

MIDAS: Basisgegevens, steekproef en verwerking

Annex 1 bij 'MIDAS 2.0: Revisie van een
dynamisch microsimulatiemodel'

Januari 2023

Gijs Dekkers, gd@plan.be,
Raphaël Desmet, rd@plan.be,
Karel Van den Bosch, kvdb@plan.be

Inhoudstafel

Inleiding	5
Steekproeftrekking en weging	5
Selectie van bruikbare observaties	6
Einheid, datum en periode van observatie	7
Vergelijking van inkomens tussen de MIDAS projectie voor 2012, de Datawarehouse gegevens en EU-SILC	9
Vergelijking van de pensioenen tussen de MIDAS projectie, en de Datawarehouse gegevens voor het jaar 2012	11

Lijst van tabellen

Tabel 1	Aantallen in de populatie, in de steekproef en wegingscoëfficiënten	6
Tabel 2	Sociaal-economische positie, naargelang criterium van weglaten	7
Tabel 3	Einheid en data van observatie	8
Tabel 4	Vergelijking bruto-inkomens tussen Datawarehouse gegevens, MIDAS projectie en EU-SILC voor het jaar 2012	10
Tabel 5	Vergelijking bruto-netto traject tussen MIDAS projectie en EU-SILC voor het jaar 2012	10
Tabel 6	Vergelijking bruto-pensioenen tussen Datawarehouse gegevens en MIDAS projectie voor het jaar 2012: detail	12

Federaal Planbureau

Het Federaal Planbureau (FPB) is een instelling van openbaar nut die beleidsrelevante studies en vooruitzichten maakt over economische, socio-economische en milieuvraagstukken. Daarnaast bestudeert het de integratie van die vraagstukken in een context van duurzame ontwikkeling. Het stelt zijn wetenschappelijke expertise onder meer ter beschikking van de regering, het Parlement, de sociale gesprekspartners, nationale en internationale instellingen.

De werkzaamheden van het FPB worden steeds gekenmerkt door een onafhankelijke benadering, transparantie en aandacht voor het algemeen welzijn. De kwaliteit van de gegevens, een wetenschappelijke methodologie en de empirische geldigheid van de analyses staan daarbij centraal. Tot slot zorgt het FPB voor een ruime verspreiding van de resultaten van zijn werkzaamheden en draagt zo bij tot het democratisch debat.

Het Federaal Planbureau is EMAS en Ecodynamische Onderneming (drie sterren) gecertificeerd voor zijn milieubeheer.

<https://www.plan.be>

e-mail: contact@plan.be

Overname wordt toegestaan, behalve voor handelsdoeleinden, mits bronvermelding.

Verantwoordelijke uitgever: Saskia Weemaes

Wettelijk Depot: D/2023/7433/3

Federaal Planbureau

Belliardstraat 14-18, 1040 Brussel

tel.: +32-2-5077311

e-mail: contact@plan.be

<https://www.plan.be>

MIDAS: basisgegevens, steekproef en verwerking

Annex 1 bij 'MIDAS 2.0: Revisie van een dynamisch microsimulatiemodel'

Januari 2023

Gijs Dekkers, gd@plan.be, Raphaël Desmet, rd@plan.be, Karel Van den Bosch, kvdb@plan.be

Abstract - In deze annex worden de basisgegevens van MIDAS meer in detail beschreven, evenals de steekproeftrekking en de verwerking.

Inleiding

In 2016 zijn de basisgegevens, die dateerden van 2001, vervangen door nieuwe gegevens voor het jaar 2011. Deze gegevens zijn voornamelijk afkomstig uit diverse instellingen van de sociale zekerheid via het Datawarehouse Arbeidsmarkt en Sociale Bescherming, beheerd door de Kruispuntbank van de Sociale Zekerheid, aangevuld met fiscale variabelen uit de IPCAL (database van het Ministerie van Financiën), en enkele gegevens (opleiding, woonstatuut) uit de administratieve Census van 2011. Ten behoeve van de simulatie van de pensioenen zijn ook retrospectieve gegevens inzake de arbeidsmarktloopbaan opgevraagd. De volledige steekproef omvat 601 683 personen, gestratificeerd naar regio. Hieronder gaan we in op de steekproeftrekking en de weging, de selectie van bruikbare cases, de data en perioden waarvoor gegevens beschikbaar zijn, een vergelijking van de inkomens tussen de gegevens afkomstig uit het Datawarehouse data, de MIDAS simulaties en EU-SILC.

