



UNIVERSITEIT VAN AMSTERDAM
Amsterdam School of Economics

Technische Appendix

bij de tweede druk van het rapport

Grenzeloze Verzorgingsstaat De Gevolgen van Immigratie voor de Overheidsfinanciën



Jan H. van de Beek
Hans Roodenburg
Joop Hartog
Gerrit W. Kreffer

Dit is de Technische Appendix bij het rapport 'Grenzeloze verzorgingsstaat: De gevolgen van immigratie voor de overheidsfinanciën'.

In deze appendix wordt de gebruikte methode toegelicht. Dit vergemakkelijkt het begrijpen en beoordelen.

De conclusie van het rapport is verontrustend. Voortgaande immigratie met de huidige omvang en kostenstructuur zal de overheidsfinanciën toenemend onder druk zetten. Afbouw van de verzorgingsstaat en/of inperking van de immigratie zijn dan onvermijdelijk.

De appendix vergroot de mogelijkheid tot replicatie van de berekeningen. De auteurs hopen, dat dit periodieke actualisatie van het rapport bevordert.

Technische Appendix

bij de tweede druk van het rapport

Grenzeloze Verzorgingsstaat

De Gevolgen van Immigratie
voor de Overheidsfinanciën

Jan H. van de Beek

Hans Roodenburg

Joop Hartog

Gerrit W. Kreffer

Copyright © 2021-2023 Uitgeverij Demo-Demo, Zeist

Tweede herziene druk, april 2023

Auteurs: Jan H. van de Beek, Hans Roodenburg, Joop Hartog, Gerrit W. Kreffer

Druk: Pumbo.nl

Omslagontwerp en vormgeving binnenwerk: Jan H. van de Beek

Omslagfoto: Brian Merrill via Pixabay

Soort uitgave: technische appendix bij onderzoeksrapport

ISBN: 9789083334813

NUR: 781

Zie voor meer informatie: <https://demo-demo.nl/>

Niets uit deze uitgave mag worden verveelvoudigd, door middel van druk, fotokopieën, geautomatiseerde gegevensbestanden of op welke andere wijze ook zonder voorafgaande schriftelijke toestemming van de uitgever.

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 micro-data@cbs.nl.

Inhoud

1	Inleiding	7
2	Generatierekening.....	8
2.1	Generatierekening en nettobijdrage.....	8
2.2	Inkomsten en uitgaven van de overheid.....	10
2.3	Generatierekening voor de eerste generatie.....	12
2.4	Generatierekening voor de tweede generatie.....	17
2.5	Gevoeligheidsanalyse.....	30
3	Onderzoekspopulatie en microdatabestanden.....	34
4	Socio-demografische variabelen	36
4.1	Leeftijd en generatie	36
4.2	Opleidingsniveau en citoscores.....	37
4.3	Migratiemotief	52
4.4	Regio.....	54
5	Kosten en baten posten	58
5.1	Kosten en baten posten	58
5.2	Veiligheidszorg	60
5.3	Onderwijs en Bruto investeringen scholen	62
5.4	Kinderopvang	70
5.5	Zorg.....	70
5.6	Uitkeringen, subsidies, premies en belastingen.....	71
5.7	IRN	75
5.8	Vermogen-gerelateerde belastingen	76
6	Aanloopkosten en kosten na remigratie	82
6.1	Asielopvang en vergunningverstrekking	82
6.2	Inburgering	83
6.3	AOW-rechten.....	86
7	Demografie en vergrijzing	88
7.1	Geboorte en sterfte.....	88
7.2	Migratie	90
7.3	Vergrijzing.....	91
7.4	Demografie.....	93
8	Verankering in CPB-studies	96

8.1	Verankering in CPB-vergrijzingsstudies en CPB-datasets.....	96
8.2	IJking op CPB-vergrijzingsstudie <i>Zorgen om morgen</i> (2019).....	98
8.3	Vergelijking met CPB-rapport <i>Immigration and the Dutch Economy</i> (2003)	101
9	Belangrijke punten bij de interpretatie van de resultaten.....	104
9.1	Beschrijving is geen toetsing, samenhang is geen oorzakelijk verband.....	104
9.2	Eén peiljaar: noodzaak van herhaling	104
10	Sampling	106
10.1	Meetfouten	106
10.2	Verdeling over celgrootteklassen.....	109
10.3	Behandeling van leeftijdsgroepen met weinig waarnemingen.....	113
10.4	Behandeling van enkele extreme <i>outliers</i> (uitbijters)	118
10.5	Hoofdpunten sampling.....	121
11	Sterke en zwakke punten en suggesties voor verder onderzoek.....	122
11.1	Sterke punten.....	122
11.2	Zwakke punten	122
11.3	Suggesties voor verder onderzoek.....	123
12	Model	126
	Figuren.....	130
	Tabellen	132

1 Inleiding

Dit is de Technische Appendix bij het rapport *Grenzeloze verzorgingsstaat: De gevolgen van immigratie Voor de overheidsfinanciën*.¹ Om verwarring te voorkomen wordt het rapport *Grenzeloze verzorgingsstaat* in deze Technische Appendix aangeduid als ‘het huidige rapport’ of ‘de huidige studie’ en deze Technische Appendix als ‘deze appendix’. 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 achterliggend 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 wordt een overzicht gegeven van 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, vergunningverstrekking 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 11 (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.

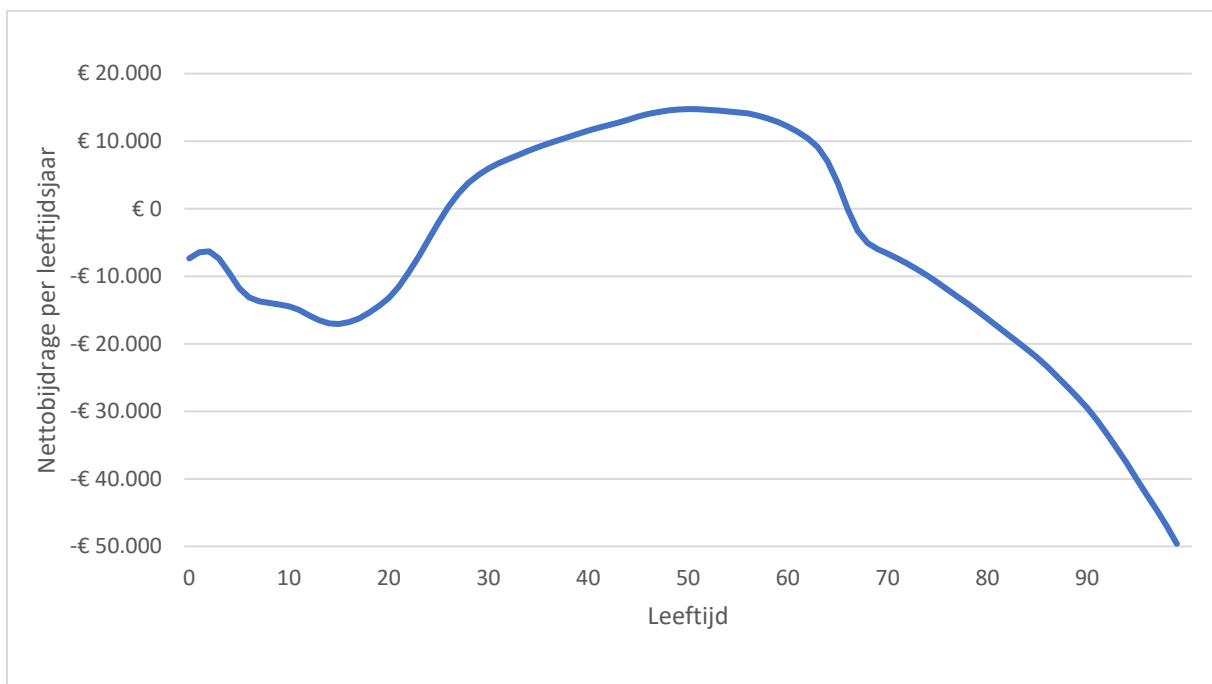
¹ Van de Beek, J. H., J. Hartog, H. J. Roodenburg & G. W. Kreffer (2021) *Grenzeloze verzorgingsstaat: De gevolgen van Immigratie voor de Overheidsfinanciën*, ISBN nummer: 9789083334806, zie ook <https://demo-demo.nl/>

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.

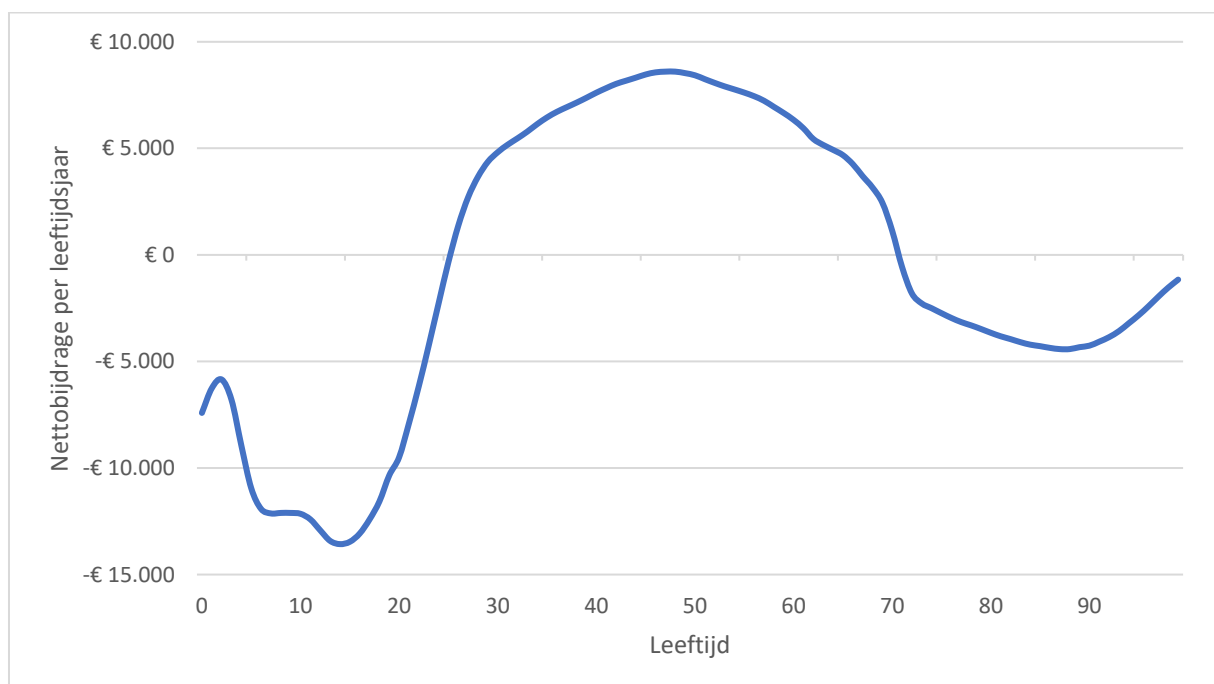
Nettobijdrage per leeftijdsjaar is een sleutelbegrip in het huidige rapport. Het is de hoeveelheid geld die per saldo in een bepaald leeftijdsjaar van een persoon naar de schatkist vloeit. Draagt iemand in een bepaald leeftijdsjaar meer bij aan de schatkist dan hij of zij ontvangt, dan is de nettobijdrage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

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. J. Roodenburg & G. W. 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 een generatie in een aantal gevallen wel min of meer 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).

bijdraagt, dan 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 het huidige rapport 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 (dus met en zonder migratieachtergrond) 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 het huidige rapport – 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 het huidige rapport. 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

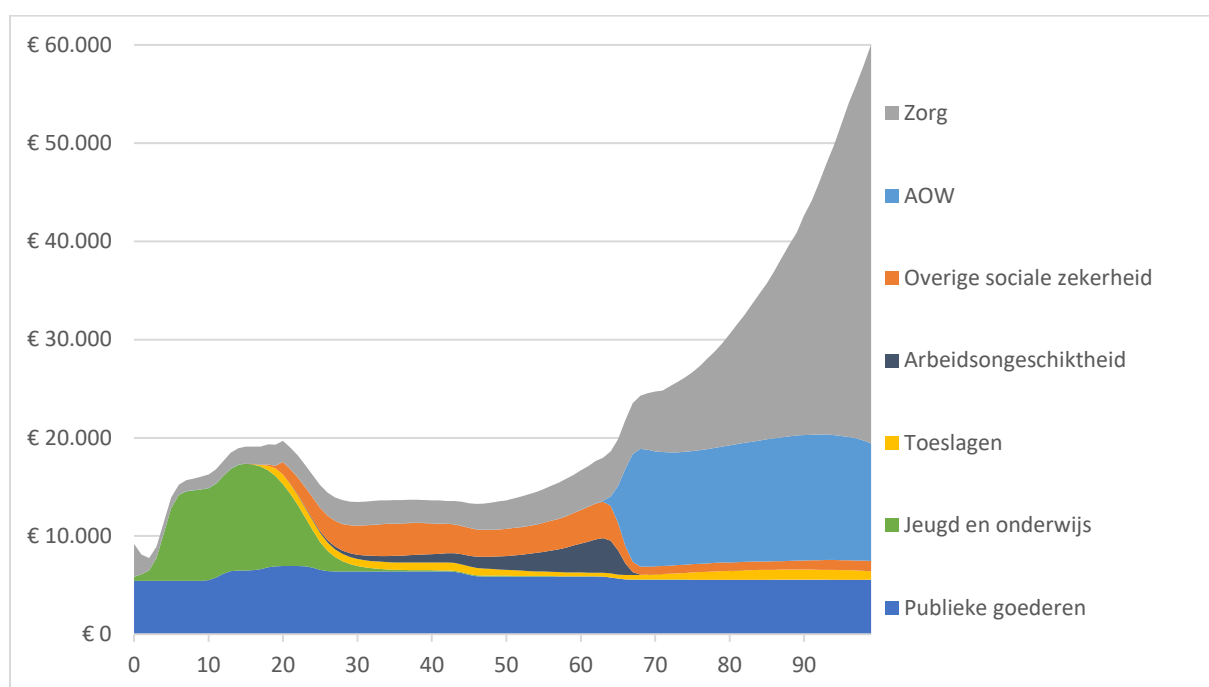
⁵ Voluit 'totale nettobijdrage over de levensloop' maar om schrijf-technische redenen wordt in het huidige rapport kortweg gesproken van nettobijdrage.

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

bedragen op hoge leeftijd en bedragen in de verre toekomst veel minder zwaar mee. In Figuur 2.2 is dit 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).

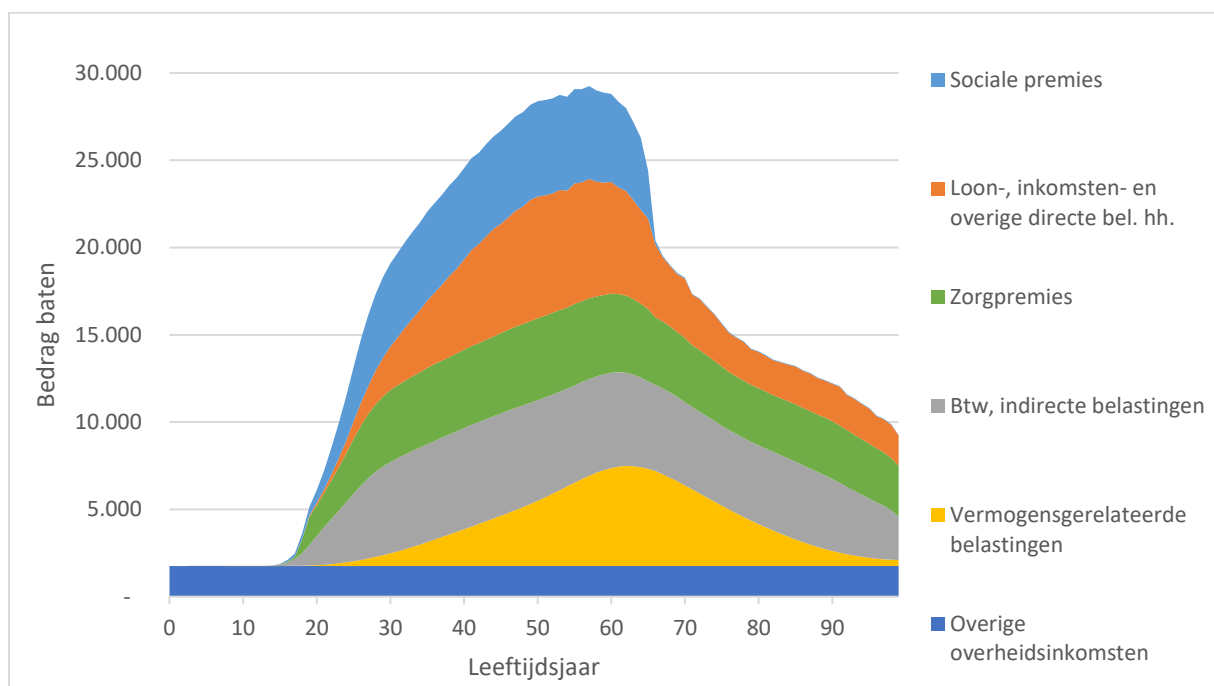
Wat betreft de uitgavenkant geeft Figuur 2.3 inzicht in de belangrijkste posten. Van een aantal posten is aangenomen dat ze voor alle leeftijden gelijk zijn. Veel van deze posten zijn samengebracht onder de noemer publieke goederen. Hieronder vallen openbaar bestuur, defensie, investeringen in gebouwen 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

een euro waarover men in de toekomst kan beschikken. Toekomstige bedragen worden daarom contant gemaakt, d.w.z. uitgedrukt in euro's van 2016 door ze tegen een bepaalde rekenrente af te waarderen. Zie verder de begrippenlijst in het huidige rapport. 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 sterftetekansen 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 sterftetekansen van 2060 gebruikt.

wel leeftijdsafhankelijk en hebben betrekking op de opsporing en eventuele vervolging en bestraffing van verdachten van misdrijven.

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 ten 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 grondverkopen, 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.

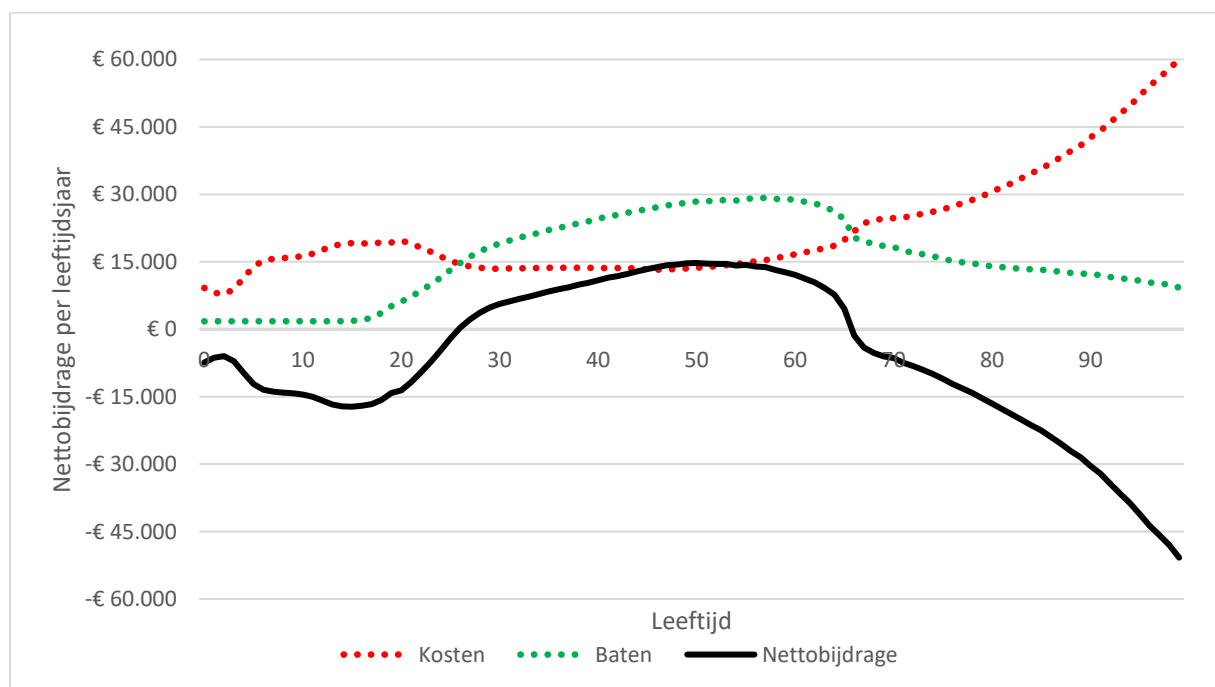
De afdrachten zijn verder duidelijk leeftijdsafhankelijk, wat vooral komt doordat loonbelasting, inkomstenbelasting en sociale premies met name tijdens het werkzame deel van het leven worden afgedragen. Vanaf de pensioenleeftijd zijn de belastingafdrachten en vooral de premieafdrachten beduidend

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

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 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 uit 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.



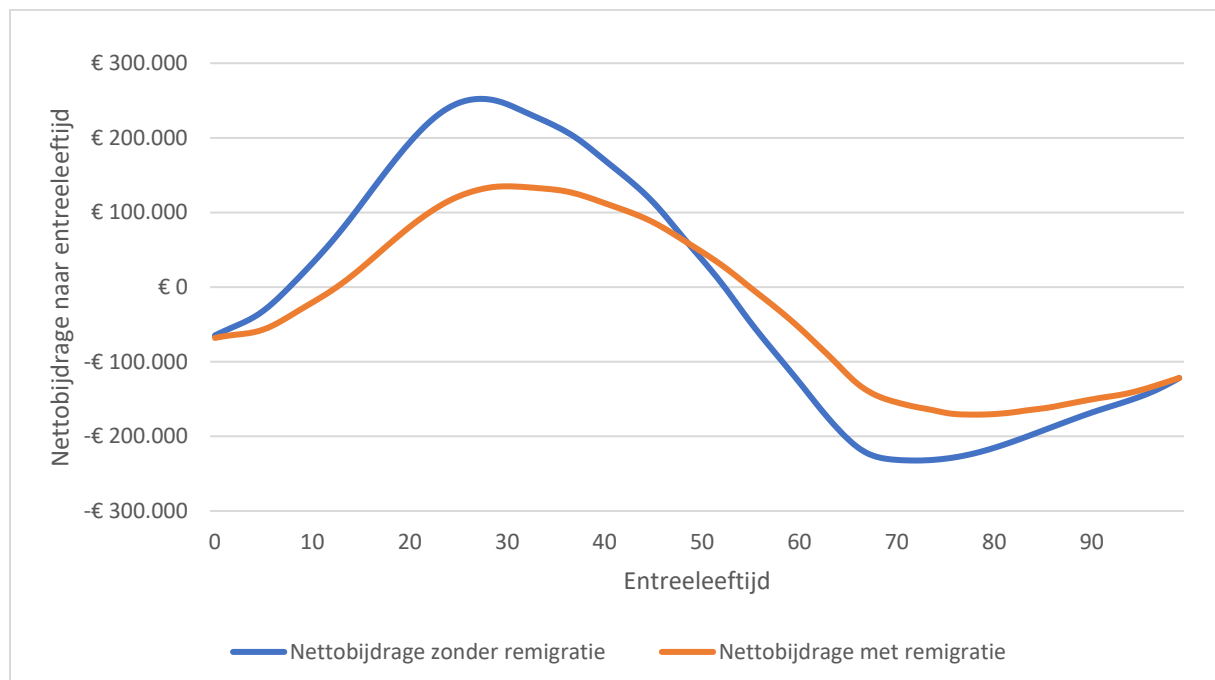
Figuur 2.5 Fiscale kosten en baten en de fiscale nettobijdrage naar leeftijd. Bron: eigen berekening op basis van CBS-microdata.

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 de Nederlandse gezondheidszorg. Daarnaast is het van belang hoe de leeftjidsverdeling 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 het huidige rapport. 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.

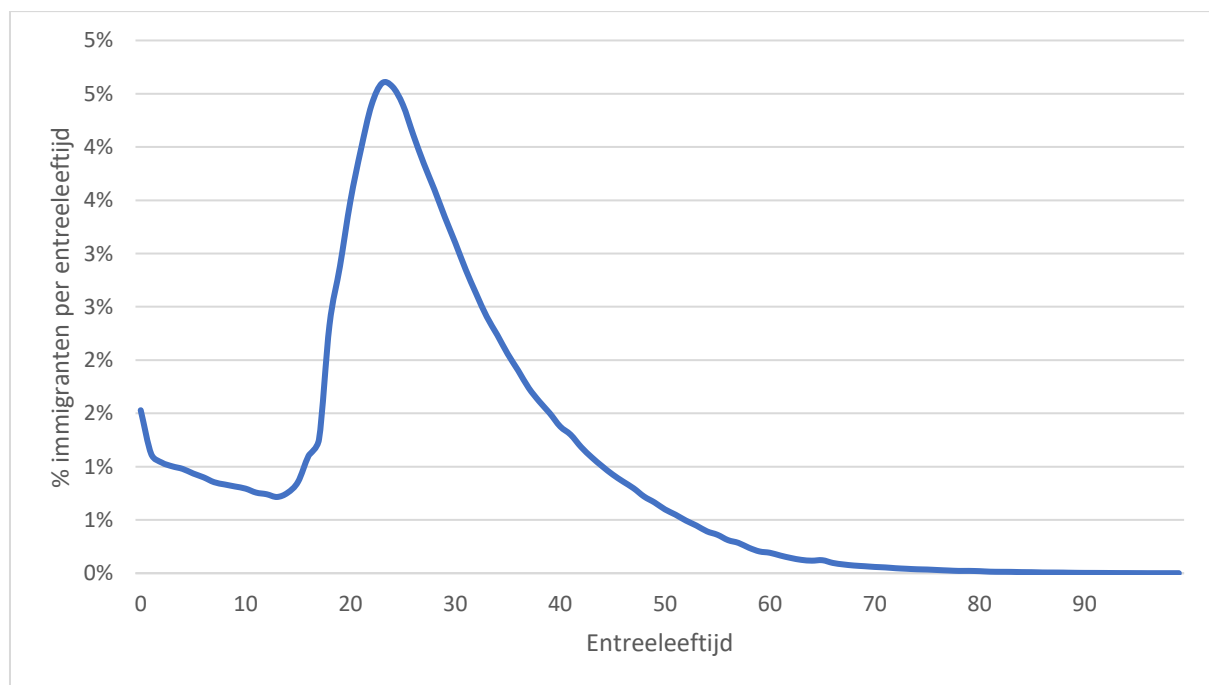
Figuur 2.6 geeft weer hoe de nettobijdrage over de resterende levensloop – dus vanaf het moment van immigratie – afhangt van de entreeleeftijd, dat wil zeggen de leeftijd die de immigrant had op het moment van immigratie.⁹ Hierbij is uitgegaan van een immigrant met de karakteristieken van de gemiddelde 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

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

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.

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 leeftijdsjaar is voor alle leeftijden vanaf 70 jaar negatief. Omdat contant maken en verdiscontering van de sterftekansen 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 vijftientigjarige leeftijd naar Nederland, dan hoeven alle kosten en baten vóór het vijftientigste levensjaar niet meegeteld te worden. Nederland heeft voor die immigrant immers geen kosten gehad voor onderwijs, gezondheidszorg en dergelijke. Voor een vijftientigjarige 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 sterftekansen – 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 vijftientigjarige een positieve waarde van bijna tweeëneenhalf ton).



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

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

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 het huidige rapport. 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

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

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.

Tabel 2.1 Verdeling over grootteklassen per leeftijdsgroep voor 46 groepen met een 2^e generatie migratieachtergrond.

leeftijdsgroep	aantal groepen dat per leeftijdsgroep in grootteklasse valt:						totaal
	0 (geen waarnemingen)	10 tot 25	25 tot 100	100 tot 400	400 tot 1600	meer dan 1600	
0 jaar	0	0	1	13	24	8	46
1 jaar	0	0	1	15	22	8	46
2 jaar	0	0	1	16	21	8	46
3 jaar	0	0	1	16	21	8	46
4 jaar	0	0	1	15	22	8	46
5 jaar	0	0	1	17	20	8	46
6 jaar	0	0	1	17	20	8	46
7 jaar	0	0	1	17	21	7	46
8 jaar	0	0	1	17	21	7	46
9 jaar	0	0	1	18	19	8	46
10 jaar	0	0	1	18	19	8	46
11 jaar	0	0	1	18	19	8	46
12 jaar	0	0	1	18	20	7	46
13 jaar	0	0	1	18	20	7	46
14 jaar	0	0	1	18	20	7	46
15 jaar	0	0	1	19	19	7	46
16 jaar	0	0	2	19	20	5	46
17 jaar	0	0	1	19	21	5	46
18 jaar	0	0	1	23	17	5	46
19 jaar	0	1	2	20	18	5	46
20 tot 22 jaar	0	0	1	9	24	12	46
22 tot 24 jaar	0	0	1	12	22	11	46
24 tot 26 jaar	0	0	2	11	23	10	46
26 tot 28 jaar	0	0	2	13	23	8	46
28 tot 30 jaar	0	0	4	13	23	6	46
30 tot 32 jaar	0	0	4	17	19	6	46
32 tot 34 jaar	0	0	7	16	17	6	46
34 tot 36 jaar	0	1	6	18	15	6	46
36 tot 40 jaar	0	0	6	12	17	11	46
40 tot 44 jaar	0	1	10	12	11	12	46
44 tot 48 jaar	0	2	10	14	10	10	46
48 tot 52 jaar	5	3	8	16	8	6	41
52 tot 56 jaar	14	2	9	9	8	4	32
56 tot 60 jaar	14	3	11	8	6	4	32
60 tot 64 jaar	18	4	8	6	6	4	28
64 tot 68 jaar	18	5	8	5	6	4	28
68 tot 72 jaar	23	4	4	5	7	3	23
72 tot 80 jaar	28	2	3	8	2	3	18
80 tot 90 jaar	31	0	6	6	1	2	15
90 jaar en ouder	39	1	3	1	1	1	7

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 het huidige rapport 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. Een overzicht van de verdeling over de verschillende grootteklassen zijn per leeftijdsgroep gegeven in Tabel 2.1.

Met name niet-westerse immigratie is 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. Immigrantengroepen 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 in het peiljaar 2016 geboren zijn.¹⁷ 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.

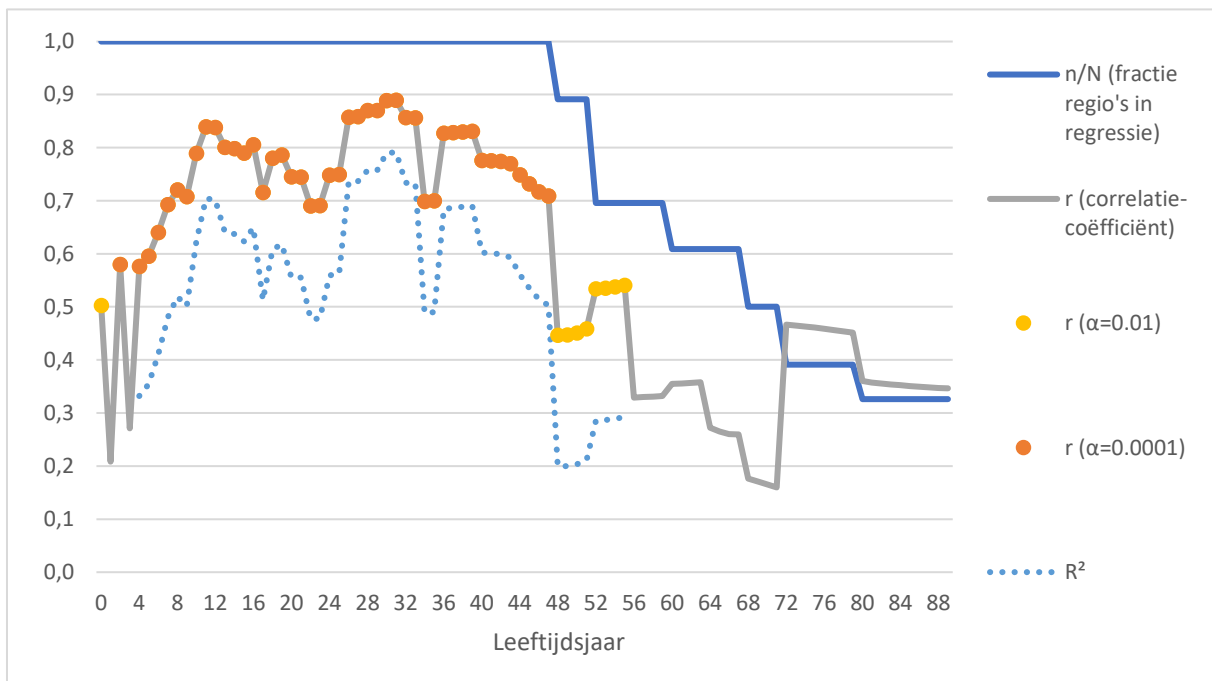
Uit het voorgaande is gebleken dat er vaak onvoldoende data zijn om nettobijdrage profielen te maken voor het hele leven van tweede generatie immigranten. Uiteraard kan men data die er niet zijn, niet maken, maar men kan wel van omliggende gegevens gebruik maken om de leemten zo goed mogelijk in te vullen.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

Bij drie van de vier hierna beschreven methoden spelen citoscores en schoolprestaties een rol bij de ijking van de methode. Zoals beschreven in Hoofdstuk 9 van het huidige rapport is zowel de citoscore als het hoogst behaalde opleidingsniveau erg bepalend voor de nettobijdrage over de levensloop. Een hoger opleidingsniveau betekent gemiddeld gezien ook een hogere nettobijdrage. Hetzelfde geldt voor citoscores, die sterk bepalend zijn voor de hoogst behaalde opleiding en daarmee voor de nettobijdrage. Voor de Nederlandse bevolking als geheel levert elke citopunt hoger bijvoorbeeld ruwweg €

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 En met het oog op de berekeningen in §7.2-§7.3 van het huidige rapport, ook een zo goed mogelijke schatting voor personen met een tweede generatie migratieachtergrond die tussen 1995 en 2019 zijn geboren.

20.000 extra nettobijdrage over de levensloop op. Van deze kennis is gebruik gemaakt om drie van de vier methoden te ijken. Voor de ijking van methode 1 en 2 zijn data gebruikt van citotoetsen afgenomen tussen 2006 en 2018. Omdat de citotoets doorgaans wordt afgenomen op 11 of 12-jarige leeftijd, betreft dit dus personen die geboren zijn in de periode 1994-2007. Voor methode 4 is het middelbare schoolniveau gebruikt van 15-jarigen, over de verslagjaren 2007 tot 2017. Dit betreft dus personen die geboren zijn tussen 1992 en 2002. Dat zijn in beide gevallen geboortejaren die redelijk dicht in de buurt liggen van het peiljaar 2016 dat in het huidige rapport gehanteerd is.¹⁸ Ter vergelijking: de directe waarneming van 46-jarigen betreft personen die omstreeks het jaar 1970 geboren zijn. De aanname onder drie van de vier hier gehanteerde methoden is dat citoscores van het geboortecohort 1994-2007 en het onderwijsniveau van het geboorte cohort 1992-2002 betere voorspellers zijn van de toekomstige nettobijdragen van personen van de tweede generatie die rond het peiljaar 2016 zijn geboren, dan directe waarnemingen van personen met een tweede generatie migratieachtergrond uit cohorten die geboren zijn in de jaren zestig en zeventig van de vorige eeuw.



Figuur 2.8 Pearsons correlatiecoëfficiënten (en de kwadraten R^2 daarvan) voor de samenhang tussen de gemiddelde citoscore voor alle leeftidsjaren samen (2006-2018) en de nettobijdrage per leeftidsjaar afzonderlijk voor 46 tweede generatie groepen, zijnde 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. De fractie n/N staat voor het quotiënt van het aantal groepen n dat is meegenomen in de regressie en het totaal aantal groepen $N = 46$. Significante correlaties op significantieniveau 0.01 en 0.0001 zijn gemarkeerd.

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 tweede generatie van

¹⁸ Daarnaast overlappen de geboortecohorten waarvoor cito en onderwijsgegevens beschikbaar zijn goed met de periode 1995-2019 waarover in §7.2-§7.3 de nettobijdrage berekend is.

de herkomstgroep. Dit is geïllustreerd in Figuur 2.8, waar voor 46 tweede generatie groepen¹⁹ voor elk leeftijdsjaar afzonderlijk de (Pearsons) correlatiecoëfficiënt is gegeven tussen enerzijds de gemiddelde citoscore voor de betreffende groep voor alle leeftijdsjaren samen (m.a.w. berekend over de gehele groep) en anderzijds de gemiddelde nettobijdrage voor de betreffende groep per leeftijdsjaar afzonderlijk.

Het patroon in Figuur 2.8 is als volgt tentatief te verklaren.²⁰ Vanaf de basisschoolleeftijd neemt de correlatie toe omdat onderwijs en zorg dan de grote uitgavenposten zijn en beiden sterk gerelateerd zijn aan citoscores (zie §9.4 van het huidige rapport). Dat blijft zo tot 16 jaar. Van 17 tot 26 jaar zijn er waarschijnlijk verschillende tegengestelde 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 (vergelijk §9.4 van het huidige rapport). 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-delings herkomstregio's (exclusief Nederland) meestal²¹ 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 op extrapolatie²² 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 301 profielen onderzocht: 99 profielen tussen de referentieprofielen van autochtonen en niet-westerse eerste generatie, 101 profielen groter dan of gelijk aan het autochtone profiel en 101 profielen kleiner dan of gelijk aan 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.

¹⁹ Het gaat bij de berekeningen in deze paragraaf om 41 herkomstregio's (de 42-delings minus Nederland), waarbij voor herkomst Indonesië, Turkije, Marokko, Suriname en Aruba en de (voormalige) Nederlandse Antillen is uitgesplitst naar het aantal in het buitenland geboren ouders.

