



UNIVERSITEIT VAN AMSTERDAM
Amsterdam School of Economics

GRENZELOZE VERZORGINGSSTAAT

DE GEVOLGEN VAN IMMIGRATIE

VOOR DE OVERHEIDSFINANCIËN

TECHNISCHE APPENDIX



Jan van de Beek, Hans Roodenburg, Joop Hartog en Gerrit Kreffer

Foto omslag: Dariusz Jemielniak / Wikimedia commons

Versie: 12-02-2021.3

Copyright © 2021, Zeist: demo-demo uitgever

http://www.demo-demo.nl/files/Grenzeloze_Verzorgingsstaat.pdf

http://www.demo-demo.nl/files/Grenzeloze_Verzorgingsstaat_Technische_Appendix.pdf

Resultaten gebaseerd op eigen berekeningen Universiteit van Amsterdam, Amsterdam School of Economics, op basis van niet-openbare microdata van het Centraal Bureau voor de Statistiek. Deze microdata zijn onder voorwaarden voor statistisch en wetenschappelijk onderzoek toegankelijk. Voor nadere informatie microdata@cbs.nl.

Inhoud

1	Inleiding	5
2	Generatierekening.....	6
2.1	Generatierekening en nettobijdrage.....	6
2.2	Inkomsten en uitgaven van de overheid.....	8
2.3	Generatierekening voor de eerste generatie.....	10
2.4	Generatierekening voor de tweede generatie.....	14
3	Onderzoekspopulatie en microdatabestanden.....	25
4	Socio-demografische variabelen	27
4.1	Leeftijd en generatie	27
4.2	Opleidingsniveau en citoscores.....	28
4.3	Migratiemotief	30
4.4	Regio.....	32
5	Kosten en baten posten	36
5.1	Kosten en baten posten	36
5.2	Veiligheidszorg	38
5.3	Onderwijs en Bruto investeringen scholen	40
5.4	Zorg.....	47
5.5	Uitkeringen, subsidies, premies en belastingen.....	49
5.6	IRN	52
5.7	Vermogen-gerelateerde belastingen	53
6	Aanloopkosten en kosten na remigratie.....	59
6.1	Asielopvang en vergunningsverstrekking.....	59
6.2	Inburgering	60
6.3	AOW-rechten.....	63
7	Demografie en vergrijzing	66
7.1	Geboorte en sterfte.....	66
7.2	Migratie	68
7.3	Vergrijzing.....	69
7.4	Demografie.....	70
8	Verankering in CPB-studies	72
8.1	Verankering in CPB-vergrijzingsstudie <i>Minder zorg om morgen</i> (2014).....	72
8.2	IJking op CPB-vergrijzingsstudie <i>Zorgen om morgen</i> (2019).....	74
8.3	Vergelijking met CPB-rapport <i>Immigration and the Dutch Economy</i> (2003)	77

9	Belangrijke punten bij de interpretatie van de resultaten.....	79
9.1	Beschrijving is geen toetsing, samenhang is geen oorzakelijk verband.....	79
9.2	Eén peiljaar: noodzaak van herhaling	79
10	Sampling	81
11	Sterke en zwakke punten en suggesties voor verder onderzoek.....	89
11.1	Sterke punten.....	89
11.2	Zwakke punten.....	89
11.3	Suggesties voor verder onderzoek.....	90
	Errata.....	91

1 Inleiding

Dit is de Technische appendix bij het rapport *Grenzeloze verzorgingsstaat: De gevolgen van immigratie Voor de overheidsfinanciën*.¹ Dit rapport wordt in deze appendix om verwarring te voorkomen aangeduid als ‘het huidige rapport’. In deze appendix wordt de gebruikte methode toegelicht. Dit heeft als doel om het begrijpen en beoordelen van het huidige rapport te vergemakkelijken. Daarnaast schept het in beginsel² de mogelijkheid tot replicatie van de berekeningen. Een achterliggende doel van het huidige rapport is immers het periodiek produceren van actuele generatierekeningen.

Deze appendix is als volgt opgebouwd. Hoofdstuk 2 handelt over de methode van generatierekening. In §2.1 wordt eerst een algemene toelichting gegeven op de gebruikte methode van generatierekening. In §2.2 wordt een illustratie gegeven van de opbouw van de nettobijdrage per levensjaar op basis van de belangrijkste inkomsten en uitgaven van de overheid. In §2.3 wordt uitgelegd hoe generatierekening is toegepast voor eerste generatie immigranten. In §2.4 wordt hetzelfde gedaan voor tweede generatie immigranten.

In Hoofdstuk 3 wordt de onderzoekspopulatie beschreven en de belangrijkste CBS-microdatabestanden die gebruikt zijn voor het onderzoek. In Hoofdstuk 4 worden enkele sociaal-demografische variabelen beschreven als leeftijd, generatie, opleidingsniveau, citoscore, migratiemotief en herkomstregio. Hoofdstuk 5 bevat een algemene beschrijving van de operationalisering van de kosten- en batenposten. Na een algemeen overzicht van de kosten- en batenposten (§5.1) volgt detailuitleg over de operationalisering van afzonderlijke posten (§5.2 t/m §5.8). Hoofdstuk 6 handelt over ‘aanloopkosten’ – kosten die optreden aan het begin van het migratieproces en kosten die na eventuele remigratie optreden. §6.1 en §6.2 handelen over de aanloopkosten voor asielopvang, vergunningsverstrekking en inburgering die kunnen spelen bij eerste generatie immigranten. §6.3 handelt over de kosten na remigratie van eerste generatie immigranten, i.h.b. opgebouwde AOW-rechten.

Hoofdstuk 7 handelt over de operationalisering van geboorte, sterfte en migratie. Tevens worden een aantal demografische berekeningen toegelicht die betrekking hebben op bevolkingsgroei en vergrijzing. In Hoofdstuk 8 wordt de verankering in de CPB-vergrijzingsstudies toegelicht en wordt een vergelijking gemaakt met het CPB-rapport *Immigration and the Dutch economy* waarvan het huidige rapport een actualisatie is.

Hoofdstuk 9 t/m 12 (nog in revisie) bevatten enkele uiteenlopende zaken. Hoofdstuk 9 gaat in op enkele belangrijke punten m.b.t. de interpretatie van de resultaten. In Hoofdstuk 10 wordt de sampling besproken. In Hoofdstuk 11 worden de sterke en zwakke punten besproken en enkele suggesties voor verder onderzoek gedaan.

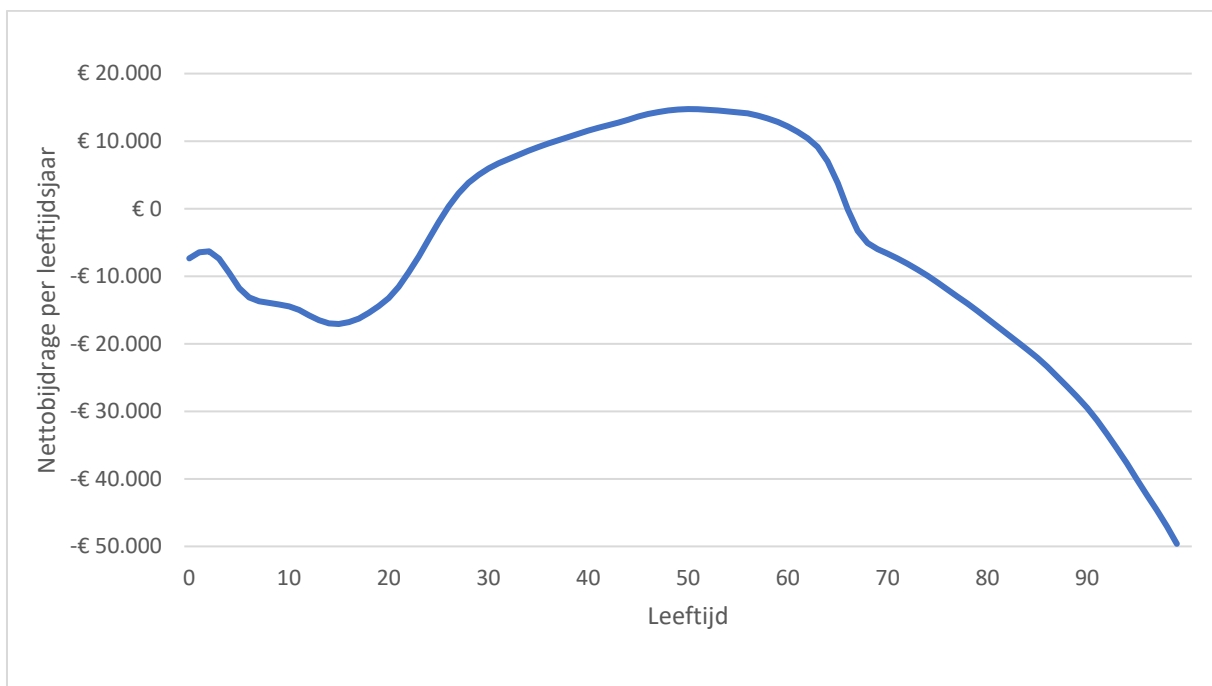
¹ Van de Beek, J. H., J. Hartog, H. Roodenburg & G. Kreffer (2021) *Grenzeloze verzorgingsstaat: De gevolgen van Immigratie voor de Overheidsfinanciën*, ISBN nummer: 9789090344034, http://www.demo-demo.nl/files/Grenzeloze_Verzorgingsstaat.pdf

² Voor replicatie van de resultaten van het onderzoek *Grenzeloze verzorgingsstaat* is toegang nodig tot CBS-microdata. De resultaten zijn gebaseerd op eigen berekeningen Universiteit van Amsterdam, Amsterdam School of Economics, op basis van niet-openbare microdata van het Centraal Bureau voor de Statistiek. Deze microdata zijn onder voorwaarden voor statistisch en wetenschappelijk onderzoek toegankelijk. Voor nadere informatie microdata@cbs.nl.

2 Generatierekening

2.1 Generatierekening en nettobijdrage

De in dit rapport gebruikte methode om de kosten en baten van migratie te berekenen staat bekend als generatierekening.³ De essentie van generatierekening⁴ is dat men voor personen van een bepaalde groep voor elk leeftijdsjaar alle bedragen optelt die een persoon gemiddeld afdraagt aan de overheid en daar alle bedragen vanaf trekt die deze persoon ontvangt van de overheid. Door dat te doen ontstaat een leeftijdsprofiel voor de *nettobijdrage per leeftijdsjaar*. Optelling van die nettobijdragen over alle (toekomstige) levensjaren geeft de nettobijdrage over de levensloop van leden van de betreffende groep.



Figuur 2.1 Nettobijdrage per leeftijdsjaar van de gemiddelde Nederlander in 2016.

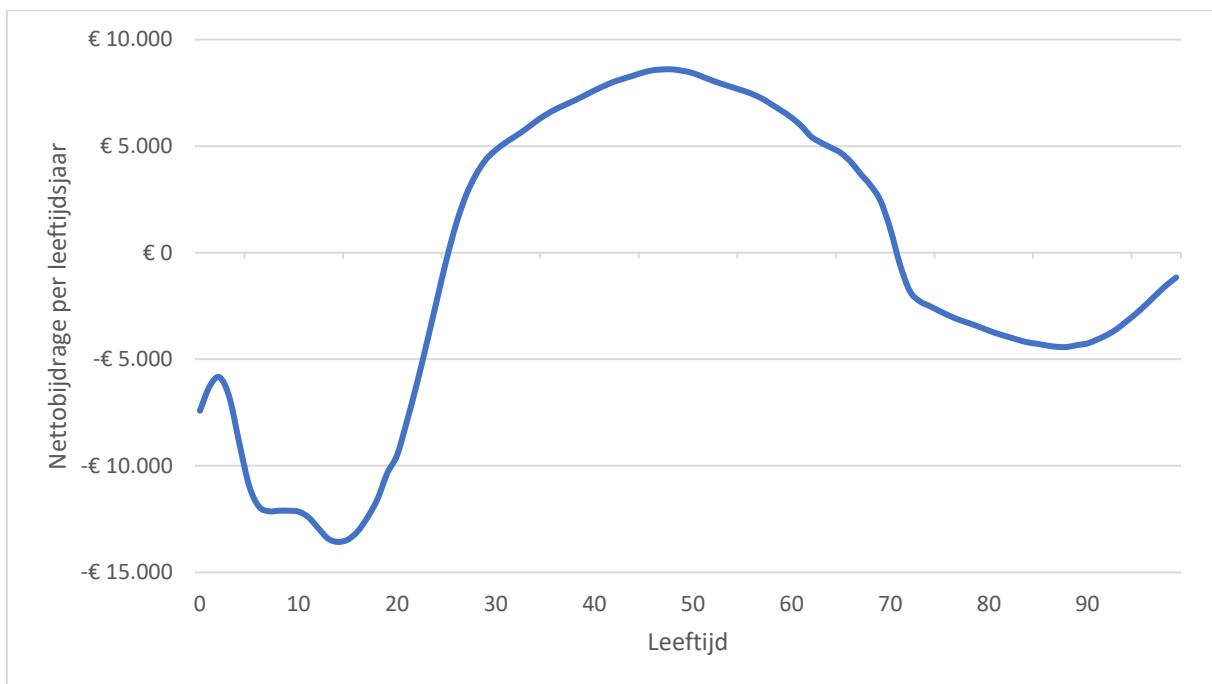
Nettobijdrage per leeftijdsjaar is een sleutelbegrip in deze studie. Het is de hoeveelheid geld die per saldo in een bepaald leeftijdsjaar van een persoon naar de schatkist vloeit. Draagt iemand in een bepaald leeftijdsjaar meer bij aan de schatkist dan hij of zij ontvangt, dan is de nettobijdrage positief in dat leeftijdsjaar. Kost iemand in een bepaald leeftijdsjaar de schatkist meer dan hij of zij bijdraagt, dan

³ Zie ook Hoofdstuk 3 en §4.1, in het bijzonder Box 4.1 en de Begrippenlijst in Van de Beek, J. H., J. Hartog, H. Roodenburg & G. Kreffer (2021)

⁴ Het woord 'generatie' in 'generatierekening' kan mogelijk tot verwarring leiden. Oorspronkelijk was deze methode ontworpen om fiscale onevenwichtigheden tussen verschillende geboortegeneraties te onderzoeken, vandaar de naam 'generational accounting'. Het ging daarbij toen dus niet om generatie in de betekenis van migratiegeneratie, maar van generaties in de bevolking in algemene zin. Deze methode kan echter ook gebruikt worden voor een dynamische benadering (zie §3.1) van de kosten en baten van immigratie, zoals in huidige studie wordt gedaan. Om met het oog op uitsplitsing naar herkomstgroepen, migratiemotief e.d. over zoveel mogelijk data te beschikken is over het algemeen niet per geboortegeneratie berekend, maar zijn alle data gebruikt. In de praktijk beslaat in een aantal gevallen wel een geboortecohort, zoals bij de nettobijdrage naar citoscore (cohort 2006-2018) en migratiemotief (cohort 1995-2015), maar dat is ingegeven door de databeschikbaarheid. Verder is in een aantal gevallen een indeling gemaakt naar migratiegeneratie, volgens de definities van het CBS (zie §13.7).

is de nettobijdrage negatief in dat leeftijdsjaar. Om allerlei praktische redenen is het erg lastig om grote groepen personen over hun gehele leven te volgen. Daarom is in deze studie gebruik gemaakt van de gegevens voor 2016 van mensen van alle leeftijden om de leeftijdsprofielen te creëren voor de nettobijdrage over de gehele levensloop.

Over het algemeen wordt een nettobijdrage profiel opgesteld voor een groep personen. In Figuur 2.1 is het leeftijdsprofiel voor de hele Nederlandse bevolking in 2016 weergegeven. Te zien is dat het profiel onder de nullijn ligt tijdens de jeugd- en studietijd en tijdens de pensioenleeftijd. Gedurende deze perioden is de nettobijdrage per leeftijdsjaar van de gemiddelde Nederlander dus negatief en kost hij of zij de overheid geld. Voor de werkzame leeftijd ligt de nettobijdrage boven de nullijn. Gedurende deze periode levert de gemiddelde Nederlander dus een positieve nettobijdrage en levert hij of zij de schatkist geld op. In 2016 – zoals gezegd het peiljaar van deze studie – was de bijdrage positief voor leeftijden van het 26^e tot en met 65^e levensjaar en negatief voor de overige leeftijden. Door alle bedragen over de hele levensloop – dus voor elk leeftijdsjaar – op te tellen ontstaat de *nettobijdrage*⁵ een ander sleutelbegrip uit deze studie. Dat is het bedrag dat iemand, over zijn hele leven gerekend, netto bijdraagt aan de schatkist of ontvangt uit de schatkist.



Figuur 2.2 Nettobijdrage naar leeftijd voor een gemiddelde Nederlander, geboren in 2016, gecorrigeerd voor sterftekansen en uitgedrukt in euro's van 2016, Bron: eigen berekening op basis van CBS-microdata.

Bij de berekening van de nettobijdrage over de levensloop wordt rekening gehouden met sterftekansen; Nederlanders die heel oud zijn kosten de schatkist gemiddeld veel geld, maar niet iedereen bereikt een hoge leeftijd. Door de sterftekansen mee te wegen ontstaat een realistisch beeld. Verder worden de bedragen uitgedrukt in euro's van 2016 (contant gemaakt).⁶ Door beide aanpassingen tellen

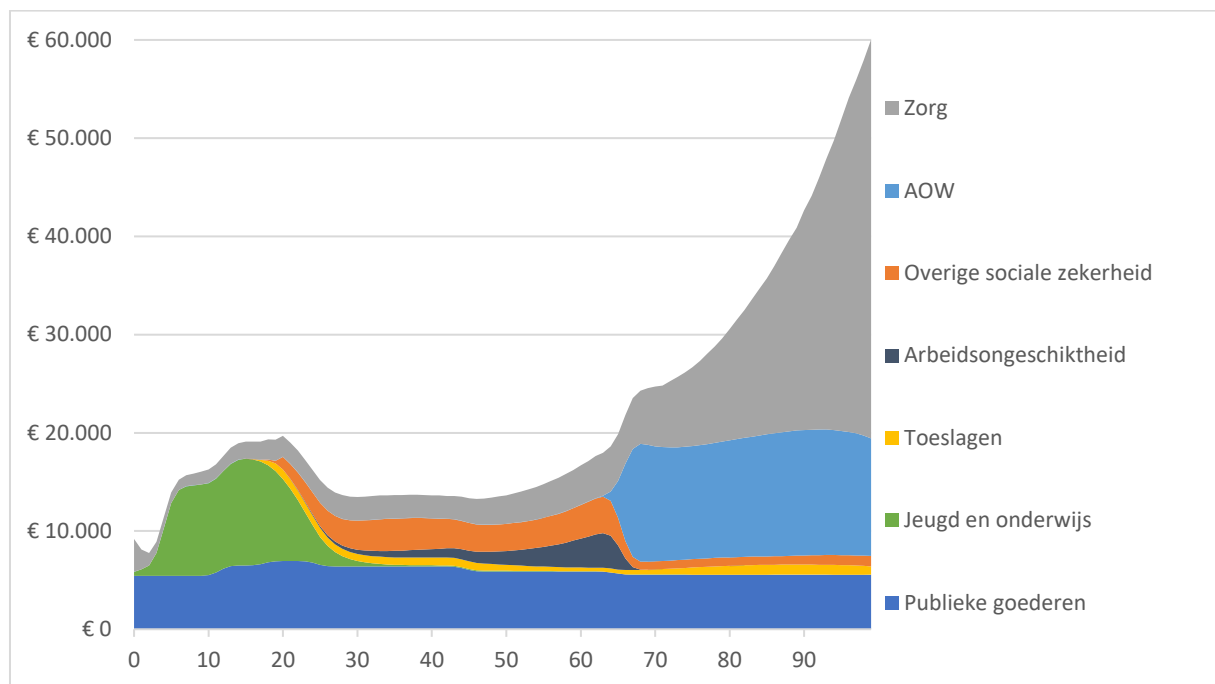
⁵ Voluit 'totale nettobijdrage over de levensloop' maar om schrijf-technische redenen wordt in het rapport kortweg gesproken van nettobijdrage.

⁶ De bedragen voor inkomsten en uitgaven van de overheid worden contant gemaakt. Er wordt rekening gehouden met zogenaamde tijdsvoordeel; een euro waarover men nu kan beschikken heeft een hogere waarde dan een euro waarover men in de toekomst kan beschikken. Toekomstige bedragen worden daarom contant

bedragen op hoge leeftijd en bedragen in de verre toekomst veel minder zwaar mee. In Figuur 2.2 dit is geïllustreerd door de nettobijdragen uit Figuur 2.1 te corrigeren voor sterfte en contant te maken.⁷ In Figuur 2.1 bedragen de kosten voor negentigers € 30.000 tot €50.000. In Figuur 2.2 bedragen de kosten voor negentigers minder dan € 5.000, door het contant maken en het feit dat lang niet iedereen zo oud wordt.

2.2 Inkomsten en uitgaven van de overheid

Om inzicht te krijgen hoe een nettobijdrageprofiel zoals in Figuur 2.2 tot stand komt, is het nodig om naar de afzonderlijke ingrediënten te kijken. Dat wordt gedaan door op hoofdlijnen te kijken naar de inkomsten en uitgaven die de overheid maakt voor Nederlandse ingezetenen.



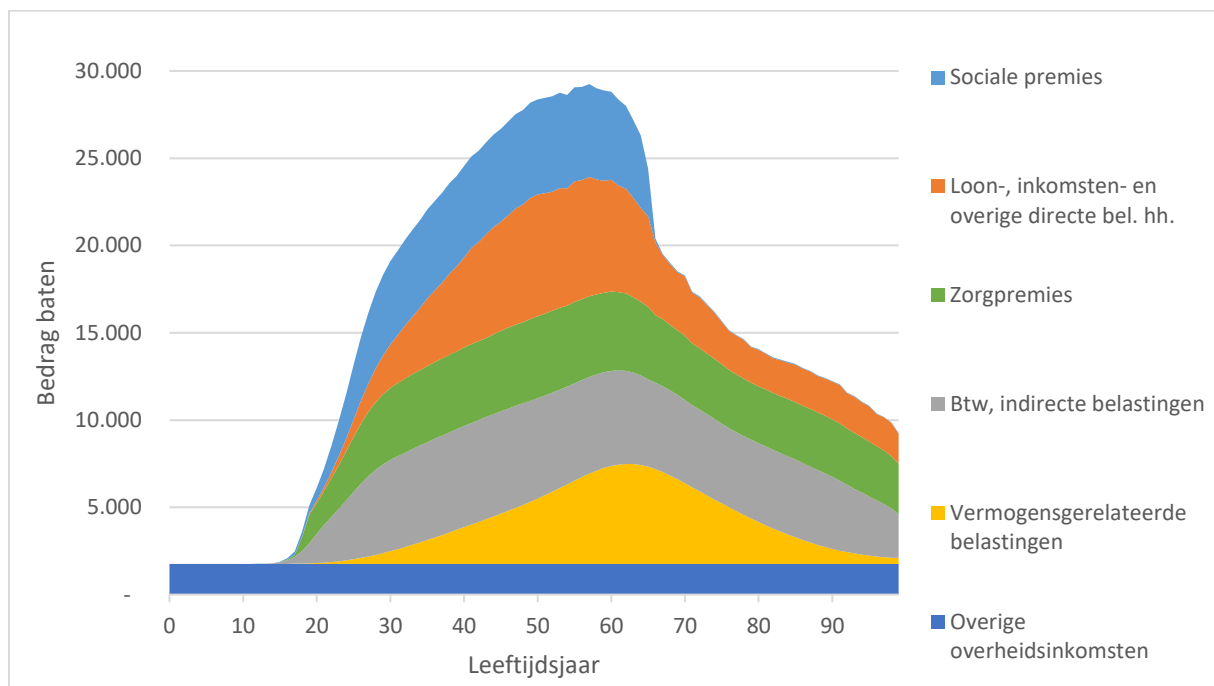
Figuur 2.3 Fiscale kosten naar leeftijd (bron: eigen berekening o.b.v. CBS-microdata).

Wat betreft de uitgavenkant geeft Figuur 2.3 inzicht in de belangrijkste posten. Van een aantal posten is aangenomen dat ze voor alle leeftijden gelijk zijn. Veel van deze posten zijn samengebracht onder de noemer publieke goederen. Hieronder vallen openbaar bestuur, defensie, investeringen in gebouwen en infrastructuur en overdrachten naar het buitenland (ontwikkelingshulp en dergelijke). De impliciete aanname is dat alle leeftijden in gelijke mate profiteren van – en bijdragen aan – deze kosten. Daarnaast zijn de kosten voor veiligheidszorg onder publieke goederen geschaard. Deze kosten zijn wel leeftijdsafhankelijk en hebben betrekking op de opsporing en eventuele vervolging en bestraffing van verdachten van misdrijven.

gemaakt, d.w.z. uitgedrukt in euro's van 2016 door ze tegen een bepaalde rekenrente af te waarderen. Zie verder de begrippenlijst. De gebruikte rekenrente is 1%, dit in navolging van de door het CPB gehanteerde rekenrente ten tijde van het schrijven van het huidige rapport.

⁷ Hierbij is uitgegaan van CBS-statline data over de verwachte sterftekansen van een in 2016 geboren persoon zoals gebruikt in de CBS-prognose 2017-2060, tabel *Prognose periode-levensverwachting; geslacht en leeftijd, 2017-2060*, opgehaald 13-9-2020 van: <https://opendata.cbs.nl/statline/portal.html?la=nl&catalog=CBS&tableId=83795NED&theme=91>. Deze data zijn beschikbaar tot 2060. Voor de jaren daarna zijn de sterftekansen van 2060 gebruikt.

Voor het overige is de levensloop zoals eerder gezegd in drie delen te verdelen. Voor de jonge jaren – en dan met name de periode tussen 4 en 20 jaar – is onderwijs de grote post. Voor het middelste deel van het leven – ongeveer tussen 20 en 65 jaar – is sociale zekerheid een grote post, waarbij opvalt dat arbeidsongeschiktheid sterk leeftijdsafhankelijk is. De post toeslagen (huursubsidie, kindgebonden budget en kind- en zorgtoeslag) is relatief gering in omvang, maar loopt wel over een groot deel van het leven. Vanaf de pensioenleeftijd zijn de AOW en de gezondheidszorg grote posten, waarbij ook weer opvalt dat gezondheidszorg sterk leeftijdsafhankelijk is. Hierbij moet wel bedacht worden dat sterfte er voor zorgt dat de torenhoge gezondheidszorgkosten op zeer hoge leeftijd in een gemiddeld leven slechts te dele meetellen. Een deel⁸ van dat effect is te zien door vergelijking van Figuur 2.1 met Figuur 2.2.



Figuur 2.4 Fiscale baten naar leeftijd. Bron: eigen berekening op basis van CBS-microdata. De overige overheidsinkomsten zijn gebaseerd op CPB-gegevens.

Aan de inkomstenkant is een vergelijkbare indeling van de levensloop in drie delen te maken (zie Figuur 2.4). De inkomsten onder de noemer overige overheidsinkomsten (netto grondverkoop, een deel van de niet-belastingmiddelen) zijn in navolging van het CPB aan alle inwoners in gelijke mate toegerekend, dus ook aan kinderen. Tot een jaar of 20 is dit zelfs vrijwel de enige bijdrage aan de schatkist. Tussen ruwweg 20 en 65 jaar zijn de afdrachten aan belastingen en premies relatief hoog met een piek rond 55 jaar. Voor jonge twintigers is er elk leeftijdsjaar een vrij sterke toename van de afdrachten, rond 65 jaar juist een vrij sterke daling vanwege de AOW-leeftijd.

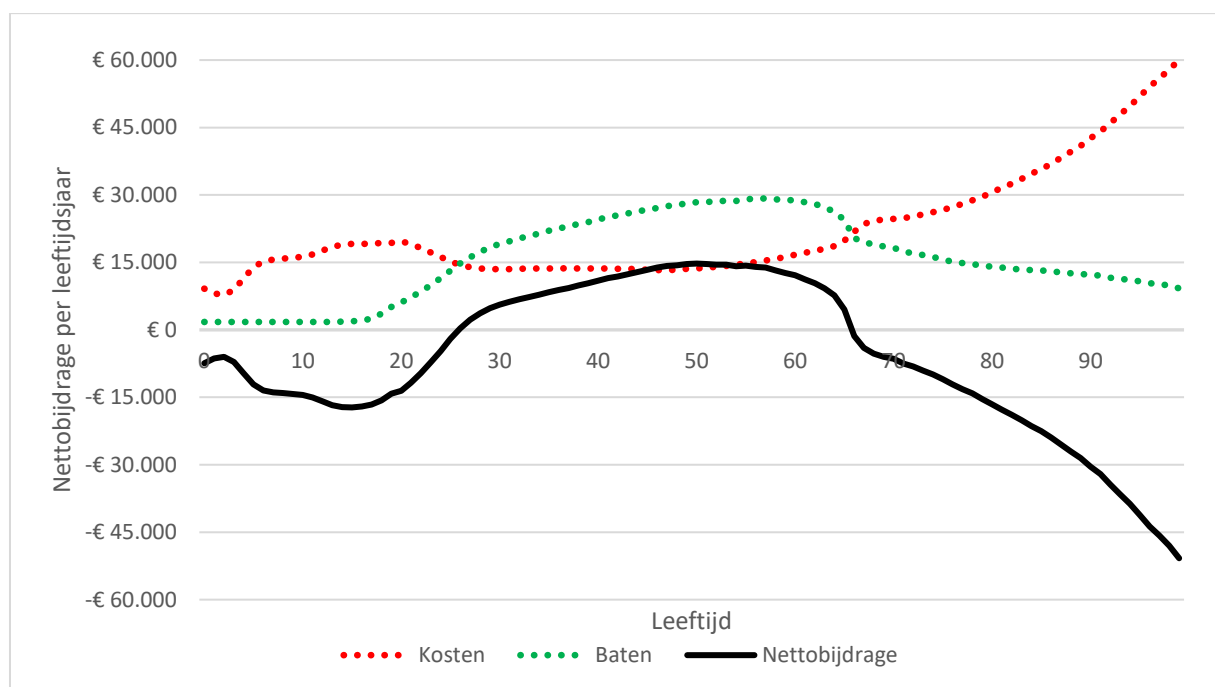
De afdrachten zijn verder duidelijk leeftijdsafhankelijk, wat vooral komt doordat loonbelasting, inkomstenbelasting en sociale premies met name tijdens het werkzame deel van het leven worden afgedragen. Vanaf de pensioenleeftijd zijn de belastingafdrachten en vooral de premieafdrachten beduidend minder. De indirecte belastingen (denk aan de BTW) en overige belastingen en niet-belastingmiddelen zijn daarentegen vanaf ongeveer 20 jaar een zich veel gelijkmatiger ontwikkelende inkomstenbron

⁸ Het andere deel van het verschil wordt veroorzaakt door contant maken.

voor de overheid. In Figuur 2.4 zijn ook de zorgpremies opgenomen, zowel de door werkgevers en werknemers betaalde premies als de verplichte verzekeringspremies voor de zorgverzekering. Tot slot zijn er vennootschapsbelasting, indirecte belastingen via bedrijven en erfbelasting welke zijn geschaard onder de verzamelterm vermogen-gerelateerde belastingen. Deze belastingen zijn gerelateerd het directe en indirecte bezit van bedrijven en hangen o.a. samen met inkomens uit BV's, huizenbezit en vooral de opbouw en afbouw van pensioenvermogens.

Door overheidsinkomsten en -uitgaven van elkaar af te trekken ontstaat het leeftijdsprofiel voor de nettobijdrage Figuur 2.1. Dit is geïllustreerd in Figuur 2.5. Het totaal van de kosten uit Figuur 2.3 is in Figuur 2.5 weergegeven met een rode stippellijn en het totaal van de baten uit Figuur 2.4 met een groene stippellijn. Door de kosten per levensjaar van de baten af te trekken ontstaat het leeftijdsprofiel voor de nettobijdrage. Dit is de zwarte lijn in Figuur 2.5 die correspondeert met Figuur 2.1.

Door de combinatie van kosten en baten wordt duidelijk hoe de driedeling jeugd, werkzaam leven en pensioen tot stand komt door. Tijdens de jeugd heeft de overheid forse kosten, maar nagenoeg geen inkomsten. Tijdens de werkzame periode heeft de overheid grote inkomsten uit belastingen en premies en relatief lage kosten. Tot slot zijn er tijdens het pensioen nog steeds inkomsten voor de overheid, maar worden deze in toenemende mate overtroffen door de kosten voor AOW en zorg.



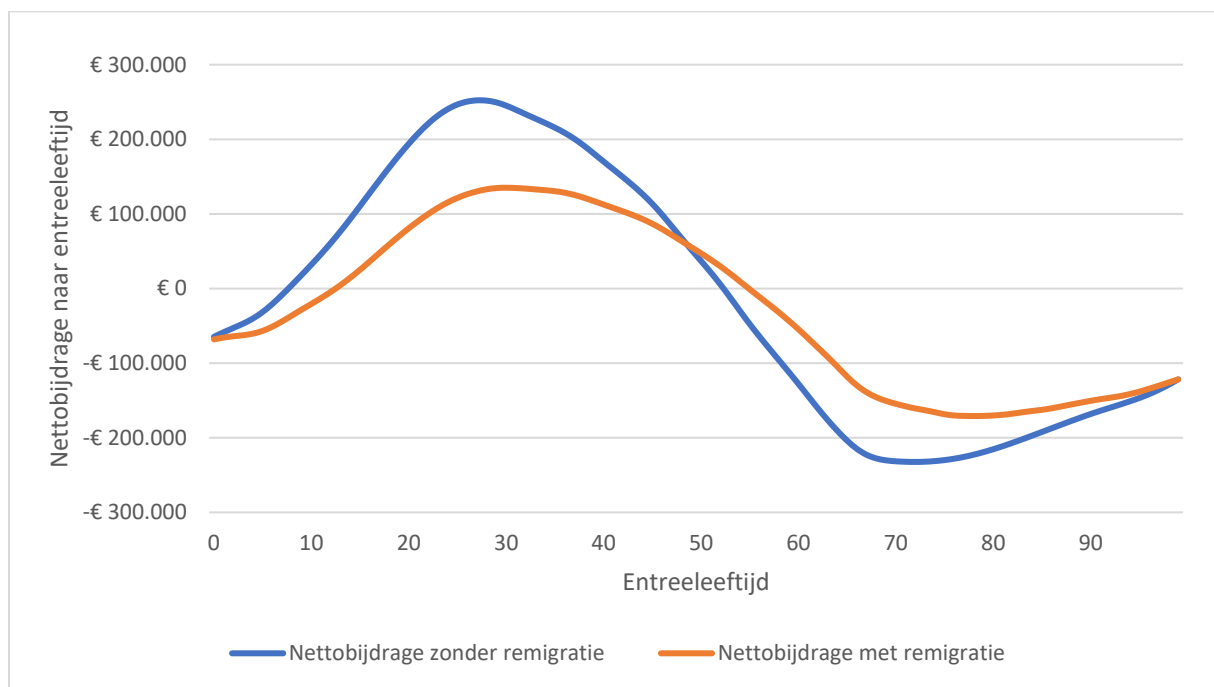
Figuur 2.5 Fiscale kosten en baten en de fiscale nettobijdrage naar leeftijd. Bron: eigen berekening op basis van CBS-microdata.

2.3 Generatierekening voor de eerste generatie

Een generatierekening kan ook voor immigrantengroepen worden gemaakt. In dat geval zijn er wel enkele aanpassingen nodig. De belangrijkste aanpassing is dat in beginsel alleen de kosten en baten geteld moeten worden voor de jaren dat een immigrant in Nederland verblijft. De periode vóór het moment van immigratie telt niet mee. Als de immigrant weer remigreert stoppen de (meeste) kosten, al behouden remigranten in het concrete geval van Nederland soms wel de opgebouwde AOW-rechten of toegang tot Nederlandse gezondheidszorg. Daarnaast is het van belang hoe de leeftijdsverdeling is

van de betreffende immigrantengroep, want de hoogte van de nettobijdrage hangt mede af van de leeftijd op het moment van immigratie. Tot slot zijn er mogelijk ook aanvangskosten die optreden rond het moment van immigratie, zoals de kosten van het verstrekken van verblijfsvergunningen en de opvang van asielzoekers. Deze zaken worden in deze paragraaf op hoofdlijnen toegelicht. Voor de eenvoud en vergelijkbaarheid gaat deze paragraaf uit van de gegevens van dezelfde groep als de vorige paragraaf, namelijk personen met de karakteristieken van de gemiddelde Nederlander.

Zoals gezegd dient het optellen van alle kosten en baten alleen te geschieden voor de leeftijden vanaf het moment van immigratie. De *entreeleeftijd* – de leeftijd op het moment van immigratie – is dus erg bepalend voor de nettobijdrage over de (resterende) levensloop. De entreeleeftijd is dan ook een belangrijk kernbegrip in deze studie. Dat is in te zien door naar verschillende entreeleeftijden te kijken.



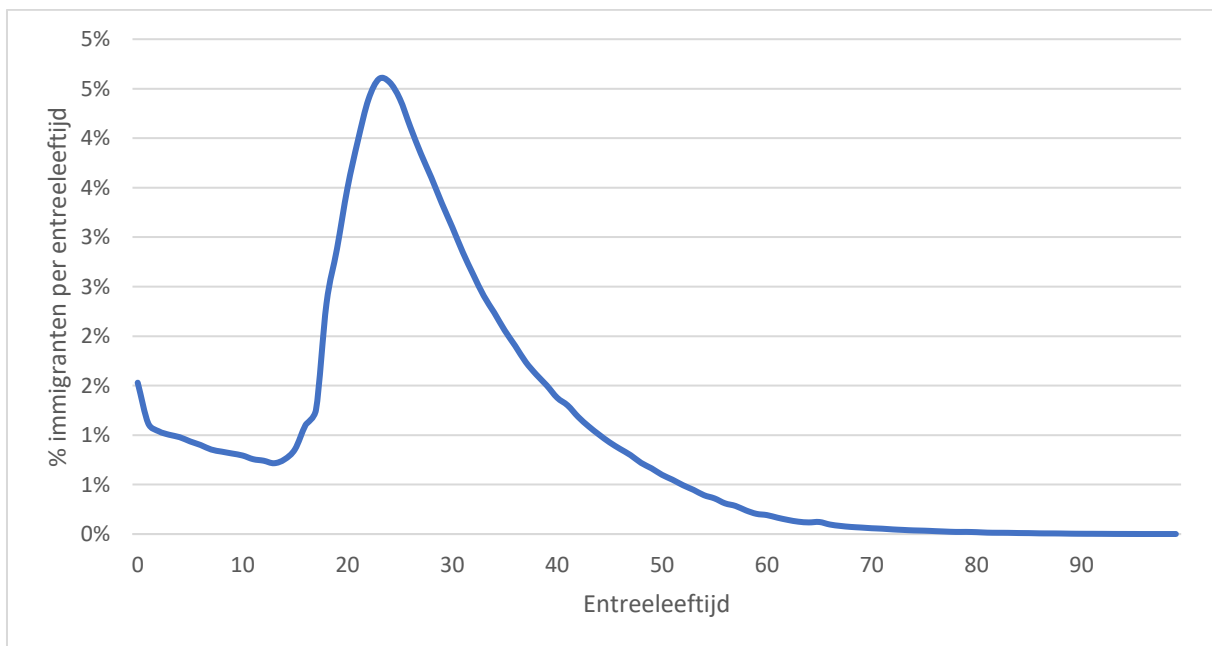
Figuur 2.6 Nettobijdrage over de levensloop naar entreeleeftijd, met en zonder remigratie. Bron: eigen berekening op basis van CBS-microdata.

Figuur 2.6 geeft weer hoe de nettobijdrage over de resterende levensloop – dus vanaf het moment van immigratie – afhangt van de entreeleeftijd, dat wil zeggen de leeftijd die de immigrant had op het moment van immigratie.⁹ Hierbij is uitgegaan van een immigrant met de karakteristieken van de gemiddelde Nederlander. De blauwe lijn geeft de nettobijdrage weer voor alle mogelijke entreeleeftijden tussen 0 en 99 jaar als er geen sprake is van remigratie. Uit Figuur 2.6 valt af te lezen dat de hoogste nettobijdrage wordt geleverd door immigranten met een entreeleeftijd tussen 25 en 30 jaar. Voor entreeleeftijden van 8 tot 53 jaar is er sprake van een positieve nettobijdrage. Immigranten die op zeer jonge leeftijd (tot 8 jaar) naar Nederland komen en immigranten die vanaf middelbare leeftijd (vanaf 53 jaar) komen, kosten volgens dit scenario de schatkist daarentegen gemiddeld genomen altijd geld, ook als ze elk leeftijdsjaar dezelfde nettobijdrage leveren als de gemiddelde Nederlander. Dit is te begrijpen aan de hand van twee concrete voorbeelden.

⁹ Bij het standaardscenario met aanpassing van de AOW leeftijd naar 69,5 jaar in 2060 en sterftেকansen constant vanaf 2060.

Als een immigrant met dezelfde nettobijdrage per jaar als de gemiddelde Nederlander bijvoorbeeld op zeventigjarige leeftijd binnenkomt, dan is de nettobijdrage over het resterende verblijf in Nederland gemiddeld genomen altijd negatief. Dit is op te maken uit Figuur 2.1 want de nettobijdrage per leeftijdjaar is voor alle leeftijden vanaf 70 jaar negatief. Omdat contant maken en verdiscontering van de sterftেকansen geen invloed hebben op het teken (positief of negatief) van de bedragen is de nettobijdrage (over de resterende levensloop) voor een immigrant die op zijn zeventigste immigrereert negatief, zoals in Figuur 2.6 is af te lezen aan de blauwe lijn die voor iemand van 70 jaar een negatieve waarde heeft van ruim twee ton (–€ 230.000).

Komt een immigrant daarentegen bijvoorbeeld op vijfentwintigjarige leeftijd naar Nederland, dan hoeven alle kosten en baten vóór het vijfentwintigste levensjaar niet meegeteld te worden. Nederland heeft voor die immigrant immers geen kosten gehad voor onderwijs, gezondheidszorg en dergelijke. Voor een vijfentwintigjarige immigrant met hetzelfde nettobijdrageprofiel als de gemiddelde Nederlander is uit Figuur 2.1 op te maken dat hij of zij direct een positieve bijdrage gaat leveren. Uiteindelijk komt ook voor deze immigrant de oude dag en de daaraan verbonden nettokosten. Echter, over het hele resterende leven gerekend – dus contant gemaakte bedragen gesommeerd¹⁰ van 25 jaar tot het moment van overlijden en met verdiscontering van de sterftেকansen – valt de nettobijdrage voor een immigrant met entreeleeftijd 25 positief uit, zoals in Figuur 2.6 is te zien (de blauwe lijn heeft voor een vijfentwintigjarige een positieve waarde van bijna tweeëneenhalf ton).



Figuur 2.7 Percentage immigranten per entreeleeftijd, eerste generatie immigranten, 2011-2017. Bron: eigen berekening op basis van CBS-microdata.

Omdat de entreeleeftijd zo bepalend is voor de nettobijdrage, is het nodig om nauwkeurig te weten hoe immigranten procentueel verdeeld zijn over de verschillende entreeleeftijden. In Figuur 2.7 is het *immigratieprofiel* – de procentuele verdeling over entreeleeftijden – gegeven voor alle eerste generatie immigranten uit de periode 2011-2017. Het immigratieprofiel is een belangrijk kernconcept in deze

¹⁰ Feitelijk wordt gesommeerd voor een figuur gelijk aan Figuur 2.2, maar dan beginnende bij 25 jaar en berekend voor de sterftেকansen die een 25-jarige heeft vanaf 2016.

studie. Het maakt namelijk veel uit of het merendeel van de immigranten bij gunstige of ongunstige entreeleeftijden binnenkomt. Te zien is dat de piek in Figuur 2.7 rond de 25 jaar ligt en dat is ook ongeveer de leeftijd waarbij de nettobijdrage in Figuur 2.6 het hoogst is. Voor het hier gehanteerde scenario vallen beide pieken dus gunstig samen. Door de nettobijdrage naar entreeleeftijd in Figuur 2.6 (de blauwe lijn, zonder remigratie) te wegen met het immigratieprofiel in Figuur 2.7 kan men de nettobijdrage over de levensloop bepalen, in dit geval voor immigranten met de karakteristieken van de gemiddelde Nederlander en zonder remigratie.

Uiteraard is er bij immigratie ook altijd sprake van forse remigratie, zeker van de eerste generatie.¹¹ Van westerse arbeidsmigranten is bijvoorbeeld bij veel groepen het merendeel binnen 10 jaar weer geëmigreerd. Ook hiermee moet rekening worden gehouden voor een realistische berekening. De oranje lijn in Figuur 2.6 geeft de nettobijdrage voor alle entreeleeftijden in geval er sprake is van remigratie. Hierbij is weer uitgegaan van immigranten met de karakteristieken van de gemiddelde Nederlander, maar nu aanvullend met de veronderstelling dat ze hetzelfde remigratiegedrag vertonen als de gemiddelde eerste generatie immigrant.¹² Deze hypothetische immigranten verdienen dus evenveel als de gemiddelde Nederlander, ontvangen even vaak een uitkering, consumeren evenveel onderwijs en zorg, kortom, ze hebben een nettobijdrage per jaar die (*grosso modo*¹³) gelijk is aan die van de gemiddelde Nederlander (zie Figuur 2.1). Verder hebben ze voor elke entreeleeftijd en verblijfsduur dezelfde remigratiekans als de gemiddelde eerste generatie immigrant. Te zien is dat remigratie in dit scenario zowel de positieve als de negatieve waarden voor de nettobijdrage minder extreem maakt.¹⁴ Daarnaast verschuift het leeftijdsinterval waarin immigranten een positieve nettobijdrage hebben enkele jaren in de richting van de hogere leeftijden.