Steekproeftrekking en weging

De populatie waaruit de steekproef is getrokken bestond uit alle personen die op 1 januari 2012 in België woonden. Dit omvat niet alleen Belgen en niet-Belgen, maar ook de personen in het wachtregister dat vooral asielzoekers registreert. De methode van steekproeftrekking werd gestuurd door twee overwegingen. Ten eerste is de huishoudcontext essentieel voor analyses van armoede en ongelijkheid, hetgeen betekent dat informatie nodig is voor alle personen in het huishouden waar een steekproefindividu deel van uitmaakt. Om deze reden is de steekproef getrokken op het niveau van huishoudens, en is informatie gevraagd voor alle personen die leefden in de steekproefhuishoudens.¹ Ten tweede moest de steekproef niet alleen representatief zijn voor de Belgische bevolking als geheel, maar ook betrouwbare resultaten kunnen opleveren voor iedere regio afzonderlijk. Daarom werd de bevolking gestratificeerd naar regio, en werd per regio hetzelfde aantal huishoudens getrokken. Om praktische redenen werd daarenboven binnen iedere regio gestratificeerd naargelang het ging om private of om collectieve huishoudens; de laatste categorie omvat onder meer verzorgingstehuizen voor ouderen, gevangenen en kloosters. Binnen de strata van collectieve huishoudens werd de steekproef op individueel niveau getrokken. Tabel 1 toont de totale bevolking per stratum, en de omvang van de steekproef. Een gelijk aantal private huishoudens in de steekproef per regio leidt tot een ongelijk aantal steekproefindividuen, omdat de omvang van de huishoudens niet dezelfde is in iedere regio.

De steekproeftrekking impliceert dat de ongewogen steekproef niet representatief is voor de Belgische bevolking als geheel; de regio Brussel is relatief oververtegenwoordigd, en Vlaanderen is relatief ondervertegenwoordigd. Om hiervoor te corrigeren werd per stratum een wegingscoëfficiënt berekend, die gelijk is aan de inverse van de verhouding tussen het aantal individuen in de steekproef en het

¹ Een alternatieve methode bestaat eruit om een steekproef van individuen te trekken, en vervolgens informatie te vragen over alle personen die deel uitmaakten van de huishoudens waarin de steekproefindividuen leefden. Deze procedure is echter vanuit statistisch oogpunt minder efficiënt; dat wil zeggen, bij dezelfde steekproefomvang (in termen van individuen waarover informatie wordt gevraagd) leidt deze procedure tot ruimere betrouwbaarheidsmarges.

aantal individuen in de populatie. Toepassing van deze wegingscoëfficiënten zorgt ervoor dat resultaten (gemiddelden, totalen, percentages) representatief zijn voor de Belgische bevolking.

Tabel 1 Aantallen in de populatie, in de steekproef en wegingscoëfficiënten (1)

Stratum		Aantal huishoudens in populatie	Aantal huishoudens in steekproef	Aantal individuen in populatie	Aantal individuen in steekproef	Wegingscoëfficiënt
Brussels Hoofdstedelijk Gewest	Private huishoudens	548 822	88 600	1 142 940	183 870	6,22
	Collectieve huishoudens	10 403		10 401	2 604	3,99
	Totaal	559 225		1 153 341	186 474	
Vlaams Gewest	Private huishoudens	2 678 095	88 600	6 285 887	207 790	30,25
	Collectieve huishoudens	79 965		79 957	2 600	30,75
	Totaal	2 758 060		6 365 844	210 390	
Waals Gewest	Private huishoudens	1 533 212	88 600	3 511 274	202 216	17,36
	Collectieve huishoudens	46 916		46 901	2 603	18,02
	Totaal	1 580 128		3 558 175	204 819	
België	Private huishoudens	4 760 129	265 800	10 940 101	593 876	
	Collectieve huishoudens	137 284		137 259	7 807	
	Totaal	4 897 413		11 077 360	601 683	

(1) In een zeer klein aantal huishoudens (16 huishoudens met daarin 58 individuen) hadden niet alle individuen dezelfde waarde op de variabele regio, of op de variabele privaat of collectief huishouden. In geval van inconsistentie naar regio (12 huishoudens), werd voor alle individuen de regio gelijkgesteld aan de regio van de referentiepersoon. In geval van inconsistentie naar privaat of collectief huishouden (4 huishoudens) werden alle individuen beschouwd als levend in een privaat huishouden, en werd de variabele lipro-positie in die zin aangepast. (Documentatie over de exacte aanpassingen is beschikbaar.)

Ter controle zijn een aantal tabellen aangemaakt met vergelijkingen tussen de gewogen steekproef en de populatie, naargelang stratum (regio en privaat of collectief huishouden), aantal leden van het huishouden, LIPRO-positie (positie in het huishouden) en leeftijdscategorie (niet getoond). In proportionele termen zijn de afwijkingen bijna altijd klein tot zeer klein.