²⁰ De lijnen in deze figuur vertonen een geblokt profiel omdat bij leeftijden vanaf 20 jaar per leeftijdsgroep is gesampeld van 2, 4, 8 of 10 jaar, zie Tabel 2.1.

²¹ Uitzondering zijn de leeftijden van 72 tot en met 79 jaar, waarbij de correlaties zijn berekend voor een klein aantal groepen en *p*-waarden hebben die net onder 5% liggen.

²² 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.

In formulevorm is elk van de extrapolatieprofielen een vector NB^{Ext} met lengte 100 (0 tot 99 jaar) waarvan de elementen gegeven worden door:

$$NB_j^{Ext} = \frac{i}{100} \cdot NB_j^{Aut} + \left[1 - \frac{i}{100}\right] \cdot NB_j^{NW1} \quad \text{voor } 0 \leq j \leq 99 \text{ en } -100 \leq i \leq 200$$

Hierin staat j voor leeftijdjaar en i voor de mate van integratie. Verder staat NB_j^{Aut} voor de nettobijdrage van autochtonen in leeftijdjaar j , NB_j^{NW1} voor de nettobijdrage van de niet-westerse 1^e generatie in leeftijdjaar j en NB_j^{Ext} voor de waarde van het extrapolatieprofiel in leeftijdjaar j .

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 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.



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. De gestippelde lijn betekent dat er weinig data zijn en/of dat er sprake is van een cohorteffect, dit t.b.v. de bespreking van methode 1.

Het bepalen van de mate van integratie is voor de eerste methode schematisch weergegeven in Figuur 2.9 voor een groep personen met een bepaalde tweede generatie migratieachtergrond die we aanduiden als $G2$. 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 de beschikbare data weer van de tweede generatie groep $G2$. Deze data heeft twee potentiële problemen: te weinig waarnemingen per leeftijdsjaar en cohorteffecten. De leeftijdsjaren met weinig waarnemingen zijn aangegeven door een deel van de grijze lijn te stippelen. Het gestippelde deel vormt in deze figuur ook een trendbreuk met de rest van het grijze profiel (hetgeen kan duiden op een cohort-effect) in die zin dat het relatief steil oploopt in verhouding tot het profiel voor autochtonen (blauw) en 1^e generatie niet-westers (oranje). De intuïtie achter methode 1 is om een leeftijdsjaar te bepalen vanaf waar het gestippelde deel van het grijze profiel niet meer wordt meegenomen bij de bepaling van het best passende (gele) extrapolatieprofiel.

De wijze waarop het probleem van weinig waarnemingen wordt opgelost in methode 1 is om leeftijdsjaren met te weinig waarnemingen niet mee te nemen in de berekening. Het is echter niet a priori vast te stellen wanneer er sprake is van ‘te weinig waarnemingen’. Om die reden is er in methode 1 gewerkt met zes zogenaamde samplenormen $SN \in (5, 10, 20, 50, 100, 200)$.²³ Vervolgens is er bij elk van de zes mogelijke waarden van SN een leeftijd L_{SN} gekozen zodanig dat geldt:

$$N_j^{G2} \geq SN \text{ voor } j < L_{SN} \text{ en } N_j^{G2} < SN \text{ voor } j = L_{SN}$$

Hierin is N_j^{G2} het aantal waarnemingen voor leeftijdsjaar j voor de onderhavige tweede generatie groep $G2$.²⁴ In woorden is L_{SN} bij een gegeven samplenorm SN de jongste leeftijd waarvoor het aantal waarnemingen N_j^{G2} kleiner is dan de betreffende samplenorm SN . Vervolgens is L_{SN} gemaximeerd op 56 jaar.

De wijze waarin het probleem van de cohorteffecten wordt opgelost in methode 1 is door het minimaliseren van de afstand tussen het extrapolatieprofiel NB^{Ext} (de gele lijn in Figuur 2.9) en de waarnemingen van de nettobijdrage NB^{G2} voor de betreffende 2^e generatiegroep (de grijze lijn in Figuur 2.9). Die afstand wordt voor een gegeven samplenorm SN geoperationaliseerd met een variant van de kleinste kwadratenmethode²⁵ en wel met de volgende formule:

²³ 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. In het geval van de Voormalige Sovjet-Unie is er sprake van een afwijkende situatie in die zin dat er voor alle leeftijden tot 80 jaar (vaak veel) meer dan 10 waarnemingen zijn per leeftijdsjaar, m.u.v. de leeftijdsjaren tussen 36 en 44 waarvoor slechts 8,25 waarnemingen per leeftijdsjaar zijn en deze jaren zijn bij methode 1 en 2 behandeld alsof het 10 waarnemingen waren teneinde de 320 waarnemingen tussen 44 en 56 ook mee te kunnen nemen in de schatting.

²⁴ 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 $40 / 4 = 10$.

²⁵ De begrenzing op 55 jaar is geïnspireerd op de observatie dat vanaf 56 jaar de correlatie tussen nettobijdragen per leeftijdsjaar met de gemiddelde citoscore (gemeten over alle leeftijdsjaren) sterk afneemt.

$$KS_{SN} = \sqrt{\frac{\sum_{j=4}^{L_{SN}-1} \left(\frac{i}{100} \cdot NB_j^{Aut} + \left[1 - \frac{i}{100} \right] \cdot NB_j^{NW1} - NB_j^{G2} \right)^2}{L_{SN} - 4}}$$

De wortel dient om de termen qua omvang beperkt te houden en de term $L_{SN} - 4$ dient om de kwadratensom KS_{SN} invariant te maken voor de maximale leeftijd $L_{SN} - 1$ waarvoor gesommeerd wordt. De aldus gemeten afstand tussen het gele extrapolatieprofiel en de grijze waarnemingen in Figuur 2.9 wordt gemaximeerd met een maximale kwadratensom KS_{max} .

Het achterliggende idee is dat een cohorteffect zoals de stijging van het gestippelde deel van de grijze lijn in Figuur 2.9 de kwadratensom KS_{SN} niet te groot mag maken. Omdat a priori niet is vast te stellen wanneer de kwadratensom 'te groot' is, is de berekening herhaaldelijk uitgevoerd voor verschillende bovengrenzen KS_{max} . Er zijn 160 bovengrenzen gebruikt, van 400 tot 16.000, aanvankelijk met stapgrootte 400, waarna rond het optimum verder is gezocht met stapgrootte 100. Er zijn dan steeds twee mogelijkheden. Ofwel er geldt dat $KS_{SN} > KS_{max}$ voor alle $-100 \leq i \leq 200$ in welk geval er geen oplossing is, ofwel er is voor de gegeven combinatie van SN en KS_{max} een i waarvoor de kwadratensom KS_{SN} minimaal is. Deze i geeft dan het best passende extrapolatieprofiel voor de betreffende tweede generatiegroep $G2$ bij deze combinatie van SN en KS_{max} .

Deze hele rekenexercitie levert dus voor elke KS_{max} maximaal²⁶ zes oplossingen, immers, er is bij een gegeven KS_{max} voor elke SN ten hoogste één oplossing en er zijn zes verschillende SN te weten 5, 10, 20, 50, 100 en 200. Van de gevonden schattingen zijn vervolgens de hoogste en de laagste schatting verwijderd en is het gemiddelde genomen van de overige schattingen. 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, Marokko, Suriname men de (voormalige) Antillen uitgesplitst naar de subgroepen met één dan wel twee in het buitenland geboren ouders. Deze correlatie was maximaal voor $KS_{max} = 3.000$, $r(44) = .91$, $p \ll .0001$.

Bij de tweede methode wordt net als bij de eerste methode een lineaire combinatie van het nettobijdrageprofiel van autochtonen en het nettobijdrage profiel van de 1^e generatie niet-westers gefit op de waarnemingen NB^{G2} voor de betreffende tweede generatiegroep $G2$. Dit gebeurt eveneens met behulp van een (andere) variant van de kleinste kwadratenmethode. Ook moeten dezelfde twee problemen worden opgelost: er zijn voor sommige groepen cohorteffecten en voor oudere leeftijdsjaren zijn er soms te weinig waarnemingen.

Het probleem van een te klein aantal waarnemingen is in methode 2 in twee stappen opgelost. Allereerst is er een leeftijd L gekozen zodanig dat geldt:

$$N_j^{G2} \geq 10 \text{ voor } j < L \quad \text{en} \quad N_j^{G2} < 10 \text{ voor } j = L$$

²⁶ In de praktijk kunnen niet altijd alle zes schattingen gemaakt worden en in een enkel geval waren er drie schattingen.

Hierin is N_j^{G2} weer het aantal waarnemingen voor leeftijdjaar j voor de onderhavige tweede generatie groep $G2$. In woorden is L de jongste leeftijd waarvoor het aantal waarnemingen kleiner is dan 10. Ook bij methode 2 is L gemaximeerd op 56.

Daarnaast is ook het gemiddeld aantal waarnemingen N_j^{G2} per leeftijdjaar direct meegewogen in de kwadratensom KS (zie formule hieronder). Als N_j^{G2} 100 of groter was, is gewicht 1 gehanteerd en als N_j^{G2} onder 100 lag is gewicht $N_j^{G2}/100$ gebruikt. Zo is gewaarborgd dat leeftijdjaren met een klein aantal waarnemingen geen al te grote invloed hebben op de kwadratensom KS .

Het probleem van eventuele cohorteffecten is in methode 2 opgelost door het direct meewegen in de kwadratensom van de verklaarde varianties R_j^2 (zie de blauwe lijn in Figuur 2.8) van de lineaire regressies van de nettobijdrage per leeftijdjaar j op de groepsgemiddelden van de citoscores (alle leeftijdjaren samen) voor de betreffende tweede generatie herkomstgroep. Het achterliggende idee is dat de gemiddelde citoscores zó bepalend zijn voor de toekomstige nettobijdrage van groepen, dat lagere correlaties met de citoscore voor dertigplussers een indicatie zijn van cohorteffecten.

Dit alles samen geeft de volgende formule voor de kwadratensom:

$$KS = \sum_{j=0}^{L-1} R_j^2 \cdot \min\left(1, \frac{N_j^{G2}}{100}\right) \cdot \left(\frac{i}{100} \cdot NB_j^{Aut} + \left[1 - \frac{i}{100}\right] \cdot NB_j^{NW1} - NB_j^{G2}\right)^2$$

Wederom is de i waarvoor de kwadratensom minimaal is genomen als maat voor de integratie voor de betreffende groep. Ook deze tweede methode zorgt voor een sterke correlatie met de gemiddelde citoscores van de 46 tweede generatie groepen, $r(44) = .90$, $p \ll .0001$.²⁷

De derde methode is vergelijkbaar met de tweede methode, met enkele verschillen. In de eerste plaats is R_j^2 niet meegewogen in de kwadratensom. In de tweede plaats is geen minimum van 10 waarnemingen per leeftijdjaar gesteld en voor de onderlinge vergelijkbaarheid is daarom gekozen om alleen leeftijdjaren tot 48 jaar mee te nemen, omdat van alle groepen tot tenminste 48 jaar data zijn. In deze methode is de kwadratensom KS dus alleen gewogen met het gemiddeld aantal personen N_j^{G2} per leeftijdjaar, zoals beschreven bij methode 2, met dien verstande dat een striktere norm is aangehouden: Als N_j^{G2} 400 of groter was, is gewicht 1 gehanteerd en als N_j^{G2} onder 400 lag is gewicht $N_j^{G2}/400$ gebruikt. Van deze methode zijn twee varianten uitgerekend. De eerste variant voor leeftijden van 0 tot 48 jaar en de tweede variant voor leeftijden van 25 tot 48 jaar. Van beide varianten is het gemiddelde genomen. De rationale hierachter is om gevoeligheid te verminderen voor verschillen in studiekosten (denk aan 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. Nadeel is echter dat eventuele cohorteffecten ook relatief zwaarder wegen.

Dit alles samen geeft de volgende formule voor de kwadratensom (van beide varianten, d.w.z. vanaf 0 jaar en vanaf 25 jaar):

²⁷ Wederom voor de 42-delings minus Nederland, met de vijf grote herkomstlanden Indonesië, Turkije, Marokko, Suriname en de (voormalige) Antillen uitgesplitst naar de subgroepen met één dan wel twee in het buitenland geboren ouders.

$$KS = \sum_{j=0 \text{ of } j=25}^{47} \min\left(1, \frac{N_j^{G2}}{400}\right) \cdot \left(\frac{i}{100} \cdot NB_j^{Aut} + \left[1 - \frac{i}{100}\right] \cdot NB_j^{NW1} - NB_j^{G2}\right)^2$$

Ook hier is de i waarvoor de kwadratensom KS minimaal is genomen als maat voor de integratie voor de betreffende groep.

Ook deze methode heeft een sterke correlatie met de gemiddelde citoscores van de 46 tweede generatie groepen, $r(44) = .90$, $p \ll .0001$.²⁸ 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 onderstreept nog eens dat de correlatie tussen integratiemaat en citoscore bij de eerste twee methoden geen artefact is van hetijken op de citoscore. Opvallend is verder dat berekening over 25-47 jaar een iets sterkere correlatie geeft ($r(44) = .90$, $p \ll .0001$) dan berekening over 0-47 jaar ($r(44) = .89$, $p \ll .0001$). Over het algemeen liggen de scores voor berekening over 0-47 en 25-47 jaar dicht bij elkaar. Voor een aantal groepen valt berekening over 25-47 jaar aanzienlijk hoger uit en de verschillen zijn vooral terug te voeren op cohorteffecten, maar toch is berekening over 25-47 jaar meegenomen om aan de 'voorzichtig positieve kant' te blijven.

Tot slot is er een vierde methode gebruikt. Voor de vierde methode zijn voor leeftijden L tussen 20 en 48 jaar de sommen $S_L^{G2} = \sum_{j=0}^L NB_j^{G2}$ bepaald door de nettobijdragen NB_j^{G2} van de betreffende tweede generatiegroep $G2$ op te tellen van 0 tot L . Op vergelijkbare wijze zijn voor autochtonen de sommen $S_L^{Aut} = \sum_{j=0}^L NB_j^{Aut}$ bepaald en evenzo voor de niet-westerse 1^e generatie²⁹ de sommen $S_L^{NW1} = \sum_{j=0}^L NB_j^{NW1}$. De mate van integratie voor een gegeven L is vervolgens berekend als een verhoudingsgetal:

$$i = 100 \cdot \frac{(S_L^{G2} - S_L^{NW1})}{(S_L^{Aut} - S_L^{NW1})}$$

Voor de onderlinge vergelijkbaarheid is net als bij methode 3 gekozen voor leeftijden tot 48 jaar, omdat van alle groepen tot tenminste 48 jaar data zijn. De maximale waarde van L is dus 47 jaar. Merk op dat deze methode fundamenteel afwijkt van de eerste drie methoden, omdat er niet wordt gewerkt met interpolatie van referentieprofielen.

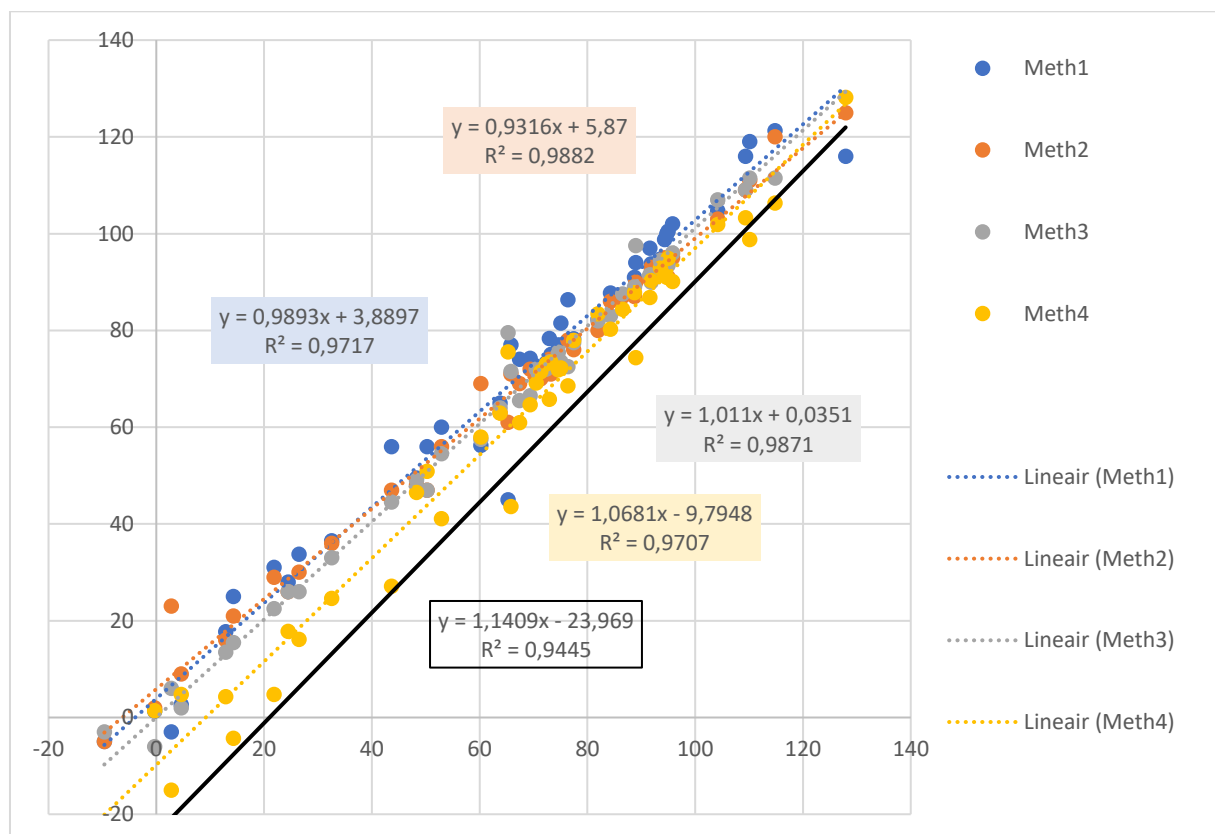
Bij methode 4 hangt i af van L (m.a.w. i is een functie van L). Het is niet *a priori* vast te stellen welke $20 \leq L < 47$ de beste integratiemaat i oplevert. Zeker voor veertigers is er soms sprake van kleine aantallen waarnemingen en daarom is alle data tot 48 jaar meenemen waarschijnlijk niet de beste keuze. Omdat niet zonder meer is te beslissen welke L de beste integratiemaat oplevert, is bij deze methode 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 tussen voornoemd verschil en de integratiemaat i is als grens aangehouden en dat blijkt het geval te zijn voor $L = 41$ jaar ($r(44) = .90$, $p \ll .0001$) (ijking op de citoscores geeft overigens ook $L = 41$ jaar, $r(44) = .90$, $p \ll .0001$). Tot 36 jaar zijn er drie groepen met kleine N: Japan, Voormalige Sovjetunie en Hoorn van Afrika en

²⁸ Zie vorige voetnoot.

²⁹ Daarbij is bij de 1^e generatie niet-westers de nettobijdrage voor de leeftijden 0 en 1 gelijkgesteld aan de nettobijdrage voor autochtonen van 0 respectievelijk 1 jaar, om zodoende rekening te houden met de kosten voor geboortezorg e.d.

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.

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.



Figuur 2.10 Samenhang tussen vier methoden om de mate van integratie van de tweede generatie te schatten voor 46 tweede generatie groepen. 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. De waarden op de horizontale as zijn de gemiddelden van methode 1 t/m 4. De dikke zwarte lijn is geeft de regressievergelijking van een schatting van de mate van integratie o.b.v. de directe waarnemingen van de nettobijdrage over de eerste 35 leeftijdjaren.

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 leeftijdsjaar 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 leeftijdsjaar 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 leerlinggewichten, criminaliteitscijfers en het aandeel niet-EER studenten wellicht weer zwaarder meewegen. Vandaar ook dat bij methode 4 gekozen is voor verankering in het onderwijs niveaus 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 leeftijdsjaren 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 leeftijdsjaren 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 2.2. 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 2.2 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	Citoscore
Gem. 1-4	1					
Methode 1	0,99	1				
Methode 2	0,99	0,98	1			
Methode 3	0,99	0,97	0,99	1		
Methode 4	0,99	0,95	0,97	0,98	1	
Citoscore	0,91	0,91	0,90	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.³⁰

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 gekijkt 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 het huidige rapport – 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 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-absolue 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

³⁰ 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.

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.

2.5 Gevoeligheidsanalyse

Uiteraard zijn de gemaakte berekeningen gevoelig voor aannamen rondom de operationalisering van variabelen. In §6.5 van het huidige rapport is voor enkele variabelen een gevoeligheidsanalyse gemaakt. Allereerst is inzichtelijk gemaakt dat de mate van verfijning van de indeling van de wereld in regio's (zie §4.4 van deze appendix) verschillen tussen landen meer of minder zichtbaar kunnen maken. Deze analyse behoeft geen nadere toelichting.

Verder is in §6.5 van het huidige rapport het effect inzichtelijk gemaakt van de discontovoet, door de berekeningen ook uit te voeren met een discontovoet van 1,5% (dus 1% lager dan de discontovoet van 2,5% in het standaardscenario). Dit is gerealiseerd door in het model de reële interest i dienovereenkomstig te verlagen (zie Hoofdstuk 12 van deze appendix; voor meer toelichting op de discontovoet zie Hoofdstuk 8 van deze appendix). Voorts is in §6.5 van het huidige rapport het effect zichtbaar gemaakt van een AOW-leeftijd van 65 jaar. Een korte toelichting op deze berekening wordt gegeven §8.1 van deze appendix.

Tabel 2.3 Toerekening van de kosten voor publieke goederen aan personen in een gemengd scenario (miljarden euro's).

Post nr.	Post	MACROBEDRAG	Toerekening naar rato:	
			Bbp	Bevolking
	TOTAAL UITGAVEN*	307,5		
	Betaalde rente* (niet aan personen toegerekend)	7,6		
	Aan personen toegerekend*	299,9		
1	Openbaar bestuur*	64,9		
	Af: veiligheidszorg (criminaliteit, politie, justitie)**	10,0		
	Bij: fundamenteel onderzoek (af van onderwijs)**	3,7 ±		
	Aangepaste post Openbaar bestuur **	58,2		
	Aangepaste post Openbaar bestuur**	58,2	20%	80%
2	Defensie*	6,9	100%	0%
10	Overdrachten buitenland*	10,5	100%	0%
13	Bruto inv. gebouwen*	8,5	20%	80%
14	Bruto inv. infrastructuur*	10,1	20%	80%
	TOTAAL	94,1	32,7	61,4
	Verdeling in %	100%	35%	65%

*Macrobedragen begroting 2016 volgens CPB2018-dataset (zie Hoofdstuk 8 van deze appendix). Post nr. en 'TOTAAL UITGAVEN' verwijzen naar Tabel 5.1. Alle met * gemarkeerde bedragen zijn ook gehanteerd in het huidige rapport.

**Macrobedragen gehanteerd in het huidige rapport die afwijken van de CPB2018-dataset macrobedragen.

Tot slot is in §4.2 en §6.5 van het huidige rapport ingegaan op een belangrijk verschil met de berekeningen in de CPB-studie *Immigration and the Dutch Economy* uit 2003. In dat CPB-rapport werden de kosten voor publieke goederen als openbaar bestuur, defensie en dergelijk, naar personen toegerekend naar rato van de mate waarin deze personen bijdroegen aan het bruto binnenlands product³¹. In

³¹ Roodenburg, H., R. Euwals & H. ter Rele (2003), blz. 67

het huidige rapport worden dergelijke publieke goederen aan iedere inwoner van Nederland in gelijke mate toegerekend en daarmee worden de kosten naar rato van de omvang bevolking toegerekend. Dit is conform de CPB-vergrijzingsstudie *Minder zorg om vergrijzing* uit 2014 en de CPB2018-dataset (zie ook §8.1) waar het huidige rapport zich grotendeels op baseert.³² Dat geeft uiteraard verschillende uitkomsten. In de gevoeligheidsanalyse in §6.5 van het huidige rapport zijn daarom twee varianten op het basisscenario doorgerekend.

De eerste variant op het basisscenario – 35% bbp-gerelateerd – gaat ervan uit dat publieke goederen voor 35% worden toegerekend naar rato van bbp en voor de overige 65% naar rato van de bevolking. Onder publieke goederen wordt verstaan de posten Defensie, Overdrachten buitenland, Bruto investeringen Gebouwen, Bruto investeringen infrastructuur en de Aangepaste post Openbaar bestuur. De Aangepaste post Openbaar bestuur is de post Openbaar bestuur uit Tabel 5.1 exclusief de kosten voor Veiligheidszorg (criminaliteit, politie en justitie) en inclusief de kosten voor fundamenteel onderzoek, dat in Tabel 5.1 onder de post Onderwijs valt (zie voor details van de berekening in Tabel 2.3).

De berekening voor de variant 35% bbp-gerelateerd is als volgt. In *Immigration and the Dutch Economy* uit 2003 werden de bbp-gerelateerde kosten toegerekend naar rato van het product van het loon relatief t.o.v. autochtonen en de arbeidsparticipatie relatief t.o.v. autochtonen.³³ In het huidige rapport is als proxy voor de bijdrage aan het bbp het persoonlijk primair inkomen (PPI) genomen, berekend op basis van CBS-microdata. Er is uitgegaan van het voor leeftijd gecorrigeerde gemiddelde van de PPI over de leeftijden 16-68 jaar. Dit PPI is uitgedrukt als het percentage $\%PPI$ van het PPI van autochtonen, zie kolom 2 in Tabel 6.4 van het huidige rapport. Vervolgens zijn voor elke groep in de 42-deling de totale kosten PG voor publieke goederen berekend. Als tot slot NB de nettobijdrage is van een bepaalde groep, dan is het bedrag NB' volgens de variant 35% bbp-gerelateerd gelijk aan $NB' = NB + 0,35 \times PG \times (\%PPI - 1)$.

De rationale achter de keuze voor 35% is als volgt. Bepaalde overheidsuitgaven zijn inderdaad zoals gesteld in *Immigration and the Dutch Economy* meer gerelateerd aan het bbp en niet zozeer aan de bevolking. Denk aan verplichtingen zoals EU-afdrachten en internationale afspraken zoals ontwikkelingssamenwerking en defensie, die zijn uitgedrukt als percentage van het bbp. De meeste overige uitgaven volgen naar onze inschatting echter in vrij grote mate de ontwikkeling van de bevolking. Vandaar de keuze om (zoals hiervoor betoogd) de aanpak van de CPB-vergrijzingsstudie *Minder zorg om vergrijzing* uit 2014 en de CPB2018 dataset te volgen. De redenering is als volgt. Het totale macrobedrag voor de betrokken posten bedraagt 94,1 miljard euro (zie Tabel 2.3). Daarvan heeft 17,4 miljard (18,5%) betrekking op de posten Defensie en Overdrachten buitenland. Van deze posten nemen we aan dat ze 100% bbp-gerelateerd zijn. De overige posten hebben betrekking op Bruto investeringen Gebouwen, Bruto investeringen infrastructuur en de Aangepaste post Openbaar bestuur. De Aangepaste post Openbaar bestuur is waarschijnlijk grotendeels afhankelijk van schaalgrootte, omdat de kosten voor een groot deel bestaan uit personeelskosten, bijvoorbeeld voor ambtenaren die werken in allerlei uitvoerende diensten – zoals de belastingdienst – die qua omvang grotendeels afhangen van

³² “De overige uitgaven, zoals openbaar bestuur en defensie, vormen een vaste fractie van het bbp. De baten ervan worden, bij gebrek aan informatie, voor alle burgers gelijk verondersteld.” Smid, B., H. ter Rele, S. Boeters, N. Draper, A. Nibbelink en B. Wouterse (2014), opgehaald 17-1-2022 van: <https://www.cpb.nl/sites/default/files/publicaties/download/cpb-boek-12-minder-zorg-om-vergrijzing.pdf> blz. 31 en de gebruikte CPB2018-dataset (zie hoofdstuk 8 van deze appendix), waarin deze posten voor alle inwoners gelijk zijn verondersteld.

³³ Roodenburg, H., R. Euwals & H. ter Rele (2003), blz. 64-67

de bevolkingsomvang. De kosten voor bruto investeringen in gebouwen zijn deels gerelateerd aan de omvang van het ambtenarenapparaat en daarmee indirect aan de bevolkingsomvang. De kosten voor infrastructuur zouden in een dichtbevolkt land als Nederland door congestieverschijnselen wel eens meer dan evenredig (d.w.z. progressief) kunnen samenhangen met bevolkingsomvang. Het voert te ver om voor deze gevoeligheidsanalyse voor elke post een functie te schatten van de mate van afhankelijkheid van het bbp en/of de bevolkingsomvang. In plaats daarvan is een rekenvoorbeeld gegeven waarin Defensie en Overdrachten buitenland 100% bbp-gerelateerd zijn en de overige posten 20% bbp-gerelateerd. Dit geeft 35% bbp-gerelateerde kosten voor publieke goederen, zie voor details van de berekening in Tabel 2.3. De resulterende variant 35% bbp-gerelateerd geeft een indruk van de nettobijdragen als een kleiner deel van de kosten voor publieke goederen bbp-gerelateerd zijn.

Over de tweede variant kunnen we kort zijn. In de hoofdtekst van het huidige rapport wordt toegelicht dat om uiteenlopende redenen veel deelposten van de post Publieke goederen wel eens onevenredig sterk zouden kunnen toenemen met bevolkingsgroei door immigratie. Voor deze variant is uitgegaan van een op zich arbitraire stijging van de kosten voor publieke goederen met 20%. Het effect hiervan op de nettobijdrage is min of meer lineair met het gekozen percentage, dus de lezer kan zelf door interpolatie of extrapolatie de effecten van grotere waarden doorrekenen.

3 Onderzoekspopulatie en microdatabestanden

Voor de berekeningen in het huidige rapport is zoveel mogelijk gebruik gemaakt van CBS-microdata. Dat zijn zeer gedetailleerde en geanonimiseerde data, die op persoonsniveau voor (vrijwel) alle inwoners van Nederland beschikbaar zijn. De CBS-microdata database bestaat uit een groot aantal tabellen, die onderling met elkaar verbonden kunnen worden via sleutelvariabelen. In deze 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.³⁴

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,

³⁴ 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 mogen 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, opgehaald 12-1-2023 van: <https://www.cbs.nl/nl-nl/onze-diensten/maatwerk-en-microdata/microdata-zelf-onderzoek-doen/catalogus-microdata>

GBAMIGRATIEBUS en VRLMIGMOTBUS voor migratie en migratiemotieven en tenslotte het pensioen-aanspraken bestand PAS.

Voor personen die in 2016 zijn geboren of in 2016 als 0-jarige zijn 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”.³⁵ De derde generatie maakt deel uit van de groep mensen met een Nederlandse achtergrond, welke gedefinieerd is als: “Persoon van wie de beide ouders in Nederland zijn geboren, ongeacht het land waar men zelf is geboren”.³⁶ 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”.³⁷ 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) Opgehaald 19-4-2023 van: <https://www.cbs.nl/nl-nl/nieuws/2016/47/wie-zijn-de-derde-generatie>. 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

³⁵ Opgehaald 19-4-2023 van: <https://www.cbs.nl/nl-nl/nieuws/2016/47/wie-zijn-de-derde-generatie>

³⁶ Opgehaald 19-4-2023 van: <https://www.cbs.nl/nl-nl/onze-diensten/methoden/begrippen?tab=p#id=persoon-met-een-nederlandse-achtergrond>

³⁷ Opgehaald 19-4-2023 van: <https://www.cbs.nl/nl-nl/onze-diensten/methoden/begrippen?tab=p#id=persoon-met-een-tweede-generatie-migratieachtergrond>

onbekende achtergrond heeft. Dan nemen we de achtergrond van de vader".³⁸ 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 KIN-DOUDERTAB 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.³⁹ Er zijn overigens weinig publieke data beschikbaar voor de derde generatie, behoudens een enkel CBS-statline bestand⁴⁰ of CBS-maatwerk bestand.⁴¹

4.2 Opleidingsniveau en citoscores

Opleidingsniveau is afgeleid uit het CBS-microdatabestand HOOGSTEOPL voor de verslagjaren 2007-2017. Dit bestand geeft onder andere het hoogst behaalde opleidingsniveau per 1 oktober van het verslagjaar. Voor de hoogst behaalde opleiding zijn in totaal waarnemingen gebruikt van 6,9 miljoen personen.

Voor een aantal deelberekeningen is daarnaast ook gebruikt gemaakt van het CBS-microdatabestand ONDERWIJSDEELNEMERSTAB voor de verslagjaren 2007-2017 om de actueel gevolgde opleiding te bepalen. Voor de actueel gevolgde opleiding zijn gegevens gebruikt van 2,7 miljoen personen.

De indeling van de onderwijsniveaus in het verslag is gebaseerd op de SOI van het CBS. SOI staat voor Standaard Onderwijs Indeling (zie voor meer informatie de term Standaard Onderwijs Indeling in de Begrippenlijst in het huidige rapport). Meestal worden de zogenaamde SOI 5-deling en 8-deling gebruikt.

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.⁴²

³⁸ Opgehaald 19-4-2023 van: <https://www.cbs.nl/nl-nl/nieuws/2016/47/wie-zijn-de-derde-generatie->

³⁹ Vergelijk met de toelichting bij de CBS-dataset Tabellen Derde Generatie over de derde generatie, opgehaald 19-4-2023 van: <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->

⁴⁰ Opgehaald 19-4-2023 van: <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/70688ned/table?dl=18BAC>

⁴¹ Voor 2016 en 2020 zijn er maatwerkbestanden zie voorgaande noten en zie ook volgende tabel opgehaald 19-4-2023 van: <https://www.cbs.nl/nl-nl/maatwerk/2020/46/personen-met-ouders-van-tweede-generatie-1-januari-2020>

⁴² 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.

Deze gegevens voor citoscores en hoogst behaalde en actueel gevolgde opleiding zijn met name gebruikt in Hoofdstuk 9 van het huidige rapport. In dit hoofdstuk zijn op hoofdlijnen vijf typen berekeningen gemaakt o.b.v. CBS-microdata: (I) berekeningen m.b.t. groepsgemiddelden van opleidingsniveau en/of citoscore, (II) berekeningen van nettobijdrage naar opleidingsniveau, (III) berekeningen van nettobijdrage naar citoscore. Hierbij is doorgaans uitgesplitst naar generatie, herkomstregio en/of migratiemotief, (IV) berekeningen van de nettobijdrage van de 2^e generatie, zoals weergegeven in Tabel 9.4 van het huidige rapport en (V) indicatieve berekeningen om te bepalen in welke mate groepsverschillen in nettobijdrage ontstaan door groepsverschillen in opleidingsniveau op het moment van immigratie, dan wel groepsverschillen in citorendement, dan wel groepsverschillen in opleidingsrendement (zie voor de genoemde begrippen eventueel de Begrippenlijst).

(Ad I) De berekeningen m.b.t. groepsgemiddelden van opleidingsniveau en citoscore zijn als volgt uitgevoerd. Voor de citoscore is daarbij gemiddeld over de jaren 2006-2018. Voor de actuele opleiding (dus degenen die actief studeren) is gemiddeld over de jaren 2007-2017. Voor de hoogst behaalde opleiding zijn gegevens voor 2016 gebruikt. Het gevolg hiervan is dat bij citoscores en actueel gevolgde opleiding de getallen voor de 1^e, 2^e en 3^e generatie over het algemeen betrekking hebben op verschillende cohorten. Het is dus doorgaans niet zo dat een groep waarvoor een gemiddelde is berekend voor de 2^e generatie daadwerkelijk de kinderen betreft van de groep waarvoor een gemiddelde is berekend voor de 1^e generatie. Concreet voorbeeld: de gemiddelde citoscore voor de Turkse 2^e generatie is gebaseerd op de gegevens van een groep personen die in de praktijk nooit de kinderen kunnen zijn van de groep personen waarop de gemiddelde citoscore voor de Turkse 1^e generatie is gebaseerd. Dit is bijvoorbeeld van belang bij de interpretatie van Figuur 9.20, 9.21, 9.23 en 9.24 uit het huidige rapport. Voor actueel gevolgde opleiding is de kans dat een deel van de 2^e generatie daadwerkelijk de kinderen zijn van de 1^e generatie iets groter, omdat zowel kinderen als volwassenen onderwijs volgen, al zal ook hier de kans klein zijn (weinig volwassen volgen bijvoorbeeld een havo-opleiding). Dit komt omdat citoscore en actueel gevolgde opleiding momentopnamen zijn en de periode waarover waarnemingen beschikbaar zijn te kort is. De hoogst behaalde opleiding betreft daarentegen mensen van alle leeftijden waarvoor data beschikbaar is. Hier is de kans bijvoorbeeld veel groter dat de groep personen waarop het gemiddelde opleidingsniveau voor de Duitse 2^e generatie migratieachtergrond is gebaseerd daadwerkelijk deels de kinderen zijn van de groep personen waarop het gemiddelde opleidingsniveau voor de Duitse 1^e generatie migratieachtergrond is gebaseerd. Hetzelfde geldt voor combinaties van gemiddeld opleidingsniveau met gemiddelde citoscore zoals in Figuur 9.11 uit het huidige rapport.

(Ad II) Verder zijn berekeningen gemaakt voor de nettobijdrage naar opleidingsniveau voor de hele bevolking en uitgesplitst naar migratieachtergrond (westers en niet-westers). Hiertoe zijn profielen geconstrueerd (zoals Figuur 9.6 uit het huidige rapport) voor de nettobijdrage per leeftijdsjaar, uitgesplitst naar opleidingsniveau. 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'.⁴³ Dit is als volgt gedaan.

⁴³ Er zijn totaal dus 16 nettobijdrageprofielen, twee voor elk niveau in de SOI 8-deling. Concreet: er is voor het niveau mbo4 een profiel voor de actueel gevolgde opleiding (ACT-profiel) voor studenten die staan ingeschreven bij een mbo en een profiel voor de hoogst behaalde opleiding (HB-profiel) voor niet-studierenden die mbo4 als hoogst behaalde opleiding hebben.