In deze hele paragraaf is gebruik gemaakt van de gegevens van de gemiddelde Nederlander. Eerste generatie immigranten wijken soms sterk in positieve of negatieve zin af van de gemiddelde Nederlander. Dat betekent dat hun nettobijdrageprofiel over het geheel genomen boven respectievelijk onder de lijn in Figuur 2.1 zal liggen, met navenant hogere of lagere nettobijdragen.

Tot slot zijn er nog enkele andere kosten- en batenposten die niet of minder spelen bij ingezetenen die geen (eerste generatie) immigrant zijn. Allereerst zijn er de aanloopkosten van immigratie. De opvang van asielzoekers, het verstrekken van verblijfsvergunningen en de inburgering van immigranten brengen overheidskosten met zich mee en die moeten idealiter voor de desbetreffende groepen worden meegenomen in de berekening. Verder lopen sommige kosten zoals bijvoorbeeld zorgkosten en AOW na remigratie mogelijk door en ook die moeten idealiter meegewogen worden voor een totaalberekening van de totale nettobijdrage over de levensloop. De empirische invulling hiervan wordt gegeven in §6.1 t/m §6.3

¹¹ Zie voor het verloop van de remigratie voor verschillende groepen §2.2 van het rapport.

¹² Zoals waargenomen op basis van de data over 1995-2017.

¹³ Niet volstrekt gelijk, want het gaat om personen met de nettobijdragekarakteristieken van de gemiddelde Nederlander en de migratiekarakteristieken van de gemiddelde immigrant en dan zijn de nettobijdragen per jaar gelijk aan die van de gemiddelde Nederlander, *behoudens pensioenopbouw en aanspraken op de AOW en het gebruik van bijstand vanaf 65 jaar*, omdat deze afhangen van de verblijfsduur en de opbouw van AOW-rechten. Op deze wijze is bijvoorbeeld de in het rapport figurerende referentie-autochtoon (een 'immigrant met de karakteristieken van de gemiddelde autochtoon') berekend.

¹⁴ Althans, in dit voorbeeld, want het precieze effect van remigratie op de nettobijdrage naar entreeleeftijd hangt o.a. mede af van het profiel voor de nettobijdrage naar leeftijdsjaar.

2.4 Generatierekening voor de tweede generatie

De fiscale impact van immigratie blijft niet beperkt tot de eerste generatie. Als de immigrant zich bijvoorbeeld blijvend vestigt en kinderen krijgt, zullen die kinderen over hun levensloop ook een nettobijdrage leveren aan de schatkist die positief of negatief uit kan pakken. Ook kan er een groot verschil bestaan tussen de eerste en tweede generatie wat betreft opleidingsniveau en arbeidsmarktprestaties. Hierdoor is het bijvoorbeeld mogelijk dat de eerste generatie een negatieve nettobijdrage levert en de tweede generatie een positieve of andersom. Daarom is het nodig tenminste ook de tweede generatie in de berekeningen mee te nemen.

In de CPB-studie uit 2003 werd aangenomen dat de tweede generatie niet-westerse immigranten qua arbeidsmarktkenmerken halverwege de niet-westerse eerste generatie en de gemiddelde Nederlander zouden zitten. In deze studie is dat idee verder uitgewerkt, door deze aanname van '50% integratie' te vervangen door een schatting van de mate van integratie op basis van CBS-microdata. Hierbij is aangenomen dat het nettobijdrageprofiel voor de tweede generatie in beginsel altijd een lineaire combinatie is van het referentieprofiel voor autochtonen (volledige integratie, ofwel 100% integratie) en het referentieprofiel voor de niet-westerse eerste generatie (geen integratie, ofwel 0% integratie). Dit is gedaan door gebruik te maken van de beschikbare data voor de nettobijdrage van de tweede generatie.

Er zijn echter bij het berekenen van de nettobijdrage van de tweede generatie enkele beperkingen. De eerste daarvan is gebrek aan data. Voor de meeste Europese herkomstregio's en -landen en voor landen als de VS, Indonesië en China is er voldoende data. Maar voor belangrijke herkomstlanden als Turkije en Marokko is het aantal tweede generatie personen voor leeftijden boven de 50 jaar zeer gering en zijn de data tussen 40 en 50 jaar beperkt. Voor sommige landen en regio's zijn er nog minder data.

Met name niet-westerse immigratie is namelijk een recent verschijnsel. Tot de Tweede Wereldoorlog was er weinig immigratie en kwamen de meeste immigranten uit omliggende landen en de koloniën. Daarna nam de immigratie toe. Vanaf eind jaren veertig was er eerst een tamelijk omvangrijke immigratie van Indische Nederlanders, die overigens door het CBS als westers gecategoriseerd worden. Vervolgens kreeg vanaf eind jaren zestig de gastarbeid substantiële omvang. Halverwege de jaren zeventig – rond de periode van de Surinaamse onafhankelijkheid – werd de migratie uit Suriname omvangrijk. Asielmigratie werd pas halverwege de jaren tachtig van betekenis, met 1985 als eerste jaar met meer dan 5.000 asielzoekers. Kortom, voor veel groepen is de tweede generatie nog jong.

Een tweede uitdaging bij een betrouwbare bepaling van de nettobijdrage van de tweede generatie zijn zogenaamde cohorteffecten.¹⁵ Onder cohort wordt hier verstaan een groep mensen die allemaal in dezelfde periode zijn geïmmigreerd. Immigranten van verschillende cohorten hebben voor sommige herkomstgroepen duidelijk andere karakteristieken waardoor de nettobijdrage van de huidige tweede generatie niet goed te bepalen is op basis van de data voor de hogere leeftijden. Immigratie uit overzeese rijkdelen als Suriname verschoof in Nederland bijvoorbeeld van beperkte elitemigratie naar meer massale migratie van midden- en laaggeschoolden rond de periode van de Surinaamse

¹⁵ Cohorteffecten en gebrek aan data hangen vaak samen, omdat het vaak gaat om de overgang van beperkte elitemigratie (weinig data) naar grootschaligere (veel data) en minder selectieve immigratie, bijvoorbeeld door asiel, werving van gastarbeiders, of koloniale migratie en de vervolgens daaruit voortvloeiende gezinsmigratie.

onafhankelijkheid.¹⁶ Deze overgang komt tot uitdrukking in verschillen in nettobijdrage tussen cohorten Surinaamse immigranten: tweede generatie immigranten van 42 jaar en ouder (in 2016) hebben een relatief hoge nettobijdrage ten opzichte van latere cohorten (die in 2016 dus jonger waren dan 42 jaar).

De cohorteffecten verschillen per categorie herkomstgroep en zijn bij de tweede generatie vanaf een bepaalde leeftijd vaak duidelijk zichtbaar. Voor Surinamers ligt het omslagpunt bij een leeftijd van rond 42 jaar en dat is ook precies de duur van 2016 teruggerekend naar de periode vóór de onafhankelijkheid in 1975. Voor Turken, Marokkanen en een aantal andere wervingslanden voor gastarbeiders ligt het omslagpunt bij een leeftijd van ongeveer 45 jaar. Voor veel qua migrantenaantallen belangrijke herkomstgebieden in Afrika en Azië is de toename van asielmigratie en de daaruit voortkomende gezinsmigratie vanaf midden jaren tachtig het moment waarop de migratie toenam. Bij deze groepen is te verwachten dat mogelijke cohorteffecten optreden bij leeftijden van 35 jaar of jonger.

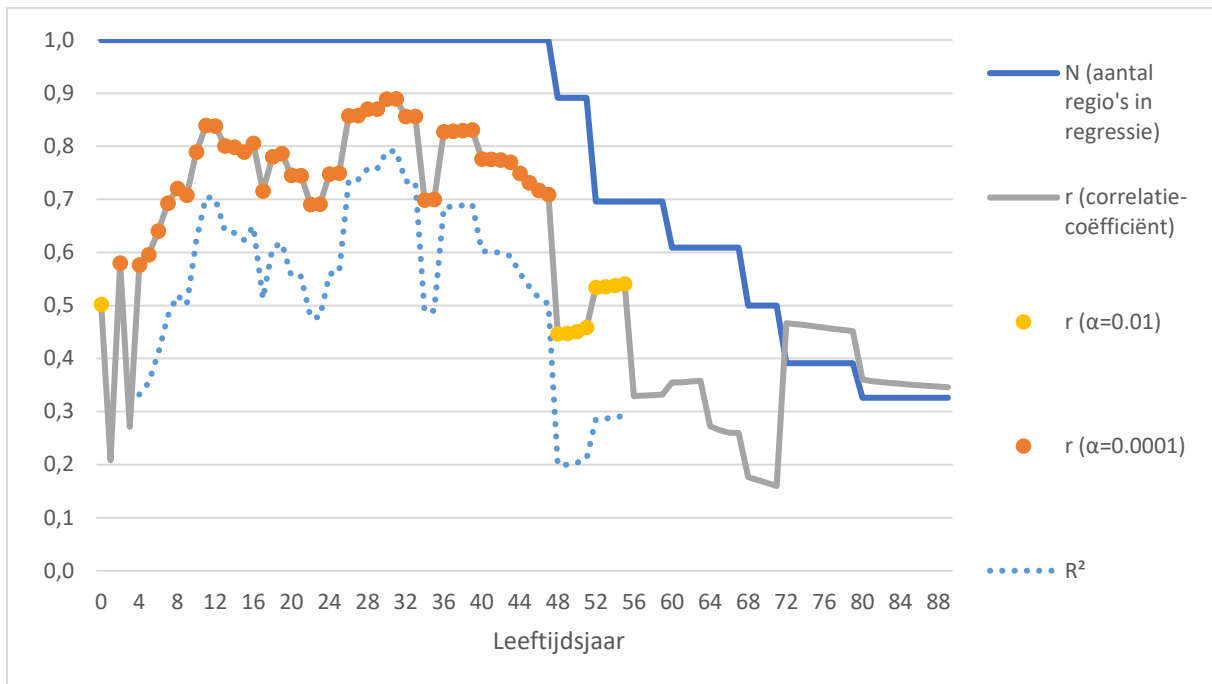
De doelstelling bij het schatten van de nettobijdrage van de tweede generatie is geweest om de beste schatting te geven voor in Nederland geboren kinderen van eerste generatie immigranten die tussen 1995 en 2040 zijn geboren of geboren zullen worden. Om dat doel te bereiken moeten alle cohorteffecten zoveel mogelijk worden uitgefilterd. *Het is dus uitdrukkelijk niet de bedoeling geweest om het nettobijdrageprofiel te vinden dat het best past bij alle beschikbare data.* Dat zou veel makkelijker zijn geweest, maar zou vrijwel zeker onzuiverheid (“bias”) geven (in relatie tot de in de eerste zin van deze alinea gegeven doelstelling). Bij de hier gepresenteerde methoden is er ongetwijfeld ook onzuiverheid, omdat de toekomst onbekend is. Onzuiverheid of vertekening is onvermijdelijk, maar de onzuiverheid die zou kunnen ontstaan door cohorteffecten e.d. in het verleden kan men proberen te vermijden en dat is bij de hierna beschreven methoden zoveel mogelijk gedaan.

Uit het voorgaande is gebleken dat er vaak onvoldoende data zijn om nettobijdrage profielen te maken voor het hele leven van tweede generatie immigranten. Uiteraard kan men data die er niet zijn, niet maken, maar men kan wel van omliggende gegevens gebruik maken om de leemten zo goed mogelijk in te vullen. In een poging om schattingsfouten te minimaliseren is de mate van integratie van de tweede generatie met vier verschillende methoden geschat.

Voor het uiteenzetten van de eerste twee methoden is het van belang te weten dat bij de tweede generatie de nettobijdrage sterk correleert met de gemiddelde citoscore van de herkomstgroep. Dit is geïllustreerd in Figuur 2.8, waar voor 41 herkomstregio's (de 42-deling minus Nederland¹⁷), voor elk leeftijdsjaar afzonderlijk de (Pearsons) correlatiecoëfficiënt is gegeven tussen enerzijds de gemiddelde citoscore per regio en anderzijds de gemiddelde nettobijdrage per regio.

¹⁶ “De tweede belangrijke immigratiebeweging in de jaren zeventig was die van de Rijksgenoten, en dan met name de Surinamers. Vanaf 1880 tot 1950 was er sprake van een beperkte migratie van Surinamers afkomstig uit de (ambtelijke) elite, vaak met het doel om na het afronden van een opleiding in Nederland weer terug te keren naar Suriname. Een deel van deze groep bleef in Nederland en kon zich door zijn goede opleiding redelijk makkelijk een positie in Nederland verwerven. Vanaf 1950 nam de immigratie van mensen uit de middenklasse toe, waaronder ook een beperkt aantal arbeidsimmigranten. Surinamers hadden op basis van het Koninkrijksstatuut (1954) vrije toegang tot Nederland. In de jaren zestig nam het migratiesaldo geleidelijk toe van 400 in 1960 tot 4400 in 1969. Vanaf 1970 breidde de immigratie zich verder uit naar laaggeschoolde Surinamers met een beperkte oriëntatie op Nederland en nam ook het aantal immigranten toe”, Van de Beek, J. H. (2010)

¹⁷ Hier en in de rest van deze paragraaf is er voor herkomst Indonesië, Turkije, Marokko, Suriname, Aruba en de (voormalige) Nederlandse Antillen verder uitgesplitst naar het aantal in het buitenland geboren ouders.



Figuur 2.8 Pearsons correlatiecoëfficiënten (en de kwadraten R^2 daarvan) voor de samenhang tussen de gemiddelde citoscore (2006-2018) en de nettobijdrage per levensjaar voor tweede generatie immigranten voor de 42-deling in herkomstregio's (exclusief Nederland), waarbij voor herkomst Indonesië, Turkije, Marokko, Suriname, Aruba en de (voormalige) Nederlandse Antillen is uitgesplitst naar het aantal in het buitenland geboren ouders. Significante correlaties op significantieniveau 0.01 en 0.0001 zijn gemarkeerd.

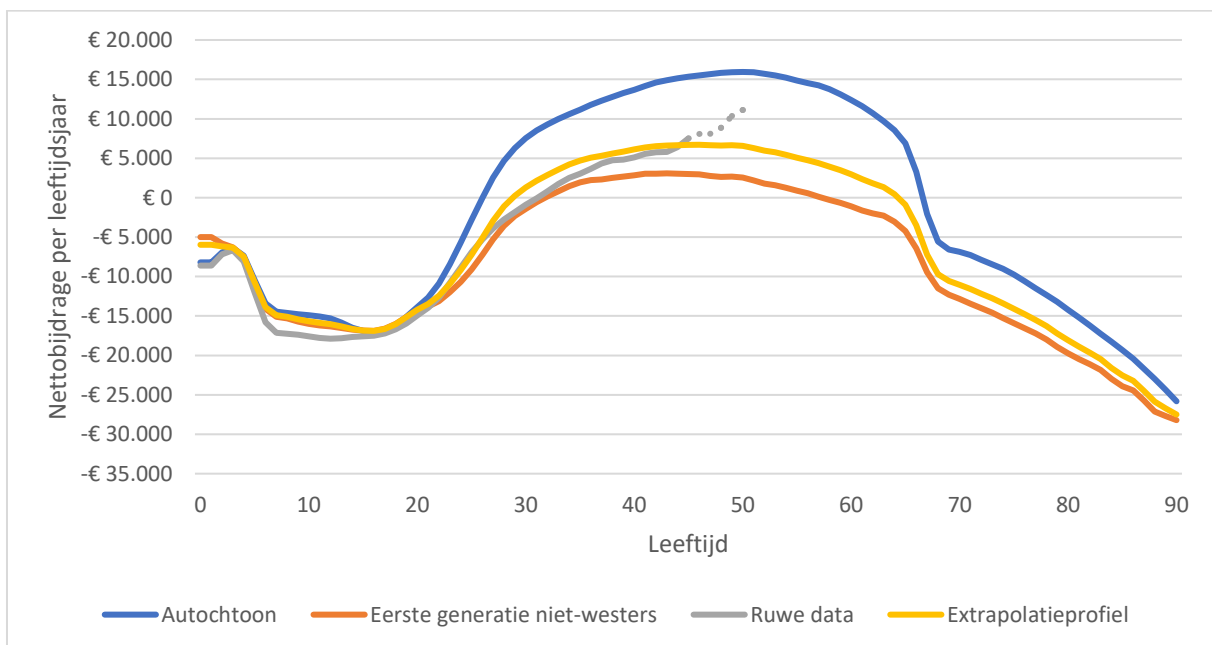
Het patroon in Figuur 2.8 is als volgt tentatief te verklaren.¹⁸ Vanaf de basisschoolleeftijd neemt de correlatie toe omdat onderwijs en zorg dan de grote uitgavenposten zijn en beiden sterk gerelateerd zijn aan citoscores. Dat blijft zo tot 16 jaar. Van 17 tot 26 jaar zijn er waarschijnlijk verschillende tegenstelde krachten werkzaam, onder andere doordat een deel van de jongeren met een citoscore rond het gemiddelde na het einde van de leerplicht gaat werken (al dan niet via een leerwerktraject) en begint met belasting betalen en mensen met een hogere citoscore gemiddeld genomen vaker langer studeren en later gaan werken. Dat is een mogelijke verklaring waarom de samenhang tussen cito en nettobijdrage in deze levensfase lager is, vooral van 20 tot 26 jaar. Van 26 tot 48 jaar beweegt de correlatiecoëfficiënt tussen 0,7 en 0,9 en is er sterke samenhang tussen de gemiddelde citoscores en de jaarlijkse nettobijdrage van de tweede generatie. Voor deze leeftijden is een groot deel van de mensen klaar met de opleiding en bepaalt het opleidingsniveau sterk het inkomen en daarmee de nettobijdrage. Voor leeftijden van 48 tot 56 jaar gaan de databeperkingen en cohorteffecten (de citoscores zijn afgenomen in de periode 2006-2018) een grote rol spelen en is de samenhang zwakker en grilliger. Vanaf 56 jaar is er voor de 42-deling in herkomstregio's (exclusief Nederland) meestal¹⁹ geen sprake meer van een significante samenhang.

De sterke samenhang tussen cito en nettobijdrage wordt vervolgens gebruikt als hulpmiddel bij het bepalen van de nettobijdrage van de tweede generatie. De hiervoor gebruikte methode is gebaseerd

¹⁸ Het geblokte profiel van de 'kromme' komt doordat de databeperkingen ook noodzakelijk maken om zeker bij kleinere groepen de wat hogere leeftijden per leeftijdsgroep te samplen.

¹⁹ Uitzonderingen: 0 jaar, 2 jaar en (berekend over een kleine groep landen) 72-79 jaar, de laatste met p-waarden net onder 5%.

op extrapolatie²⁰ van de waarnemingen die er tussen 4 en maximaal 55 jaar zijn, naar de rest van het leven. Daarbij zijn (geautomatiseerd, met een *brute force* methode) steeds 300 profielen onderzocht: 100 profielen tussen de referentieprofielen van autochtonen en niet-westerse eerste generatie, 100 profielen boven het autochtone profiel en 100 profielen onder het profiel van de niet-westerse eerste generatie. Elk profiel is zoals eerder gezegd een lineaire combinatie van het referentieprofiel voor autochtonen en het referentieprofiel voor de niet-westerse eerste generatie. Elk profiel is gepast (gefit) op de waarnemingen voor de nettobijdrage van de betreffende tweede generatiegroep voor de leeftijden tussen 4 en 55 jaar. Bij het maken van de profielen is steeds dezelfde stapgrootte aangehouden: 1 procent van het verschil tussen het profiel voor autochtoon en het profiel voor niet-westers eerste generatie.



Figuur 2.9 Schematische weergave van een mogelijk extrapolatieprofiel voor de data voor het nettobijdrageprofiel (ruwe data, zonder remigratie, contant maken e.d.) voor personen met een tweede generatie migratieachtergrond. Gestippelde lijn betekent dat er weinig data zijn.

Het getal behorend bij het best passende profiel kan geïnterpreteerd worden als de mate van integratie. Een waarde van 100 staat hier voor perfecte integratie en zou geïnterpreteerd kunnen worden als '100% geïntegreerd', in die (beperkte) zin dat de nettobijdrage gelijk staat aan die van de gemiddelde autochtoon. Een waarde groter dan 100 betekent een hogere nettobijdrage dan de gemiddelde autochtoon. Men is dan dus als het ware 'méér dan geïntegreerd', met andere woorden, presteert beter dan een autochtoon als het gaat om de nettobijdrage aan de schatkist over de levensloop. Een waarde kleiner dan 100 betekent een lagere nettobijdrage dan de gemiddelde autochtoon. Een waarde van 0 staat hier voor een nettobijdrage gelijk aan de gemiddelde eerste generatie niet-westerse immigrant en kan geïnterpreteerd worden als '0% geïntegreerd'. Een waarde van 40 zou men kunnen

²⁰ In wezen wordt ook van interpolatie gebruik gemaakt (er worden profielen gefit tussen twee referentieprofielen voor autochtonen en eerste generatie niet-westerse immigranten), maar omdat er ook steeds sprake is van extrapolatie (zowel naar de hogere leeftijden als ook naar profielen die bijvoorbeeld boven het referentieprofiel voor autochtonen liggen) is consequent voor de term extrapoleren gekozen.

interpreteren als ‘voor 40% geïntegreerd’, weer in de beperkte zin van de mate van nettobijdrage, want dit gaat niet over sociaal-culturele integratie en dergelijke.

Het bepalen van de mate van integratie is voor de eerste methode schematisch weergegeven in Figuur 2.9 voor personen met een tweede generatie Marokkaanse migratieachtergrond. De blauwe lijn geeft de nettobijdrage van autochtonen weer. De oranje lijn de nettobijdrage van alle eerste generatie niet-westerse migranten. De gele lijn is een mogelijke (lineaire) combinatie van autochtonen en eerste generatie niet-westerse immigranten. De grijze lijn geeft het patroon weer volgens de beschikbare data voor een groep personen met een tweede generatie migratieachtergrond. Het gestippelde deel is gebaseerd op een te laag aantal personen per leeftijd volgens een vooraf bepaalde samplenorm. Er zijn zes samplenormen gebruikt: minimaal 5, 10, 20, 50, 100 en 200 personen per leeftijd.²¹

Het extrapoleren bestaat er nu uit om voor elk van de zes samplenormen net zo lang te variëren met de gele lijn in Figuur 2.9 tot de best passende lijn (van de 300 mogelijke lijnen) gevonden is. Dit is gedaan op basis van een variant van de kleinste kwadratenmethode.²² De kwadratensom KS_L is daarbij herhaaldelijk bepaald door de kwadraten te sommeren over de leeftijden van 4 tot L jaar, voor alle leeftijden L van 25 tot 56 jaar. Voor elke run van 4 tot L jaar is onderzocht of KS_L onder een vooraf ingestelde bovengrens KS_{max} bleef. Indien dat niet het geval was, is de schatting niet meegenomen in de verdere analyse.²³ Er zijn 160 bovengrenzen gebruikt, van 400 tot 16.000, aanvankelijk met stapgrootte 400, waarna rond het optimum verder is gezocht met stapgrootte 100. Vervolgens is voor elke KS_{max} (met de eerder beschreven *brute force* aanpak) de best passende lijn berekend.²⁴

Op voorgaande wijze is voor een bepaalde KS_{max} voor elk van de zes samplenormen een schatting gemaakt. Van deze (maximaal²⁵) zes schattingen zijn vervolgens de hoogste en de laagste schatting verwijderd en is het gemiddelde genomen van de overige schattingen.²⁶ Dit is voor elke KS_{max} gedaan. Vervolgens is onderzocht bij welke van de KS_{max} de beste extrapolatie wordt gevonden. Het criterium daarbij was de sterkte van de correlatie met de gemiddelde citoscores voor de tweede generatie. Dit is gedaan voor de 42-deling minus Nederland, met de vijf grote herkomstlanden Indonesië, Turkije,

²¹ Bij het minimum van 5 wordt de facto tot 36 jaar alle beschikbare data meegenomen, want het minimale aantal waarnemingen per leeftijdsgroep (niet per leeftijd!) is 10, vanwege de CBS-eisen m.b.t. onthullingsrisico. Tot 20 jaar beslaan de leeftijdsgroepen één jaar en van 20 tot 36 jaar twee jaar en is het minimale aantal waarnemingen dus altijd vijf per leeftijd. Van 36 tot 56 jaar beslaan de leeftijdsgroepen vier jaar. In dat laatste geval zijn er minstens $10 / 2 = 2,5$ personen per levensjaar gesampled en zal een deel van de waarnemingen bij het de drempelwaarde van vijf waarnemingen per leeftijd dus in een aantal gevallen niet worden meegenomen.

²² Hierbij is uitgegaan van een gestandaardiseerde kleinste kwadratensom. Met gestandaardiseerde kleinste kwadratensom wordt bedoeld de kwadratensom gedeeld door het aantal waarnemingen en daarvan de wortel. De begrenzing op 55 jaar is geïnspireerd op de observatie dat vanaf 56 jaar de correlatie met citoscores sterk afneemt. Om geen optimale oplossing te missen is desondanks voor de zekerheid alles ook nog eens uitgerekend voor alle leeftijden tussen 26 en 96 jaar, maar beperking tot 56 jaar gaf in alle gevallen de beste resultaten.

²³ Om het algoritme niet vroegtijdig te laten termineren op kleine spikes zijn de leeftijdsprofielen voor de nettobijdrage en het aantal waarnemingen vooraf gesmoothed door tweemaal achter elkaar toepassen van de functie $(f(L - 1) + 2 \times f(L) + f(L + 1))/4$ voor leeftijden $0 < L < 99$ waarin $f(L)$ de te smoothen functie is.

²⁴ Voor zover deze voor de betreffende KS_{max} bestond, want als de kwadratensom van meet af aan te groot is termineert het algoritme direct.

²⁵ In de praktijk kunnen vaak niet alle zes de schattingen worden gemaakt wegens een gebrek aan data, zie ook vorige voetnoot.

²⁶ In een aantal gevallen waren er vier schattingen en in één geval (Japan) drie schattingen en is het gemiddelde genomen van die schattingen zonder de extremen te verwijderen.

Marokko, Suriname men de (voormalige) Antillen uitgesplitst naar de subgroepen met één dan wel twee in het buitenland geboren ouders. Deze was maximaal voor $KS_{max} = 3.000$, $r(44) = .91$, $p \ll .0001$.

De rationale achter methode 1 is dat mensen met een tweede generatie migratieachtergrond die vanaf 2016 worden geboren qua nettobijdrage het meest zullen lijken op jonge mensen met een tweede generatie migratieachtergrond zoals waargenomen in 2016. Dat is ook wat de – met de leeftijd afnemende – correlatie met de citoscore laat zien. Feitelijk probeert dit algoritme een steeds groter deel 4 tot L van het nettobijdrageprofiel te fitten, beginnend $L = 25$ jaar en eindigend bij $L = 55$ jaar. Als beste profiel wordt dan het profiel genomen met de hoogste L waarvoor geldt dat $KS_L < KS_{max}$ en KS_L minimaal voor de betreffende L . Het algoritme termineert als het ware als de nettobijdrage voor het laatst toegevoegde levensjaar de gestandaardiseerde kwadratensom te groot maakt (d.w.z. $\geq KS_{max}$).

De rationale achter de tweede methode is om de sterkst correlerende periode van het leven (zie Figuur 2.8) het zwaarst te wegen en daarnaast het aantal waarnemingen per leeftijd mee te wegen. Dit is gedaan door wederom met de kleinste kwadratensom te fitten, maar nu uitsluitend voor de hele periode van 4 tot 56 jaar en per leeftijd de kwadraten te wegen met de verklaarde variantie (R^2 , zie de blauwe lijn in Figuur 2.8) van de lineaire regressies van de nettobijdrage per levensjaar op de gemiddelde citoscore voor de betreffende tweede generatie herkomstgroep (wederom voor de 42-deling minus Nederland).

Daarnaast is ook gewogen met het gemiddeld aantal personen N per leeftijd.²⁷ Als $N \geq 100$ of groter was, is gewicht 1 gehanteerd en als N onder 100 lag is gewicht $N / 100$ gebruikt. Zo is gewaarborgd dat kleine aantallen een te grote invloed zouden hebben. Ook deze tweede methode zorgt voor een sterke correlatie met de gemiddelde citoscores van de tweede generatie groepen, $r(44) = .90$, $p \ll .0001$.²⁸

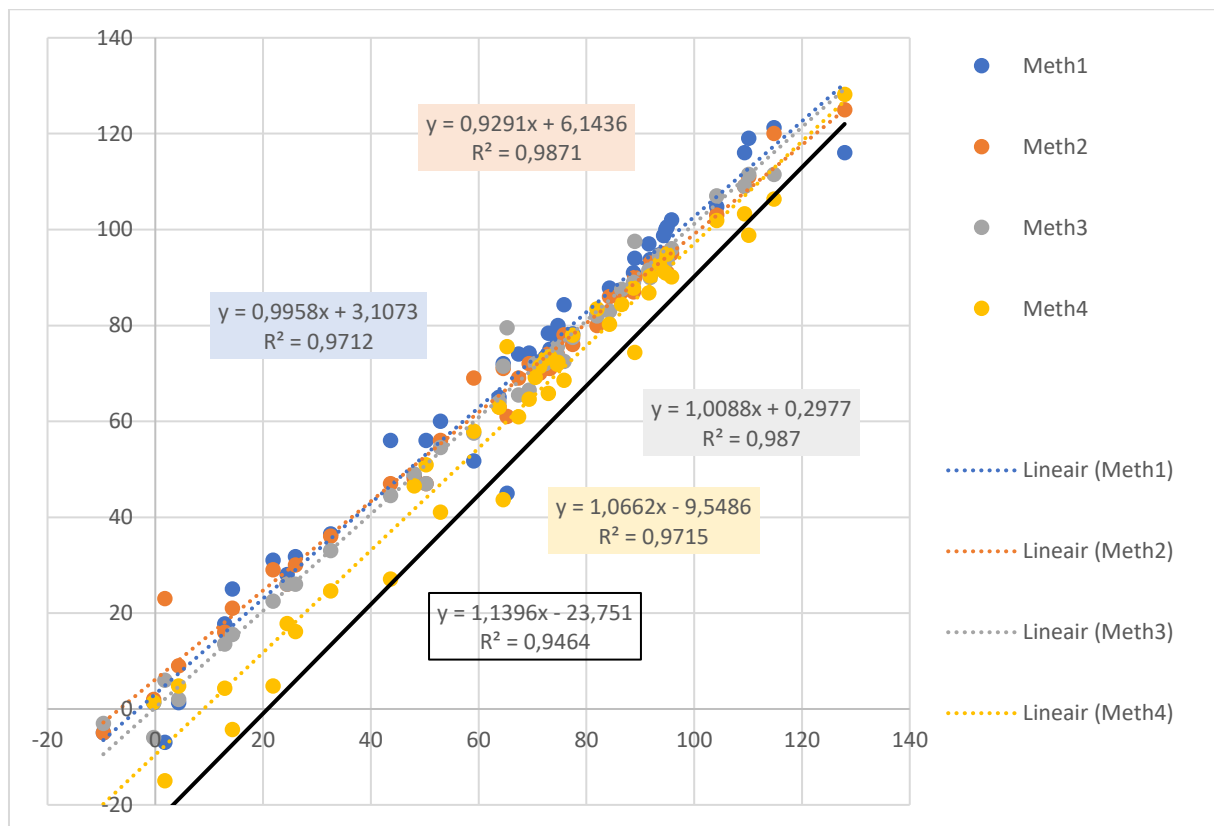
De derde methode is vergelijkbaar met de tweede, met als verschil dat R^2 niet is meegewogen in de kwadratensom. Voor de vergelijkbaarheid is nu gekozen voor de leeftijden tot 48 jaar, omdat van alle groepen tot tenminste 48 jaar data zijn. In deze methode is ook gewogen met het gemiddeld aantal personen N per leeftijd, zoals beschreven bij methode 2. Van deze methode zijn twee varianten uitgerekend en gemiddeld. De eerste variant voor leeftijden van 0 tot 48 jaar en de tweede variant voor leeftijden van 25 tot 48 jaar. De rationale hierachter is om gevoeligheid te verminderen voor verschillen in studiekosten (bij groepen met relatief veel niet-EER studenten) of relatief hoge leerlinggewichten. De periode 25 tot 48 jaar geeft een goed beeld van de carrièreopbouw zonder dergelijke verstoringen. Ook deze methode heeft een sterke correlatie met de gemiddelde citoscores van de tweede generatie groepen voor de 41-deling met uitsplitsing van de big five, $r(44) = .90$, $p \ll .0001$. Merk op dat bij deze methode anders dan bij de eerste twee methoden niet a priori gebruik is gemaakt van informatie over de correlatie met de citoscores, en dat de correlatie evengoed erg sterk is. Dat

²⁷ Vanwege de geringe hoeveelheid data is er voor hogere leeftijden niet per leeftijd gesampled, maar per leeftijdsgroep. Als leeftijd L behoort tot leeftijdsgroep G dan is de gestandaardiseerde samplegrootte N voor leeftijd L gelijk aan het aantal personen in leeftijdsgroep G gedeeld door het aantal jaren dat leeftijdsgroep G beslaat. Dus als er voor leeftijdsgroep 36 tot 40 jaar 40 waarnemingen zijn is de gestandaardiseerde samplegrootte $40 / 4 = 10$. Als minimum aantal waarnemingen per leeftijd is net als bij methode 1 vijf aan gehouden. Omdat voor leeftijden vanaf ongeveer 45 jaar vaak sprake is van cohorteffecten is de kwadratensom net als bij methode 1 gewogen met de sigmoïde functie $(1 + e^{L-50})^{-1}$ voor leeftijden L .

²⁸ Wederom voor de 42-deling minus Nederland, met de vijf grote herkomstlanden Indonesië, Turkije, Marokko, Suriname men de (voormalige) Antillen uitgesplitst naar de subgroepen met één dan wel twee in het buitenland geboren ouders.

onderstreept nog eens dat de correlatie tussen integratiemaat en citoscore bij de eerste twee methoden geen artefact is van het ijken op de citoscore.

Tot slot is er een vierde methode gebruikt. Voor de vierde methode is de som S_L bepaald voor de nettobijdragen tot leeftijd L . Voor de vergelijkbaarheid is net als bij methode 3 gekozen voor de leeftijden tot 48 jaar, omdat van alle groepen tot tenminste 48 jaar data zijn. De maximale leeftijd L is dus 47 jaar. Vervolgens is de mate van integratie bepaald als verhoudingsgetal ten opzichte van de overeenkomstige sommen voor autochtonen en de eerste generatie niet-westers. Daarbij is bij de laatste groep het bedrag voor de leeftijden 0 en 1 gelijk gesteld aan autochtonen, om zodoende rekening te houden met de kosten voor geboortezorg e.d. Deze methode wijkt dus fundamenteel af van de eerste drie omdat er niet werkt gewerkt met interpolatie van referentieprofielen. Omdat er zeker voor veertigers soms sprake is van zeer kleine aantallen is bij deze methode voor de 42-delings (minus Nederland) geijkt op het aandeel havo/vwo-leerlingen minus het aandeel PRO/SVO/LWOO-leerlingen voor de leeftijd 15 jaar. Dit verschil geeft een zeer goede indicatie van de latere hoogst behaalde opleiding en daarmee van de nettobijdrage. De leeftijd L die de hoogste correlatie geeft is als grens aangehouden en dat blijkt het geval te zijn voor $L = 41$ jaar (ijking op de citoscores geeft overigens ook $L = 41$ jaar). Tot 36 jaar zijn er drie groepen met kleine N : Japan, Voormalige Sovjetunie en Hoorn van Afrika en Soedan. Deze hebben tot 40 jaar een vrij regelmatig verloop. Vanaf 36 jaar en vooral vanaf 40 jaar zijn er meer groepen met een kleine N en bij deze methode wordt daar niet voor gewogen. Kennelijk wordt het voordeel van meer data vanaf 42 jaar overtroffen door het nadeel van cohorteffecten en toenemende variantie door kleine N en verlaagt dat de samenhang met het schoolniveau van de 15-jarigen.



Figuur 2.10 Samenhang tussen vier methoden om de mate van integratie van de tweede generatie te schatten. De vierde methode (Meth4) is direct gebaseerd op de waargenomen nettobijdrage over de eerste 41 leeftijdjaren en vormt de laagste schatting; de overige drie methoden geven derhalve een overschatting van de mate van integratie.

Deze vierde methode geeft ook een indicatie van de mate van overschatting van de eerste drie methoden. Immers, deze methode is direct gebaseerd op de waargenomen nettobijdrage voor de eerste 41 leeftijdjaren. De regressielijn voor de vierde methode ligt zeker voor de groepen met een lage schatting voor de mate van integratie ruim onder die van de andere drie methoden. Dit geeft aan dat het gemiddelde van de vier methoden zeker voor de groepen met een geringe geschatte integratie vrijwel zeker een overschatting is.

Indien men alleen uit zou gaan van de waarnemingen over de leeftijdjaren dat er voor alle groepen (m.u.v. Japan, Voormalige Sovjetunie en Hoorn van Afrika en Soedan) redelijk veel data is, dan zou dit leiden tot veel lagere schattingen voor de mate van integratie. De doorgetrokken zwarte lijn Figuur 2.10 geeft bijvoorbeeld de trendlijn (zonder puntenwolk) als men alleen uitgaat van de waarnemingen tot 36 jaar. Dan valt het integratiepercentage voor de groepen met de geringste geschatte mate van integratie zo'n 25 procentpunten lager uit dan bij de nu gebruikte integratiemaat het geval is.

Als er mettertijd meer data beschikbaar komt, zal blijken welke schattingen het best de werkelijkheid benaderen. In ieder geval is zeker dat in methode 2 – die gebaseerd zijn op leeftijden tot 56 jaar – een deel van de cohorteffecten is meegewogen in de kwadratensom, waardoor dit hoogstwaarschijnlijk een overschatting is. Voor methode 1 geldt hetzelfde, maar die methode is zo ingericht dat hij minder gevoelig is voor cohorteffecten en dat levert met name voor de groepen met een geringe (geschatte) mate van integratie lagere uitkomsten op. Ook in methode 3 – die gebaseerd is op leeftijden tot 48 jaar – worden cohorteffecten meegewogen, maar minder en dat levert een lagere schatting op. Methode 4 gebruikt directe waarnemingen van de nettobijdrage per leeftijdjaar tot 42 jaar en geeft weer een lagere schatting dan methode 3. En de variant van methode 4 die is gebaseerd op de waarnemingen tot 36 jaar geeft een nog lagere schatting. Duidelijk is: hoe lager de maximale leeftijd, hoe minder cohorteffecten en hoe lager de integratieschatting.

Hoe lager de maximale leeftijd, hoe dichter men ook in de buurt zit van de waarnemingen voor 'doelgroep' van deze rekenexercitie: de in Nederland geboren (tweede generatie) kinderen van immigranten die gedurende de periode 1995-2040 immigrerden of zullen immigreren. Het is in zijn algemeenheid ook weer niet zo dat data over een lager maximaal leeftijdjaar altijd per se een betere schatting zal geven voor de voornoemde tweede generatiegroep. Neemt men een lagere bovengrens, dan gaan mogelijke verstoringen in het referentieprofiel eerste generatie niet-westers door bijvoorbeeld leeringgewichten, criminaliteitscijfers en het aandeel niet-EER studenten wellicht weer zwaarder meewegen. Vandaar ook dat bij methode 4 gekozen is voor verankering in het onderwijs niveau op de middelbare school, waarvan het voorspellend vermogen voor de nettobijdrage over de levensloop in ieder geval vaststaat.

In methode 1 wordt voor het maximaal te verantwoorden aantal leeftijdjaren data meegenomen met wegfilteren van de cohorteffecten, correctie voor kleine N en ijking op citoscores. Deze methode is te beschouwen als een optimistische bovengrens. Methode 4 is een in middelbare schoolprestaties verankerde methode waarin een gemiddeld aantal leeftijdjaren is meegenomen, waarbij 42 jaar ook ongeveer samenvalt met voor cruciale groepen (Surinamers, Turken, Marokkanen) de cohorteffecten duidelijk zichtbaar zijn. Deze methode is te beschouwen als een voorzichtig optimistische ondergrens. De variant van methode 4 gebaseerd op waarnemingen tot 36 jaar (de doorgetrokken zwarte lijn Figuur 2.10) is te beschouwen als een wellicht wat pessimistische ondergrens.

Uit simulatie blijkt dat in voor het basisscenario met remigratie het effect per procentpunt integratie op het totaalbedrag gemiddeld orde grootte € 1.000 bedraagt en voor de groepen met de geringste geschatte mate van integratie maximaal orde grootte € 2.000. Er van uitgaande dat de range meest waarschijnlijke uitkomsten ligt tussen de eerste en vierde methode, kunnen de bedragen maximaal € 6.000 hoger of € 20.000 lager zijn. Die range geldt vooral voor de groepen met de geringste geschatte mate van integratie, voor de overige groepen is de range veel kleiner.

Vervolgens zijn de uitkomsten van de vier methoden met elkaar vergeleken. Ze blijken onderling zeer sterk te correleren en ook sterk te correleren met de citoscore, zie Tabel 2.1. Het gemiddelde van de vier in het voorgaande besproken methoden is gebruikt als operationalisering van de mate van integratie van de tweede generatie. Dit gemiddelde vertoont een sterke samenhang ($r(44) = .91, p \ll .0001$) met de citoscores van de tweede generatie.

Tabel 2.1 Correlaties tussen vier methoden om de mate van integratie te schatten van de tweede generatie, het gemiddelde van die vier methoden (Gem. 1-4) en de citoscores, voor de 42-deling, exclusief Nederland, waarbij voor Indonesië, Turkije, Marokko, Suriname en de (voormalige) Antillen is uitgesplitst naar het aantal in het buitenland geboren ouders.

	Gem. 1-4	Methode 1	Methode 2	Methode 3	Methode 4	Cito
Gem. 1-4	1					
Methode 1	0,96	1				
Methode 2	0,97	0,97	1			
Methode 3	0,98	0,96	0,99	1		
Methode 4	0,98	0,95	0,96	0,97	1	
Cito	0,91	0,89	0,88	0,90	0,90	1

De in het voorgaande uiteengezette operationalisering van de mate van integratie van de tweede generatie wordt gebruikt om voor (ruwweg) de tweede helft van het leven het nettobijdrageprofiel van de tweede generatie in te vullen. Dit is als volgt gedaan. Als eerste stap is voor elke groep een extrapolatieprofiel *Extrapol* gemaakt op basis van de schatting van de mate van integratie. Dit *Extrapol* profiel is een lineaire combinatie van het nettobijdrageprofiel voor autochtonen en het nettobijdrageprofiel voor de eerste generatie niet-westers. Deze profielen zijn scenario-afhankelijk gemaakt, dus met gebruik van de discontovoet, sterftekansen, enzovoort voor het betreffende scenario (en dus niet zoals in het voorgaande op basis van de ruwe data voor de nettobijdrage). Feitelijk worden hiervoor de profielen gebruikt van personen met de karakteristieken (qua inkomen, belastingen, voorzieningengebruik, zorgkosten, enzovoort) van de gemiddelde autochtoon c.q. eerste generatie niet-westerse immigrant, maar dan zónder remigratie en mét volledige AOW-rechten. Daarbij wordt voor een nul-jarige de contante waarde van de nettobijdrage berekend voor elk toekomstig levensjaar *afzonderlijk*, met verdiscontering van sterftekansen en dergelijke.

Als tweede stap is een nettobijdrageprofiel *Gen2* gemaakt voor de tweede generatie van de betreffende groep volgens hetzelfde scenario, op basis van de waarnemingen. Van de profielen *Extrapol* en *Gen2* is vervolgens het gewogen gemiddelde *NBgen2* genomen. Er is gewogen naar het aantal waarnemingen N per leeftijd. Als N 100 of groter was, is voor profiel *Gen2* het gewicht 1 gehanteerd en als N onder 100 lag is gewicht $N / 100$ gebruikt. Zo is voorkomen dat kleine aantallen een te grote invloed zouden hebben. Daarnaast is er gewogen met de sigmoïde functie $(1 + e^{L-35})^{-1}$ voor leeftijden L voor

een soepele overgang tussen waarnemingen *Gen2* (tellen vrijwel volledig mee tot 30 jaar) en extrapolatieprofiel *Extrapol* (telt vrijwel volledig mee voor leeftijden vanaf 40 jaar).

Op de hiervoor geschetste manier is het profiel zoveel mogelijk gebaseerd op de echte data voor de eerste een derde deel van het leven en verder aangevuld op basis van de geschatte mate van integratie. Het voordeel hiervan is dat bedragen in het resterende deel van het leven door sterfte en het contant maken minder zwaar meetellen dan de bedragen in de eerste helft van het leven, waardoor het effect van eventuele schattingsfouten verkleind zal worden. Ter illustratie: bij een discontovoet van 2,5% en een groeivoet van 1,0% is het gewicht van de CBS-tafelbevolking van 2016 gewogen tegen de in vorige alinea gedefinieerde sigmoïde functie 58%. Anders gezegd: 58% van de nettobijdrage van de tweede generatie is gebaseerd op directe waarneming en 42% is gebaseerd op extrapolatie, gebaseerd op directe waarnemingen tot (afhankelijk van de methode) het 42^e ... 56^e leeftijdjaar.²⁹ Voorgaande methoden leiden tot een schatting van de mate van integratie voor de 42-deling in herkomstregio's (minus Nederland). Voor herkomst Nederland is de mate van integratie op 100 gesteld (volledige integratie), dit te behoeve van berekeningen voor de referentie-autochtoon, d.w.z. de hypothetische immigranten met de karakteristieken van de gemiddelde autochtoon. Deze schatting voor de 42-deling vormt de basis voor de mate van integratie voor de overige herkomstgroepen naar herkomst en/of motief. Dit is als volgt gedaan.