Selectie van bruikbare observaties

Een deel van de betaalde arbeid door inwoners van België wordt niet geregistreerd in het Datawarehouse, onder meer door grensarbeid, personeel van internationale instellingen en huispersoneel. Daardoor observeren we in de MIDAS startdata aanzienlijk minder werkenden dan volgens zowel SILC als de MALTESE cijfers het geval is. Deze onderschatting creëert problemen voor de alignering (zie de tekst van het WP), en leidt uiteraard tot vertekeningen in de resultaten. Deze onderschatting is gecorrigeerd door bepaalde gevallen uit de steekproef weg te laten, en de overige te herwegen tot de populatie. Terwille van de coherentie van de data voor het berekenen van de armoederisicostatus worden alleen volledige huishoudens weggelaten.

Drie criteria werden gebruikt om de huishoudens te selecteren die weggelaten kunnen worden:

1. Alle personen 16+ in het huishouden zijn in de loop van 2011 geïmmigreerd. Deze personen waren slechts een deel van het jaar in België, waardoor hun jaarinkomen onderschat is.
2. Het huishouden ontvangt geen enkel inkomen, behalve kinderbijslag, dus ook geen residueel inkomen (leefloon, IVT). Hoewel situaties van 0 inkomen in feite kunnen voorkomen, zal het in de meeste gevallen om ontbrekende inkomensgegevens gaan. Bij ontbreken van arbeidsmarkt- en/of

inkomensinformatie in het startjaar is het niet mogelijk de relevante kenmerken van individuen in latere jaren te simuleren.

- De verhouding tussen enerzijds het totaal van de netto-inkomsten in het huishouden volgens de fiscale (IPCAL) gegevens, en anderzijds het totaal in het huishouden van de bruto belastbare inkomens volgens het Datawarehouse is groter dan 1,5. Normaal gesproken is deze verhouding uiteraard kleiner dan 1. Een hoge waarde wijst erop dat belangrijke inkomsten van het huishouden niet in het Datawarehouse zijn opgenomen, bijvoorbeeld het loon van een grensarbeider. De drempelwaarde van 1,5 is gekozen, omdat uit analyse bleek dat een ratio van 1,5 vrij extreem is. De netto-inkomsten in deze huishoudens zijn vaak aanzienlijk.

Tabel 2 toont de sociaal-economische situatie van de individuen, volgens de drie selectiecriteria. In totaal werden 47 691 individuen (8,0 % van de totale steekproef) weggelaten, de meerderheid volgens criterium 2. Van alle weggelaten individuen waren 47 % geklassificeerd als 'anders niet-actief', wat betekent dat er geen informatie beschikbaar was om ze in één van de andere categorieën in te delen²; 31 % zijn kinderen of jongeren die schoolgaan of studeren.

Tabel 2 Sociaal-economische positie (1), naargelang criterium van weglaten

	Weggelaten individuen			Behouden individuen
	Criterium 1	Criterium 2	Criterium 3	
1. Werknemer in private sector	986	0	1 339	126 138
2. Werknemer in publieke sector	95	0	283	25 522
3. Ambtenaar	2	0	189	27 546
4. Zelfstandige	382	1 802	1 482	28 665
5. Schoolgaand of student	3 495	7 866	3 707	143 744
6. Werkloos, tijdskrediet	82	0	945	34 865
7. Invalide werknemer	8	0	395	17 607
8. Invalide zelfstandige	0	0	4	159
9. CELS (brugpensioen)	0	0	293	4 536
10. Gepensioneerd	3	0	2 185	96 264
11. Anders niet-actief	3 348	16 054	3 016	48 676
Totaal	8 401	25 722	13 838	553 722

(1) Deze variabele speelt een centrale rol in MIDAS, omdat het gesimuleerde type inkomen dat het individu ontvangt van deze variabele afhangt.

Na weglating van de betreffende huishoudens werden de wegingscoëfficiënten van de overgebleven individuen aangepast, zodat het opgewogen totaal aantal personen naargelang leeftijd en geslacht ongewijzigd bleef.

Eenheid, datum en periode van observatie

Alle variabelen zijn gemeten op het niveau van het individu (Tabel 3). Voor sommige variabelen wordt nog een onderscheid gemaakt binnen één individu naargelang bijvoorbeeld de job (RSZ variabelen) of het pensioenrecht (RVP). Sommige variabelen in het bestand zijn gemeten op één of meer momenten in de tijd, andere hebben betrekking op een bepaalde periode of periodes. De meeste demografische variabelen zijn vastgelegd op het moment van steekproeftrekking (1/1/2012). Variabelen uit de Census 2011, namelijk onderwijsniveau en het woonstatuut (eigenaar of huurder), hebben betrekking op

² Deze categorie komt overeen met code 4 ('Andere') van de Datawarehouse variabele 'nomenclatuur van de sociaal-economische situatie'.