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 citoscore⁴⁴ van de uiteindelijke hoogst behaalde opleiding. Voor de leeftijden van 12 tot maximaal 23 jaar is in beginsel aangenomen 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 studiepadi 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 studiepadi 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 studiepadi 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 het nettobijdrageprofiel gebaseerd op de gemiddelde citoscore bij de betreffende hoogst behaalde opleiding.⁴⁵

Deze twee nominale studiepadi 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 21⁴⁶ tot en met 38 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, zoals die zijn geobserveerd o.b.v. CBS-microdata. 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.⁴⁷ Voor leeftijden vanaf 39 jaar is altijd alleen het HB-profiel bij de betreffende hoogst behaalde opleiding gebruikt. Het kleine aantal mensen van 39 jaar en ouder dat nog ingeschreven staat voor een opleiding op niveau *s*, wordt verondersteld om dezelfde nettobijdrage te hebben als mensen die *s* als hoogst behaalde opleiding hebben. 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

⁴⁴ De nettobijdrage hangt mede af van de citoscore, zie onder (Ad III).

⁴⁵ Deze gemiddelde citoscore naar hoogst behaalde opleiding is verkregen uit matrixberekeningen die onder (Ad III) worden besproken. 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 betreffende gemiddelde citoscore en de twee scores direct erboven en eronder.

⁴⁶ Bij hbo-, wo-master vanaf 22 jaar.

⁴⁷ 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.

uitsplitsing naar migratieachtergrond met gewicht 0,25 is meegewogen in de kwadratensom.⁴⁸ Verder zijn op basis van CBS-maatwerkdata de sterftekansen aangepast aan het opleidingsniveau.⁴⁹

(Ad III) Daarnaast zijn er in Hoofdstuk 9 van het huidige rapport enkele berekeningen gemaakt voor de nettobijdrage per citoscore (zie Figuur 9.3, 9.8, 9.9 en 9.18). Voor deze berekeningen zijn longitudinale data gebruikt, te weten voor de citoscore over de periode 2006-2018 en voor de actuele en hoogst behaalde opleiding over de periode 2007-2017. Deze perioden zijn te kort voor een berekening in één keer. Daarom zijn de gegevens van verschillende cohorten gecombineerd. De genoemde figuren zijn dan ook synthetisch (zie het gelijknamige begrip in de Begrippenlijst van het huidige rapport), d.w.z. samengesteld uit verschillende componenten. Dit is op hoofdlijnen als volgt gedaan:

- i. 0 tot 4 jaar: o.b.v. observaties van de nettobijdrage voor de hele bevolking.
- ii. 4 tot 12 jaar: o.b.v. observaties van de nettobijdrage per citoscore.
- iii. 12 tot 21 jaar: o.b.v. observaties van de nettobijdrage per citoscore.
- iv. 21 tot 39 jaar: o.b.v. observaties van de nettobijdrage per opleidingsniveau i.c.m. transformatiematrices die tot 39 jaar per citoscore de verdeling over opleidingsniveaus geven.
- v. 39 jaar en ouder: o.b.v. observaties van de nettobijdrage per opleidingsniveau i.c.m. de voor 38-jarigen geobserveerde verdeling over opleidingsniveaus per citoscore.

Tabel 4.1 Samenstelling citocohorten naar leeftijd en jaar van afname.

Aantal personen waarvoor CITOJJJTAB data bevat, naar jaar van afname (2006-2013).										
leeftijd ultimo			aantal personen per jaar van afname							
2015	2016	2017	2006	2007	2008	2009	2010	2011	2012	2013
16	17	18				34	1.502	85.252	49.730	2.820
17	18	19			26	1.468	81.219	49.428	2.716	31
18	19	20		26	1.433	80.338	46.929	2.681	23	16
19	20	21	31	1.338	80.011	49.260	2.779	21	35	
20	21	22	1.351	80.015	49.995	3.066	48	17		
21	22	23	81.921	51.742	3.409	35	29			
22	23	24	51.152	3.331	53	25				

Stap i (0 tot 4 jaar) vergt geen extra rekenwerk omdat dit direct overgenomen kan worden uit in het voorgaande beschreven berekeningen van de nettobijdrage per leeftijdsjaar voor de bevolking als geheel. Voor *Stap ii* (4-12 jaar) is een aanvullende berekening gemaakt om per citoscore de gemiddelde onderwijskosten (er zijn verschillen in kosten tussen onderwijstypen als gewoon en bijzonder basisonderwijs en ook de zogenaamde leerlinggewichten leiden tot kostenverschillen, zie ook §5.3 van deze appendix) en zorgkosten te bepalen (zie Figuur 9.7 uit het huidige rapport voor details). Dit zijn namelijk de enige kostenposten die sterk kunnen verschillen tussen personen voor deze leeftijdsgroep.⁵⁰ Voor *Stap iii* (12 tot 21 jaar) zijn de kosten- en batenposten die nodig zijn voor de berekening van de nettobijdrage per leeftijdsjaar geaggregeerd naar de citoscores (waarnemingen 2006-2018), hetgeen

⁴⁸ De methode is vergelijkbaar zoals toegepast voor de tweede generatie, zie §2.4.

⁴⁹ 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.

⁵⁰ Overige kosten- en batenposten zoals belastingen, premies, pensioenen, uitkeringen, toeslagen, enz. zijn voor deze leeftijdsgroep nog niet van toepassing of zijn voor alle groepen even hoog.

een berekening is die vergelijkbaar is met de berekeningen toegelicht in Hoofdstuk 2 van deze appendix.

De berekening in *Stap iv* (21 tot 39 jaar) is aanmerkelijk complexer en bestaat uit verschillende onderdelen. Eerst zijn drie cohorten – citocohorten genaamd – uit CITOTAB gevolgd, die 20 jaar oud waren ultimo 2015, 2016 en 2017 respectievelijk. Dat ging in totaal om 399.413 personen waarvan de cito eindtoets tussen 2006 en 2013 werd afgenomen, zie Tabel 4.1. In deze tabel heeft elk cohort zijn eigen kleur: blauw voor het cohort dat ultimo 2017 20 jaar oud is, groen voor het cohort dat ultimo 2016 20 jaar oud is en geel voor het cohort dat ultimo 2015 20 jaar oud is. Van deze cohorten is de distributie bepaald over de Hoogst Behaalde dan wel Actueel gevolgde opleiding (afgekort HBA) op 20-jarige leeftijd.⁵¹ Anders gezegd: van elk van deze drie cohorten afzonderlijk is bepaald of zij op 20-jarige leeftijd studeerden of niet studeerden, waarbij van de studerende de actueel gevolgde opleiding is bepaald en van de niet-studerende de hoogst behaalde opleiding.

Tabel 4.2 Schematisch overzicht van de gebruikte HBA-cohorten voor de bepaling van de nettobijdrage per cito-score. De percentages hebben betrekking op de beschikbaarheid van HBA-data.

data				percentage data waarvoor HBA ontbreekt naar verslagjaar										
Totaal	HBA onbekend	gebruikte data	leeft. U17	zonder NIRWO						met NIRWO				
				2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
209.680				7,4%	6,8%	6,2%	5,6%	4,9%	4,2%	2,4%	1,5%	0,6%	0,1%	0,0%
201.353	6,5%	188.342	28	7,7%	7,0%	6,5%	5,9%	5,2%	4,3%	2,4%	1,5%	0,7%	0,1%	0,0%
198.011	7,0%	184.065	29	7,7%	7,0%	6,5%	5,8%	5,1%	4,2%	2,2%	1,4%	0,7%	0,1%	0,0%
197.883	8,7%	180.642	30	8,7%	8,0%	7,4%	6,6%	5,9%	5,0%	2,1%	1,4%	0,7%	0,1%	0,0%
195.122	Totaal	553.049		12,4%	11,6%	10,7%	9,9%	9,0%	8,1%	2,2%	1,4%	0,7%	0,1%	0,0%
186.545	2,4%	182.050	32	16,7%	15,6%	14,7%	13,6%	12,5%	11,3%	2,4%	1,5%	0,8%	0,2%	0,0%
177.831	2,6%	173.270	33	20,3%	19,1%	18,1%	16,8%	15,4%	14,1%	2,6%	1,7%	0,9%	0,2%	0,0%
166.724	2,8%	161.979	34	22,7%	21,5%	20,3%	18,9%	17,1%	15,5%	2,8%	1,9%	1,0%	0,2%	0,0%
162.185	Totaal	517.299		23,0%	21,6%	20,4%	18,8%	16,8%	14,9%	2,9%	2,0%	1,1%	0,3%	0,0%
161.689	3,1%	156.746	36	22,0%	20,5%	19,3%	17,6%	15,5%	13,3%	3,1%	2,1%	1,2%	0,3%	0,0%
160.056	3,2%	154.932	37	22,6%	21,1%	19,8%	18,0%	15,7%	13,3%	3,2%	2,2%	1,4%	0,3%	0,0%
151.471	3,3%	146.444	38	23,5%	22,0%	20,7%	18,8%	16,4%	13,9%	3,3%	2,3%	1,4%	0,3%	0,0%
149.479	Totaal	458.122		24,0%	22,6%	21,3%	19,3%	16,8%	14,2%	3,3%	2,3%	1,5%	0,4%	0,0%

Vervolgens is op basis van drie andere cohorten – HBA-cohorten genaamd – een transformatiematrix bepaald van de HBA voor 20-jarigen naar de HBA voor de leeftijden tot maximaal 30 jaar. In Tabel 4.2 is schematisch weergegeven hoe dit is uitgevoerd. De HBA ontwikkelt zich over de tijd: naarmate de leeftijd vordert is er een steeds groter deel dat niet meer studeert en neemt ook het gemiddelde opleidingsniveau toe. Het blauwe citocohort uit Tabel 4.1 is 20 jaar oud ultimo 2017. Om hun HBA-ontwikkeling te volgen vanaf 20 jaar is het HBA-cohort genomen dat ultimo 2017 28 jaar oud is, zie het bovenste blauwe HBA-cohort in Tabel 4.2.⁵² Dit cohort is ultimo 2009 20 jaar oud. Met het middelste

⁵¹ Voor deze stap is dit gedaan voor een *customized* HBA-indeling in 10 actueel gevolgde opleidingsniveaus en 9 hoogst behaalde opleidingsniveaus, samengesteld uit een combinatie van de SOI 8-deling en de SOI 18-deling, dit omdat enerzijds het precieze schoolniveau voor deze leeftijden nog zeer bepalend is, maar anderzijds het gebruik van alleen de SOI 18-deling zou hebben geleid tot teveel cellen met een te klein aantal waarnemingen.

⁵² Voor deze stap is voor elk van de betrokken leeftijden 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).

blauwe cohort in Tabel 4.2 wordt de ontwikkeling verder gevolgd tot 32 jaar oud (ultimo 2017) en met het onderste blauwe cohort tot 36 jaar oud (ultimo 2017). Voor de groene cohorten geldt iets vergelijkbaars, met dien verstande dat alles één jaar is opgeschoven: dit cohort is in 2008 al 20 jaar oud en ultimo 2017 29 jaar oud. Met het middelste en onderste groene cohort kan de HBA-ontwikkeling worden gevolgd tot 33 respectievelijk 37 jaar oud (ultimo 2017). Voor de gele cohorten geldt iets vergelijkbaars, zij het dat alles nóg een jaar opgeschoven is. Met de gele cohorten kan men de HBA-ontwikkeling volgen tot maximaal 38 jaar oud (ultimo 2017).

Merk op dat voor de onderste zes cohorten in Tabel 4.2 gebruik is gemaakt van de zogenaamde ‘met NIRWO’ variant, hetgeen wordt aanbevolen door het CBS.⁵³ Bij deze cohorten konden slechts vijf jaren worden gebruikt omdat de ‘met NIRWO’ variant slechts beschikbaar was voor de jaren 2013-2017.

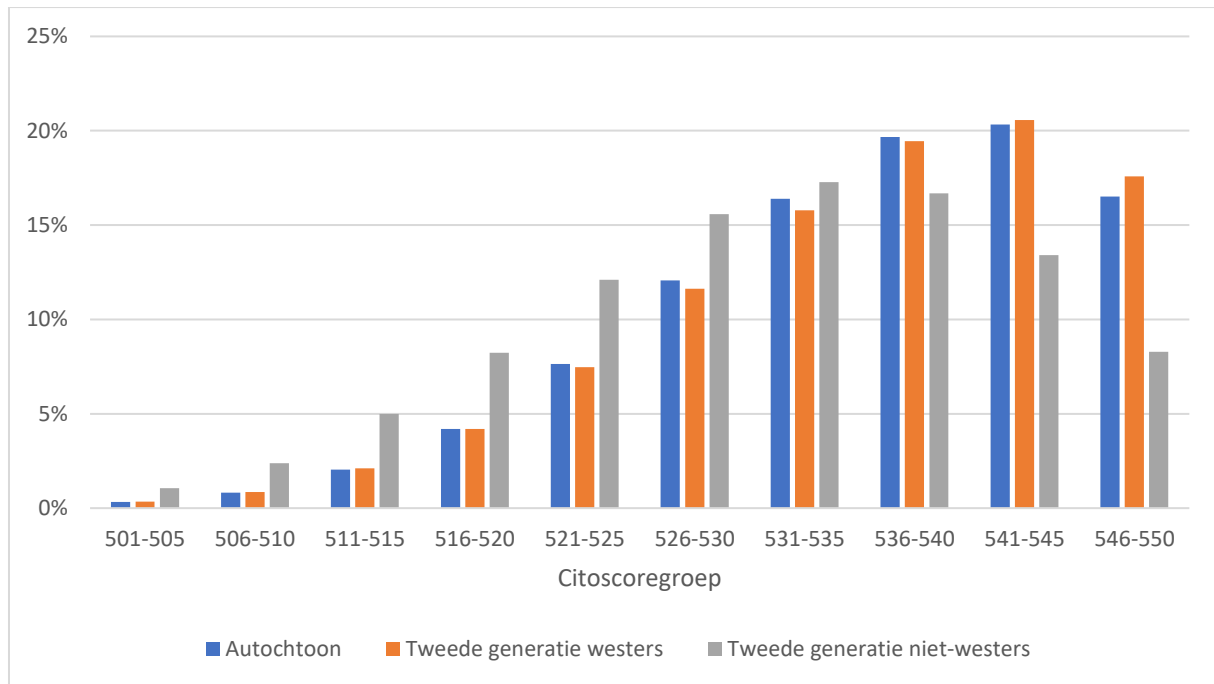
Door matrixvermenigvuldiging (uitgevoerd in Excel) zijn vervolgens transformatiematrices bepaald van citoscores naar HBA voor de leeftijden 21-38 jaar. Met deze matrices kan dus op basis van een bepaalde citoscore voor alle genoemde leeftijden de verdeling over de hoogst behaalde dan wel actueel gevolgde opleiding worden voorspeld. Dit is gedaan voor de bevolking als geheel, voor autochtonen en voor de tweede generatie, al dan niet uitgesplitst naar westers en niet-westers. Omdat voor elk van deze groepen ook de nettobijdrage per actueel gevolgde (ACT-profiel) en per hoogst behaalde (HB-profiel) opleiding bekend is, kan door matrixvermenigvuldiging ook de nettobijdrage per citoscore worden bepaald, dat wil zeggen, door vermenigvuldiging van de eerder genoemde transformatiematrices met de matrix van leeftijdsspecifieke nettobijdragen per actueel gevolgde en per hoogst behaalde opleiding.

Tot slot is *Stap v* (vanaf 39 jaar) als volgt uitgevoerd. Een zeer kleine fractie mensen studeert nog op 38-jarige leeftijd. Op basis van beschikbare microdata over onderwijsdeelname D_L voor leeftijden $0 \leq L \leq 99$ kan een profiel \vec{P}_L worden vastgesteld met $P_L = 1$ voor leeftijden $L \leq 38$ en $P_L = \frac{D_L}{D_{38}} < 1$ voor leeftijden $L > 38$. Het profiel \vec{P}_L neemt geleidelijk af en is vanaf 85 jaar gelijk aan 0. Dit profiel geeft het aandeel studenten in de totale bevolking voor 39 jaar en ouder, en wordt gebruikt (door vectorvermenigvuldiging) om het aandeel studenten geleidelijk te verminderen en het aandeel niet-studenten dienovereenkomstig te verhogen. Hoe dat gebeurt wordt uitgelegd aan de hand van een concreet voorbeeld. Stel dat van een bepaalde groep 1% van de 38-jarigen een mbo4-opleiding volgt en 10% mbo4 als hoogst behaalde opleiding heeft. Voor leeftijden $L > 38$ wordt dan het aandeel personen dat een mbo4-opleiding volgt gesteld op $1\% \times P_L$ en het aandeel personen dat mbo4 als hoogst behaalde opleiding heeft wordt gesteld op $10\% + 1\% \times (1 - P_L)$. Vervolgens kan ook voor de leeftijden vanaf 39 jaar de nettobijdrage per leeftijdsjaar verkregen worden door matrixvermenigvuldiging met de leeftijdsspecifieke nettobijdragen per actueel gevolgde en per hoogst behaalde opleiding.

Nettobijdrageprofielen naar citoscore zoals bijvoorbeeld weergegeven in Figuur 9.8 van het huidige rapport kunnen vervolgens (synthetisch) gemaakt worden door de informatie verkregen uit de hiervoor beschreven *Stap i* tot en met *Stap v* te combineren. Dergelijke profielen zijn gemaakt voor de bevolking als geheel, autochtonen en voor 2^e generatie westerse en niet-westerse migratieachtergrond. Op basis van deze profielen kan men met de in §2.1 van deze appendix beschreven methoden

⁵³ Zie voor details de CBS-microdata *Documentatie Hoogst behaald en hoogst gevolgd opleidingsniveau en opleidingsrichting van de bevolking in Nederland (HOOGSTEOPLTAB)*, opgehaald 19-4-2023 van: <https://www.cbs.nl/-/media/cbs-op-maat/microdatabestanden/documents/2021/40/hogsteopltab.pdf>

de nettobijdrage (over de levensloop) berekenen voor elke citoscore afzonderlijk, al dan niet uitgesplitst naar migratieachtergrond (zie Figuur 9.9 en 9.18 in het huidige rapport).



Figuur 4.2 Verdeling over citoscore-groepen (citodistributie) naar migratieachtergrond.

Omdat ook de verdeling over citoscores per migratieachtergrond bekend is (zie Figuur 4.2) kan men hieruit ook de totale nettobijdrage berekenen (als somproduct). Deze totalen komen aanzienlijk hoger uit dan de op directe observaties gebaseerde nettobijdrage (zonder emigratie), zoals gegeven in Hoofdstuk 4 tot 6 van het huidige rapport (zie Tabel 4.3 van het huidige rapport). De verschillen bedragen circa € 150.000 voor autochtonen, € 160.000 voor 2^e generatie westers en € 180.000 voor 2^e generatie niet-westers. Mogelijke verklaringen van deze verschillen zijn toename van citoscores⁵⁴ en een steeds hoger opleidingsniveau,⁵⁵ zowel in het algemeen als ook voor personen met een migratieachtergrond door een ‘inhaalslag’ en vooral bij niet-westers mogelijk ook door veranderende samenstelling (groter aandeel met migratieachtergrond met hoger opleidingsniveau en hogere citoscores). Omdat de voor deze rekenexercitie gebruikt data vooral betrekking hebben op de tamelijk recente periode 2006-2018, ligt het voor de hand dat deze mogelijke effecten zullen leiden tot hogere uitkomsten dan de berekeningen in Hoofdstuk 4-6 van het huidige rapport, die immers gebaseerd zijn op de gehele in 2016 in Nederland aanwezige bevolking (die dus waarschijnlijk een lager opleidingsniveau heeft en daarmee ook lagere nettobijdragen dan de citocohorten vanaf 2006). In Figuur 9.9 en 9.18 van het huidige rapport is gecorrigeerd door ijking op de nettobijdrage voor de bevolking als geheel. De onderlinge verhoudingen tussen autochtonen en westerse en niet-westerse tweede generatie migratieachtergrond blijven daarbij ongewijzigd. De rationale achter deze correctie is in de eerste plaats om de uitkomsten beter aan te laten sluiten op de overige uitkomsten in het huidige rapport. Daarnaast leidt niet corrigeren tot veel hogere totale nettobijdragen aan de schatkist van de bevolking en

⁵⁴ Vergelijk CBS-statline *Score eindtoets basisonderwijs Cito; gezinskenmerken, 2005-2010*, opgehaald 19-4-2023 van: <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/81788NED/table?dl=68067>

⁵⁵ Vergelijk CBS Jaarrapport integratie 2020, Hoofdstuk 2, opgehaald 19-4-2023 van: <https://longreads.cbs.nl/integratie-2020/onderwijs/>

de ongekend grote, structurele overschotten op de begroting die daarmee geassocieerd zijn, leveren ook geen realistisch scenario op. Tot slot gaat het bij deze hele rekenexercitie vooral om het demonstreren van het grote effect van citoscore op nettobijdrage en de verschillen die daarbij optreden tussen autochtonen en westerse en niet-westerse migratieachtergrond en daarbij gaat het dus vooral om de onderlinge verhoudingen van de bedragen.

Tot slot de toelichting op enkele deelbeweringen bij Figuur 9.18 van het huidige rapport. Het gemiddelde verschil in nettobijdrage per citopunt tussen autochtonen enerzijds en tweede generatie westers respectievelijk niet-westers anderzijds is berekend als:

$$\frac{1}{50} \cdot \sum_{c=501}^{550} (NB_{cAU} - NB_{cm})$$

Waarin NB_{cAU} de nettobijdrage is voor autochtonen (AU) met citoscore c en NB_{cm} de nettobijdrage voor personen met een tweede generatie migratieachtergrond m (in de berekening westers en niet-westers) met citoscore c , $501 \leq c \leq 550$. Dit levert € 59.000 verschil voor westers en € 170.000 verschil voor niet-westers. In de hoofdtekst zijn deze uitkomsten afgerond op veelvouden van € 10.000.

De toename in nettobijdrage per citopunt hoger is berekend als:

$$\left\{ \sum_{c=502}^{550} (NB_{(c+1)m} - NB_{cm}) \cdot \frac{1}{2} (N_{(c+1)m} + N_{cm}) \right\} / \sum_{c=502}^{550} \frac{1}{2} (N_{(c+1)m} + N_{cm})$$

Waarin N_{cm} het aantal waarnemingen is voor migratieachtergrond m en citoscore c . De N_{cm} vormen dus de distributie over citoscores $501 \leq c \leq 550$ voor de groep met migratieachtergrond m . Dit levert € 21.400 toename voor autochtonen, € 20.700 toename voor tweede generatie westers en € 16.400 voor tweede generatie niet-westers. In de hoofdtekst zijn deze uitkomsten afgerond op veelvouden van € 1.000.

Deze berekening is ook uitgevoerd voor tweede generatie westers en niet-westers samen (het gewogen gemiddelde). Dat levert € 17.600 toename in nettobijdrage per citopunt hoger. In de hoofdtekst is deze uitkomst afgerond op € 18.000.

Daarnaast is de toename in nettobijdrage per citopunt hoger ook berekend door voor migratieachtergrond m regressie van de termen

$$\overline{NB}_m = \sum_{c=501}^{550} NB_{cAU} \cdot N_{cm} / \sum_{c=501}^{550} N_{cm}$$

op de gemiddelde citoscore

$$\bar{c}_m = \sum_{c=501}^{550} c \cdot N_{cm} / \sum_{c=501}^{550} N_{cm}$$

uit te voeren voor de 42-deling (minus Nederland). Dit is met het oog op gevoeligheid voor operationalisering meermalen uitgevoerd (voor m in de 19-deling, voor m in de 42-deling en ook nog met de nettobijdragen van de bevolking als geheel i.p.v. autochtonen. In alle gevallen is de uitkomst een

toename van circa € 19.000 per citopunt hoger. Het resultaat van de vorige alinea en deze alinea zijn in de hoofdtekst aangeduid als € 18.000 à € 19.000 toename in nettobijdrage per citopunt hoger.

Ad (IV) De nettobijdrage per opleidingsniveau van de 2^e generatie is met een algoritme geschat. Met opleidingsniveau wordt in de rest van onderdeel (IV) van deze toelichting de CBS SOI 8-deling bedoeld. Dit algoritme geeft voor elk opleidingsniveau van de 1^e generatie afzonderlijk, een schatting van de verdeling over opleidingsniveaus voor de 2^e generatie. In Tabel 4.3 is dit schematisch weergegeven voor de niet-westerse migratieachtergrond. In het lichtgrijze blok van deze figuur zijn de waarnemingen gegeven voor de verdeling over opleidingsniveaus voor de 1^e generatie en in het hemelsblauwe blok voor de 2^e generatie. Concreet: van de 1^e generatie heeft volgens waarnemingen 30,9% (ten hoogste) basisschool en van de 2^e generatie is dat 4,2%. In het (witte) blok getallen in het midden is per opleidingsniveau voor de 1^e generatie een schatting gegeven van de verdeling over opleidingsniveaus voor de 2^e generatie. Concreet: bij de 1^e generatie met (ten hoogste) basisschool is de geschatte verdeling voor de 2^e generatie: 4,2% basisschool, 12,5% vmbo b/k, 1,6% vmbo g/t, enzovoorts. Het lavendelkleurige blok geeft ter controle de gewogen gemiddelden die deze schatting oplevert en te zien is dat de afwijkingen ten opzichte van de waarnemingen in het hemelsblauwe blok gering zijn. In Tabel 4.4 is hetzelfde gedaan voor de westerse migratieachtergrond en in Tabel 4.5 voor de referentie-autochtoon. Eerst volgt hieronder de uitleg voor westers en niet-westers, dan de uitleg voor de referentie-autochtoon.

Tabel 4.3 Kansverdeling voor het opleidingsniveau van 2^e generatie kinderen per opleidingsniveau van 1^e generatie immigranten, voor niet-westerse migratieachtergrond.

Verdeling over opleidingsniveaus (niet-westers)										
		waarneming 1 ^e generatie								controle
waarneming 2 ^e generatie		basisschool	vmbo b/k	vmbo g/t	mbo2, 3	mbo4	havo/vwo	bachelor	master	
		30,9%	12,5%	4,9%	15,2%	10,0%	7,5%	11,6%	7,4%	
basisschool	4,2%	4,2%	5,8%	6,2%	3,9%	5,5%	3,7%	2,7%	2,8%	4,3%
vmbo b/k	9,3%	12,5%	11,9%	10,9%	7,7%	6,7%	6,4%	6,9%	4,1%	9,3%
vmbo g/t	4,5%	1,6%	6,2%	9,6%	5,7%	6,6%	6,6%	4,3%	4,9%	4,6%
mbo2, 3	17,7%	23,8%	16,9%	15,6%	17,2%	14,6%	14,7%	12,6%	11,1%	17,7%
mbo4	22,3%	30,3%	21,4%	14,3%	22,4%	18,1%	15,8%	17,9%	14,7%	22,3%
havo/vwo	8,9%	3,7%	10,7%	13,5%	10,7%	12,0%	15,6%	8,3%	10,9%	8,9%
bachelor	21,9%	20,6%	19,0%	17,0%	21,5%	20,7%	20,5%	28,9%	27,6%	21,8%
master	11,1%	3,3%	8,1%	12,9%	10,9%	15,7%	16,7%	18,6%	23,7%	11,1%
Totaal		100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%
Gewogen gemiddelde opleidingsniveau 2 ^e generatie, weging o.b.v. rangorde SOI 8-deling										
		4,73	4,85	4,94	5,19	5,31	5,45	5,69	5,93	13,0%
Niet-westers Schatting nettobijdrage 2 ^e generatie per opleidingsniveau 1 ^e generatie										
		-€ 85.489	-€ 73.631	-€ 64.759	-€ 61.607	-€ 51.960	-€ 46.668	-€ 36.130	-€ 23.785	38,0%

Het gebruikte algoritme is gebaseerd op verschillende observaties en aannamen. De eerste aanname is dat er een strikt monotoon stijgend verband is tussen het opleidingsniveau x van de 1^e generatie en het opleidingsniveau $f(x)$ van de 2^e generatie. De achterliggende gedachte is dat hoger opgeleide 1^e

generatie ouders gemiddeld genomen hoger opgeleide 2^e generatie kinderen zullen hebben.⁵⁶ Hierbij is toegestaan dat $f(x)$ een knik vertoont. Dit laatste is gebaseerd op de observatie dat er op regioniveau in de 42-deling (minus Nederland) een positieve samenhang is tussen het opleidingsniveau van de 1^e en 2^e generatie, maar dat de stijging van de regressielijn voor de regio's met een gemiddeld⁵⁷ opleidingsniveau kleiner dan 5 ($\hat{y} = 0,43 + 3,85, R^2 = .47, N = 26, \text{Spearman's } \rho = .63, p < .001$) duidelijk verschilt van de regio's met een gemiddeld opleidingsniveau groter dan 5 ($\hat{y} = 0,99 + 0,05, R^2 = .73, N = 15, \text{Spearman's } \rho = .84, p < .001$).⁵⁸ Een mogelijke verklaring hiervoor is dat immigranten met een gemiddeld laag opleidingsniveau vaker 'onbenut opleidingspotentieel' hebben, dat bij hun in Nederland geboren 2^e generatie kinderen tot uitdrukking komt. Het relatief hoge aandeel met ten hoogste basisschool onder (met name niet-westerse) 1^e generatie immigranten is hiervoor een indicatie. Om de zaak niet onnodig te compliceren zijn er vier varianten doorgerekend waarbij is aangenomen dat geldt:

$$f(x) = \begin{cases} ax + b & \text{voor } x < 5 \\ x & \text{voor } x \geq 5 \end{cases} \text{ met } a = 0,4 \text{ of } 0,6 \text{ of } 0,8 \text{ of } 1,0 \text{ en } b = 3, 2, 1 \text{ respectievelijk } 0$$

Tabel 4.4 Kansverdeling voor het opleidingsniveau van 2^e generatie kinderen per opleidingsniveau van 1^e generatie immigranten, voor westerse migratieachtergrond.

Verdeling over opleidingsniveaus (westers)										
waarneming 2 ^e generatie		waarneming 1 ^e generatie								controle
		basisschool	vmbo b/k	vmbo g/t	mbo2, 3	mbo4	havo/vwo	bachelor	master	
		16,9%	9,6%	3,8%	13,6%	10,1%	10,4%	18,7%	16,8%	
basisschool	5,1%	8,7%	9,4%	7,4%	6,2%	5,2%	4,5%	2,1%	1,5%	5,1%
vmbo b/k	8,1%	11,1%	12,5%	10,6%	10,4%	7,0%	5,9%	6,9%	3,6%	8,1%
vmbo g/t	4,1%	4,8%	5,8%	9,3%	3,2%	5,4%	4,3%	3,5%	2,0%	4,2%
mbo2, 3	14,5%	21,3%	15,3%	14,2%	15,0%	13,5%	14,2%	10,6%	12,0%	14,5%
mbo4	15,7%	18,6%	14,3%	13,3%	17,6%	17,5%	14,6%	14,8%	13,3%	15,7%
havo/vwo	8,7%	6,0%	9,4%	12,1%	8,8%	11,6%	11,6%	7,4%	8,1%	8,7%
bachelor	25,3%	20,9%	19,4%	16,4%	22,1%	22,2%	24,8%	31,3%	32,8%	25,2%
master	18,5%	8,5%	13,9%	16,8%	16,8%	17,6%	20,1%	23,4%	26,7%	18,4%
Totaal		100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%
Gewogen gemiddelde opleidingsniveau 2 ^e generatie, weging o.b.v. rangorde SOI 8-deling										
		4,74	4,88	5,01	5,26	5,43	5,63	5,93	6,21	10,0%
Westers Schatting nettobijdrage 2 ^e generatie per opleidingsniveau 1 ^e generatie										
		-€ 48.971	-€ 38.307	-€ 32.084	-€ 24.879	-€ 19.616	-€ 10.892	€ 2.427	€ 13.329	33,0%

Verder is aangenomen dat per afzonderlijk opleidingsniveau van de 1^e generatie de verdeling over opleidingsniveaus van de 2^e generatie zodanig moet zijn dat een hoger opleidingsniveau voor de 1^e generatie altijd geassocieerd is met een hoger gewogen gemiddelde van het opleidingsniveau van de 2^e generatie. Deze aanname ligt in het verlengde van de eerste aanname. Dit laatste is voor niet-

⁵⁶ Hiervoor is in het algemeen voldoende empirisch bewijs, zie bijvoorbeeld: CBS-statline, *Jongeren (15 tot 25 jaar); onderwijsniveau, kenmerken ouders*, opgehaald 15-5-2022 van: <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/84337NED/table?dl=68046>

⁵⁷ Hiermee wordt in de rest van deze paragraaf het gewogen gemiddelde genoemd waarbij is gewogen met de rangorde van de SOI 8-deling, waarbij 'basisschool' rang 1 heeft en hbo, wo master, doctor rang 8 heeft.

⁵⁸ Iets vergelijkbaars geldt ook voor de 87-deling in herkomstregio's.

westerse migratieachtergrond weergegeven in Tabel 4.3, waar het (gewogen) gemiddelde opleidingsniveau van de 2^e generatie kinderen varieert van 4,73 voor kinderen van 1^e generatie ouders met (ten hoogste) basisschool tot 5,93 voor kinderen van ouders met een master. In de tabel is te zien dat een hoger gemiddeld opleidingsniveau bij de 1^e generatie altijd een hoger gemiddeld opleidingsniveau bij de 2^e generatie geeft (zie voor westerse migratieachtergrond Tabel 4.4). Tot slot is op vergelijkbare wijze aangenomen dat per afzonderlijk opleidingsniveau van de 1^e generatie de verdeling over opleidingsniveaus van de 2^e generatie zodanig moet zijn dat een hoger opleidingsniveau voor de 1^e generatie altijd geassocieerd is met een hoger gewogen gemiddelde van de nettobijdrage van de 2^e generatie.

Het algoritme werkt nu als volgt. Op basis van de voorgaande aannamen is de volgende kwadraten-som KS geminimaliseerd:

$$KS = \sum_{i=1}^8 \sum_{j=1}^8 \frac{P_{ij} \cdot (f(j) - i)^2}{100} + \sum_{j=1}^7 \text{Min}(0, S_{j+1} - S_j - 100)^2 + \sum_{j=1}^8 (T_j - 1000)^2 + \sum_{i=1}^8 2(O_i - E_i)^2$$

Hierin is $P_{i,j}$ de kans dat een 2^e generatie kind opleidingsniveau i behaalt als de 1^e generatie ouder opleidingsniveau j heeft. Deze kans wordt in de eerste deelsom gebruikt om de randvoorwaarde af te dwingen dat er bij benadering een strikt monotoon stijgend verband $f(x)$ (één van de vier opties zoals hiervoor beschreven) bestaat tussen het opleidingsniveau van de 1^e generatie en de 2^e generatie. Verder is S_j het somproduct $S_j = \sum_{i=1}^8 i \cdot P_{i,j}$ dat wordt gebruikt om in de tweede deelsom de randvoorwaarde af te dwingen dat een hoger opleidingsniveau voor de 1^e generatie altijd geassocieerd is met een hoger gewogen gemiddelde van het opleidingsniveau van de 2^e generatie. Voorts wordt de som $T_j = \sum_{i=1}^8 P_{i,j}$ in de derde deelsom gebruikt om er voor te zorgen dat de kolomtotalen ongeveer 100% bedragen.⁵⁹ Tot slot is O_i de observatie van het aandeel 2^e generatie met opleidingsniveau i (zie de hemelsblauwe blokken in Tabel 4.3 en Tabel 4.4) en E_i de schatting daarvan (zie de lavendelkleurige blokken) die gegeven wordt door $E_i = \sum_{j=1}^8 P_{i,j} \cdot \text{Gen}1_j$ waarin $\text{Gen}1_j$ het percentage 1^e generatie immigranten is met opleidingsniveau j (zie de lichtgrijze blokken). Dit wordt gebruikt in de vierde deelsom om er voor te zorgen dat de geschatte percentages zodanig zijn dat de werkelijke verdeling over de opleidingsniveaus voor de 2^e generatie goed wordt benaderd. De lavendelkleurige blokken in Tabel 4.3 en Tabel 4.4 laten ter controle zien dat dit het geval is. De kansen $P_{i,j}$ in Tabel 4.3 en Tabel 4.4 zijn verkregen door 500 maal de kwadraten-som KS te minimaliseren met een *bruteforce* genetisch algoritme (1 miljoen generaties, alleen mutaties op de waarden (gehele getallen⁶⁰) van $0 \leq P_{i,j} \leq 1000$ voor $1 \leq i, j \leq 8$) en hiervan het gemiddelde te nemen. Dit is vier keer gedaan: één keer voor elke variant van $f(x)$.

Nadat op deze wijze voor elk opleidingsniveau van de 1^e generatie afzonderlijk de verdeling over opleidingsniveaus van de 2^e generatie is bepaald, zijn de kansen $P_{i,j}$ genormaliseerd tot kansen $P'_{i,j}$ zodanig dat geldt $\sum_{i=1}^8 P'_{i,j} = 1$.⁶¹ Met deze genormaliseerde kansen $P'_{i,j}$ is voor elk opleidingsniveau van de 1^e generatie afzonderlijk, een schatting berekend voor het gewogen gemiddelde van de

⁵⁹ De berekening is uitgevoerd in promilles, vandaar dat 1.000 wordt gebruikt i.p.v. 100.

⁶⁰ De berekening is uitgevoerd door alles uit te drukken in op gehele getallen afgeronde promilles, vandaar $0 \leq P_{i,j} \leq 1000$.