Eerst zijn de integratieschattingen voor de regio-indelingen die boven de 42-deling liggen gemaakt, synthetisch als gewogen gemiddelde van de schattingen voor de 42-deling. Bij die weging is gebruik gemaakt van de omvang van de bevolking tot 25 jaar. De bedoeling is immers om de beste schatting te maken voor in Nederland geboren (tweede generatie) kinderen van immigranten die immigrerden of zullen immigreren in de periode 1995-2040, vandaar de nadruk op jongen mensen. Voor de vijf grote herkomstlanden Indonesië, Turkije, Marokko, Suriname en de (voormalige) Antillen is de schatting het gewogen gemiddelde van de groep met één en de groep met twee in het buitenland geboren ouders. Voor de overige samengestelde regio's in de 12-deling – zoals bijvoorbeeld de Europese Unie of Afrika (excl. Marokko) – is het gewogen gemiddelde op basis van de 42-deling de preferente schatting, omdat die voor deze regio's het meeste detail biedt. Voor deze groepen is juist de schatting op basis van de uitsplitsing naar het aantal in het buitenland geboren ouders geijkt op de synthetische schatting op basis van de 42-deling.

Voorts is op basis van de schatting van de mate van integratie voor de 42 wereldregio's ook een schatting gemaakt voor combinaties van migratiemotief en herkomstregio's, zoals bijvoorbeeld 'arbeidsmigratie uit de Europese Unie'. De voorgaande methoden kunnen niet rechtstreeks worden toegepast op migratiemotief vanwege een gebrek aan gegevens. Migratiemotief wordt immers pas sinds 1995 bijgehouden, zodat er in 2016 – het peiljaar van deze studie – slechts voor ruim 20 jaar gegevens zijn.

Om ook voor migratiemotieven tot een zo adequaat mogelijk schatting te komen, is daarom gebruik gemaakt van de observatie dat voor de 42-deling in herkomstregio's een sterke samenhang bestaat tussen de citoscore van de tweede generatie en de mate van integratie van de tweede generatie. Voor de tweede generatie laat enkelvoudige regressie van het geschatte integratiepercentage op de

²⁹ Dezelfde simulatie laat zien dat bij gebruik van de sigmoïde functie $(1 + e^{L-42})^{-1}$ (42 jaar is gebruikt als grenswaarde in methode 4) zelfs 66% van de totale nettobijdrage van de tweede generatie direct op waarnemingen is gebaseerd, maar in dat geval gaan de cohorteffecten zichtbaar een rol spelen. Vandaar dat is ingeschat dat minder waarnemingen meenemen toch leidt tot een betere schatting.

citoscore zien dat elke punt toename in de gemiddelde citoscore, een gemiddelde stijging van het integratiepercentage van circa 10,0% oplevert. Daarom is gecorrigeerd voor de citoscores voor de tweede generatie, waarvoor wel per deelgroep ruimschoots voldoende data voorhanden zijn. Dit is als volgt gedaan. Eerst is het absolute verschil in citoscore bepaald tussen de regio als geheel en de betreffende combinatie van motief en regio. Van dit verschil is 1,645 maal de standaardfout afgetrokken en als het resulterende getal positief was is de mate van integratie dienovereenkomstig aangepast, uiteraard met in achtneming van het teken van het niet-absolute verschil. Op dezelfde wijze is voor onderliggende regio's, dat wil zeggen verfijningen van de 42-deling (zijnde de 87-deling), door correctie voor het verschil in citoscores een schatting gemaakt van de nettobijdrage voor de tweede generatie, indien er voor de hogere leeftijden weinig of geen data beschikbaar zijn.

3 Onderzoekspopulatie en microdatabestanden

Voor de berekeningen in het huidige rapport is zoveel mogelijk gebruik gemaakt van CBS-microdata. Dat is zeer gedetailleerde en geanonimiseerde data, die op persoonsniveau voor (vrijwel) alle inwoners van Nederland beschikbaar is. De CBS-microdata database bestaat uit een groot aantal tabellen, die onderling met elkaar verbonden kunnen worden via sleutelvariabelen. In deze Technische appendix wordt voortdurend verwezen naar de gebruikte CBS-microdatabestanden. Op de website van het CBS is onder de noemer Microdata catalogus een uitgebreide documentatie te vinden.³⁰

Voor de fiscale kostenbatenanalyse is in beginsel uitgegaan van de op 1 januari 2016 levende bevolking van 16.979.120 personen, overeenkomstig het CBS-statline gegeven. Van deze bevolking is het CBS microdatabestand INPATAB beschikbaar met daarin inkomensgegevens van vrijwel de gehele bevolking. Dit bestand bevat echter niet de kinderen die in 2016 zijn geboren en de nul-jarigen die in 2016 zijn geïmmigreerd. Om deze ook mee te nemen in de berekening zijn ook alle nul-jarigen meegenomen uit het CBS microdatabestand ZVWZORGKOSTENTAB. Zodoende kunnen de zorgkosten voor nul-jarigen in de analyse worden meegenomen. Op die wijze zijn 170.976 nul-jarigen toegevoegd. Deze twee groepen – de op 1 januari 2016 levende bevolking van 16.979.120 personen en de 170.976 in 2016 geboren personen waarvan zorgkosten bekend zijn – vormen in het huidige rapport de onderzoekspopulatie, zie Tabel 3.1.

Tabel 3.1 Onderzoekspopulatie.

	Onderzoekspopulatie		Niet behorend tot de bevolking op 1 januari 2016	Totaal
	0 jaar*	1 jaar en ouder*		
Onderzoekspopulatie	170.976	16.979.120		17.150.096
Waarvan in CBS-microdatabestand:				
INPA2016TABV1		16.979.120		16.979.120
ZVWZORGKOSTEN2016TABV1	170.976	16.765.505	148.391	17.084.872

*Leeftijd ultimo 2016

Voor de onderzoekspopulatie zijn – naast INPATAB – verder de gegevens gebruikt uit een aantal andere CBS-microdata bestanden. Op basis van eveneens ZVWZORGKOSTENTAB zijn ook de zorgkosten van personen behorende tot de onderzoekspopulatie die ultimo 2016 één jaar of ouder waren toegevoegd. Voor de persoonsgegevens is gebruik gemaakt van GBAPERSOONTAB en KINDOUDERTAB. Verder zijn gegevens gebruikt over onderwijsdeelname, citoscores en de hoogst behaalde opleiding is gebruikt gemaakt ONDERWIJSDEELNEMERSTAB, CITOTAB en HOOGSTEOPL. Daarnaast is gebruik gemaakt van een groot aantal andere microdatabestanden: INHATAB voor de inkomens van huishoudens, KINDEROPVANG voor aan aanvragers en partners toegekende bedragen voor kinderopvang, GBAMIGRATIEBUS en VRLMIGMOTBUS voor migratie en migratiemotieven en tenslotte het pensioenaanspraken bestand PAS.

³⁰ De toegang tot deze data kan aangevraagd worden bij het CBS. Bij het gebruik blijft alle microdata om privacy redenen binnen het CBS. Alleen geaggregeerde data die niet herleidbaar zijn tot personen mag onder strikte voorwaarden en na controle door het CBS worden gepubliceerd. Voor replicatie is dus toegang tot deze microdata nodig. Zie voor meer informatie de website van het CBS en de catalogus microdata: <https://www.cbs.nl/nl-nl/onze-diensten/maatwerk-en-microdata/microdata-zelf-onderzoek-doen/catalogus-microdata>.

Voor personen die in 2016 zijn geboren of geïmmigreerd zijn alleen de zorgkosten meegenomen alsmede alle kosten die in gelijke mate aan alle personen worden toegekend, zoals de kosten voor openbaar bestuur. Voor personen die ultimo 2016 één jaar of ouder waren zijn de gegevens uit alle gebruikte microdatabestanden alleen meegenomen voor zover de betreffende personen voorkwamen in INPATAB, verslagjaar 2016, dit omdat het weinig zinvol werd geacht om personen in de berekening mee te nemen waarvoor geen informatie is over voor de kosten en baten cruciale zaken als inkomen, belastingen, premies, uitkeringen, subsidies en dergelijke. Zoals gezegd resulteert die aanpak in een onderzoekspopulatie die gelijk is aan het CBS-statline gegeven voor 2016 plus alle nul-jarigen (ultimo 2016) uit het CBS microdatabestand ZVWZORGKOSTENTAB. In de volgende hoofdstukken wordt de operationalisering van de verschillende variabelen verder toegelicht. Als in het vervolg aan ZVWZORGKOSTENTAB e.d. wordt gerefereerd, wordt geduid op de data voor zover deze betrekking heeft op personen in de onderzoekspopulatie.

4 Socio-demografische variabelen

4.1 Leeftijd en generatie

De sociaal-demografische variabelen zijn toegevoegd op basis van het CBS-microdatabestand GBAPERSOONTAB voor het verslagjaar 2017, dat onder andere de variabelen herkomst (eerste en tweede) generatie en geboortjaar en -maand bevat. Uit geboortjaar en geboortemaand is de leeftijd te berekenen. De geboortedag ontbreekt in het microdatabestand en als benadering is verondersteld dat iedereen op de 15^e van de maand geboren is. Hieruit is de leeftijd ultimo 2016 berekend. Ook de variabele generatie is opgenomen in GBAPERSOONTAB, voor zover het de eerste en tweede generatie betreft, evenals het aantal in het buitenland geboren ouders van de tweede generatie.

De derde generatie is afgeleid door combinatie van verschillende CBS-microdatabestanden. Het CBS geeft de volgende [definitie van de derde generatie](#): “Iemand behoort tot de derde generatie als beide ouders in Nederland zijn geboren en ten minste één van de vier grootouders in het buitenland is geboren. Mensen die in Nederland zijn geboren en van wie minstens één ouder in het buitenland geboren is, behoren dus niet tot de derde generatie, maar tot de tweede.” De derde generatie maakt deel uit van de groep mensen met een [Nederlandse achtergrond](#), die gedefinieerd zijn als: “Persoon van wie de beide ouders in Nederland zijn geboren, ongeacht het land waar men zelf is geboren.” Behalve een Nederlandse achtergrond geldt voor iemand van de derde generatie dat één of beide ouders een [tweede generatie migratieachtergrond](#) hebben, hetgeen is gedefinieerd als een “Persoon die in Nederland is geboren en van wie ten minste één ouder in het buitenland is geboren.” Combinatie van de twee laatste definities levert de eerste definitie.



Figuur 4.1 Schematische weergave van personen met een Nederlandse achtergrond.

Bron: www.cbs.nl. (2018) <https://www.cbs.nl/nl-nl/nieuws/2016/47/wie-zijn-de-derde-generatie>- Benaderd: 25-11-2018. Bewerkte uitsnede uit oorspronkelijke afbeelding.

Het [CBS specificeert de achtergrond van de derde generatie](#) nader aan de hand van de achtergrond van de ouders “waarbij de achtergrond van de moeder bepalend is, tenzij zij een Nederlandse of onbekende achtergrond heeft. Dan nemen we de achtergrond van de vader.” Deze systematiek is ook in huidige studie gevolgd.

De derde generatie is afgeleid uit de informatie uit het bestand GBAPERSOONTAB dat de herkomst van vader en moeder bevat voor zover bekend. Hieruit is met behulp van het CBS microdatabestand

KINDOUDERTAB af te leiden wie tot de derde generatie behoort. Per definitie behoren diegenen tot de derde generatie waarvan beide ouders tweede generatie zijn of waarvan één ouder tweede generatie is en één ouder een Nederlandse achtergrond heeft. Vervolgens is de herkomst van het kind gelijkgesteld aan de herkomst van de moeder, tenzij zij een Nederlandse of onbekende achtergrond had, in die gevallen is de herkomst van het kind gelijkgesteld aan de herkomst van de vader. Er is ook een groep waarvan niet is vast te stellen of ze tot de derde generatie behoren. In de praktijk is in huidige studie de derde generatie beperkt tot personen waarvan de generatie van beide ouders bekend is, waarbij onderscheid is gemaakt tussen één tweede generatie ouder en twee tweede generatie ouders.³¹ Er zijn overigens weinig publieke data beschikbaar voor de derde generatie, behoudens een enkel CBS-statline bestand³² of CBS-maatwerk bestand.³³

4.2 Opleidingsniveau en citoscores

Opleidingsniveau is afgeleid uit het CBS-microdatabestand HOOGSTEOPL voor de verslagjaren 2009-2017. Dit bestand geeft onder andere het hoogst behaalde opleidingsniveau per 1 oktober van het verslagjaar. Voor een aantal deelberekeningen is daarnaast ook gebruikt gemaakt van het CBS-microdatabestand ONDERWIJSDEELNEMERSTAB voor de verslagjaren 2009-2017 om de actueel gevolgde opleiding te bepalen. In beginsel is daarbij gebruik gemaakt van de CBS SOI 8-deling.

De citoscores zijn afgeleid uit het CBS-microdatabestand CITOTAB voor de verslagjaren 2006-2018. Deze bestanden bevatten de citoscores van in totaal 1,8 miljoen personen. Deze data zijn geaggregeerd voor de variabele CitoStandaardScore (bereik 501-550) en andere relevante variabelen.³⁴

Voor het bepalen van de hoogste opleiding naar citoscore zijn drie cohorten uit CITOTAB gevolgd, die 20 jaar oud waren ultimo 2015, 2016 en 2017 respectievelijk (totaal 399.413 personen). Van deze cohorten is de distributie bepaald over de Hoogst Behaalde dan wel Actueel gevolgde opleiding (afgekort HBA) op 20-jarige leeftijd.³⁵ Vervolgens is op basis van drie andere cohorten een transformatiematrix bepaald van de HBA voor 20-jarigen naar de HBA voor de leeftijden 21-28 (voor de laatste vijf jaren is de 'met NIRWO' variant gebruikt).³⁶ Deze stap is tweemaal herhaald voor de leeftijden 29-38 jaar. Bij de laatste twee stappen konden slechts vijf jaren worden gebruikt doordat de 'met NIRWO'³⁷ variant slechts beschikbaar was voor de jaren 2013-2017. Door matrixvermenigvuldiging (uitgevoerd in Excel) zijn vervolgens transformatiematrices bepaald van citoscores naar HBA voor de leeftijden 21-38 jaar.

³¹ Vergelijk met de toelichting bij de CBS-dataset Tabellen Derde Generatie over de derde generatie, <https://www.cbs.nl/nl-nl/maatwerk/2020/46/personen-met-ouders-van-tweede-generatie-1-januari-2020> en https://www.cbs.nl/-/media/_excel/2016/47/maatwerktabellen%20derde%20generatie%20-%201%20januari%202016.xlsx zie ook: <https://www.cbs.nl/nl-nl/nieuws/2016/47/wie-zijn-de-derde-generatie->

³² <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/70688ned/table?dl=18BAC>

³³ Voor 2016 en 2020 zijn er maatwerkbestanden zie voorgaande noten en zie: <https://www.cbs.nl/nl-nl/maatwerk/2020/46/personen-met-ouders-van-tweede-generatie-1-januari-2020>

³⁴ Met het oog op eventuele replicatie is het van belang te weten dat de regio's Israël en Arabisch schiereiland, Jordanië en Libanon aanvankelijk één regio vormde die achteraf is opgesplitst naar twee afzonderlijke regio's en in een enkel geval de data voor deze regio's door imputatie of afleiding zijn verkregen.

³⁵ Voor deze stap is dit gedaan voor een *customized* HBA indeling in 10 actueel gevolgde opleidingsniveaus en 9 hoogst behaalde opleidingsniveaus, op basis van celvullingen samengesteld uit een combinatie van SOI 8-deling en de SOI 18-deling, dit omdat het precieze schoolniveau voor deze leeftijden nog zeer bepalend is, maar het gebruik van alleen de SOI 18-deling voor een aantal gevallen tot te weinig gevulde cellen zou leiden.

³⁶ Voor deze stap is een transformatiematrix gemaakt voor de in vorige voetnoot toegelichte *customized* HBA indeling in 19 groepen naar de reguliere HBA-indeling in 16 groepen (8 voor Hoogst Behaalde en 8 voor Actueel gevolgde opleiding).

³⁷ Zie de CBS-microdatadocumentatie.

Met deze matrices kan dus op basis van een bepaalde citoscore voor alle genoemde leeftijden de verdeling over de hoogst behaalde dan wel actueel gevolgde opleiding worden voorspeld. Dit is gedaan voor de bevolking als geheel, voor autochtonen en voor de eerste en tweede generatie, al dan niet uitgesplitst naar westers en niet-westers (vanaf nu in deze paragraaf kortweg aangeduid als 'voor de hele bevolking en naar migratieachtergrond').

Op basis van aggregatie van nettobijdragedata naar citoscore (tot 24 jaar) en het gebruik van voornoemde transformatiematrices (voor leeftijden vanaf 24 jaar) zijn nettobijdrageprofielen opgesteld per citoscore, voor de hele bevolking en naar migratieachtergrond. De eerste stap was mogelijk door directe aggregatie van de nettobijdragedata naar citoscore en eventuele andere variabelen.³⁸ Voor leeftijden vanaf 24 jaar is dit gedaan door per leeftijd een matrixvermenigvuldiging uit te voeren van voornoemde transformatiematrices met matrices met contant gemaakte nettobijdragen per levensjaar (24 – 99 jaar) naar hoogst behaalde opleiding (volgens de SOI 8-deling). Dit is gedaan voor de hele bevolking en naar migratieachtergrond.

Verder zijn leeftijdsprofielen geconstrueerd voor de nettobijdrage naar het hoogst behaalde opleidingsniveau voor de hele bevolking en naar migratieachtergrond. Deze profielen zijn nodig om de stappen in de vorige alinea te kunnen uitvoeren en daarnaast om inzicht te krijgen in de nettobijdrage naar opleidingsniveau en migratieachtergrond. Deze profielen zijn in beginsel samengesteld uit de nettobijdrageprofielen voor de actuele (ACT) en hoogst behaalde (HB) opleidingen volgens de CBS SOI 8-deling, die vanaf nu in de rest van deze paragraaf zullen worden aangeduid als 'ACT-profiel' respectievelijk 'HB-profiel'.³⁹ Dit is als volgt gedaan.

Voor de leeftijden 0 tot 4 jaar is bij de constructie van de profielen uitgegaan van de bevolking als geheel. Voor de leeftijden 4 tot 11 jaar is het ACT profiel gebruikt voor basisonderwijs, gecorrigeerd voor de gemiddelde citoscore van de uiteindelijke hoogst behaalde opleiding. Voor de leeftijden van 12 tot maximaal 23 jaar is in beginsel uitgegaan dat één van de volgende twee nominale studiepaden gevolgd wordt tot de betreffende hoogst behaalde opleiding (ten behoeve van de leesbaarheid zijn de niveaus in de CBS SOI 8-deling onderstreept):

- Het eerste studiepad is basisonderwijs, eventueel gevolgd door vijf jaar vmbo b/k, mbo1, eventueel gevolgd door twee jaar mbo2 en mbo3, eventueel gevolgd door één jaar mbo4.
- Het tweede studiepad is drie jaar vmbo g/t, havo-, vwo-onderbouw, eventueel gevolgd door drie jaar havo, vwo, eventueel gevolgd door drie jaar hbo-, wo-bachelor, eventueel gevolgd door één jaar hbo-, wo-master, doctor.

Indien het eerste studiepad zonder diploma wordt afgesloten is de hoogst behaalde opleiding basisonderwijs, waarbij in verband met de leerplicht verondersteld is dat er altijd vijf jaar vmbo b/k, mbo1 gevolgd is. Vervolgens zijn voor leeftijden tot 21 of 22 jaar alle profielen indien nodig aangevuld van

³⁸ Hiervoor was data voor personen van 4 tot 24 jaar. Deze data zijn voor personen van 0 tot 4 jaar aangevuld met data van de bevolking als geheel.

³⁹ Er zijn totaal dus 16 nettobijdrageprofielen, twee voor elk niveau in de SOI 8-deling. Concreet: er is voor het niveau havo, vwo een profiel voor de actueel gevolgde opleiding (ACT-profiel) en een profiel voor de hoogst behaalde opleiding (HB-profiel).

het nettobijdrageprofiel gebaseerd op de gemiddelde citoscore bij de betreffende hoogst behaalde opleiding.⁴⁰

Deze twee nominale studiepaden zijn vervolgens als volgt aangevuld voor de resterende levensjaren. Voor de hoogst behaalde opleidingen mbo2 en mbo3, mbo4, hbo-, wo-bachelor en hbo-, wo-master, doctor is voor alle leeftijden van 21⁴¹ tot 39 jaar het gewogen gemiddelde genomen van het HB-profiel en het ACT-profiel voor de betreffende hoogst behaalde opleiding. Hierbij is gewogen op basis van de verhouding tussen hoogst behaalde en actueel gevolgde opleiding. Voor havo, vwo is alleen voor 21-jarigen het gewogen gemiddelde genomen. Voor de overige drie niveaus is geen gewogen gemiddelde genomen en is vanaf 22 jaar uitsluitend het HB-profiel gebruikt.⁴² Voor leeftijden vanaf 39 jaar is altijd alleen het HB-profiel bij de betreffende hoogst behaalde opleiding gebruikt. Bij uitsplitsing naar migratieachtergrond was er voor leeftijden vanaf 72 jaar weinig data. In deze gevallen is het nettobijdrageprofiel aangevuld door het best passende profiel, waarbij het profiel van de desbetreffende opleiding zonder uitsplitsing naar migratieachtergrond met gewicht 0,25 is meegewogen in de kwadratensom.⁴³ Tot slot zijn op basis van CBS-maatwerkdata de sterftekansen aangepast aan het opleidingsniveau.⁴⁴

4.3 Migratiemotief

Het migratiemotief voor de eerste generatie is afgeleid uit het CBS-microdatabestand VRLMIGMOTBUS voor de verslagjaren 2015 en 2017 en gebaseerd op de indeling in de vijf categorieën van de variabele VRLVERBLIJFSDOEL, te weten arbeid, studie, asiel, gezinsmigratie en overig. De variabele VRLVERBLIJFSDOEL is het “migratiemotief van de IND dat wordt gepubliceerd na imputatie (en andere aanpassingen).”⁴⁵

De term ‘migratiemotief’ is hier een administratief-juridisch begrip. Het is het motief op basis waarvan de immigrant een verblijfsvergunning heeft aangevraagd en verkregen en dat sluit niet uit dat andere beweegredenen een rol hebben gespeeld bij de migratiebeslissing. Daarnaast betreft het hier door het CBS geïmputeerde data, dat wil zeggen dat het CBS op basis van de karakteristieken van de immigranten waarvan het motief bekend is, een schatting heeft gemaakt van het meest waarschijnlijke motief

⁴⁰ Deze gemiddelde citoscore naar hoogst behaalde opleiding is verkregen uit de eerder beschreven matrixberekeningen. Dit wordt wegens gebrek aan data tot leeftijden van 21 of 22 jaar gebruikt als proxy voor het gewogen gemiddelde van het ACT-profiel en het HB-profiel dat voor leeftijden vanaf 21 jaar wordt gebruikt. Om meetfouten te minimaliseren is bij de berekening het gemiddelde genomen van de betreffend gemiddelde citoscore en de twee scores direct erboven en eronder.

⁴¹ Bij hbo-, wo-master vanaf 22 jaar.

⁴² NB: dit levert geen grote vertekening op: dat er voor bepaalde leeftijden geen data zijn voor deze ACT-profielen, impliceert immers dat het volgen van deze opleidingen voor de betreffende leeftijden in de praktijk niet of nauwelijks voorkomt en dus dat het HB-profiel derhalve een adequate beschrijving van de werkelijkheid geeft.

⁴³ De methode is vergelijkbaar zoals toegepast voor de tweede generatie, zie §2.4.

⁴⁴ De levensverwachting is lager is voor lager opgeleiden en hoger voor hoger opgeleiden. Op basis van publiek beschikbare data is een inschatting gemaakt van de grootte van dit effect en de sterftekansen zijn dienovereenkomstig gecorrigeerd.

⁴⁵ CBS, *Vrlmigmotbus: Migratiemotieven*, opgehaald 12-2-2021 van: <https://www.cbs.nl/nl-nl/onze-diensten/maatwerk-en-microdata/microdata-zelf-onderzoek-doen/microdatabestanden/vrlmigmotbus-migratiemotieven>

voor de immigranten waarvan het motief niet bekend is.⁴⁶ Tot slot zijn deze data pas vanaf 1995 beschikbaar.⁴⁷

Het migratiemotief is een lastige variabele die regelmatig door het CBS wordt herzien.⁴⁸ Er zijn verschillende variabelen die verwant zijn met het migratiemotief en die onder verschillende benamingen in de CBS-documentatie voorkomen, waaronder migratiemotief, migratiereden, verblijfsdoel en afgeleid migratiedoel. Vergelijkingen van opeenvolgende, op het eerste gezicht ogenschijnlijk verwante tabellen tonen dan ook voor hetzelfde verslagjaar en voor vergelijkbare categorieën zoals ‘Aziatische arbeidsmigranten’ andere uitkomsten, mede omdat tabellen soms migratiemotief en soms migratiereden rapporteren en er soms wordt uitgesplitst naar geboorteland⁴⁹, soms naar nationaliteit⁵⁰ en soms naar migratieachtergrond⁵¹.

Daarbij komt dat voor bepaalde groepen geldt dat nogal wat immigranten andere migratiemotieven hebben naast of in plaats van het migratiemotief dat gebaseerd is op de IND-statistieken. Het CBS publiceert dan ook statistieken over het afgeleid migratiedoel gebaseerd op het daadwerkelijke gedrag.⁵² Bij deze statistieken wordt niet gekeken naar het motief (anders dan asiel) dat mensen opgaven bij de IND, maar naar andere gegevens zoals de belangrijkste bron van inkomsten of het feit dat iemand studeert.

In deze studie is voor de eenduidigheid alleen gebruik gemaakt van de variabele VRLVERBLIJFSDOEL uit VRLMIGMOTBUS, d.w.z. van het “migratiemotief van de IND”, zowel voor het maken van de leeftijdsprofielen voor de nettobijdrage als ook voor het schatten van de verdeling over migratiemotieven per combinatie van migratiejaar en herkomstgroep. Dit is ook de meest beleidsrelevante benadering, omdat deze operationalisering direct gebaseerd is op de door de IND gecontroleerde gronden van het toelatingsbeleid.

Tabel 4.1 Het voorkomen van combinaties van migratiemotief gezinsmigratie met andere motieven (arbeid, asiel, studie en overig) als percentage van alle mogelijke combinaties, voor vaders en moeders van tweede generatie immigranten bij gelijktijdige en niet-gelijktijdige immigratie.

	percentage van alle mogelijke combinaties in geval de immigratiedata van de ouders		
	verschillend zijn	gelijk zijn	totaal
moeder gezinsmigratie, vader ander motief	39%	36%	38%
vader gezinsmigratie, moeder ander motief	10%	6%	9%
verhouding	3,9	6,1	4,2

⁴⁶ Voor niet-EU/EFTA immigranten is dit in ongeveer 1% van de gevallen het motief onbekend, CBS *Statistiek Migratiemotieven*, opgehaald 12-2-2021 van: <https://www.cbs.nl/nl-nl/onze-diensten/methoden/onderzoeksomschrijvingen/korte-onderzoeksbeschrijvingen/statistiek-migratiemotieven>

⁴⁷ Zie verder de CBS-microdata documentatie.

⁴⁸ Vergelijk bijvoorbeeld Bevolkingsprognose 2017-2060: veronderstellingen migratie, CBS voor meer over de migratiereden / het migratiemotief en de CBS-microdata documentatie.

⁴⁹ <https://opendata.cbs.nl/#/CBS/nl/dataset/70693ned/table?dl=3F3D4>

⁵⁰ <https://opendata.cbs.nl/#/CBS/nl/dataset/84809NED/table?dl=3F3D5>

⁵¹ <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/84140NED/table?dl=3F3D6>

⁵² CBS, *Statistiek Migratiemotieven*, opgehaald 12-2-2021 van: <https://www.cbs.nl/nl-nl/onze-diensten/methoden/onderzoeksomschrijvingen/korte-onderzoeksbeschrijvingen/statistiek-migratiemotieven>

Omdat mensen soms meer dan eens immigreren, is alleen het laatst opgegeven motief gebruikt. Het motief is afzonderlijk bepaald voor nul-jarigen (ultimo 2016) en voor personen van één jaar of ouder (ultimo 2016). Voor nul-jarigen zijn de in 2016 opgegeven motieven uit VRLMIGMOTBUS verslagjaar 2017 gebruikt. Voor personen van 1 jaar of ouder is gebruik gemaakt van het laatst opgegeven motief tot ultimo 2015, d.w.z. de motieven na ultimo 2015 zijn eerst uit VRLMIGMOTBUS verslagjaar 2017 verwijderd en daarna is het laatste migratiemotief bepaald. Dit is gedaan omdat voor leeftijden vanaf één jaar de analyse de in Nederland verblijvende populatie op 1 januari 2016 betreft en migratiebewegingen vanaf 1 januari 2016 derhalve irrelevant zijn.

Er is ook een migratiemotief afgeleid voor de tweede generatie. Dat is gedaan door via het CBS-microdatabestand KINDOUDERTAB de motieven van de ouders aan het kind toe te kennen. Bij het bepalen van het migratiemotief van het kind is gekozen voor de volgende definitie:

- het motief van het kind is gelijk aan het motief van de vader;
- als het motief van de vader gezinshereniging is, dan wordt het motief van de moeder gebruikt.

Deze aanpak is eenvoudig, eenduidig en makkelijk te repliceren en doet goed recht aan de empirie.⁵³ Empirisch blijkt de vader namelijk vaak doorslaggevend voor het motief. Een indicatie is dat de combinaties met 'motief moeder = gezinsmigratie, motief vader is anders' veel talrijker zijn dan de combinaties met 'motief vader = gezinsmigratie, motief moeder is anders'. Dit is zowel het geval als beide ouders een verschillende immigratiedatum hebben als ook als beide ouders dezelfde immigratiedatum hebben, zie Tabel 4.1. Kennelijk is gezinsmigratie vaak een secundair motief, waarbij één ouder komt met (hoofd)doel arbeid, asiel, studie of overig en de andere ouder komt met doel gezinsmigratie. Om die reden is het kind toegewezen aan het hoofddoel (arbeid, asiel, studie of overig) omdat dit het achterliggend motief is van het feit dat het kind een ingezetene is van Nederland.

4.4 Regio

Om te kunnen onderzoeken in welke mate de nettobijdrage afhankelijk is van de herkomstregio, is een geografische indeling van de wereld gemaakt, zie Tabel 4.2. In beginsel is hierbij uitgegaan van bestaande indelingen door het CBS. Daar waar mogelijk is verder aangesloten bij de indeling in wereldregio's die de Verenigde Naties hanteren.⁵⁴ Daarnaast is vooral ook gekeken of het aantal waarnemingen voor elke regio voldoende zou zijn. Verder is er daar waar mogelijk gekozen voor geografisch aaneengesloten gebieden om misleidende effecten bij het interpreteren van de wereldkaarten te voorkomen. Binnen de hiervoor genoemde randvoorwaarden is verder geprobeerd gelijksoortige landen bij elkaar te voegen, waarbij het expliciet gaat om economische ontwikkeling en belangrijke asielherkomstregio's.

De meest grove indeling is Nederlandse achtergrond versus migratieachtergrond. Dit is feitelijk een tweedeling in Nederland en de rest van de wereld. Verder is er de CBS-indeling westers en niet-westers (het CBS hanteert hiervoor in de microdatabestanden de term 'landtype'). Daarnaast is de CBS 12-deling gehanteerd die het CBS bijvoorbeeld ook gebruikt voor de bevolkingsprognoses. Deze indeling is een verdere verfijning van de regio's westers en niet-westers. De westerse regio's zijn: Europese Unie, Overig Europa, Indonesië, Overig buiten Europa. De niet-westerse regio's zijn Azië (excl.

⁵³ De aanpak is geïnspireerd door de toekenning van de migratieachtergrond bij de tweede generatie volgens de CBS-definitie, met het verschil dat hier op basis van de empirie het primaat bij de vader is gelegd i.p.v. de moeder.

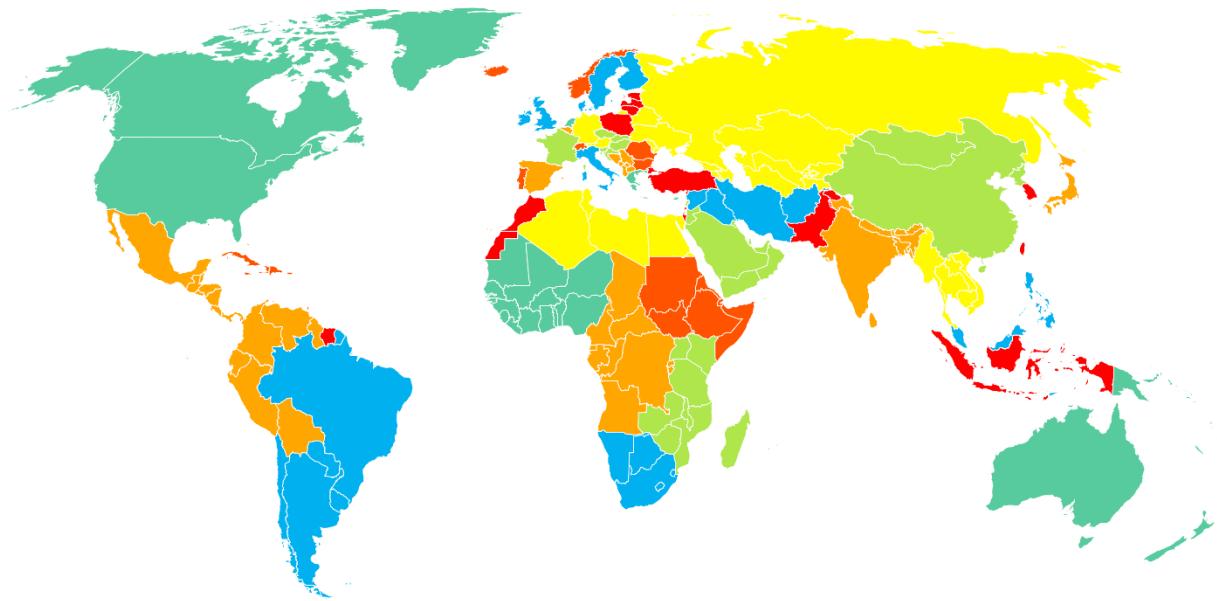
⁵⁴ Bijvoorbeeld in de bevolkingsstatistieken **BRON**.

Indonesië en Japan), Turkije, Marokko, Afrika (excl. Marokko), Suriname, Aruba en (voormalige) Nederlandse Antillen en Latijns-Amerika (excl. Suriname, Aruba en Antillen).

Tabel 4.2 Regio-indeling.

Landtype	CBS 12-deling	19-deling	42-deling
Westers	Nederland	Nederland	Nederland
	Europese Unie	West-Europa (VN-regio)	België en Luxemburg
			Duitsland en Oostenrijk
			Frankrijk
			Denemarken, Zweden en Finland
		VK, Ierland, Denemarken, Zweden en Finland	VK en Ierland
		GIPS-landen, Malta en Cyprus	Griekenland en Cyprus
			Italië en Malta
			Portugal
	MOE-landen	Spanje	
		Bulgarije en Roemenië	
		Hongarije, Tsjechië, Slowakije, Slovenië en Kroatië	
	Overig Europa	Overig Europa	Polen en Baltisch staten
EFTA, dwergstaten en crown dependencies			
Voormalig Joegoslavië (excl. Slovenië en Kroatië) en Albanië			
Overig buiten Europa	Overig buiten Europa	Voormalig Sovjet-Unie (excl. Baltische staten)	
		Noord-Amerika	
		Oceanië	
Indonesië	Indonesië	Indonesië	
Niet-westers	Azië (excl. Indonesië en Japan)	Oost-Azië	Zuid-Korea, Taiwan, Hong Kong en Singapore
		Zuidoost-Azië	China, Mongolië en Noord-Korea
		Filippijnen, Maleisië, Brunei en Timor Leste	
		Thailand, Indochina en Myanmar	
	Indisch subcontinent	Indisch subcontinent excl. Pakistan	
		Pakistan	
	West-Azië	Afghanistan, Iran, Syrië en Irak	
		Israël	
		Arabisch schiereiland, Jordanië en Libanon	
	Turkije	Turkije	Turkije
	Marokko	Marokko	Marokko
	Afrika (excl. Marokko)	Noord-Afrika (excl. Marokko)	Noord-Afrika (excl. Marokko)
			West-Afrika
		Sub-Sahara Afrika	Centraal-Afrika
			Hoorn van Afrika en Soedan
			Oost-Afrika
	Zuidelijk Afrika		
Suriname	Suriname	Suriname	
Aruba en (voormalige) Nederlandse Antillen	Aruba en (voormalige) Nederlandse Antillen	Aruba en (voormalige) Nederlandse Antillen	
Latijns-Amerika (excl. Suriname, Aruba en Antillen)	Latijns-Amerika (excl. Suriname, Aruba en Antillen)	Brazilië, Argentinië, Paraguay, Uruguay, Chili en Frans Guyana	
		Caraïben	
		Midden-Amerika en Zuid-Amerika Overig	

Deze indeling is in het huidige rapport verder onderverdeeld in een 42-deling, zoals weergegeven in Tabel 4.2 en Figuur 4.2. Dat is met een tussenstap gedaan (kolom 19-deling in de tabel). Voor Europa is daarbij zoveel mogelijk aangesloten bij bestaande indelingen van het CBS zoals GIPS-landen (Griekenland, Italië, Portugal, Spanje, in deze studie aangevuld met Malta en Cyprus) en MOE-landen (Midden- en Oost-Europese landen, in deze studie Estland, Letland, Litouwen, Polen, Tsjechië, Slowakije, Hongarije, Slovenië, Kroatië, Roemenië en Bulgarije). Verder zijn twee landen, namelijk voormalig Joegoslavië en de voormalige Sovjetunie in beginsel zoveel mogelijk niet opgesplitst, dit omdat er vaak weinig immigranten zijn uit de landen die ontstonden na het uiteenvallen. Uitzondering is gemaakt voor Kroatië, Slovenië en de Baltische staten die bij de MOE landen zijn gevoegd. Om die reden zijn de Centraal-Aziatische republieken van de Voormalige Sovjetunie ook niet apart onderscheiden. Voormalig Joegoslavië en de voormalige Sovjetunie zijn apart geanalyseerd. Deze regio's onderscheiden zich verder van de overige Europese landen omdat er veel asielzoekers uit afkomstig zijn. De regio EFTA, dwergstaten en crown dependencies is een restcategorie die min of meer dwingend voortvloeit uit bestaande CBS-indelingen.⁵⁵ Turkije is conform CBS-indelingen apart ingedeeld.



Figuur 4.2 Regio-indeling volgens de 42-deling.

Oceanië is als aparte regio opgevoerd conform de indeling in werelddelen van het CBS. In Azië zijn Japan en Indonesië in navolging van het CBS tot aparte westerse regio's bestempeld. Verder zijn de VN regio-indelingen Oost-Azië en Zuidoost Azië aangehouden (maar dan exclusief Japan en Indonesië). Binnen de regio Oost-Azië zijn de zogenaamde Aziatische tijgers Zuid-Korea, Taiwan, Hong Kong en Singapore apart genomen als sub-regio.

De VN-regio Centraal-Azië is komen te vervallen vanwege de eerder genoemde beperkingen ten aanzien van de data voor de voormalige Sovjetunie. De VN-regio West-Azië (zonder Turkije) is mede daarom uitgebreid met Afghanistan, Iran en Irak. Bij de verdere indeling van de regio West-Azië is Israël apart ingedeeld omdat het in veel opzichten verschilt van omliggende landen. Daarnaast zijn de regio's Arabisch schiereiland, Jordanië en Libanon (overwegend relatief welvarende landen) en Afghanistan,

⁵⁵ Dit heeft te maken met het vermijden van onthullingsrisico dat zou kunnen ontstaan met het combineren van regio's.

Iran, Irak en Syrië (overwegend asielmigratie) in het leven geroepen. Het resterende deel van Azië is omgedoopt tot de regio Indiaas subcontinent, verder onderverdeeld in Pakistan en de rest van het Indiaas subcontinent.

In Afrika is de indeling in Noord-, West-, Centraal- en Zuidelijk Afrika van de VN in beginsel gehanteerd. Echter, daarnaast is er een regio Hoorn van Afrika en Soedan gemaakt (Somalië, Eritrea, Ethiopië, Djibouti en Noord- en Zuid-Soedan). Dit is gedaan omdat uit deze regio veel asielmigranten afkomstig zijn. Op die manier kunnen verschillen tussen asielherkomstregio's en overige regio's inzichtelijk worden gemaakt. Daarnaast zijn Burundi en Rwanda bij Centraal-Afrika gevoegd, om te kleine aantallen waarnemingen voor bepaalde leeftijden te voorkomen. Tot slot is Marokko apart geanalyseerd, vanwege bestaande CBS-indelingen.

Voor Amerika zijn de VN-regio's Noord-Amerika en de Caraïben aangehouden. Daarnaast zijn Suriname en de voormalige Antillen onderscheiden. De VN-regio's Midden-Amerika en Zuid-Amerika van de VN zijn vanwege gebrek aan data heringedeeld. Daarbij zijn de meer ontwikkelde zuidelijke landen, te weten Chili, Frans Guyana en de landen die behoren tot de zogenaamde Mercosur douane-unie (Brazilië, Argentinië, Paraguay en Uruguay) als regio apart genomen en is de rest van Latijns-Amerika bestempeld tot de regio Midden-Amerika en Zuid-Amerika Overig.

Bij de indeling in regio speelt een afweging tussen nauwkeurigheid en betrouwbaarheid. Met een grove indeling zijn de aantallen groot, maar tegelijkertijd worden mogelijk ongelijksoortige herkomstlanden bij elkaar gevoegd. Met andere woorden, de betrouwbaarheid is hoog, maar de nauwkeurigheid niet. Bij een fijnere indeling is er meer detail, maar dat gaat als de groepen té klein zijn wel ten koste van de betrouwbaarheid. In §6.5 is gebruik gemaakt van een indeling in 87 regio's die een verfijning is van de 42-deling.

5 Kosten en baten posten

5.1 Kosten en baten posten

Bij de in het huidige rapport gemaakte berekening is zoveel mogelijk aangesloten bij de generatierekeningen van het CPB. Uitgangspunt van de berekeningen vormen leeftijdsprofielen (0-99 jaar) voor de Nederlandse bevolking voor de belangrijkste kosten- en batenposten, die ter beschikking zijn gesteld door het CPB voor de jaren 2016, 2021, 2030, 2040, 2050 en 2060. Deze leeftijdsprofielen lopen van 0 tot en met 99 jaar en geven voor elke leeftijd het bedrag voor de betreffende post in één van de zes genoemde jaren. In deze profielen zitten de verwachtingen met betrekking tot beleid en (trendmatige) groei verwerkt die het CPB had ten tijde van de vervaardiging. Vandaar dat in de huidige studie voor het peiljaar 2016 is gekozen, dat het begin vormt van de door het CPB ter beschikking gestelde tijdreeksen. Voor 2016 zijn alle posten en de bijbehorende macrobedragen (dus niet de bedragen per leeftijd) weergegeven in Tabel 5.1.

Deze CPB-(leeftijds)profielen zijn door interpolatie en extrapolatie uitgebreid naar alle voor de berekening benodigde jaren.⁵⁶ Door deling van alle bedragen door het corresponderende bedrag voor 2016 is voor elke post, elke leeftijd en elk jaar vanaf 2016 een gewicht bepaald waarmee een bedrag in 2016 vermenigvuldigd kan worden om de waarde van dat bedrag voor een toekomstig jaar te verkrijgen voor de betreffende post en leeftijd. Deze gewichtenmatrix stelt in staat om de CPB-profielen in te vullen met empirische gegevens op basis van bijvoorbeeld CBS-microdata en vervolgens de door het CPB verwachte veranderingen door economische groei, beleidsveranderingen en dergelijke door te voeren.

De indeling in posten en de CPB-profielen vormen het uitgangspunt van de berekening. Bij de operationalisering van elke post is steeds de fundamentele keuze gemaakt tussen zelf berekenen of uitgaan van de CPB-profielen. Dit is in Tabel 5.1 weergegeven in de kolom Operationalisering. Als daar alleen 'CPB-profiel' staat is onverkort het CPB-profiel voor 2016 gebruikt (dit betreft de posten nr. 2, 10, 13, 14, 22 en 23). In alle andere gevallen heeft de operationalisering geheel of gedeeltelijk plaatsgevonden op basis van empirische data zoals CBS-microdata en CBS-statline data. In die gevallen is het CPB-profiel voor 2016 geheel of gedeeltelijk vervangen door een zelf berekend leeftijdsprofiel.

Vervolgens is het resulterende totaalbedrag van het model voor het huidige rapport geïjkt om voor elk van de 23 posten het betreffende CPB-macrobedrag uit te verkrijgen. Hierbij is gewogen naar de leeftijdsopbouw van de in Hoofdstuk 3 beschreven totale onderzoekspopulatie.

