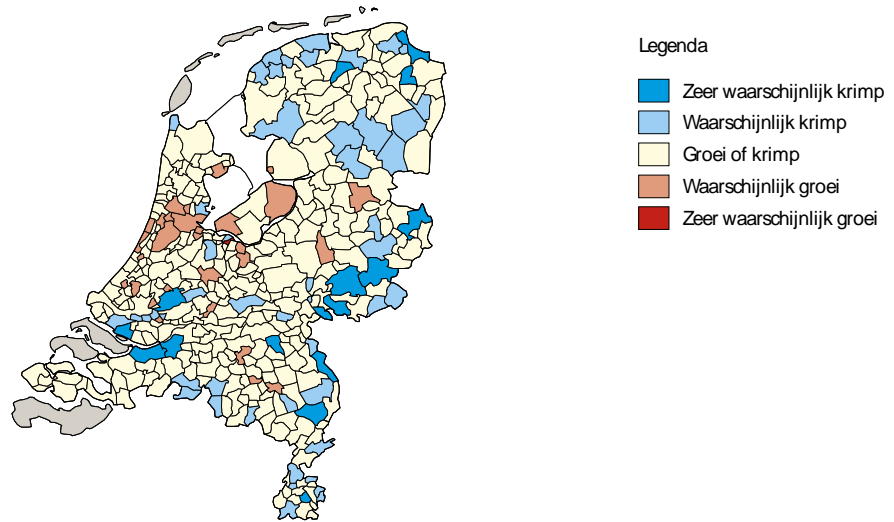
Bron: PBL/CBSregionale bevolkings- en huishoudensprognose 2015-2040

5.5 Verwachte ontwikkeling aantal 20- tot 65-jarigen per gemeente, 2015-2040



Bron: PBL/CBSregionale bevolkings- en huishoudensprognose 2015-2040

5.6 Verwachte ontwikkeling aantal paren per gemeente, 2015-2040



Bron: PBL/CBS regionale bevolkings- en huishoudensprognose 2015-2040

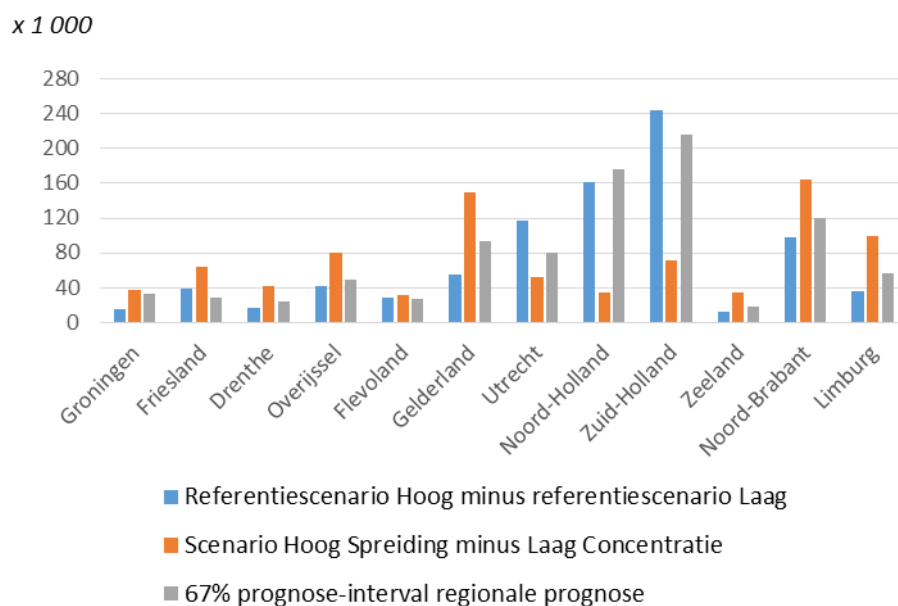
6. Vergelijking met de WLO scenario's van 2015

In het verleden zijn er diverse regionale scenario's gepubliceerd die een aantal alternatieve toekomstige ontwikkelingen beschreven. In de Welvaart en Leefomgeving scenario's (WLO scenario's) van 2015 (Manders en Kool, 2015) zijn twee toekomstbeelden geschetst: één passend bij een sterke en één bij een zwakke welvaartsontwikkeling. In deze scenario's werd een zwakkere welvaartsontwikkeling geassocieerd met een relatief lage immigratie, lage geboorten en hoge sterfte, terwijl bij een sterke welvaartsontwikkeling het omgekeerde werd verondersteld. De WLO-scenario's zijn breder dan demografie alleen, ze beschrijven ook de toekomstige ontwikkelingen op het gebied van macro economie, mobiliteit, klimaat en energie. Hun belangrijkste toepassing is niet het beschrijven van onzekerheid, maar het in kaart brengen van beleidsuitdagingen en het doorrekenen van de effecten van voorgenomen beleid onder verschillende denkbare toekomstige omstandigheden. Door de Jong (2015) is de demografische toekomst van Nederland verkend aan de hand van een sterkere dan wel lagere bevolkingsgroei dan voorzien in de CBS prognose (van 2012). In de regionale scenario's (Ritsema van Eck et al., 2015) is in het referentiescenario Hoog verondersteld dat een hoge nationale bevolkingsgroei samengaat met een sterke concentratie van wonen en werken in de Randstad, waardoor de bevolkingsgroei zich concentreert in de grote steden. In het referentiescenario Laag is verondersteld dat een lagere bevolkingsgroei gepaard gaat met een afzwakkende concentratie, waardoor de bevolkingsgroei meer gelijkmatig over de regio's wordt verspreid. In twee aanvullende onzekerheidsscenario's is gekeken naar de combinatie van hoge bevolkingsgroei met meer bevolkingsverspreiding (scenario Hoog Spreiding) en een lagere bevolkingsgroei met meer concentratie (scenario Laag Concentratie); zie Van Gemeren et al., 2016.

De scenario methodiek brengt met zich mee, dat het niet mogelijk is om waarschijnlijkheden te koppelen aan de WLO scenario's, aangezien de onzekerheid rondom de (regionale) prognose niet kan worden gevangen in slechts vier scenario's. Wel geldt dat op nationaal niveau de bandbreedte van de twee referentiescenario's Hoog en Laag voor het inwonertal en aantal huishoudens precies overeenkomen met het 67% prognose-interval voor deze variabelen van de toen geldende CBS nationale bevolkings- en huishoudensprognoses (zie De Jong et al., 2016). Aangezien de regionale scenario's geaggregeerd naar het landelijk niveau consistent zijn met de nationale scenario's, doet dit vermoeden dat op het regionale niveau de bandbreedte tussen de twee lage scenario's enerzijds en de twee hoge scenario's anderzijds, in zekere mate het twee-derde betrouwbaarheidsinterval weerspiegelen. In dit kader worden hieronder de marge tussen de referentiescenario's Laag en Hoog en Laag en de marge tussen de alternatieve scenario's Laag Concentratie en Hoog Spreiding vergeleken met de marge tussen de ondergrens en de bovengrens van het 67% prognose-interval van de regionale prognose. De regionale scenario's van de WLO zijn enkel zijn gepubliceerd op de regionale niveaus van COROP-gebieden en

provincies, en slechts voor twee zichtjaren, te weten 2030 en 2050. De prognose horizon van de regionale prognose loopt echter tot 2040. Dit betekent dat de vergelijking enkel voor het kalenderjaar 2030 kan worden uitgevoerd.