⁶¹ Het algoritme werkt met op gehele getallen afgeronde promilles dus eerst zijn alle getallen gedeeld door 1000 en vervolgens vermenigvuldigd met een kleine correctiefactor om kleine afwijkingen te compenseren en ervoor te zorgen dat alles optelt tot 1.

nettobijdrage voor de 2^e generatie en wel als volgt: $\overline{NB}_j = \sum_{i=1}^8 NB_i \cdot P'_{i,j}$, waarin \overline{NB}_j het gewogen gemiddelde is van de 2^e generatie kinderen van 1^e generatie ouders met opleidingsniveau j en NB_i de nettobijdrage van de 2^e generatie met opleidingsniveau i . Vervolgens is onderzocht bij welke variant van $f(x)$ de geschatte nettobijdrage \overline{NB}_j van de 2^e generatie het meest gelijkmatig⁶² toeneemt met het opleidingsniveau van de 1^e generatie. Dit bleek voor niet-westerse immigranten het geval te zijn voor $f(x) = 0,6x + 2$ voor $x < 5$. Voor westerse immigranten was dit voor elk van de vier varianten van $f(x)$ het geval. Omdat de opwaartse stijging voor westerse immigranten geringer is (16,8% heeft basisschool bij de 1^e generatie tegen 30,9% bij niet-westerse immigranten) is voor westers gekozen voor $f(x) = 0,8x + 1$ voor $x < 5$. Daarnaast is voor niet-westers ook nog 50 maal het hiervoor beschreven algoritme uitgevoerd met $f(x) = 0,4x + 3$ voor $x < 5$ en 50 maal met $f(x) = 0,8x + 1$ voor $x < 5$. Idem is voor westers nog 50 maal het hiervoor beschreven algoritme uitgevoerd met $f(x) = 0,6x + 2$ voor $x < 5$ en 50 maal met $f(x) = 1,0x + 0$ voor $x < 5$.

Tabel 4.5 Kansverdeling voor het opleidingsniveau van 2^e generatie kinderen per opleidingsniveau van 1^e generatie immigranten, voor de referentie-autochtoon.

Verdeling over opleidingsniveaus (referentie-autochtoon)										
		waarneming 45-65 jaar ('1 ^e generatie')								controle
waarneming 25-45 jaar ('2 ^e generatie')		basisschool	vmbo b/k	vmbo g/t	mbo2, 3	mbo4	havo/vwo	bachelor	master	
		6,6%	10,6%	7,2%	13,4%	17,2%	8,1%	22,3%	14,5%	
basisschool	4,0%	10,3%	9,2%	4,0%	5,4%	2,4%	4,0%	2,3%	0,5%	4,0%
vmbo b/k	4,7%	10,2%	6,6%	8,1%	3,5%	4,2%	4,1%	4,0%	2,2%	4,7%
vmbo g/t	3,2%	8,8%	4,6%	4,5%	4,2%	1,5%	3,0%	1,9%	2,3%	3,2%
mbo2, 3	11,1%	18,5%	15,0%	14,6%	12,7%	15,8%	7,7%	5,0%	7,3%	11,1%
mbo4	16,2%	13,9%	17,0%	17,2%	16,8%	17,9%	13,1%	16,5%	14,7%	16,2%
havo/vwo	6,9%	8,0%	8,3%	10,1%	6,0%	4,5%	10,7%	5,7%	7,2%	6,9%
bachelor	31,8%	17,4%	25,5%	24,9%	32,9%	32,7%	32,2%	36,7%	35,6%	31,7%
master	22,3%	12,9%	13,7%	16,6%	18,5%	21,0%	25,1%	27,9%	30,1%	22,2%
Totaal		100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%
Gewogen gemiddelde opleidingsniveau 2 ^e generatie, weging o.b.v. rangorde SOI 8-deling										
		4,74	5,20	5,46	5,74	5,92	6,08	6,33	6,48	7,0%
Referentie-autochtoon		Schatting nettobijdrage 2 ^e generatie per opleidingsniveau 1 ^e generatie								controle
		-€ 36.629	-€ 21.664	-€ 11.762	-€ 1.183	€ 5.831	€ 15.039	€ 25.256	€ 30.867	28,0%

Tot slot is voor de berekening van de nettobijdragen \overline{NB}_j van de 'tweede generatie' referentie-autochtoon uitgegaan van de verdeling over opleidingsniveaus volgens de SOI8-deling van de bevolking als geheel.⁶³ Uiteraard is er bij de referentie-autochtoon niet echt sprake van een 1^e en 2^e generatie. Als proxy voor de 1^e generatie is het ongewogen gemiddelde genomen van de leeftijdsgroepen 45 tot 55 en 55 tot 65 jaar en als proxy voor de 2^e generatie het ongewogen gemiddelde van de leeftijdsgroepen 25 tot 35 en 35 tot 45 jaar. Verder is het hierboven beschreven algoritme gebruikt dat ook is aangevend voor westers en niet-westers, met dien verstande dat het algoritme 500 keer is uitgevoerd voor de volgende varianten van $f(x)$:

⁶² Bij de andere drie varianten vielen opeenvolgende niveaus soms nagenoeg samen.

⁶³ CBS-statline, *Bevolking; hoogstbehaald onderwijsniveau en onderwijsrichting*, opgehaald 18-7-2022 van: <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/85184NED/table?dl=6BB0B>

$$f(x) = ax + b \text{ voor } 1 \leq x \leq 8 \text{ met } a = 0,6 \text{ of } 0,8 \text{ of } 1,0 \text{ en } b = 2,1 \text{ respectievelijk } 0$$

Daarnaast is het algoritme ook 500 maal uitgevoerd voor:

$$f(x) = \begin{cases} 0,8x + 1 & \text{voor } x < 5 \\ x & \text{voor } x \geq 5 \end{cases}$$

d.w.z. voor de variant die voor westers ook 500 maal is doorgerekend.

Als laatste stap is er een selectie gemaakt uit de vele honderden transformatiematrices die met het algoritme zijn uitgerekend om de groepsverschillen kleiner te maken. Daarbij is uitgegaan van de volgende uitgangspunten. (i) de 2^e generatie kinderen van een 1^e generatie referentie-autochtoon met ten hoogste basisschool hebben geen substantieel hoger gemiddeld opleidingsniveau dan voor westers en niet-westers het geval is, dit omdat westers en vooral niet-westers een relatief hoog aandeel met ten hoogste basisschool hebben, wat doet vermoeden dat bij hen het onbenut opleidingspotentieel relatief hoog is. (ii) de groepsverschillen voor de tertiair opgeleiden (bachelor en master) moeten zoveel als beperkt worden omdat de waarde van deze opleidingen voor zover in het buitenland genomen het best te bepalen is. Als ijkpunt is genomen het percentage 2^e generatie kinderen van hoogopgeleide 1^e generatie (SOI 3-deling) ouders die zelf een laag, dan wel middelbaar onderwijsniveau hebben. Deze percentages (evenals de cellen waarop hun berekening is gebaseerd) zijn in Tabel 4.3 t/m 4.5 rood/groen/zwart omrand en gemarkeerd met een gele cel met daarin het woord 'controle'. Er is voor gekozen het verschil tussen niet-westers en westers respectievelijk westers en referentie-autochtoon zodanig te beperken dat het percentage laagopgeleide 2^e generatie kinderen van hoogopgeleide 1^e generatie ouders bij niet-westers 13% is, bij westers 10% en bij referentie-autochtoon 7%. Voor het percentage middelbaar opgeleide 2^e generatie kinderen van hoogopgeleide 1^e generatie ouders is aangehouden bij niet-westers 38% is, bij westers 33% en bij referentie-autochtoon 28%. Deze percentages zijn goed te verenigen met de waargenomen verschillen in het gewogen gemiddelde van het opleidingsniveau van de 2^e generatie zoals weergegeven in Tabel 4.3 t/m 4.5. Bij de selectie zijn randvoorwaarden opgelegd om de aanname van monotoon stijgend opleidingsniveau te waarborgen (d.w.z. de aanname te waarborgen dat per afzonderlijk opleidingsniveau van de 1^e generatie de verdeling over opleidingsniveaus van de 2^e generatie zodanig moet zijn dat een hoger opleidingsniveau voor de 1^e generatie altijd geassocieerd is met een hoger gewogen gemiddelde van het opleidingsniveau van de 2^e generatie). Uiteindelijk zijn Tabel 4.3 t/m 4.5 het gemiddelde van 117, 214 en 86 transformatiematrices. Het verder verkleinen van de groepsverschillen tussen niet-westers, westers en referentie-autochtoon leverde te kleine aantallen op.

Tot slot enkele opmerkingen bij deze berekeningen. (a) Bij vervolgonderzoek zouden deze transformatiematrices beter direct uit CBS-microdata afgeleid kunnen worden, op basis van de KINDOUDERTAB. Er was echter op het moment dat het belang van deze berekening werd onderkend geen toegang meer tot CBS-microdata. (b) Deze in het voorgaande beschreven schatting van de nettobijdrage van de 2^e generatie is gebruikt voor de herberekening in Tabel 9.4 in de tweede druk van het huidige rapport. In het Tijdschrift voor Politieke Economie verscheen een eerdere versie van de eerste vier kolommen van deze tabel, die op een iets andere manier berekend was (o.a. was de nettobijdrage van de '2^e generatie' referentie-autochtoon berekend op basis van Tabel 4.4 voor westers) waardoor de uitkomsten net anders zijn. (c) Dat brengt ons gelijk op het punt van gevoeligheid voor aannamen. Er is geëxperimenteerd met verschillende algoritmen en aannamen, maar de effecten daarvan zijn beperkt (orde grootte

€ 10.000 en vaak veel minder) en dat valt in het niet bij de grote verschillen naar migratieachtergrond en opleidingsniveau die zijn waargenomen voor de 1^e generatie (zie Tabel 9.4 van het huidige rapport).

Ad (V) Om inzicht te krijgen in welke mate groepsverschillen in nettobijdrage ontstaan door groepsverschillen in opleidingsniveau op het moment van immigratie, dan wel groepsverschillen in citorendement, dan wel groepsverschillen in opleidingsrendement (zie voor de onderstreepte begrippen de Begrippenlijst) zijn in §9.12 van het huidige rapport Tabel 9.5, Tabel 9.6 en Tabel 9.7 gepresenteerd. Deze tabellen verschillen van de overeenkomstige tabellen in de 1^e druk. In de 1^e druk werd gebruik gemaakt van de nettobijdragen van nul-jarigen zonder remigratie. In de 2^e druk is gebruik gemaakt van de herziene Tabel 9.4 die de nettobijdragen bevat per 1^e generatie immigrant met remigratie. De berekeningswijze in de 2^e druk is beleidsrelevanter, omdat toelatingsbeleid primair de eerste generatie betreft. Er is vanaf gezien om de uitkomsten van beide berekeningswijzen in de 2^e druk te presenteren, omdat dit zou leiden tot grotere tabellen en veel uitleg. De in de 2^e druk gehanteerde berekeningswijze zorgt voor andere, doorgaans lagere uitkomsten. In de eerste plaats omdat remigratie de verschillen doorgaans kleiner maakt (zie Tabel 4.3 van het huidige rapport). Verder is de berekening voor de tweede generatie in de 2^e druk niet meer per individu, maar per eerste generatie immigrant. Daardoor gaan verschillen in kindertal een rol spelen. Ook wordt in de berekening slechts ongeveer $\frac{2}{3}$ van dat kindertal tot de tweede generatie gerekend en $\frac{1}{3}$ tot de eerste generatie. Hierna volgt per tabel een uitleg van de berekeningswijze.

In Tabel 9.5 wordt een indicatie gegeven van effect op de nettobijdrage door verschillen in citorendement, d.w.z. groepsverschillen in hoogst behaalde opleiding per gegeven citoscore die in het onderwijssysteem ontstaan bij de eerste en tweede generatie. Concreet gaat het om verschillen als 'autochtonen met citoscore 530 hebben op 38-jarige leeftijd voor 30% mbo4 als hoogst behaalde opleiding en westerse tweede generatie met citoscore 530 hebben voor 35% mbo4 als hoogst behaalde opleiding. Het effect van dit soort verschillen in citorendement wordt uitgedrukt als het verschil in nettobijdrage t.o.v. de referentie-autochtoon. In essentie wordt gekeken hoe voor een bepaalde groep de verdeling over citoscores leidt tot een verdeling over opleidingsniveaus en vervolgens wordt die verdeling over opleidingsniveaus gewogen met de bedragen voor referentie-autochtoon.

De berekening is als volgt uitgevoerd. Per combinatie van citoscore, migratieachtergrond en generatie (autochtoon en 1^e en 2^e generatie westers en niet-westers) is de kansverdeling over opleidingsniveaus bekend voor 38-jarigen in de vorm van de eerder in deze paragraaf besproken transformatiematrices. Deze kansen worden aangeduid als P_{csmg} waarin $501 \leq c \leq 550$ de betreffende citoscore is, s het betreffende opleidingsniveau in de SOI 8-deling, m de betreffende migratieachtergrond en g de betreffende generatie. Hierbij hebben de P_{csmg} betrekking op de fractie die op 38-jarige leeftijd opleiding s als hoogst behaalde opleiding heeft of als actuele opleiding (m.a.w. de studerende, een zeer gering deel van het totaal voor deze leeftijd). De onderscheiden migratieachtergronden zijn referentie-autochtoon (aangeduid als RA), westers en niet-westers. De generatie g kan aan deze aanduidingen worden toegevoegd als cijfer, dus $RA1$ staat voor 1^e generatie referentie-autochtoon. Verder is N_{cmg} het aantal waarnemingen voor citoscore c , migratieachtergrond m en generatie g waarop de berekening van de kansen P_{csmg} gebaseerd is. Tot slot is NB_{smg} de nettobijdrage volgens kolom 2, 5 en 6 van Tabel 9.4 uit het huidige rapport. NB_{sWE1} staat dus voor de nettobijdrage van eerste generatie westerse migratieachtergrond voor opleiding s . Voor elke m en g wordt nu de gewogen nettobijdrage berekend als:

$$GNB_{mg} = \left(\sum_{c=501}^{550} N_{cmg} \cdot \sum_{s \in SOI8} P_{csmg} \cdot NB_{sRAg} \right) / \sum_{c=501}^{550} N_{cmg}$$

De gerapporteerde verschillen t.o.v. referentie-autochtonen $V_{mg}^{9.5}$ in Tabel 9.5 worden nu berekend als: $V_{mg}^{9.5} = GNB_{mg} - GNB_{RAg}$. De verschillen in Tabel 9.5 zijn ook robuust: bij andere operationaliseringen, zoals wegen tegen de SOI 5-deling zijn de uitkomsten vrijwel hetzelfde.

De in Tabel 9.5 van het huidige rapport gerapporteerde verschillen met de referentie-autochtoon zijn zeer gering. De verschillen in nettobijdrage t.o.v. de referentie-autochtoon ontstaan in de praktijk vooral door groepsverschillen in opleidingsrendement (Tabel 9.6 van het huidige rapport) en verschillen in de verdeling over opleidingsniveaus (Tabel 9.7 van het huidige rapport).

In Tabel 9.6 van het huidige rapport wordt een indicatie gegeven van het effect dat groepsverschillen in opleidingsrendement hebben op de nettobijdrage relatief t.o.v. de referentie-autochtoon. In Tabel 9.4 van het huidige rapport is te zien dat er forse groepsverschillen zijn in opleidingsrendement, d.w.z. forse verschillen in gemiddelde nettobijdrage naar migratieachtergrond voor personen met hetzelfde opleidingsniveau. Zo is de nettobijdrage van 1^e generatie referentie-autochtonen met mbo4 € 42.000 tegen -€ 78.000 voor 1^e generatie westerse immigranten met mbo4. In Tabel 9.7 wordt een indicatie gegeven van het effect op nettobijdrage door (zelf)selectie op opleidingsniveau, d.w.z. door groepsverschillen in de verdeling over opleidingsniveaus die ontstaan bij de 1^e generatie door (zelf)selectieprocessen op het moment van immigratie. De berekeningen in Tabel 9.6 en 9.7 gaan als volgt.

Uitgaande van Tabel 9.4 kan voor generatie g het verschil V_{mg} in nettobijdrage tussen migratieachtergrond m en de referentie-autochtoon RA worden berekend als:

$$V_{mg} = \sum_{s \in SOI8} NB_{smg} \cdot P_{sm1} - \sum_{s \in SOI8} NB_{sRAg} \cdot P_{sRA1}$$

Hierin staat P_{sm1} voor de op CBS-microdata gebaseerde kans dat een persoon met eerste generatie migratieachtergrond m opleidingsniveau s heeft. Er is voor de referentie-autochtoon uiteraard geen sprake van een 1^e generatie. Als proxy zijn de P_{sRA1} geoperationaliseerd op basis van het ongewogen gemiddelde van de verdeling over opleidingsniveaus van de Nederlandse bevolking volgens de SOI8-deling⁶⁴ voor de leeftijdsgroepen 45 tot 55 en 55 tot 65 jaar.

Bij deze berekeningen wordt eveneens gebruik gemaakt van de eerder genoemde nettobijdragen NB_{smg} uit Tabel 9.4 van het huidige rapport. Merk op dat alleen de verdeling over opleidingsniveaus van de eerste generatie een rol speelt, want de daaruit voortvloeiende verdeling over opleidingsniveaus van de tweede generatie is zoals in het voorgaande uiteengezet reeds verdisconteerd in de resultaten van Tabel 9.4.

Merk op dat voorgaande expressie kan worden herschreven tot:

$$V_{mg} = V_{mg}^{9.6} + V_{mg}^{9.7}$$

⁶⁴ CBS-statline, *Bevolking; hoogstbehaald onderwijsniveau en onderwijsrichting*, opgehaald 18-7-2022 van: <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/85184NED/table?dl=6BB0B>

waarin:

$$\begin{cases} V_{mg}^{9.6} = \sum_{s \in SO18} NB_{smg} \cdot P_{sm1} - \sum_{s \in SO18} NB_{sRAg} \cdot P_{sm1} \\ V_{mg}^{9.7} = \sum_{s \in SO18} NB_{sRAg} \cdot P_{sm1} - \sum_{s \in SO18} NB_{sRA1} \cdot P_{sRA1} \end{cases}$$

De verschillen $V_{mg}^{9.6}$ worden gerapporteerd in Tabel 9.6 van het huidige rapport. Bij gegeven generatie g zijn de $V_{mg}^{9.6}$ te interpreteren als dat deel van het groepsverschil in nettobijdrage t.o.v. de referentie-autochtoon dat ontstaat door verschillen tussen de nettobijdragen NB_{smg} voor personen met migratieachtergrond m en opleidingsniveau s en de nettobijdrage NB_{sRAg} voor referentie-autochtonen met hetzelfde opleidingsniveau s .

De verschillen $V_{mg}^{9.7}$ worden gerapporteerd in Tabel 9.7 van het huidige rapport. Bij gegeven generatie g zijn de $V_{mg}^{9.7}$ te interpreteren als dat deel van het groepsverschil in nettobijdrage t.o.v. de referentie-autochtoon dat ontstaat door verschillen in de verdeling over opleidingsniveaus, d.w.z. de verschillen tussen de kans P_{sm1} dat personen met migratieachtergrond m opleidingsniveau s hebben en de kans P_{sRA1} dat referentie-autochtonen opleidingsniveau s hebben.

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).”⁶⁵

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 voor de immigranten waarvan het motief niet bekend is.⁶⁶ Tot slot zijn deze data pas vanaf 1995 beschikbaar.⁶⁷

Het migratiemotief is een lastige variabele die regelmatig door het CBS wordt herzien.⁶⁸ 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

⁶⁵ 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>

⁶⁶ 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>

⁶⁷ Zie verder de CBS-microdata documentatie.

⁶⁸ Vergelijk bijvoorbeeld Bevolkingsprognose 2017-2060: veronderstellingen migratie, CBS voor meer over de migratiereden / het migratiemotief en de CBS-microdata documentatie.

tabellen tonen dan ook voor hetzelfde verslagjaar en voor vergelijkbare categorieën zoals ‘Aziatische arbeidsmigranten’ andere uitkomsten, mede omdat tabellen soms migratiemotief en soms migratiedoel rapporteren en er soms wordt uitgesplitst naar geboorteland⁶⁹, soms naar nationaliteit⁷⁰ en soms naar migratieachtergrond⁷¹.

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.⁷² 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 het huidige rapport is voor de eenduidigheid alleen gebruik gemaakt van de variabele VRLVERBLIJFS-DOEL 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 4.6 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

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-micro-databestand 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:

⁶⁹ Opgehaald 19-4-2023 van: <https://opendata.cbs.nl/#/CBS/nl/dataset/70693ned/table?dl=3F3D4>

⁷⁰ Opgehaald 19-4-2023 van: <https://opendata.cbs.nl/#/CBS/nl/dataset/84809NED/table?dl=3F3D5>

⁷¹ Opgehaald 19-4-2023 van: <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/84140NED/table?dl=3F3D6>

⁷² CBS, *Statistiek Migratiemotieven*, opgehaald 12-2-2021 van: <https://www.cbs.nl/nl-nl/onze-diensten/methoden/onderzoeksomschrijvingen/korte-onderzoeksbeschrijvingen/statistiek-migratiemotieven>

- 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.⁷³ Empirisch blijkt de vader namelijk vaak doorslaggevend voor het motief. Een indicatie is dat de combinaties met 'motief moeder = gezinmigra­tie, motief vader is anders' veel talrijker zijn dan de combinaties met 'motief vader = gezinmigra­tie, 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 4.6. Kennelijk is gezinmigra­tie vaak een secundair motief, waarbij één ouder komt met (hoofd)doel arbeid, asiel, studie of overig en de andere ouder komt met doel gezinmigra­tie. 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 4.7. In beginsel is hierbij uitgegaan van bestaande indelingen door het CBS. Daar waar mogelijk en nuttig is verder aangesloten bij de indeling in wereldregio's die de Verenigde Naties hanteren.⁷⁴ Daarnaast is vooral ook gekeken of het aantal waarnemingen voor elke regio voldoende zou zijn voor de benodigde berekeningen. 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. Indonesië en Japan), Turkije, Marokko, Afrika (excl. Marokko), Suriname, Aruba en (voormalige) Nederlandse Antillen en Latijns-Amerika (excl. Suriname, Aruba en Antillen).

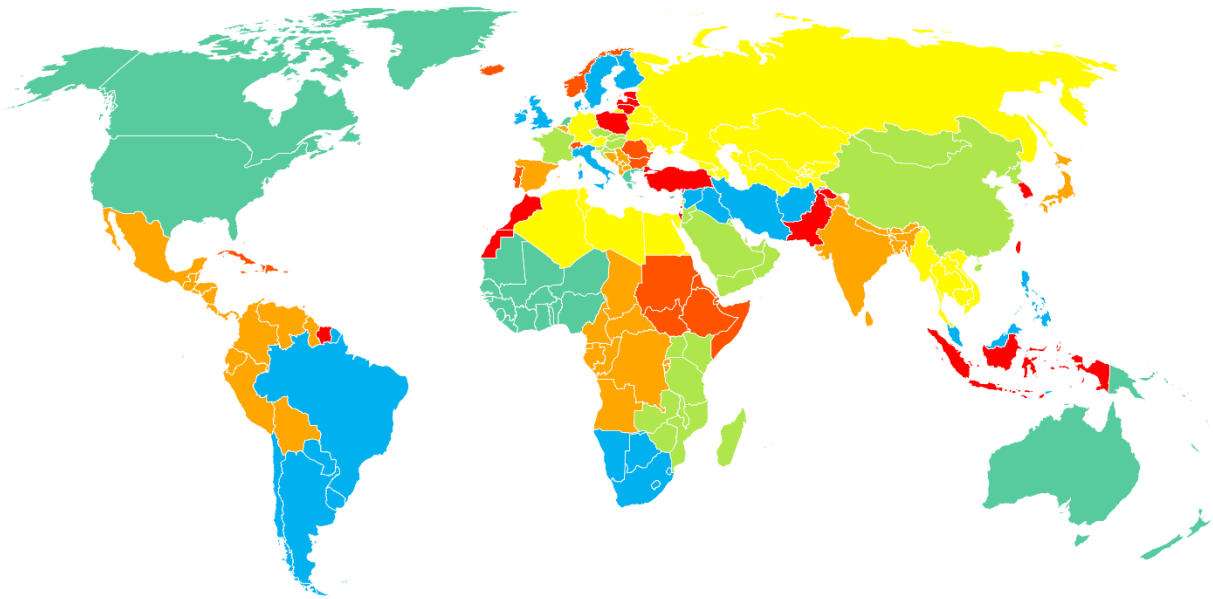
Deze indeling is in het huidige rapport verder onderverdeeld in een 42-deling, zoals weergegeven in Tabel 4.7 en Figuur 4.3. Dat is met een tussen­stap gedaan (kolom 19-deling in de tabel). Voor Europa is daarbij zoveel mogelijk aangesloten bij bestaande indelingen van het CBS zoals GIPS-landen (Griekenland, Italië, Portugal, Spanje, in het huidige rapport aangevuld met Malta en Cyprus) en MOE-landen (Midden- en Oost-Europese landen, in het huidige rapport Estland, Letland, Litouwen, Polen, Tsjechië, Slowakije, Hongarije, Slovenië, Kroatië, Roemenië en Bulgarije).

⁷³ 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.

⁷⁴ Zie bijvoorbeeld United Nations Statistics Division *Methodology: Standard country or area codes for statistical use (M49 Standard)*, *Geographic Regions*, opgehaald 4-7-2021 van: <https://unstats.un.org/unsd/methodology/m49/>

Tabel 4.7 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
		VK, Ierland, Denemarken, Zweden en Finland	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
			China, Mongolië en Noord-Korea
		Zuidoost-Azië	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
	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	
		Carabiën	
		Midden-Amerika en Zuid-Amerika Overig	



Figuur 4.3 Regio-indeling volgens de 42-deling.

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.⁷⁵ Turkije is conform CBS-indelingen apart ingedeeld. 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, 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

⁷⁵ Dit heeft te maken met het vermijden van onthullingsrisico dat zou kunnen ontstaan met het combineren van regio's.

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 van het huidige rapport 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 data is afkomstig uit een projectie⁷⁶ tot 2060 voor de CPB *Actualisatie Middellangetermijnverkenning 2018-2021*⁷⁷ (zie §8.1).

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.⁷⁸ 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 geijkt 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.

⁷⁶ Projectie voor het 'houdbare basispad model', versie 4, 15-8-2017, ISIS-versie 23.8.0.

⁷⁷ Opgehaald 19-4-2023 van: <https://www.cpb.nl/publicatie/actualisatie-middellangetermijnverkenning-2018-2021>

⁷⁸ 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 groefactor 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

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 verzorging* en de in dit rapport veronderstelde discontovoet van 3% reëel en 5% nominaal en productiviteitsgroei van 1,5% reëel en 3,5% nominaal. De ijking is uitgevoerd met gebruikmaking van de stelpost die nodig is om het effect van de geleidelijke aanpassing van de AOW-leeftijd⁷⁹ op de sociale zekerheid in te stellen (zie ook §5.6).

NB: in Hoofdstuk 8 van het huidige rapport zijn een aantal posten anders gerangschikt, gesplitst of samengenomen. Hoe dat is gedaan is in de inleiding van de betreffende paragrafen van Hoofdstuk 8 toegelicht.

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.⁸⁰ 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.⁸¹ 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.⁸²

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 verdachten naar leeftijd (vijf groepen) en migratieachtergrond.⁸³ Deze data zijn beschikbaar voor 70 herkomstlanden (eerste en tweede generatie) en daarnaast voor personen met Nederlandse en (niet)-

⁷⁹ Volgens het groeipad zoals dat gold tot het zogenaamde pensioenakkoord, zie ook §8.1.

⁸⁰ Feitelijk zijn dit alle COFOG-functies die niet al afgedekt zijn door de overige uitgavenposten in Tabel 5.1, opgehaald 19-4-2023 van: <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/82902NED/table?dl=308CD>

⁸¹ Opgehaald 19-4-2023 van: <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/82902NED/table?dl=308CD>

⁸² Opgehaald 19-4-2023 van: <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/80162NED/table?dl=31700>

⁸³ 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>

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 combinaties⁸⁴ van herkomstland en leeftijdsgroep was er onvoldoende data. In die gevallen zijn de data genomen van beide generaties samen. In een beperkt aantal gevallen⁸⁵ 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 tabellen⁸⁶ berekend hoe vaak er bij een geregistreeerde 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 'geregistreeerde 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.^{87 88} Ervan 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 geregistreeerde verdachten en krijgen gewicht 1 van het stadium 'geregistreeerde verdachte'.⁸⁹ 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 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

⁸⁴ 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.

⁸⁵ Tweemaal bij de leeftijdscategorie 12 tot 18 jaar en 27 maal in de leeftijdscategorie 65 jaar of ouder.

⁸⁶ 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>

⁸⁷ CBS-statline, *Overheidsuitgaven; naar functies 1995-2016*, opgehaald 12-02-2021 van: <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/82902NED/table?dl=308CD>

⁸⁸ CBS-statline, *Veiligheidszorg; kerncijfers 2002-2015*, opgehaald 12-02-2021 van: <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/80162NED/table?dl=31700>

⁸⁹ Bij de geregistreeerde verdachten zijn er dus geen kostenverschillen tussen groepen, bij de overige vier stadia in de rechtsketen wel.

‘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 geaggregeerd 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 zijn aanvullend gegevens over verdachten naar opleidingsniveau als proxy gebruikt.⁹⁰

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 Hoofdstuk 3 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.⁹¹ 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 OPLEIDINGSNRREFV20 en CTORREFV8 zijn aan de ONDERWIJSDEELNEMERSTAB bestanden alle variabelen uit OPLEIDINGSNRREFV20 en CTORREFV8 toegevoegd. Vervolgens zijn dummyvariabelen toegevoegd

⁹⁰ 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>

⁹¹ 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.

om tellingen te kunnen maken voor de onderwijssoorten primair onderwijs, voortgezet onderwijs (VMBO, HAVO, VWO), MBO, HBO, WO-bachelor en WO-master, dit ten behoeve van de berekening van de kosten per (cluster van) onderwijstype(n) zoals weergegeven in Tabel 8.2 van het huidige rapport. 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, maatwerktabellen⁹² van de kosten per diploma van het CBS en CBS-statline tabellen over hetzelfde onderwerp.⁹³ Hieronder wordt steeds de berekening voor 2016 toegelicht. De berekening voor 2015 is vergelijkbaar.⁹⁴

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 2016.⁹⁵ 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 2015.⁹⁶ 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 2016⁹⁷ voor overheidsuitgaven van € 1.297 miljoen en het totale deelnemersaantal voor 2016/2017 van 100,9 duizend leerlingen.⁹⁸ 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

⁹² Opgehaald 19-4-2023 van: <https://www.cbs.nl/nl-nl/nieuws/2016/51/mbo-diploma-nu-net-zo-duur-als-universitaire-master>

⁹³ Vanwege budgettaire beperkingen en/of het niet aanwezig zijn van de data was het niet altijd mogelijk om zelf de aantallen uit CBS-microdatatabanden 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.

⁹⁴ 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 huidige rapport. Voor het overige zijn kleine afwijkingen veroorzaakt doordat de onderwijsdeelnemers op twee peilmomenten zijn gepeild (1 oktober 2015 en 1 oktober 2016), zie de volgende alinea.

⁹⁵ Opgehaald 19-4-2023 van: <https://www.rijksfinancien.nl/jaarverslag/2016>

⁹⁶ Opgehaald 19-4-2023 van: <https://www.cbs.nl/-/media/excel/2016/51/onderwijsuitgaven-diplomas-maatwerk.xls>

⁹⁷ Opgehaald 19-4-2023 van: <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/80393ned/table?dl=1528B>

⁹⁸ Opgehaald 19-4-2023 van: <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/37220/table?dl=1528C>

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.⁹⁹ Dit bedrag is gedeeld door het totale deelnemersaantal voor 2016/2017 van 1.427,3 duizend leerlingen¹⁰⁰, 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 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 2016.¹⁰¹ 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.¹⁰² 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 € 4.111.¹⁰³ 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 € 11.810.¹⁰⁴

De kosten voor het MBO zijn afgeleid uit het jaarverslag bij de rijksbegroting voor 2016 en vastgesteld op € 8.000 per leerling.¹⁰⁵ De kosten voor het HBO zijn eveneens afgeleid uit het jaarverslag bij de rijksbegroting voor 2016 en vastgesteld op € 6.800 per leerling.¹⁰⁶ 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.¹⁰⁷ Dit leidt tot € 6.788 per leerling

⁹⁹ Opgehaald 19-4-2023 van: <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/80393ned/table?dl=1528B>

¹⁰⁰ Opgehaald 19-4-2023 van: <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/37220/table?dl=1528C>

¹⁰¹ Opgehaald 19-4-2023 van: http://www.rijksbegroting.nl/2016/verantwoording/jaarverslag,kst233799_8.html

¹⁰² Opgehaald 19-4-2023 van: <https://www.cbs.nl/nl-nl/nieuws/2016/51/mbo-diploma-nu-net-zo-duur-als-universitaire-master>

¹⁰³ Opgehaald 19-4-2023 van: <https://wetten.overheid.nl/BWBR0038601/2016-10-13>

¹⁰⁴ Een alternatieve berekening is gegeven in een studie door Panteia naar onderwijsbekostiging door OCW en EZ, opgehaald 19-4-2023 van: <https://zoek.officielebekendmakingen.nl/blg-583591.pdf> In het huidige rapport 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, opgehaald 19-4-2023 van: <https://www.rekenkamer.nl/publicaties/rapporten/2014/06/30/bekostiging-voortgezet-onderwijs>). Berekening op basis van deze factoren levert vergelijkbare bedragen op met iets grotere verschillen tussen regulier VO enerzijds en LWOO en Praktijkonderwijs anderzijds.

¹⁰⁵ Opgehaald 19-4-2023 van: <https://www.rijksfinancien.nl/jaarverslag/2016>

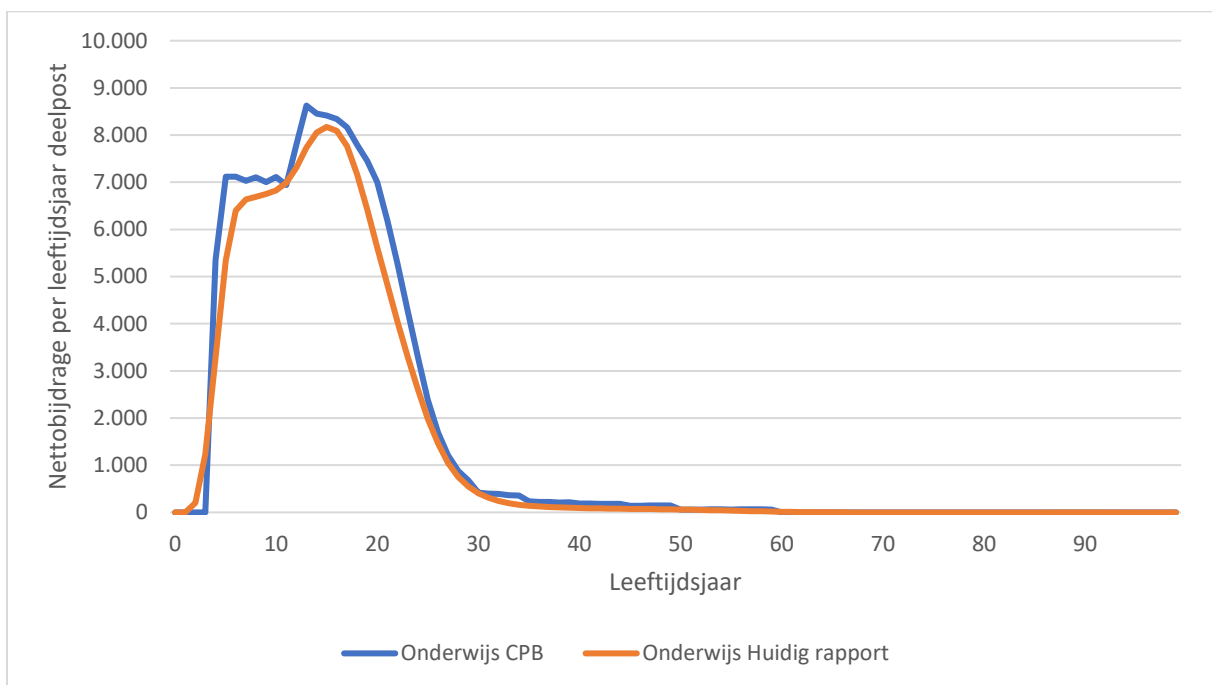
¹⁰⁶ Opgehaald 19-4-2023 van: <https://www.rijksfinancien.nl/jaarverslag/2016>

¹⁰⁷ Opgehaald 19-4-2023 van: https://www.cbs.nl/-/media/_excel/2016/51/onderwijsuitgaven-diplomas-maatwerk.xls

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, waardoor terugrekenen naar studiekosten per jaar ook niet mogelijk is.¹⁰⁸ 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 2016.¹⁰⁹

Voor het WO wordt in de het jaarverslag bij de rijksbegroting voor 2016¹¹⁰ 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.