Vervolgens is de nettobijdrage (d.w.z. de contant gemaakte nettobijdrage over de levensloop) geïjkt. Hiervoor is als richtwaarde de nettobijdrage N over de levensloop van een in 2016 geboren Nederlandse ingezetene gehanteerd ($N = -€ 54.839$), uitgaande van de basisdataset van *Minder zorg om vergrijzing* en de in dit rapport veronderstelde discontovoet van 3% reëel en 5% nominaal en productiviteitsgroei van 1,5% reëel en 3,5% nominaal.

⁵⁶ De profielen gaan uit van een economische groei van 3,5%, dus de extrapolatie vanaf 2060 is in beginsel gedaan met deze groeivoet. Naderhand is dit vervolgens teruggerekend naar meer recente CPB-schattingen van de groeivoet en discontovoet, zie Hoofdstuk 8. De interpolatie is indien mogelijk exponentieel gedaan door schatting van de groeifactor en voor een beperkt aantal gevallen waarbij dat niet mogelijk was lineair.

Tabel 5.1 Inkomsten en uitgaven van de overheid in 2016. Operationalisering per post en deelpost en macrobedrag per (deel)post in miljarden euro's. Alleen de belangrijkste CBS-microdatabestanden zijn weergegeven. Bron indeling in posten en macrobedragen per deelpost: CPB.

Nr.	Post	Operationalisering	Bedrag
TOTAAL UITGAVEN			299,9
1	Openbaar bestuur	CBS-statline en maatwerk data over Veiligheidszorg, CPB-profiel	64,9
2	Defensie	CPB-profiel	6,9
3	Onderwijs	CBS-microdata voor 2015 en 2016 ONDERWIJSDEEL-NEMERSTAB, CBS-Statline, CBS-maatwerkberekeningen, rijksbegrotingen	31,4
4	AKW/WSF Kinderbijslag/Wet Studiefinanciering)	CBS-microdata INPATAB 2016	5,2
5	Arbeidsongeschiktheid/ZW (Ziektewet)	CBS-microdata INPATAB 2016	13,4
6	Werkloosheid	CBS-microdata INPATAB 2016	6,9
7	Bijstand/ANW (Algemene nabestaanden wet)	CBS-microdata INPATAB 2016	7,4
8	SZ-rest (Sociale zekerheid rest)	CBS-microdata INPATAB 2016, CPB-profiel	17,3
9	AOW	CBS-microdata INPATAB 2016	36,9
10	Overdrachten buitenland	CPB-profiel	10,5
11	Subsidies	CBS-microdata INPATAB 2016	9,4
12	Zorg	CBS-microdata ZFWZORGKOSTENTAB en Kinderopvang, Vektis-data, CPB-profiel	65,3
13	Bruto investeringen gebouwen	CPB-profiel	8,5
14	Bruto investeringen infrastructuur	CPB-profiel	10,1
15	Bruto investeringen scholen	CBS-microdata voor 2015 en 2016 ONDERWIJSDEEL-NEMERSTAB	5,9
TOTAAL INKOMSTEN			298,8
16	LIS (Loon- en inkomstenbelasting en sociale premies)	CBS-microdata INPATAB 2016	153,2
17	Overige directe belastingen huishoudens	CBS-microdata INPATAB 2016	7,9
18	Erfbelasting	CPB-profiel i.c.m. CBS-microdata INPATAB 2016 en CBS-Statline data	1,7
19	VPB/dividendbelasting (vennootschapsbelasting en dividendbelasting van Nederlandse bedrijven)	CPB-profiel i.c.m. CBS-microdata INPATAB 2016 en CBS-Statline data	21,8
20	IRN (Indirecte Resterende belastingen en Niet-belastingmiddelen)	CBS-microdata INPATAB 2016, INHATAB 2016 en CBS-Statline data naar rato van percentiel bruto inkomen	68,1
21	IRN van bedrijven	CPB-profiel i.c.m. CBS-microdata INPATAB 2016 en CBS-Statline data	16,1
22	Netto grondverkopen	CPB-profiel	2,3
23	Niet-belastingmiddelen-rest	CPB-profiel	27,8

De ijking is uitgevoerd met gebruikmaking van de stelpost die nodig is om het effect van de geleidelijke aanpassing van de AOW-leeftijd⁵⁷ op de sociale zekerheid in te stellen (zie ook §0). Dit is gedaan met een iteratief proces. Hiervoor is herhaaldelijk voor een AOW-leeftijd van 66, respectievelijk 71 jaar de nettobijdrage uitgerekend. De toename van de nettobijdrage Δ per jaar toename van de AOW-leeftijd wordt verkregen door het verschil in nettobijdrage een AOW-leeftijd van 66, respectievelijk 71 te delen door 5. Regressie voor verschillende AOW-leeftijden laat zien dat dit vrijwel lineair is voor de leeftijden 65-71. Op basis van Δ en N kan geschat worden welke ijkwaarde hoort bij de verlaging van de AOW-leeftijd met 1,75 jaar i.v.m. het pensioenakkoord.⁵⁸ Door te kalibreren op deze ijkwaarde verandert echter Δ zelf weer. Door echter steeds op basis van de nieuwe Δ een nieuwe ijkwaarde te schatten, ontstaat een steeds preciezere benadering. Dit proces is getermineerd bij een verschil kleiner dan € 0,49.

5.2 Veiligheidszorg

Bij de meeste uitgaven die vallen onder de post Openbaar bestuur (Tabel 5.1, post nr. 1) wordt verondersteld dat het profijt en de kosten gelijk verdeeld zijn over alle inwoners. Het betreft uiteenlopende zaken als uitvoerende en wetgevende organen, openbare orde en veiligheid, economische aangelegenheden en milieubescherming.⁵⁹ Hiervan is alleen naar migratieachtergrond gedifferentieerd voor de COFOG-functie openbare orde en veiligheid en wel voor een bedrag van 10 miljard euro, dat als volgt wordt beargumenteerd.

Het macrobedrag voor COFOG-functie openbare orde en veiligheid bedraagt circa 13 miljard euro. Daarvan heeft 1,7 miljard betrekking op de brandweer. De rest heeft betrekking op criminaliteit en dergelijke. Daarnaast wordt 1,2 miljard euro besteed aan onderzoek en andere aspecten van openbare orde veiligheid. De resterende COFOG-posten politie, rechtspraak en gevangeniswezen beslaan samen ongeveer 10 miljard euro.⁶⁰ Dit laatste bedrag kan genomen worden als conservatieve proxy voor de totale op criminaliteit betrekking hebbende overheidsuitgaven.

Ook op de volgende wijze kan men argumenteren dat 10 miljard een redelijke en ook conservatieve schatting is voor de totale naar migratieachtergrond te differentiëren veiligheidskosten. Uit de (in 2015 stopgezette) tabellen over de gemiddelde kosten voor veiligheidszorg over de jaren 2010-2015 blijkt dat gemiddeld ongeveer 7 miljard euro wordt besteed aan opsporing, vervolging, berechting, tenuitvoerlegging en ondersteuning van slachtoffers, verdachten en daders. Daarnaast wordt 6 miljard euro besteed aan preventie. Conservatief geschat heeft de helft (3 miljard euro) van dat geld betrekking op de preventie van strafbare feiten. Samen met de eerder genoemde 7 miljard is dit eveneens een bedrag van 10 miljard euro.⁶¹

Vervolgens is het bedrag van 10 miljard euro naar rato van het aantal geregistreerde verdachten per 10.000 inwoners verdeeld over personen. Hiervoor is CBS-statline data gebruikt voor het aantal

⁵⁷ Volgens het groeipad zoals dat gold tot het zogenaamde pensioenakkoord.

⁵⁸ Excelwerkblad 8.1L uit de dataset bij de CPB-vergrijzingsstudie *Zorgen om morgen*, 2019, opgehaald 10-2-2021 van: <https://www.cpb.nl/sites/default/files/omnidownload/CPB-Vergrijzingsstudie-2019-Data-figuren.xlsx>

⁵⁹ Feitelijk zijn dit alle COFOG-functies die niet al afgedekt zijn door de overige uitgavenposten in Tabel 5.1, <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/82902NED/table?dl=308CD>

⁶⁰ <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/82902NED/table?dl=308CD>

⁶¹ <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/80162NED/table?dl=31700>

verdachten naar leeftijd (vijf groepen) en migratieachtergrond.⁶² Deze data zijn beschikbaar voor 70 herkomstlanden (eerste en tweede generatie) en daarnaast voor personen met Nederlandse en (niet-)westerse migratieachtergrond en enkele andere categorieën. Van deze data is het gemiddelde genomen voor de jaren 2014-2018. In beginsel is dat gedaan voor de eerste en tweede generatie afzonderlijk. Voor een aantal combinaties⁶³ van herkomstland en leeftijdsgroep was er onvoldoende data. In die gevallen zijn de data genomen van beide generaties samen. In een beperkt aantal gevallen⁶⁴ was ook die data niet voorhanden en is het gemiddelde afgeleid uit het aantal verdachten voor alle leeftijden van desbetreffende groep in combinatie met de verhoudingen in het aantal verdachten tussen de verschillende leeftijdsgroepen in de bevolking als geheel.

Voorgaande exercitie levert het aantal verdachten per leeftijdsgroep, uitgesplitst naar migratieachtergrond. Er zijn echter aanzienlijke groepsverschillen als het gaat om het aandeel verdachten dat de verschillende stadia van de strafrechtketen doorloopt. Het gaat dan bijvoorbeeld om verschillen in het percentage verdachten dat gedetineerd wordt. Om deze verschillen te verdisconteren is in navolging van CBS-maatwerk tabellen⁶⁵ berekend hoe vaak er bij een geregistreerde verdachte sprake is van vervolging, dagvaarding, schuldigverklaring met straf en tenuitvoerlegging met gevangenisstraf. Zo ontstaan vijf gewichten voor de vijf stadia van de rechtsketen, waarbij het stadium 'geregistreerde verdachte' altijd gewicht 1 heeft en de overige genoemde stadia het gewicht dat overeenkomt met de oververtegenwoordiging ten opzichte van autochtonen. De genoemde CBS-maatwerktafel bevat geen gegevens voor westerse migranten; deze zijn geschat op basis van CBS-statline data.

Vervolgens zijn de zodoende per groep berekende relatieve gewichten voor de vijf stadia van de rechtsketen gewogen tegen het aandeel in de kosten van elk stadium. Hiervoor zijn de gegevens gebruikt over de gemiddelde kosten voor veiligheidszorg over de jaren 2010-2015 en de COFOG-gegevens voor 2016.^{66 67} Uitgaande dat de helft van het totaalbedrag voor preventie toegerekend wordt naar personen, komt dit neer op 3 miljard euro, oftewel 30% van de totale 10 miljard euro veronderstelde kosten. Deze worden toegerekend aan alle geregistreerde verdachten en krijgen gewicht 1 van het stadium 'geregistreerde verdachte'.⁶⁸ Verder wordt 23% besteed aan detentie en dit krijgt het gewicht van het stadium 'tenuitvoerlegging met gevangenisstraf'. Vervolging beslaat ongeveer 7% van de kosten en krijgt het gewicht van het stadium 'vervolging'. Opsporing beslaat 22% van de kosten en krijgt het

⁶² CBS-statline, *Verdachten; geslacht, leeftijd, migratieachtergrond en generatie*, opgehaald 25-12-2020 van: <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/81959NED/table?dl=486C3>

⁶³ Het gaat in ruim twee derde van de gevallen om tieners (12 tot 18 jaar, overwegend eerste generatie) en ouderen (65 jaar of ouder, overwegend tweede generatie), vaak van herkomstgroepen met een relatief kleine aanwezigheid in Nederland. Om twee redenen beperkt dit de vertekening. Bij het aggregeren over alle landen naar (tenminste) de CBS 42-deling tellen migratieachtergronden met relatief kleine groepen ook weinig mee in het totaal per herkomstregio. Verder tellen nettobijdragedata van tieners en 65-plussers bij het bepalen van de nettobijdrage over de levensloop veel minder mee dan de tussenliggende leeftijden, zoals elders in deze appendix uiteengezet is.

⁶⁴ Tweemaal bij de leeftijdscategorie 12 tot 18 jaar en 27 maal in de leeftijdscategorie 65 jaar of ouder.

⁶⁵ CBS-maatwerk, *Migratieachtergrond van personen in de strafrechtketen*, opgehaald 12-02-2021 van: <https://www.cbs.nl/nl-nl/maatwerk/2019/10/migratieachtergrond-van-personen-in-de-strafrechtketen>

⁶⁶ CBS-statline, *Overheidsuitgaven; naar functies 1995-2016*, opgehaald 12-02-2021 van: <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/82902NED/table?dl=308CD>

⁶⁷ CBS-statline, *Veiligheidszorg; kerncijfers 2002-2015*, opgehaald 12-02-2021 van: <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/80162NED/table?dl=31700>

⁶⁸ Bij de geregistreerde verdachten zijn er dus geen kostenverschillen tussen groepen, bij de overige vier stadia in de rechtsketen wel.

gewicht van het stadium 'dagvaarding'. De kosten voor berechting en ondersteuning van slachtoffers, verdachten en daders beslaan de resterende 11% en krijgen het gewicht van het stadium 'schuldigverklaring met straf'. Alhoewel de kostenposten wellicht niet perfect aansluiten bij de stadia, worden zo kostenverschillen tussen de verschillende typen verdachten in grote lijnen verdisconteerd.

Tot slot worden de aantallen verdachten per migratieachtergrond vermenigvuldigd met de gewichten die in de voorgaande alinea's zijn besproken. Op basis van de zo verkregen gegevens is daarna geagegreerd naar eerste en tweede generatie en herkomstregio voor de CBS 42-deling, de 19-deling, de 12-deling, de 2-deling en naar de totalen voor Nederlandse achtergrond en eerste en tweede generatie migratieachtergrond. Vervolgens zijn op basis hiervan leeftijdsprofielen gemaakt, welke op basis van de gegevens op basis van het leeftijdsprofiel voor de bevolking als geheel zijn geïjkt op het eerder geschetste macrobedrag van 10 miljard euro. Voor opleiding en combinaties van opleiding met de indeling westers/niet-westers is aanvullend gegevens over verdachten naar opleidingsniveau als proxy gebruikt.⁶⁹

5.3 Onderwijs en Bruto investeringen scholen

In deze paragraaf wordt de operationalisering van de posten Onderwijs en Bruto investeringen scholen toegelicht (Tabel 5.1, post nr. 3 en 15). De meeste uitleg wordt gewijd aan de post Onderwijs. Vooraf worden twee andere daarmee samenhangende posten besproken.

De eerste post wordt gevormd door de bruto-investeringen in scholen. Deze zijn naar rato toegewezen aan onderwijsdeelnemers: bij onderwijsdeelname gedurende 2015 én 2016 met gewicht één, bij deelname in alleen 2015 met gewicht twee derde en bij deelname in alleen 2016 met gewicht één derde. Hierbij is deelname geoperationaliseerd als inschrijving op de peildatum 1 oktober van het verslagjaar van het CBS-microdata bestand ONDERWIJSDEELNEMERSTAB. De resulterende variabele is daarna gewogen naar de leeftijdsopbouw van de in §0 beschreven totale onderzoekspopulatie en vervolgens is het resulterende totaalbedrag geïjkt om het CPB-macrobedrag voor de post Bruto investeringen scholen uit Tabel 5.1 te verkrijgen.

De tweede post wordt gevormd door het niet bekostigen van het onderwijs voor niet-EER studenten. Hiervoor is een lumpsumbedrag van een geschatte gemiddelde studieduur van 2,9 jaar à € 6.700 in mindering gebracht op de totale kosten voor de betreffende studenten.⁷⁰ Met hogere of lagere collegegelden is geen rekening gehouden, want dat loopt niet via het fiscale kanaal. Er is ook geen rekening gehouden met de mate van deelname aan kostbare studierichtingen zoals technische studies en medicijnen. Dit is een eenmalige bate, die na het berekenen van de nettobijdrage in mindering wordt gebracht.

De rest van deze paragraaf handelt over de operationalisering van de post Onderwijs. De kosten voor het onderwijs zijn gebaseerd op eigen berekening van de kosten per onderwijssoort in combinatie met de gegevens voor deelname aan de verschillende onderwijssoorten op basis van het CBS-microdatabestand ONDERWIJSDEELNEMERSTAB voor de verslagjaren 2015 en 2016. In ONDERWIJSDEELNEMERSTAB zijn alle deelnemers aan het bekostigde onderwijs opgenomen. Via de hulpbestanden

⁶⁹ CBS-statline, *Verdachte jongeren; geslacht, herkomst, opleiding en recidive, 2006-2014*, opgehaald 10-2-2021 van: <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/81978NED/table?dl=3C860>

⁷⁰ Vergelijk CPB-notitie Economische effecten van internationalisering in het hoger onderwijs en mbo, 2019. Figuur 3.3, aantal niet-EER studenten voor het jaar 2016 en de gemiddelde studieduur zoals gegeven in Tabel 5.2. Het bedrag van € 6.700 is ontleend aan de Rijksbegroting 2016.

OPLEIDINGSNRREFV20 en CTOREFV8 zijn aan de ONDERWIJSDEELNEMERSTAB bestanden alle variabelen uit OPLEIDINGSNRREFV20 en CTOREFV8 toegevoegd. Vervolgens zijn dummyvariabelen toegevoegd om tellingen te kunnen maken voor de onderwijssoorten PO (primair onderwijs), VO (voortgezet onderwijs), MBO, HBO, WO Bachelor en WO Master. Hierin is MBO1 tot MBO gerekend. Verder zijn er dummyvariabelen gemaakt voor het al dan niet volgen van LWOO, Praktijkonderwijs en het speciaal (basis)onderwijs. Daarna zijn er per onderwijssoort kosten toegerekend aan leerlingen. Uitgangspunt bij de berekeningen vormen de jaarverslagen bij de begroting van 2016, [maatwerktabellen](#) van de kosten per diploma van het CBS en CBS-statline tabellen over hetzelfde onderwerp.⁷¹ Hieronder wordt steeds de berekening voor 2016 toegelicht. De berekening voor 2015 is vergelijkbaar.⁷²

Om meetfouten zoveel mogelijk uit te middelen is de berekening drie keer uitgevoerd. Ten eerste zijn de bedragen voor 2015 en voor 2016 afgeleid uit voor deze jaren genoemde bedragen in [het jaarverslag bij de rijksbegroting voor 2016](#). Ten tweede zijn de bedragen voor zover mogelijk berekend op basis van [een maatwerkberkening van het CBS voor de kosten per diploma voor de standaardroute voor 2015](#). Vervolgens zijn de beide bedragen voor 2015 gecombineerd met de gegevens over de onderwijsdeelnemers in 2015 en is het bedrag voor 2016 gecombineerd met de onderwijsdeelnemers in 2016. Tenslotte zijn deze drie bedragen gemiddeld waarbij elk voor één derde meetelde.

Hieronder worden de operationalisering van achtereenvolgens het primair onderwijs, het voortgezet onderwijs en het MBO, HBO en WO.

Onder het primair onderwijs valt het basisonderwijs en het speciaal (basis)onderwijs. In het jaarverslag bij de begroting wordt dit verder niet uitgesplitst, maar omdat de verschillen in kosten aanzienlijk zijn is dat in het huidige rapport wel gedaan. Voor de kosten van het speciaal (basis)onderwijs is voor 2016 uitgegaan van het op CBS-statline gegeven [macrobedrag voor 2016](#) voor overheidsuitgaven van € 1.297 miljoen en [het totale deelnemersaantal voor 2016/2017](#) van 100,9 duizend leerlingen. Dit geeft een bedrag van € 12.854 per leerling.

Voor de kosten van het reguliere primaire onderwijs is uitgegaan van de aantallen leerlingen genoemd in het jaarverslag bij de begroting voor 2016. Belangrijk hierbij is de zogenaamde gewichtenregeling; scholen kunnen volgens een bepaalde systematiek extra geld krijgen voor leerlingen van laagopgeleide ouders. In totaal hebben de basisscholen in het jaar 2016/2017 voor 45.950 virtuele leerlingen extra bekostiging ontvangen volgens deze gewichtenregeling.

Voor de kosten van het regulier basisonderwijs is uitgegaan van het op CBS-statline gegeven [macrobedrag voor 2016](#) voor overheidsuitgaven van € 9.402 miljoen. Dit bedrag is gedeeld door [het totale deelnemersaantal voor 2016/2017](#) van 1.427,3 duizend leerlingen, vermeerderd met 45.950 virtuele leerlingen waarvoor scholen in 2016 extra bekostiging ontvingen in verband met voornoemde gewichtenregeling. Dit geeft een bedrag van € 6.382 per leerling voor het reguliere basisonderwijs. Dit bedrag

⁷¹ Vanwege budgettaire beperkingen en/of het niet aanwezig zijn van de data was het niet altijd mogelijk om zelf de aantallen uit CBS-microdatatabestanden te halen omdat niet over alle bestandsonderwerpen kon worden beschikt. Soms zijn op veelvoud van € 100 afgeronde bedragen overgenomen uit bepaalde bronnen, soms zijn bedragen zelf berekend en zijn op euro's afgeronde bedragen gebruikt. Zonder schijnprecisie te willen nastreven zijn bij berekende bedragen de op euro's afgeronde bedragen gebruikt.

⁷² De in deze paragraaf gerapporteerde bedragen voor 2015 wijken enigszins af. Mede hierom wijken de in deze paragraaf gerapporteerde bedragen af van de gemiddelde in de berekening gebruikte bedragen zoals gegeven in Tabel 8.2 van het rapport. Voor het overige zijn kleine afwijkingen veroorzaakt doordat de onderwijsdeelnemers op twee peilmomenten zijn gepeild (1 oktober 2015, respectievelijk 2016), zie de volgende alinea.

is per groep afzonderlijk verhoogd volgens de systematiek van de gewichtenregeling. De naar personen toegerekende bedragen inzake de gewichtenregeling worden aan het eind van de paragraaf toegelicht, na de bespreking per afzonderlijk schoolniveau.

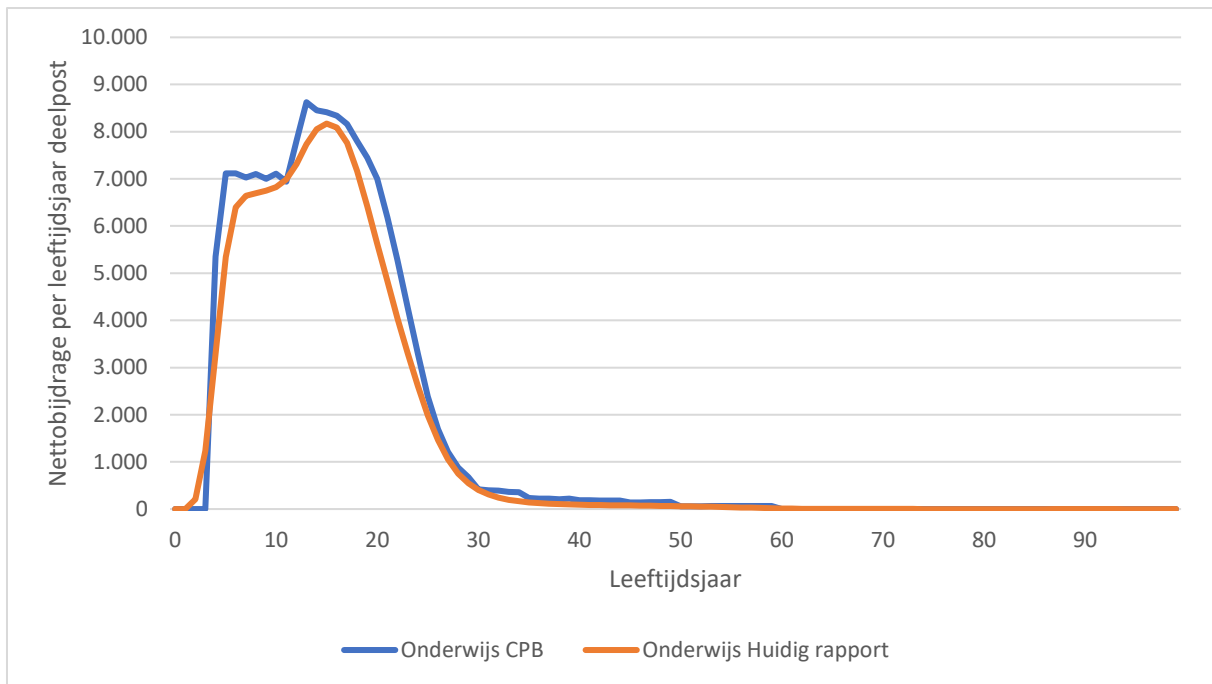
De kosten voor het voortgezet onderwijs zijn in beginsel afgeleid uit het [jaarverslag bij de begroting voor 2016](#). In Tabel 3.2 in het jaarverslag zijn leerlingaantallen te vinden en ook de gemiddelde kosten die per leerling op € 8.194 zijn gesteld. Uiteraard zijn er allerlei verschillen tussen onderwijssoorten die ook kostenverschillen met zich meebrengen. Deze verschillen zijn in navolging van [gangbare berekeningen van o.a. het CBS](#) genegeerd. De kosten voor LWOO en Praktijkschool liggen echter beduidend hoger dan de gemiddelde kosten voor de overige onderwijsvormen in het VO, met name door hogere personeelskosten. Een aanvullende bekostiging is voor het jaar 2016 bij [ministeriële regeling](#) vastgesteld op € 4.111. Uitgaande van de bedragen en aantallen die in het jaarverslag bij de begroting gegeven zijn komen hiermee de kosten voor reguliere VO leerlingen op € 7.699 en voor LWOO en Praktijkschoolleerlingen op € 11.810.⁷³

De kosten voor het MBO zijn afgeleid uit [het jaarverslag bij de rijksbegroting voor 2016](#) en vastgesteld op € 8.000 per leerling. De kosten voor het HBO zijn eveneens afgeleid uit [het jaarverslag bij de rijksbegroting voor 2016](#) en vastgesteld op € 6.800 per leerling. Ter controle zijn voor het HBO de kosten eveneens berekend door de kosten uit [een maatwerkberekening van het CBS voor de kosten per diploma voor de standaardroute](#) te herleiden tot kosten per jaar. Dit leidt tot € 6.788 per leerling voor het jaar 2015, vrij dicht in de buurt van het in het jaarverslag genoemde bedrag van € 6.800 voor zowel 2015 als 2016.

Voor het MBO is een dergelijke berekening helaas niet mogelijk omdat volgens het CBS “geen betrouwbare theoretische studieduur” is vast te stellen – [zie toelichting bij deze tabel](#) – waardoor terugrekenen naar studiekosten per jaar ook niet mogelijk is. Echter, als er vanuit gegaan wordt dat de gemiddelde vertraging gelijk is aan de vertraging bij HBO-studenten, dan komt een herleiding voor 2015 op een bedrag van € 7.957 per student per jaar, dicht in de buurt van de € 8.000 die voor 2016 en de € 8.100 die voor 2015 genoemd wordt in [het jaarverslag bij de rijksbegroting voor 2016](#).

Voor het WO wordt in de [het jaarverslag bij de rijksbegroting voor 2016](#) niet gedifferentieerd naar de Bachelor en de Master fase; er wordt één bedrag genoemd van € 6.700 per student voor het WO in het algemeen. Berekening van de gemiddelde masterduur op basis van CBS-microdata gaf aanleiding om de kosten in de bachelor fase een paar honderd euro hoger in te schatten dan in de masterfase.

⁷³ Een alternatieve berekening is gegeven in [een studie door Panteia](#) naar onderwijsbekostiging door OCW en EZ. In deze studie worden de vermenigvuldigingsfactoren 1,67 en 1,58 genoemd voor respectievelijk OCW en EZ bekostigd onderwijs (zie voetnoot 6). Deze factoren lijken echter voor het jaar 2016 te hoog te zijn en daarnaast is de berekening in Tabel 2.5 gebaseerd op de verkeerde (te hoge) bedragen, namelijk de bedragen voor directiesalarissen i.p.v. salarissen onderwijzend personeel (vergelijk p. 53 van het [rapport bekostiging voortgezet onderwijs](#) van de Rekenkamer of de betreffende ministeriële regelingen). Berekening op basis van deze factoren levert vergelijkbare bedragen op met iets grotere verschillen tussen regulier VO enerzijds en LWOO en Praktijkonderwijs anderzijds.



Figuur 5.1 Onderwijs: vergelijking met het CPB-profiel.

Bij de post Onderwijs bedraagt het verschil tussen het totaalbedrag op basis van CBS-microdata en het CPB-macrobedrag voor 2016 uit Tabel 5.1 ongeveer 3 miljard euro. Voor vrijwel alle leeftijden ligt het CPB-profiel boven het profiel op basis van CBS-microdata, zie Figuur 5.1. Gezien de hogere kosten voor de basisschoolleeftijd rekent het CPB de kosten van de leerlinggewichten kennelijk voor 100% toe (dus op basis van de maximale extra vergoeding volgens het leerlinggewicht) en niet zoals in huidige studie voor de empirisch waargenomen 50%. Het CPB gaat verder in haar schattingen gezien de studiekosten voor met name twintigers waarschijnlijk uit van een hogere toerekening van studiekosten naar WO-studenten, waarbij kennelijk zowel onderwijs- als onderzoekskosten aan studenten worden toegerekend. De onderzoekskosten vormen een aanzienlijk bedrag in de totale kosten van de academies. In de huidige studie wordt wetenschappelijk onderzoek als een publiek goed behandeld dat gelijkelijk aan de bevolking wordt toegerekend. Dit gebeurt door voor de hele bevolking het verschil tussen het leeftijdsprofiel uit huidige studie en het CPB-profiel toe te rekenen naar een voor alle inwoners gelijke restpost, die wordt opgeteld bij de post Openbaar bestuur. (NB: vergelijk Tabel 5.3 in [deze studie](#) of met de post 'fundamenteel onderzoek' die in de COFOG-indeling onder de functie Algemeen overheidsbestuur wordt geschaard.⁷⁴

Tot slot een toelichting op de zogenaamde gewichtenregeling. Met de gewichtenregeling wordt aan basisscholen extra bekostiging toegekend voor kinderen van laagopgeleide ouders. Deze zogenaamde leerlinggewichten zijn geoperationaliseerd op basis van de CBS-microbestanden CITOTAB voor de verslagjaren 2010 t/m 2014 waarin de in 2016 nog steeds geldende leerlinggewichten (0,3 en 1,2) zijn opgenomen. In beginsel krijgt een school voor een leerling met leerlinggewicht 0,3 (30% extra bekostiging) een vergoeding die 1,3 maal hoger is dan het standaardbedrag per leerling en voor een leerling met leerlinggewicht 1,2 (120% extra bekostiging) een vergoeding die 2,2 maal hoger is. De in CITOTAB gegeven leerlinggewichten zijn geaggregeerd naar de in deze studie gebruikte groepen, met uitsplitsing naar generatie, motief en herkomstregio (maximaal toegepaste verfijning volgens de 42-deling).

⁷⁴ <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/82902NED/table?dl=318DD>

De zo verkregen gemiddelden zijn gebruikt om de bedragen per leerling voor de reguliere basisschoolkosten voor het jaar 2016 op te hogen, zoals de regeling bedoeld is. Als in het vervolg van deze paragraaf over leerlinggewichten wordt gesproken gaat het over gemiddelde groepsgewichten gebaseerd op deze CBS-microdata.

Omdat de gewichtenregeling voor het jaar 2016 niet is bijgehouden in de voor dit onderzoek gebruikte CBS-microdatabestanden zijn de cijfers voor leerlingaantallen overgenomen uit datasets van DUO. In totaal werden scholen bekostigd voor 45.950 extra virtuele leerlingen op basis van de gewichtenregeling. Voor de daadwerkelijke toekenning van het aantal virtuele leerlingen A dat aan een school wordt toegekend wordt de volgende formule gebruikt:

$$A = \lfloor 0,3 \times N^{G0.3} + 1,2 \times N^{G1.2} - 0,06 \times N \rfloor$$

Hierin is $N^{G0.3}$ het aantal leerlingen met gewicht 0,3, $N^{G1.2}$ het aantal leerlingen met gewicht 1,2 en N het totaal aantal leerlingen in de betreffende school en symboliseren de haken een afronding naar beneden.

Merk op dat M , het gewogen gemiddelde van het leerlinggewicht voor een school gelijk is aan

$$M = \frac{0,3 \times N^{G0.3} + 1,2 \times N^{G1.2}}{N}$$

en dat daarmee de formule voor A herschreven kan worden tot:

$$A = \lfloor MN - 0,06 \times N \rfloor = N \times \lfloor M - 0,06 \rfloor = MN \times \left\lfloor 1 - \frac{0,06}{M} \right\rfloor$$

Omdat het aantal virtuele extra leerlingen op de school *voor de toepassing van de toekenningsformule* gelijk is aan MN kan (bij het negeren van de afronding) de toekenningsfractie T_M geschreven worden als:

$$T_M = \begin{cases} 0 & \text{als } M < 0,06 \\ 1 - \frac{0,06}{M} & \text{als } M \geq 0,06 \end{cases}$$

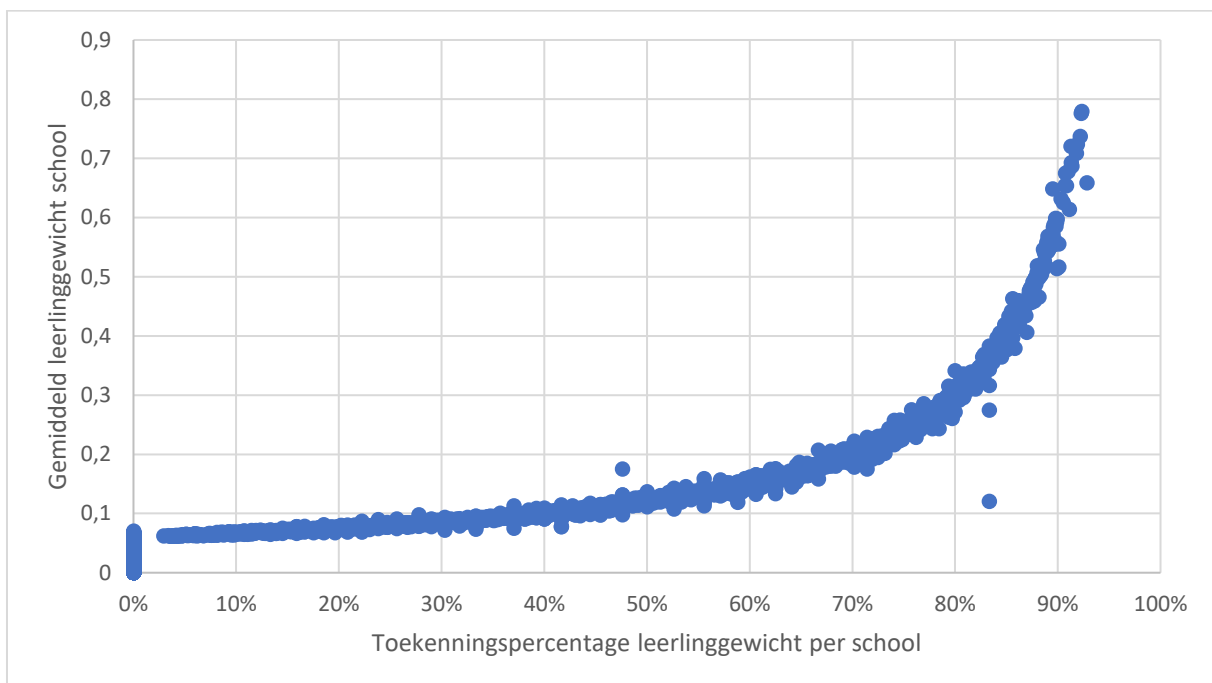
Deze formule is zodanig dat in de praktijk gemiddeld slechts een deel (voor 2016: 50,1%, berekend op basis van data van DUO) van het leerlinggewicht wordt meegeteld bij het bepalen van de extra bekostiging van basisscholen. Voor scholen met een lage concentratie kinderen met laagopgeleide ouders is de toekenningsgraad lager dan voor scholen met een hoge concentratie kinderen met laagopgeleide ouders.⁷⁵ De maximaal waargenomen toekenningsgraad was in 2016 92,8%.⁷⁶

Deze waarnemingen leggen een operationaliseringsprobleem bloot. Het volledig meerekenen van het op basis van CBS-microdata berekende leerlinggewicht geeft zeker een overschatting. Echter, berekening op basis van de gemiddelde (berekend over de hele populatie) daadwerkelijk toekenning in 2016

⁷⁵ Deze gewichtenregeling zou (ten tijde van het uitwerken van deze operationalisering) herzien gaan worden en vervangen door een systematiek gebaseerd op CBS modellen waarbij de gemiddelde bedragen naar verwachting voor de grotere gemeenten lager zullen zijn.

⁷⁶ Ter indicatie: berekening leert dat het toekenningspercentage mede afhangt van de verhouding tussen de leerlinggewichten 0,3 en 1,2 en het maximum ligt op 95% bij uitsluitend leerlinggewicht 1,2 en 92% als de ene helft van de leerlingen gewicht 0,3 heeft en de andere helft leerlinggewicht 1,2.

geeft vrijwel zeker een forse onderschatting voor niet-westerse immigranten. De reden voor deze laatste veronderstelling is dat de procentuele daadwerkelijke toekenning vooral hoog is voor scholen met een hoge concentratie kinderen met laagopgeleide ouders (zie Figuur 5.2, maar dit volgt ook uit bovenstaande formule). Ter illustratie: als het gemiddeld leerlinggewicht van een school gelijk is aan het gemiddelde leerlinggewicht van autochtonen (0,03), dan wordt 0% van het extra geld daadwerkelijk toegekend. Is het gemiddeld leerlinggewicht van een school gelijk aan het Nederlands gemiddelde (0,09) dan wordt 33% daadwerkelijk toegekend. Is het gemiddelde leerlinggewicht op een school gelijk aan het gemiddelde van alle eerste en tweede generatie immigranten (0,30) respectievelijk niet-westerse immigranten (0,41) dan wordt 80% respectievelijk 85% van het extra geld daadwerkelijk toegekend.



Figuur 5.2 Toekenning extra financiering als percentage van het totale leerlinggewicht versus gemiddeld leerlinggewicht per school. Berekening op basis van data van DUO voor 6494 scholen.

Door het gemiddeld lage opleidingsniveau van niet-westerse immigranten in combinatie met het bestaan van zogenaamde ‘zwarte scholen’ – scholen met een hoge concentratie voornamelijk niet-westerse immigranten – ligt het in de lijn der verwachting dat onder niet-westerse immigranten de toekenning relatief hoog zal zijn. Het bestaan van zwarte scholen wordt deels veroorzaakt door zelfselectie: autochtone ouders vermijden het om in wijken te wonen met hoge concentraties niet-westerse immigranten en/of vermijden het om hun kinderen op zwarte scholen te plaatsen. Dit, in combinatie met het gemiddeld hogere opleidingsniveau van autochtone ouders, zal er naar verwachting toe leiden dat kinderen van laagopgeleide autochtone ouders minder vaak op een school zullen zitten met een hoog gemiddeld leerlinggewicht, waardoor het toekenningspercentage voor hen lager zal zijn dan voor kinderen van niet-westerse immigranten met hetzelfde opleidingsniveau als de voornoemde autochtone ouders. De scheve verdeling over scholen van de extra bekostiging is ook te zien in Tabel 5.2 met de voor 2016 waargenomen toekenningspercentages.

Tabel 5.2 Verdeling van het toekenningspercentage van het leerlinggewicht, 2016. Eigen berekening op basis van data van DUO.

Toekenningspercentage	Aantal scholen (abs.)	Aantal scholen (%)
0*	475	7%
0	4079	63%
1	0	0%
2	0	0%
3-5	25	0%
6-10	75	1%
11-20	177	3%
21-30	197	3%
31-40	214	3%
41-50	218	3%
51-60	219	3%
61-70	245	4%
71-80	301	5%
81-90	253	4%
91 of meer	16	0%
Totaal	6494	100%

*Geen kinderen met leerlinggewicht groter dan 0.

Bij gebrek aan data over de exacte distributie van leerlingen met en zonder migratieachtergrond over scholen is de volgende heuristiek toegepast. Stel dat de helft van de leerlingen op een school zit die perfect gemengd is en de andere helft op een school met alleen de eigen migratieachtergrond. Op een perfect gemengde school is het leerlinggewicht 0,09 (het landelijk gemiddelde) en de toekenningsfractie $\frac{1}{3}$. De toekenningsfractie voor een school met gemiddeld leerlinggewicht M wordt dan:

$$T_M = \begin{cases} \frac{1}{6} & \text{als } M < 0,06 \\ \frac{2}{3} - \frac{0,03}{M} & \text{als } M \geq 0,06 \end{cases}$$

Als meer in het algemeen $0 < C < 1$ het aandeel leerlingen is dat op een perfect gemengde school zit en de $1 - C$ resterende leerlingen op een perfect homogene school waar iedereen hetzelfde leerlinggewicht heeft, dan wordt de toekenning voor een leerling met leerlinggewicht M :

$$T_{CM} = \begin{cases} C \times \frac{1}{3} + (1 - C) \times 0 = \frac{C}{3} & \text{als } M < 0,06 \\ C \times \frac{1}{3} + (1 - C) \times \left(1 - \frac{0,06}{M}\right) = \frac{3-2C}{3} + \frac{0,06(C-1)}{M} & \text{als } M \geq 0,06 \end{cases}$$

Het is echter wenselijk dat de formule voor de toekenning van het leerlinggewicht zodanig is dat de toekenning op basis van het Nederlandse gemiddelde (0,09) gelijk is aan de som van de toekenning van alle deelgroepen bij een uitsplitsing naar migratieachtergrond. Voor de uitsplitsing naar de 42-deligen en eerste en tweede generatie is dit voor geen enkele C het geval bij bovenstaande formule. Daarnaast zou het macrobedrag moeten corresponderen. Gaan we uit van het macrobedrag bij 50,1% de facto toekenning zoals voor 2016 geldt, dan levert dit uitgaande van een leerlinggewicht van 0,09 voor de gemiddelde bevolking een macrobedrag op van circa 475 miljoen euro. Gaan we echter uit van

de 45.950 in 2016 daadwerkelijk toegekende virtuele leerlinggewicht leerlingen, dan komt het macrobedrag op 293 miljoen euro. Als we beide bedragen middelen, dan komt dit op 385 miljoen euro. Als we aannemen dat succesvolle migrantengroepen vaak op scholen met veel autochtonen zitten, dan kunnen we de toekenning voor autochtonen als minimum nemen. Als we verder – geïnspireerd door bovenstaande verkenning – aannemen dat de toekenningsformule voor een groep met gemiddeld leerlinggewicht M van de vorm $T_M = a - \frac{b}{M}$ is met $0 \leq a, b \leq 1$ (en met als minimum waarde de waarde voor autochtonen) en dat de toekenningsformule voor de hele bevolking en voor de bevolking uitgesplitst naar migratieachtergrond volgens de 42-delingsformule een macrobedrag van circa 385 miljoen euro moet opleveren, dan voldoet de volgende formule:

$$T_M = \text{Max} \left(\frac{217}{1000}, \frac{508}{1000} - \frac{0,01}{M} \right)$$

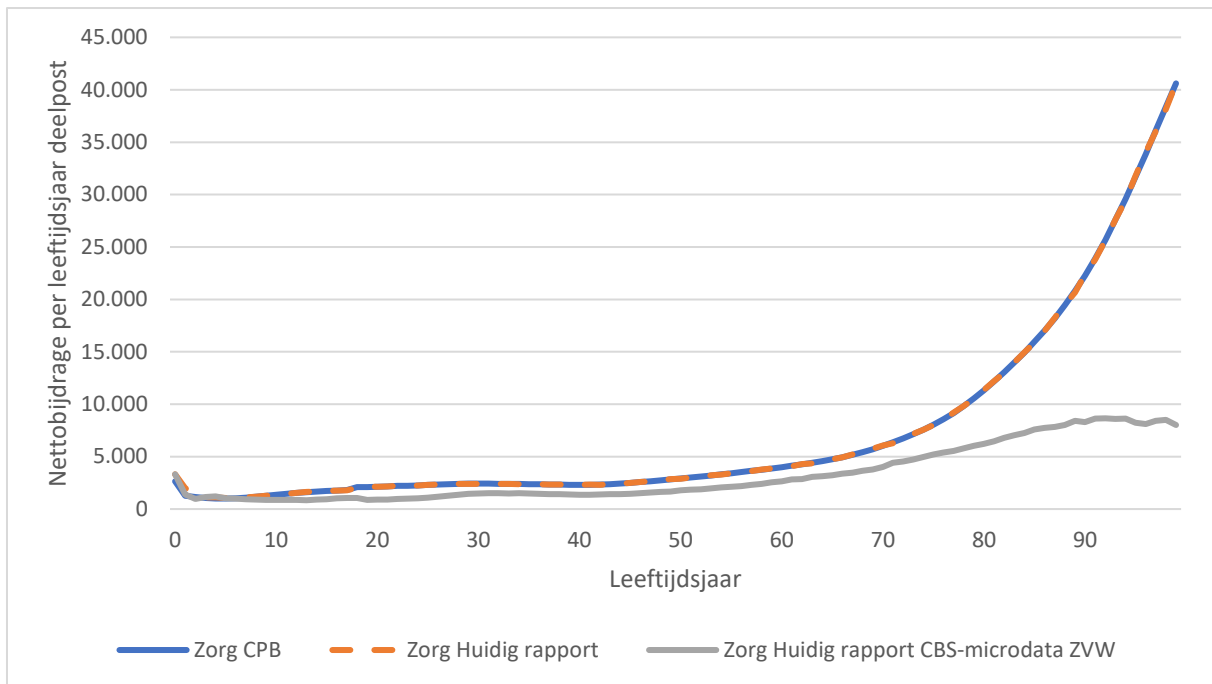
Het macrobedrag voor de hele bevolking is met deze formule 381 miljoen en voor de bevolking uitgesplitst naar migratieachtergrond volgens de 42-delingsformule 388 miljoen. Met deze veronderstelling is het toekenningspercentage bij autochtonen 21,7% (dit is ook direct het minimum voor alle groepen) en is het hoogste toekenningspercentage 49,7%. Het toekenningspercentage voor de bevolking als geheel is 40,0%. Door deze fractie te vermenigvuldigen met het waargenomen leerlinggewicht en het bedrag per basisschoolleerling à € 6.382 per jaar, wordt het daadwerkelijk aan scholen toegekende extra onderwijsgeld geschat. Kinderopvang

In deze paragraaf wordt de operationalisering van de Kinderopvang toegelicht. De bedragen voor kinderopvang zijn afgeleid uit de CBS-microdatabestanden Kinderopvang en AanvragerKinderopvang voor het verslagjaar 2016, door deze te combineren. Eerst is gecontroleerd of zonder verlies van gegevens aan iedere aanvrager met een partner één unieke partner kon worden toegekend en dat bleek het geval. Vervolgens is op basis van het microdatabestand INPATAB het persoonlijk bruto inkomen toegekend van aanvrager en de eventuele partner en is de toegekende toeslag naar rato toegerekend aan aanvrager en eventuele partner. Daarna is gecontroleerd of de macrobedragen klopten en dat bleek het geval te zijn. De kosten voor kinderopvang zijn bij de post overige sociale zekerheid geteld, omdat het onder Sociale Zaken en Werkgelegenheid op de begroting van 2016 stond.⁷⁷

5.4 Zorg

In deze paragraaf wordt de operationalisering van de post Zorg (Tabel 5.1, post nr. 12) toegelicht. De zorgkosten zijn berekend voor zover ze betrekking hebben op de Zorgverzekeringswet (ZVW) en zijn berekend op basis van het CBS-microbestand ZORGKOSTENTAB. De bedragen zijn daarbij opgehoogd met een ophoogfactor die door het CBS bij dit bestand wordt geleverd en die dient ter correctie van de data. Deze ophoogfactor wordt gebruikt in CBS-statline tabellen. De bedragen die voor het huidige rapport zijn berekend op basis van het gebruikte microdatabestand komen exact overeen met de door CBS gepubliceerde [Statline gegevens](#).