6.1 Bevolking per provincie in 2030: 67% prognose-interval PBL/CBS regionale prognose, marge tussen WLO Referentiescenario's Laag en Hoog en tussen WLO scenario's Laag Concentratie en Hoog Spreiding



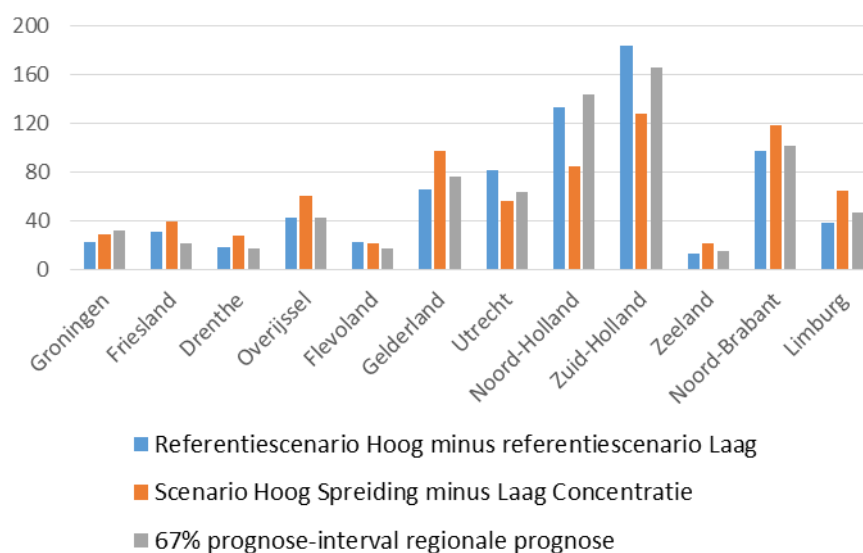
In grafiek 6.1. wordt per provincie nagegaan in hoeverre de bevolking op 2030 varieert volgens het 67% prognose interval van de PBL/CBS regionale prognose en de vier WLO scenario's. Voor de WLO scenario's geldt dat de twee referentiescenario's Laag en Hoog zodanig zijn opgesteld dat de bevolking in de Randstad sterk varieert en die in de overige provincies matig varieert, terwijl in de twee alternatieve scenario's het omgekeerde het geval is. De provincie Flevoland vormt hierbij een uitzondering aangezien de twee marges volgens de vier WLO scenario's vrijwel hetzelfde uitvallen. Gezien de wijze van constructie van de WLO scenario's zou dit betekenen dat in principe het 67% prognose-interval van de regionale prognose tussen de twee marges van de WLO scenario's zou dienen uit te vallen. Grafiek 6.1 laat zien dat dit inderdaad in de meeste provincies het geval is; enkel in de provincies Friesland en Flevoland ligt het 67% prognose-interval wat onder de twee marges volgens de WLO scenario's. Verder valt op dat in Noord-Holland en Zuid-Holland het 67% prognose-interval van de regionale prognose dicht tegen de marge tussen de WLO referentiescenario's Laag en Hoog ligt. Voor het WLO scenario Hoog geldt dat hoge bevolkingsgroei is geconcentreerd in de Randstad, terwijl in het WLO scenario Laag geldt dat de lage bevolkingsgroei veel minder sterk dan nu het geval is in de Randstad terecht komt. Volgens de regionale prognose kennen de provincies Noord-Holland en Zuid-Holland de grootste onzekerheid (in het absoluut niveau van het inwonertal), dat wordt weerspiegeld door de marge tussen de twee WLO referentiescenario's. Verder valt op dat enkel voor de provincie Groningen het 67% prognose-interval van

de regionale prognose dicht bij de marge van de twee alternatieve WLO scenario's ligt.

Niet alleen voor de bevolking, maar ook voor de huishoudens is deze verschil analyse uitgevoerd; zie grafiek 6.2. De bevindingen sporen grotendeels met het bovenbeschreven patroon. Wederom geldt dat de marge van het 67% prognose interval van de regionale prognose dicht bij de marge tussen de twee WLO referentiescenario's ligt. Enkel voor de provincie Groningen ligt de marge van het 67% prognose-interval van de regionale prognose dicht bij de marge tussen de twee alternatieve WLO scenario's.

6.2 Huishoudens per provincie in 2030: 67% prognose-interval PBL/CBS regionale prognose, marge tussen WLO Referentiescenario's Laag en Hoog en tussen WLO scenario's Laag Concentratie en Hoog Spreiding