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

¹⁰⁸ Zie de toelichting bij deze tabel, opgehaald 19-4-2023 van: <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/82299NED/table?dl=15295>

¹⁰⁹ Opgehaald 19-4-2023 van: <https://www.rijksfinancien.nl/jaarverslag/2016>

¹¹⁰ Opgehaald 19-4-2023 van: <https://www.rijksfinancien.nl/jaarverslag/2016>

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.¹¹¹

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 het huidige rapport gebruikte groepen, met uitsplitsing naar generatie, motief en herkomstregio (maximaal toegepaste verfijning volgens de 42-delings). 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.¹¹² 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:

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

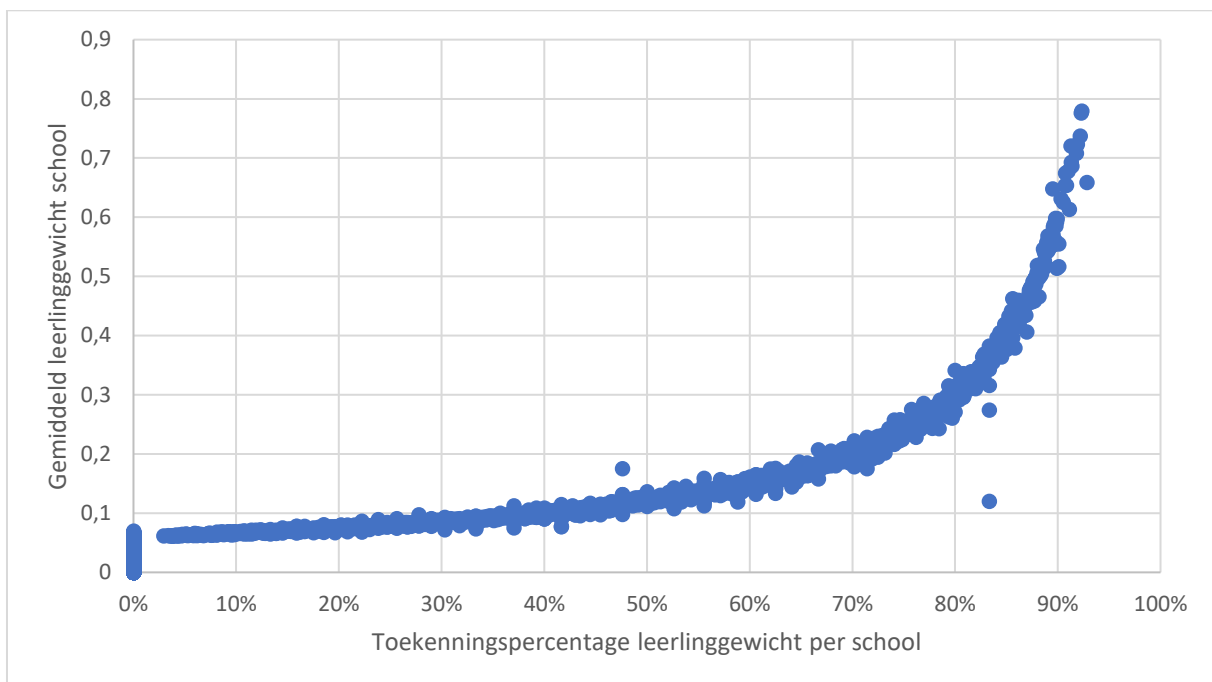
¹¹¹ Vergelijk Tabel 5.3 in de studie *Onderwijsbekostiging OCW en EZ: Vergelijking uitgaven en systematiek 2004 – 2014*, opgehaald 4-7-2021 van: <https://zoek.officielebekendmakingen.nl/blg-583591.pdf> of met de post 'fundamenteel onderzoek' die in de COFOG-indeling onder de functie Algemeen overheidsbestuur wordt geschaard in CBS-statline tabellen als *Overheidsuitgaven, naar functies 1995-2016*, opgehaald 4-7-2021 van: <https://open-data.cbs.nl/statline/#/CBS/nl/dataset/82902NED/table?dl=318DD>

¹¹² De Excel dataset "02.-leerlingen-bo-svw-vestiging,-gewicht,-impulsgebied,-school-gewicht-2015-2016" die in het verleden te downloaden was op <https://duo.nl> (opvraagbaar bij auteurs).

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 DUO¹¹³) 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.¹¹⁴ De maximaal waargenomen toekenningsgraad was in 2016 92,8%.¹¹⁵



Figuur 5.2 Toekenning extra financiering als percentage van het totale leerlinggewicht versus gemiddeld leerlinggewicht per school. Berekening op basis van data van DUO¹¹⁶ voor 6494 scholen.

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 geeft vrijwel zeker een forse onderschatting voor niet-westerse immigranten. De reden voor deze

¹¹³ Zie vorige voetnoot.

¹¹⁴ 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, opgehaald 19-4-2023 van: <https://www.cbs.nl/nl-nl/publicatie/2017/06/herziening-gewichtenregeling-primair-onderwijs>

¹¹⁵ 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.

¹¹⁶ De Excel dataset "02.-leerlingen-bo-svv-vestiging,-gewicht,-impulsgebied,-school-gewicht-2015-2016" die in het verleden te downloaden was op <https://duo.nl> (opvraagbaar bij auteurs).

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.

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.¹¹⁷

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.

¹¹⁷ Zie vorige voetnoot.

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-deling 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-deling een macrobedrag van circa 385 miljoen euro moet opleveren, dan voldoet de volgende formule:

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

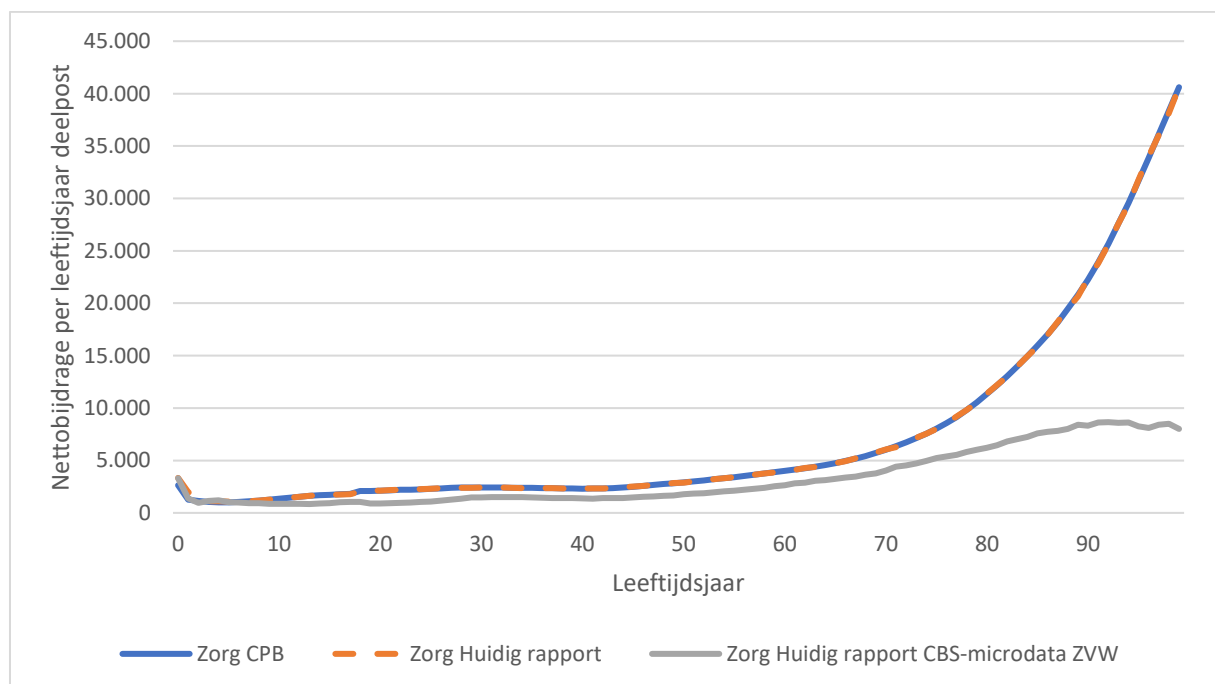
Het macrobedrag voor de hele bevolking is met deze formule 381 miljoen en voor de bevolking uitgesplitst naar migratieachtergrond volgens de 42-deling 388 miljoen. Met deze veronderstelling is het toekenningspercentage bij autochtonen 21,7% (dit is ook direct het minimum voor alle groepen) en is het 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.

5.4 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 ze onder Sociale Zaken en Werkgelegenheid op de begroting van 2016 stonden.¹¹⁸

5.5 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.¹¹⁹



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.¹²⁰ Op basis van deze tabel is het eigen risico naar leeftijd berekend voor 18 t/m 90 jarigen. Vanaf 90 jaar is het

¹¹⁸ Rijksbegroting, opgehaald 21-02-2021 van: <https://www.rijksbegroting.nl/2016/voorbereiding/begroting?hoofdstuk=40.22>

¹¹⁹ Opgehaald 19-4-2023 van: <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/81827NED/table?dl=1535C>

¹²⁰ Dat is gedaan op basis van een bewerkte tabel van Vektis behorend bij het Vektis-rapport *Verzekerden in Beeld* 2016, opgehaald 19-4-2023 van:

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).

5.6 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.

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

<https://www.vektis.nl/uploads/Publicaties/Zorgthermometer/Verzekerden%20in%20beeld%202016%20-%20bijlage.xlsx>, <https://www.vektis.nl/intelligence/publicaties/verzekerden-in-beeld-2016>

2015 voor het UWV geschat op gemiddeld 6,8% van de totale kosten¹²¹ en werden voor dat jaar geschat op totaal 2,9 miljard euro¹²². 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 AKW¹²³ en Stufi¹²⁴. 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 Ziektewet, WIA en re-integratie van (gedeeltelijke) arbeidsongeschikten die het overgrote deel van het verschil dekken¹²⁵. Bij Bijstand/ANW (18% van het CPB-macrobedrag) is een heel klein deel terug te voeren op de uitvoering van de ANW door het Rijk¹²⁶ en is de rest geïnterpreteerd als uitvoeringskosten van de gemeenten, die rond 21% liggen¹²⁷. 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 uitvoeringskosten¹²⁸ 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 uitvoeringskosten¹²⁹ 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.

¹²¹ Opgehaald 19-4-2023 van: https://www.seor.nl/Cms_Media/S1183-Kosten-en-opbrengsten-terugbrengen-AOW-leeftijd-naar-65-jaar.pdf, blz. 72.

¹²² Opgehaald 19-4-2023 van: <https://digitaal.scp.nl/publiekvoorzien/sociale-zekerheid/>

¹²³ Opgehaald 19-4-2023 van: https://www.rijksbegroting.nl/2016/voorbereiding/begroting,kst212308_21.html

¹²⁴ Opgehaald 19-4-2023 van: https://www.rijksbegroting.nl/2016/voorbereiding/begroting,kst212220_23.html

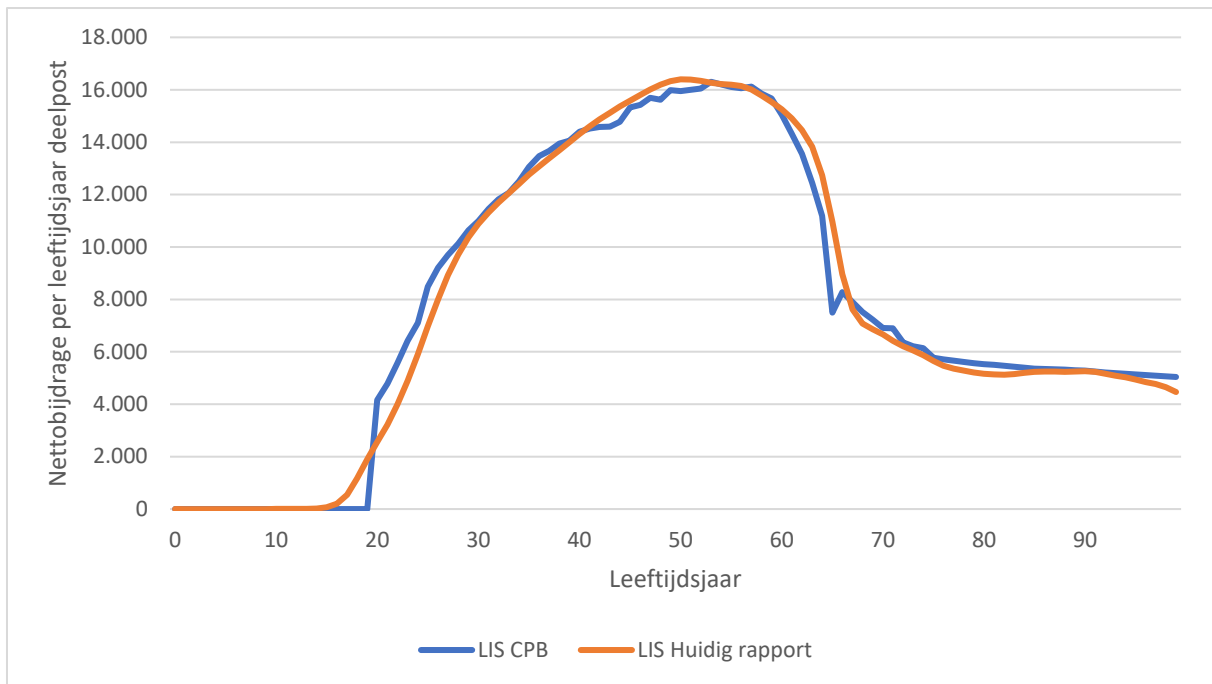
¹²⁵ Opgehaald 19-4-2023 van: https://www.rijksbegroting.nl/2016/voorbereiding/begroting,kst212308_21.html

¹²⁶ Opgehaald 19-4-2023 van: https://www.rijksbegroting.nl/2016/voorbereiding/begroting,kst212308_21.html

¹²⁷ Opgehaald 19-4-2023 van: <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/>

¹²⁸ Opgehaald 19-4-2023 van: https://www.rijksbegroting.nl/2016/voorbereiding/begroting,kst212308_21.html

¹²⁹ Opgehaald 19-4-2023 van: https://www.rijksbegroting.nl/2016/voorbereiding/begroting,kst212308_21.html



Figuur 5.4 LIS: vergelijking met het CPB-profiel.

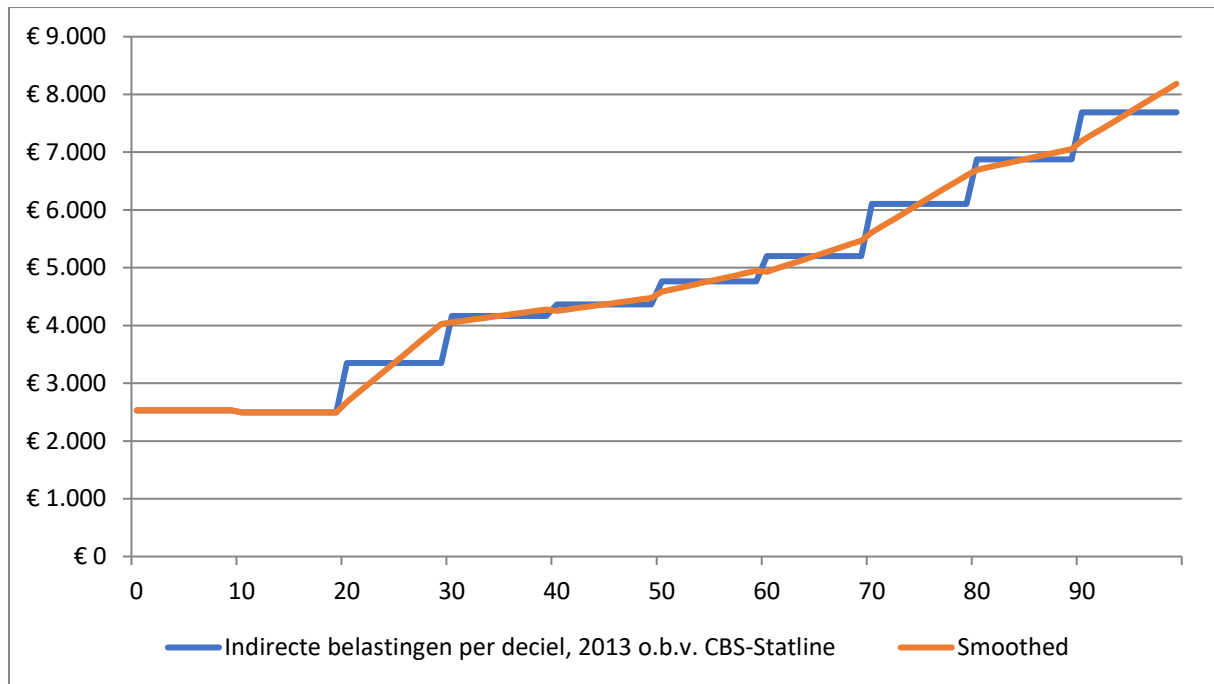
De post Subsidies is geoperationaliseerd door de INPATAB categorieën Huurtoeslag, Zorgtoeslag, Kindtoeslag en kindgebonden budget te sommeren. Vervolgens is deze post naar leeftijd gewogen tegen de in Hoofdstuk 3 beschreven onderzoekspopulatie en geijkt op het CPB-macrobedrag voor Subsidies in Tabel 5.1. Het verschil (2% van het CPB-macrobedrag) stemt overeen met de uitvoeringskosten¹³⁰ 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.¹³¹ 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

¹³⁰ Opgehaald 19-4-2023 van: <https://open.overheid.nl/repository/ronl-archief-4e7714cb-640a-4aa1-93ad-390204415c4b/1/pdf/bijlage-beleidsdoorlichting-uitvoering-toeslagen-door-belastingdienst.pdf> blz. 58

¹³¹ Deze wat cryptische omschrijving komt voort uit het feit dat de leeftijdscategorieën verschillen in grootte.

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.



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

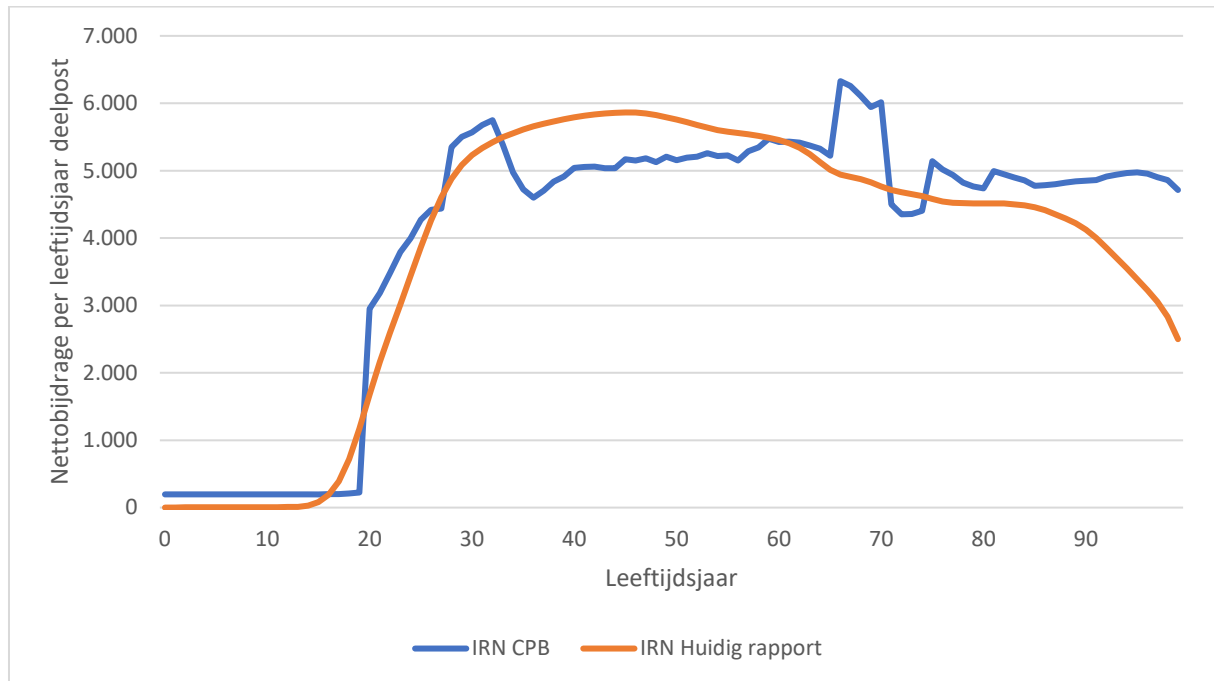
De in Hoofdstuk 8 beschreven bedragen voor de WW zijn vrij hoog. Dat heeft verschillende oorzaken. Allereerst betaalden autochtonen in het peiljaar 2016 gesommeerd over alle leeftidsjaren circa € 14.000 meer aan premie dan er werd ontvangen aan WW-uitkeringen.¹³² Daarnaast betalen jongeren relatief veel WW-premie en ontvangen ouderen relatief veel WW-uitkering, zodat contant maken het verschil nog vergroot¹³³. 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 groeiwoet 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 groeiwoet 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 groeiwoet, maar er is voor gekozen om de berekening te laten zien zoals hij is uitgevoerd.

¹³² 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.

¹³³ Immers, de ontvangen bedragen aan WW-uitkering liggen gemiddeld verder in de toekomst en worden dus sterker verkleind door contant maken.

5.7 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.



Figuur 5.6 IRN: vergelijking met CPB-profiel.

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.¹³⁴ 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.

Deze operationalisering levert een totaalbedrag op van 37 miljard. Vervolgens is door vermenigvuldiging met een factor geijkt 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

¹³⁴ Opgehaald 19-4-2023 van: <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/81290ned/table?dl=15588>

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.¹³⁵

5.8 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 met 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.¹³⁶ Totaal beslaan deze posten dus circa $16 + 22 = 38$ miljard euro.

Deze posten worden door het CPB bij generatierекeningen voor bijvoorbeeld de vergrijzing naar personen toegerekend. Het CPB zegt over de invulling van de vennootschapsbelasting bijvoorbeeld:¹³⁷

“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.

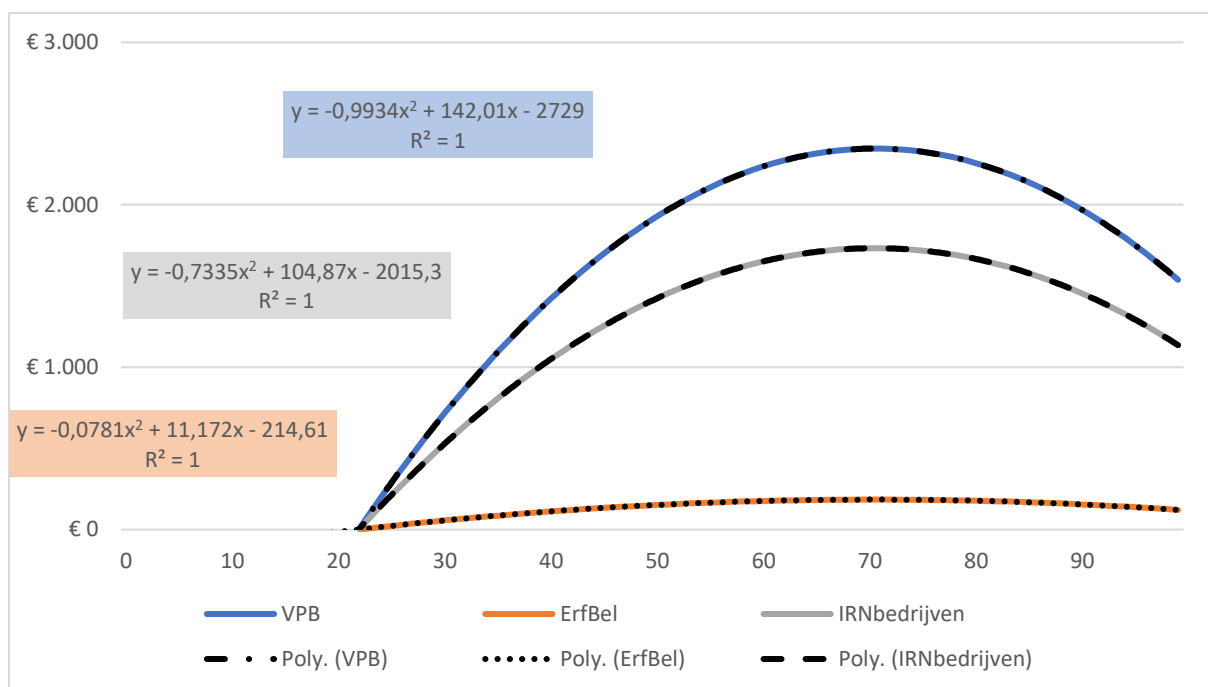
In de huidige studie wordt het totaalbedrag van 37,8 miljard euro van deze posten op verschillende manieren aan personen toegerekend. De hoofdmoot wordt in navolging van het CPB nog steeds gevormd door het (pensioen)vermogen, maar daarnaast wordt ook het directe bezit van bedrijven meegenomen. In de 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.¹³⁸ Voor Nederlandse bedrijven is de totale VPB, DB en IRN van bedrijven als volgt te schatten.

¹³⁵ De verschillen in de profielen middelen elkaar door het iken 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.

¹³⁶ Vergelijk met CBS-Statline, opgehaald 19-4-2023 van: <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.

¹³⁷ Bron: CPB-rapport Minder zorg om vergrijzing. blz. 22, opgehaald 19-4-2023 van: <https://www.cpb.nl/sites/default/files/publicaties/download/cpb-boek-12-minder-zorg-om-vergrijzing.pdf>

¹³⁸ Ontvangen binnenlands en/of buitenlands dividend, toegerekend rendement bij deelneming in buitenlandse beleggingsmaatschappijen en dividend van aangewezen participatiemaatschappijen, zie de microdata documentatie, opgehaald 19-4-2023 van: <https://www.cbs.nl/nl-nl/onze-diensten/maatwerk-en-microdata/microdata-zelf-onderzoek-doen/microdatabestanden/inpatab-inkomen-van-personen>



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

Grote niet-financiële Nederlandse bedrijven keerden 3,1 miljard euro dividend uit en betaalden 1,2 miljard euro vennootschapsbelasting.¹³⁹ Bij een kwart dividendbelasting¹⁴⁰ komt dit op 2 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 schijnnaauwkeurigheid 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 gebruteerde¹⁴¹ dividenden aanmerkelijk belang beslaat 7,0 miljard euro. Een kwart daarvan is dividendbelasting, ongeveer 1,8 miljard

¹³⁹ Opgehaald 19-4-2023 van: <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/80262ned/table?dl=31854>

¹⁴⁰ Opgehaald 19-4-2023 van: <https://www.berekenhet.nl/belastingtarieven/belastingtarieven-2016.html>

¹⁴¹ "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."

euro. Dit is al verrekend met de loon en inkomstenbelasting.¹⁴² 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 bedrijven¹⁴³ 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.

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

¹⁴² Dit is gecontroleerd met een regressie over ruim een half miljoen DGA: elke euro dividend levert 24,9 cent extra loon- en inkomstenbelasting op.

¹⁴³ SBI 2008 indeling, branches B t/m J en N, opgehaald 19-4-2023 van: <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/81156NED/table?dl=3617E>.

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 maken 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.

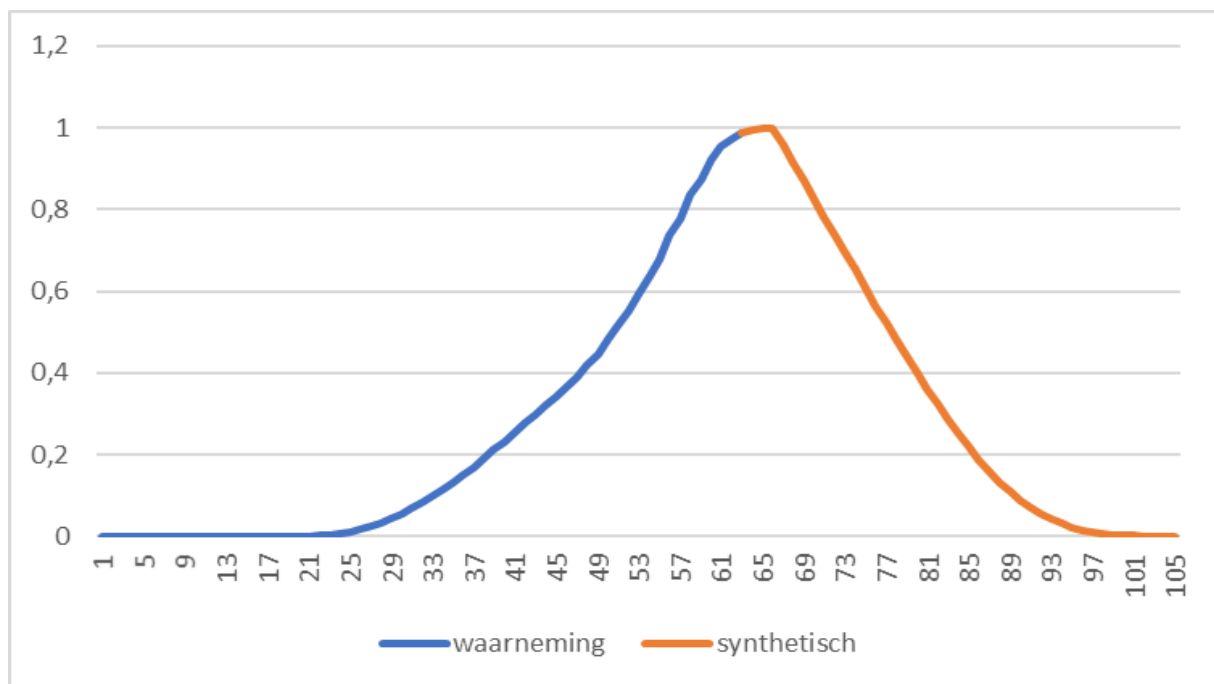
Op basis van voorgaande systematiek is voor leeftijden tot 64 jaar de pensioenopbouw geschat. In huidige studie is daarbij uitgegaan van een gemiddelde reëel rendement op aandelen van 2% (een conservatieve schatting). Tot 60 jaar is het berekende profiel gebruikt. Vanaf 60 jaar is de pensioenopbouw fictief. Omdat bij het samplen van groepen met minder data vaak gesampled is voor de leeftijds-categorie 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.¹⁴⁴ Hiermee wordt de rest van het synthetische pensioenvermogen aangevuld.

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

¹⁴⁴ De tafelbevolking is hiervoor synthetisch aangevuld op basis van CBS-statline data voor 2014-2018 voor aantallen personen van 99-105 jaar, opgehaald 19-4-2023 van: <https://opendata.cbs.nl/statline/#/CBS/nl/data-set/37325/table?dl=2F6F2>.

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.



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

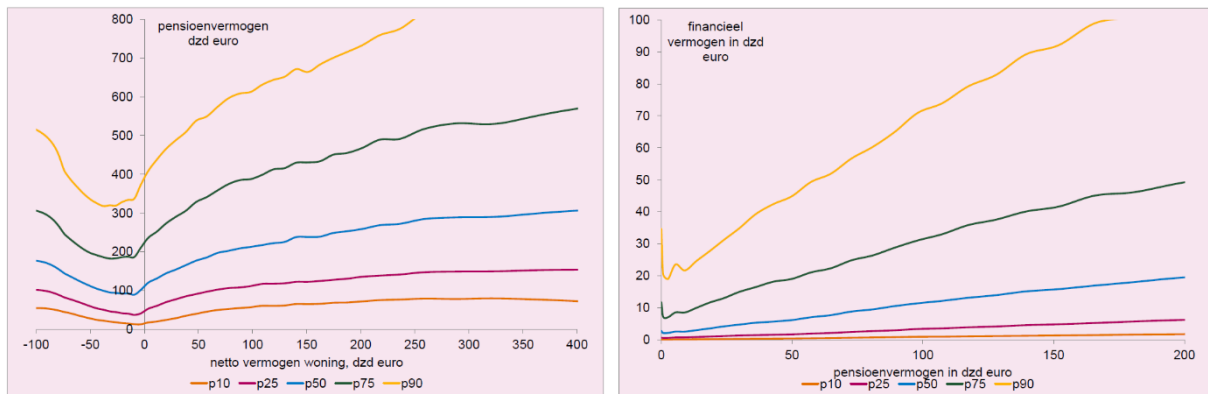
De leeftijdsprofielen voor het directe aandelenvermogen, de belastingafdrachten van ondernemers en het pensioenvermogen zijn vervolgens opgeteld tot een totaal profiel dat in de rest van deze paragraaf wordt aangeduid als ‘vermogensprofiel’. Dit vermogensprofiel is als basis gebruikt voor de posten VPB/dividendbelasting en IRN van bedrijven, door het te wegen tegen de leeftijdsopbouw van de hele onderzoekspopulatie en vervolgens te kijken 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 geijkt op het CPB-macrobedrag voor erfbelasting uit Tabel 5.1, door te wegen tegen de leeftijdsopbouw van de hele onderzoekspopulatie.

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 herkomstsgroepen. 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.

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>



6 Aanloopkosten en kosten na remigratie

6.1 Asielopvang en vergunningverstrekking

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 het huidige rapport zijn hiervoor – op basis van de Rijksbegrotingen 2013-2018¹⁴⁵ – tentatieve berekeningen gemaakt, die een orde van grootte aangeven.¹⁴⁶ Voor inburgeringskosten en vergunningverstrekking 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 2017¹⁴⁷ is geschat dat de kosten voor asiel circa 2,4 maal zwaarder zijn dan de kosten voor regulier. Door het naar reguliere vergunningen toe te schrijven deel van de IND kosten over de periode 2013-2018 (na inflatiecorrectie¹⁴⁸) 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.

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 (jeugdbescherming, voogdij, e.d. voor minderjarige asielzoekers) en een lumpsum van 30 miljoen per jaar voor de posten Vluchtelingenwerk en de Dienst Terugkeer en Vertrek. 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 inflatiecorrectie¹⁴⁹) uit te drukken in euro's van 2016 en vervolgens te delen door het totaal aantal verblijfsvergunningen voor asiel in de periode 2013-2018.¹⁵⁰ 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

¹⁴⁵ Zie de volgende begrotingsstukken, opgehaald 19-4-2023 van:

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

¹⁴⁶ Het is niet moeilijk om op basis van de tabellen-set bij dit rapport de totalen aan te passen met een eigen berekening.

¹⁴⁷ TK 34 550 VI, Nr. 2 Tabel 37.4.

¹⁴⁸ Loonvoet overheid, CEP2018, opgehaald 19-4-2023 van: https://www.cpb.nl/sites/default/files/omnidownload/Verzamelde_bijlagen_CEP_2018.xlsx

¹⁴⁹ Loonvoet overheid, CEP2018, opgehaald 19-4-2023 van: https://www.cpb.nl/sites/default/files/omnidownload/Verzamelde_bijlagen_CEP_2018.xlsx

¹⁵⁰ 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>.

toelating op basis van het VN-vluchtelingenverdrag nu eenmaal is. Over de periode 2013-2018 is het verschil overigens miniem omdat het aantal asielzoekers het aantal afgegeven verblijfsvergunningen asiel niet veel ontloopt.¹⁵¹

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). 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 2001.¹⁵² In prijzen van 2016 komt dit bedrag – afhankelijk van de gehanteerde prijsindex – op ongeveer € 10.000. 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.”¹⁵³

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:

¹⁵¹ 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>

¹⁵² Roodenburg et al. (2003) *Immigration and the Dutch Economy*, blz. 67

¹⁵³ 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.”¹⁵⁴

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.”¹⁵⁵ Het kabinet is voornemens om de wijzigingen in het jaar 2020 te doen ingaan.¹⁵⁶

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.¹⁵⁷ 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

¹⁵⁴ (2017) Vertrouwen in de toekomst, Regeerakkoord 2017-2021, blz. 55

¹⁵⁵ Zie noot III blz. 54

¹⁵⁶ Rijksoverheid (2018) Inburgering op de schop

¹⁵⁷ 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).¹⁵⁸ 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.”¹⁵⁹

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).¹⁶⁰

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

- Asielmigranten: $((6.000 \times 60 + 7.400 \times 35) / 95) \times 0,94 =$ afgerond € 6.100
- Gezinsmigranten: $((2.200 \times 60 + 3.600 \times 35) / 95) \times 0,33 =$ 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 $7.400 \times 0,125 =$ 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 harde cijfers wordt een ‘diffuse prior’ gehanteerd, te weten afbetaling voor de helft. Tabel 6.1 geeft een samenvatting van de bevindingen:

¹⁵⁸ Die ontheffing wordt achteraf verleend als de inburgeraar aan zijn inspanningsverplichting heeft voldaan,

¹⁵⁹ Zie noot II blz. 58

¹⁶⁰ Zie noot II blz. 4, diagram

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

	[1] lening	[2] kwijschelding	[3] = [1] – [2] restant lening	[4] = [3] × 0,5 oninbaar	[2] + [4] 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 voornamelijk 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.”¹⁶¹ 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 veelvoud van € 50. Dit geeft de volgende bedragen:

Asielmigrant:	$€ 5.650 \times (1 + 0,17 / 2) =$ (afgerond):	€ 6.150
Gezinsmigrant:	$€ 450 \times (1 + 0,17 / 2) =$ (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.¹⁶² Vervolgens is op basis van de remigratiekansen en het immigratieprofiel berekend hoeveel AOW-rechten immigranten opgebouwd hebben op het moment van remigratie.

¹⁶¹ Zie noot III, blz. 2

¹⁶² Zie de volgende bestanden opgehaald 19-4-2023 van: 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/handhavingsverdrag-naar-welke-landen-kan-uitkering-mee/detail/uitzonderingen-wajong-uitkering-ww-uitkering>
<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/

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.¹⁶³ 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.¹⁶⁴ 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.¹⁶⁵ 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 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.

¹⁶³ SVB, *Overzicht verdragslanden*, opgehaald 12-12-2020 van: <https://www.svb.nl/nl/aow/verdragslanden>

¹⁶⁴ Het is allemaal nog iets ingewikkelder, want de AOW-leeftijd verschuift, maar dit volstaat voor een tentatieve berekening.

¹⁶⁵ Staatscourant, stcrt-2015-10678 stcrt-2016-24815.

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 CBS¹⁶⁶ 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 gebruikt¹⁶⁷. 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 immigreert. 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

¹⁶⁶ 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>

¹⁶⁷ 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 sterfttekansen zijn gebaseerd op gegevens uit CBS-statline. In het standaardscenario is uitgegaan van de CBS-sterfttekansen tot 2060. De sterfttekansen zijn vervolgens vanaf 2060 constant gehouden. In het gebruikte model worden leeftijden van 0 tot 100 jaar gebruikt. De facto is de sterfttekans 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 verzawaard 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 sterfttekansen niet vanaf 2060 constant worden gehouden, is op basis van de CBS-prognoses voor sterfttekansen 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.¹⁶⁸ De nettobijdrage wordt negatiever als ook na 2060 de sterfttekansen 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.¹⁶⁹ De sterfttekansen zijn lager voor mensen met een hoger opleidingsniveau en *vice versa*. De verhouding in sterfttekansen per leeftijd(sgroep) tussen het onderwijsniveau op de SOI 3-deling (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 sterfttekansen 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/telers 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 gesmoothed.¹⁷⁰

Het effect op de nettobijdrage over de levensloop van differentiëren van sterfttekansen 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 sterfttekansen. 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

¹⁶⁸ Om reden dat deze afwijken van het verder vrijwel perfecte log-lineaire model.