⁷⁷ Rijksbegroting, opgehaald 21-02-2021 van: <https://www.rijksbegroting.nl/2016/voorbereiding/begroting?hoofdstuk=40.22>



Figuur 5.3 Zorg: vergelijking met het CPB-profiel.

Vervolgens is het bedrag voor het eigen risico geschat op basis van een tabel van Vektis.⁷⁸ Op basis van deze tabel is het eigen risico naar leeftijd berekend voor 18 t/m 90 jarigen. Vanaf 90 jaar is het eigen risico constant gehouden. Deze data zijn beschikbaar voor 2015 en op basis van de bekende verschillen tussen 2015 en 2016 is een berekening gemaakt voor 2016. De verschillen tussen 2015 en 2016 zijn overigens miniem (per leeftijd maximaal 3 euro hogere of lagere eigen bijdrage). Het eigen risico is vervolgens toegepast op alle zorg, m.u.v. huisartsenzorg (inclusief zogenaamde multidisciplinaire zorg), geboortezorg en wijkverpleging en m.u.v. zorg voor jongeren tot 18 jaar. Voor jongeren die in 2016 18 werden is het eigen risico toegepast naar rato van de tijd voor en na het midden van de maand waarin men 18 werd.

Het op deze wijze berekende profiel ligt voor vrijwel alle leeftijden onder het CPB-profiel (zie de grijze lijn in Figuur 5.3). Dit komt omdat een groot deel van de zorgkosten niet onder de ZVW vallen. De reden is dat wegens beperkingen van data en databudget alleen de kosten voor de verplichte zorgverzekering zijn meegenomen. Dit is als volgt opgelost. Eerst is het verschilprofiel tussen het CBS-microdata leeftijdsprofiel en het CPB-profiel voor Zorg berekend (dit is dus het verschil tussen het blauwe profiel en het grijze profiel in Figuur 5.3). Dit profiel bevatte enkele relatief kleine negatieve bedragen. Deze bedragen zijn op nul gesteld en vervolgens is het profiel zodanig herijkt dat na optelling bij het CBS-microdata leeftijdsprofiel en weging tegen de onderzoekspopulatie het totaalbedrag gelijk is aan het CPB-macrobedrag voor 2016. Te zien is dat het na optelling resulterende profiel (de oranje lijn) vrijwel samenvalt met het CPB-profiel (de blauwe lijn).

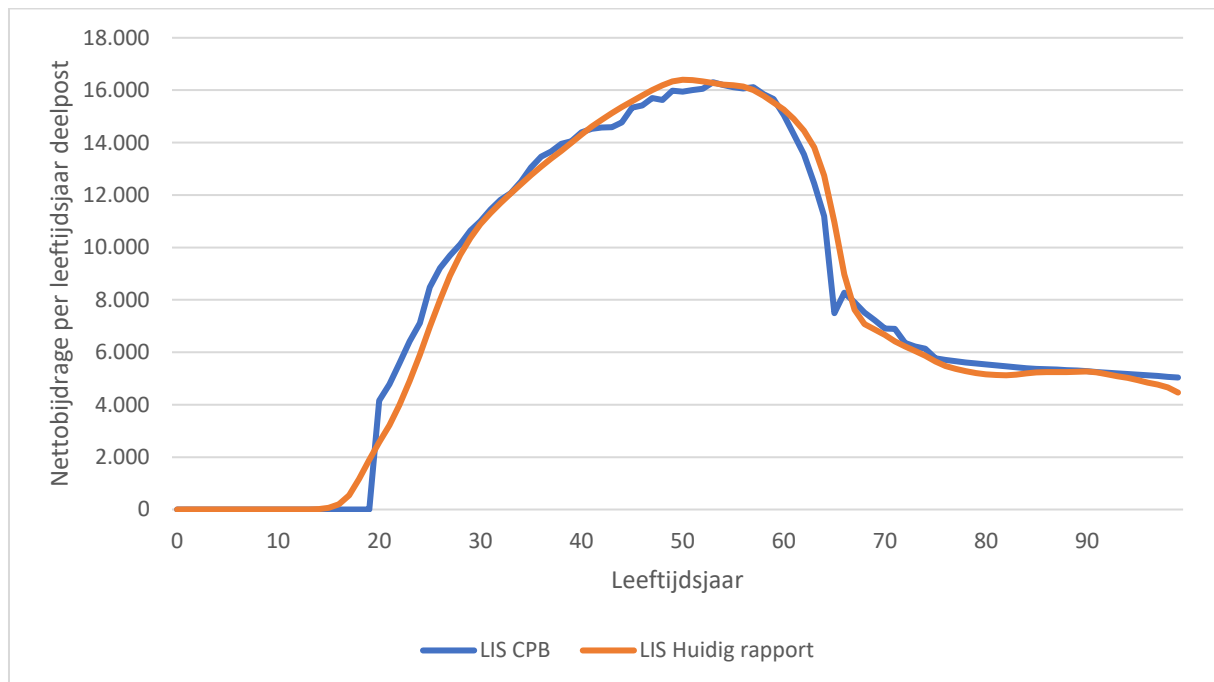
⁷⁸ Dat is gedaan op basis van een bewerkte tabel van Vektis behorend bij het Vektis-rapport Verzekerden in Beeld 2016.

5.5 Uitkeringen, subsidies, premies en belastingen

In deze paragraaf worden de operationalisering van een groot aantal posten uit Tabel 5.1 (post nr. 4 t/m 9, 11, 16, 17 en 20) uiteengezet. Het gaat hier om AKW/WSF, Arbeidsongeschiktheid/ZW, Werkloosheid, bijstand/ANW, AOW, SZ-rest, Subsidies, LIS en Overige directe belastingen huishoudens. Al deze posten hebben gemeen dat ze zijn berekend op basis van het CBS-microdatabestand INPATAB.

Bij de posten AKW/WSF, Arbeidsongeschiktheid/ZW, Werkloosheid, bijstand/ANW en AOW zijn de CBS-microdata gegevens over uitkeringsontvangsten als leidend genomen. Deze posten komen direct voor in INPATAB, zij het dat de samengestelde posten als AKW/WSF nog wel opgeteld moeten worden.

Bij een aantal uitkeringen (WW, Arbeidsongeschiktheid, bijstand, ANW en SZ-rest) is sprake van premies die worden betaald door de uitkeringsinstantie. Deze worden naar rato (soms hebben mensen meer dan één uitkering) bij de uitkering(en) van de ontvanger opgeteld. Hierbij is onderscheid gemaakt tussen sociale premies voor inkomensverzekeringen en voor sociale voorzieningen en de premies voor zorgkosten. Omdat alle premies worden meegenomen aan de inkomstenkant onder de post LIS, vallen deze premies weg tegen de bij de uitkering opgetelde bedragen en kan tegelijkertijd de indeling in posten van het CPB worden aangehouden.



Figuur 5.4 LIS: vergelijking met het CPB-profiel.

De totale kosten lagen voor deze posten vaak lager dan het CPB-macrobedrag in Tabel 5.1. Het verschil tussen het door het CPB becijferde macrobedrag en de voor dit rapport op basis van CBS-microdata berekende som van daadwerkelijk aan personen uitgekeerde bedragen is geïnterpreteerd als uitvoeringskosten.

In de hierna volgende bespreking wordt dit verschil uitgedrukt als percentage van het CPB-macrobedrag. Uitvoeringskosten vormen in het algemeen een aanzienlijk deel van de kosten en worden voor

2015 voor het UWV geschat op gemiddeld 6,8% van de totale kosten⁷⁹ en werden voor dat jaar geschat op totaal 2,9 miljard euro⁸⁰. Voor een deel zijn de uitvoeringskosten direct herleidbaar tot bijvoorbeeld begrotingsstukken.

Voor AKW/WSF is het verschil (4% van het CPB-macrobedrag) grotendeels herleidbaar tot de uitvoeringskosten betreffende AKW⁸¹ en Stufi⁸². Ook bij Arbeidsongeschiktheid/ZW (8% van het CPB-macrobedrag) is dit het geval met een flink aantal posten die verband houden met de uitvoering van Ziekte-wet, WIA en re-integratie van (gedeeltelijke) arbeidsongeschikten die het overgrote deel van het verschil dekken⁸³. Bij Bijstand/ANW (18% van het CPB-macrobedrag) is een heel klein deel terug te voeren op de uitvoering van de ANW door het Rijk⁸⁴ en is de rest geïnterpreteerd als uitvoeringskosten van de gemeenten, die rond 21% liggen⁸⁵. Voor deze posten is het totaalbedrag gewogen naar leeftijd tegen de in Hoofdstuk 3 beschreven totale onderzoekspopulatie en geijkt op de CPB-macrobedragen in Tabel 5.1.

Voor de AOW (3% van het CPB-macrobedrag) is het relatief geringe verschil gedeeltelijk terug te voeren tot uitvoeringskosten⁸⁶ en is de rest geïnterpreteerd als overige apparaatskosten en als AOW-uitkeringen aan in het buitenland woonachtige niet-ingezetenen (niet in het GBA ingeschreven betekent immers dat men niet in de CBS-microdata voorkomt). Van dit verschil is in het huidige rapport ongeveer een kwart (305 miljoen euro) toegerekend aan geremigreerde immigranten. De post AOW is geijkt op de in Hoofdstuk 3 beschreven totale onderzoekspopulatie en geijkt op de CPB-macrobedrag minus 305 miljoen euro zoals weergegeven in Tabel 5.1. De 305 miljoen euro zijn vervolgens naar rato van remigratiekansen e.d. toegerekend aan eerste generatie immigranten.

Bij de WW is het verschil tussen het CBS-microdata bedrag en het CPB-macrobedrag nihil (0% van het CPB-macrobedrag). Echter, de uitvoeringskosten⁸⁷ bedragen 10,4% van het gevonden CBS-microdata bedrag. Hier is het CBS-microdata-bedrag opgehoogd met een lumpsumopslag van 10,4% en komt het hoger uit dat het CPB-macrobedrag, waarbij het verschil ongeveer even groot is als de in de huidige studie bij de WW-uitkering getelde 'premie WW ten laste van de uitkeringsinstantie'.

De post SZ-rest kon maar voor een deel worden ingevuld door CBS-microdata. Het totaalbedrag uit de CBS-microdata wijkt daarom sterk af van het CPB-macrobedrag voor 2016 uit Tabel 5.1. Het CBS-microdata-bedrag is vermeerderd met een opslag van 6% voor uitvoeringskosten, welke herleid is aan de hand van de Wajongkosten en kinderopvangkosten. Het verschil tussen het opgehoogde CBS-microdata-bedrag en het CPB-macrobedrag van circa 13 miljard euro is – conform het CPB-profiel voor SZ-rest – in gelijke mate aan alle inwoners vanaf 20 jaar toegeschreven, ongeacht leeftijd, herkomst e.d. door het totaalbedrag te delen door de totale onderzoekspopulatie vanaf 20 jaar.

⁷⁹ https://www.seor.nl/Cms_Media/S1183-Kosten-en-opbrengsten-terugbrengen-AOW-leeftijd-naar-65-jaar.pdf, blz. 72.

⁸⁰ <https://digitaal.scp.nl/publiekvoorzien/sociale-zekerheid/>

⁸¹ https://www.rijksbegroting.nl/2016/voorbereiding/begroting,kst212308_21.html

⁸² https://www.rijksbegroting.nl/2016/voorbereiding/begroting,kst212220_23.html

⁸³ https://www.rijksbegroting.nl/2016/voorbereiding/begroting,kst212308_21.html

⁸⁴ https://www.rijksbegroting.nl/2016/voorbereiding/begroting,kst212308_21.html

⁸⁵ <https://www.consultancy.nl/nieuws/24796/gemeente-2900-kwijt-aan-uitvoering-bijstandsuitkering>, <https://www.gemeente.nu/sociaal/uitvoering-bijstand-kost-bijna-2900-euro-per-uitkering/>

⁸⁶ https://www.rijksbegroting.nl/2016/voorbereiding/begroting,kst212308_21.html

⁸⁷ https://www.rijksbegroting.nl/2016/voorbereiding/begroting,kst212308_21.html

De post Subsidies is geoperationaliseerd door de INPATAB categorieën Huurtoeslag, Zorgtoeslag, Kindtoeslag en kindgebonden budget te sommeren. Vervolgens is deze post naar leeftijd gewogen tegen de in Hoofdstuk 3 beschreven onderzoekspopulatie en geijkt op het CPB-macrobedrag voor Subsidies in Tabel 5.1. Het verschil (2% van het CPB-macrobedrag) stemt overeen met de uitvoeringskosten⁸⁸ van de toeslagen. Ook deze post is gewogen naar leeftijd tegen de in Hoofdstuk 3 beschreven totale onderzoekspopulatie en geijkt op het CPB-macrobedrag in Tabel 5.1. Voor de posten LIS en Overige directe belastingen huishoudens is een profiel gebaseerd op de INPATAB variabele INPV3900INK (Belasting op inkomen). Daarnaast zijn de premies voor WW, Arbeidsongeschiktheid en AOW die eveneens als INPATAB-variabelen voorkomen hieronder begrepen. Tevens zijn de premies voor zorg meegemeld. Het gaat om AWBZ en de premies voor de ZFW. Bij de ZFW-premies zijn zowel de werknemers- als werkgeverspremies en ook de premies van de verplichte basisverzekering meegenomen. Dit laatste is gedaan omdat de financiering van zorg ondanks de zogenaamde marktwerking e.d. feitelijk vooral fiscaal verloopt. Het zodoende verkregen totaalbedrag komt vrijwel exact overeen met het totaal van de posten LIS en Overige directe belastingen huishoudens en is na weging naar leeftijd tegen de onderzoekspopulatie geijkt op de overeenkomstige CPB-macrobedragen. Het resulterende leeftijdsprofiel voor LIS is getoond in Figuur 5.4.

Bij de posten AOW, Arbeidsongeschiktheid/ZW, Werkloosheid, bijstand/ANW en SZ-rest en de posten LIS, Overige directe belastingen en IRN zijn de data voor de leeftijdsgroep 65-71 jaar synthetisch samengesteld op basis van de gesampelde data voor de leeftijdscategorie die 64 jaar bevat en de leeftijdscategorie die 72 jaar bevat.⁸⁹ Op basis van de gegevens voor de hele bevolking is het huidige verloop van deze posten voor de leeftijdsgroep 65-71 jaar vertaald naar het meest waarschijnlijke verloop bij andere gemiddelde AOW-leeftijden in het bereik 65-71 jaar. Hierbij is uitgegaan van een in 2060 te bereiken AOW-leeftijd van 69,5 jaar.

De in Hoofdstuk 8 beschreven bedragen voor de WW zijn vrij hoog. Dat heeft verschillende oorzaken. Allereerst betaalden autochtonen in het peiljaar 2016 gesommeerd over alle leeftijdsjaren circa € 14.000 meer aan premie dan er werd ontvangen aan WW-uitkeringen.⁹⁰ Daarnaast betalen jongeren relatief veel WW-premie en ontvangen ouderen relatief veel WW-uitkering, zodat contant maken het verschil nog vergroot⁹¹. Verder is uitgegaan van CPB veronderstellingen over de ontwikkelingen vanaf het peiljaar 2016. Deze zijn specifiek voor de ontwikkeling van elke uitkering (WW, Arbeidsongeschiktheid, Bijstand, WAO, enz.), waarbij het macrobedrag voor de WW-uitkeringen aanvankelijk tot 2021 wordt verondersteld te dalen met circa 4,0% per jaar om zich vervolgens uiteindelijk te ontwikkelen richting een generieke groeivoet van 3,5% nominaal. De voornoemde CPB-veronderstellingen zijn echter generiek voor de ontwikkeling van de post LIS (Loonbelasting, Inkomstenbelasting en sociale premies), dus alle premies ontwikkelen zich in de jaren vanaf 2016 op dezelfde wijze, waarbij het macrobedrag voor LIS en dus ook voor de WW-premie juist wordt verondersteld te stijgen met circa 4,7% tot 2021, om zich daarna eveneens geleidelijk te ontwikkelen richting de voornoemde generieke

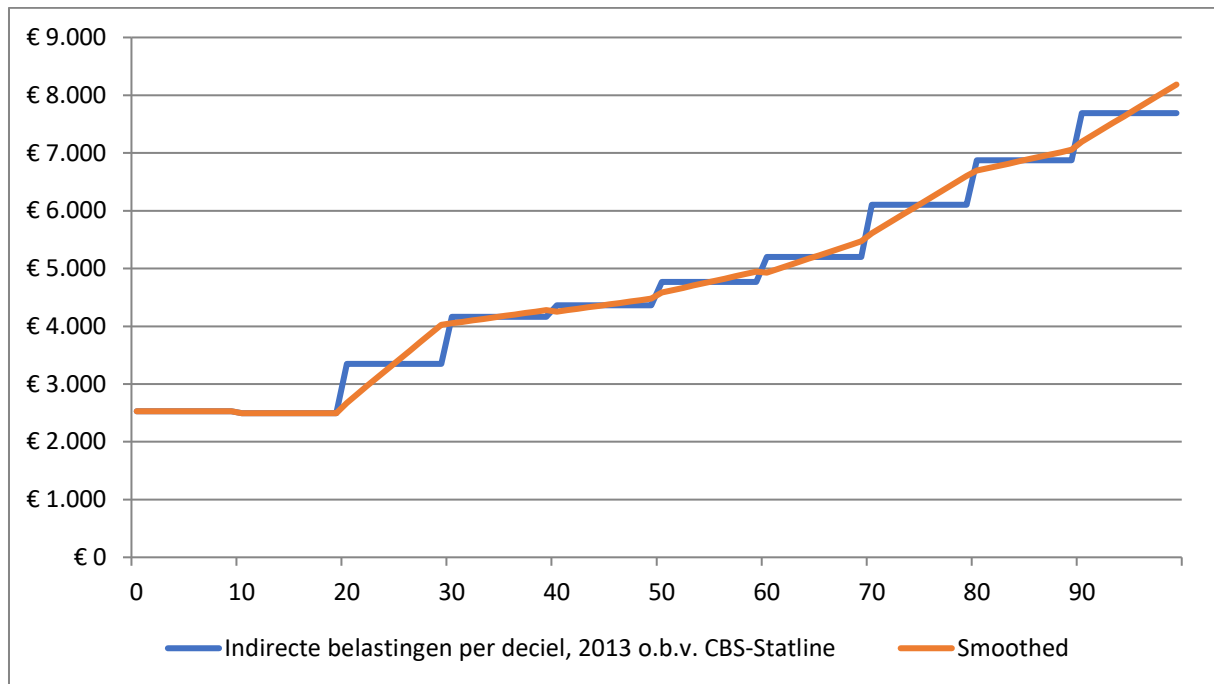
⁸⁸ <https://www.rijksoverheid.nl/binaries/rijksoverheid/documenten/rapporten/2016/12/15/bijlage-beleidsdoorlichting-uitvoering-toeslagen-door-belastingdienst/bijlage-beleidsdoorlichting-uitvoering-toeslagen-door-belastingdienst.pdf>, blz. 58

⁸⁹ Deze wat cryptische omschrijving komt voort uit het feit dat de leeftijdscategorieën verschillen in grootte.

⁹⁰ En vergelijkbaar: voor de gemiddelde ingezetenen €9.000 meer, voor de gemiddelde tweede generatie bijna €3.000 meer en voor de gemiddelde eerste generatie €0 meer.

⁹¹ Immers, de ontvangen bedragen aan WW-uitkering liggen gemiddeld verder in de toekomst en worden dus sterker verkleind door contant maken.

groeivoet van 3,5% nominaal. Dit maakt het verschil tussen de WW-premies en WW-uitkeringen in de jaren vanaf 2016 nog groter. Voor de totale nettobijdrage berekening maakt dit niets uit, omdat hierbij alleen het totaalbedrag voor LIS een rol speelt en de ontwikkeling van dit totaalbedrag is ontleend aan het CPB. Bij het uitsplitsten naar het nettobedrag voor bijvoorbeeld WW geeft het mogelijk echter wel een overschatting. Een alternatief zou zijn om deze berekening uit te voeren met een constante nominale groeivoet, maar er is voor gekozen om de berekening te laten zien zoals hij is uitgevoerd.



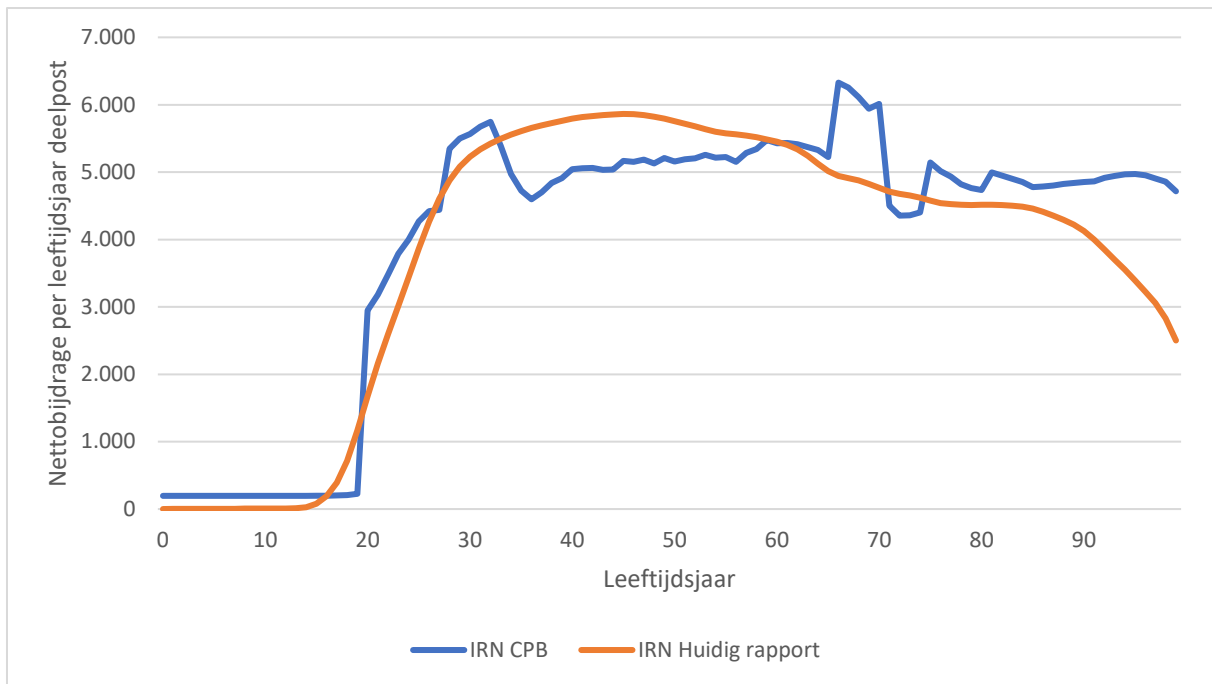
Figuur 5.5 Indirecte belastingen per inkomensdecieel en gesmoothed. Berekening op basis van CBS-statline.

5.6 IRN

In deze paragraaf wordt de operationalisering toegelicht van de post IRN (indirecte belastingen, resterende belastingen en niet belastingmiddelen) waaronder indirecte belastingen vallen als de BTW (zie Tabel 5.1, post nr. 20). Indirecte belastingen zijn een grote inkomstenpost voor de overheid. Het CPB stelt voor 2016 de post IRN op 84 miljard (zie de post belastingen op productie en invoer op CBS-statline of de nationale rekeningen voor 2016) waarvan 68 miljard wordt toegerekend aan huishoudens en de rest aan bedrijven.

Er was geen toegang tot een microdatabestand voor indirecte belastingen en uit de aard der zaak is er ook helemaal geen directe persoonsgebonden registratie van veel indirecte belastingen zoals de BTW. In plaats daarvan is gerekend met een proxy op basis van een [CBS-statline tabel](#) waarin gemiddelde bedragen zijn opgenomen voor de decielen van het bruto huishoudinkomen. Deze decielen zijn eerst per deciel zodanig met een lineaire methode over de percentielen verdeeld dat een zo 'glad' mogelijk profiel ontstond (Figuur 5.5).

Vervolgens zijn de CBS-microdatabestanden INHATAB en INPATAB via de hoofdkostwinnaar aan elkaar gekoppeld, op zodanige wijze dat aan elk persoon in INPATAB het percentiel van het bruto huishoudinkomen (voor zover bekend) kon worden toegewezen. Daarna is het bedrag dat bij percentiel hoort naar rato van het persoonlijk bruto inkomen toegerekend aan de individuele leden van het huishouden. Daarbij zijn negatieve inkomens genegeerd.



Figuur 5.6 IRN: vergelijking met CPB-profiel.

Deze operationalisering levert een totaalbedrag op van 37 miljard. Vervolgens is door vermenigvuldiging met een factor geïjkt op het CPB-macrobedrag. Feitelijk wordt er dus een macrobedrag van 68 miljard ingevuld met een operationalisering die in directe zin een macrobedrag van 37 miljard oplevert. Dit komt omdat in de gebruikte CBS-data niet alle posten zijn begrepen die onder IRN geschaard worden. Vergelijking van het CPB-profiel met deze operationalisering laat zien dat de profielen in grote lijnen overeenstemmen. Wel heeft het voor de huidige studie vervaardigde CBS-microdata profiel een wat gladder en natuurlijker verloop (zie Figuur 5.6), waarschijnlijk omdat het indirect op de ruwe, niet-geaggregeerde inkomensdata van miljoenen personen is gebaseerd. Deze operationalisering levert resultaten op die in lijn liggen met de resultaten die met het CPB-profiel zouden worden verkregen.⁹²

5.7 Vermogen-gerelateerde belastingen

In deze paragraaf wordt de operationalisering toegelicht van de posten Erfbelasting, VPB/dividendbelasting en de IRN via bedrijven (zie Tabel 5.1, post nr. 18, 19 en 21). Deze belastingen hangen onder meer samen vermogen, bijvoorbeeld in de vorm van aandelenbezit via vermogensopbouw in pensioenfondsen. De IRN van bedrijven komt overeen met dat deel van de post belastingen op productie en invoer in de nationale rekeningen en CBS-statline dat via de bedrijven in de schatkist komt. Van de totale IRN wordt in 2016 volgens het CPB 68 miljard direct opgebracht door de huishoudens en de overige 16 miljard indirect via de bedrijven. Bij VPB/dividendbelasting gaat het om de vennootschapsbelasting en dividendbelasting, maar dan beperkt tot Nederlandse bedrijven, in totaal een bedrag van bijna 22 miljard euro.⁹³ Totaal beslaan deze posten dus circa $16 + 22 = 38$ miljard euro.

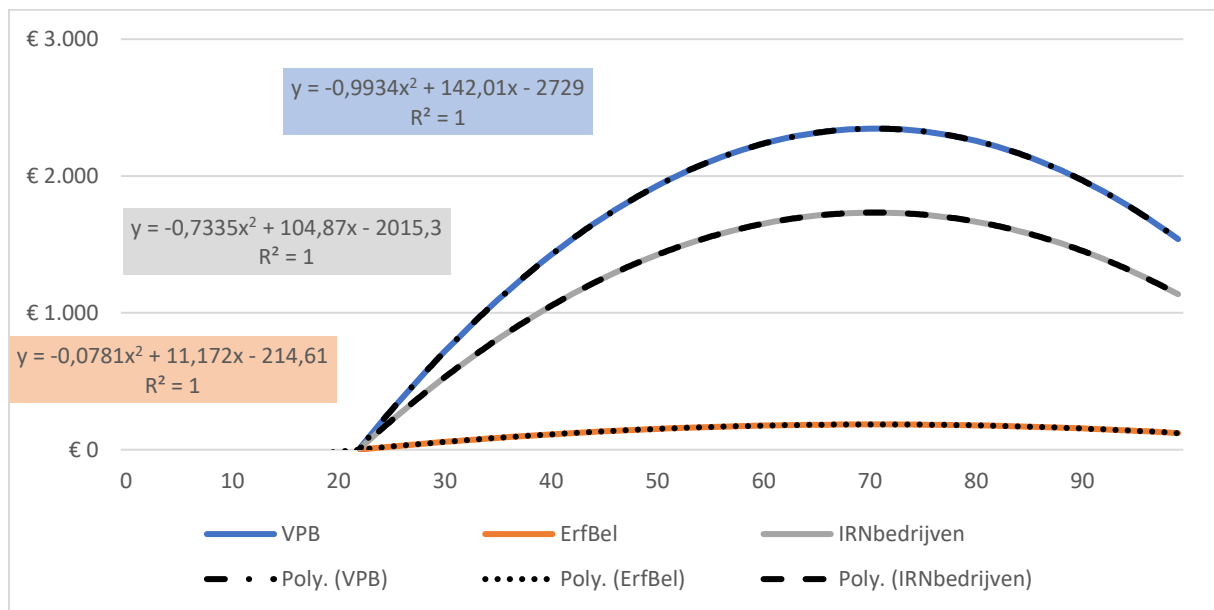
⁹² De verschillen in de profielen middelen elkaar door het ijkten op het CPB-macrobedrag in beginsel in grote lijnen uit. Eventuele verschillen zullen vooral ontstaan als door contant maken, remigratie en het wegen tegen het immigratieprofiel leeftijden waarvoor de verschillen het grootst zijn een relatief groot gewicht krijgen.

⁹³ Vergelijk: <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/82569ned/table?dl=30C53> echter de bedragen zijn anders omdat niet alles wordt meegeteld door het CPB.

Deze posten worden door het CPB bij generatierekeningen voor bijvoorbeeld de vergrijzing naar personen toegerekend. Het CPB zegt over de invulling van de vennootschapsbelasting bijvoorbeeld:⁹⁴

“Het profiel van de vennootschapsbelasting loopt, in lijn met het bezit van vermogen, geleidelijk op tot de leeftijd van 60 jaar om daarna weer te dalen. Aangenomen wordt dat deze belasting uiteindelijk wordt opgebracht door aandeelhouders.”

De essentie is dus dat het CPB het leeftijdsprofiel voor de VPB gelijk laat lopen met het leeftijdsprofiel voor de vermogensontwikkeling van personen/huishoudens. Het CPB rekent zowel de post VPB/dividendbelasting als de post IRN via bedrijven toe aan personen via een ideaaltypisch leeftijdsprofiel (te weten een parabool) zoals afgebeeld in Figuur 5.7. Hetzelfde geldt voor de post Erfbelasting.



Figuur 5.7 CPB-profiel voor posten IRN via bedrijven, VPB/dividendbelasting en erfbelasting met vergelijking en R^2 voor trendlijn.

In de huidige studie wordt het totaalbedrag van 37,8 miljard euro van deze posten op verschillende manieren aan personen toegerekend. De hoofdmoot wordt in navolging van het CPB nog steeds gevormd door het (pensioen)vermogen, maar daarnaast wordt ook het directe bezit van bedrijven meegenomen. In de eerste plaats is er direct aandelenbezit van personen. Het totale bedrag à 1,8 miljard euro aan ontvangen dividenden is bekend uit INPATAB op basis van de variabele INPT3120DIV. Het gaat hierbij om binnenlandse en buitenlandse bedrijven; een onderscheid wordt niet gemaakt.⁹⁵ Voor Nederlandse bedrijven is de totale VPB, DB en IRN van bedrijven als volgt te schatten.

Grote niet-financiële Nederlandse bedrijven keerden 3,1 miljard euro dividend uit en betaalden 1,2 miljard euro vennootschapsbelasting.⁹⁶ Bij een kwart dividendbelasting (zie noot 61) komt dit op 2

⁹⁴ Bron: CPB-rapport Minder zorg om vergrijzing. p. 22. <https://www.cpb.nl/sites/default/files/publicaties/download/cpb-boek-12-minder-zorg-om-vergrijzing.pdf>

⁹⁵ Ontvangen binnenlands en/of buitenlands dividend, toegerekend rendement bij deelneming in buitenlandse beleggingsmaatschappijen en dividend van aangewezen participatiemaatschappijen, zie: <https://www.cbs.nl/nl-onze-diensten/maatwerk-en-microdata/microdata-zelf-onderzoek-doen/microdatabestanden/inpatab-inkomen-van-persoon>

⁹⁶ <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/80262ned/table?dl=31854>

miljard euro aan belastingen. Gaan we uit van de verhouding tussen IRN bedrijven en VPB/DB die het CPB hanteert voor de macrobedragen dan is er voor elke euro VPB/DB 73,8 cent IRN bedrijven. Totaal komt dit op 3,4 miljard euro belasting op 3,1 miljard euro dividend, m.a.w. 112 cent belasting per euro dividend. Ruwweg komt het er in geval van grote niet-financiële Nederlandse bedrijven op neer dat één euro dividend volgens INPATAB-variabele INPT3120DIV in totaal ruim één euro belastingen voor de Nederlandse schatkist oplevert. Echter, een aanzienlijk deel van de dividendbelasting e.d. zal geen betrekking hebben op Nederlandse bedrijven. Bij gebrek aan nauwkeurige gegevens en om schijn-nauwkeurigheid te voorkomen wordt er vanuit gegaan dat de Nederlandse staat voor elke euro dividend 50 cent belasting ontvangt onder de noemers VPB/DB en IRN van bedrijven. Voor het genoemde bedrag van 1,8 miljard komt dat neer op een macrobedrag van 0,9 miljard euro. Blijft over 36,9 miljard euro.

Voor een deel zijn (natuurlijke) personen direct eigenaar van de bedrijven. Hierbij moet onderscheid gemaakt worden tussen directeur-groot aandeelhouders (DGA-s) en overige ondernemers. Voor DGA is middels de INPATAB-variabele 'dividend aanmerkelijk belang' (INPT3110DAB) uit INPATAB de betaalde dividendbelasting te achterhalen. Het macrobedrag voor de gebruteerde⁹⁷ dividenden aanmerkelijk belang beslaat 7,0 miljard euro. Een kwart daarvan is dividendbelasting, ongeveer 1,8 miljard euro. Dit is al verrekend met de loon en inkomstenbelasting.⁹⁸ Nog niet verrekend zijn de belastingen die betaald worden onder de noemers VPB en IRN via bedrijven. Het is echter ondoenlijk dit te herleiden uit de betaalde dividendbelasting, omdat gegevens over de verhouding tussen de verschillende typen belasting bij kleinere bedrijven te schaars zijn. Hetzelfde geldt voor belastingen die worden ontvangen van personen met inkomen uit onderneming die géén DGA zijn. Deze personen hebben allerlei bedrijfskosten waarover btw, bpm, accijnzen e.d. betaald moeten worden. Berekening over een groot deel van de Nederlandse bedrijven⁹⁹ laat zien dat er per euro bedrijfsresultaat naast de bedrijfskosten voor inkoop en voor personeel ook 213 cent 'overige bedrijfskosten' zijn. Onder deze laatste post vallen onder andere voor 22 cent kosten voor vervoersmiddelen, een post met een forse totale belastingplicht, voor onder andere motorrijtuigenbelasting, autobrandstoffen (accijnzen), verzekeringen (assurantiebelasting) en de BPM. Daarnaast zijn er andere belastingen zoals energiebelasting en milieubelastingen en -heffingen. Het is op basis van de beschikbare informatie onmogelijk om echt te berekenen hoeveel belastingen onder de noemers VPB en IRN via bedrijven wordt ontvangen van DGA en andere ondernemers, temeer daar er grote verschillen kunnen zijn tussen bijvoorbeeld een zzp'er en een DGA. Daarom is uitgegaan van een diffuse prior: voor elke euro winst uit onderneming en elke euro dividend dan wel loon van DGA's wordt 10 cent belasting betaald die betrekking heeft op bedrijfskosten. Op basis van de INPATAB-variabelen INPT3110DAB (dividenden uit aanmerkelijk belang, 7,0 miljard), INPT1030DGN (loon DGA, 13,7 miljard) en INPT2070WIN (inkomen uit eigen onderneming, 34,7 miljard) wordt een schatting gemaakt van belastingen via dit kanaal. De genoemde inkomstenbronnen beslaan 55,3 miljard euro. 10% daarvan is 5,5 miljard euro. Blijft over 31,4 miljard euro.

⁹⁷ "De behandeling van de dividendbelasting vloeit voort uit de brutoregistratie van dividend, d.w.z. inclusief dividendbelasting. Dit betekent dat de dividendbelasting geboekt dient te worden bij de sector die het dividend ontvangt. Dit heeft tot gevolg dat er ook dividendbelasting aan het buitenland wordt betaald en uit het buitenland wordt ontvangen."

⁹⁸ Dit is gecontroleerd met een regressie over ruim een half miljoen DGA: elke euro dividend levert 24,9 cent extra loon- en inkomstenbelasting op.

⁹⁹ SBI 2008 indeling, branches B t/m J en N <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/81156NED/table?dl=3617E>.

Het resterende bedrag van de VPB, DB en IRN via bedrijven wordt ingevuld met het indirecte bezit van ondernemingen via de pensioenvermogens. Voor het berekenen van het pensioenvermogen is de variabele PENSAAANSPRBedragOpgebouwdPensioen uit het CBS-microdatabestand PAS (PENSIOENAANSPRAAK2016ANA) herrekend naar de contante waarde behorende bij de leeftijd en pensioenleeftijd van de betreffende persoon. Dit is gedaan voor verschillende discontovoeten, waarbij met discontovoet in dit specifieke geval het rendement op aandelen wordt bedoeld. Uitgangspunt waren de tabellen met factoren van de pensioenfederatie voor 2018 en 2020 voor de negen pensioenleeftijden van 60 tot en met 68 jaar. Deze factoren zijn voor elke combinatie van leeftijd en pensioenleeftijd afzonderlijk (ongewogen) gemiddeld voor mannen en vrouwen. Daarna zijn de factoren als volgt geanalyseerd.

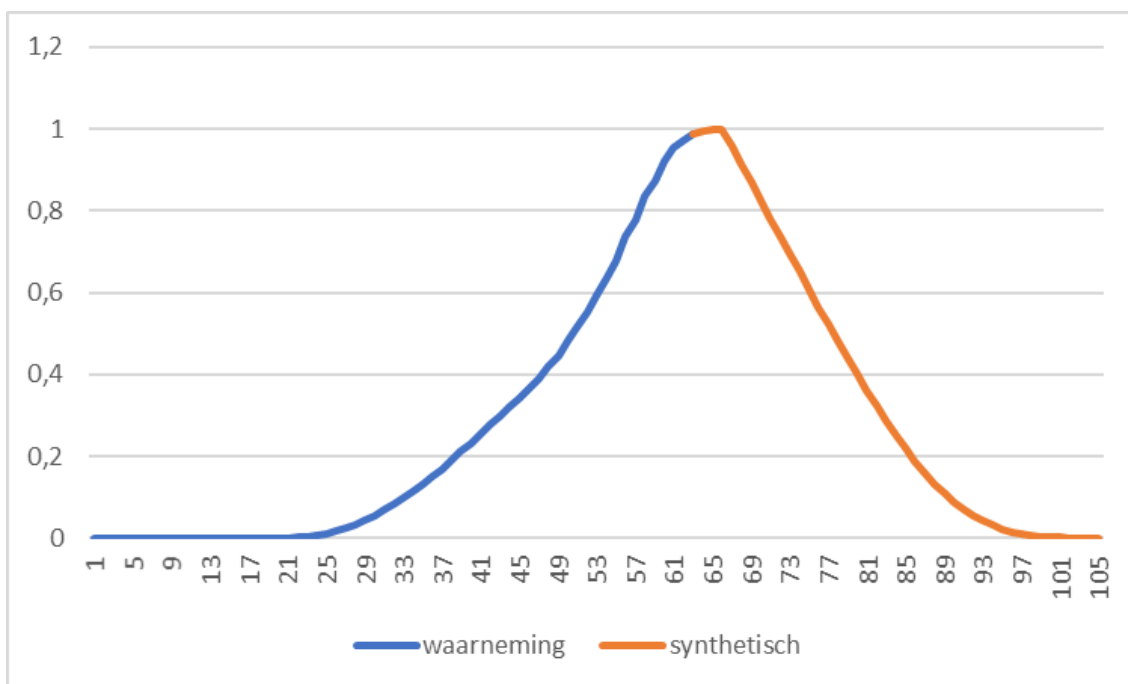
1. Eerst is voor elk van de 9 pensioenleeftijden afzonderlijk de verhouding tussen de factoren berekend van opeenvolgende leeftijden. Hiervan is de discontovoet voor het betreffende jaar (1,648% voor 2018 respectievelijk 0,290% voor 2020) afgetrokken. Het restprofiel is opslagprofiel genoemd. Dit resulteert in negen maal een vector met de hiervoor benoemde verschillen voor alle leeftijden van 16 tot 60-68 jaar (de hoogste leeftijd hangt af van de pensioenleeftijd). Deze opslagprofielen zijn voor alle pensioenleeftijden vrijwel identiek (voor zover ze waarden gemeenschappelijk hebben, wat vanaf 60 jaar niet altijd het geval is). Hiervan is het gemiddelde genomen, gemiddeld opslagprofiel genaamd.
2. Vervolgens is door interpolatie en extrapolatie voor 0, 1, 2, 3 en 4% een vergelijkbaar geschat gemiddeld opslagprofiel gemaakt. Van elk opslagprofiel is met Excel een vierdegraadspolynoom geschat. Ter controle: door optellen van dit vierdegraadspolynoom met de discontovoet worden de verhoudingsgetallen van levensjaar op levensjaar van de omrekenfactoren zeer accuraat verkregen voor de twee rentes waarvoor tabellen beschikbaar zijn (dus voor 2018 1,648% en voor 2020 0,290%).
3. Door voor een bepaalde pensioenleeftijd uit te gaan van de factor in het levensjaar dat de pensioenleeftijd bereikt wordt, genaamd eindfactor, kan door successievelijk delen door de voornoemde verhoudingsgetallen ook de factor berekend worden voor alle levensjaren voorafgaand aan de pensioenleeftijd.
4. De eindfactoren zelf zijn voor de discontovoeten 0, 1, 2, 3 en 4% verkregen door lineaire extrapolatie en interpolatie uit de eindfactoren behorende bij de discontovoeten voor 2018 en 2020 (voor 2018 1,648% en voor 2020 0,290%). Vervolgens is voor elke discontovoet de waarde van de eindfactoren als functie van de pensioenleeftijd geschat met lineaire regressie.

De vier stappen maakt het automatisch genereren van de factoren mogelijk. Op basis van stap 4 wordt de eindfactor voor een bepaalde discontovoet en een bepaalde pensioenleeftijd berekend. Op basis van stap 2 en 3 worden alle factoren berekend voor de jaren voorafgaand aan de pensioenleeftijd. De berekening is uiteindelijk op twee manieren uitgevoerd, éénmaal door uit te gaan van de leeftijd op de peildatum en éénmaal door uit te gaan van de leeftijd ultimo 2016.

Op basis van voorgaande systematiek is voor leeftijden tot 64 jaar de pensioenopbouw geschat. In huidige studie is daarbij uitgegaan van een gemiddelde reëel rendement op aandelen van 2% (een conservatieve schatting). Tot 60 jaar is het berekende profiel gebruikt. Vanaf 60 jaar is de pensioenopbouw fictief. Omdat bij het samplen van groepen met minder data vaak gesampled is voor de leeftijdscategorie 60-63 jaar, is voor de leeftijden 60 tot en met 65 jaar uitgegaan van het gemiddelde over die leeftijd, M_{60-63} genaamd. Vervolgens is het profiel van de pensioenopbouw synthetisch samengesteld.