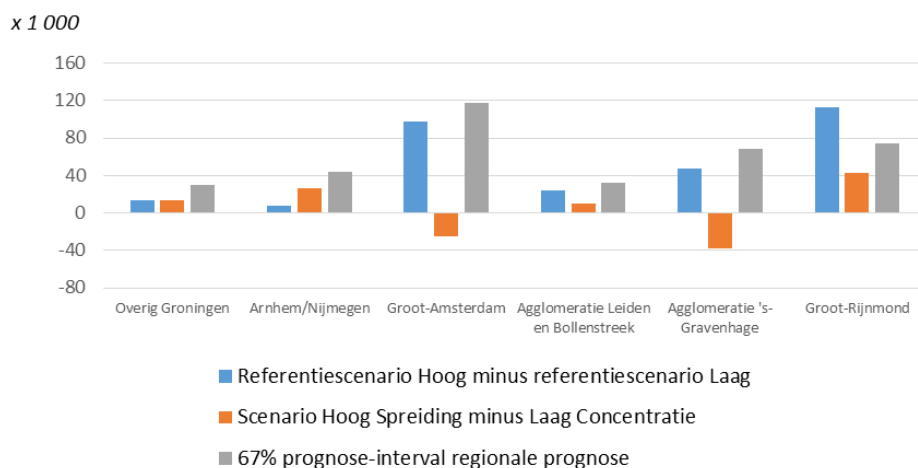
x 1 000



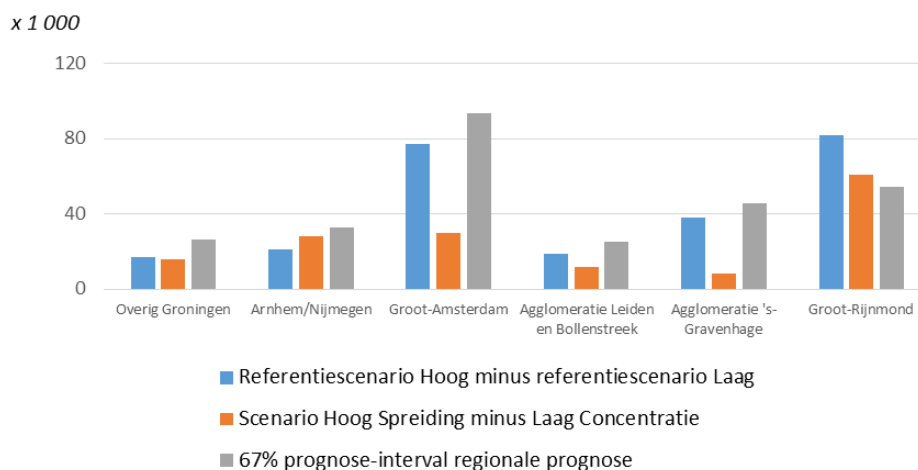
Ook voor de COROP gebieden is de verschil analyse uitgevoerd. Het voert te ver om de verschillen voor alle regio's te tonen. Voor de regio's Overig Groningen (dat ook de gemeente Groningen bevat) en Arnhem/Nijmegen geldt dat de marge van het 67% prognose-interval van de regionale prognose onder de twee marges van de WLO scenario's ligt. De regio Groot-Amsterdam kent in de regionale prognose de grootste marge van het 67% prognose-interval; deze valt zelfs groter uit dan de marge tussen de WLO referentiescenario's Laag en Hoog. Opvallend is dat het inwonertal volgens het scenario Laag Concentratie hoger uitvalt dan volgens het scenario Hoog Spreiding. Dit heeft te maken met de verhaallijnen van deze aanvullende scenario's, waarbij de veronderstelde aantrekkelijkheid van de stad in het eerste scenario veel groter is dan in het tweede. Het patroon van het regio Groot-Amsterdam is ook van toepassing op de Agglomeratie 's-Gravenhage. Ook in de agglomeratie Leiden en Bollenstreek is het 67% prognose-interval groter dan de marge tussen de twee WLO referentiescenario's Laag en Hoog. De regio Groot-Rijnmond kent de grootste marge tussen de twee WLO

referentiescenario's Laag en Hoog; deze marge is zelfs beduidend groter dan het 67% prognose-interval van de regionale prognose.

6.3 Bevolking voor geselecteerde COROP gebieden in 2030: 67% prognose-interval PBL/CBS regionale prognose, marge tussen WLO Referentiescenario's Laag en Hoog en tussen WLO scenario's Laag Concentratie en Hoog Spreiding



6.4 Huishoudens voor geselecteerde COROP gebieden in 2030: 67% prognose-interval PBL/CBS regionale prognose, marge tussen WLO Referentiescenario's Laag en Hoog en tussen WLO scenario's Laag Concentratie en Hoog Spreiding



Voor dezelfde geselecteerde COROP gebieden is nagegaan hoe het 67% prognose-interval en de twee marges wat betreft de aantallen huishoudens uitvallen. De bevindingen sporen grotendeels met het bovenbeschreven patroon voor de bevolking. Wel valt op dat er bij de regio's Groot-Amsterdam en Agglomeratie 's-Gravenhage het aantal huishoudens volgens het scenario Laag Concentratie lager ligt dan volgens het scenario Hoog Spreiding, terwijl voor de bevolking het omgekeerde het geval is. Opmerkelijk is voorts dat de marge volgens het 67% prognose-interval van de regionale prognose onder de twee marges van de WLO scenario's ligt. Dit komt doordat de twee marges van de WLO scenario's voor de huishoudens vrij dicht bij elkaar liggen, terwijl dit bij de bevolking niet het geval is.

7. Conclusies en evaluatie

Met de introductie van de quasi stochastische prognose in de PBL/CBS regionale bevolkings- en huishoudensprognose is het mogelijk geworden om uitspraken te doen over de onzekerheden in de uitkomsten. Het biedt de mogelijkheid om onzekerheidsintervallen voor alle belangrijke outputvariabelen te berekenen. Bij de modellering van de onzekerheid wordt meegenomen dat op hogere regionale niveaus de onzekerheid minder groot is. Op nationaal niveau komen de prognose-intervallen die met de quasi stochastische prognose worden bepaald, vrijwel overeen met de prognose-intervallen van de nationale stochastische prognoses van het CBS. Deze overeenkomst wordt veroorzaakt doordat de veronderstellingen in de quasi stochastische prognose over hoe onzekerheid zich ontwikkelt, consistent zijn met de veronderstellingen die in de stochastische prognose worden gebruikt.

Bij de ontwikkeling van de quasi stochastische methode zijn diverse belangrijke keuzes gemaakt. Vanuit pragmatische overwegingen zijn vereenvoudigingen toegepast. Het is mogelijk de methodiek op die punten verder te ontwikkelen. Zo worden de binnenlandse migratie-varianten nu berekend met een vereenvoudigd model. Het valt te overwegen om de onzekerheid van de binnenlandse migratie intern via PEARL door te rekenen. Het voordeel is dat dit de mogelijkheid biedt om het effect van de beschikbaarheid van woningen op de onzekerheid van de binnenlandse migratie mee te nemen (indien er bijvoorbeeld te weinig woningen beschikbaar zijn in een bepaalde gemeente, dan kan dit de vestiging in deze gemeente afremmen). In het verlengde hiervan, biedt dit ook de mogelijkheid om alternatieve woningbouwveronderstellingen in de berekening van de onzekerheid in te brengen. Het ligt in de bedoeling bij de update van de regionale prognose in 2019 na te gaan in hoeverre een dergelijke verfijning meerwaarde heeft ten opzichte van de huidige methodiek.

Literatuur

Alho, J. en B.D. Spencer (1985) Uncertain population forecasting, *Journal of the American Statistical Association*, 80, pp. 306-314.

Pflaumer, P. (1988) Confidence intervals for population projections based on Monte Carlo methods, *International Journal of Forecasting*, 4, pp. 135-142.

Alders, M. en J. De Beer (2005) An expert knowledge approach to stochastic mortality forecasting in the Netherlands, *Perspectives on Mortality Forecasting. II Probabilistic Models*, 2, 39–60.

Beer, J. de, en M. Alders (1999) Stochastische bevolkingsprognose, *Maandstatistiek van de bevolking* 47 (10), blz. 10–17.

Carolina, N. en C. van Duin (2010) Onzekerheidsmarges voor de sterfteprognose van het CBS, *Bevolkingstrends* 2e kwartaal 2010, blz. 32-37.

Duin, C. van (2017), Quasi stochastic population forecasts, CBS discussion paper 2017|02.

Gemerén, J. van, J. Ristema van Eck, P. Zwanenveld, D. Snellen, G. Verweij, O. van Gerwen (2016) *Toekomstverkenning Welvaart en Leefomgeving*,

Achtergronddocument Regionale ontwikkelingen en verstedelijking, CPB en PBL, PBL publicatienr. 1773.

Jong, A. de, M. Alders, P. Feijten, P. Visser, I. Deerenberg, M. van Huis en D. Leering (2005) *Achtergronden en veronderstellingen bij het model PEARL*. NAI Uitgevers, Rotterdam.

Jong, A. de, J. de Beer en N. van der Gaag (2016) *Toekomstverkenning Welvaart en Leefomgeving, Achtergrondstudie Demografie*, CPB en PBL, PBL publicatienr. 1772.
Kooiman, N, A. de Jong, C. Huisman, C. van Duin en L. Stoeldraijer (2016) *PBL/CBS Regionale bevolkings- en huishoudensprognose 2016-2040: sterke regionale verschillen*, *Bevolkingstrends* 2016|08.