1/1/2011. De variabele ‘nomenclatuur van de socio-economische positie’ wordt door het Datawarehouse bepaald aan de hand van de situatie op de laatste dag van ieder kwartaal. De meeste inkomensvariabelen zijn gemeten per kwartaal of per jaar. Alle variabelen zijn beschikbaar voor de jaren 2011 en 2012, uitgezonderd die met bron RIZIV of Nationaal Intermutualistisch College. Een aantal gegevens die van belang zijn voor de pensioenopbouw of om transitie in arbeidsmarktstatus te schatten, zoals de ‘nomenclatuur van de socio-economische positie’, en de variabelen met als bron de RSZ, de RSZ-PPO, de RSVZ en de Sigedis databank, zijn voor een langere periode gevraagd.

De beschikbare gegevens bevatten veel minder informatie over de vroegere loopbaan van zelfstandigen en ambtenaren dan het geval is voor werknemers. Door deze beperking is de simulatie van pensioenen in de systemen voor zelfstandigen en ambtenaren veel minder nauwkeurig dan de berekening van de werknemerspensioenen. Omdat deze informatie recent via het Datawarehouse beschikbaar is gekomen, of beschikbaar zal worden, kan in de toekomst bij een nieuwe data-aanvraag deze beperking wellicht overwonnen of gemilderd worden.

Tabel 3 Eenheid en data van observatie

Variabele	Observatie-eenheid	Moment (M) of Duur (D)	Tijdseenheid of moment	Periode
Demografische variabelen (geboortejaar, geslacht, burgerlijke staat, positie in huishouden, migratiegeschiedenis)	Individu	M	01-01-12	
Onderwijsniveau, woonstatuut (eigenaar of huurder)	Individu	M	01-01-11	
Nomenclatuur van de socio-economische positie	Individu	M	Laatste dag van ieder kwartaal	2003-2012
<i>Lonen, inkomens, uitkeringen per instelling van SZ:</i>				
RIZIV	Uitkeringsepisode x individu	D	Per kwartaal	2012
PDOS	Individu	D	Per jaar	2011-2012
RVP	Pensioenrecht x individu	D	Per jaar	2011-2012
Sigedis	Career_code x Individu	D	Per jaar	1954-2012
RSVZ	Individu	D	Per jaar	1997-2012
RSZ	Job x individu	D	Per kwartaal	1997-2012
RSZ-PPO	Job x individu	D	Per kwartaal	1997-2012
RVA	Individu	D	Per kwartaal	1999-2009, 2011-2012
POD-MI	Individu	D	Per kwartaal	2011-2012
FBZ	Individu	D	Per jaar	2011-2012
FAO	Individu	D	Per jaar	2011-2012
FOD-SZ	Individu	D	Per jaar	2011-2012
NIC	Uitkeringsepisode x individu	D	Per kwartaal	2012
<i>Fiscale gegevens</i>				
IPCAL	Individu	D	Per jaar	2011-2012
<i>Inkomensvariabelen geconstrueerd door DWH:</i>				
KB-RKW, KB-RSVZ	Individu	D	Per jaar	2005-2012
FAO	Individu	D	Per jaar	2007-2012
Andere inkomens	Individu	D	Per jaar	2011-2012

Vergelijking van inkomens tussen de MIDAS projectie voor 2012, de Datawarehouse gegevens en EU-SILC

De inkomens zoals gesimuleerd in MIDAS wijken om verschillende redenen af van de geobserveerde inkomens in de Datawarehouse gegevens. Allereerst worden niet alle inkomens gesimuleerd: in de MIDAS-projectie wordt geen rekening gehouden met de inkomsten uit kapitaal en aanvullende pensioenen uit de tweede of derde pijler, periodieke overdrachten tussen huishoudens, studiebeurzen, het inkomen uit loondienst betaald in natura en, ten slotte, bepaalde specifieke uitkeringen in verband met invaliditeit, tijdskrediet en loopbaanonderbreking³. Ten tweede heeft in projectie iedere persoon per jaar slechts een inkomen uit ten hoogste één inkomensbron. Individuen met werk zijn dus het gehele jaar tewerkgesteld, werklozen zijn gedurende het hele jaar werkloos, enz., en de inkomens worden op basis van volledige jaren berekend. Deze assumptie is realistisch voor de meeste gepensioneerden, maar veel minder voor werkenden en werklozen. Om na te gaan hoe groot de afwijkingen zijn die hierdoor ontstaan, vergelijken we de MIDAS projectie voor het jaar 2012 (het eerste projectiejaar), de Datawarehouse gegevens voor hetzelfde jaar, en de EU-SILC golf 2013, met inkomens van 2012.