Voor de leeftijden 60 en 61 jaar is de pensioenopbouw op basis van observatie voor de hele bevolking op 100% resp. 102% van M_{60-63} gesteld. Op basis van de waarnemingen neemt het pensioenvermogen vanaf 62 af, maar dat zou vrijwel zeker vertekening geven, vanwege het oplopen van de pensioenleeftijd. Vandaar dat voor 62, 63 en 64 jaar de pensioenopbouw op 103,5%, 105,0 resp. 106,0% van M_{60-63} gesteld is.

Voor leeftijden vanaf 65 jaar is een afname van het pensioenvermogen geschat, wederom op basis van een rendement op aandelen van 2%. Dit is gedaan door op basis van de CBS-tafelbevolking voor 2016 een factor te schatten voor het startvermogen dat nodig is om vanaf 66 jaar jaarlijks één euro pensioenuitkering te bekostigen over de resterende levensloop. Deze factor is berekend op 16,3228. Voor deze factor is er voor 105-jarigen een restvermogen van nul euro.¹⁰⁰ Hiermee wordt de rest van het synthetische pensioenvermogen aangevuld.



Figuur 5.8 Operationalisering pensioenopbouw en pensioenafname Nederlandse bevolking naar leeftijd, synthetisch profiel.

Al met al resulteert dit in het verloop van het pensioenvermogen voor de gemiddelde Nederlander dat is weergegeven in Figuur 5.8. Het pensioenvermogen is tot 16 jaar nul. Van 16 tot 60 jaar is het verloop geschat op basis het PAS-bestand. Voor 60 tot 66 jaar is het synthetisch aangevuld op basis van het gemiddelde M_{60-63} over de leeftijden 60-63 jaar. Vanaf 66 jaar is een gestage afname verondersteld totdat het pensioenvermogen aan het einde van het 105^e levensjaar gelijk is aan 0.

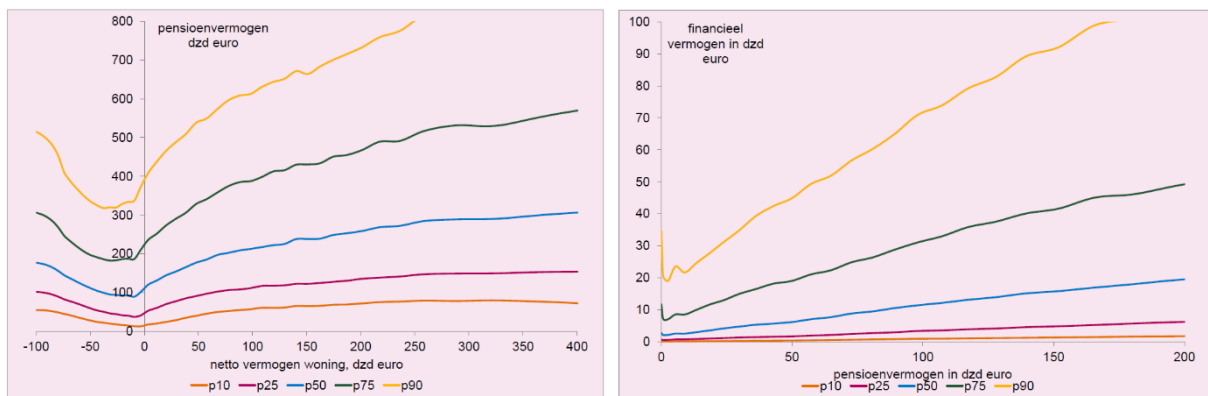
De leeftijdsprofielen voor het directe aandelenvermogen, de belastingafdrachten van ondernemers en het pensioenvermogen zijn vervolgens opgeteld tot een totaal profiel dat in de rest van deze paragraaf wordt aangeduid als 'vermogensprofiel'. Dit vermogensprofiel is als basis gebruikt voor de posten

¹⁰⁰ De tafelbevolking is hiervoor synthetisch aangevuld op basis van CBS-statline data voor 2014-2018 voor aantallen personen van 99-105 jaar <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/37325/table?dl=2F6F2>.

VPB/dividendbelasting en IRN van bedrijven, door het te wegen tegen de leeftijdsopbouw van de hele onderzoekspopulatie en vervolgens te ijking op de betreffende CPB-macrobedragen uit Tabel 5.1.

Voor de erfbelasting is een iets andere operationalisering gekozen. Voor wat betreft de leeftijd-distributie is in navolging van het CBP in beginsel het CPB-profiel bij de post 'ErfBel' uit Tabel 5.1 aangehouden. Vervolgens is dit profiel gewogen naar groepsverschillen in inkomen, vermogen en eigen woningbezit. Hiervoor zijn drie operationalisering gemiddeld. De eerste operationalisering gaat uit van het in de vorige paragraaf gedefinieerde leeftijdsprofiel voor vermogen, waarbij de bedragen zijn gesommeerd over de leeftijden van 28 tot 48 jaar (voor de tweede generatie, dit wegens gebrek aan data voor hogere leeftijden) of 28 tot 68 jaar (voor de overige groepen). Daarna is voor wat betreft deze som de verhouding bepaald tussen de betreffende groep en de bevolking als geheel. Bij de tweede en derde operationalisering zijn op dezelfde wijze ook verhoudingsgetallen bepaald op basis voor het Persoonlijk Bruto Inkomen (variabele INPPERSBRUT uit INPATAB) en de betaalde rente op hypotheekschulden (variabele INPT317ORBW uit INPATAB). De verhoudingsgetallen van deze drie operationalisering zijn vervolgens gemiddeld. Daarna is het voornoemde CPB-profiel 'ErfBel' uit Tabel 5.1 met dit gemiddelde vermenigvuldigd. Tot slot is het resulterende profiel geïjkt op het CPB-macrobedrag voor erfbelasting uit Tabel 5.1, door te wegen tegen de leeftijdsopbouw van de hele onderzoekspopulatie.

Figuur 5.9 Samenhang tussen pensioenvermogen (werknemers) vermogen in de eigen woning (links) respectievelijk financieel vermogen (rechts), 2014. Facsimile van Figuur 3.4 respectievelijk 3.5 uit CPB, De verscheidenheid van de vermogens van huishoudens, opgehaald 12-12-2020 van: <https://www.cpb.nl/publicatie/de-verscheidenheid-van-vermogens-van-huishoudens>



De rationale achter deze drie operationalisering is dat er over het algemeen een redelijk samenhang lijkt te bestaan tussen enerzijds het financiële vermogen en het netto vermogen in de eigen woning (beide belangrijke vermogensbestanddelen bij erfenissen) en anderzijds het pensioenvermogen (zie Figuur 5.9) en verder dat het inkomen behoorlijk sterk samenhangt met het vermogen van verschillende herkomstgroepen. De beperking tot leeftijden van 28 tot 48 jaar voor de tweede generatie is omdat er bij de tweede generatie niet voor alle groepen data is voor personen vanaf 48 jaar.

6 Aanloopkosten en kosten na remigratie

6.1 Asielopvang en vergunningsverstrekking

Naast kosten en baten die voor alle burgers over de hele levensloop gelden, zijn er enkele kostenposten die alleen spelen voor eerste generatie immigranten. Allereerst zijn er de aanloopkosten van immigratie. De opvang van asielzoekers, het verstrekken van verblijfsvergunningen en de inburgering van niet-westerse immigranten brengen overheidskosten met zich mee en die kunnen voor de desbetreffende groepen worden meegenomen in de berekening.

In deze studie zijn hiervoor tentatieve berekeningen gemaakt, die een orde van grootte aangeven.¹⁰¹ Voor inburgeringskosten en vergunningsverstrekking is een scheiding gemaakt tussen asielmigranten en overige immigranten. Allereerst zijn de IND-kosten verdeeld over verblijfsvergunningen voor asiel en reguliere immigratie. Op basis van het budgettair kader 2017¹⁰² is geschat dat de kosten voor asiel circa 2,4 maal zwaarder zijn dan de kosten voor regulier. Door het naar regulier toe te schrijven deel van de IND kosten over de periode 2013-2018 (na inflatiecorrectie¹⁰³) over te slaan over het totaal aantal reguliere verblijfsvergunningen over de periode 2013-2018 zijn de kosten voor een reguliere vergunning geschat op circa € 3.100. Hierbij is uitgegaan van de voor IND-kosten genoemde bedragen in de Rijksbegrotingen 2013-2018.¹⁰⁴

Voor asiel zijn de opvangkosten als volgt berekend. Eerst zijn de grootste posten opgeteld om te komen tot een totaalbedrag per jaar. Meegenomen in deze berekening zijn de kosten uit de Rijksbegrotingen 2013-2018 voor COA, Nidos en een lumpsum van 30 miljoen per jaar voor de posten Vluchtelingenwerk en Terugkeer en bewaring.¹⁰⁵ Daarnaast zijn de kosten van IND meegenomen voor zover toerekenbaar naar asiel. De kosten per vergunning zijn berekend door de totale kosten over de periode 2013-2018 (na inflatiecorrectie¹⁰⁶) uit te drukken in euro's van 2016 en vervolgens te delen door het totaal aantal verblijfsvergunningen voor asiel in de periode 2013-2018.¹⁰⁷ De kosten per vergunning komen zo op circa € 53.700.

Het eindbedrag drukt dus niet de kosten van de opvang van één asielzoeker uit, maar de kosten verbonden aan het toelaten van één asielmigrant, inclusief de kosten van het selectiemechanisme dat de toelating op basis van het VN-vluchtelingenverdrag nu eenmaal is. Over de periode 2013-2018 is het

¹⁰¹ Het is niet moeilijk om op basis van de tabellen-set bij dit rapport de totalen aan te passen met een eigen berekening.

¹⁰² TK 34 550 VI, Nr. 2 Tabel 37.4.

¹⁰³ Loonvoet overheid, CEP2018, Bijlage 08, https://www.cpb.nl/sites/default/files/omnidownload/Verzamelde_bijlagen_CEP_2018_0.xlsx

¹⁰⁴ http://www.rijksbegroting.nl/2013/voorbereiding/begroting,kst173859_15.html
http://www.rijksbegroting.nl/2014/voorbereiding/begroting,kst186612_12.html
http://www.rijksbegroting.nl/2015/voorbereiding/begroting,kst199430_12.html
http://www.rijksbegroting.nl/2016/voorbereiding/begroting,kst212222_11.html
http://www.rijksbegroting.nl/2017/voorbereiding/begroting,kst225850_17.html
http://www.rijksbegroting.nl/2018/voorbereiding/begroting,kst236777_17.html

¹⁰⁵ Zie voetnoot 104.

¹⁰⁶ Loonvoet overheid, CEP2018, Bijlage 08, https://www.cpb.nl/sites/default/files/omnidownload/Verzamelde_bijlagen_CEP_2018_0.xlsx

¹⁰⁷ Het aantal verblijfsvergunningen asiel bedraagt over de periode 2013-2018: 131.760, zie CBS-statline *Verblijfsvergunningen voor bepaalde tijd; verblijfsgrond en nationaliteit*, opgehaald 18-12-2020 van: <https://open-data.cbs.nl/statline/#/CBS/nl/dataset/82027NED/table?dl=2D47B>.

verschil overigens miniem omdat het aantal asielzoekers het aantal afgegeven verblijfsvergunningen asiel niet veel ontloopt.¹⁰⁸

Het bedrag voor de asielopvang is vrij hoog, vooral vanwege de forse kosten voor het COA. Deze bedroegen volgens de Rijksbegrotingen 2008-2018 over de periode 2013-2018 5,1 miljard euro voor 132.000 verblijfsvergunningen (€ 39.000 per verblijfsvergunning) en over de periode 2008-2018 7,4 miljard euro voor 169.000 vergunningen (€ 43.000 per vergunning).¹⁰⁹ Daar komen de kosten voor o.a. de IND, Vluchtelingenwerk, Nidos en de dienst Terugkeer en Vertrek nog bij.

6.2 Inburgering

In deze paragraaf wordt een tentatieve berekening gegeven voor de kosten van inburgering. In 1998 werd de Wet Inburgering Nieuwkomers (WIN) van kracht. Sindsdien bestaat de formele verplichting voor immigranten om zich te bekwamen in de Nederlandse taal en kennis te vergaren over de Nederlandse samenleving. Aanvankelijk werden de kosten gedragen door de overheid. In de CPB-studie uit 2003 wordt voor de inburgeringskosten zonder nadere toelichting een bedrag genoemd van € 7.000 in 2001.¹¹⁰ In prijzen van 2016 komt dit bedrag, afhankelijk van de gehanteerde prijsindex, ergens rond de € 10.000.¹¹¹ Omdat de inhoud van de inburgeringscursussen sinds 2001 *grosso modo* hetzelfde is gebleven, geeft dit cijfer ook een indicatie van de orde van grootte van de huidige kosten. Sindsdien is echter de bekostigingssystematiek gewijzigd.

Met de Wet Inburgering van 2006 (ingegaan 2007) werd eigen verantwoordelijkheid meer van belang. Op het niet tijdig slagen voor het inburgeringsexamen werden naast boetes ook verblijfsrechtelijke sancties in de wet opgenomen. In 2013 volgde een ander bekostigingssysteem in de vorm van een leenstelsel:

“Voor asielmigranten is het maximaal te lenen bedrag gelijk aan € 10.000,-. Voor gezinsmigranten is het maximaal te lenen bedrag inkomensafhankelijk, maar ook nooit hoger dan € 10.000,-. Alleen van asielmigranten, dus niet van gezinsmigranten, kan de lening na het voldoen aan de inburgeringsplicht onder voorwaarden worden kwijtgescholden.”¹¹²

Het cursusaanbod werd aan de markt overgelaten. Mede omdat de verblijfsrechtelijke sancties, die een sterke prikkel konden vormen voor de inburgeraar om zich in te spannen, in de praktijk zelden werden toegepast, bleven de met deze wet beoogde resultaten uit.

Een en ander was aanleiding om in het regeerakkoord voor het kabinet Rutte III de inburgering andermaal op de schop te nemen. De verblijfsrechtelijke sancties blijven – althans op papier – overeind:

¹⁰⁸ Het aantal 1^e asielaanvragen over de periode 2013-2018 is nagenoeg gelijk aan het aantal verblijfsvergunningen asiel: 130.770, zie *Asielverzoeken en nareizigers; nationaliteit, geslacht en leeftijd*, opgehaald 18-12-2020 van: <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/83102NED/table?dl=2370E>

¹⁰⁹ Zie voetnoot 104.

¹¹⁰ Roodenburg et al. (2003) *Immigration and the Dutch Economy*

¹¹¹ Zie het Excel-bestand *Inburgeringskosten_volgens_CPB2003 in prijzen van 2016.xlsx*

¹¹² Blom, Michiel et al. (2018) *Inburgering: systeemwereld versus leefwereld*, Evaluatie Wet Inburgering 2018, blz. 18

“Verwijtbaar niet inburgeren heeft consequenties, zoals het verliezen van de verblijfstatus voor reguliere migranten en het niet verkrijgen van een sterkere verblijfsstatus [lees: permanente verblijfsvergunning] voor asielvergunninghouders.”¹¹³

De bekostiging wordt ingrijpend gewijzigd. Het leenstelsel wordt afgeschaft en de bekostiging geschiedt opnieuw door de overheid. De uitvoering wordt neergelegd bij de gemeenten. Inhoudelijk is sprake van een zwaarder programma, omdat “[d]e taaleis wordt aangescherpt van A2 naar B1. Hiertoe wordt ook taalles op niveau B1 gefinancierd door de rijksoverheid.”¹¹⁴ Het kabinet is voornemens om de wijzigingen in het jaar 2020 te doen ingaan.¹¹⁵

Bij de schatting van de inburgeringskosten per immigrant voor zover deze ten laste komen van de overheid, zijn de volgende algemene overwegingen gemaakt. Volgens de systematiek van de huidige studie worden de gevolgen van immigratie voor de collectieve financiën berekend uitgaande van de bestaande instituties. In de praktijk betekent dit dat de situatie in 2016 uitgangspunt vormt. Dit is het meest recente jaar waarvoor alle of in ieder geval de belangrijkste gegevens beschikbaar zijn. Voor de inburgeringskosten wordt echter rekening gehouden met de beleidswijzigingen die met ingang van 2020 zijn voorzien (zie het voorgaande). Bedragen na 2016 worden uitgedrukt in prijzen van 2016. Een en ander betekent dat er twee schattingen gemaakt moeten worden, te weten voor de jaren 2016-2019 en voor 2020 en latere jaren.

Voor de periode 2016-2019 zijn de volgende overwegingen gemaakt. Onder het huidige leenstelsel kan maximaal € 10.000 worden geleend. Gezinsmigranten moeten het met een inkomensafhankelijk maximum doen. Tenzij ze in de laagste inkomenscategorie vallen ligt dat lager dan de genoemde € 10.000. Dat is echter niet zozeer een indicatie dat voor deze categorie lagere kosten gelden, maar het lagere bedrag vloeit veeleer voort uit de gedachte dat zij een eigen bijdrage kunnen leveren.

De kosten die ten laste komen van de overheid worden bepaald door drie factoren:

- Het gemiddelde bedrag dat wordt geleend (= totaal bedrag aan leningen gedeeld door het aantal immigranten inclusief niet-leners);
- Het percentage kwijtscheldingen wegens voldoen aan de inburgeringsplicht (alleen asielmigranten);
- Het percentage oninbare vorderingen inzake niet kwijtgescholden leningen.

Wat betreft de punten 1 (geleend bedrag) en 2 (kwijtscheldingen) biedt het evaluatierapport van Blom et al.¹¹⁶ enige aanknopingspunten:

“Van de inburgeraars die in 2013 inburgeringsplichtig zijn geworden blijkt uit deze evaluatie, dat op peildatum 1 september 2017 ongeveer driekwart heeft voldaan aan de verplichtingen die voortvloeien uit de inburgeringswet. Deze groep bestaat uit: geslaagden, personen met een vrijstelling en ontheven personen. Van de inburgeraars die in 2013 inburgeringsplichtig werden, heeft circa 60% het inburgeringsexamen (of Staatsexamen NT2) behaald op de peildatum. Aan 5% van de inburgeraars is een vrijstelling toegekend. Inburgeraars kunnen een vrijstelling krijgen wanneer zij een Nederlandse opleiding hebben afgerond met een

¹¹³ (2017) Vertrouwen in de toekomst, Regeerakkoord 2017-2021, blz. 55

¹¹⁴ Zie noot III blz. 54

¹¹⁵ Rijksoverheid (2018) Inburgering op de schop

¹¹⁶ Zie noot II blz. 5 en blz. 19

startkwalificatie (MBO-2-niveau of hoger). Inburgeraars kunnen ontheven worden van de inburgeringsplicht: op medische gronden, op basis van aantoonbaar geleverde inspanningen of omdat zij aantoonbaar voldoende zijn ingeburgerd (zie hoofdstuk 7).¹¹⁷ Ruim één op de tien inburgeraars (11%) uit het eerste cohort heeft een ontheffing gekregen. Het resterende kwart heeft een verlenging gekregen van de reguliere inburgeringstermijn van drie jaar en is nog bezig met inburgeren. De helft van die verlengingen is het gevolg van een zogeheten verwijtbare termijnoverschrijding. De inburgeraars die dat betreft hebben een boete gekregen voor het niet tijdig voldoen aan de inburgeringsplicht en een nieuwe inburgeringstermijn van twee jaar. ... Indien de inburgeraar verwijtbaar niet tijdig aan de verplichtingen uit de wet heeft voldaan, dan legt DUO [= de betrokken overheidsinstantie] een boete op en gaat een nieuwe termijn van twee jaar lopen. Het recht op kwijtschelding van de lening vervalt.”

Het genoemde evaluatierapport geeft ook informatie over de verstrekte leningen:

“Uit de gegevens van DUO blijkt dat het overgrote deel (94%) van de inburgeringsplichtige asielmigranten uit de cohorten 2013 een lening heeft (zie Figuur 5-9). Slechts een derde van de gezinsmigranten heeft een lening. Het gemiddelde leenbedrag van gezinsmigranten ligt sterk lager dan dat van asielmigranten (Figuur 5- 10). Gezinsmigranten die zijn geslaagd voor het (inburgerings- of staats)examen lenen gemiddeld € 2200,- ten opzichte van geslaagde asielmigranten die gemiddeld € 6000,- lenen. Figuur 5-11 geeft inzicht in de verdeling van de geleende bedragen voor de diverse groepen inburgeraars. Hieruit blijkt dat bijna de helft van de asielmigranten meer dan € 7000,- leent, en een kwart het maximale bedrag van € 9000 - € 10.000,-. De meerderheid van de gezinsmigranten leent tussen de € 1000,- en € 3000.”¹¹⁸

Uit de in bovenstaand citaat genoemde figuur valt af te lezen dat de niet-geslaagden onder zowel asiel- als gezinsmigranten circa € 1.400 meer lenen dan de geslaagden. De verhouding geslaagd/niet geslaagd is ongeveer 60/35 (voor beide categorieën samen).¹¹⁹

Het gemiddelde bedrag dat wordt geleend laat zich als volgt berekenen:

- Asielmigranten: $((6000 \times 60 + 7400 \times 35) / 95) \times 0,94 =$ afgerond € 6.100
- Gezinsmigranten: $((2200 \times 60 + 3600 \times 35) / 95) \times 0,33 =$ afgerond € 900

Welk deel van deze bedragen wordt kwijtgescholden (punt 2) is bepalend voor de overheidsfinanciën. Dit is alleen van toepassing op asielmigranten. Bekend is dat een kwart van alle inburgeringsplichtigen (helaas niet alleen asielmigranten) niet binnen de gestelde termijn slaagt, en dat in de helft (12,5%) van deze gevallen sprake is van ‘verwijtbare tijdsoverschrijding’, met als gevolg dat de kwijtschelding vervalt. Gemiddeld per inburgeringsplichtige asielmigrant komt dit neer op $7400 \times 0,125 =$ afgerond € 900

Tenslotte punt 3 (oninbare vorderingen). Gezien het lage inkomen van veel inburgeringsplichtige immigranten zullen niet alle (niet kwijtgescholden) leningen uiteindelijk worden afbetaald. Bij gebrek aan

¹¹⁷ Die ontheffing wordt achteraf verleend als de inburgeraar aan zijn inspanningsverplichting heeft voldaan,

¹¹⁸ Zie noot II blz. 58

¹¹⁹ Zie noot II blz. 4, diagram

harde cijfers wordt een 'diffuse prior' gehanteerd, te weten afbetaling voor de helft. Tabel 6.1 geeft een samenvatting van de bevindingen:

Tabel 6.1 Inburgeringskosten 2016-2019 per jaar per immigrant (in €)

	[1]	[2]	[3] = [1] – [2]	[4] = [3] × 0,5	[2] + [4]
	lening	kwijtschelding	restant lening	oninbaar	Kosten overheid
Asielmigrant	6100	5200	900	450	5650
Gezinsmigrant	900	0	900	450	450

Voor de kostenontwikkeling vanaf 2020 zijn de volgende overwegingen gemaakt. Zoals is voorzien in het regeerakkoord wordt in het stelsel dat in 2020 moet ingaan het niveau voor het onderdeel taal verhoogd. Wat die verhoging van het taalniveau betekent voor de kosten is voorsnog onduidelijk. De volgende cryptische zin in het regeerakkoord brengt niet veel duidelijkheid: "De huidige financiële kaders, samen met de middelen die in het regeerakkoord zijn vrijgemaakt voor de verhoging van het taalniveau, gelden als harde randvoorwaarden."¹²⁰ In de financiële bijlage van het regeerakkoord staat een post nummer 93, waaruit men zou kunnen opmaken dat het gaat om een stijging van 60 naar 70 miljoen (=17%) voor het onderdeel taal. Bij gebrek aan helderheid is uitgegaan van een stijging van de helft van dit percentage en is vervolgens afronding op veelvouden van € 50. Dit geeft de volgende bedragen:

- Asielmigrant: € 5.650 × (1 + 0,17 / 2) = (afgerond) € 6.150
- Gezinsmigrant: € 450 × (1 + 0,17 / 2) = (afgerond) € 500

Met deze bedragen zijn de inburgeringskosten berekend. Daarbij is per groep een schatting gemaakt van de fractie personen met een inburgeringsplicht, die afhankelijk is van variabelen als leeftijd en migratieachtergrond. Op basis van deze fractie en het immigratieprofiel is vervolgens het gemiddelde bedrag per persoon berekend voor elke groep. Tot slot is het leeftijdsprofiel voor de post Openbaar bestuur met dit bedrag opgehoogd. Om dit te compenseren is het initiële leeftijdsprofiel voor de post Openbaar bestuur voorafgaand aan de berekening zodanig verlaagd dat voor de bevolking als geheel het macrobedrag voor 2016 weer in overeenstemming is met het bedrag in Tabel 5.1.

6.3 AOW-rechten

Naast kosten die optreden rond de immigratie, zijn er ook kosten die doorlopen na eventuele remigratie. Een (r)emigrant houdt veelal recht op (een deel van) de opgebouwde AOW-rechten en die kosten voor de schatkist lopen in dat geval dus deels door na vertrek uit Nederland. Ook hiervoor is een tentatieve berekening gemaakt. Hiervoor is per herkomstgroep, migratiemotief of combinatie van herkomst en motief berekend wat het aandeel zogenaamde verdragslanden voor het meenemen van AOW-rechten is.¹²¹ Vervolgens is op basis van de remigratiekansen en het immigratieprofiel berekend hoeveel AOW-rechten immigranten opgebouwd hebben op het moment van remigratie.

¹²⁰ Zie noot III, blz. 2

¹²¹ https://www.svb.nl/int/nl/algemeen/eu_eer.jsp

<https://www.svb.nl/int/nl/algemeen/verdragslanden.jsp>

<https://www.uvw.nl/particulieren/overige-onderwerpen/internationaal/handavingsverdrag-naar-welke-landen-kan-uitkering-mee/detail/uitzonderingen-wajong-uitkering-ww-uitkering>

Vervolgens is de contante waarde van de opeisbare AOW-rechten bepaald tegen de bij het betreffende scenario behorende discontovoet en sterfttekansen. Hierbij zijn twee lumpsumbedragen gehanteerd: één voor mensen uit verdragslanden en één voor mensen uit de niet-verdragslanden, die minder rechten hebben op het 'meenemen van de AOW' in geval van remigratie.¹²² Hierbij is in beide gevallen er van uitgegaan dat er over het bruto bedrag 18,65% belasting betaald wordt in Nederland en deze belasting is in mindering gebracht.

Het bedrag voor verdragslanden is berekend door het gewogen gemiddelde te nemen van de in 2016 uitbetaalde bedragen voor de AOW, gewogen naar de CBS-tafelbevolking 2016, waarbij de leeftijd 68 jaar het uitgangspunt was (gewicht 1,0). Het wegen is gedaan om te compenseren voor het feit dat de verhouding samenwonenden / alleenstaanden wijzigt met de leeftijden en daarmee het AOW bedrag. Het bedrag voor 67 jaar is half geteld, om te compenseren voor het feit dat gemiddeld genomen het bedrag in het jaar dat de AOW begint lager zal zijn.¹²³ Na aftrek van voornoemde belasting levert dit een bedrag op van € 9.331 per jaar.

Voor het bedrag uit de niet-verdragslanden is uitgegaan van 50% van het minimumloon inclusief vakantiegeldopslag. De bedragen voor het minimumloon tot 1 juli 2016 en vanaf 1 juli 2016 zijn daarbij gemiddeld.¹²⁴ Er is vanuit gegaan dat het bedrag in het jaar dat het recht intreedt de helft is van het bedrag in de daarop volgende jaren. Na aftrek van een in Nederland te betalen belasting van 18,65% levert dit een bedrag van € 7.817 per jaar.

Vervolgens is de netto contante waarde van de AOW-rechten voor verdragslanden en niet-verdragslanden berekend, tegen de in het betreffende scenario geldende discontovoet en sterfttekansen. Dit zijn overigens *gemiddeld per immigrant* geen heel grote bedragen (veelal minder dan € 10.000), omdat veel immigranten jong komen (gemiddeld vaak rond 25 jaar) en kort blijven (een groot deel is binnen 10 jaar weer weg). Bij remigratie zijn de opgebouwde rechten daarom nog beperkt en doordat de AOW bij jongeren nog ver in de toekomst ligt worden de bedragen door het contant maken en tussentijdse sterfttekansen nog een stuk lager. Bij een groeivoet van 1% en discontovoet van 2,5% – zoals de CPB hanteert in haar laatste vergrijzingsstudie (2019) – ligt na verdiscontering van sterfttekansen, de contante waarde van de AOW van jonge immigranten – twintigers en (jonge) dertigers bijvoorbeeld – in de orde van grootte van een ton. Een immigrant die vijf verblijfsjaren heeft die meetellen voor de AOW-opbouw heeft recht op 10% van de AOW en de contante waarde daarvan op het moment van immigratie ligt dus slechts in de orde van grootte van € 10.000.

Deze berekeningen zijn tentatief, bedoeld om een orde van grootte aan te geven. De werkelijke bedragen van de kosten na remigratie liggen waarschijnlijk anders. Wellicht worden lang niet alle AOW-rechten opgeëist en wellicht valt door allerlei wettelijke en praktische bezwaren het gemiddelde opeisbare AOW-recht in niet-verdragslanden nog veel lager uit. Daar staat echter tegenover dat de kosten

<https://statline.cbs.nl/Statweb/publication/?DM=SLNL&PA=71093ned&D1=1&D2=0&D3=0&D4=a&D5=a&D6=l&HDR=T&STB=G1,G2,G3,G4,G5&VW=T>
https://www.belastingtips.nl/kennisbank/fiscaaladvies_dossier/emigratie/recht_op_aow_na_emigratie/

¹²² SVB, *Overzicht verdragslanden*, opgehaald 12-12-2020 van: <https://www.svb.nl/nl/aow/verdragslanden>

¹²³ Het is allemaal nog iets ingewikkelder, want de AOW-leeftijd verschuift, maar dit volstaat voor een tentatieve berekening.

¹²⁴ Staatscourant, stcrt-2015-10678 stcrt-2016-24815.

na emigratie ook hoger zouden kunnen liggen dan hier verondersteld. Voor een deel houden remigranten bijvoorbeeld recht op Nederlandse zorg en dat is hier niet meegeteld.

7 Demografie en vergrijzing

7.1 Geboorte en sterfte

De leeftijdsspecifieke vruchtbaarheid is gebaseerd op CBS-statline gegevens voor de jaren 2014-2018 van het aantal levend geboren kinderen per duizend van de gemiddelde vrouwelijke bevolking per leeftijd. Deze data zijn gesommeerd over de jaren 2014-2018 voor elk van de volgende 35 leeftijdscategorieën afzonderlijk: 'jonger dan 16 jaar', de afzonderlijke leeftijden van 16 tot 49 jaar en de leeftijdscategorie '49 jaar of ouder'. Op basis van deze data is een leeftijdsprofiel gemaakt voor de leeftijdsspecifieke vruchtbaarheid. Dit profiel heeft waarde 0 voor leeftijden van 0 tot 15 jaar en voor leeftijden van 50 tot 100 jaar. De waarde voor de overige leeftijden (15 tot 50 jaar) is als volgt bepaald. De waarde voor 15 jaar is gelijk gesteld aan de (hiervoor beschreven) som voor de categorie 'jonger dan 16 jaar'. De waarden voor 16 tot 49 jaar zijn gelijk gesteld aan de (hiervoor beschreven) afzonderlijke sommen voor 16 tot 49 jaar. De waarde voor 49 jaar is gelijk gesteld aan de (hiervoor beschreven) som voor de categorie '49 jaar of ouder'. Tot slot is het resulterende leeftijdsprofiel met een normalisatiefactor vermenigvuldigd zodanig dat een kansverdeling ontstond (m.a.w. zodanig dat de som over alle leeftijden gelijk is aan 1).

Voor het kindertal per vrouw is uitgegaan van data van het CBS¹²⁵ voor het kindertal van de eerste generatie. Deze data zijn echter alleen beschikbaar voor tamelijk grove geografische categorieën. Daarom is ook data van de Verenigde Naties voor de Total Fertility Rate in herkomstregio's gebruikt¹²⁶. Vervolgens is het gewogen gemiddelde ($\frac{2}{3}$ CBS, $\frac{1}{3}$ VN) van beide cijfers gebruikt. De vruchtbaarheid is een dynamisch gegeven, maar voor de eenvoud is verondersteld dat ze constant blijft. Dit levert uiteraard vertekening op, vooral voor geboorten die verder in de toekomst liggen. Echter, omdat de piek in de distributie van entreeleeftijden (meestal rond 25 jaar) ongeveer samenvalt met de piek in de leeftijdsspecifieke vruchtbaarheid (rond 30 jaar), zal een groot deel van de tweede generatie kind zijn van redelijk recent gearriveerde eerste generatie immigranten, waardoor deze vertekening naar verwachting beperkt blijft.

Om de berekening te vereenvoudigen is geen onderscheid gemaakt tussen mannen en vrouwen. Het aantal geboorten is voor elke groep als volgt bepaald. Het kindertal van de betreffende groep is eerst gehalveerd. Vervolgens is de uitkomst vermenigvuldigd met het leeftijdsprofiel voor de leeftijdsspecifieke vruchtbaarheid. Uit de resulterende vector en de vector voor de leeftijdsopbouw van de betreffende groep in het betreffende jaar is door matrixvermenigvuldiging het aantal geboorten voor het betreffende jaar bepaald.

Daarnaast is rekening gehouden met het feit dat in het buitenland geboren kinderen van eerste generatie immigranten zelf ook tot de eerste generatie behoren. Ook is verdisconteerd dat een deel van de immigranten tijdens of na de vruchtbare leeftijd immigrereert. Hierbij is aangenomen dat kinderen tot 18 jaar altijd met de ouders mee emigreren en kinderen vanaf 18 jaar altijd in Nederland blijven. Mogelijke levert dit vertekening op als er zelfselectie op zou treden ten aanzien van het al dan niet hebben

¹²⁵ CBS-statline, *Geboorte; vruchtbaarheid, migratieachtergrond en generatie moeder*, opgehaald 12-5-2020 van: <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/83307NED/table?dl=39911>

¹²⁶ United Nations, Department of Economic and Social Affairs, Population Division (2019). *World Population Prospects 2019*, Online Edition.

van kinderen. Als mensen met kinderen minder vaak remigreren, zou dit bijvoorbeeld een onderschatting van de (positieve of negatieve) nettobijdrage kunnen geven.

De sterftekansen zijn gebaseerd op gegevens uit CBS-statline. In het standaardscenario is uitgegaan van de CBS-sterftekansen tot 2060. De sterftekansen zijn vervolgens vanaf 2060 constant gehouden. In het gebruikte model worden leeftijden van 0 tot 100 jaar gebruikt. De facto is de sterftekans voor 99-jarigen dus 100% (er zijn geen 100-jarigen in het model dus in het model verdwijnen alle 99-jarigen na één jaar volledig uit het model). Dat is niet realistisch, want een groeiende groep mensen wordt 100 jaar of ouder. Op basis van CBS-statline data is daarom geschat dat het aantal 100-plussers 2,4 maal het aantal 99-jarigen is. Met deze factor worden de kosten en baten van 99-jarigen verzwaaard om de kosten voor 100-plussers mee te kunnen nemen in de berekening. Hierbij is uitgegaan van een effectieve discountfactor van 1% en de aanname dat de kosten en baten voor 99-jarigen gelijk zijn aan de kosten en baten voor 100-plussers. Het gaat om een relatief klein deel van het totaalbedrag dus deze benadering volstaat.

Om te onderzoeken wat het effect is als de sterftekansen niet vanaf 2060 constant worden gehouden, is op basis van de CBS-prognoses voor sterftekansen voor 2017-2060 met SPSS per leeftijd een exponentiële curve gefit welke is geëxtrapoleerd naar de jaren vanaf 2060. Daarbij zijn voor hogere leeftijden de data voor de eerste jaren niet meegenomen voor zeventigers, de eerste 5 jaar, voor tachtigers de eerste 10 jaar en voor negentigers de eerste 15 jaar.¹²⁷ De nettobijdrage wordt negatiever als ook na 2060 de sterftekansen blijven dalen. Het verschil ligt in de orde van grootte van € 10.000.

Voor de verschillende onderwijsniveaus is rekening gehouden met verschillen in sterfte op basis van CBS-statline data.¹²⁸ De sterftekansen zijn lager voor mensen met een hoger opleidingsniveau en *vice versa*. De verhouding in sterftekansen per leeftijd(sgroep) tussen het onderwijsniveau op de SOI 3-delig (laag, middelbaar en hoog) is hierbij als uitgangspunt genomen. Als referentiegroep is hiervoor het middelbaar onderwijsniveau gebruikt (havo, vwo en mbo 2, 3 en 4). Omdat zeker voor kinderen en jongeren de sterftekansen in de buurt zitten van de kleinste eenheid, te weten 0,0001, waardoor afwijkingen door afronden een grote rol kunnen spelen, zijn in de verhoudingsgetallen de noemers/tellers op zodanige wijze verhoogd/verlaagd met 0,00005 dat een effect van afronden uitgesloten is. De verhoudingsgetallen tussen hoog/laag scholingsniveau enerzijds en middelbaar scholingsniveau anderzijds zijn op 0 gesteld als ze voor hoog/laag scholingsniveau boven/onder 1 waren. Tot slot zijn de uitkomsten gesmootherd.¹²⁹

Het effect op de nettobijdrage over de levensloop van differentiëren van sterftekansen naar opleidingsniveau wordt groter naarmate het opleidingsniveau lager is. Ten opzichte van deze referentiegroep is het verschil het grootst voor mensen met basisschool als hoogste opleiding, met een nettobijdrage over de levensloop die circa € 18.000 hoger is als rekening wordt gehouden met sterftekansen. Voor mensen met vmbo b/k, mbo1 resp. vmbo g/t, havo-, vwo-onderbouw als hoogste opleiding is de nettobijdrage € 17.000 resp. € 14.000 hoger als wordt gedifferentieerd naar opleidingsniveau. Voor personen met hbo-, wo-bachelor als hoogst behaalde opleiding levert rekening houden met verschillen

¹²⁷ Om reden dat deze afwijken van het verder vrijwel perfecte log-lineaire model.

¹²⁸ <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/83780NED/table?dl=3CF9C>

¹²⁹ Door vijfmaal achtereenvolgens toepassen van de functie $(f(L - 1) + 2 \times f(L) + f(L + 1))/4$ voor leeftijden $0 < L < 99$ waarin $f(L)$ de te smoothen functie is.

in sterftetekansen juist een nettobijdrage op die € 8.000 lager is. Bij het niveau hbo-, wo-master, doctor is het verschil verwaarloosbaar.

Er is geen rekening gehouden met verschillen in sterftetekansen naar motief of herkomst, omdat hiervoor geen geschikte data gevonden konden worden. Er is overwogen om verschillen in sterftetekansen naar inkomen als proxy te gebruiken, maar daarvan is afgezien omdat niet helder was hoe goed deze benadering zou zijn. Voorgaande resultaten geven echter wel een indicatie van de te verwachten verschillen die vermoedelijk in de orde van grootte van € 10.000 zullen liggen.

7.2 Migratie

In deze paragraaf wordt de operationalisering van de immigratieprofielen en remigratiekansen besproken. Omdat de immigratieprofielen nogal bepalend zijn voor de nettobijdrage (over de levensloop) is er zoveel mogelijk voor elke groep afzonderlijk een immigratieprofiel gemaakt. In totaal zijn er voor 139 verschillende groepen een immigratieprofiel gemaakt. Dat betreft herkomstgroepen, groepen ingedeeld naar migratiemotief (asiel, arbeid, studie, enzovoort) en combinaties van beiden. Remigratiekansen zijn eveneens zeer bepalend voor de nettobijdrage en daarom ook zoveel als mogelijk voor elke groep afzonderlijk bepaald. In totaal zijn er voor 76 verschillende groepen 407 verschillende remigratieprofielen berekend voor verschillende intervallen van entreeleeftijden.

Voor de operationalisering zijn op basis van GBAMIGRATIEBUS 2017 alle migratiebewegingen vanaf 1995 genomen voor de eerste generatie (regulier en overige op- en afvoer). Bij meerdere migratiebewegingen van dezelfde persoon per jaar zijn in- en uitgaande bewegingen die elkaar neutraliseren verwijderd. Voor de immigratieprofielen zijn de immigratiebewegingen geaggregeerd naar leeftijd ultimo 2016 en de betreffende herkomstgroep (regio en/of migratiemotief). Voor de remigratiekansen zijn de emigratiebewegingen geaggregeerd naar entreeleeftijd, verblijfsduur en de betreffende herkomstgroep (regio en/of migratiemotief) en gedeeld door het aantal op 1 januari van het betreffende jaar in Nederland verblijvende personen van de betreffende herkomstgroep.

Bij het bepalen van de remigratiekansen zijn per groep meerdere profielen gemaakt omdat de remigratiekansen sterk afhangen van de entreeleeftijd. In beginsel zijn voor entreeleeftijden tot 60 jaar voor de leeftijdsgroepen 0 tot 20 jaar, 20 tot 30 jaar, 30 tot 40 jaar en 40 tot 60 jaar en 60 tot 70 jaar aparte remigratieprofielen gemaakt, voor motief, regio (maximale verfijning tot de 18-deling) en de combinatie motief en regio (maximale verfijning tot de 12-deling). Indien de data dit toelieten zijn er ook voor de tussenliggende 10-jaars (entree)leeftijdsgroepen remigratieprofielen gemaakt. In enkele gevallen zijn twee van de vernoemde leeftijdsgroepen juist samengevoegd. Ook is voor de hogere leeftijden bij de combinatie motief en regio vaak beperkt tot alleen het onderscheid westers en niet-westers. Voor de leeftijdsgroep 70 tot 80 jaar is alleen uitgesplitst naar motief en de CBS 12-deling van herkomstregio's. Voor de leeftijdsgroep 80 tot 90 jaar is alleen uitgesplitst naar westers en niet-westers. Voor de leeftijdsgroep 90 tot 100 jaar is alleen een remigratieprofiel gemaakt voor alle eerste generatie immigranten samen. In totaal zijn zo 407 verschillende remigratieprofielen gemaakt.

De remigratieprofielen zijn voor zover de data dat toelieten gemaakt tot en met 23 verblijfsjaren. Van 23 jaar tot 50 jaar zijn de profielen geëxtrapoleerd. Dat is als volgt gedaan. Het verloop van de remigratiekansen naar verblijfsjaar is bij benadering exponentieel (met een negatieve exponent; de kansen nemen af). De factor van het verval is geschat voor acht 10-jaars entreeleeftijdsgroepen voor de

remigratiekansen van alle eerste generatie immigranten gezamenlijk.¹³⁰ Dit is gedaan door het verval te berekenen voor 11 verschillende 10-jaars perioden (1-10 verblijfsjaren, 2-11 verblijfsjaren, ... , 11-20 verblijfsjaren).¹³¹ Vervolgens is het gemiddelde hiervan genomen en is dit gebruikt om alle 407 profielen (voor zover mogelijk en van toepassing) te extrapoleren tot 50 jaar.

Uiteraard is zover vooruit extrapoleren met onzekerheden omgeven.¹³² Om een betere schatting te kunnen geven voor de remigratiekansen na 23 verblijfsjaren zijn er ook drie profielen geschat voor de lange termijn remigratiekansen; één voor westerse immigranten, één voor niet-westerse immigranten en één voor alle immigranten gezamenlijk. Dit is gedaan door voor deze drie groepen afzonderlijk per verblijfsduur zowel het aantal emigranten als ook het aantal op 1 januari van het betreffende jaar in Nederland verblijvende personen te bepalen en vervolgens de eerste grootheid te delen door de tweede grootheid. Voor verblijfsduren tot 24 jaar zijn de op empirische gegevens gebaseerde profielen gebruikt. Voor de verblijfsduren van 24 tot 30 jaar de extrapolaties. Voor verblijfsduren van 30 tot 50 jaar is het gewogen gemiddelde genomen van de extrapolaties en het van toepassing zijnde lange termijn remigratieprofiel. Er is gewogen met de functie $(\text{verblijfsduur} - 30) / 20$ voor een geleidelijke overgang. Voor verblijfsduren vanaf 50 jaar is alleen het van toepassing zijnde lange termijn remigratieprofiel gebruikt, dat wil zeggen het profiel voor eerste generatie immigranten, al dan niet uitgesplitst naar de herkomstregio westers of niet-westers.

7.3 Vergrijzing

Voor de vergrijzings simulatie is een demografisch model gebouwd gebaseerd op leeftijdsopbouw van de Nederlandse bevolking op 1 januari 2019 gebaseerd op CBS-Statline data. Dit is aangevuld met gegevens over de leeftijdsopbouw van de derde generatie op 1 januari 2016 die is gebaseerd op CBS-microdata. Hierbij is uitgegaan van de CBS-definities van personen met een Nederlandse achtergrond en een eerste, tweede, dan wel derde generatie migratieachtergrond.

Voor de operationalisering van geboorte, sterfte en migratie is in beginsel uitgegaan van de operationalisering in de voorgaande twee paragrafen. De verdeling over entreeleeftijden – het immigratieprofiel – is gelijk gesteld aan die van de gemiddelde immigrant. Hetzelfde is gedaan voor de remigratiekansen. Van de tweede generatie wordt – net als bij de berekening van de nettobijdrage – verondersteld dat ze tot 18 jaar altijd met hun ouders mee meegaan in geval die remigreren. Voor de derde en volgende generaties is verondersteld dat ze niet migreren. Hetzelfde is aangenomen voor de autochtone groep.