Manders, T. en C. Kool, 2015, *Toekomstverkenning Welvaart en Leefomgeving, Nederland in 2030 en 2050: twee referentiescenario's*, PBL en CPB, PBL publicatienr. 1689.

PBL (2011), *Ruimtelijke Verkenning 2011: Een land van Regio's*. Den Haag: Planbureau voor de Leefomgeving (PBL).

Appendix 1: Schatten van de onzekerheidsmarges uit de uitkomsten van de varianten

In de stochastische bevolkingsprognose worden de veronderstellingen voor vruchtbaarheid, sterfte en migratie onafhankelijk van elkaar getrokken uit een kansverdeling en uitgaand van die veronderstellingen wordt een variant doorgerekend. Door dit een groot aantal keer te doen wordt een kansverdeling voor de uitkomsten van de prognose gegenereerd. Hier wordt iets soortgelijks gedaan, maar daarbij wordt gewerkt met een kansverdeling voor de veronderstellingen die bestaat uit slechts twee waarden: één hoger en één lager dan volgens de prognose, beide met een realisatiekans van ½. Uitgaand van 8 componenten (3 demografische, 3 voor huishoudensontwikkeling, 2 voor binnenlandse migratie) levert dat $2^8=256$ mogelijke combinaties van de veronderstellingen op, en dus een verdeling met 256 verschillende waarden voor elke uitkomst Y van de prognose. De onzekerheidsmarges worden afgeleid uit de gemiddelde afwijking (Bias) en de gemiddelde kwadratische afwijking (MSE) van de uitkomsten in de 256 combinatievarianten ten opzichte van de uitkomsten van de prognose.

$$\text{Bias}(\Delta Y(t)) = \frac{1}{2^8} \sum_{r_1=Laag}^{Hoog} \sum_{r_2=Laag}^{Hoog} \dots \sum_{r_8=Laag}^{Hoog} \Delta \tilde{Y}(r_1, r_2, \dots, r_8, t), \quad (\text{A1.1})$$

$$\text{MSE}(\Delta Y(t)) = \frac{1}{2^8} \sum_{r_1=Laag}^{Hoog} \sum_{r_2=Laag}^{Hoog} \dots \sum_{r_8=Laag}^{Hoog} \Delta \tilde{Y}(r_1, r_2, \dots, r_8, t)^2,$$

waarbij

$$\Delta \tilde{Y}(r_1, r_2, \dots, r_8, t) = \tilde{Y}(r_1, r_2, \dots, r_8, t) - \hat{Y}(t) \quad (\text{A1.2})$$

met

$$\tilde{Y}(r_1, r_2, \dots, r_8, t)$$

de waarde voor uitkomst Y wanneer voor de vruchtbaarheid de aangepaste veronderstellingen met richting r_1 (= hoog of laag) worden gebruikt, voor de levensverwachting met richting r_2 en zo verder. $\hat{Y}(t)$ is de waarde van Y volgens de prognose.

Om de doorlooptijd te beperken worden niet alle combinaties van de veronderstellingen doorgerekend. Voor de demografische en huishoudensdynamiek componenten worden maar één hoge en één lage variant berekend, waarbij alleen de input van die component ten opzichte van de prognosewaarde wordt aangepast. Voor de verhuiscomponenten worden de vier combinaties van hoge en lage mobiliteit met een hoog of laag vestigingssaldo doorgerekend. Deze aanpak reduceert het aantal varianten dat per gemeente moet worden doorgerekend van 256 naar 16. Uit de uitkomsten van deze 16 varianten worden die voor de 256 combinatievarianten geschat onder de aanname dat er weinig interactie is tussen de prognosefouten in de verschillende componenten (behalve tussen de componenten

vestigingssaldo en mobiliteit). ‘Weinig interactie’ betekent dat het effect van een gelijktijdige verandering van de veronderstellingen voor meerdere componenten op de prognose-uitkomsten ongeveer hetzelfde is als de som van de effecten op de uitkomsten van de aanpassingen in elke component afzonderlijk. De geen-interactie aanname is in het algemeen goed wanneer het effect van de aanpassingen in de componenten op het inwonertal klein zijn ten opzichte van het inwonertal. Omdat in de regionale prognose slechts 25 jaar vooruit wordt gekeken, is dat naar verwachting het geval. Omdat de interactie-effecten evenveel combinatievarianten omhoog als omlaag duwen vallen positieve en negatieve interactie-effecten bovendien voor een groot deel tegen elkaar weg in de berekening van de onzekerheidsmarges.¹

Onder de aanname van zwakke interactie geldt:

$$\Delta\tilde{Y}(r_1, r_2, \dots, r_8, t) \approx \sum_{s=1}^6 \Delta\tilde{Y}_s(r_s, t) + \Delta\tilde{Y}_{BM}(r_7, r_8, t), \quad (A1.3)$$

met

$$\Delta\tilde{Y}_s(r_s, t)$$

de uitkomsten voor de demografische en huishoudensdynamiek varianten waarin de veronderstellingen voor component s worden aangepast in de richting rs en

$$\Delta\tilde{Y}_{BM}(r_7, r_8, t)$$

de uitkomsten voor de binnenlandse migratievarianten waarbij de mobiliteit en het vestigingssaldo zijn aangepast in de richtingen r7 en r8.

Invullen van (A1.3) in de uitdrukkingen voor de Bias en MSE (A1.1,A1.2) geeft

$$\text{Bias}(\Delta Y(t)) = \frac{1}{2^8} \sum_{r_1=Laag}^{Hoog} \dots \sum_{r_8=Laag}^{Hoog} \left(\sum_{s=1}^6 \Delta\tilde{Y}_s(r_s, t) + \Delta\tilde{Y}_{BM}(r_7, r_8, t) \right) \quad (A1.4)$$

$$\text{MSE}(\Delta Y(t)) = \frac{1}{2^8} \sum_{r_1=Laag}^{Hoog} \dots \sum_{r_8=Laag}^{Hoog} \left(\sum_{s=1}^6 \Delta\tilde{Y}_s(r_s, t) + \Delta\tilde{Y}_{BM}(r_7, r_8, t) \right)^2,$$

De hoge en lage marges voor Y worden geschat door op te leggen dat ze de Bias en MSE reproduceren.

$$\begin{aligned} \Delta\tilde{Y}_{HOOG}(t) &= \text{Bias}(\Delta Y(t)) + \sqrt{\text{MSE}(\Delta Y(t)) - \text{Bias}(\Delta Y(t))^2}, \\ \Delta\tilde{Y}_{LAAG}(t) &= \text{Bias}(\Delta Y(t)) - \sqrt{\text{MSE}(\Delta Y(t)) - \text{Bias}(\Delta Y(t))^2}. \end{aligned} \quad (A1.5)$$

Voor de bevolkingsprognose met alleen de drie demografische componenten geeft deze werkwijze onzekerheidsmarges voor de standcijfers die sterk lijken op die volgens de stochastische prognose, als van dezelfde inputveronderstellingen als in de stochastische prognose wordt uitgegaan (Van Duin, 2017).