Tabel 4 toont de resultaten van deze vergelijking voor een aantal bruto-inkomenscomponenten. Voor iedere component en voor iedere bron wordt vermeld het percentage van alle personen dat het type inkomen ontvangt, het gemiddelde van het type inkomen voor degenen die het ontvangen, en het gemiddelde per hoofd van de bevolking. De laatste grootte kan beschouwd worden als een indicatie van de schatting van de totale massa van het type inkomen. De volgende punten zijn vermeldenswaard:

- Het percentage personen met een loon is lager in MIDAS dan in DWH en in SILC, maar het gemiddelde loon is hoger. De reden hiervoor is de assumptie dat werkenden een volledig jaar werken.
- Per saldo verschilt de totale loonmassa in de MIDAS projectie nauwelijks van het geobserveerde bedrag in DWH 2012, en is deze ongeveer 8 % lager dan in EU-SILC 2013.
- Het percentage zelfstandigen in de MIDAS projectie is enigszins te hoog vergeleken met DWH en EU-SILC; het gemiddeld inkomen is hoger dan in de beide andere bronnen.
- De MIDAS projectie voor de pensioenen ligt zeer dichtbij de EU-SILC resultaten, maar geeft zowel een lager percentage ontvangers als een kleiner gemiddeld bedrag vergeleken met de DWH observaties. Hierbij moet aangetekend worden dat de pensioenen in DWH meer omvatten dan alleen de Belgische wettelijke pensioenen; zie volgende sectie.
- De MIDAS projectie van de uitkeringen door de RVA geeft een veel te laag percentage ontvangers, vergeleken met DWH; ten opzichte van de EU-SILC is het verschil kleiner. Anderzijds is het gemiddelde uitkeringsbedrag te hoog. De reden is de assumptie dat iedere persoon slechts één inkomensbron heeft, en dit gedurende het gehele jaar. Het bedrag per hoofd van de bevolking komt wel overeen.

³ Voor meer details verwijzen we naar: Frère, J.-M., *‘De bevolking met een risico op armoede of sociale uitsluiting in België. Projectie tot 2030’*, WP 12-16, Federaal Planbureau, november 2016

- Een gelijkaardige conclusie geldt voor de ziekte- en invaliditeitsuitkeringen (ZIV), hoewel hier ook het door MIDAS geprojecteerde bedrag per hoofd van de bevolking te laag is. Naast bovengenoemde redenen speelt ook een rol dat niet alle typen van ziekte- en invaliditeitsuitkering worden gesimuleerd.
- De categorie ‘overige uitkeringen’ omvat vooral middenengetoetste uitkeringen, zoals het leefloon, de inkomensvervangende tegemoetkoming (IVT) en het Zorgbudget voor ouderen met een zorgnood (de vroegere tegemoetkoming voor hulpbehoevende bejaarden, THAB). Van deze wordt alleen het leefloon gesimuleerd.
- Het totaal inkomen per hoofd van de bevolking in de MIDAS projectie ligt slechts 2,4 % onder de observatie in DWH 2012, en slechts 1,7 % onder het totaal van de inkomenscomponenten per hoofd in EU-SILC. Het totale inkomen in EU-SILC is 4,3 % hoger dan de som van de hier vermelde inkomenscomponenten.

Tabel 4 Vergelijking bruto-inkomens tussen Datawarehouse gegevens, MIDAS projectie en EU-SILC voor het jaar 2012

	DWH data 2012			Projectie MIDAS 2012**			SILC 2013 (inkomens 2012)		
	% > 0	Gem. >0	Gem. over allen	% > 0	Gem. >0	Gem. over allen	% > 0	Gem. >0	Gem. over allen
Lonen	42,4%	29 609	12 563	33,46%	36 551	12 230	39,4%	34 354	13 530
Inkomens zelfstandigen	5,7%	25 659	1 457	5,24%	30 546	1 602	4,8%	23 842	1 126
Pensioenen	19,5%	19 452	3 790	18,62%	18 194	3 389	18,0%	18 826	3 395
Uitkeringen RVA	17,0%	5 320	906	7,86%	9 765	768	9,1%	8 398	764
Uitkeringen ZIV	9,8%	6 434	628	4,60%	14 182	653	5,1%	9 900	501
Kinderbijslag			539	13,90%		518			552
Overige uitkeringen	4,9%	5 185	253	2,93%	4 916	144	2,6%	4 433	117
Totaal van de inkomenscomponenten			20 135			19 303			19 986

* Per hoofd: per hoofd van de totale bevolking. Dit is in feite een indicatie van de totale massa.