¹⁶⁹ Opgehaald 19-4-2023 van: <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/83780NED/table?dl=3CF9C>

¹⁷⁰ 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 immigratieprofielen 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.¹⁷¹ Dit is gedaan door het verval te berekenen voor 11 verschillende 10-jaars perioden (1-10 verblijfsjaren, 2-11 verblijfsjaren, ... , 11-20 verblijfsjaren).¹⁷² 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.¹⁷³ 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 2020 gebaseerd op CBS-Statline data. Dit is aangevuld met gegevens over de leeftijdsopbouw van de derde generatie op 1 januari 2020 die is gebaseerd op CBS-data. 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.

Voor de remigratiekansen van de eerste en tweede generatie is een leeftijdsvector geconstrueerd. Voor de eerste generatie is dit gedaan door het immigratieprofiel van de gemiddelde eerste generatie immigrant te combineren met de remigratiekansen van de gemiddelde immigrant voor de 10 leeftijdsgroepen 0 tot 10 jaar t/m 90 tot 100 jaar (§7.2 van deze appendix). Concreet is de (fractionele) populatie die ontstaat door de instroom van 1 immigrant in 2020 (verdeeld over de entreeleeftijden 0-99

¹⁷¹ 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).

¹⁷² Voor de periode na 20 verblijfsjaren was er weinig data en daarom is dat niet gebruikt voor de extrapolatie.

¹⁷³ 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.

jaar volgens het voornoemde immigratieprofiel), per entreeleeftijd van 0-99 jaar afzonderlijk gevolgd tot de leeftijd van 99 jaar. Een immigrant met een entreeleeftijd van 0 jaar wordt dus gevolgd tot het jaar 2119 wat neerkomt op een leeftijd van 99 jaar. Een immigrant met een entreeleeftijd van 99 jaar wordt alleen in 2020 gevolgd. Deze populatie ontwikkelt zich volgens de in §7.1 van deze appendix besproken sterftেকansen en de in §7.2 van deze appendix besproken remigratiekansen. De ontwikkeling over de tijd van deze populatie is dus een 100×100 matrix met alleen nullen onder de diagonaal die loopt van 'entreeleeftijd=99 \times kalenderjaar=2020' naar 'entreeleeftijd=0 \times kalenderjaar=2119'. In de rest van deze paragraaf wordt met een 100×100 matrix steeds een matrix bedoeld van entreeleeftijd (0-99 jaar) en kalenderjaar (2020-2119). Op basis van deze matrix is een tweede 100×100 matrix berekend, met het (fractionele) aantal emigranten. Door vervolgens beide matrices te aggregeren naar de leeftijd van de immigranten, is voor elke leeftijd de emigratiekans te berekenen. De leeftijd is te berekenen uit de combinatie van kalenderjaar en de entreeleeftijd in het jaar 2020. Dit levert het leeftijdsprofiel met remigratiekansen voor de eerste generatie.

Voor de tweede generatie is eerst de voornoemde 100×100 populatiematrix van de eerste generatie immigranten gecombineerd met het leeftijdsprofiel voor de leeftijdsspecifieke vruchtbaarheid (zie §7.1 van deze appendix). Dit levert een 100×100 populatiematrix met daarin het (fractionele) aantal kinderen dat geboren is voor elke combinatie van kalenderjaar en entreeleeftijd (met uit de aard der zaak ook weer veel cellen die nul zijn). Op deze populatie voor de tweede generatie is nu successievelijk 18 maal de volgende bewerking toegepast voor kind-leeftijden $0 \leq k \leq 17$: (i) eerst is de populatie door scalaire vermenigvuldiging genormaliseerd op een populatieomvang van 1; (ii) vervolgens is in de vorm van een 100×100 remigratiematrix op basis van de remigratiekansen van de eerste generatie, het fractionele aantal remigrerende tweede generatie bepaald voor kind-leeftijd k en voor elke combinatie van kalenderjaar en entreeleeftijd; (iii) daarna is de som bepaald van voornoemde matrix met remigrerende tweede generatie van kind-leeftijd k en deze som is vanwege het normaliseren in stap (i) ook direct de remigratiekans voor kind-leeftijd k ; (iv) tot slot is de populatiematrix aangepast door aftrekking van de remigratiematrix (de sterftেকansen van de tweede generatie zijn verwaarloosd) en is een volgende cyclus begonnen vanaf stap (i). Dit levert het leeftijdsprofiel met remigratiekansen voor de tweede generatie.

De remigratieprofielen voor de eerste en tweede generatie zijn gebaseerd op het gemiddelde emigratiegedrag over de periode 1995-2017. Immigratie is echter in de loop der tijd dynamischer geworden, met hogere emigratiekansen. Door de remigratieprofielen voor de eerste en tweede generatie te vermenigvuldigen met een juist gekozen scalair, is de steady state remigratiekans aan te passen. Met steady state remigratiekans wordt bedoeld de stabiele verhouding tussen immigranten en emigranten die op lange termijn optreedt als men constant dezelfde immigratie heeft.

Het model sluit zoveel mogelijk aan bij de rekenexercitie van de Verenigde Naties, waaruit wordt geciteerd in §10.2 van het huidige rapport en welke in §10.3 van het huidige rapport wordt gerepliceerd. In de berekeningen van de VN gaat het om de additionele bevolkingsgroei door 'vervangingsmigratie' (letterlijk: "replacement migration") bedoeld om de vergrijzing constant te houden. Hierin wordt de ontwikkeling van de bevolking in 1995 gesteld tegenover de ontwikkeling van de groep 'post-1995 immigrants and their descendants'. In het model gebruikt voor §10.3 van het huidige rapport wordt de ontwikkeling van de Nederlandse bevolking in 2020 en hun nakomelingen vergeleken met de ontwikkeling van de post-2020 immigranten en hun nakomelingen. Daarbij wordt het aantal post-2020

immigranten steeds zo gekozen dat de verhouding tussen de groep 70-plussers en de groep 20 tot 70 jaar constant op het niveau van 2020 blijft.

Om het model eenvoudig te houden zijn enkele veronderstellingen gedaan:

- 1) Er zijn evenveel mannen als vrouwen.
- 2) Het kindertal is voor alle groepen constant 1,7 kind per vrouw.
- 3) Alle kinderen die vanaf 2020 in Nederland worden geboren zijn ófwel nakomelingen van de mensen die in 2020 in Nederland woonden, ófwel nakomelingen van de post-2020 immigranten.
- 4) Nakomelingen van post-2020 immigranten zijn ófwel het kind van twee post-2020 1^e generatie immigranten (= post-2020 2^e generatie), ofwel het kind van twee post-2020 2^e generatie ouders (= post-2020 3^e generatie), ofwel het kind van twee post-2020 3^e generatie ouders (= post-2020 4^e generatie).

Punt 1) maakt het model veel eenvoudiger, omdat er qua sterftekansen en (re)migratiekansen gewerkt kan worden met gemiddelden van mannen en vrouwen. Punt 3) behoeft enige toelichting. Het gaat in de berekening alleen maar om de additionele bevolkingsgroei vanaf 2020 en vanwege punt 2) – het kindertal is voor alle groepen altijd 1,7 – heeft rekeninghouden met allerlei mogelijke vormen van ‘gemengde relaties’¹⁷⁴ helemaal geen effect op die additionele bevolkingsgroei, terwijl het de berekening wel stukken ingewikkelder maakt. Het werd echter wel relevant gevonden om inzicht te geven hoe groot het aandeel 1^e en 2^e generatie post-2020 is. Vandaar dat er in Figuur 10.2 van het huidige rapport vier groepen worden onderscheiden, te weten, naast ‘Ingezetenen in 2020 en hun nakomelingen’ ook ‘Post-2020 immigranten’ (= post-2020 1^e generatie), ‘Kinderen van post-2020 immigranten’ (= post-2020 2^e generatie) en ‘(Achter)kleinkinderen van post-2020 immigranten’ (= post-2020 3^e en 4^e generatie). Die laatste groep omvat naast de post-2020 3^e generatie en de post-2020 4^e generatie ook de nakomelingen van de post-2020 4^e generatie, maar dat laatste gaat tot het jaar 2100 om een volstrekt verwaarloosbare fractie van de totale bevolking en is daarom in de figuur niet apart onderscheiden of genoemd. Punt 4) dient om de berekening voor het inzichtelijk maken van de relatieve omvang van 1^e en 2^e generatie post-2020 immigranten zo eenvoudig mogelijk te houden.

In het model is eerst emigratie afgetrokken en vervolgens sterfte. Daarna is veroudering toegepast en geboorte berekend. Vervolgens is de resulterende totale bevolking berekend. Als laatste stap is de immigratie berekend en vervolgens is de resulterende bevolking opnieuw berekend. Deze aanpak heeft als voordeel dat het relatief makkelijk is om te berekenen hoeveel immigranten er nodig zijn voor het constant houden van vergrijzing (of voor andere doelen als een constant migratiesaldo of een stationaire bevolking).

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

¹⁷⁴ NB: deze aanname is puur gemaakt om het rekenwerk te vereenvoudigen en hangt niet samen met voorkeuren over de wenselijkheid of onwenselijkheid van gemengde relaties.

toelaten van de zogenaamde *high potentials* om de kosten van vergrijzing op te vangen) van het huidige rapport.

Het demografische model uit de vorige paragraaf is ook gebruikt voor de schattingen voor bevolkingsgroei die zijn gegeven in §1.1 van het huidige rapport. In het model levert een jaarlijks migratiesaldo dat 10 duizend hoger ligt in 2100 1,23 miljoen extra inwoners op. In §1.1 van het huidige rapport wordt gesteld dat het verschil in bevolkingsomvang in het jaar 2100 van een migratiesaldo van 0 personen versus een migratiesaldo van 80.000 personen gelijk is aan ongeveer 10 miljoen. Dat is een afronding van $8 \times 1,23 = 9,84$ miljoen. 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.¹⁷⁵ 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.¹⁷⁶ 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 neer op 1,19 miljoen extra inwoners in 2100, dus dicht in de buurt van het in het huidige rapport gehanteerde model.

Bij de schatting van het inwoneraantal zonder immigratie (§2.2) is uitgegaan van de bevolking op 1 december 2022.¹⁷⁷ De totale bevolking was toen 17,81 miljoen. Daarvan is de eerste generatie (2,64 miljoen) evident het directe gevolg van immigratie. Hetzelfde geldt voor de tweede generatie met twee in het buitenland geboren ouders (0,93 miljoen). De tweede generatie met één in het buitenland geboren ouder (1,13 miljoen) is met gewicht 0,5 meegeteld (=0,56 miljoen) als bijdrage aan de bevolkingsgroei door immigratie. Het effect van de derde generatie tot 50 jaar (omvang op 1 januari 2020: 0,88 miljoen) op de bevolkingsgroei is o.b.v. CBS maatwerkdata¹⁷⁸ geschat op 0,32 miljoen, op 1 januari 2020 en op 0,36 miljoen eind 2022. Bij deze berekening is gebruik gemaakt van de observatie dat circa $\frac{1}{3}$ van de bevolking met een westerse tweede generatie migratieachtergrond twee in het buitenland geboren ouders heeft en circa $\frac{2}{3}$ van de bevolking met een niet-westerse tweede generatie migratieachtergrond twee in het buitenland geboren ouders.¹⁷⁹ Al met al geeft dit $2,64 + 0,93 + 0,56 + 0,36 = 4,49$ miljoen bevolkingsgroei die is toe te schrijven aan immigratie, afgerond 4,5 miljoen.

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

¹⁷⁵ United Nations, Department of Economic and Social Affairs, Population Division (2019).

¹⁷⁶ United Nations. *Net migration rate (per 1,000 population)*. Opgehaald 12-1-2023 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>

¹⁷⁷ CBS-maatwerk. *Bevolking op eerste van de maand; geslacht, leeftijd, migratieachtergrond*. Opgehaald 12-1-2023 van: <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/83482NED/table?dl=76461>

¹⁷⁸ CBS-maatwerk. *Personen met ouders van de tweede generatie*. Opgehaald 12-1-2023 van: <https://www.cbs.nl/nl-nl/maatwerk/2020/46/personen-met-ouders-van-tweede-generatie-1-januari-2020>

¹⁷⁹ CBS-Statline. *Bevolking; geslacht, leeftijd, generatie en migratieachtergrond, 1 januari*. Opgehaald 12-1-2023 van: <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/37325/table?dl=46C51>

bedroeg.¹⁸⁰ 2,5% van 708 miljard is 17,7 miljard. Bij de veronderstelde € 100.000 nettobijdrage per *high potential* levert dit op dat er jaarlijks 177.000 *high potentials* nodig zijn om 2,5% bbp vergrijzingskosten te dekken. Simulatie met hetzelfde demografische model dat in de vorige paragraaf uiteengezet is, maar dan met de geobserveerde remigratiekansen (§7.2 van deze appendix) van arbeidsmigranten¹⁸¹, geeft tot het jaar 2080 een additionele bevolkingsgroei van 5,7 miljoen personen.

Verder is uitgegaan dat er ‘voor elke 3,4 extra arbeidsmigranten één extra gezinsmigrant naar Nederland komt’.¹⁸² Dat komt neer op ongeveer 0,3 gezinsmigrant per arbeidsmigrant, oftewel 52.000 gezinsmigranten per jaar. Uitgaande van de remigratiekansen van gezinsmigranten (§7.2 van deze appendix) geeft dit 2,6 miljoen extra inwoners in het jaar 2080. In totaal geeft deze immigratie dus 8,3 miljoen extra bevolkingsgroei tot het jaar 2080.

De vraag is uiteraard wat de nettobijdrage van deze gezinsmigranten zal zijn. Voor een deel gaat dit om minderjarige gezinsmigranten. Op basis van CBS-statline data^{183 184} is geschat dat ongeveer ¼ van de gezinsmigranten minderjarig, oftewel 13.000 van de 52.000 gezinsmigranten. Als de minderjarige gezinsmigranten lijken op autochtonen zal hun nettobijdrage ongeveer € 0 zijn. Als we daarvanuit gaan zijn er dus nog 52.000 – 13.000 = 39.000 meerderjarige gezinsmigranten waarvan de nettobijdrage verdisconteerd moet worden. Als ze negatief bijdragen doet dat een deel van de nettobijdrage van de arbeidsmigranten teniet en moeten er jaarlijks meer dan 177.000 arbeidsmigranten worden toegelaten voor het gewenste effect. Andersom: als de gezinsmigranten positief bijdragen dan zijn er minder dan 177.000 arbeidsmigranten nodig.

Een rekenvoorbeeld kan de effecten verduidelijken. Stel dat de selectie op *high potentials* zodanig is dat er ook dusdanig (impliciete) selectie is op de nettobijdrage van de gezinsmigranten dat deze € 80.000 bedraagt, dan bedraagt de additionele bevolkingsgroei tot het jaar 2080 ruim 7 miljoen. Omgekeerd: indien er geen (impliciete) selectie plaatsvindt op de nettobijdrage van de gezinsmigranten en ze kosten de schatkist over de levensloop netto € 80.000 – in lijn met de huidige situatie (zie Tabel 6.1 van het huidige rapport) – dan bedraagt de extra bevolkingsgroei tot 2080 ruim 10 miljoen personen.

¹⁸⁰ 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>

¹⁸¹ Dat moet, omdat de generatierekening voor het bepalen van de nettobijdrage gebaseerd is op diezelfde remigratiekansen.

¹⁸² Zie CBS, *Veronderstellingen Immigratie* bij de bevolkingsprognose 2014-2060, opgehaald 7-1-2023 van: <https://www.cbs.nl/-/media/imported/documents/2015/17/2015bt06-bevolkingsprognose-2014-2060.pdf?la=nl-nl>

¹⁸³ Opgehaald 19-4-2023 van: <https://opendata.cbs.nl/#/CBS/nl/dataset/84808NED/table?dl=760DC>

¹⁸⁴ Opgehaald 19-4-2023 van: <https://opendata.cbs.nl/#/CBS/nl/dataset/84809NED/table?dl=760DB>

8 Verankering in CPB-studies

8.1 Verankering in CPB-vergrijzingsstudies en CPB-datasets

De methode van generatierekening die in het huidige rapport wordt gebruikt is dezelfde methode als gebruikt wordt in de CPB-vergrijzingsstudies. Bij de berekeningen is in beginsel uitgegaan van de methoden en aannamen¹⁸⁵ van de CPB-vergrijzingsstudie *Minder zorg om vergrijzing*¹⁸⁶ uit 2014 en voor zover mogelijk de CPB-vergrijzingsstudie *Zorgen om morgen*¹⁸⁷ uit 2019. Alhoewel *Minder zorg om vergrijzing* vertrekpunt is voor het huidige rapport is in de praktijk gebruik gemaakt van een actuelere dataset, namelijk een projectie¹⁸⁸ tot 2060 voor de CPB *Actualisatie Middellangetermijnverkenning 2018-2021*¹⁸⁹.

In de CPB-vergrijzingsstudies draait het om het bepalen van het zogenaamde “houdbaarheidssaldo” voor de overheidsfinanciën. Dit houdbaarheidssaldo is – met een aantal correcties¹⁹⁰ – te herleiden tot het begrip “netto profijt”, dat wil zeggen, het netto fiscaal profijt dat een individu heeft 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, dat wil zeggen, hetzelfde bedrag, maar met een tegengesteld teken.

De berekeningen in het huidige rapport zijn gebaseerd op het basisjaar 2016. Dat was bij aanvang van het onderzoek in 2018 het meest recente jaar waarvoor alle gegevens beschikbaar waren. Om de ontwikkeling van beleid en economie vanaf het jaar 2016 mee te nemen in de berekening is in het huidige rapport gebruik gemaakt van de voornoemde projectie tot 2060 voor de CPB *Actualisatie Middellangetermijnverkenning 2018-2021*. Dit betreft een dataset met leeftijdsprofielen (0-99 jaar) voor alle overheidsuitgaven en -inkomsten (m.u.v. financiële baten en lasten en ontvangsten uit het buitenland) verdeeld over 23 kosten- en batenposten (zie Tabel 5.1). Deze dataset – in de rest van dit document kortweg aangeduid als CPB2018-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 voorzag tijdens het schrijven van de *Actualisatie Middellangetermijnverkenning 2018-2021*. Voor twee veronderstellingen zijn echter aanpassingen gedaan.

¹⁸⁵ Zie de technische appendix van *Minder zorg om vergrijzing* opgehaald 19-4-2023 van: <https://www.cpb.nl/sites/default/files/publicaties/bijlagen/dp170-technische-bijlage.pdf>

¹⁸⁶ Smid, B., H. ter Rele, S. Boeters, N. Draper, A. Nibbelink en B. Wouterse (2014), *Minder zorg om vergrijzing*, opgehaald 17-1-2022 van: <https://www.cpb.nl/sites/default/files/publicaties/download/cpb-boek-12-minder-zorg-om-vergrijzing.pdf>

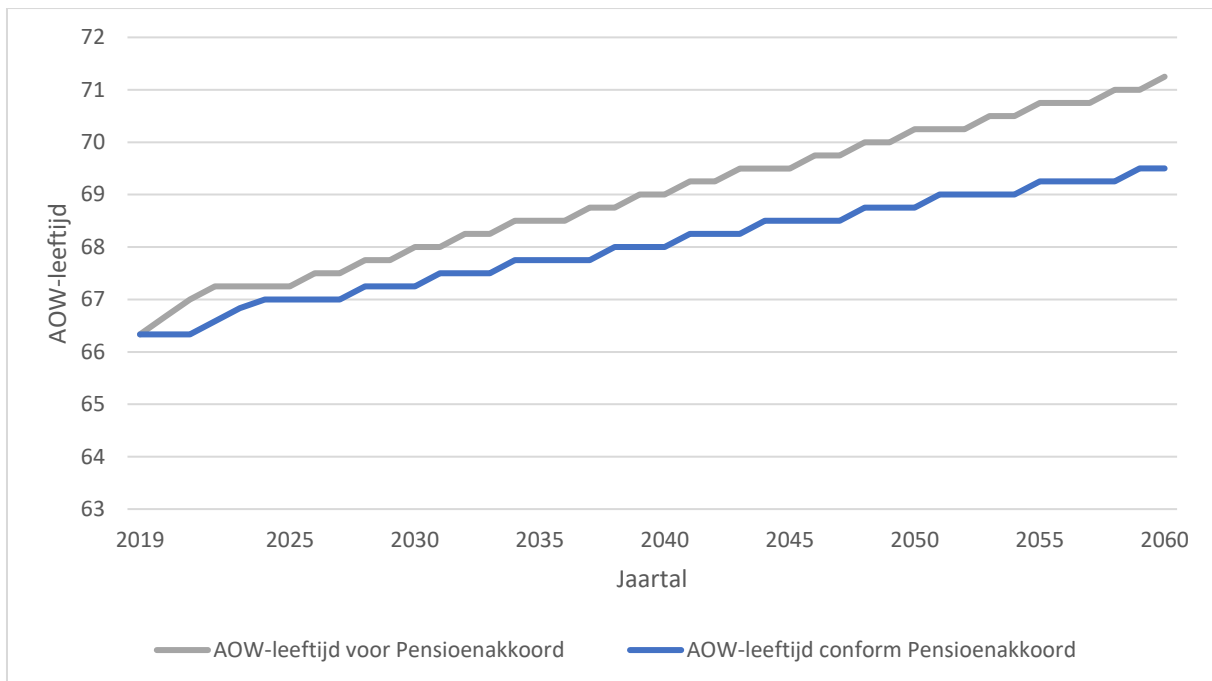
¹⁸⁷ Adema, Y., & I. van Tilburg (2019), *Zorgen om morgen*, opgehaald 17-1-2022 van: <https://www.cpb.nl/sites/default/files/omnidownload/CPB-Vergrijzingsstudie-2019-Zorgen-om-morgen.pdf>

¹⁸⁸ Projectie voor het ‘houdbare basispad model’, versie 4, 15-8-2017, ISIS-versie 23.8.0.

¹⁸⁹ Opgehaald 19-4-2023 van: <https://www.cpb.nl/publicatie/actualisatie-middellangetermijnverkenning-2018-2021>

¹⁹⁰ Smid, B., H. ter Rele, S. Boeters, N. Draper, A. Nibbelink en B. Wouterse (2014), kader ‘EMU-saldo en netto profijt’, blz. 41, opgehaald 17-1-2022 van: <https://www.cpb.nl/sites/default/files/publicaties/download/cpb-boek-12-minder-zorg-om-vergrijzing.pdf>

De eerste aanpassing ten opzichte van CPB2018 betreft de veronderstelde productiviteitsgroei en discontovoet. Het onderzoeksvoorstel voor het huidige rapport dateert van medio 2018 en hierin werd in navolging van de CPB-vergrijzingsstudie *Minder zorg om vergrijzing*¹⁹¹ en CPB2018 uitgegaan van een reële discontovoet van 3% en een productiviteitsgroei van 1,5% per jaar.¹⁹² Tijdens het onderzoek verscheen de CPB-vergrijzingsstudie *Zorgen om morgen* (2019), dat uitging van een reële discontovoet 2,5%¹⁹³ en productiviteitsgroei van 1%¹⁹⁴. Om zoveel mogelijk aan te sluiten bij het CPB zijn deze veronderstellingen overgenomen in het huidige rapport.



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-Vergrijzingsstudie-2019-Data-figures.xlsx>

De tweede aanpassing ten opzichte van CPB2018 betreft de AOW-leeftijd die 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 hogere en lagere leeftijden die niet beïnvloed zijn door de aanpassing van de AOW-leeftijd is de aanpassing van de AOW leeftijd volledig uit de oorspronkelijke CPB2018-dataset verwijderd. In de praktijk is vanwege het indelen in leeftijdsgroepen vooral gebruik gemaakt van de ontwikkeling in de leeftijdsgroepen 60 tot 64 jaar en 72 tot 80 jaar. In de resulterende dataset is de ontwikkeling over de tijd van de betreffende posten voor het interval 64 tot 72 jaar in lijn met de omliggende hogere en lagere

¹⁹¹ Smid, B., H. ter Rele, S. Boeters, N. Draper, A. Nibbelink en B. Wouterse (2014), opgehaald 17-1-2022 van: <https://www.cpb.nl/sites/default/files/publicaties/download/cpb-boek-12-minder-zorg-om-vergrijzing.pdf>

¹⁹² Dit betekende een nominale discontovoet van 5% en nominale productiviteitsgroei met 3,5% per jaar, want het CPB veronderstelde een inflatie van 2%, zie Smid, B., H. ter Rele, S. Boeters, N. Draper, A. Nibbelink en B. Wouterse (2014), blz. 27

¹⁹³ Zie Adema, Y., & I. van Tilburg (2019), blz. 70-72

¹⁹⁴ Zie Adema, Y., & I. van Tilburg (2019), blz. 42

leeftijden. Deze herziene dataset is onder andere gebruikt voor het ‘AOW bij 65 jaar scenario’ in de gevoeligheidsanalyse in §6.5 van het huidige rapport.

Het basisscenario voor het huidige rapport wordt gevormd door de CPB2018-dataset in combinatie met de in het voorgaande beschreven aanpassingen; de verlaging van de reële discontovoet (van 3,0% naar 2,5%) en productiviteitsgroei (van 1,5% naar 1%) en de aanpassing van de AOW-leeftijd. Omdat *Minder zorg om vergrijzing* uit 2014 een houdbaarheidssaldo van +0,4% bbp heeft en het pensioenakkoord een verslechtering van 0,4% bbp oplevert (zie § 8.2) heeft het basisscenario voor het huidige rapport een houdbaarheidssaldo van ongeveer 0% bbp.

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

Figuur 8.2 Verschillen tussen de CPB-vergrijzingsstudies Minder zorg om vergrijzing uit 2014 en Zorgen om morgen uit 2019. Facsimile van Tabel 2.2 uit Adema, Y., & I. van Tilburg (2019), Zorgen om morgen, opgehaald 17-1-2022 van: <https://www.cpb.nl/sites/default/files/omnidownload/CPB-Vergrijzingsstudie-2019-Zorgen-om-morgen.pdf>

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). 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. Het CPB noemt de volgende verschillen (voor een volledig overzicht zie Figuur 8.2) met de CPB-vergrijzingsstudie *Minder zorg om vergrijzing* uit 2014:

- 1) Het CPB heeft de reële discontovoet verlaagd van 3,0% naar 2,5%;¹⁹⁵
- 2) De veronderstelde productiviteitsgroei is door het CPB verlaagd van 1,5% naar 1%;¹⁹⁶
- 3) Door hogere zorgkosten is het houdbaarheidssaldo t.o.v. 2014 verslechterd met 0,6% bbp;¹⁹⁷
- 4) Het pensioenakkoord verslechtert het houdbaarheidssaldo t.o.v. 2014 met 0,4% bbp;¹⁹⁸

¹⁹⁵ Zie Adema, Y., & I. van Tilburg (2019), blz. 70-72

¹⁹⁶ Zie Adema, Y., & I. van Tilburg (2019), blz. 42

¹⁹⁷ Zie Adema, Y., & I. van Tilburg (2019), blz. 17-18

¹⁹⁸ Zie Adema, Y., & I. van Tilburg (2019), blz. 17

- 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.¹⁹⁹
- 6) Het arbeidsaanbod stijgt over de lange termijn.

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 1% zijn daarom overgenomen. Ook de effecten van het pensioenakkoord zijn in het huidige rapport verwerkt (zie §8.1). De overige verschillen konden echter niet zonder meer worden overgenomen omdat de datasets CPB2018 en CPB2019 verschillen. Onder meer is de CPB2019 geaggregeerd tot 8 deelposten in plaats van 23 deelposten van CPB2018 (zie Tabel 5.1), waarbij er te weinig bekend is van de precieze methoden e.d., wat tot inconsistenties kan leiden als beide datasets worden gecombineerd.

Wel is de CPB2019-dataset²⁰⁰ 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 CPB2018-dataset. Hierbij is het profiel voor langdurige zorg (Wlz, Wmo e.d.) – d.w.z. het zogenaamde verschilprofiel uit §5.5 van deze appendix – vervangen, door het verschilprofiel dat ontstaat door somming van de profielen Wmo, Wlz en overige zorg in Excelwerkblad 2.4 van de 2019-dataset²⁰¹. Dit levert op lange termijn een 0,3% bbp verslechtering op, namelijk de helft²⁰² 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 2020.²⁰³ 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.²⁰⁴ 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 bbp²⁰⁵, waarbij is uitgegaan van 3,0% groei t.o.v. 2019²⁰⁶ 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

¹⁹⁹ Zie Adema, Y., & I. van Tilburg (2019), blz. 16

²⁰⁰ Opgehaald 19-4-2023 van: <https://www.cpb.nl/sites/default/files/omnidownload/CPB-Vergrijzingsstudie-2019-Data-figuren.xlsx>

²⁰¹ Zie vorige voetnoot

²⁰² Zie Adema, Y., & I. van Tilburg (2019), blz. 17

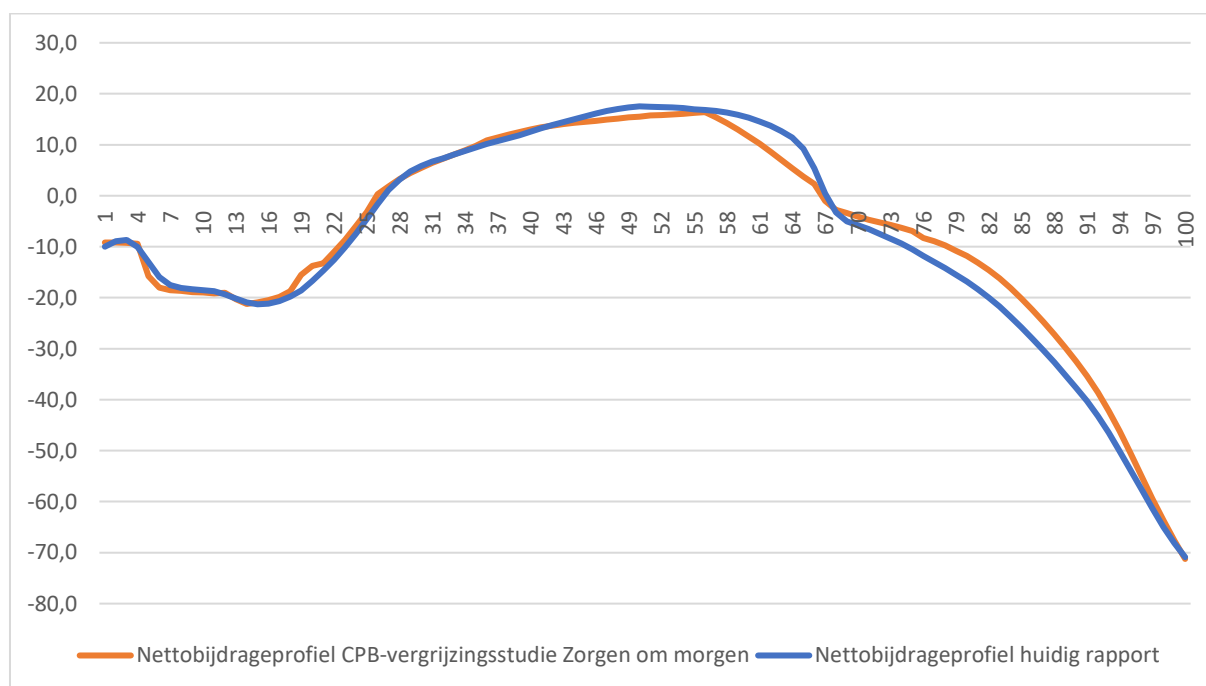
²⁰³ Figuur 4.1 blz. 7, opgehaald 19-4-2023 van: <https://www.cpb.nl/sites/default/files/omnidownload/CPB-Achtergronddocument-dec2019-Arbeidsparticipatie-en-gewerkte-uren-tot-en-met-2060.pdf>

²⁰⁴ "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", opgehaald 19-4-2023 van: https://www.cpb.nl/sites/default/files/omnidownload/CPB-Notitie-4okt2017-Analyse-economische-en-budgettaire-effecten-Regeerakkoord_0.pdf

²⁰⁵ Door Corona zal het bbp voor 2020 lager uitvallen, maar hier wordt vergeleken met de voorspelling van het CPB en niet met de werkelijkheid.

²⁰⁶ Zie dit bestand, gepubliceerd 1 dag voor *Zorgen om morgen*, Opgehaald 19-4-2023 van: <https://www.cpb.nl/sites/default/files/omnidownload/CPB-Policy-Brief-Decemberraming-2019-voorzicht-2020.pdf>

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 CBP-vergrijzingsstudie uit 2019.



Figuur 8.3 Vergelijking van het nettobijdrageprofiel van het huidige rapport met het nettobijdrageprofiel uit de CPB-vergrijzingsstudie Zorgen om morgen uit 2019.

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 (zie Figuur 8.3). 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.

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 2016-2021²⁰⁷ 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 CPB2018 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:

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

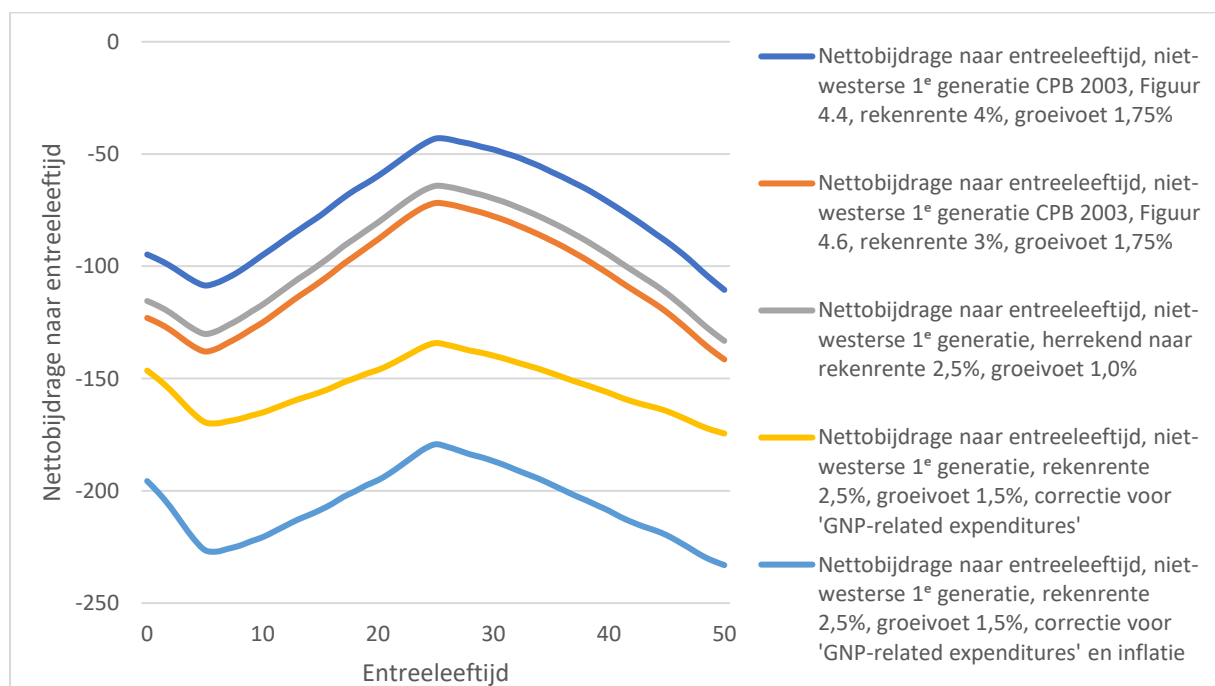
²⁰⁷ Vergelijk CPB, *Verzamelde bijlagen bij de raming november 2020*, opgehaald 15-12-2020 van: <https://www.cpb.nl/sites/default/files/omnidownload/verzamelde-bijlagen-raming-nov-2020-plus-mlt.xlsx>

Dit levert een nieuwe dataset op met daarin alle veronderstellingen uit CPB2018, 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.²⁰⁸ 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*²⁰⁹ en *Figure 4.6*²¹⁰ 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 8.4*) 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 8.4*).



Figuur 8.4 Vertaling van de resultaten voor niet-westerse eerste generatie immigranten uit Roodenburg et al. naar het huidige rapport.

²⁰⁸ Roodenburg, H., R. Euwals & H. ter Rele (2003), blz. 68, voetnoot 12

²⁰⁹ Roodenburg, H., R. Euwals & H. ter Rele (2003), blz. 70

²¹⁰ Roodenburg, H., R. Euwals & H. ter Rele (2003), blz. 74

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 leeftijdjaar (uit het CBS-microdatabestand INPATAB, variabele INPPERSPRIM, verslagjaar 2016) het verschil tussen beide methoden geschat. Er is gekozen voor het PPI omdat dit goed de bijdrage aan het BNP weerspiegelt en net als in de nettobijdrageprofielen van het CPB (zie *Figure 4.3*²¹¹) 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*.²¹² Voor de sterftekansen is de tafelbevolking uit 2016 als proxy genomen. Vervolgens is het resulterende profiel *P* geijkt met een discountfactor als stelpost en als ijkpunten de gegeven referentiewaarden voor ‘GDP-related expenditures’ voor de groep ‘non-Western’ in *Table 4.3*²¹³. De ‘GDP-related expenditures’ waren volgens *Table 4.2*²¹⁴ 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 *Figure 8.4*. Het resultaat is de gele lijn in *Figure 8.4*. Tot slot is gecorrigeerd voor inflatie, door vermenigvuldiging met een factor 1,336^{215 216} en dat levert de lichtblauwe lijn op in *Figure 8.4*.

De lichtblauwe 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 (in absolute termen) één vijfde hoger dan het resultaat voor eerste generatie niet-westerse immigranten in het huidige rapport.