¹ Bij een test op de nationale prognose bleken de verschillen tussen de combinatievarianten berekend met en zonder de geen-interactie aanname inderdaad klein en waren de marges die uit beide soorten combinatievarianten werden berekend vrijwel identiek.

Appendix 2: Afleiden inputveronderstellingen voor de nationale demografische varianten

De veronderstellingen voor de vruchtbaarheid, migratie en het sterftecijfer in de demografische varianten zijn gebaseerd op de 67%-onder en bovengrens voor het kindertal per vrouw, het migratiesaldo en de levensverwachting die worden gebruikt in de nationale stochastische bevolkingsprognose van het CBS. Die onder en bovengrenzen schatten de waarden waarbinnen die drie indicatoren zich in de toekomst waarschijnlijk (met 2/3 zekerheid) zullen bevinden. De tijdreeks voor de afwijking van het vruchtbaarheidscijfer (TFR) en de levensverwachting (e_0) ten opzichte van de prognosewaarde wordt in de stochastische prognose gemodelleerd als een random walk, die afwijking in het migratiesaldo (N) als een AR(1) model

$$\begin{aligned}\Delta TFR(t) &= \Delta TFR(t-1) + \varepsilon_{TFR}(t), \\ \Delta e_0(g, t) &= \Delta e_0(g, t-1) + \varepsilon_{e_0}(t), \\ \Delta N(t) &= \varphi \Delta N(t-1) + \varepsilon_N(t).\end{aligned}\tag{A2.1}$$

waarbij de $\varepsilon(t)$'s ruistertermen zijn die onderling onafhankelijk zijn en ook geen correlatie hebben in de tijd. Ze worden getrokken uit een symmetrische normaalverdeling² met tijdsafhankelijke standaarddeviaties $\sigma(\text{TFR})=0,04$ kind/vrouw, $\sigma(e_0)=0,4$ levensjaar en $\sigma(N)=15$ duizend migranten. De autocorrelatieparameter φ in de tijdreeks voor ΔN beschrijft in hoeverre de afwijking ten opzichte van de prognosewaarde in een jaar doorwerkt in de afwijkingen in de daarop volgende jaren. De random walk modellen voor de vruchtbaarheid en de levensverwachting zijn hetzelfde als een AR(1) model met $\varphi = 1$ (de afwijking werkt volledig door in het volgende jaar). De tijdreeks voor netto migratie in de stochastische prognose 2015-2060 heeft een autocorrelatieparameter van $\varphi = 0,77$.

De 67% onzekerheidsintervallen voor TFR, e_0 en N volgens (A2.1) worden gegeven door

² In de stochastische bevolkingsprognose worden de afwijkingen in het aantal immigranten (naar geboorteland) en de emigratiekansen (naar geboorteland) elk apart als een AR(1) model gemodelleerd met een niet-normale verdeling. Voor de totale netto migratie levert dat echter een tijdreeks op die goed benaderd kan worden door een AR(1) model met een normaal verdeelde ruisterterm.

$$\begin{aligned}
P_{5/6}[\Delta TFR(t_0 + k)] - P_{1/6}[\Delta TFR(t_0 + k)] &= 0,967 \cdot 2\sigma_{TFR} \sqrt{k}, \\
P_{5/6}[\Delta e_0(g, t_0 + k)] - P_{1/6}[\Delta e_0(g, t_0 + k)] &= 0,967 \cdot 2\sigma_{e_0} \sqrt{k}, \\
P_{5/6}[\Delta N(t_0 + k)] - P_{1/6}[\Delta N(t_0 + k)] &= 0,967 \cdot 2\sigma_N \sqrt{\frac{1 - \varphi^{2k}}{1 - \varphi^2}}.
\end{aligned} \tag{A2.2}$$

waarbij $P_x[Y]$ het percentiel van rang $100 \cdot x$ van de verdeling van Y is. De bandbreedte van TFR en e_0 neemt volgens deze veronderstellingen toe met de wortel van de prognoseduur. De bandbreedte van de netto migratie neemt sneller dan dat toe in de eerste prognosejaren, maar groeit op de langere termijn nauwelijks meer.

Voor de twee Mig-varianten worden de volgende inputveronderstellingen gebruikt

$$\begin{aligned}
\tilde{N}(\text{MigHigh}, t) &= \hat{N}(t) + \psi_\varphi(t) P_{5/6}[\Delta N(t)], \\
\tilde{N}(\text{MigLow}, t) &= \hat{N}(t) + \psi_\varphi(t) P_{1/6}[\Delta N(t)],
\end{aligned} \tag{A2.3}$$

waarbij $\hat{N}(t)$ het migratiesaldo volgens de prognose is. De versmallingsfactor $\psi_\sigma(t)$ corrigeert voor het effect dat fluctuaties in het migratiesaldo die na een paar jaar uitmiddelen geen bijdrage leveren aan de onzekerheid in de standcijfers van de bevolking, terwijl ze wel bijdragen aan de bandbreedte van het migratiesaldo. De versmallingsfactor is berekend door op te leggen dat de Mig-varianten het 67% interval voor de gecumuleerde afwijkingen in het migratiesaldo volgens het tijdsreeksmodel (A2.1) moeten reproduceren. Dus

$$\begin{aligned}
\sum_{k=1}^{t-t_0} \Delta \tilde{N}(\text{MigHoog}, t_0 + k) &= P_{5/6} \left[\sum_{k=1}^{t-t_0} \Delta N(t_0 + k) \right] \\
\sum_{k=1}^{t-t_0} \Delta \tilde{N}(\text{MigLaag}, t_0 + k) &= P_{1/6} \left[\sum_{k=1}^{t-t_0} \Delta N(t_0 + k) \right]
\end{aligned} \tag{A2.4}$$

Het idee achter deze eis is dat in de percentielen van de gecumuleerde afwijkingen in het rechterlid het effect van uitmiddelen van fluctuaties is meegenomen.

De eis betekent dat moet gelden dat

$$\sum_{k=1}^{t-t_0} \psi_\varphi(t_0 + k) P_{5/6}[\Delta N(t_0 + k)] = P_{5/6} \left[\sum_{k=1}^{t-t_0} \Delta N(t_0 + k) \right] \tag{A2.5}$$

en een soortgelijke conditie voor de ondergrens. Hieruit kan de volgende uitdrukking voor de versmallingsfactor worden afgeleid (Van Duin, 2017)

$$\psi_\varphi(t) = \frac{\sqrt{\sum_{j=1}^{t-t_0} \left(\sum_{k=j}^{t-t_0} \varphi^{k-j} \right)^2} - \sqrt{\sum_{j=1}^{t-t_0-1} \left(\sum_{k=j}^{t-t_0-1} \varphi^{k-j} \right)^2}}{\sqrt{\sum_{k=0}^{t-t_0-1} \varphi^{2k}}}. \tag{A2.6}$$

In de afleiding is gebruikt dat de ruisterm in de tijdreeks voor N symmetrisch normaal verdeeld is.