Het kindertal is voor alle groepen constant op 1,7 kind per vrouw gesteld. Om het model eenvoudig te houden is verder verondersteld dat er voor wat betreft de migratie vanaf 2020 geen gemengde ouderparen zijn.¹³³ Daarnaast sluit deze handelwijze aan bij de rekenexercitie van de Verenigde Naties –

¹³⁰ Voor de groepen 80 tot 90 en 90 tot 100 jaar is extrapolatie niet nodig, want in het model wordt gerekend met leeftijden tot 100 jaar en de remigratie kansen voor de eerste 23 jaar zijn reeds bekend (voor zover de data dat toelaat).

¹³¹ Voor de periode na 20 verblijfsjaren was er weinig data en daarom is dat niet gebruikt voor de extrapolatie.

¹³² Hierbij moet wel opgemerkt worden dat de remigratiekansen na 23 verblijfsjaren vaak al erg klein zijn, want de meeste remigratie vindt de eerste 10 jaar plaats. Het effect van extrapolatiefouten zal daardoor meestal gering zijn, mede omdat bedragen door het contant maken minder zwaar wegen naarmate ze verder in de toekomst liggen.

¹³³ De laatste aanname is puur gemaakt om het rekenwerk te vereenvoudigen en hangt niet samen met voorkeuren over de wenselijkheid of onwenselijkheid van gemengde relaties.

waaruit wordt geciteerd in §10.2 van het huidige rapport en welke in §10.3 wordt gerepliceerd – waarin wordt gesproken van ‘immigranten en hun nakomelingen’, met andere woorden, waarin reeds ingezetenen en nieuwe immigranten en hun nakomelingen gescheiden categorieën zijn. Ieder kind heeft vanaf dat moment dus ofwel twee autochtone ouders en is dan zelf autochtoon, ofwel twee eerste generatie ouders en is dan zelf tweede generatie, ofwel twee ouders die nakomelingen zijn van twee tweede generatie ouders en is dan zelf derde generatie. De vierde en volgende generaties zijn opgeteld bij de autochtonen. Omdat voor alle groepen dezelfde vruchtbaarheid (1,7 kind per vrouw) is verondersteld maakt dit weinig verschil voor de totale omvang van de migratie en het benodigde aantal immigranten om de vergrijzing constant te houden, wat het doel van de rekenexercitie is. Door deze vereenvoudigingen zal het aantal personen met een tweede en derde generatie migratieachtergrond lager uitvallen dan bij gemengde relaties het geval zal zijn.¹³⁴

7.4 Demografie

In deze paragraaf worden enkele demografische passages toegelicht uit Hoofdstuk 1 (bevolkingsgroei in de 21^e eeuw bij verschillende migratiesaldi), Hoofdstuk 2 (schatting huidige bevolkingsomvang indien er geen immigratie was geweest) en Hoofdstuk 11 (schatting additionele bevolkingsgroei bij het toelaten van de zogenaamde *high potentials* om de kosten van vergrijzing op te vangen).

Het demografische model uit de vorige paragraaf is ook gebruikt voor de schattingen voor bevolkingsgroei die zijn gegeven in §1.1. In het model levert een jaarlijks migratiesaldo dat 10 duizend hoger ligt in 2100 1,19 miljoen extra inwoners op. Vergelijk de in deze paragraaf gegeven inwoneraantallen en de bevolkingstoename per 10 duizend toename van het jaarlijks migratiesaldo met de Medium Variant en het Zero-Migration scenario van de Verenigde Naties.¹³⁵ In de Medium Variant is het tegen de populatieomvang gewogen gemiddelde migratiesaldo 1,19 immigranten per 1000 inwoners, hetgeen neerkomt op een migratiesaldo van 20,0 duizend personen per jaar.¹³⁶ Het verschil in bevolkingsomvang in 2100 tussen de Medium Variant en het Zero-Migration scenario bedraagt 2,37 miljoen personen en dat komt per 10 duizend migratiesaldo eveneens neer op 1,19 miljoen extra inwoners in 2100.

Bij de schatting van het inwoneraantal zonder immigratie (§2.2) zijn de eerste generatie en de tweede generatie met twee in het buitenland geboren ouders niet meegeteld. De tweede generatie met één in het buitenland geboren ouder is met gewicht 0,5 meegeteld. De derde generatie is gebaseerd op CBS maatwerkdata.¹³⁷ De bevolking van de derde generatie is eerst in mindering gebracht op de bevolking met Nederlandse achtergrond. Vervolgens is de derde generatie meegewogen door gebruikt

¹³⁴ Concreet geldt voor ouders met een eerste of tweede generatie migratieachtergrond volgens CBS-definities het volgende: als uit twee personen met een Nederlandse achtergrond in Nederland één kind geboren wordt en uit twee personen met een migratieachtergrond in Nederland één kind geboren wordt, dan is er één kind met een Nederlandse achtergrond en één kind met een (tweede of derde generatie) migratieachtergrond. Maar als vier personen – twee met een migratieachtergrond en twee met een Nederlandse achtergrond – twee gemengde paren vormen die elk één kind krijgen, dan zijn er twee kinderen met een migratieachtergrond. Gemengde relaties leiden volgens de CBS-definities dus tot minder personen met een Nederlandse achtergrond en meer personen met een migratieachtergrond.

¹³⁵ United Nations, Department of Economic and Social Affairs, Population Division (2019).

¹³⁶ United Nations. (2020, 15 december) *Net migration rate (per 1,000 population)*. Opgehaald van: <http://data.un.org/Data.aspx?q=migration&d=PopDiv&f=variableID:85;crID:528;timeID:104,110,116,122,128,134,140,146,152,158,164,170,176,74,80,86,92,98&c=2,4,6,7&s=crEngNameOrderBy:asc,timeEngNameOrderBy:desc,varEngNameOrderBy:asc&v=1>

¹³⁷ CBS-maatwerk. (2020, 15 december) *Personen met ouders van de tweede generatie*. Opgehaald van: <https://www.cbs.nl/nl-nl/maatwerk/2020/46/personen-met-ouders-van-tweede-generatie-1-januari-2020>

te maken van combinatoriek en de observatie dat circa $\frac{1}{5}$ van de bevolking met een westerse tweede generatie migratieachtergrond twee in het buitenland geboren ouders heeft en circa $\frac{2}{5}$ van de bevolking met een niet-westerse tweede generatie migratieachtergrond twee in het buitenland geboren ouders.¹³⁸

Bij de schatting van de additionele bevolkingsgroei bij het toelaten van de zogenaamde *high potentials* om de kosten van vergrijzing op te vangen (§11.3), is uitgegaan van de bedragen voor 2016, het peiljaar van de huidige studie. Om vergelijking te vergemakkelijken is uitgegaan van een even lange periode als in de rekenexercitie op basis van *Immigration and the Dutch economy* die in de hoofdtekst wordt gegeven (60 jaar), vandaar de berekening van 2020 tot 2080. Omdat alle bedragen in het huidige rapport zijn uitgedrukt in euro's van 2016, wordt uitgegaan van het bbp in 2016, dat 708 miljard euro bedroeg.¹³⁹ Afgerond tot 700 miljard euro levert dit op dat er 175.000 *high potentials* nodig zijn die elk een nettobijdrage aan de schatkist leveren van een ton. Simulatie met hetzelfde demografische model dat in de vorige paragraaf uiteengezet is, levert op dat dit leidt tot een additionele bevolkingsgroei van 7,2 miljoen personen. Zou men het bbp van 2016 niet naar beneden afronden, dan zou het gaan om 7,3 miljoen personen.

¹³⁸ CBS-Statline. (2020, 15 december) *Bevolking; geslacht, leeftijd, generatie en migratieachtergrond, 1 januari*. Opgehaald van: <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/37325/table?dl=46C51>

¹³⁹ CBS-statline, *Opbouw binnenlands product (bbp); nationale rekeningen*, opgehaald 12-02-2021 van: <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/84087NED/table?dl=4D597>

8 Verankering in CPB-studies

8.1 Verankering in CPB-vergrijzingsstudie *Minder zorg om morgen* (2014)

De methode van generatierekening die in het huidige rapport wordt gebruikt is dezelfde methode als gebruikt wordt in de CPB-vergrijzingsstudies. In de CPB-vergrijzingsstudies draait het om het bepalen van het zogenaamde “houdbaarheidssaldo” voor de overheidsfinanciën. Dit houdbaarheidssaldo is – met een aantal correcties¹⁴⁰ – te herleiden tot het begrip “netto profijt” van de overheid. Het begrip netto profijt is spiegelbeeldig aan het begrip nettobijdrage dat centraal staat in de huidige studie. Een ingezetene die bijvoorbeeld voor +€ 50.000 netto profijt heeft van de overheid, heeft een (negatieve) nettobijdrage van –€ 50.000, d.w.z. hetzelfde bedrag met het tegengestelde teken.

De berekeningen in het huidige rapport zijn gebaseerd op het basisjaar 2016. Om de ontwikkeling van beleid en economie vanaf het jaar 2016 mee te nemen in de berekening is gebruikt gemaakt van aannamen die ten grondslag lagen aan de berekeningen van de CPB-vergrijzingsstudie *Minder zorg om vergrijzing*¹⁴¹ uit 2014. Voor de korte termijn ging het in deze CPB-vergrijzingsstudie om aannamen omtrent de ontwikkeling van de conjunctuur e.d. Voor de lange termijn ging het om de basisassumptie van een reële discontovoet van 3% en een productiviteitsgroei van 1,5% per jaar.¹⁴² Op basis van die aannamen heeft het CPB de ontwikkeling (in nominale bedragen) over de periode 2016-2060 van 23 kosten- en batenposten bepaald, die gebruikt zijn in de generatierekeningen in deze vergrijzingsstudie. In het huidige rapport is uitgegaan van deze 23 kosten- en batenposten en hun door het CPB veronderstelde ontwikkeling over de periode 2016-2060.

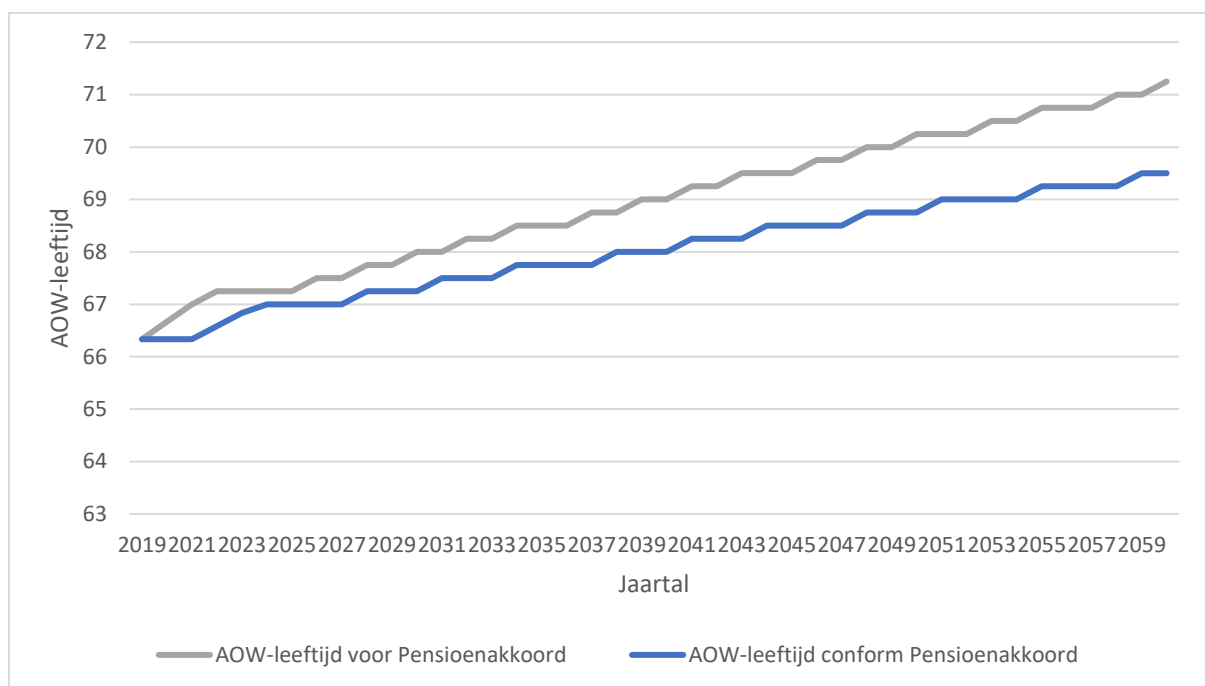
De ontwikkeling over de tijd van de 23 kosten- en batenposten van het CPB is overgenomen van de data uit de CPB-vergrijzingsstudie *Minder zorg om morgen*. Deze dataset – in de rest van dit document kortweg aangeduid als CPB2014-dataset – bestaat uit $6 \times 23 \times 100$ nominale bedragen. Concreet gaat het om leeftijdsprofielen (0-99 jaar) voor elk van de 6 jaren 2016, 2021, 2030, 2040, 2050 en 2060 en elk van de 23 posten in Tabel 5.1.

De ontwikkeling van deze bedragen over de tijd is gebaseerd op een aantal veronderstellingen van het CPB. Veel van deze veronderstellingen worden zondermeer overgenomen, omdat ze uiteindelijk op lange termijn uitkomen op bedragen die passen bij het lange termijn groeipad van de economie dat het CPB tijdens het schrijven van de CPB-vergrijzingsstudie *Minder zorg om morgen* voorzag. Voor twee veronderstellingen zijn echter aanpassingen gedaan. De AOW-leeftijd loopt door het zogenaamde pensioenakkoord minder snel op, zie Figuur 8.1. De aanpassing van de AOW-leeftijd heeft betrekking op bepaalde posten (zoals de AOW zelf, maar ook belastingen en uitkeringen) en op bepaalde leeftijden (vooral de leeftijden 65-71 jaar). Door gebruik te maken van de ontwikkeling over de tijd voor leeftijden die niet beïnvloed zijn door de aanpassing van de AOW-leeftijd is de aanpassing van de AOW leeftijd volledig uit de dataset verwijderd. In de resulterende dataset is de ontwikkeling over de tijd van de betreffende posten voor de *range* 64-72 jaar in lijn met de omliggende hogere en lagere leeftijden.

¹⁴⁰ Smid, B., H. ter Rele, S. Boeters, N. Draper, A. Nibbelink en B. Wouterse (2014), kader ‘EMU-saldo en netto profijt’, blz. 41

¹⁴¹ Smid, B., H. ter Rele, S. Boeters, N. Draper, A. Nibbelink en B. Wouterse (2014)

¹⁴² Dit betekent een nominale discontovoet van 5% en nominale productiviteitsgroei met 3,5% per jaar, want het CPB veronderstelt een inflatie van 2%, zie Smid, B., H. ter Rele, S. Boeters, N. Draper, A. Nibbelink en B. Wouterse (2014), blz. 27



Figuur 8.1 Ontwikkeling AOW-leeftijd, oorspronkelijk en volgens het pensioenakkoord. Bron: Excelwerkblad 8.1L uit de dataset bij de CPB (2019) Zorgen om morgen, CPB-vergrijzingsstudie, opgehaald 10-2-2021 van: <https://www.cpb.nl/sites/default/files/omnidownload/CPB-Vergrizingsstudie-2019-Data-figures.xlsx>

Daarnaast zijn er veronderstellingen met betrekking tot de productiviteitsgroei. Voor de lange termijn (vanaf 2022) gaat het om een veronderstelde productiviteitsgroei van 1,5% reëel en 3,5% nominaal. Voor de korte termijn (2016-2021) gaat het om conjunctuur gerelateerde ontwikkeling. Uit de technische bijlage¹⁴³ en bijbehorende Excelbestanden (onder de werkbladen 'Aannamen' bij *Minder zorg om morgen* is af te leiden dat er voor 2016-2021 geen afwijkende aannamen zijn gedaan.

De gerealiseerde dan wel geschatte nominale productiviteitsgroei in de CPB-raming uit november 2020¹⁴⁴ ligt overigens lager (zie Tabel 8.1). Berekend over 2017-2021 geeft dit een schatting van 2,1% inflatie en 0,3% arbeidsproductiviteitsgroei (samen 2,4%).

Tabel 8.1 Waargenomen inflatie en productiviteitsgroei.

		2016	2017	2018	2019	2020	2021
Prijs bruto bbp (% per jaar)	Bijlage 1	0,5	1,3	2,4	3,0	2,2	1,5
Arbeidsproductiviteit bedr. (per uur)	Bijlage 1	-0,2	0,6	-0,1	-0,5	-0,2	1,8
Gewerkte uren per werkende	Bijlage 10	1437	1437	1436	1440	1395	1435

¹⁴³ <https://www.cpb.nl/sites/default/files/publicaties/bijlagen/dp170-technische-bijlage.pdf>

¹⁴⁴ CPB Verzamelde bijlagen bij de raming november 2020, opgehaald 15-12-2020 van: <https://www.cpb.nl/sites/default/files/omnidownload/verzamelde-bijlagen-raming-nov-2020-plus-mlt.xlsx>

8.2 IJking op CPB-vergrijzingsstudie *Zorgen om morgen* (2019)

Het onderzoeksvoorstel voor het huidige rapport dateert van medio 2018, maar voor publicatie verscheen er een nieuwe CPB-vergrijzingsstudie *Zorgen om morgen* (2019). Hierom zijn enkele aanpassingen gedaan. De belangrijkste aanpassingen zijn een verlaging van de veronderstelde productiviteitsgroei en discontovoet met 0,5%.

Om verder zoveel mogelijk te ijken op berekeningen van het CPB is gebruik gemaakt van de dataset bij de CPB-vergrijzingsstudie *Zorgen om morgen* uit 2019, vanaf nu kortweg aangeduid als CPB2019-dataset. In *Zorgen om morgen* worden de verschillen beschreven¹⁴⁵ met de eerdere CPB-vergrijzingsstudie *Minder zorg om vergrijzing* uit 2014, te weten:

- 1) Het CPB heeft de reële discontovoet verlaagd van 3,0% naar 2,5%;¹⁴⁶
- 2) De veronderstelde productiviteitsgroei is door het CPB verlaagd van 1,5% naar 1%;¹⁴⁷
- 3) Door hogere zorgkosten is het houdbaarheidssaldo t.o.v. 2014 verslechterd met 0,6% bbp;¹⁴⁸
- 4) Het pensioenakkoord verslechtert het houdbaarheidssaldo t.o.v. 2014 met 0,4% bbp;¹⁴⁹
- 5) In totaal is het houdbaarheidssaldo t.o.v. 2014 met 2,0% bbp verslechterd, waarvan 0,9% door beleid van Rutte III, anders dan het pensioenakkoord en 0,2% bbp door Rutte II.¹⁵⁰
- 6) Het arbeidsaanbod stijgt over de lange termijn.

Een volledig overzicht wordt gegeven in Tabel 2.2. uit *Zorgen om morgen*:

Tabel 2.2 Verschillen in het houdbaarheidssaldo tussen de huidige en de vorige vergrijzingsstudie

	% bbp
Vorige vergrijzingsstudie 2014	+0,4
Ontwikkelingen tijdens Rutte II	-0,2
Beleid Rutte III - regeerakkoord	-0,6
Beleid Rutte III – lastenverlichting en Pensioenakkoord	-0,7
Aanpassingen zorguitgaven	-0,6
Arbeidsaanbod	+0,5
Overig, waaronder revisie van de Nationale rekeningen	-0,4
Lagere discontovoet en productiviteitsgroei	0,0
Modelwijzigingen Gamma	0,0
Huidige vergrijzingsstudie 2019	-1,6

Uitgangspunt bij de huidige studie is om het CPB zoveel mogelijk te volgen. De verlaging van de reële discontovoet met 0,5%-punt naar 2,5% en de verlaging van de productiviteitsgroei met 0,5%-punt naar

¹⁴⁵ Zie met name Tabel 2.2. in *Zorgen om morgen* en de bijbehorende toelichting.

¹⁴⁶ Zie Adema, Y., & I. van Tilburg (2019), blz. 70-72

¹⁴⁷ Zie Adema, Y., & I. van Tilburg (2019), blz. 42

¹⁴⁸ Zie Adema, Y., & I. van Tilburg (2019), blz. 17-18

¹⁴⁹ Zie Adema, Y., & I. van Tilburg (2019), blz. 17

¹⁵⁰ Zie Adema, Y., & I. van Tilburg (2019), blz. 16

1% zijn daarom overgenomen. Ook de effecten van het pensioenakkoord zijn in het huidige rapport verwerkt (zie vorige paragraaf). Dit vormt samen met de CPB2014-dataset het basisscenario. Omdat *Minder zorg om vergrijzing* uit 2014 een houdbaarheidssaldo van +0,4% bbp heeft en het pensioenakkoord een verslechtering van 0,4% bbp oplevert heeft dit basisscenario een houdbaarheidssaldo van ongeveer 0% bbp.

De overige beleidswijzigingen kunnen echter niet zonder meer overgenomen worden omdat de datasets van beide CPB-vergrijzingsstudies verschillen (onder meer is de CPB2019-dataset geaggregeerd tot 8 deelposten i.p.v. 23) en omdat te weinig bekend is van de precieze mechanismen wat tot inconsistenties kan leiden als beide datasets worden gecombineerd.

Wel is de CPB2019-dataset¹⁵¹ gebruikt om de berekening in het huidige rapport te controleren. Volgens de documentatie bij *Zorgen om morgen* heeft de dataset betrekking op het jaar 2020. Daarom is bij de vergelijking uitgegaan van de nominale bedragen voor 2020 uit de CPB2014-dataset. Hierbij is het profiel voor langdurige zorg (Wlz, Wmo e.d.) – d.w.z. het zogenaamde verschilprofiel uit §5.5 – vervangen, door het verschilprofiel dat ontstaat door sommering van de profielen Wmo, Wlz en overige zorg in Excelwerkblad 2.4 van de 2019-dataset¹⁵². Dit levert op lange termijn een 0,3% bbp verslechtering op, namelijk de helft¹⁵³ van het totale verschil van 0,6% bbp in punt 3) in bovenstaande opsomming. Echter, over de periode 2016-2020 is het verschil klein (0,02% bbp) want de grootte van het effect loopt op met de grijze druk. Iets dergelijks geldt voor het arbeidsaanbod, dat is ook iets dat pas op de lange termijn effect heeft, en niet al in 2020.¹⁵⁴ Van de revisie nationale rekeningen is een aanzienlijk deel in het huidige rapport mogelijk al verwerkt door gebruik van microdata, al is dat niet zeker. Derhalve resteert tenminste een verslechtering van 0,2% bbp (Rutte II) plus 0,7% – 0,4% = 0,3% bbp (Rutte III minus pensioenakkoord) en het regeerakkoord Rutte III van minstens 0,6% bbp, want het korte termijn effect is groter dan het lange termijn effect.¹⁵⁵ Totaal is dat minimaal 1,1% bbp. Omdat niet duidelijk is waarop dit precies betrekking heeft, is dit gerealiseerd door optelling van een constante bij de post Openbaar bestuur die na weging tegen de bevolking op 1 januari 2020 een macrobedrag oplevert ter grootte van 1,1% bbp van het voor 2020 voorspelde bbp¹⁵⁶, waarbij is uitgegaan van 3,0% groei t.o.v. 2019¹⁵⁷ en een bbp voor 2020 van 859 miljard euro. Dit levert vrijwel dezelfde macrobedragen. Bij optelling van 1,13% bbp (i.p.v. 1,1% bbp) is het verschil tussen het nettobijdrage profiel van het huidige rapport voor 2020 en van het profiel op basis van de CPB2019-dataset, gewogen tegen de bevolking van 2020, zelfs nihil. Dat geeft aan dat met de veronderstelde aanpassingen de resultaten in het huidige rapport maar weinig afwijken van de resultaten in de CPB-vergrijzingsstudie uit 2019.

¹⁵¹ <https://www.cpb.nl/sites/default/files/omnidownload/CPB-Vergrijzingsstudie-2019-Data-figuren.xlsx>

¹⁵² <https://www.cpb.nl/sites/default/files/omnidownload/CPB-Vergrijzingsstudie-2019-Data-figuren.xlsx>

¹⁵³ Zie Adema, Y., & I. van Tilburg (2019), blz. 17

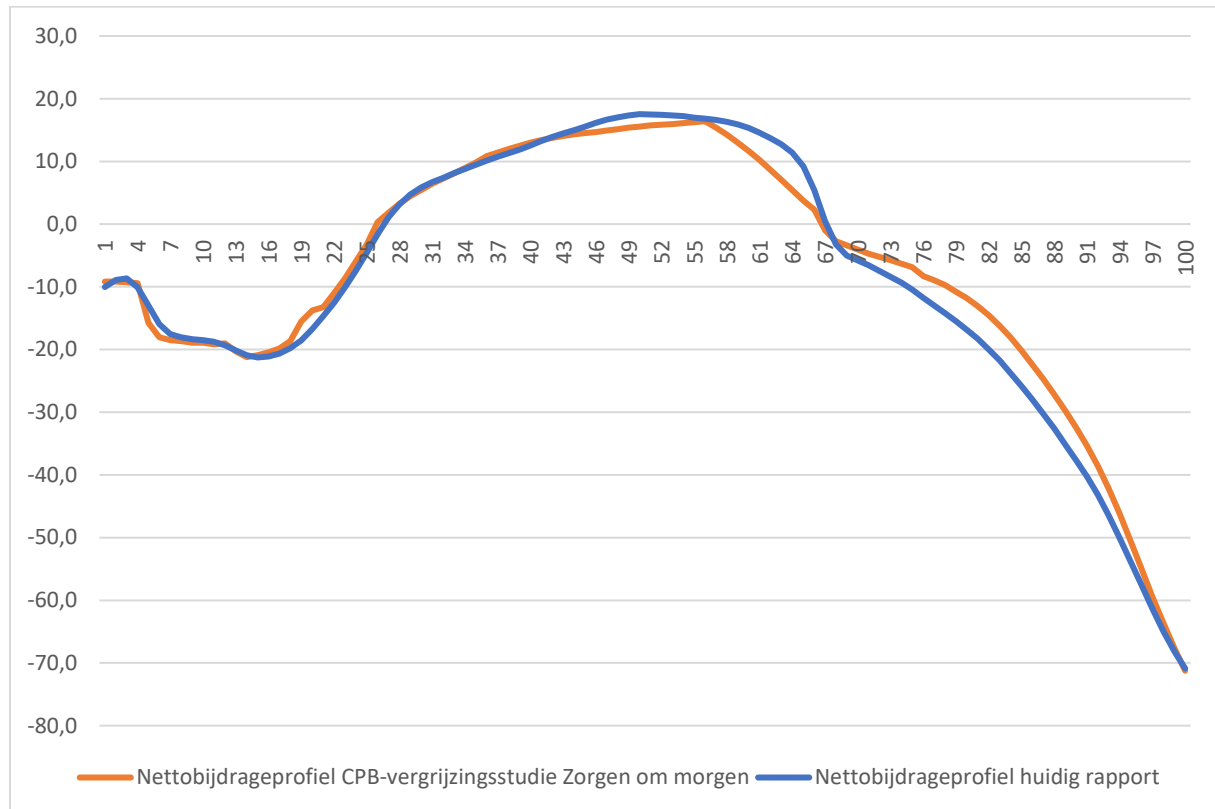
¹⁵⁴ Figuur 4.1 blz. 7 <https://www.cpb.nl/sites/default/files/omnidownload/CPB-Achtergronddocument-dec2019-Arbeidsparticipatie-en-gewerkte-uren-tot-en-met-2060.pdf>

¹⁵⁵ "Direct effect, zie blz 22: Het negatieve houdbaarheidseffect wordt volledig gerealiseerd binnen de kabinetsperiode. Het beleidspakket heeft een effect op het EMU-saldo in 2021 van -1,2% bbp. Na 2021 volgt een positief effect van 0,6% bbp." https://www.cpb.nl/sites/default/files/omnidownload/CPB-Notitie-4okt2017-Analyse-economische-en-budgettaire-effecten-Regeerakkoord_0.pdf

¹⁵⁶ Door Corona zal het bbp voor 2020 lager uitvallen, maar hier wordt vergeleken met de voorspelling van het CPB en niet met de werkelijkheid.

¹⁵⁷ <https://www.cpb.nl/sites/default/files/omnidownload/CPB-Policy-Brief-Decemberraming-2019-voorzicht-2020.pdf> NB: gepubliceerd 1 dag voor *Zorgen om morgen*

Het resulterende profiel voor 2020 wijkt qua verloop af van het profiel uit CPB2019, vooral doordat voor de huidige studie microdata is gebruikt, wat een gladder en natuurlijker verloop geeft. Daarnaast zijn bijvoorbeeld de vermogen-gerelateerde belastingen in het huidige rapport direct afgeleid van de opgebouwde (en na pensionering ook weer geleidelijk afgebouwde) pensioenvermogens, hetgeen hogere waarden geeft rond de AOW-leeftijd.



Figuur 8.2 Vergelijking van het nettobijdrageprofiel van het huidige rapport met het nettobijdrageprofiel uit de CPB-vergrijzingsstudie Zorgen om morgen uit 2019.

Omdat het netto leeftijdsprofiel voor 2020 van de huidige studie – na voornoemde correcties – gewogen tegen de bevolking van 2020 nagenoeg hetzelfde macrobedrag oplevert als het netto leeftijdsprofiel voor 2020 uit *Zorgen om morgen* (2019), is aangenomen dat een onzekere correctie over de waargenomen inflatie en productiviteitsgroei voor de jaren 2016-2021 (zie Tabel 8.1) geen betrouwbaardere resultaten oplevert dan de aanname dat de productiviteitsgroei voor alle jaren gelijk is. Daarom is voor alle jaren de productiviteitsgroei van 3,5% nominaal uit de CPB-vergrijzingsstudie *Minder zorg om morgen* (2014) omgezet naar een productiviteitsgroei van 1,0% reëel uit de CPB-vergrijzingsstudie *Zorgen om morgen* (2019). Dit is gedaan door alle nominale bedragen te vermenigvuldigen met de volgende correctiefactoren:

$$\left(\frac{1,010}{1,020 * 1,015} \right)^{J-2016} \quad \text{voor } 2016 \leq J \leq 2200$$

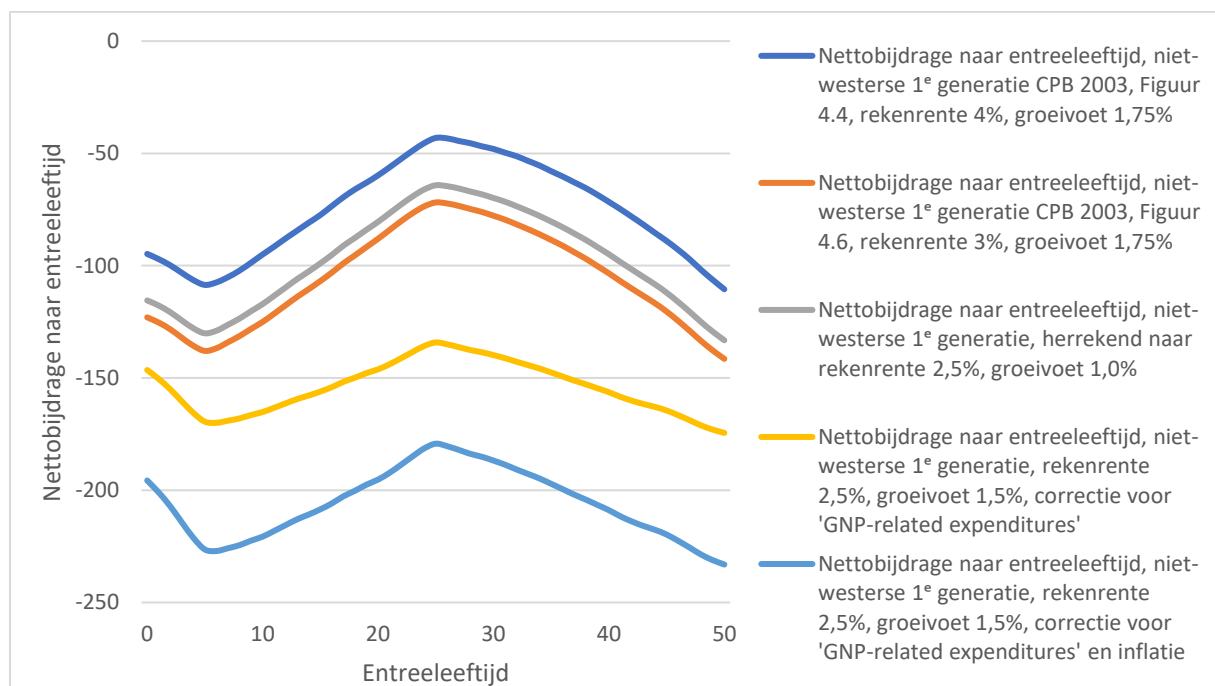
Dit levert een nieuwe dataset op met daarin alle veronderstellingen uit de CPB-vergrijzingsstudie *Minder zorg om morgen* uit 2014, aangevuld met de aanpassing van de AOW-leeftijd aan het pensioenakkoord en herrekent naar een reële productiviteitsgroei van 1%. Rekenen in reële termen maakt het overbodig om veronderstellingen te doen over toekomstige inflatie. Deze nieuwe dataset voor de

ontwikkeling over de tijd van de 23 posten uit Tabel 5.1 wordt gebruikt voor het basisscenario van het huidige rapport. De discontovoet is in het basisscenario 2,5% reëel, in navolging van de CPB- vergrijzingsstudie *Zorgen om morgen*.

8.3 Vergelijking met CPB-rapport *Immigration and the Dutch Economy* (2003)

De resultaten van het CPB-rapport en het huidige rapport kunnen niet zonder meer vergeleken worden, vanwege de in tussentijd opgetreden inflatie, een andere berekeningsmethodiek en een andere discontovoet en groeivoet.

Allereerst de discontovoet en groeivoet. Het CPB rekende in 2003 met een groeivoet van 1,75% reëel en een discontovoet van 4,0% reëel.¹⁵⁸ Dit levert een effectieve discontovoet op van 2,25%. In het huidige rapport is – in navolging van de huidige CPB-standaard – gerekend met een groeivoet van 1,0% reëel en een discontovoet van 2,5% reëel. Dit levert een effectieve discontovoet op van 1,5%. Het verschil in discontovoet is opgelost door uit *Figure 4.4*¹⁵⁹ en *Figure 4.6*¹⁶⁰ uit het CPB-rapport de informatie te extraheren voor de eerste generatie niet-westerse immigranten en dit om te rekenen naar de situatie in het huidige rapport. In voornoemde *Figure 4.4* wordt gerekend met een discontovoet van 4,0% en in voornoemde *Figure 4.6* met een discontovoet van 3% (de donkerblauwe en oranje lijn Figuur 8.3) en dat levert een effectieve discontovoet op van 1,25%. Deze informatie is omgerekend naar de groeivoet en discontovoet in het huidige rapport (de grijze lijn in Figuur 8.3).



Figuur 8.3 Vertaling van de resultaten voor niet-westerse eerste generatie immigranten uit Roodenburg et al. naar het huidige rapport. Vervolgens is het in §4.2 beschreven methodologische verschil van het meten van publieke goederen gecorrigeerd. Daartoe is op basis van het Persoonlijk Primair Inkomen (PPI) per leeftijdsjaar (uit het CBS-microdatabestand INPATAB, variabele INPPERPRIM, verslagjaar 2016) het verschil tussen beide methoden geschat. Er is gekozen voor het PPI omdat dit goed

¹⁵⁸ Roodenburg, H., R. Euwals & H. ter Rele (2003), blz. 68, voetnoot 12

¹⁵⁹ Roodenburg, H., R. Euwals & H. ter Rele (2003), blz. 70

¹⁶⁰ Roodenburg, H., R. Euwals & H. ter Rele (2003), blz. 74

de bijdrage aan het BNP weerspiegelt en net als in de nettobijdrageprofielen van het CPB (zie *Figure 4.3*¹⁶¹) vooral voor de leeftijden 15-65 jaar verschil uitmaakt. Op basis van de PPI per leeftijdjaar is een profiel *P* berekend voor de nettobijdrage naar entreeleeftijd voor entreeleeftijden tot en met 50 jaar. Bij de berekening van *P* is uitgegaan van de door het CPB in 2003 gebruikte discontovoet en groeivoet. De remigratiekansen zijn geëxtraheerd uit *Figure A.3.1*.¹⁶² Voor de sterftekansen is de tabelbevolking uit 2016 als proxy genomen. Vervolgens is het resulterende profiel *P* geïjkt met een discountfactor als stelpost en als ijkpunten de gegeven referentiewaarden voor ‘GDP-related expenditures’ voor de groep ‘non-Western’ in *Table 4.3*¹⁶³. De ‘GDP-related expenditures’ waren volgens *Table 4.2*¹⁶⁴ door het CPB reeds voor 52% meegenomen in de berekening. De resterende 48% is bijgevoegd door 12/13^e deel van profiel *P* af te trekken van de grijze lijn in *Figuur 8.3*. Het resultaat is de gele lijn in *Figuur 8.3*. Tot slot is gecorrigeerd voor inflatie, door vermenigvuldiging met een factor 1,336^{165 166} en dat levert de lichtblauwe lijn op in *Figuur 8.3*.

De gele lijn is vervolgens gewogen tegen het in huidige studie gebruikte immigratieprofiel en dat levert –€ 188.000 nettobijdrage op voor de leeftijden tot en met 50 jaar. Vervolgens is op basis van de verhouding tussen de betreffende leeftijdsgroepen voor niet-westerse immigranten zoals waargenomen in de huidige studie, de nettobijdrage voor de leeftijden boven 50 jaar geschat op –€ 13.000. Totaal levert dat een nettobijdrage op voor eerste generatie niet-westerse immigranten van circa –€ 199.000. Dat is één vijfde hoger dan het resultaat voor eerste generatie niet-westerse immigranten in het huidige rapport.

¹⁶¹ Roodenburg, H., R. Euwals & H. ter Rele (2003), blz. 68

¹⁶² Roodenburg, H., R. Euwals & H. ter Rele (2003), blz. 120

¹⁶³ Roodenburg, H., R. Euwals & H. ter Rele (2003), blz. 71

¹⁶⁴ Roodenburg, H., R. Euwals & H. ter Rele (2003), blz. 66

¹⁶⁵ Hierbij is voor het CPB-rapport uitgegaan van het basisjaar 2000, dat wordt genomen voor het BNP (Roodenburg, H., R. Euwals and H. ter Rele (2003), Table 3.3, voetnoot b, blz. 49) en verder expliciet genoemd wordt op blz. 117 (“base year”) en impliciet op blz. 32 “It remains to be seen whether these results still apply to the year 2000”. De beschrijvende statistieken zijn gebaseerd op peiljaar 2000 (*Figure 2.5, 2.6, 2.7 en 2.8*). De gebruikte datasets zijn van de jaren (voetnoot *Table 4.1*) 2001, 2000 en (voetnoot *Table 4.2*) 2002, 1999, 2001 en 2001. Ook bij de demografische projecties wordt 2000 aangehouden, zie o.a. *Table A.3.1*, het Huidige rapport heeft als peiljaar 2016. De consumentenprijzen index van het CBS laat over deze periode een groei zien van 33,6%, CBS-statline, *Consumentenprijzen; prijsindex 2015=100*, opgehaald 14-12-2020 van: <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/83131NED/table?dl=493F7> en de cumulatieve mutaties jaar op jaar over de periode 2000-2015 bedragen volgens de bijlagen bij het Centraal Economisch Plan (CEP) van het CPB 33,1%, zie CPB, *Verzamelde bijlagen CEP 2018*, opgehaald 15-12-2020 van: https://www.cpb.nl/sites/default/files/omnidownload/Verzamelde_bijlagen_CEP_2018.xlsx

¹⁶⁶ CBS-statline, *Consumentenprijzen; prijsindex 2015=100*, opgehaald 14-12-2020 van: <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/83131NED/table?dl=8943>

9 Belangrijke punten bij de interpretatie van de resultaten

9.1 Beschrijving is geen toetsing, samenhang is geen oorzakelijk verband

Het huidige onderzoek is sterk verkennend en beschrijvend van aard. Hoofdzakelijk wordt op basis van allerlei data de fiscale nettobijdrage over de levensloop beschreven van immigranten, uitgesplitst naar herkomstgroepering, migratiemotief en generatie. Tijdens het onderzoek werden allerlei samenhangen tussen variabelen opgemerkt en deze zijn vaak ook gerapporteerd. Ook dit is echter voornamelijk beschrijvend gebeurd, bedoeld als aanknopingspunten voor verder onderzoek. Incidenteel is de significantie van de correlaties gerapporteerd of is bij kleine N voor de volledigheid ook Spearman's rho gerapporteerd. Er is verder niet beoogd toetsend onderzoek te doen. Ook is niet ingezet op het aantonen van oorzakelijke verbanden.

9.2 Eén peiljaar: noodzaak van herhaling

Zoals opgemerkt in Box 4.1, in Hoofdstuk 4, bepalen we de verwachte netto fiscale bijdragen van immigranten over hun resterende verblijfsduur in Nederland (dus over hun toekomstige leeftijden, in toekomstige jaren) uit de bedragen die gelden voor een migrant van de relevante leeftijd in 2016. Zoals daar opgemerkt: 'een immigrant die in 2016 als 30 jarige binnenkomt, krijgt in 2026 de bedragen toegewezen die in 2016 voor een 40 jarige golden'. De berekening geldt dus voor de opbrengsten en lasten in 2016 van kenmerken die we kunnen waarnemen (zoals genoten onderwijs en herkomstregio) en voor de kenmerken van de immigrantenpopulatie die in 2016 aanwezig is en die we niet kunnen waarnemen (zoals ambitie, creativiteit etc.). We kunnen de bedragen bij toepassing op toekomstige jaren differentiëren naar de waarneembare kenmerken, maar niet naar de mogelijke verandering van de lasten en baten die met die kenmerken samenhangen (loonverschillen naar opleidingsniveau kunnen bijvoorbeeld veranderen). Over veranderingen die samenhangen met niet waargenomen kenmerken kunnen we per definitie niets zeggen. Essentieel is ook de veronderstelling dat de verschillen die we in 2016 waarnemen *tussen* de cohorten kunnen worden gehanteerd als verschillen *binnen* cohorten. Het verschil tussen immigranten, uit een bepaalde regio, met een bepaalde opleiding, die vijf jaar in leeftijd verschillen hoeft niet gelijk te zijn aan de verandering die een immigrant uit dezelfde regio, met dezelfde opleiding kan verwachten over vijf jaar, als hij inderdaad vijf jaar ouder zal zijn. Het eerste verschil geldt tussen verschillende personen, het tweede verschil is het leeftijdseffect van dezelfde persoon. Een immigrant die in 2016 vijf jaar ouder is hoeft niet representatief te zijn voor een statistisch vergelijkbare immigrant die in 2021 vijf jaar ouder geworden is.

Omwille van de niet-waarneembare verschillen tussen individuen en de dynamiek die onze maatschappij eigen is, is het belangrijk om na te gaan in hoeverre resultaten voor een bepaald peiljaar ook geldig zijn voor een ander peiljaar. Dat impliceert een pleidooi voor regelmatige herhaling voor een berekening als deze, als een monitoring van de fiscale implicaties van de ontwikkeling van immigratie. We verwachten niet dat de grote verschillen in fiscale effecten tussen herkomstregio's en migratiemotieven erg volatiel in de tijd zullen blijken. Maar zulke monitoring zal wel voortdurend gedetailleerde, goed gefundeerde informatie leveren voor het beleid, en tijdig verschuivingen aan het licht brengen. En het zal, zo nodig, sceptici kunnen overtuigen dat basisresultaten robuust zijn.

Een vergelijking met de resultaten van het CPB onderzoek uit 2003 – zoals gedaan in Hoofdstuk 4 van het huidige rapport en in §8.3 – geeft al aan dat bepaalde essentiële conclusies uit ons onderzoek toen ook golden. Ook volgens die studie is de bijdrage van de niet-westerse immigrant op elke entreeleeftijd negatief en is de voordeligste entreeleeftijd kort na voltooiing van een opleiding (25 à 30 jaar). De

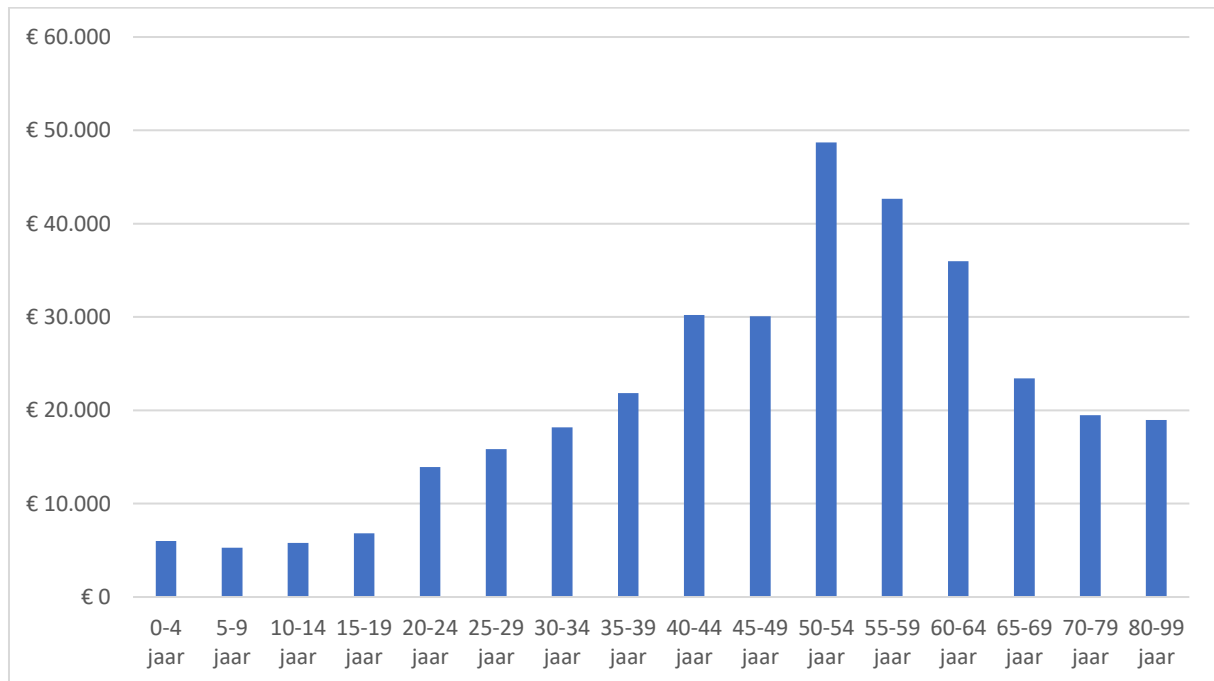
disproportionele afhankelijkheid van niet-westerse immigranten van sociale uitkeringen is ook vaak vastgesteld.¹⁶⁷ De deelname van leerlingen met niet-westerse migratieachtergrond vertoont een stijgende lijn¹⁶⁸, wat zou kunnen leiden tot een geringer negatief resultaat voor niet-westerse immigranten.