** Bedragen in prijzen van 2012 (zoals de andere bronnen)

*** Totaal inkomen: in DWH en MIDAS som van de inkomenscomponenten; in SILC variabele hy010 (totaal bruto inkomen huishouden)

% > 0: % van alle individuen die bedrag > 0 ontvangen; Gem.>0: gemiddelde van degenen die bedrag > 0 ontvangen

Tabel 5 Vergelijking bruto-netto traject tussen MIDAS projectie en EU-SILC voor het jaar 2012

	Gehele bevolking gemiddelden			Gehele bevolking medianen			Bevolking 65+ gemiddelden			Bevolking 65+ medianen	
	MIDAS 2012	SILC 2012	IPCAL 2012	MIDAS 2012	SILC 2012	IPCAL 2012	MIDAS 2012	SILC 2012	IPCAL 2012	MIDAS 2012	SILC 2012
Totaal bruto inkomen per hoofd	19 104	20 846		16 653	18 039		16 247	18 607		14 389	16 362
Totaal netto inkomen per hoofd	13 610	15 811		12 820	14 200		13 652	15 772		13 610	14 300
%netto / bruto	71,2%	75,8%		77,0%	78,7%		84,0%	84,8%		94,6%	87,4%
Sociale bijdragen per hoofd	1 956			332			329			0	
Belastingen per hoofd	3 913			3 957			2 265			1 125	
Belastingen + bijdragen	5 869	5 035		4 289	2 229		2 595	2 834		1 125	590
"Belasting staat" per hoofd			3 016			0			1 708		20

* Per hoofd: per hoofd van de totale bevolking. Dit is in feite een indicatie van de totale massa.

Tabel 5 geeft een vergelijking van het bruto-netto traject in MIDAS, vergeleken met EU-SILC 2012 (in de DWH data zijn alleen bruto en bruto belastbare inkomens beschikbaar. Bruto belastbaar wil zeggen na afhouding van sociale bijdragen). Tevens wordt de ‘belasting staat’ uit de IPCAL belastinggegevens vermeld. Dit bedrag is niet gelijk aan de in totaal betaalde belasting – niet alle belastingkortingen zijn afgetrokken en de gemeentelijke opcentiemen zijn er niet bij opgeteld – en geeft dus slechts een indicatie van het werkelijke verschil tussen bruto- en netto-inkomen. Het netto inkomen in de MIDAS projectie ligt verder beneden het netto-inkomen in EU-SILC dan het geval is voor het bruto-inkomen. Ook de andere cijfers wijzen op een overschatting van de gesimuleerde belastingen in MIDAS, wellicht omdat in de simulatie niet met alle aftrekposten rekening kan worden gehouden bij gebrek aan gegevens. Het verschil is kleiner voor de mediane bedragen dan voor de gemiddelden, hetgeen er op wijst dat dit probleem waarschijnlijk vooral gesitueerd is bij de hogere inkomens.

Samenvattend: de MIDAS projectie van inkomens heeft een aantal beperkingen, die vooral voor de bevolking beneden 65 jaar gelden: sommige bronnen van inkomen worden niet gesimuleerd, in de projectie heeft ieder individu één type inkomen per kalenderjaar, waardoor sommige type inkomens worden onderschat, en mogelijk wordt de belasting overschat, zodat het netto-inkomen te laag uitkomt. Om deze reden worden alleen projecties gepubliceerd voor de gepensioneerden, en niet voor andere categorieën van de bevolking.

Vergelijking van de pensioenen tussen de MIDAS projectie, en de Datawarehouse gegevens voor het jaar 2012

De pensioenen waarvan sprake in de vorige sectie omvatten alle pensioenen waarover informatie bekend is bij de RVP.⁴ In MIDAS worden alleen de Belgische wettelijke rust- en overlevingspensioenen en de IGO gesimuleerd. Tabel 6 laat zien dat deze inkomenscomponenten 89 percent van alle pensioenen vertegenwoordigen. De rest bestaat grotendeels uit tweedepijlerpensioenen, en voor een kleiner deel uit pensioenen van buitenlandse oorsprong; de overige soorten pensioen waarover de RVP informatie heeft zijn verwaarloosbaar.

⁴ De resultaten voor pensioenen in de vorige sectie waren gebaseerd op de variabele *uitkering_rvp* uit het door het Datawarehouse geconstrueerde bestand “Inkomens”. De documentatie voor deze variabele op de website van het Datawarehouse vermeldt alle typen pensioen, behalve buitenlandse pensioenen. Een vergelijking van de betreffende variabele met de meer gedetailleerde variabelen vanuit de RVP toont echter aan dat ook buitenlandse pensioenen in *uitkering_rvp* zijn opgenomen.