Voor een random walk ($\varphi = 1$) vereenvoudigt (A2.6) tot

$$\psi_1(t_0 + T) = \frac{\sqrt{\sum_{k=0}^T (T - k + 1)^2} - \sqrt{\sum_{k=0}^{T-1} (T - k)^2}}{\sqrt{T+1}} \approx \begin{cases} 1; T = 0 \\ 0,87; T > 0 \end{cases} \quad (\text{A2.7})$$

Dus voor een RW model is de versmallingsfactor (vrijwel) constant na het eerste prognosejaar. Als de autocorrelatie parameter kleiner dan 1 is neemt de versmallingsfactor af met de prognoseduur, wat aangeeft dat met elk extra prognosejaar de onzekerheid in het migratiesaldo voor dat jaar steeds minder bijdraagt aan de onzekerheid in de standcijfers (omdat de fluctuaties in het migratiesaldo uitmiddelen). Grafiek A2.1 toont de versmallingsfactor voor verschillende waarde van de autocorrelatieparameter.

De aanpassingen in de input voor het migratiesaldo moet nog worden uitgesplitst naar een aanpassing in de input voor de immigratie (I) en de emigratie (E). Omdat voor Nederland het leeftijds patroon van immigranten en emigranten veel op elkaar lijkt, heeft de manier waarop dit gedaan wordt weinig effect op de resulterende onzekerheidsmarges voor de bevolkingcijfers. Omdat de immigratie grotere jaar-op-jaar fluctuaties laat zien dan de emigratie, is besloten om voor de immigratie de eigen bandbreedte volgens de stochastische prognose als uitgangspunt te nemen en de emigratie vervolgens zo te kiezen dat de gewenste hoge en lage waarde voor de netto migratie wordt verkregen.

$$\begin{aligned} \tilde{I}(\text{MigHoog}, t) &= \hat{I}(t) + \psi_{0,77}(t) P_{5/6}[\Delta I(t)], \\ \tilde{E}(\text{MigHoog}, t) &= \tilde{I}(\text{MigHoog}, t) - \tilde{N}(\text{MigHoog}, t), \\ \tilde{I}(\text{MigLaag}, t) &= \hat{I}(t) + \psi_{0,77}(t) P_{1/6}[\Delta I(t)], \\ \tilde{E}(\text{MigLaag}, t) &= \tilde{I}(\text{MigLaag}, t) - \tilde{N}(\text{MigLaag}, t). \end{aligned} \quad (\text{A2.8})$$

Voor de Geb-varianten is opgelegd dat ze het 67%-interval van de gecumuleerde afwijkingen in het totale vruchtbaarheidscijfer volgens de tijdreeks moeten reproduceren. Dat geeft

$$\begin{aligned} T\tilde{F}R(\text{GebHoog}, t) &= T\hat{F}R(t) + \psi_1(t) P_{5/6}[\Delta TFR(t)], \\ T\tilde{F}R(\text{GebLaag}, t) &= T\hat{F}R(t) + \psi_1(t) P_{1/6}[\Delta TFR(t)]. \end{aligned} \quad (\text{A2.9})$$

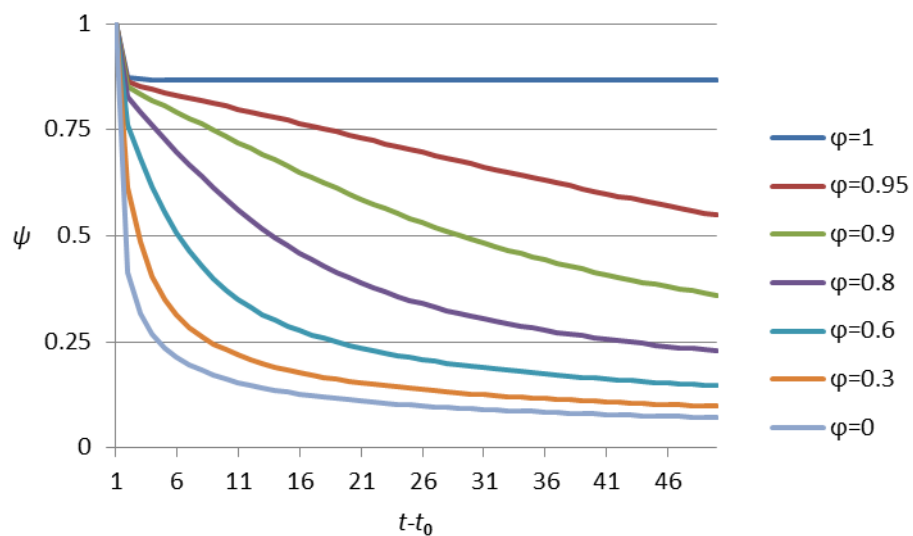
Voor de Sterfte-varianten wordt gebruikt

$$\begin{aligned} \tilde{F}(\text{SterfteHoog}, g, t) &= 1 + \psi_1(t) P_{5/6}[\Delta F(g, t)], \\ \tilde{F}(\text{SterfteLaag}, g, t) &= 1 + \psi_1(t) P_{1/6}[\Delta F(g, t)], \end{aligned} \quad (\text{A2.10})$$

waarbij F de leeftijdsafhankelijke aanpassingsfactor is die, wanneer toegepast op de sterftetekansen uit de prognose, de levensverwachting volgens de variant reproduceert. De prognosewaarde van F is dus gelijk aan 1.

Voor een uitgebreidere onderbouwing van de veronderstellingen (A2.9) en (A2.10), zie (Van Duin, 2017).

A2.1 Versmallingsfactor voor een AR(1) model met verschillende waarden van de autocorrelatie parameter



Appendix 3: Toevoegen extra regionale onzekerheid aan de nationale varianten

Op het niveau van gemeenten is de onzekerheid in de demografische kerncijfers groter dan op nationaal niveau, omdat tegengestelde ontwikkelingen in verschillende regio's op nationaal niveau uit kunnen middelen. Om dit te modelleren worden de gemeentelijke boven- en ondergrenzen in de varianten relatief ruimer ingesteld dan de nationale. Als voorbeeld: in de hoge TFR-variant van de regionale prognose gebruiken we voor het vruchtbaarheidscijfer van gemeente c in jaar t .

$$\begin{aligned} \tilde{TFR}(\text{GebHoog}, c, t) = & \tilde{TFR}(\text{GebHoog}, t) \left(\frac{\hat{TFR}(c, t)}{\widehat{TFR}(NL, t)} \right) \\ & + \Delta \tilde{TFR}(\text{GebHoog}, c, t), \end{aligned} \quad (\text{A3.1})$$

waarbij de laatste term rechts de bijdrage van extra fluctuaties in de gemeentelijke TFR modelleert. De variant met input volgens (A3.1) wordt de 'aangeklede' hoge variant voor de TFR genoemd, die waarin de eerste term in het rechterlid als input wordt genomen de 'kale' variant. In de kale variant is de relatieve bandbreedte van de TFR voor alle gemeenten hetzelfde als in de nationale variant, in de aangeklede variant niet meer.