¹⁶⁷ Zorlu, A., J. Hartog en M. Beentjes (2010), Uitkeringsgebruik van migranten, Working Paper 10-101, AIAS, Universiteit van Amsterdam.

¹⁶⁸ <https://longreads.cbs.nl/integratie-2018/onderwijs/>

10 Sampling

Om zo objectief mogelijk te kunnen beoordelen hoe bij de indeling in regio's (en andere aggregaties) de afweging tussen betrouwbaarheid en nauwkeurigheid uitpakt is de sampling voor de verschillende leeftijdsgroepen bestudeerd. Voorafgaand aan de regio-indeling, het indelen in leeftijdsgroepen en dergelijke, is een schatting gemaakt van de standaarddeviatie van de uiteindelijke nettobijdrage per leeftijdsgroep, zie Figuur 10.1.¹⁶⁹ Deze grafiek laat zien dat de groepsverschillen van 0 tot 20 jaar relatief klein zijn. Vanaf 20 jaar neemt de standaarddeviatie toe. Tussen 40 en 65 jaar is de standaarddeviatie het hoogst om daarna weer te dalen. Dit laat zien dat het belangrijk is om voor leeftijden vanaf 20 jaar en vooral voor leeftijden van 40 tot 65 jaar veel waarnemingen per leeftijdsgroep te hebben.



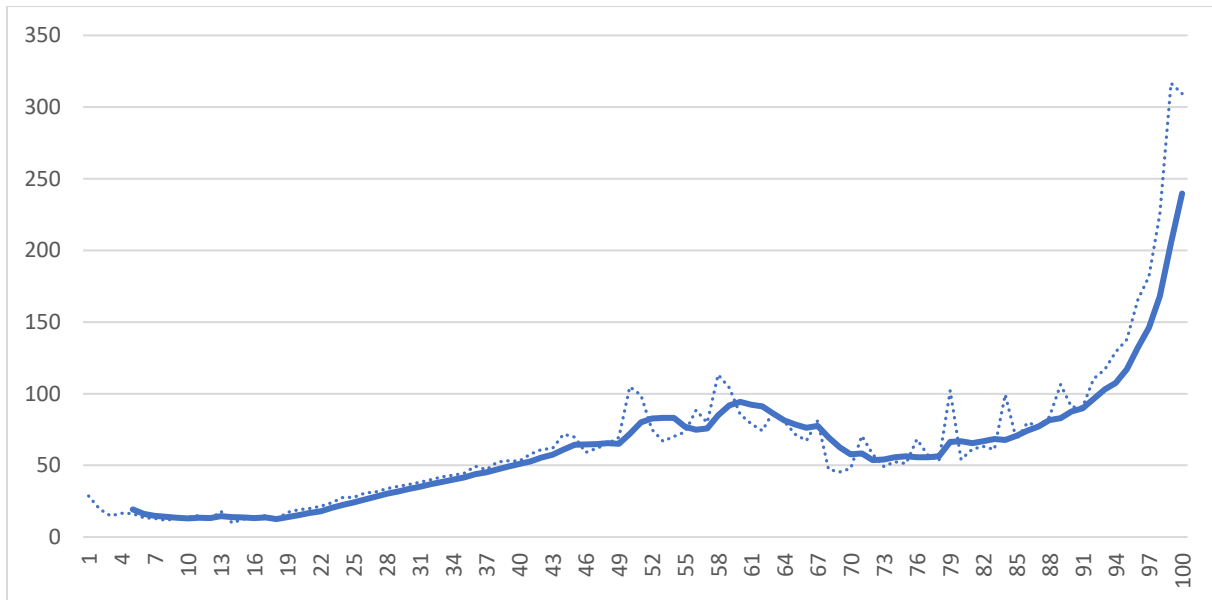
Figuur 10.1 Standaarddeviatie van de nettobijdrage op basis van CBS-microdata¹⁷⁰ per vijfjaarsleeftijdsgroep. Deze voorgaande observatie is gebruikt bij het indelen in regio's en in het bijzonder bij de vraag hoezeer de betrouwbaarheid in het geding komt als er door die indeling het aantal waarnemingen voor bepaalde leeftijdsgroepen klein wordt. Bij de beantwoording van deze vraag zijn de volgende overwegingen betrokken.

Voor kinderen jonger dan de basisschoolleeftijd (0-3 jaar) zijn lage aantallen waarnemingen per leeftijdsgroep geen probleem vanwege de geringe groepsverschillen. Eigenlijk verschillen op die leeftijd alleen de zorgkosten. Kleine aantallen of het samenvoegen van regio's zijn in dit geval geen probleem. Ook tijdens de rest van de leerplichtperiode zijn de verschillen relatief gering. Voor ouderen in de

¹⁶⁹ Zie voetnoot bij de titel van Figuur 10.1.

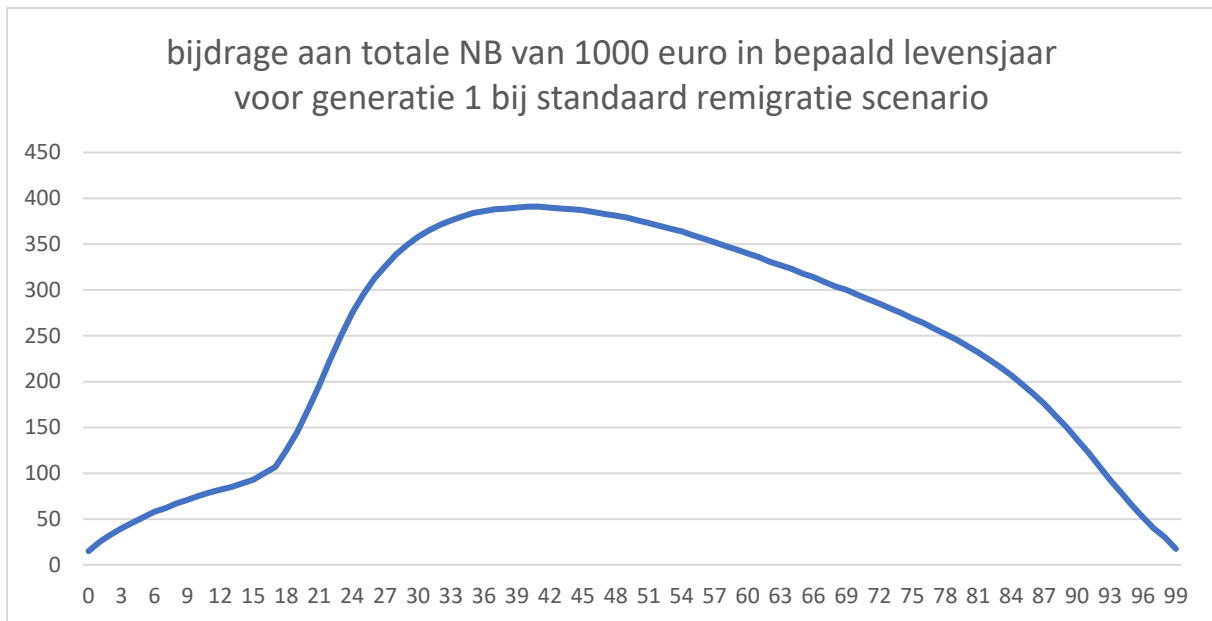
¹⁷⁰ Deze berekening is aan het begin van het proces, voorafgaand aan de indeling in regio's uitgevoerd met als doel vooraf een inschatting te kunnen maken van het effect van de indeling in herkomstregio's en leeftijdsgroepen. Deze berekening bevat alleen de posten die berekend zijn op basis van CBS-microdata, want in principe zijn dat de enige posten die direct bijdragen aan de variantie. Verder betreft dit de ruwe CBS-microdata bedragen; in de uiteindelijke berekening zijn deze veelal met een factor gewogen om te kijken op de CPB macrobedragen voor 2016. Ook waren in deze berekening de bruto investeringen in scholen, vennootschapsbelasting en IRN via bedrijven nog niet toegerekend naar personen en telden de leerlinggewichten voor 100% mee. Controle op basis van de definitieve berekeningen geeft echter vergelijkbare resultaten.

pensioenleeftijd is de standaarddeviatie gemiddeld drie tot vier keer zo hoog als voor leeftijden tot 20 jaar. In tegenstelling tot jongeren betalen ouderen belasting en hebben ze gemiddeld hoge zorgkosten. Daar komt bij dat de standaardfout mede afhangt van de groeps grootte per leeftijdsgroep en die is voor hoge leeftijden gemiddeld laag door toenemende sterftekans. Dit effect is geïllustreerd in Figuur 10.2 aan de hand van de bevolking als geheel. Te zien is dat de standaardfout na een piek rond 60 jaar aanvankelijk daalt, maar met name voor hoge leeftijden weer sterk stijgt, vanwege het effect van afnemende aantallen waarnemingen voor hogere leeftijden.



Figuur 10.2 Standaarddeviatie van de nettobijdrage voor de bevolking en 5-jaars voortschrijdend gemiddelde.

Toch zijn er enkele redenen waarom juist bij ouderen ‘onder-sampling’ een minder groot probleem is dan het op het eerste gezicht wellicht lijkt. De eerste reden hangt samen met de reden van onder-sampling zelf: namelijk dat er maar relatief weinig mensen echt heel oud worden en de nettobijdrage over de laatste levensfase daarom ook maar beperkt meeweegt in de totale nettobijdrage. In het geval van migranten speelt daarnaast ook remigratie (indien van toepassing in het betreffende scenario) een rol bij het geringe aantal ouderen. Om deze reden is het gewicht van ouderen in de totale nettobijdrage van de groep als geheel geringer. Verder komen de meeste immigranten als ze jong zijn, rond de 25 jaar oud, en slechts een klein deel is op het moment van immigratie ouder dan 50 jaar. De directe nettobijdrage van ouderen in de eerste verblijfsjaren is als geheel van de nettobijdrage van de groep dus beperkt. Uiteraard worden immigranten wel ouder en komen ze dan vanzelf in de beperkt gesampled leeftijdscategorieën. Echter, door het contant maken en door sterfte en eventuele remigratie wegen toekomstige bedragen minder zwaar mee. Dit alles is geïllustreerd in onderstaande figuur, waar voor de groep van alle eerste generatie immigranten in het standaardscenario met remigratie, voor elke leeftijd afzonderlijk, de bijdrage aan de totale nettobijdrage over de levensloop is berekend van een fout van +€ 1.000. Te zien is dat het effect voor jongeren gering is, vanaf 20 jaar snel toeneemt en rond 40 jaar piekt op ongeveer € 400 om daarna weer gestaag af te nemen.



Figuur 10.3 Totale toename van de totale nettobijdrage over de levensloop als gevolg van € 1.000 nettobijdrage in een bepaald levensjaar.

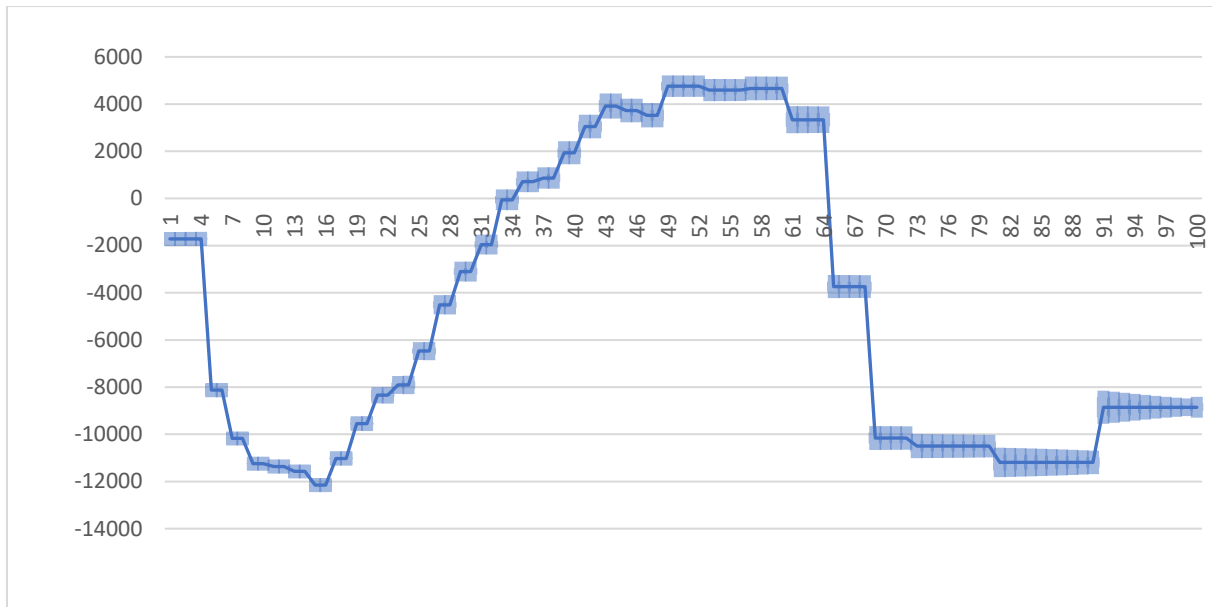
Om inzicht te geven hoe de foutmarges doorwerken in de berekening is in Figuur 10.4 de nettobijdrage per levensjaar weergegeven voor personen met een eerste generatie Marokkaanse migratieachtergrond. Het getrapte patroon komt door het samplen per leeftijdsgroep. De blauwe lijn geeft de gemiddelde bedragen voor het peiljaar 2016. De blauwe foutbalken geven de effectieve standaardfouten weer. Daarmee wordt bedoeld de standaardfout uitgaande van de (steekproef)standaarddeviaties en groeps grootte voor elke specifieke leeftijdsgroep, gewogen naar het effect op de nettobijdrage over de levensloop zoals weergegeven in Figuur 10.3.

Al met al blijkt de som over alle leeftijdsgroepen van de effectieve standaardfouten in de regel relatief klein te zijn, voor de CBS 12-deling meestal tussen € 5.000 en € 10.000. Uitzonderingen zijn Latijns-Amerika (€ 16.000) en vooral de regio Overig buiten Europa, die Japan, Noord-Amerika en Oceanië omvat (€ 47.000). Nadere analyse wijst uit dat vooral arbeidsmigranten soms een zeer hoge nettobijdrage hebben wat leidt tot een grotere standaarddeviatie, met name als de leeftijdsgroepen wat kleiner zijn zoals o.a. bij de regio Overig buiten Europa het geval is. Over het algemeen is dit geen groot probleem, omdat de meeste leeftijdsgroepen veel waarnemingen bevatten.

Bij arbeidsmigranten uit Latijns-Amerika en Overig buiten Europa leidt dit in een aantal gevallen tot vertekening doordat een klein aantal extreme grootverdieners in een leeftijdsgroep met geringe omvang vielen. Op die manier hangt de totale bijdrage van de groep teveel af van toeval: zou men bijvoorbeeld (ceteris paribus) één jaar eerder of later samplen, dan zou de bijdrage van de groep als geheel substantieel kunnen toe of afnemen.¹⁷¹ Verschillende oplossingen zijn overwogen. Het

¹⁷¹ Om dit concreet te maken een voorbeeld aan de hand van fictieve, ronde getallen: stel iemand uit een bepaalde groep draagt 10 miljoen euro bij in het peiljaar 2016. Stel dat deze persoon in 2016 in een leeftijdsgroep met 1000 personen valt, maar een jaar eerder of later in een groep met 100 personen. Stel verder dat het steeds om vierjaarsleeftijdsgroepen gaat. Valt de persoon in de groep van 1000 personen dan is de bijdrage van deze ene persoon aan het totale nettobijdrageprofiel $4 \times € 10.000.000 / 1000 = € 40.000$. Dat is een forse (aan één individu toe te schrijven) nettobijdrage, maar het is kennelijk een empirisch feit dat er zich onder de betreffende groep soms personen met een zeer hoog inkomen e.d. bevinden. Verwijderen van deze data is dus geen optie.

verwijderen van extreme uitbijters viel af omdat men dan ook de positieve bijdrage van de groep als geheel gaat verminderen. Andere oplossingen als het verzamelen van data voor meer dan één peiljaar werden te tijdrovend geacht. Uiteindelijk is er voor gekozen om voor arbeidsmigranten uit deze twee herkomstgebieden het gemiddelde van de nettobijdrage van de top 100 apart te berekenen en de originele records te vervangen door de gemiddelden, waarbij de leeftijdsdistributie wel gehandhaafd is (voor elk persoon van leeftijd $36 < L < 68$ in de top 100 is er een record met leeftijd L met de gemiddelde waarden over de top 100 voor alle voor alle variabelen).



Figuur 10.4 Sampling van de nettobijdrage voor personen met een eerste generatie Marokkaanse migratieachtergrond. gesampelde waarden en effectieve standaardfout, d.w.z. de standaardfout op basis van de grootte en standaarddeviatie van de samplegroepen gewogen naar de effectieve doorwerking in de totale nettobijdrage over de levensloop, als gevolg van het contant maken en wegen naar het immigratieprofiel.

Voor zover er bij de regio-indelingen sprake was van kleine aantallen ouderen en jongeren is dat verholpen door het gebruik van de gegevens van de bovenliggende regio. Dat is op twee manieren

Maar valt deze persoon in de groep met 100 personen, dan ontstaat een zeer grote spike op het profiel omdat de aan één persoon toegeschreven bijdrage aan de hele groep dan € 400.000 bedraagt. Effectief komt daarvan in het standaardscenario met remigratie al gauw één derde (zie Figuur 10.3) terecht in de totale nettobijdrage per groep. Om die reden is voor combinaties van arbeid en motief, steeds gekeken wat de ruwe nettobijdrage per individu zou zijn. Voor een beperkt aantal personen was die substantieel. Het betrof arbeidsmigranten uit de Europese Unie, Overig buiten Europa en Latijns-Amerika en vooral bij de laatste twee regio's was het effect soms groot omdat de aantallen wat oudere arbeidsmigranten uit die regio's veel kleiner zijn dan bij de EU het geval is. Vandaar dat gehandeld is zoals in de hoofdstuk tekst geschetst. De kern van het probleem is dat het aantal echte 'grootverdieners' zo gering is, dat een kans-gebeurtenis zoals 'er valt een persoon met een nettobijdrage van 5 miljoen of meer in leeftijdscategorie X' tot onaanvaardbaar grote door toeval veroorzaakte fouten leidt. De leeftijdsdistributie van de top 100 hoogste nettobijdragen in 2016 was bij benadering normaal verdeeld voor leeftijden tussen 36 en 71 met een piek voor vijftigers, maar een enkele keer valt zo'n persoon in oudere leeftijdsgroep met relatief weinig waarnemingen. Verwijderen van deze personen is ontkennen van de empirie, maar herverdeling van hun nettobijdrage naar rato van de top 100 leeftijdsdistributie binnen hun groep, is slechts een middel om het toeval dat ontstaat door zeer zeldzame gebeurtenissen te verkleinen. Het effect is naar verwachting vergelijkbaar met het nemen van een groot aantal peiljaren, maar eenvoudiger uit te voeren. De berekening is uitgevoerd voor verschillende Top-X tussen 10 en 100. Er zijn geen grote verschillen.

gebeurd: handmatig en met een algoritme. Het algoritme wordt verderop in deze paragraaf beschreven, in deze en de volgende alinea's volgt de beschrijving van de handmatige invulling vanuit de bovenliggende regio.

Bij de 19-deling is in twee gevallen¹⁷² voor de leeftijdsgroep 90-99 jaar gebruik gemaakt van de data van de bovenliggende regio. Voor de 42-deling is eveneens in een aantal gevallen voor de leeftijdsgroepen 0-3 jaar,¹⁷³ 80-89 jaar¹⁷⁴ en 90-99 jaar¹⁷⁵ data van de bovenliggende regio gebruikt. Deze profielen zijn dus synthetisch, in die zin dat ze opgebouwd zijn uit profielen voor verschillende groepen. Om synthetische profielen zoveel mogelijk te voorkomen, zijn de regio's zodanig ingedeeld dat er voor leeftijden tussen 20 en 72 jaar veel waarnemingen waren. Hierop zijn twee uitzonderingen gemaakt: Japan en Israël zijn door bestaande CBS-indelingen en grote verschillen met omliggende landen noodgedwongen apart ingedeeld in de 42-deling. Bij Japan zijn om die reden voor het midden van het leven de leeftijdsgroepen groter gemaakt door een aantal malen twee opeenvolgende leeftijdsgroepen samen te nemen.

In de huidige studie is ook gesampled naar migratiemotief. Ook hierbij zijn in een aantal gevallen ontbrekende data handmatig aangevuld vanuit het bovenliggende niveau. Bij de berekening voor de nettobijdrage per migratiemotief speelt het probleem van kleine aantallen vooral bij studie- en arbeidsmigranten. Er zijn – uit de aard der zaak – weinig zeer jonge studie- en arbeidsmigranten. Met name het aantal waarnemingen in de leeftijdsgroepen tot 16 jaar is gering of nul. Daarnaast is het aantal 68-plussers onder studiemigranten laag en is er geen sample voor de groep 90 jaar en ouder. Het gebrek aan data is in dit geval als volgt opgelost. Voor studie- en arbeidsmigranten tot 16 jaar zijn de data voor autochtonen gebruikt. De rationale hierachter is dat studie- en arbeidsmigranten vrij hoge gemiddelde citoscores hebben en voor deze leeftijden er een sterke correlatie is tussen citoscore en nettobijdrage per levensjaar. Overigens is in dit specifieke geval het effect van om het even welke (redelijke) keuze verwaarloosbaar, omdat het aandeel studie- en arbeidsmigranten tot 16 jaar zeer klein is, zodat ook het effect op de totale nettobijdrage gewogen naar entreeleeftijd verwaarloosbaar is. Voor studiemigranten vanaf 90 jaar ontbreken eveneens data en hiervoor zijn de gegevens van de eerste generatie als geheel gebruikt.

Daarnaast is ook geaggregeerd naar combinaties van migratiemotief en regio. Voor combinaties van motief met de indeling westers en niet-westers is er – behoudens het in de vorige alinea besproken probleem met de jonge studie- en arbeidsmigranten en de oude studiemigranten – over het algemeen voldoende data. Er is echter ook geaggregeerd naar combinaties van motief met de CBS-12 deling, dit met het oog op aansluiting bij de indeling naar herkomstregio's die het CBS onder andere gebruikt bij haar bevolkingsprognoses. Door het combineren van motief en regio wordt het aantal waarnemingen per leeftijdsgroep uiteraard nog kleiner. Hier is per geval bekeken hoe de (synthetische) nettobijdrageprofielen moesten worden samengesteld voor een zo betrouwbaar mogelijke berekening. Voor de herkomstregio Aruba en (voormalige) Nederlandse Antillen waren er alleen voor de categorie

¹⁷² Voor herkomstregio's Noord-Afrika, Sub-Sahara Afrika en Indisch subcontinent.

¹⁷³ Voor de deelregio's van herkomstregio's Sub-Sahara Afrika, Indisch subcontinent, Oost-Azië en Latijns-Amerika (excl. Suriname, Aruba en Antillen).

¹⁷⁴ Voor de deelregio's van herkomstregio's Sub-Sahara Afrika, Indisch subcontinent en Overig buiten Europa.

¹⁷⁵ Voor de deelregio's van herkomstregio's Sub-Sahara Afrika, Indisch subcontinent, Europese Unie, Overig (buiten) Europa, Oost- en West-Azië en Latijns-Amerika (excl. Suriname, Aruba en Antillen).

onbekend¹⁷⁶ voldoende waarnemingen. Voor alle andere motiefcategorieën waren er nagenoeg geen waarnemingen en deze zijn derhalve niet meegenomen in de analyse. Voor asielmigranten is een indeling gemaakt in de regio's Afrika (excl. Marokko), West-Azië, overig niet-westers en westers, omdat de hoeveelheid data een verdere verfijning niet toestond.

Met name bij verfijningen van de regio-indeling – al dan niet in combinatie met motief of opleidingsniveau – is ook na aanvullen vanuit het bovenliggende niveau soms te weinig data en/of geeft aanvullen vanuit het bovenliggende niveau trendbreuken in het leeftijdsprofiel voor de nettobijdrage. Om dit probleem om te lossen en met name voor de hogere leeftijden te komen tot een betere schatting is een algoritme ontwikkeld. Dit algoritme verloopt in twee stappen.

In stap 1 worden de waarnemingen aangevuld tot een minimum uit het bovenliggende niveau. Hierbij zijn drie verschillende hiërarchieën onderscheiden voor het bovenliggende niveau. Voor de regio's is de indeling AlleHerkomsten > Landtype > CBS 12-deling > 19-deling > 42 deling > 87-deling (zie §4.4) aangehouden. Bij combinaties van regio met motief is de indeling Alle Motieven × AlleRegios > Motief × AlleRegios > Motief × Landtype > Motief × CBS 12-deling > Motief × CBS 19-deling¹⁷⁷ aangehouden. Voor Opleiding is de indeling Opleiding > Opleiding × Generatie > Opleiding × Generatie × Landtype aangehouden. Bij de aanvulling uit het bovenliggende niveau is 25 waarnemingen per leeftijdsjaar¹⁷⁸ als norm aangehouden. Als er reeds sprake was van handmatig aanvullen uit een bovenliggend niveau is zijn de aantallen voor de betreffende groepen op 25 gesteld.

In stap 2 is vervolgens voor de leeftijden met weinig waarnemingen een naar samplegrootte gewogen gemiddelde genomen van een best passend nettobijdrageprofiel en de waarnemingen. De techniek die daarvoor is gebruikt is dezelfde als beschreven in §2.4, waarbij als norm net als bij de tweede generatie 100 waarnemingen per levensjaar is aangehouden. Voor leeftijden tot 72 jaar staat dat meestal gelijk aan 400 waarnemingen per vierjaarsleeftijdsgroep.

Voor de uitvoering in stap 1 is dus gekozen voor een gelaagde oplossing wat betreft de regio's, waarbij indien nodig data van het 'bovenliggende niveau' gebruikt zijn. Om dit concreet te maken de uitwerking voor de combinaties van motief en regio. Eerst is de best mogelijke schatting gemaakt voor motief (voor alle regio's samen), daarna voor motief i.c.m. landtype (westers/niet-westers), vervolgens voor motief i.c.m. de CBS 12-deling in wereldregio's en tot slot voor motief i.c.m. de 19-deling (zie eventueel de uitleg over de verschillende niveaus van regio-indeling). Hierbij is motief¹⁷⁹ dus het bovenliggende niveau voor motief × landtype, vervolgens is motief × landtype weer het bovenliggende niveau voor

¹⁷⁶ Dat wil zeggen, motief onbekend vanaf 1995, immigranten met onbekend motief van voor 1995 zijn in de analyse niet meegenomen in verband met de vergelijkbaarheid met de overige motieven die pas vanaf 1995 beschikbaar zijn.

¹⁷⁷ Alleen voor de motieven arbeid en gezinsmigratie.

¹⁷⁸ In de praktijk voor leeftijden tot 72 jaar meestal 100 waarnemingen per vierjaarsleeftijdsgroep en (minder vaak) 50 waarnemingen per tweejaarsleeftijdsgroep, 200 waarnemingen per achtjaarsleeftijdsgroep of 250 waarnemingen per tienjaarsleeftijdsgroep.

¹⁷⁹ In geval van aggregatie naar motief is er geen echt 'bovenliggend niveau'. Voor studie- en arbeidsmigranten tot 16 jaar ontbreken data en hiervoor zijn data voor autochtonen gebruikt. De rationale hierachter is dat studie- en arbeidsmigranten vrij hoge gemiddelde citoscores hebben en voor deze leeftijden er een sterke correlatie is tussen citoscore en nettobijdrage per levensjaar. Overigens is in dit specifieke geval het effect van om het even welke (redelijke) keuze verwaarloosbaar, omdat het aandeel studie- en arbeidsmigranten tot 16 jaar zeer klein is, zodat ook het effect op de totale nettobijdrage gewogen naar entreeleeftijd verwaarloosbaar is. Voor studiemigranten vanaf 90 jaar ontbreken eveneens data en hiervoor zijn de gegevens van de eerste generatie als geheel gebruikt.

motief \times 12-deling en tot slot is motief \times 12-deling weer het bovenliggende niveau voor motief \times 19-deling.¹⁸⁰ Voor combinaties van motief en regio is indien nodig het profiel voor het bovenliggende niveau gebruikt om de leeftijdsgroepen met weinig waarnemingen aan te vullen tot de eerder genoemde norm van 25 waarnemingen per levensjaar. Dit is gedaan door de samplegroottes per leeftijdsgroep te transformeren tot een gewichtenvector die maximaal 1 is voor groepen met 25 of meer waarnemingen per levensjaar.¹⁸¹

Voor de uitvoering van stap 2 is eerst (net als voor de tweede generatie, zie §2.4) de best passende combinatie van het nettobijdrageprofiel voor autochtonen en voor eerste generatie niet-westerse immigranten gefit. Vervolgens is het gewogen gemiddelde van het profiel in stap 1 en het voornoemde gefitte profiel genomen. Het wegen is gedaan door de samplegroottes per leeftijdsgroep te transformeren tot een gewichtenvector die maximaal 1 is voor groepen met 100 of meer waarnemingen per levensjaar en minimaal 0,25 voor groepen met 25 of minder waarnemingen per levensjaar.¹⁸² Op die manier worden bij 100 of meer waarnemingen per levensjaar alleen de waargenomen data gebruikt en naarmate het aantal waarnemingen verder onder 100 per levensjaar komt wordt dit aangevuld met data uit het bovenliggende niveau en het gefitte profiel.

Deze aanpak is bewerkelijk, maar daar staan enkele voordelen tegenover. In de eerste plaats wordt alle data die er wel zijn meegenomen in de berekening. Stap 1 zorgt er verder voor dat voor alle leeftijden er een verankering plaatsvindt in het bovenliggende niveau die in stap 2 uiteindelijk voor 25% meeweegt. Dit is een waarborg tegen extreme uitslagen voor leeftijdsgroepen waarvoor er echt heel weinig of zelfs geen data zijn. Doordat bij het fitten in stap 2 het aantal waarnemingen meegewogen wordt in de kwadratensom (zie §2.4) wordt net als bij de tweede generatie vertekening door cohort-effecten gedempt. Immers, recente cohorten met een korte verblijfsduur zijn sterker vertegenwoordigd in de leeftijden waarin het immigratieprofiel gewoonlijk piekt (en worden daarom zwaarder meegewogen bij het fitten) dan eerdere cohorten met een verblijfsduur van bijvoorbeeld 20 jaar. Voor bijvoorbeeld Turkse arbeidsmigranten leidt deze aanpak tot beduidend hogere uitkomsten dan een aanpak waarin men de samplegrootte niet meeweegt. Dit weerspiegelt het feit dat in de gemaakte dwarsdoorsnede Turkse arbeidsmigranten in de relatief omvangrijke jonge leeftijdsgroepen een beduidend hogere nettobijdrage hebben dan in de oudere leeftijdsgroepen die veel minder waarnemingen tellen.

Bij de berekeningen voor de nettobijdrage per opleiding speelt het probleem van kleine aantallen waarnemingen vooral bij de actueel gevolgde opleidingen voor ouderen en de hoogst behaalde leeftijden voor jongeren. Weinig dertigers volgen een middelbare schoolopleiding bijvoorbeeld en er zijn ook maar weinig tieners met een afgeronde masteropleiding. Vooral voor de eerste generatie was er voor een aantal leeftijdsgroepen weinig data. Omdat voor eerste generatie immigranten met hogere

¹⁸⁰ Concreet: voor westerse arbeidsmigranten is Arbeid het bovenliggende niveau; voor arbeidsmigranten uit de Europese Unie is Arbeid \times Westers het bovenliggende niveau; voor arbeidsmigranten uit de GIPS-landen is Arbeid \times Europese Unie het bovenliggende niveau.

¹⁸¹ De samplegroottes zijn gedeeld door de omvang van de leeftijdsgroep in jaren, vervolgens gedeeld door 25 en als laatste stap gemaximeerd op 1.

¹⁸² De samplegroottes zijn gedeeld door de omvang van de leeftijdsgroep in jaren, vervolgens gedeeld door 100 en daarna gemaximeerd op 1. Hierbij is verdisconteerd dat in stap 1 het aantal waarnemingen al op basis van het bovenliggende niveau is aangevuld voor groepen met minder dan 25 waarnemingen per levensjaar. Vandaar dat het minimale gewicht 0,25 is.

entreeleeftijden sowieso geen betrouwbare opleidingsgegevens zijn is deze groep geheel buiten beschouwing gelaten. De analyse is derhalve beperkt tot autochtonen en de tweede generatie. Indien bij deze groepen sprake was van kleine aantallen waarnemingen is dit in eerste instantie opgelost door de leeftijdsgroepen te vergroten. Als dat onvoldoende soelaas bood zijn voor de tweede generatie bij de indeling westers/niet-westers de data van alle tweede generatie immigranten gebruikt of indien nodig van de hele bevolking. Uiteraard levert dit voor deze groepen vertekening op. Echter, deze data worden gebruikt om de nettobijdrage naar citoscore te berekenen en daarbij is (om bij het eerder genoemde voorbeeld te blijven) het aandeel dertigers dat een middelbare schoolopleiding volgt of het aantal tieners met een afgeronde masteropleiding zo gering dat het effect van eventuele meetfouten in de totaalberekening verwaarloosbaar is.¹⁸³ Sterke en zwakke punten en suggesties voor verder onderzoek

¹⁸³ Vergelijk dit met het door het wegen tegen het immigratieprofiel verwaarloosbare effect van meetfouten door het ontbreken van data voor arbeidsmigranten tot 16 jaar en studiemigranten tot 12 jaar.

11 Sterke en zwakke punten en suggesties voor verder onderzoek

11.1 Sterke punten

Een sterk punt van het huidige onderzoek is dat vrijwel alle posten waarbij dat mogelijk en zinvol was zoveel mogelijk zijn ingevuld met CBS-microdata. De mate van detail in uitsplitsing naar herkomstregio en/of migratiemotief geeft daarnaast veel inzicht.

Hetzelfde geldt voor de berekening van de nettobijdrage naar hoogst behaalde opleiding en citoscore. Dat laat, samen met de aangetoonde samenhang van citoscores tussen de generaties, uitstekend het langetermijneffect zien van het al dan niet selecteren op opleidingsniveau en/of opleidingspotentie.

Ook sterk is dat de nettobijdrage van de tweede generatie niet zoals in de CPB-studie uit 2003 gebaseerd is op de aanname dat de tweede generatie voor de helft geïntegreerd zou zijn, maar dat per groep op verschillende manieren een schatting is gemaakt van de mate van integratie. Voor het integratiebeleid levert dit een waardevolle sociaaleconomische integratiemaat op voor de tweede generatie – die tot nu toe nog niet is gehanteerd – en die bij regelmatig actualiseren gebruikt zou kunnen worden om de voortgang van de integratie te monitoren.

Ook sterk is het kwantificeren van de oorzaken van de grote verschillen in nettobijdrage tussen (met name) niet-westerse immigranten en autochtonen in §9.12. Dit laat zien dat de verschillen zeker niet gezocht moeten worden in het Nederlandse onderwijssysteem, maar vooral in gebrekkige absorptie op de arbeidsmarkt en gebrek aan selectie op opleidingsniveau bij toelating tot Nederland.

De maatschappelijk implicaties van de resultaten in dit rapport zijn groot. De resultaten bieden handvatten aan beleidsmakers voor een rationeel migratiebeleid. Het maakt inzichtelijk hoe de kosten en baten van immigratie samenhangen met de herkomstregio en het migratiemotief. Daarnaast maakt het duidelijk hoe bepalend het opleidingsniveau is voor de fiscale nettobijdrage van de eerste generatie en – via de citoscores – ook voor de nettobijdrage van de tweede en volgende generaties. Selectie op opleidingsniveau is gangbaar in klassieke immigratielanden als Canada en Australië en de effecten van een dergelijke selectie zijn naar verwachting in Nederland nog groter, door de uitgebreide en relatief toegankelijke verzorgingsstaat.

De aanbeveling om deze generatierekeningen vanaf nu periodiek te herhalen zou ook inzicht geven in het effect van beleidsmaatregelen. Regelmatig actualiseren is daarnaast ook nodig omdat het overheidsbeleid wat betreft inkomsten en uitgaven, voortdurend in beweging is en de kosten- en batenposten mee veranderen.

11.2 Zwakke punten

Tegelijkertijd kent het huidige rapport ook een aantal zwakten. Nog niet alle posten zijn volledig ingevuld met CBS-microdata. In de meeste van die gevallen – defensie, overheidsinvesteringen in gebouwen en infrastructuur, overdrachten naar het buitenland, grondverkoop en overige niet-belastingmiddelen – is toerekening naar individuen door operationalisering met CBS-microdata niet mogelijk en/of niet zinvol (zie post 2, 10, 13, 14, 22 en 23). Er zijn echter drie kostenposten waarvoor het nuttig zou zijn om te onderzoeken of er verdere invulling met microdata en/of macrogegevens mogelijk is, namelijk openbaar bestuur, zorg en een deel van de sociale zekerheid. Bij openbaar bestuur zou het bijvoorbeeld kunnen gaan om de kosten van beleid dat specifiek op immigranten is gericht of de kosten van lokaal beleid (gemeenten, provincies, etc.), maar niet onder een van de andere 22 posten valt. Bij

zorg gaat het om de nog niet met CBS-microdata ingevulde kosten voor (bijzondere en langdurige) Zorg (ca. 15 miljard euro). Bij sociale zekerheid gaat het om een deel van de kosten voor de post SZ-rest (ca. 11 miljard euro). In totaal gaat het om een aanzienlijk deel van de overheidsuitgaven in 2016 (ca. 300 miljard).

Verder zijn er bij de tweede generatie voor een aantal groepen nog weinig data en dat maakt de schattingen voor die groepen onzekerder. Het probleem van beperkte data geldt ook voor een aantal combinaties van migratiemotief en herkomstregio in de 12-deling en 19-deling. Dit heeft overigens een verwaarloosbaar effect op de berekeningen van de totale nettobijdrage voor historische en toekomstige migratie in Hoofdstuk 7 van het huidige rapport, omdat kleine groepen ook maar beperkt meewegen in de totale nettobijdrage.

11.3 Suggesties voor verder onderzoek

Deze observaties leiden tot een aantal suggesties voor verder onderzoek. Allereerst verdient het aanbeveling om vooral de post Zorg verder met CBS-microdata in te vullen, bijvoorbeeld met de bestedingen uit de AWBZ-gelden. Overigens zal mettertijd de hoeveelheid data toenemen voor groepen – zoals de tweede generatie en de migratiemotiefgroepen – waarvoor nu soms beperkt data beschikbaar zijn. Dit alles zal tot een (steeds) betere schatting van de nettobijdrage leiden.

In de sfeer van toepassing kunnen de tabellen met nettobijdragen uitgesplitst naar regio, immigratiemotief en/of opleidingsniveau handvatten bieden voor vele overheidsonderdelen. Allerlei beleidsvoornemens kunnen door gebruik van deze bedragen worden doorgerekend op hun budgettaire effect.

Verder zou het goed zijn deze berekeningen periodiek te herhalen. Zo wordt de gevoeligheid van de berekeningen voor keuzes wat betreft data, peiljaar en dergelijke steeds beter inzichtelijk. Ook wordt dan het verloop van de integratie en de effecten van beleid beter zichtbaar. Naarmate er mettertijd meer data beschikbaar komt, zou ook de derde generatie meer in de analyse betrokken kunnen worden, om te bezien in hoeverre de geconstateerde achterblijvende integratie bij verschillende herkomstgroepen zich ook in de derde generatie doorzet. Relatief lage citoscores voor een aantal derde generatie groepen suggereren dat dit waarschijnlijk is, maar direct onderzoek naar de nettobijdrage is nu nog slechts beperkt mogelijk.

Vanwege de doelstelling en het daaruit voortvloeiende beschrijvend karakter van deze studie is het toetsen van oorzakelijke verbanden geen doel op zich geweest. In een aantal gevallen werd bij het beschrijven van de kosten en baten een (mogelijke) samenhang gevonden tussen twee of meer variabelen. Een samenhang tussen variabelen kan wijzen op een oorzakelijk verband, maar vormt op zich geen bewijs dat er een oorzakelijk verband is. Daarvoor zou nader onderzoek gedaan moeten worden.

Nader onderzoek naar zelfselectiemechanismen bij remigratie zou zeer nuttig zijn om meer grip te krijgen op de causale mechanismen die groepsverschillen bestendigen of zelfs versterken. Zo zou er bijvoorbeeld nader onderzoek gedaan kunnen worden naar de in §2.2 getoonde associatie tussen uitkeringsafhankelijkheid en remigratiekansen. Er lijkt zoiets te zijn als een ‘omgekeerd welvaartsmagneeteffect’ waarbij de ‘welvaartsmagneet’ uitkeringsafhankelijke immigranten relatief vaak in de Nederlandse welvaartstaat houdt terwijl andere groepen relatief vaak remigreren. Iets vergelijkbaars geldt voor de negatieve zelfselectie op citoscore die kennelijk optreedt bij remigratie (§9.12) en daaraan mogelijk gerelateerd de aanzienlijke verschillen in citoscore die er bestaan tussen de motieven en tussen westerse en niet-westerse immigranten (§9.7).

Een ander belangrijk punt dat nader onderzoek verdient is de registratie van het migratiemotief door de IND (zie §6.4) Dat weerspiegelt vaak niet het daadwerkelijke gedrag van immigranten en de uitkeringsafhankelijkheid onder geregistreerde studie- en arbeidsmigranten uit onder meer Afrika en het Midden-Oosten is onwaarschijnlijk hoog. Dit vraagt om een nader onderzoek naar de registratie en toetsing van verblijfsaanvragen door de IND.

Opvallend is tevens de buitengewoon hoge consumptie van arbeidsongeschiktheidsuitkeringen door immigranten met migratieachtergrond in Turkije, Marokko en voormalig Joegoslavië. Het is nader onderzoek waard of er daadwerkelijk zo'n groot verschil in arbeidsgeschiktheid bestaat tussen de verschillende groepen en welke ziekten, gebreken of anderszins er zorgen ervoor dat zij in zulke grote getalen niet kunnen werken. Iets vergelijkbaars geldt voor het bovengemiddeld hoge beroep op WW-uitkeringen door Poolse immigranten, dat samenhangt met aangetoonde fraude (zie §8.6-§8.7).

Tot slot suggereren de resultaten dat sociaaleconomische en sociaal-culturele integratie hand in hand gaan. Duidelijk is dat kinderen met één in Nederland geboren ouder (bij de tweede generatie) en kinderen met één ouder met een Nederlandse achtergrond (bij de derde generatie) veelal hogere netto-bijdragen hebben. Het is nader onderzoek waard om deze verschillen te verklaren. Het zou kunnen dat zelfselectie en partnerselectie hierbij een rol spelen in die zin dat mensen die voor een partner kiezen met een Nederlandse achtergrond andere karakteristieken hebben dan mensen die dat niet doen. Andere mogelijke verklaringen zijn verschillen naar migratieachtergrond in kindgebonden en oudergebonden factoren die bepalend zijn voor schoolsucces en maatschappelijke positie en bejegening.

In het verlengde hiervan is ook de statistische samenhang tussen in het herkomst land heersende waardenvoorkeuren en sociaal-culturele afstand enerzijds en sociaaleconomische integratie en citoscores anderzijds een nader onderzoek waard. Mogelijk kunnen ze inzichten opleveren die beleidsmakers kunnen integreren in het toelatingsbeleid met het oog op een soepele integratie van toekomstige immigranten.

Errata

Errata in het huidige rapport worden zo snel mogelijk verbeterd in de hoofdtekst.

Update: 13-3-2021, 8:08 uur

- Blz. 30 taalfouten bij begrip Referentie-autochtoon
- Blz. 82 spaties ontbreken in Box 4.2 tussen € en bedragen
- Blz. 162 Tabel 9.2, bedragen niet afgerond op veelvouden van € 1.000
- Blz. 207 taalfouten in blauwe box
- Blz. 212 "100.000 personen" moet zijn "175.000 personen", ter verduidelijking verwijzing naar Technische appendix toegevoegd