²¹¹ Roodenburg, H., R. Euwals & H. ter Rele (2003), blz. 68

²¹² Roodenburg, H., R. Euwals & H. ter Rele (2003), blz. 120

²¹³ Roodenburg, H., R. Euwals & H. ter Rele (2003), blz. 71

²¹⁴ Roodenburg, H., R. Euwals & H. ter Rele (2003), blz. 66

²¹⁵ 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

²¹⁶ 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 van het huidige rapport, 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.²¹⁷ De deelname van leerlingen met niet-westerse migratieachtergrond vertoont een stijgende lijn²¹⁸, wat zou kunnen leiden tot een geringer negatief resultaat voor niet-westerse immigranten.

De methode – gegevens baseren op één peiljaar en dan op basis van CPB-verwachtingen extrapoleren naar de toekomst – noopt tot periodieke herhaling van de berekeningen. Dit zou bijvoorbeeld goed kunnen in het kader van de jaarlijkse ‘staat van immigratie’, de integrale immigratierapportage die van overheidswege wordt vervaardigd (zie ook §1.2 van het huidige rapport). Men zou dan bijvoorbeeld het verslagjaar in de gebruikte CBS-microdatabestanden telkens één jaar kunnen opschuiven. De CPB-verwachtingen zouden net als in het huidige rapport kunnen worden ontleend aan de periodiek te verschijnen CPB-vergrijzingsstudies. Zodoende zou ook een tijdreeks ontstaan waarmee de fiscale en economische effecten van immigratie- en integratiebeleid verder kunnen worden gemonitord.

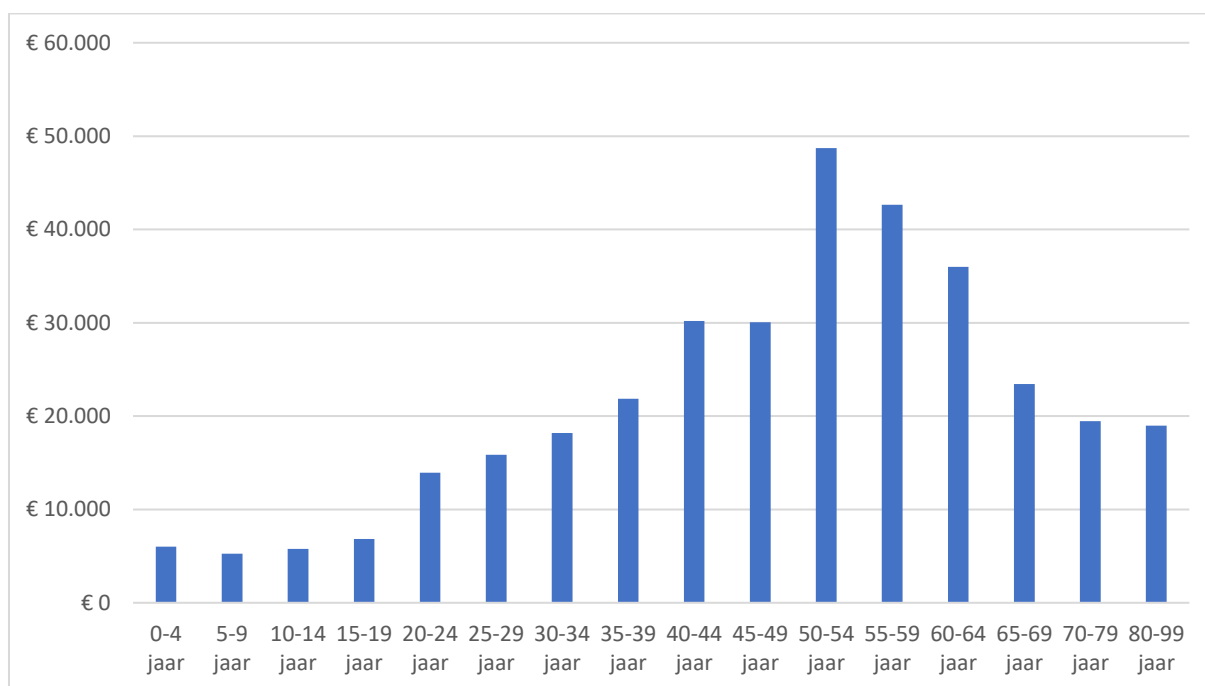
²¹⁷ Zorlu, A., J. Hartog en M. Beentjes (2010), Uitkeringsgebruik van migranten, Working Paper 10-101, AIAS, Universiteit van Amsterdam.

²¹⁸ Opgehaald 19-4-2023 van: <https://longreads.cbs.nl/integratie-2018/onderwijs/>

10 Sampling

10.1 Meetfouten

In de huidige studie bestaat een zekere spanning tussen de wens om een betrouwbare meting te geven en de wens om een gedetailleerd beeld te geven. Voor betrouwbaarheid zijn grote groepen nodig, maar bij de trapsgewijze verfijning van regio-indelingen (zie §4.4 van de appendix) worden de groepen uiteraard kleiner. Hetzelfde gebeurt bij het combineren van regio-indelingen met andere variabelen als opleiding of migratiemotief.



Figuur 10.1 Standaarddeviatie van de nettobijdrage op basis van CBS-microdata²¹⁹ per vijfjaarsleeftijdsgroep.

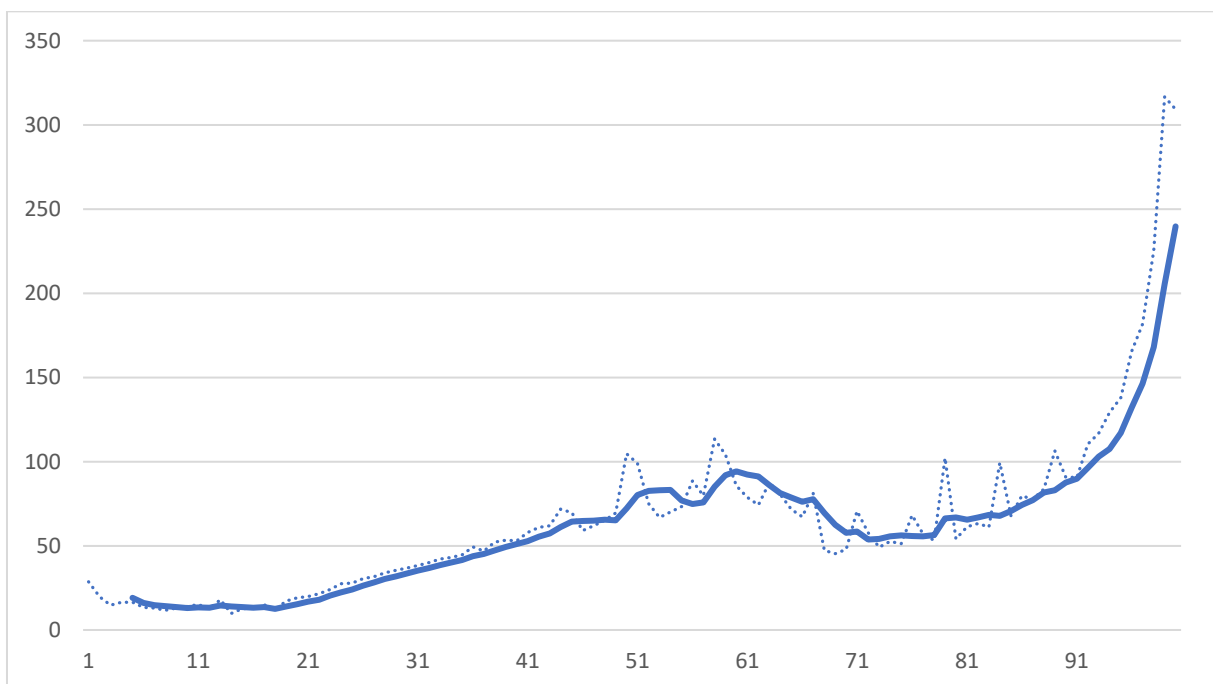
Om zo objectief mogelijk te kunnen beoordelen hoe bij de indeling in regio's (en andere aggregaties) de afweging tussen betrouwbaarheid en gedetailleerde analyse uitpakt is de sampling voor de verschillende leeftijdsgroepen bestudeerd. Voorafgaand aan de regio-indeling, het indelen in leeftijdsgroepen en dergelijke, is een schatting gemaakt van de standaarddeviatie van de uiteindelijke nettobijdrage per leeftijdsgroep, zie Figuur 10.1. Deze grafiek laat zien dat de standaarddeviaties van 0 tot 20 jaar relatief klein zijn. Vanaf 20 jaar neemt de standaarddeviatie toe. Tussen 40 en 65 jaar – en dan met name voor vijftigers – is de standaarddeviatie het hoogst om daarna weer te dalen. Dit laat zien dat het belangrijk

²¹⁹ 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 het zijn vooral die 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 veiligheidszorg, bruto investeringen in scholen, vennootschapsbelasting en IRN via bedrijven nog niet toegerekend naar personen en telden de leerlinggewichten voor 100% mee. Uiteindelijk is ook geen indeling in leeftijdsgroepen van 5, 10 en 20 jaar toegepast, maar in leeftijdsgroepen van 1, 2, 4, 8 en 10 jaar. Controle achteraf op basis van de definitieve berekeningen geeft echter vergelijkbare resultaten v.w.b. de spreiding.

is om voor leeftijden vanaf 20 jaar en vooral voor leeftijden van 40 tot 65 jaar veel waarnemingen per leeftijdsgroep te hebben.

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 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 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 groepsgrootte per leeftijdsgroep en die is voor hoge leeftijden gemiddeld laag door toenemende sterftekansen. 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.

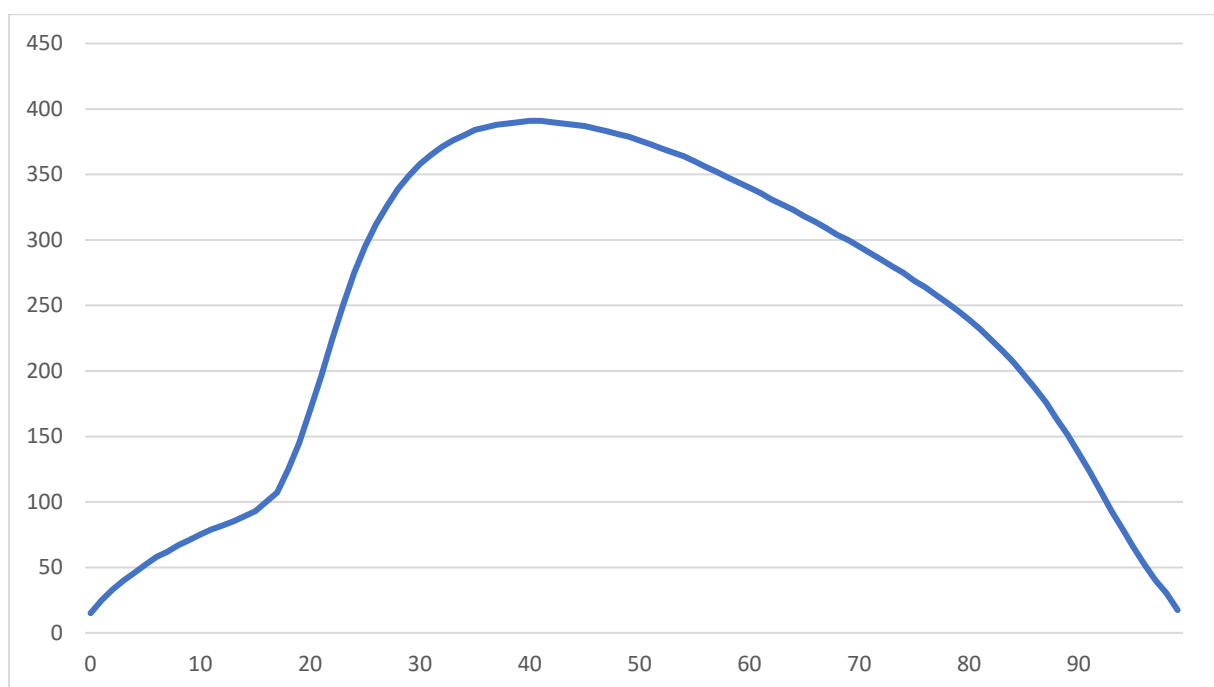


Figuur 10.2 Standaardfout van de nettobijdrage voor de gehele bevolking en 5-jaars voortschrijdend gemiddelde.

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 een rol.²²⁰ Mede door emigratie is het aantal ouderen

²²⁰ Indien van toepassing in het betreffende scenario. In het standaardscenario is sprake van emigratie.

gering, maar om dezelfde reden is het gewicht van ouderen in de totale nettobijdrage van de groep als geheel doorgaans ook gering. 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 elk leeftijdjaar afzonderlijk, de bijdrage aan de totale nettobijdrage over de levensloop is berekend van een fout van +€ 1.000 in het betreffende leeftijdjaar. 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. Effectief komt er dus maximaal 40% van een meetfout in een bepaald leeftijdjaar terecht in de nettobijdrage over de levensloop, en meestal (veel) minder.

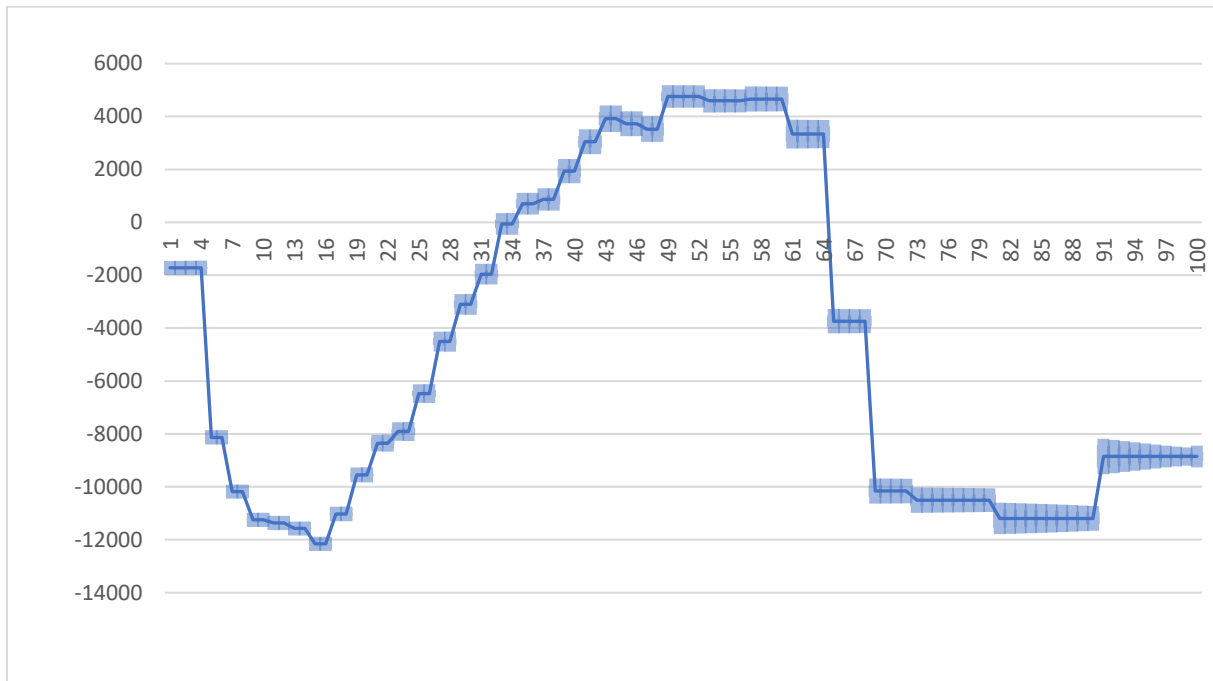


Figuur 10.3 Toename van de totale nettobijdrage over de levensloop (verticale as) als gevolg van een toename met € 1.000 van de nettobijdrage in een bepaald leeftijdjaar (horizontale as), met het immigratie- en remigratiegedrag van de gemiddelde eerste generatie immigrant en sterfte en discontering zoals in het standaardscenario.

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 groepsgrootte 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-delings 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. De oplossing van dit probleem is in §10.4 uiteengezet.



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 en gewogen naar het effect (zie Figuur 10.3) van het immigratie- en emigratiegedrag van de gemiddelde eerste generatie immigrant, en sterfte en contant maken volgens het standaardscenario.

10.2 Verdeling over celgrootteklassen

Verder is nagegaan tot welke verdeling over verschillende celgrootteklassen de sampling leidde. Met celgrootte wordt bedoeld het aantal waarnemingen per cel naar leeftijd, herkomstregio, generatie, migratiemotief en andere relevante variabelen. De celgrootte is van belang omdat de standaardfout halveert bij elke verviervoudiging van de celgrootte. Voor het indelen in celgrootteklassen is vanwege de scheve verdeling een quasi-logaritmische klassenindeling gehanteerd, zie Tabel 10.1.

De in Tabel 10.1 getoonde frequenties betreffen de belangrijkste datasets waarbij sampling een rol speelde. Om het geheel enigszins compact te kunnen weergeven zijn er drie lagen met kolomkoppen weergegeven. De eerste rij koppen geeft aan of er geaggregeerd is naar migratiemotieven of opleidingen. NVT betekent hier dat er niet geaggregeerd is naar opleiding of motief. De tweede rij koppen geeft aan of er geaggregeerd is naar generatie. NVT betekent dat er niet geaggregeerd is naar generatie. Een enkel getal geeft aan dat er selectie heeft plaatsgevonden op alleen die generatie. Meerdere getallen geven aan dat er geaggregeerd is naar de genoemde generaties. Generatie 0 staat gelijk met autochtonen (Nederlandse achtergrond). De derde rij koppen geeft weer voor welke indeling naar herkomst geaggregeerd is. NVT geeft weer aan dat er niet geaggregeerd is naar herkomst.

Tot slot is in kolom Cito voor elke tweede generatiegroep volgens de 42-delung aangegeven van hoeveel personen er citoscores beschikbaar zijn. Dit is de enige kolom waarvoor er niet ook geaggregeerd is naar leeftijd, omdat citotoetsen rond het elfde of twaalfde levensjaar worden afgenomen en in de context van huidige studie is leeftijdsonderscheid niet relevant. Voor alle andere groepen is er geaggregeerd naar leeftijdsgroepen. Daarbij zijn in beginsel (bijvoorbeeld voor de 42-delung) de volgende leeftijdsgroepen gemaakt: van 0 tot 72 jaar per vier jaar, van 72 tot 80 jaar per vier jaar, van 80 tot 100 jaar per 10 jaar. In voorkomende gevallen is hiervan afgeweken, als dat nodig was i.v.m. de sampling of de data een verdere verfijning toelieten, zoals bij heel grote groepen als westers en niet-westerse 1^e generatie immigranten of bij jonge 2^e generatie immigranten.

Tabel 10.1 Aantal waarnemingen per leeftijdsgroep. De modale klasse is vetgedrukt.

Uitsplitsing naar	Motief, Opleiding, Cito	NVT						Motief			Opleiding		Cito ^a
	Generatie ^b	NVT	0, 1, 2	1			2 ^c	1			NVT	0,1,2	1,2
	Herkomst ^d	NVT	NVT	Land-type	12-delung	42-delung	42-delung ^e	NVT	Land-type	12-delung	NVT	Land-type	42-delung
Celgrootteklasse	10-30	0	0	0	0	10	6	1	5	60	0	10	0
	31-100	0	1	0	9	42	74	4	13	144	3	55	1
	101-300	0	5	2	24	139	326	7	18	213	14	130	4
	301-1.000	0	8	10	61	320	646	8	38	276	18	308	25
	1.001-3.000	0	15	25	73	297	289	18	45	183	36	623	26
	3.001-10.000	4	45	72	120	213	73	34	75	93	60	855	19
	10.001-30.000	5	76	80	64	77	12	35	31	8	126	938	4
	30.001-100.000	11	72	1	1	1	0	10	5	2	215	1.021	3
	> 100000	80	78	0	0	0	0	0	0	0	15	9	0
Totaal	100	300	190	352	1.099	1.426	117	230	979	487	3.949	82	
10-100 als % totaal	0%	0%	0%	3%	5%	6%	4%	8%	21%	1%	2%	0%	

^a Dit betreft het totaal aantal personen van de tweede generatie waarvan de citoscore bekend is; dit is de enige variabele waarbij niet ook nog geaggregeerd is naar leeftijd. ^b Generatie 0 staat voor autochtoon. ^c Tot 48 jaar, zie §4.2 voor meer details. ^d Landtype is de CBS-aanduiding voor de indeling westers, niet-westers. ^e Waarbij voor herkomst Indonesië, Turkije, Marokko, Suriname en Aruba en de (voormalige) Nederlandse Antillen verder uitgesplitst is naar het aantal in het buitenland geboren ouders.

Voor elke celgrootteklasse is in de tabel het waargenomen aantal groepen gegeven dat binnen de betreffende celgrootteklasse valt, met andere woorden, hoeveel combinaties van leeftijd en/of herkomst en/of generatie en/of opleiding, cito dan wel motief in de betreffende klasse vallen. In verband met de leesbaarheid zijn deze combinaties hierna onderstreept.

De modale klasse (vetgedrukt in de tabel) is in alle gevallen de klasse 301-1.000 of hoger. Bij de bespreking van de tabel nemen we als uitgangspunt het percentage celgroottes dat in de klassen 10-30 of 31-100 valt, kortweg aangeduid als de klasse 10-100 (zie voor die percentages de rij '10-100 als % totaal' in de tabel).

Vier combinaties hebben (afgerond) 0% waarnemingen in de klasse 10-100, te weten: NVT × NVT × NVT (dit is de hele bevolking), NVT × Generatie 0, 1, 2 × NVT (dit is de bevolking uitgesplitst naar generatie), NVT × Generatie 1 × Landtype (dit zijn 1^e generatie immigranten uitgesplitst naar westers en niet-westers) en Cito × Generatie 2 × 42-deling. Deze combinaties laten we buiten beschouwing.

Voor de indeling Generatie 2 × 42-deling is in §2.4 reeds beschreven hoe het gebrek aan data is ondervangen. Deze indeling blijft hier verder buiten beschouwing. Voor de combinaties Opleiding × NVT × NVT en Opleiding × Generatie 0, 1, 2 × Landtype is in §4.2 beschreven hoe voor elk onderwijsniveau van de SOI 8-deling een synthetisch nettobijdrageprofiel is gemaakt op basis van de relevante nettobijdrageprofielen voor actieve onderwijsdeelnemers en personen die niet meer aan het onderwijs deelnemen. Hierbij wordt voor leeftijden tot 38 jaar bij MBO2, MBO3, MBO4, bachelor- en masteropleidingen het gewogen gemiddelde genomen van studerende en niet studerende. Een flink deel van de kleine celgroottes betreft personen die op oudere leeftijd studeren. Weinig dertigers volgen een MBO opleiding bijvoorbeeld. Indien bij deze groepen sprake was van kleine aantallen waarnemingen is dit in eerste instantie opgelost door de leeftijdsgroepen te vergroten. Uiteraard levert dit voor deze groepen vertekening op, maar die weegt nauwelijks mee in de totale nettobijdrage over de levensloop. Immers, als het aandeel dertigers dat een MBO opleiding volgt zo gering is (om bij het eerder genoemde voorbeeld te blijven), dan is het effect van eventuele meetfouten in de totaalberekening ook verwaarloosbaar.²²¹ De combinaties met opleiding blijven hierna ook buiten beschouwing.

De bespreking focust op de vijf resterende groepen. Hierbij gaat het om 1^e generatie immigranten uitgesplitst naar 12- of 42-deling (NVT × Generatie 1 × 12-deling en NVT × Generatie 1 × 42-deling). Daarnaast draait het om de combinaties voor 1^e generatie immigranten met migratiemotief en regio, te weten Motief × Generatie 1 × NVT (uitsplitsing naar alleen motief), Motief × Generatie 1 × Landtype (uitsplitsing naar combinaties van motief en westers/niet-westers) en Motief × Generatie 1 × 12-deling (uitsplitsing naar motief en regio volgens de 12-deling). Voor de laatste combinatie is het aandeel celgroottes in de klasse 10-100 met 21% het grootst.

Het probleem van kleine celgroottes bestaat uit twee kwesties. De eerste kwestie is hoe men het beste cellen kan aanvullen met weinig of geen waarnemingen. Dit wordt in §10.3 behandeld. De tweede kwestie is hoe men het beste kan omgaan met grote spikes die het gevolg kunnen zijn van weinig waarnemingen. De oplossing voor dat vraagstuk wordt in §10.4 uiteengezet. Voordat we dat bespreken enkele opmerkingen over het probleem van kleine celgroottes.

Allereerst zijn op basis van de verkenning in §10.1 de regio's in de 42-deling zodanig ingedeeld dat er voor leeftijden tussen 20 en 72 jaar veel waarnemingen zijn voor de 1^e generatie. Bij de regio-indelingen heeft de relatief kleine celgrootte dan ook voor een groot deel betrekking op groepen waarvoor de consequenties van meetfouten relatief klein zijn. Veelal betreft dit jongeren (vaak de leeftijdsgroep 0-3 jaar) en ouderen (vaak de leeftijdsgroepen 80-89 en 90-99 jaar). Zoals in §10.1 uiteengezet is het effect van meetfouten voor zeer jonge en oude mensen op de totale nettobijdrage gering. Wel zijn er enkele wat uitzonderlijke groepen. Japan (bijna 2.000 waarnemingen) is vanwege bestaande CBS-indelingen min of meer noodgedwongen apart ingedeeld in de 42-deling. Israël (ruim 4.000 waarnemingen) is eveneens ondanks weinig waarnemingen apart ingedeeld vanwege de grote verschillen met

²²¹ Vergelijk dit met het door het wege tegen het immigratieprofiel verwaarloosbare effect van meetfouten door het ontbreken van data voor arbeidsmigranten tot 16 jaar en studiemigranten tot 12 jaar.

omliggende landen. Voor de overige groepen in de 42-deling varieert de groepsomvang van 6.000 tot 260.000 personen.

Bij motief gaat het bij de kleine celgroottes voor een deel om zeer jonge studie- en arbeidsmigranten. Dit is echter voor de uiteindelijke berekening van de nettobijdrage geen enkel probleem. Door het wegen tegen het immigratieprofiel is het effect op de totale nettobijdrage over de levensloop verwaarloosbaar en wel om dezelfde reden: er komen weinig zeer jonge studie- en arbeidsmigranten, dus in het gewogen gemiddelde is hun aandeel als percentage van alle studie- respectievelijk arbeidsmigranten verwaarloosbaar.

Tabel 10.2 Aantal waarnemingen voor 1^e generatie immigranten, voor combinaties van migratiemotief en herkomstregio volgens de 12-deling.

Regio/Motief	Asiel	Gezin	Studie	Arbeid	Overig	Onbekend	Totaal
Westers	10.532						10.532
Overig niet-westers	5.106						5.106
West-Azië	69.345						69.345
Afrika	31.884	42.996	7.884	9.086	13.427	12.783	118.060
Azië		84.206	27.323	26.279	21.739	14.562	174.109
Latijns-Amerika		20.977	7.381	6.491	3.003	11.440	49.292
Europese Unie		91.247	24.253	186.223	41.859	30.280	373.862
Overig Europa		30.040	9.079	11.682	16.118	4.424	71.343
Overig buiten Europa		12.068	4.894	9.373	3.219	5.566	35.120
Indonesië		7.247	3.529	713	917	5.227	17.633
Marokko		43.847	1.861	2.382	1.938	13.523	63.550
Suriname		19.806	3.143	794	1.767	25.227	50.737
Turkije		52.154	2.424	3.719	1.960	10.908	71.164
(Voormalige) Antillen						53.606	53.606
Totaal	116.867	404.588	91.769	256.742	105.946	187.546	1.163.458

Een groter probleem is dat er voor studie- en arbeidsmigranten ook weinig waarnemingen zijn voor hogere leeftijden, want dat bemoeilijkt de berekening over de resterende levensloop, al gelden ook hier de relativerende observaties van §10.1 (meetfouten voor hogere leeftijden tellen niet zwaar in de totale nettobijdrage). Vooral bij de sampling van combinaties van migratiemotief met de CBS-12 deling in herkomstregio's speelt dat er enkele groepen zijn met een laag totaal aantal waarnemingen. I.h.b. gaat het om arbeidsmigranten uit Suriname en Indonesië en immigranten met Overig motief uit Indonesië. Van deze drie groepen ligt het totaal aantal waarnemingen tussen 700 en 900. Gemiddeld ligt het aantal waarnemingen per leeftijdsgroep voor deze drie groepen op 67 en per leeftijdsjaar op 16. Zoals in §10.3 uiteengezet zal worden is voor deze groepen het gebruikte nettobijdrageprofiel voor een groot deel van het leven synthetisch is. Dat wil zeggen, dat de waarnemingen voor relatief veel leeftijdsgroepen zijn aangevuld met gegevens van de bovenliggende groep en/of een profiel dat gefit is op de waarnemingen die er wel zijn. Hiermee moet rekening gehouden worden bij de interpretatie van de resultaten voor deze groepen. Tegelijkertijd moet hierbij een belangrijke relativerende kanttekening worden geplaatst. Het uitsplitsen naar de combinaties van motief met de CBS 12-deling in regio's is vooral gedaan om de totale nettobijdrage van recente en toekomstige immigratie te kunnen bepalen (zie Hoofdstuk 7 van het huidige rapport) en daarbij aan te sluiten bij de CBS-prognoses die volgens de 12-deling worden uitgevoerd. Op de berekening van de totale nettobijdrage heeft

eventuele vertekening een verwaarloosbare invloed en wel juist omdat het probleem van de lage aantallen voor genoemde combinaties van migratiemotief en herkomstregio ontstaat omdat er zo weinig immigranten komen voor die combinaties. Eventuele meetfouten tellen dus slechts voor een heel klein deel mee in de totaalberekening. Voor de meeste combinaties van motief en herkomst is het aantal waarnemingen overigens ruimschoots voldoende, zie Tabel 10.2.

10.3 Behandeling van leeftijdsgroepen met weinig waarnemingen

Voor zover er bij de regio-indelingen sprake was van kleine aantallen ouderen en jongeren is dat bij de eerste generatie²²² verholpen door het gebruik van de gegevens van de bovenliggende regio. Hierbij zijn twee verschillende hiërarchieën onderscheiden voor het bovenliggende niveau. Voor de regio's is de indeling AlleHerkomsten²²³ > Landtype > CBS 12-deling > 19-deling > 42 deling > 87-deling (zie §4.4) aangehouden. Bij combinaties van regio met motief is de indeling AlleMotieven × AlleRegios²²⁴ > Motief × AlleRegios > Motief × Landtype > Motief × CBS 12-deling > Motief × CBS 19-deling aangehouden.²²⁵

Er 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 motief dus het bovenliggende niveau voor motief × landtype, vervolgens is motief × landtype weer het bovenliggende niveau voor motief × 12-deling en tot slot is motief × 12-deling weer het bovenliggende niveau voor motief × 19-deling.²²⁶ 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.

Het aanvullen van te weinig of ontbrekende data uit het bovenliggende niveau is op twee manieren 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 gevallen²²⁷ 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

²²² Voor de tweede generatie is een andere oplossing gekozen, zie §2.4.

²²³ Hierin is 'AlleHerkomsten' gelijk aan alle 1^e generatie immigranten.

²²⁴ Alleen voor de motieven arbeid en gezinsmigratie.

²²⁵ Hierin is 'AlleMotieven × AlleRegios' gelijk aan alle 1^e generatie immigranten.

²²⁶ Concreet: voor westerse arbeidsmigranten is Arbeid het bovenliggende niveau; voor arbeidsmigranten uit de Europese Unie is Arbeid x Westers het bovenliggende niveau; voor arbeidsmigranten uit de GIPS-landen is Arbeid x Europese Unie het bovenliggende niveau.

²²⁷ Voor herkomstregio's Noord-Afrika, Sub-Sahara Afrika en Indisch subcontinent.

leeftijdsgroepen 0-3 jaar,²²⁸ 80-89 jaar²²⁹ en 90-99 jaar²³⁰ data van de bovenliggende regio gebruikt. Deze profielen zijn dus synthetisch, in die zin dat ze opgebouwd zijn uit profielen voor verschillende groepen.

Ook bij migratiemotief 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 onbekend²³¹ 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 – 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 op te lossen en met name voor de hogere leeftijden te komen tot een betere schatting is een algoritme ontwikkeld. Dit algoritme verloopt in drie stappen: aanvullen van ontbrekende gegevens, schatten van

²²⁸ Voor de deelregio's van herkomstregio's Sub-Sahara Afrika, Indisch subcontinent, Oost-Azië en Latijns-Amerika (excl. Suriname, Aruba en Antillen).

²²⁹ Voor de deelregio's van herkomstregio's Sub-Sahara Afrika, Indisch subcontinent en Overig buiten Europa.

²³⁰ 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).

²³¹ 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.

een best passend profiel en een naar aantallen waarnemingen gewogen gemiddelde van beiden nemen.

Als er helemaal geen waarnemingen zijn, worden voor de eerste stap de betreffende gegevens zonder meer overgenomen van de overeenkomstige gegevens van het bovenliggende niveau. Dit gebeurt – net als bij handmatig aanvullen – volgens de twee hiervoor beschreven hiërarchieën.

Indien er te weinig data zijn is de aanpak iets complexer. ‘Te weinig data’ is geoperationaliseerd als ‘50 waarnemingen of minder. De aanpak is per groep van deelposten verschillend. Een aantal posten zijn per groep semihandmatig berekend. Het gaat om de posten Openbaar Bestuur (inclusief Veiligheidszorg), Defensie, Overdrachten buitenland, Bruto investeringen gebouwen en infrastructuur, Erfbelasting, VPB/dividendbelasting, IRN van bedrijven, Netto grondverkopen en Niet-belastingmiddelen-rest (Post 1, 2, 10, 13, 14, 18, 19, 21, 22, 23 in Tabel 5.1). Deze posten zijn ofwel gelijk voor elke groep, of ze zijn met een aantal tussenstappen per groep apart semihandmatig berekend zoals beschreven in Hoofdstuk 6 van deze appendix. Deze posten blijven hier daarom verder buiten beschouwing.

Voor de overige posten is de volgende aanpak gevolgd. Stel dat P_j een element is van leeftijdsprofiel \vec{P} voor post p voor een bepaalde groep, voor leeftijdsjaren $0 \leq j \leq 99$. Stel verder dat N_j een element is van leeftijdsprofiel \vec{N} van de aantallen waarnemingen per leeftijdjaar²³² voor de betreffende groep, ook weer met $0 \leq j \leq 99$.

Als eerste stap worden bij weinig waarnemingen zowel \vec{P} als \vec{N} aangepast door voor de hogere leeftijden de waargenomen bedragen respectievelijk aantallen te vervangen door het gemiddelde bedrag voor post p respectievelijk het gemiddeld aantal waarnemingen voor de betreffende groep. ‘Weinig waarnemingen’ is geoperationaliseerd als ‘minder dan 100’ en ‘hogere leeftijden’ is geoperationaliseerd als ‘vanaf 90 jaar’, ‘vanaf 80 jaar’ of ‘vanaf 72 jaar’. De leeftijden 72, 80, 90 zijn gekozen omdat 72-79, 80-89 en 90-99 leeftijdsgroepen zijn. Daarbij is uitgegaan van de jongste leeftijd $j' \in \{72, 80, 90\}$ waarvoor gold dat $\sum_{j=j'}^{99} N_j < 100$. Als een dergelijke j' niet bestond zijn er geen aanpassingen doorgevoerd. Als een dergelijke j' wel bestond, is er eerst als volgt een ondergrens L bepaald: $L = 68$ als $j' = 72$, $L = 72$ als $j' = 80$ en $L = 80$ als $j' = 90$. Vervolgens zijn de bedragen P_j respectievelijk waarnemingen N_j voor $L \leq j \leq 99$ vervangen door $\sum_{j=L}^{99} N_j \cdot P_j / \sum_{j=L}^{99} N_j$ respectievelijk $\sum_{j=L}^{99} N_j / (100 - L)$.²³³ Wat deze stap doet is een laag aantal waarnemingen voor leeftijdsgroepen die in de pensioenleeftijd vallen, middelen om zodoende voor hogere leeftijden eventuele grote ‘spikes’ op het betreffende leeftijdsprofiel te verminderen.

Vervolgens is er een wegingsprofiel \vec{W} gemaakt op basis van de oorspronkelijke²³⁴ vector \vec{N} :

²³² De samplegroottes zijn hiervoor gedeeld door de omvang van de leeftijdsgroep in jaren.

²³³ Een alternatief is om te berekenen $\sum_{j=L}^U N_j / (U + 1 - L)$ met $U > L$ het kleinste gehele getal waarvoor geldt dat $N_U > 0$. De uitkomsten verschillen dan echter marginaal (voor de 77 groepen waarvoor er een verschil is, is het absolute verschil in 70 gevallen kleiner dan € 1.000 in zes gevallen tussen € 1.000 en € 2.000 en in één geval is het verschil circa € 2.200. Weer een ander alternatief is om de aantallen N_j helemaal niet te middelen en ook dan zijn de verschillen beperkt.

²³⁴ Dit is gedaan om er voor te zorgen dat er altijd ‘verankering’ in het bovenliggende profiel plaatsvindt als er helemaal geen waarnemingen zijn voor een bepaalde leeftijdsgroep. Vaak betreft dit arbeidsmigranten of immigranten met ‘overig motief’.

$$W_j = \text{Min}\left(1, \frac{N_j}{50}\right) \text{ als } N_j > 0 \text{ en } W_j = 0,5 \text{ als } N_j = 0 \text{ voor } 0 \leq j \leq 99$$

Dit is dus een vector die 1 is als er minstens 50 waarnemingen zijn, 0,5 is als er 0 waarnemingen zijn en $\frac{N_j}{50}$ is als het aantal waarnemingen N_j groter dan 0 en kleiner dan de drempelwaarde 50 is. Het onderdeel $W_j = 0,5 \text{ als } N_j = 0$ van bovenstaande definitie zorgt in de hiernavolgende stap voor ‘verankering’ in het bovenliggende profiel voor die (eventuele) waarden van j waarvoor er helemaal geen waarnemingen voor P_j zijn, m.a.w. waarvoor $N_j = 0$.