Om de omvang van de regionale fluctuaties in vruchtbaarheid, sterfte en migratie te schatten is gekeken naar de voorspelfout voor die kernindicatoren als we aannemen dat de nationale en gemeentelijke indicator een vaste verhouding hebben (constante regionale verschillen). Dit is gedaan voor gemeentelijke tijdreeksen over de periode 2005-2013. Hoewel dit een korte periode is, is er door het grote aantal gemeenten toch een flinke massa om de fluctuaties op te schatten. De voorspelfouten zijn vervolgens gefit aan een tijdreeksmodel. Daarna is middels regressie de standaarddeviatie van de ruisterm in het tijdreeksmodel gerelateerd aan eigenschappen van de gemeente die de mate van onzekerheid beïnvloeden, zoals het inwonertal.

Voor het schatten van de regionale fluctuaties is uitgegaan van het volgende model

$$K(c, t) = K(t) \rho(c) e^{\varepsilon(c, t)}, \quad (\text{A3.2})$$

waarbij K de kernindicator is (TFR, het gestandaardiseerde sterftecijfer m , het aantal immigranten of emigranten) en $\rho(c)$ de gemiddelde verhouding tussen de waarde voor gemeente c en voor Nederland. De ruis-term $\varepsilon(c, t)$ is normaal verdeeld, met gemiddelde nul en standaard deviatie $\sigma(c)$. De nationale kernindicator $K(t)$ wordt in deze analyse bekend verondersteld, aangezien de onzekerheid daarin al via de

nationale hoge- en lage veronderstellingen wordt meegenomen. De tijdreeks van $K(c,t)$ kan worden geschreven als.

$$\ln(f(c,t)) = \varepsilon(c,t), \quad (\text{A3.3})$$

waarbij f de factor is waarmee de kernindicator wordt aangepast ten opzichte van de “prognosewaarde” $\rho(c) K(t)$.

De tijdreeksen voor de regionale fluctuaties in TFR en m laten geen structurele autocorrelatie zien, daarom wordt de regionale ruis term $\varepsilon(c,t)$ voor verschillende jaren onafhankelijk verondersteld. Bij gemeenten met een bovengemiddeld aandeel inwoners met een eerste generatie migratieachtergrond wordt in de tijdreeksen voor de immigratiefluctuaties wel autocorrelatie gevonden. Deze zijn gemodelleerd als AR(1) reeksen met een autocorrelatieparameter $\frac{1}{2}$. Voor gemeenten met een benedengemiddeld aandeel eerste generatie allochtonen werd geen sterke autocorrelatie gevonden. De regionale fluctuaties in de emigratie zijn positief gecorreleerd met die in de immigratie. Die samenhang is goed te verklaren, aangezien veel immigranten al na een kort verblijf weer emigreren. De tijdreeksen voor de regionale fluctuaties zijn gecorrigeerd voor deze samenhang. Na die correctie werd geen autocorrelatie gevonden.

Voor de verschillende componenten is geschat hoe de bandbreedte van de regionale fluctuaties samenhangt met eigenschappen van de gemeente (tabel A3.1). Voor TFR en m is gefit aan het model

$$\ln \sigma(c) = A_0 + A_1 \ln P(c) + A_2 \ln \rho(c) + \chi(c), \quad (\text{A3.4})$$

waarbij $P(c)$ het inwonertal van de gemeente is en $\rho(c)$ de regionale factor, oftewel de verhouding tussen de gemeentelijke en nationale waarde voor de kernindicator. $\chi(c)$ is een normaal verdeelde, symmetrische ruisterm.

De standaarddeviatie voor de fluctuaties in de migratie zijn gefit aan

$$\begin{aligned} \ln \sigma_I(c) &= B_0 + B_1 \ln P_{1eGen}(c) + B_2 \ln CI(c) + B_3 \Theta_{\%1eGen} + \chi_I(c), \\ \ln \sigma_{E_{onafh}}(c) &= B_0 + B_1 \ln P_{1eGen}(c) + B_2 \ln CI(c) + \chi_{E_{onafh}}(c), \end{aligned} \quad (\text{A3.5})$$

waarbij $P(1eGen)$ het aantal inwoners met een migratieachtergrond van de eerste generatie is en CI de concentratie-index voor de immigratie: de verhouding tussen het jaarlijks aantal immigranten per inwoner in de gemeente en in Nederland. $\Theta_{\%1eGen}$ is een dummyvariabele met waarde 1 voor gemeenten met een bovengemiddeld aandeel inwoners met een eerste generatie migratieachtergrond en 0 voor de overige gemeenten.

De bandbreedte van de regionale fluctuaties in het vruchtbaarheids- en sterftcijfer neemt af met de inverse wortel van het inwonertal (tabel A3.1). De regionale fluctuaties in de immi- en emigratie hangen sterker samen met het aantal inwoners

met een eerste generatie migratieachtergrond dan met het totale aantal inwoners. Meer eerste generatie inwoners betekent minder fluctuaties. Een hogere regionale factor voor vruchtbaarheid of sterfte, dus een relatief hoog kindertal of lage levensverwachting, hangt samen met minder regionale fluctuaties in vruchtbaarheid of sterfte. Een hoge concentratie-index voor immigratie hangt juist samen met sterkere fluctuaties in zowel de immi- als de emigratie.

Voor de huishoudensovergangen kan in principe een soortgelijke procedure gevolgd worden, waarbij de gestandaardiseerde overgangsfrequenties tussen huishoudensposities per gemeente als kernindicator worden gebruikt. Deze analyse is voor de regionale prognose van 2015 nog niet gedaan. In plaats daarvan is als aanname dat de onzekerheid in de huishoudensovergangen dezelfde afhankelijkheid van de gemeentegrootte heeft als de onzekerheid in het vruchtbaarheidscijfer.

A3.1. Effect van gemeentelijke kenmerken op de omvang van de regionale fluctuaties

vruchtbaarheid	regionale factor	-0.8**
	inwonertal	-0.5**
sterfte	regionale factor	-0.4*
	inwonertal	-0.5**
immigratie	hoog aandeel eerstegeneratieallochtonen (dummy)	-0.3**
	concentratie-index immigratie	0.3**
	aantal eerstegeneratieallochtonen	-0.3**
emigratie†	concentratie-index immigratie	0.1*
	aantal eerstegeneratieallochtonen	-0.3*

significantiëniveau: *=0,05; **=0,01

† exclusief door immigratie verklaarde deel van de fluctuaties

Uitgaand van de hierboven beschreven tijdreeksen voor de regionale fluctuaties wordt rond de kale hoge of lage variant voor elke component een verhoogde en verlaagde variant geconstrueerd waarin de extra bandbreedte door regionale onzekerheid is meegenomen. Dit gebeurt op dezelfde manier als waarop de input voor de nationale varianten werd afgeleid uit de input voor de stochastische prognose (Appendix 2). Het effect dat de regionale fluctuaties in de tijd uitmiddelen wordt meegenomen middels een versmallingsfactor. Voor vruchtbaarheid en sterfte geeft deze procedure per gemeente vier varianten: één voor elke hoog/laag combinatie van nationale en regionale afwijkingen in de input. Voor het migratiesaldo geeft het 8 varianten, 2 keer zoveel omdat zowel de regionale fluctuaties in de immigratie als in de emigratie worden meegenomen. Met de methode die in Appendix 1 is beschreven wordt uit de 4 of 8 varianten één boven- en één ondermarge voor elke component afgeleid. De input van de aangeklede hoge en lage variant voor de gemeente is vervolgens op deze boven- en ondermarge gebaseerd.