Tabel 6 Vergelijking bruto-pensioenen tussen Datawarehouse gegevens en MIDAS projectie voor het jaar 2012: detail

Soort pensioen	DWH data 2012				Projectie MIDAS 2012**			
	% > 0	Gem. >0	Totaal (M***)	% van alle pensioenen	% > 0	Gem. >0	Totaal (M***)	% van DWH totaal
Rustpensioen	17,25%	15 729	30 022	71,57%	16,43%	15 621	28 400	94,60%
Overlevingspensioen	5,14%	12 111	6 890	16,42%	4,70%	12 648	6 585	95,57%
GIB/IGO	1,09%	3 787	457	1,09%	0,96%	4 128	439	96,07%
Tweedepijlerpensioen	5,66%	6 078	3 805	9,07%	-	-	-	-
Buitenlands rust- of overlevingspensioen	1,80%	3 527	703	1,68%	-	-	-	-
Tegemoetkoming mindervaliden	0,02%	3 863	8	0,02%	-	-	-	-
Afgeleid pensioen	0,06%	8 614	61	0,14%	-	-	-	-
Overige	0,01%	3 750	5	0,01%	-	-	-	-
Alle pensioenen	19,48%	19 452	41 948	100,00%	18,28%	17 512	35 425	84,45%

** Bedragen in prijzen van 2012 (zoals in het Datawarehouse)

*** In miljoenen Euro

% > 0: % van alle individuen die bedrag > 0 ontvangen; Gem.>0: gemiddelde van degenen die bedrag > 0 ontvangen

Uit Tabel 6 blijkt ook dat het aantal personen die een rust- of overlevingspensioen of de IGO ontvangen in MIDAS wat lager ligt dan volgens de Datawarehouse gegevens⁵. De assumptie van één type inkomen per persoon per jaar is waarschijnlijk hiervoor de reden. De gemiddelde bedragen verschillen weinig tussen de DWH-data en de simulatie van MIDAS voor 2012⁶. De onderschatting in MIDAS van het aantal ontvangers leidt tot een zekere onderschatting van de geaggregeerde totalen per soort pensioen. Uit een verdeling van de ontvangers naar leeftijd blijkt dat, zoals te verwachten was, het verschil in aantal ontvangers van een rustpensioen vooral gesitueerd is bij de 64-jarigen, en daarnaast ook bij de 59- en 60-jarigen. Het verschil betreft zeer waarschijnlijk personen die tijdens het jaar 2012 met pensioen zijn gegaan, en dus niet het volledige jaar een pensioen hebben ontvangen. Voor de IGO vinden we vooral verschillen in ontvangst bij de 64- en 65-jarigen, om analoge redenen. Het verschil in ontvangst voor de overlevingspensioenen is meer verspreid over de leeftijdsgroepen.

⁵ De percentages in de kolom “% > 0” tellen niet op dat percentage in de rij voor “Alle pensioenen”, omdat er uiteraard veel overlapping is in de ontvangen soorten pensioen: veel personen met een overlevingspensioen hebben ook een rustpensioen en bijna alle ontvangers van de IGO combineren dit met een ander soort pensioen.

⁶ Hierbij moet worden opgemerkt dat voor alle personen die ook al in 2011 gepensioneerd waren, de geobserveerde pensioenbedragen worden overgenomen, met uprating volgens de lange-termijn assumpties van de Studiecommissie voor de Vergrijzing.

MIDAS: Immigration and emigration

Annex 2 bij 'MIDAS 2.0: Revisie van een dynamisch microsimulatiemodel'

January 2023

Gijs Dekkers, gd@plan.be,
Raphaël Desmet, rd@plan.be,
Karel Van den Bosch, kvdb@plan.be

Table of contents

- 1. Introduction 1**
- 2. Simulating immigration: the methodology..... 3**
 - 2.1. The definition of immigrant 3
 - 2.2. Simulating immigration: general approach 4
 - 2.3. Selecting households and individuals for migration 5
- 3. Input data 7**
 - 3.1. Migration characteristics in the starting dataset 7
 - 3.2. The migration projections 14
- 4. Results..... 17**
 - 4.1. Indicators about the procedure to simulate migration 17
 - 4.2. Population and migrant status 19
 - 4.3. Socio-economic position, earnings and pensions by immigrant status 24
- 5. Conclusions 28**
- References 30**