Vervolgens wordt met een *brute force* methode de i gezocht waarvoor de onderstaande gewogen kwadratensom KS minimaal is:

$$KS = \sum_{j=0}^{99} W_j \cdot \left(P_j - \frac{i \times P_j^{Up}}{100}\right)^2 \text{ voor } i = 1, 2, 3, \dots \text{ en } 0 \leq j \leq 99$$

Hierin is P_j^{Up} een element van profiel \vec{P}^{Up} voor post p voor het bovenliggende niveau (zie eerdere uitleg over de twee hiërarchieën voor de indeling van regio’ en combinaties van regio’s en migratiemotief). De i waarvoor KS minimaal is wordt i' genoemd. Deze stap zoekt een veelvoud van het bovenliggende profiel – d.w.z. het scalaire product $FIT_{i'} = i' \times P^{Up}/100$ van het bovenliggende profiel met i' – dat het best past bij het profiel \vec{P} . Daarbij wordt rekening gehouden met leeftijdsjaren met ‘weinig’ waarnemingen, d.w.z. minder dan 50 waarnemingen. Als er voor een bepaald leeftijdsjaar j helemaal geen waarnemingen zijn voor post p dan geldt door het aanvullen vanuit het bovenliggende profiel voor dat jaar dat $P_j = P_j^{Up}$ en wordt het kwadraat van het verschil voor de helft meegewogen.

Tot slot wordt het uiteindelijke profiel \vec{P}' voor post p als volgt berekend:

$$P'_j = W_j \cdot P_j + (1 - W_j) \cdot \frac{i' \times P_j^{Up}}{100} = W_j \cdot P_j + (1 - W_j) \cdot FIT_{i'} \text{ voor } 0 \leq j \leq 99$$

Dit profiel \vec{P}' is gelijk aan de waarnemingen voor leeftijdsjaren waarvoor er minstens 50 waarnemingen zijn. Voor leeftijdsjaren waarvoor er minder dan 50 waarnemingen zijn is het profiel \vec{P}' het gewogen gemiddelde van de waarnemingen en het gefitte profiel $FIT_{i'} = i' \times P^{Up}/100$.

Bij de aanvulling uit het bovenliggende niveau is dus 50 waarnemingen per leeftijdsjaar²³⁵ als drempelwaarde aangehouden. Om dit te verduidelijken een concreet voorbeeld met fictieve, ronde getallen. Stel dat voor arbeidsmigranten uit Indonesië 120 waarnemingen zijn voor de vierjaarsleeftijdsgroep 60-63 jaar. Per leeftijdsjaar zijn dat dus $120 / 4 = 30$ waarnemingen. Dat is onder de norm van 50 waarnemingen per leeftijdsjaar. In dit geval is het bovenliggende niveau ‘Westerse arbeidsmigranten’ (Indonesië is door het CBS als westers land geclassificeerd). Stel dat de belastingpost LIS voor Indonesische arbeidsmigranten van 60-63 jaar gemiddeld € 40.000 bedraagt en voor Westerse arbeidsmigranten van 60-63 jaar gemiddeld € 30.000. Dan wordt de post LIS_{60} voor 60-jarigen als volgt berekend:

²³⁵ In de praktijk voor leeftijden tot 72 jaar meestal 200 waarnemingen per vierjaarsleeftijdsgroep en (minder vaak) 100 waarnemingen per tweejaarsleeftijdsgroep, 400 waarnemingen per achtjaarsleeftijdsgroep of 500 waarnemingen per tienjaarsleeftijdsgroep.

$$LIS_{60} = \frac{30}{50} \times \text{€ } 40.000 + \frac{20}{50} \times \text{€ } 30.000 = \text{€ } 24.000 + \text{€ } 12.000 = \text{€ } 36.000$$

Het achterliggende idee van het algoritme is dat de leeftijdsprofielen van kosten en batenposten die grotendeels of geheel gebaseerd zijn op microdata voor grote groepen (zoals de hele bevolking of alle eerste generatie immigranten samen) een kenmerkend en 'glad' verloop hebben (zie voor enkele voorbeelden §5.3 en §5.5-5.7 van deze appendix). Bij heel kleine groepen (zoals arbeidsmigranten uit Suriname en Indonesië) zijn de profielen vaak grillig door weinig waarnemingen en/of ontbreken er stukken (geen waarnemingen). Wat het algoritme ruwweg doet is de stukken met geen of weinig waarnemingen 'gladmaken' als ze in de pensioenfase vallen en vervolgens zowel aanvullen vanuit het bovenliggende niveau als ook aanvullen vanuit een profiel dat geijkt is op alle beschikbare waarnemingen en dezelfde vorm heeft als het overeenkomstige profiel voor het bovenliggende niveau. De profielen P_j' worden achtereenvolgens van het hoogste niveau tot het laagste niveau berekend en opgeslagen, zodat het profiel van het bovenliggende niveau altijd een geijkt en aangevuld profiel is als het wordt gebruikt voor een onderliggend niveau.

Het effect van dit algoritme hangt af van de omvang van groepen en de verdeling van de beschikbare data over de leeftijdsgroepen. Hierna wordt dat besproken voor de nettobijdrage voor de 1^e generatie volgens het standaardscenario met remigratie. De getallen worden voor de leesbaarheid afgerond. Voor alle 1^e generatie immigranten samen is het verschil tussen wel of niet het algoritme toepassen € 0. Kijken we naar de 42-delings dan is het effect voor grote herkomstgroepen als Marokko ($N \approx 220.000$) en Indonesië ($N \approx 260.000$) verwaarloosbaar ($\approx \text{€ } 250$ respectievelijk $\approx \text{€ } 50$). Bij herkomstgroepen met een mediane omvang ($N \approx 17.000$) zoals Spanje en India zijn de verschillen vergelijkbaar of groter ($\approx \text{€ } 150$ respectievelijk $\approx \text{€ } 900$). De twee kleinste herkomstgroepen in de 42-delings zijn Japan en Israël. Bij Japan ($N \approx 2.000$) – dat apart is geclassificeerd i.v.m. de bestaande CBS-indeling – zijn de verschillen tussen niet ($\text{€ } 197.000$) of wel ($\text{€ } 180.000$) toepassen van het algoritme met $\text{€ } 17.000$ fors. Ook voor Israël ($N \approx 4.000$) – dat apart is ingedeeld omdat het zo verschilt van buurlanden – is het verschil aanzienlijk ($\approx \text{€ } 2.300$). Vooral bij combinaties van motief en herkomstregio zijn de verschillen groot en dan met name voor motieven als arbeid en studie waarbij data voor lagere en (belangrijker) hogere leeftijden ontbreken.²³⁶ Als voorbeeld nemen we het motief arbeid. Bij arbeid is het bovenliggende niveau 'alle 1^e generatie immigranten'. Alleen aanvullen vanuit dit bovenliggende niveau levert een nettobijdrage van $\text{€ } 142.000$. Aanvullend ook nog het algoritme toepassen levert een nettobijdrage van $\text{€ } 153.000$, dus een verschil van $\text{€ } 11.000$. Dat komt omdat alleen aanvullen vanuit de gemiddelden voor 1^e generatie immigrant onderschatting geeft (arbeidsmigranten presteren bovengemiddeld, dus aanvullen vanuit de gemiddelde 1^e generatie immigrant levert te lage bedragen) en het algoritme vermindert die onderschatting. Bij de schatting van bijvoorbeeld niet-westerse arbeidsmigranten ($N = 85.000$) levert het algoritme een schatting op die $\text{€ } 9.000$ hoger is, puur doordat ontbrekende data voor hogere leeftijden worden aangevuld. Bij arbeid is de herkomst nog verder onderverdeeld naar de CBS 12-delings. De grootste groep wordt gevormd door arbeidsmigranten uit de EU ($N \approx 190.000$) waar toepassing

²³⁶ 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.

van het algoritme een daling van circa € 1.000 veroorzaakt. Voor arbeid zijn er ook twee heel kleine groepen namelijk arbeidsmigranten uit Indonesië ($N \approx 700$) en Suriname ($N \approx 800$). Voor arbeidsmigranten uit Indonesië levert toepassing van het algoritme bijvoorbeeld een stijging op van € 9.000 naar € 21.000, ofwel € 11.000 hoger. Uiteraard zijn de effecten van het algoritme ook gevoelig voor de gemaakte keuzen. Wordt de drempelwaarde bijvoorbeeld verlaagd van 50 naar 25, dan verschilt de uitkomst bij Israël circa € 100, bij Japan circa € 1.500 en bij arbeidsmigranten uit Indonesië circa € 3.000.

De hiervoor geschetste aanpak is bewerkelijk, maar daar staan enkele voordelen tegenover. Verankerung in het bovenliggende niveau is een waarborg tegen extreme uitslagen voor leeftijdsgroepen waarvoor er echt heel weinig of zelfs geen data zijn. Doordat bij het fitten van het profiel FIT_j het aantal waarnemingen meegewogen wordt in de kwadratensom worden cohorteffecten gedempt. Immers, recente cohorten met een korte verblijfsduur zijn sterker vertegenwoordigd in de leeftijden (twintigers en jonge dertigers) waarin het immigratieprofiel gewoonlijk piekt (en worden daarom zwaarder meegewogen bij het fitten) dan eerdere cohorten met een verblijfsduur van bijvoorbeeld 20 jaar die een stuk ouder zijn. 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.

10.4 Behandeling van enkele extreme *outliers* (uitbijters)

De kleine aantallen waarnemingen leidt in een aantal gevallen tot een ander probleem: extreme *outliers* – in de praktijk immigranten met een heel grote positieve nettobijdrage – kunnen een onevenredig grote invloed krijgen op de nettobijdrage van de hele groep. Dit treedt met name op bij de arbeidsmigranten uit Latijns-Amerika en de regio Overig buiten Europa, die Japan, Noord-Amerika en Oceanië omvat.

Om die reden is voor combinaties van arbeid en motief, steeds gekeken wat de 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. Vooral bij de laatste twee regio's was het effect soms groot omdat de aantallen oudere arbeidsmigranten uit die regio's klein zijn. Onder arbeidsmigranten uit de EU bevinden zich ook behoorlijk wat grootverdieners, maar daar zijn de leeftijdsgroepen zo groot dat het niet zoveel effect heeft op het gemiddelde.

Bij arbeidsmigranten uit Latijns-Amerika en Overig buiten Europa leidt dit in een aantal gevallen echter tot vertekening doordat een klein aantal extreme grootverdieners in een leeftijdsgroep met geringe omvang vallen. 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 nettobijdrage van de groep als geheel al substantieel kunnen toe- (of afnemen), doordat één grootverdiener in een andere leeftijdsgroep valt die fors kleiner (of groter) is. Opeenvolgende leeftijdsgroepen kunnen in omvang een factor 2 à 3 schelen, niet opeenvolgende leeftijdsgroepen een veelvoud daarvan. Toeval krijgt zo een onaanvaardbaar groot effect op het eindresultaat.

Om dit concreet te maken een voorbeeld aan de hand van fictieve, ronde getallen voor een denkbeeldige groep van 10.000 arbeidsmigranten uit regio X. Stel dat iemand uit deze groep een nettobijdrage

heeft van 10 miljoen euro. Stel verder dat deze persoon in een leeftijdsgroep met 1000 personen valt en dat het om een vierjaarsleeftijdsgroep gaat (de grootverdieners waar het hier om gaat vallen in de praktijk allemaal in vierjaarsleeftijdsgroepen). De bijdrage van deze ene persoon aan het totale nettobijdrageprofiel van de betreffende groep bedraagt dan $4 \times \text{€ } 10.000.000 / 1000 = \text{€ } 40.000$. Effectief komt daarvan in het standaardscenario (mét remigratie) al gauw 30% (zie Figuur 10.3, de meeste grootverdieners zijn in de praktijk veertiger, vijftigers of zestigers) terecht in de totale nettobijdrage over de levensloop van de betreffende groep. Stel voor het gemak in de rest van dit voorbeeld dat het percentage precies 30% is. Dat betekent in dit voorbeeld dus die ene grootverdiener die in een leeftijdsgroep met 1000 personen valt in zijn of haar ééntje de gemiddelde nettobijdrage van die 10.000 arbeidsmigranten uit regio X verhoogt met € 12.000 (30% van € 40.000). Dat is op zich al een fors effect. Valt diezelfde persoon echter in een leeftijdsgroep met 100 personen, dan verhoogt dat de totale gemiddelde nettobijdrage van zijn groep met $4 \times \text{€ } 10.000.000 / 100 = \text{€ } 400.000$. Effectief is dat dan € 120.000 (30% van € 400.000). Er ontstaat dan dus 'spike' op het nettobijdrageprofiel die zo groot is dat de gemiddelde nettobijdrage, berekend over 10.000 personen, ruim een ton hoger uitvalt, puur omdat die ene grootverdiener in een kleinere leeftijdscategorie valt.

Dat is een forse (aan één individu toe te schrijven) nettobijdrage, maar het is kennelijk een empirisch feit dat er zich onder bepaalde groepen soms personen met een zeer hoge nettobijdrage bevinden. Verwijderen van deze personen is dus geen goede optie.

Om tot een oplossing te komen is het probleem verder geanalyseerd. Opvallend is het grote verschil in leeftijdsverdeling van alle arbeidsmigranten uit regio X (kortweg de 'totale populatie') en de Top 100 arbeidsmigranten met de grootste nettobijdrage uit regio X (kortweg de 'Top 100'). Dit is schematisch weergegeven in Figuur 10.5.²³⁷ De kans dat een arbeidsmigrant in de leeftijdscategorie 36-39 valt is vele malen groter dan de kans dat een arbeidsmigrant in de leeftijdscategorie 68-71 valt. De verdeling van de totale populatie is dus erg scheef voor de getoonde leeftijdsgroepen.

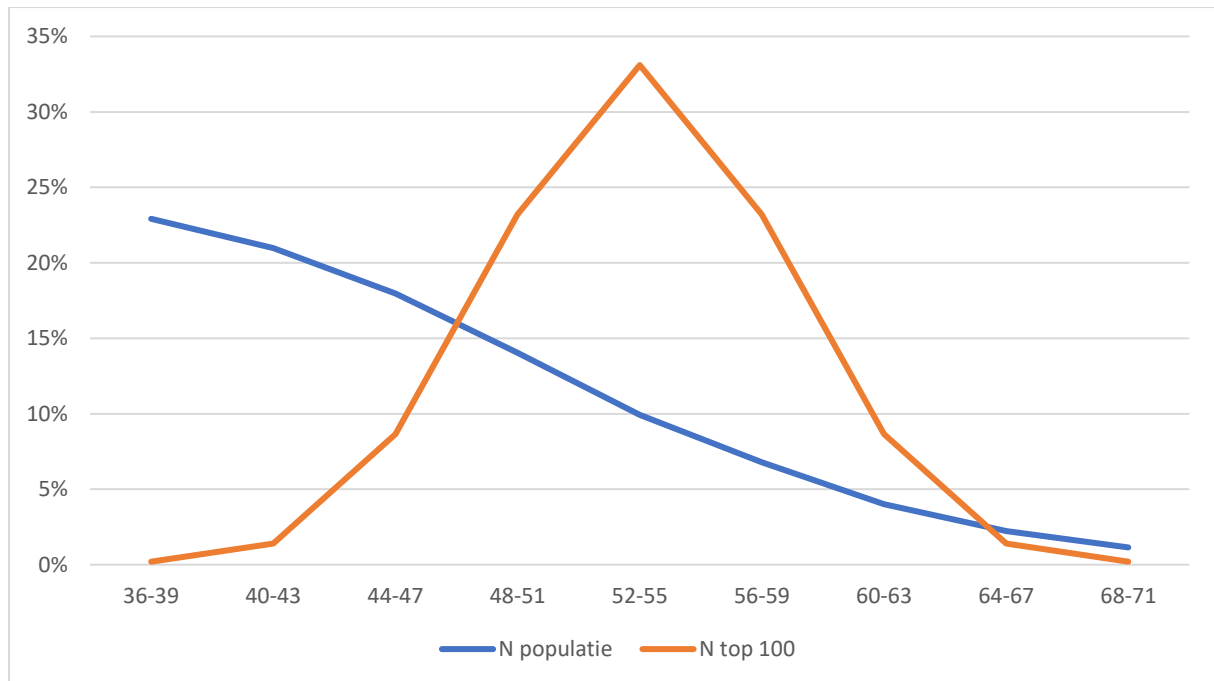
De leeftijdsverdeling van de Top 100 met hoogste nettobijdragen is echter (bij benadering) normaal verdeeld voor leeftijden tussen 36 en 71 met een piek in de leeftijdscategorie 52-55. Een normale verdeling is symmetrisch. De kans dat een veelverdiener in de leeftijdsgroep 40-43 valt is (ongeveer) even groot als de kans dat hij of zij in de leeftijdsgroep 64-67 valt. Het probleem is nu dat binnen de Top 100 de verdeling van nettobijdragen scheef is: er zijn maar een beperkt aantal extreem hoge nettobijdragen. De kans op een extreem hoge nettobijdrage is daarom zeer gering, maar het potentiële effect groot. Daarom gaat toeval – bijvoorbeeld: valt die ene persoon in de relatief grote leeftijdsgroep 40-43 of de relatief kleine leeftijdsgroep 64-67 – een grote rol spelen. Dat ondermijnt de betrouwbaarheid van de meting.

De kern van het probleem is dus dat het aantal echte 'grootverdieners' zo gering is, dat een kansgebeurtenis zoals 'er valt een persoon met een nettobijdrage van 5 miljoen of meer in leeftijdscategorie Y' tot onaanvaardbaar grote, door toeval veroorzaakte meetfouten leidt.

Er zijn verschillende oplossingen overwogen. Zoals gezegd is verwijderen van de (extreme) grootverdieners geen optie, want dat is ontkennen van de empirie. Een andere optie is het vergroten van het aantal peiljaren om zo de kans op toeval te verminderen, maar dat stuitte op allerlei praktische bezwaren. Het is ook mogelijk om de leeftijdsgroepen fors te vergroten, bijvoorbeeld naar 12, 16 of 20 jaar.

²³⁷ Deze figuur is losjes gebaseerd op de waarnemingen voor Latijns-Amerika en de regio Overig buiten Europa.

Dat zou echter onaanvaardbare vervorming van het nettobijdrageprofiel opleveren wat in de verdere analyse voor allerlei problemen zorgt.



Figuur 10.5 Schematische weergave van de verdeling over leeftijdscategorieën voor een fictieve populatie arbeidsmigranten (blauw) en de Top-100 met de hoogste nettobijdrage binnen deze fictieve populatie (oranje).

Daarom is de oplossing gezocht in de aanname²³⁸ dat de verdeling over leeftijdscategorieën van de Top 100 (zoals schematisch getoond in Figuur 10.5) de feitelijke kansverdeling voor leeftijd van de Top 100 is. Op basis van die aanname kan men extreem hoge nettobijdragen volgens die kansverdeling verdelen over de leeftijdsgroepen 36-39 t/m 68-71. Dit ‘smeert grote spikes’ uit over die leeftijdscategorieën.

Praktisch is gekozen voor de volgende oplossing. De oplossing wordt uitgelegd voor arbeidsmigranten uit Latijns-Amerika, maar voor arbeidsmigranten uit de regio Overig buiten Europa is hetzelfde gedaan. Voor elk individu bevat de dataset één record met alle variabelen. Voor de top 100 grootverdieners onder de arbeidsmigranten uit Latijns-Amerika is voor elke variabele die een bedrag als waarde heeft het gemiddelde bedrag berekend. Vervolgens zijn de bedragen in de originele records vervangen door die gemiddelden, maar de leeftijd werd niet veranderd. Op die manier blijft de totale nettobijdrage voor arbeidsmigranten uit Latijns-Amerika gelijk, maar verandert alleen de verdeling van die nettobijdrage over de verschillende leeftijdsgroepen. Deze berekening is uitgevoerd voor verschillende Top-X met X tussen 10 en 100. Bij Latijns-Amerika waren de veranderingen vanaf Top 20 beperkt en bij de regio Overig buiten Europa vanaf Top 50. Het corrigerend effect is vanaf Top 50 naar schatting € 63.000 lager voor arbeidsmigranten uit Latijns-Amerika en € 28.000 lager voor arbeidsmigranten uit de regio Overig buiten Europa.

²³⁸ Uiteraard is het mogelijk dat zal blijken dat de zeer hoge nettobijdragen bij herhaling voor een groot aantal peiljaren scheef verdeeld zijn in de richting van de hogere leeftijden (bijvoorbeeld als effect van carrière-opbouw) en in dat geval is deze aanname wellicht te optimistisch, hetgeen zou pleiten voor het nemen van een kleinere Top X om zo beter die eventueel scheve verdeling te benaderen of eventuele andere oplossingen.

10.5 Hoofdpunten sampling

Resumerend: bij de meeste indelingen is er meer dan voldoende data en komen kleine groepen weinig voor. Bij de 42-deling en bij combinaties van regio met migratiemotief speelt weinig of geen data bij jongere en/of oudere leeftijden een rol. Bij al deze groepen is het effect op de nettobijdrage over het leven meestal gering omdat eventuele fouten (i) relatief klein blijven vanwege een lage variantie, (ii) niet volledig meewegen in de einduitkomst vanwege het contant maken, sterfte en remigratie en/of (iii) slechts voor een klein deel meewegen in de einduitkomst omdat gewogen wordt tegen een kenmerk zoals entreeleeftijd dat samenhangt met de lage frequentie (anders gezegd: bij weging tegen entreeleeftijd wegen fouten die voortkomen uit lage aantallen voor een bepaalde entreeleeftijd door het wegen nauwelijks mee). Om het probleem van kleine aantallen op te lossen is data aangevuld uit bovenliggende regio's en zijn algoritmes ontwikkeld die het effect van grote 'spikes' dempen. Bij de kleinste groepen volgens de 42-deling (met herkomst Japan), bij enkele kleine groepen bij combinaties van motief met de 12-deling (met name arbeidsmigratie uit Indonesië en Suriname) en bij kleine groepen i.c.m. de aanwezigheid van personen met een heel grote positieve nettobijdrage (met name arbeidsmigratie uit Latijns-Amerika en de regio Overig buiten Europa) is de onzekerheid rondom de uitkomsten groter dan bij andere groepen. Hiermee moet bij de interpretatie rekening gehouden worden.

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 in Tabel 5.1 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 de mate waarin de toelating van immigranten selectief is ten aanzien van het opleidingsniveau. Voor groepen waarbij de 1^e generatie qua citoscore gemiddeld heel goed dan wel heel slecht scoort is in de 2^e en vaak zelfs in de 3^e generatie een matige voorsprong respectievelijk hardnekkige achterstand te bespeuren.

Ook sterk is dat de nettobijdrage van de tweede generatie niet zoals in de CPB-studie *Immigration and the Dutch Economy* (2003) gebaseerd is op de aanname dat de tweede generatie voor de helft geïntegreerd zou zijn, maar dat per groep op verschillende manieren (er is sprake van methoden- en datatringulatie) 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 onderwijsstelsel, maar vooral in gebrekkige absorptie op de arbeidsmarkt en gebrek aan selectie op opleidingsniveau bij toelating tot Nederland.

De maatschappelijke implicaties van de resultaten in het huidige 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 regeeringsbeleid wat betreft inkomsten en uitgaven, voortdurend in beweging is en de kosten- en batenposten mee veranderen. Ook zou zodoende meer inzicht worden vergaard in eventuele systematische en toevallige meetfouten.

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 in Tabel 5.1). Wel zouden een deel van deze posten naar rato van bijdrage aan het bbp berekend kunnen worden. Voor de hand liggende kandidaten zijn Defensie en Ontwikkelingshulp waarvan de hoogte van het macrobedrag een percentage van het bbp is (althans, dat is de systematiek, maar die wordt niet altijd gevolgd in de praktijk). Van andere posten zoals overheidsinvesteringen in gebouwen lijkt toerekening naar rato van bevolking zoals gedaan in het huidige rapport meer voor de hand te liggen. Posten als infrastructuur groeien waarschijnlijk eerder progressief in relatie tot de bevolkingsomvang als de bevolkingsdichtheid toeneemt door immigratie. Hetzelfde geldt vermoedelijk voor andere kosten die het gevolg zijn van extra bevolkingsgroei door immigratie, zoals allerlei (lastig te becijferen) externe kosten door zaken als congestie, extra milieuvervuiling, verlies van landschap, natuur en open ruimte of extra kosten voor het behalen van klimaatdoelen. Wellicht is de schatting van de nettobijdrage verder te verbeteren door de genoemde items te wegen naar de mate waarin ze gerelateerd zijn aan bevolkingsomvang dan wel bbp.

Daarnaast zijn er 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.), dat 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). Alles bij elkaar gaat het om een aanzienlijk deel van de totale overheidsuitgaven in 2016 van ruim 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 die in het huidige rapport zijn opgenomen 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 het huidige rapport 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 welvaartsmagneet' 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 opleidingsniveau en citoscore die kennelijk optreedt bij remigratie (§9.12) en daaraan mogelijk gerelateerd de aanzienlijke verschillen in citoscore die er bestaan tussen de migratiemotieven 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 loopt op met de verblijfsjaren. 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 nettobijdragen 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.

12 Model

We start out from observed immigrants' contributions paid to or benefits received from the government in 2016:

$$X_{gac}$$

With:

- c category of contributions or benefits
- a immigrant age, $0 \leq a \leq 99$
- g group defined by generation, source region, migration motive, education

We distinguish 23 categories of contributions/benefits, following the CPB 2014 study *Minder zorg om vergrijzing* on consequences of ageing. We take observations from Statistics Netherlands microdata, Statistics Netherlands StatLine and the CPB2018-dataset.

The net contribution profile \overline{P}_g for group g is a profile by age, a vector with elements P_{ga} given by:

$$P_{ga} = \sum_c X_{gac} \text{ for } 0 \leq a \leq 99$$

P_{ga} is the net contribution for group g and age a .

Future contributions/benefits for year $y \geq 2016$ are obtained from multiplying by W_{yac} taken from the CPB2018-dataset. In its ageing study, CPB has predicted levels of contributions and benefits up to 2060, based on anticipated economic developments and policy measures. We use these predictions to set values by category relative to their value in 2016. The CPB is an acknowledged authority in the Netherlands on economic policy analysis in political debates, providing a common reference in the debates. Of course, CPB predictions are not beyond doubt or discussion, but using them puts our calculations within the common frame of reference.

As contributions/benefits in category c for group g at age a in year y are given by $W_{yac} \cdot X_{gac}$, the net contribution of group g at age a in year y is given by:

$$NCA_{gay} = \sum_c W_{ayc} \cdot X_{gac}$$

The net contribution NCE_{ge} by age of entry e and group g is given by:

$$NCE_{ge} = \sum_{a=e}^{99} \left\{ \left(\frac{1+p}{1+i} \right)^{a-e} \cdot NCA_{ga(2016+a-e)} \cdot SURV_{ea} \cdot STAY_{gea} \right\}$$

With:

- p rate of productivity growth (percentage)
- i real rate of interest (percentage)
- $SURV_{ea}$ probability of survival up to age a when arriving at entry age e

- $STAY_{gea}$ probability of staying in the Netherlands up to age a when arriving at entry age e , for a member of group g

Note the implicit assumption that someone belonging to a certain group and arriving at age e will upon arrival be assigned the contributions/benefits of a comparable immigrant of age e , i.e., an arriving immigrant is assumed to have the age profile of the average residing immigrant of the same group and age. This eliminates cohort effects by year of arrival, which is unavoidable as we have no data by year of entry.

The net lifetime contribution LNC_g for group g is given by:

$$LNC_g = \sum_{e=0}^{99} \{NCE_{ge} \cdot I_{ge}\}$$

I_{ge} is the fraction of immigrants in group g immigrating at entry age e . The fractions are taken from immigrants arriving between 1-1-1995 en 31-12-2016.

The immigration profile \vec{I}_g is a profile by age for group g , a vector with elements I_{ge} , $0 \leq e \leq 99$.

The fraction $SURV_{ea}$ is given by:

$$SURV_{ea} = \prod_{m=e}^a (1 - S_{m(2016+m-e)})$$

S_{ay} is the probability of dying at age a in year y . The probabilities are taken from CBS predictions up to 2060 and kept constant thereafter.

The fraction $STAY_{gea}$ is given as:

$$STAY_{gea} = \prod_{m=e}^a (1 - R_{ge(m-e)})$$

R_{ged} is the probability of remigration for an immigrant in group g arriving at age e after a stay of duration d for $d + e \leq 99$. The probabilities for $0 \leq d \leq 23$ have been calculated from CBS-microdata on emigration of immigrants arriving in the Netherlands since 1-1-1995, as is described in §7.2 of this appendix.

For individuals with first generation immigration background, the net contribution is augmented by initial cost of immigration: integration courses, processing of immigration application, cost of housing and monetary allowances for refugees. Present values of entitlement to state pension (AOW) for immigrants who have left The Netherlands are also added, based on anticipated length of stay derived from remigration probabilities and weighted by the entry age distribution (pension entitlements build up by year of presence). For details see Chapter 6 of this appendix.

For individuals without immigration background and individuals with second generation immigration background who will not remigrate, the net lifetime contribution is determined as:

$$LNC_g = \sum_{a=0}^{99} \left\{ \left(\frac{1+p}{1+i} \right)^a \cdot NCA_{ga(2016+a)} \cdot SURV_{0a} \right\}$$

Thus, in this case the calculation for LNC_g equals the calculation for NCE_{ge} for entry age $e = 0$ and $STAY_{gea} = 1$ for all $0 \leq a \leq 99$.

For individuals with second generation immigration background, we assume that after age 18 they will never remigrate and before age 18 they will always remigrate with their migrating parents. In the latter case we assign them the remigration probabilities of their parents. The net lifetime contribution is then calculated as:

$$LNC_g = \sum_{a=0}^{99} \left\{ \left(\frac{1+p}{1+i} \right)^a \cdot NCA_{ga(2016+a)} \cdot SURV_{0a} \cdot STAY_{g0a} \right\}$$

Hence, in this case the calculation of LNC_g equals the calculation of NCE_{ge} for entry age $e = 0$ and and probabilities to stay $STAY_{gea}$ taken as their parents' values. Birth of (by definition of Statistics Netherlands) Dutch-born second generation immigrants is predicted from age-specific fertility rates of the mother (as described in §7.1 of this appendix.) combined with the immigration profiles (as described in §2.3 and §7.2 of this appendix.).

Thus, to summarise, our basic data are observations on contributions and benefits observed at individual levels in 2016. They are carried forward by remigration probabilities observed in 2015-2017, mortality rates predicted by CBS and macro-economic developments predicted by CPB. Differentiation by age of entry is obtained by assuming that an immigrant enters the lifetime age profile of contributions and benefits observed in 2016 for a residing immigrant at his age at entry.

Figuren

Figuur 2.1 Nettobijdrage per leeftijdsjaar van de gemiddelde Nederlander in 2016.	8
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.	9
Figuur 2.3 Fiscale kosten naar leeftijd (bron: eigen berekening o.b.v. CBS-microdata).	10
Figuur 2.4 Fiscale baten naar leeftijd. Bron: eigen berekening op basis van CBS-microdata. De overige overheidsinkomsten zijn gebaseerd op CPB-gegevens.	11
Figuur 2.5 Fiscale kosten en baten en de fiscale nettobijdrage naar leeftijd. Bron: eigen berekening op basis van CBS-microdata.	12
Figuur 2.6 Nettobijdrage over de levensloop naar entreeleeftijd, met en zonder remigratie. Bron: eigen berekening op basis van CBS-microdata.	13
Figuur 2.7 Percentage immigranten per entreeleeftijd (immigratieprofiel), eerste generatie immigranten, 2011-2017. Bron: eigen berekening op basis van CBS-microdata.	14
Figuur 2.8 Pearsons correlatiecoëfficiënten (en de kwadraten R^2 daarvan) voor de samenhang tussen de gemiddelde citoscore voor alle leeftijdsjaren samen (2006-2018) en de nettobijdrage per leeftijdsjaar afzonderlijk voor 46 tweede generatie groepen, zijnde de 42-delings 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. De fractie n/N staat voor het quotiënt van het aantal groepen n dat is meegenomen in de regressie en het totaal aantal groepen $N = 46$. Significante correlaties op significantieniveau 0.01 en 0.0001 zijn gemarkeerd.	19
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. De gestippelde lijn betekent dat er weinig data zijn en/of dat er sprake is van een cohorteffect, dit t.b.v. de bespreking van methode 1.	21
Figuur 2.10 Samenhang tussen vier methoden om de mate van integratie van de tweede generatie te schatten voor 46 tweede generatie groepen. De vierde methode (Meth4) is direct gebaseerd op de waargenomen nettobijdrage over de eerste 41 leeftijdsjaren en vormt de laagste schatting; de overige drie methoden geven derhalve een overschatting van de mate van integratie. De waarden op de horizontale as zijn de gemiddelden van methode 1 t/m 4. De dikke zwarte lijn is geeft de regressievergelijking van een schatting van de mate van integratie o.b.v. de directe waarnemingen van de nettobijdrage over de eerste 35 leeftijdsjaren.	26
Figuur 4.1 Schematische weergave van personen met een Nederlandse achtergrond.	36
Figuur 4.2 Verdeling over citoscore-groepen (citodistributie) naar migratieachtergrond.	43
Figuur 4.3 Regio-indeling volgens de 42-delings.	56
Figuur 5.1 Onderwijs: vergelijking met het CPB-profiel.	65
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. ..	67
Figuur 5.3 Zorg: vergelijking met het CPB-profiel.	70
Figuur 5.4 LIS: vergelijking met het CPB-profiel.	73
Figuur 5.5 Indirecte belastingen per inkomensdeciël en gesmoothed. Berekening op basis van CBS-statline.	74
Figuur 5.6 IRN: vergelijking met CPB-profiel.	75

Figuur 5.7 CPB-profiel voor posten IRN via bedrijven, VPB/dividendbelasting en erfbelasting met vergelijking en R^2 voor trendlijn.	77
Figuur 5.8 Operationalisering pensioenopbouw en pensioenafname Nederlandse bevolking naar leeftijd, synthetisch profiel.....	80
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	81
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-Vergrijzingsstudie-2019-Data-figuren.xlsx	97
Figuur 8.2 Verschillen tussen de CPB-vergrijzingsstudies Minder zorg om vergrijzing uit 2014 en Zorgen om morgen uit 2019. Facsimile van Tabel 2.2 uit Adema, Y., & I. van Tilburg (2019), Zorgen om morgen, opgehaald 17-1-2022 van: https://www.cpb.nl/sites/default/files/omnidownload/CPB-Vergrijzingsstudie-2019-Zorgen-om-morgen.pdf	98
Figuur 8.3 Vergelijking van het nettobijdrageprofiel van het huidige rapport met het nettobijdrageprofiel uit de CPB-vergrijzingsstudie Zorgen om morgen uit 2019.	100
Figuur 8.4 Vertaling van de resultaten voor niet-westerse eerste generatie immigranten uit Roodenburg et al. naar het huidige rapport.	101
Figuur 10.1 Standaarddeviatie van de nettobijdrage op basis van CBS-microdata per vijfjaarsleeftijdsgroep.....	106
Figuur 10.2 Standaardfout van de nettobijdrage voor de gehele bevolking en 5-jaars voortschrijdend gemiddelde.....	107
Figuur 10.3 Toename van de totale nettobijdrage over de levensloop (verticale as) als gevolg van een toename met € 1.000 van de nettobijdrage in een bepaald leeftijdsjaar (horizontale as), met het immigratie- en remigratiegedrag van de gemiddelde eerste generatie immigrant en sterfte en discontering zoals in het standaardscenario.....	108
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 en gewogen naar het effect (zie Figuur 10.3) van het immigratie- en emigratiegedrag van de gemiddelde eerste generatie immigrant, en sterfte en contant maken volgens het standaardscenario.	109
Figuur 10.5 Schematische weergave van de verdeling over leeftijdscategorieën voor een fictieve populatie arbeidsmigranten (blauw) en de Top-100 met de hoogste nettobijdrage binnen deze fictieve populatie (oranje).....	120

Tabellen

Tabel 2.1 Verdeling over grootteklassen per leeftijdsgroep voor 46 groepen met een 2 ^e generatie migratieachtergrond.	16
Tabel 2.2 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.	28
Tabel 2.3 Toerekening van de kosten voor publieke goederen aan personen in een gemengd scenario (miljarden euro's).	30
Tabel 3.1 Onderzoekspopulatie.....	34
Tabel 4.1 Samenstelling citocohorten naar leeftijd en jaar van afname.	40
Tabel 4.2 Schematisch overzicht van de gebruikte HBA-cohorten voor de bepaling van de nettobijdrage per citoscore. De percentages hebben betrekking op de beschikbaarheid van HBA-data.....	41
Tabel 4.3 Kansverdeling voor het opleidingsniveau van 2 ^e generatie kinderen per opleidingsniveau van 1 ^e generatie immigranten, voor niet-westerse migratieachtergrond.....	45
Tabel 4.4 Kansverdeling voor het opleidingsniveau van 2 ^e generatie kinderen per opleidingsniveau van 1 ^e generatie immigranten, voor westerse migratieachtergrond.	46
Tabel 4.5 Kansverdeling voor het opleidingsniveau van 2 ^e generatie kinderen per opleidingsniveau van 1 ^e generatie immigranten, voor de referentie-autochtoon.....	48
Tabel 4.6 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.	53
Tabel 4.7 Regio-indeling	55
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.	59
Tabel 5.2 Verdeling van het toekenningspercentage van het leerlinggewicht, 2016. Eigen berekening op basis van data van DUO.....	68
Tabel 6.1 Inburgeringskosten 2016-2019 per jaar per immigrant (in €)	86
Tabel 10.1 Aantal waarnemingen per leeftijdsgroep. De modale klasse is vetgedrukt.....	110
Tabel 10.2 Aantal waarnemingen voor 1 ^e generatie immigranten, voor combinaties van migratiemotief en herkomstregio volgens de 12-deling.	112