Appendix 4: Schatten van de onzekerheid in de binnenlandse migratie

Het vereenvoudigde model waarmee de binnenlandse migratievarianten worden doorgerekend gebruikt als input het bruto vestigings- en vertrekcijfer, oftewel het aantal binnenlandse migranten dat zich in de gemeente vestigt of eruit vertrekt gedeeld door het aantal inwoners. Om de onzekerheid in de inputwaarden voor die prognose te schatten is gekeken naar de voorspelfouten van een eenvoudige projectie voor de jaren 2006-2014, waarbij ervan is uitgegaan dat beide cijfers voor elke gemeente constant zouden blijven op de waarde van 2005. Dit geeft waarschijnlijk een kleine overschatting van de onzekerheid. Doordat de prognose de invloed van de ontwikkeling van het leeftijdsprofiel van de bevolking in de gemeente zelf en in de omgeving op de binnenlandse migratie modelleert en informatie benut over plannen voor woningbouw en daarvoor afgegeven vergunningen, zou die beter dan deze projectie moeten kunnen presteren. Het eerste effect zal naar verwachting echter klein zijn, omdat we hier naar korte tijdreeksen kijken, waardoor de leeftijdsamenstelling niet veel tijd heeft om te veranderen. Voor wat betreft de woningbouw kan worden opgemerkt dat die in de periode 2006-2014 grote schommelingen liet zien en dat de veronderstellingen als gevolg daarvan in opeenvolgende prognoses sterk zijn bijgesteld.

De tijdreeksen voor het vestigings- en vertrekcijfers zijn sterk positief gecorreleerd. In jaren dat veel mensen uit een gemeente vertrekken vestigen zich er ook veel. Dit ligt voor de hand, aangezien gemeenten een redelijk vast aantal plekken voor inwoners hebben dat slechts geleidelijk veranderd door nieuwbouw, sloop, of veranderingen in de woningbezetting. Binnenlandse migratie is daardoor vooral een proces van doorstroming, wat een sterke correlatie tussen het aantal vertrekkers en vestigers afdwingt. Dit betekent dat in de binnenlandse migratie varianten de componenten vestiging en vertrek niet onafhankelijk van elkaar gevarieerd mogen worden, omdat dat te wijde bandbreedtes in de uitkomsten zou geven. In plaats daarvan is met twee andere indicatoren voor de binnenlandse migratie gewerkt waarvoor de tijdreeksen vrijwel ongecorreleerd zijn: het ratio en product van het vestigings- en vertrekcijfer

$$\begin{aligned} \text{ratio}(c,t) &= \frac{a(c,t)}{d(c,t)} \\ \text{prod}(c,t) &= a(c,t)d(c,t) \end{aligned} \tag{A4.1}$$

Het ratio is het aantal vestigers per vertrekkers, dus een maat voor het saldo van de binnenlandse migratie. Het product is het kwadraat van de (geometrisch) gemiddelde mobiliteit $\sqrt{a(c)d(c,t)}$: een maat voor de intensiteit van de verhuisstromen over de gemeentegrens.

De afwijkingen in ratio en prod ten opzichte van de waarde in 2005 zijn gemodelleerd als een AR(1) tijdreeks met een log-normaal verdeelde storingsterm

$$\begin{aligned}\Delta \ln \text{ratio}(c,t) &= \varphi \Delta \ln \text{ratio}(c,t-1) + \varepsilon_{\text{ratio}}(c,t), \\ \Delta \ln \text{prod}(c,t) &= \varphi \Delta \ln \text{prod}(c,t-1) + \varepsilon_{\text{prod}}(c,t),\end{aligned}\tag{A4.2}$$

waarbij de Δ aangeeft dat het om het verschil met de waarde in 2005 gaat. De tijdreeksen laten enige autocorrelatie zien, maar minder sterk van voor internationale migratie. Voor de binnenlandse migratie reeksen is $\varphi = 0,5$ verondersteld ($\varphi = 0,77$ voor internationale migratie). Bij de binnenlandse migratie is er dus meer sprake van tijdelijke fluctuaties die voor een belangrijk deel uitmiddelen dan bij buitenlandse migratie.

De standaarddeviaties zijn gefit aan de volgende regressiemodellen

$$\begin{aligned}\ln \sigma_{\text{ratio}}(c) &= A_0 + A_1 \ln P(c) + A_2 \Theta_{<50K} + A_3 \Theta_{g4} + \chi_{\text{ratio}}(c), \\ \ln \sigma_{\text{product}}(c) &= B_0 + B_1 \Theta_{<50K} \ln P(c) + B_2 (1 - \Theta_{<50K}) \ln P(c) \\ &\quad + B_3 \Theta_{<50K} + B_4 \Theta_{g4} + \chi_{\text{product}}(c),\end{aligned}\tag{A4.3}$$

waarbij de dummyvariabele “<50k” selecteert voor gemeenten met minder dan 50 duizend inwoners en “g4” voor de grootste 4 gemeenten.

De standaarddeviaties voor de ruistermen in de tijdreeksen van ratio en prod nemen af met het inwonertal van de gemeente (tabel A4.1). De relatieve onzekerheid in de mobiliteit en het vestigingsratio is dus kleiner voor grotere gemeenten.

A4.1 Effect van het inwonertal op de onzekerheid in de binnenlandse migratiestromen

vestigingsratio	inwonertal	-0.2**
	minder dan 50 duizend inwoners (dummy)	0.2**
	g4-gemeente (dummy)	0.4*
mobiliteit	inwonertal, voor minder dan 50 duizend inwoners	-0.2**
	inwonertal, voor meer dan 50 duizend inwoners	-0.5**
	minder dan 50 duizend inwoners (dummy)	-2.9*
	g4-gemeente (dummy)	0.4

significantienniveau: *=0,05; **=0,01

Verklaring van tekens

Niets (blanco)	Een cijfer kan op logische gronden niet voorkomen
.	Het cijfer is onbekend, onvoldoende betrouwbaar of geheim
*	Voorlopig cijfer
**	Nader voorlopige cijfer
2015–2016	2015 tot en met 2016
2015/2016	Het gemiddelde over de jaren 2015 tot en met 2016
2015/'16	Oogstjaar, boekjaar, schooljaar enz., beginnend in 2015 en eindigend in 2016
2013/'14–2015/'16	Oogstjaar, boekjaar, enz., 2013/'14 tot en met 2015/'16

In geval van afronding kan het voorkomen dat het weergegeven totaal niet overeenstemt met de som van de getallen.

Colofon

Uitgever
Centraal Bureau voor de Statistiek
Henri Faasdreef 312, 2492 JP Den Haag
www.cbs.nl

Vormgeving: Centraal Bureau voor de Statistiek, Studio BCO
Ontwerp: Edenspiekermann

Inlichtingen
Tel. 088 570 70 70, fax 070 337 59 94
Via contactformulier: www.cbsl.nl/infoservice

© Centraal Bureau voor de Statistiek, Den Haag/Heerlen, 2017.
Verveelvoudigen is toegestaan, mits het CBS als bron wordt vermeld.