List of tables

Table 1	Distribution of clones over original donors	18
---------	---	----

List of graphs

Graph 1	Country of origin (own or parents') by immigration status in the MIDAS data of 2011	8
Graph 2	Nationality and length of stay in Belgium of by country of birth of foreign-born.....	9
Graph 3	Prevalence of immigration status by region in the MIDAS data of 2011	9
Graph 4	Age distribution by immigration status in the MIDAS data of 2011	10
Graph 5	Prevalence of educational attainment level by immigration state in the MIDAS data of 2011 (only persons 20+ not in full-time education).....	11
Graph 6	Socio-economic status of migrant status groups by age and gender, MIDAS starting data of 2011.....	12
Graph 7	Distribution of gross hourly wage rate by immigration status; men and women	13
Graph 8	Distribution of yearly hours of work by immigration status; men and women	14
Graph 9	Immigration and emigration by nationality in Belgium, 2012-2070	15
Graph 10	Net immigration numbers of individuals by age and nationality in 2030	15
Graph 11	Years of residence in Belgium of immigrants (foreign-born), by whether they emigrate or not	18
Graph 12	Population Pyramid for Belgium, various years; simulations without migration	20
Graph 13	Population Pyramid for Belgium, various years; simulations with migration	20
Graph 14	Demographic dependency ratio, with and without migration	21
Graph 15	Proportion in four migrant statuses, 2012-2070, with and without migration.....	22
Graph 16	Proportion in four migrant statuses, 2012-2070, with and without migration, by age group ..	24
Graph 17	Evolution of employment and "other non-active" status by immigrant status, population aged 15-66	25
Graph 18	Pensioners and immigrant status	26
Graph 19	Average earnings of people at work, by immigrant status	26
Graph 20	Gross pension benefit of pensioners, by immigrant status and gender.....	27

Federal Planning Bureau

The Federal Planning Bureau (FPB) is a public agency that carries out, in support of political decision-making, forecasts and studies on economic, social-economic and environmental policy issues and examines their integration into a context of sustainable development. It shares its expertise with the government, parliament, social partners, national and international institutions.

The FPB adopts an approach characterised by independence, transparency and the pursuit of the general interest. It uses high-quality data, scientific methods and empirical validation of analyses. The FPB publishes the results of its studies and, in this way, contributes to the democratic debate.

The Federal Planning Bureau is EMAS-certified and was awarded the Ecodynamic enterprise label (three stars) for its environmental policy.

<https://www.plan.be>

e-mail: contact@plan.be

With acknowledgement of the source, reproduction of all or part of the publication is authorized, except for commercial purposes.

Responsible publisher: Saskia Weemaes

Legal deposit: D/2023/7433/3

Federal Planning Bureau

Rue Belliard – Belliardstraat 14-18, 1040 Brussels

phone: +32-2-5077311

e-mail: contact@plan.be

<https://www.plan.be>

MIDAS: Immigration and emigration

Annex 2 bij ‘MIDAS 2.0: Revisie van een dynamisch microsimulatiemodel’

January 2023

Gijs Dekkers, gd@plan.be, Raphaël Desmet, rd@plan.be, Karel Van den Bosch, kvdb@plan.be

Abstract - This annex presents the modules on immigration and emigration in MIDAS, with more technical detail than is possible in the Working Paper.

1. Introduction

The dynamic microsimulation model MIDAS aims to simulate the prospective adequacy of social security provisions, specifically pensions. The current version of the model does not include immigration or emigration (though immigrants in the starting data were included in the simulation sample). This had two consequences. First, it meant that the simulation results lose representativeness over time. More specifically, due to the lack of the inflow of relatively young immigrants, the simulation sample showed ‘too much’ demographic ageing. Secondly, the simulation results missed the impact of immigration on the adequacy of social security and pensions.

This document describes the way that immigration and emigration is now modelled in MIDAS and presents and discusses the simulation results in comparison with the current version of the model without migration. MIDAS is a dynamic microsimulation model designed to simulate the long-term prospective impact of demographic ageing and social policies on income inequality and poverty. It incorporates demographic and macroeconomic projections from the meso-economic MALTESE model of the Federal Planning Bureau. This alignment of MIDAS on results of MALTESE allows projecting the adequacy indicators of pensions in accordance with the assessment of budgetary sustainability of the pension system. This work draws heavily on previous research in the MIMBEL project (Dekkers *et al.*, 2015; Dekkers, 2014c), financed by –and in collaboration with – the Belgian Federal Science Policy Office and the Federal Public Service Social Security.

It is important to take account of immigration and emigration when projecting poverty and inequality among pensioners. Immigration can soften the impacts of demographic ageing (Gil Alonso, 2009, Vasquez Grenno, 2009, Casarico and Devillanova, 2003; Dustmann *et al.*, 2010; Chojnicki *et al.*, 2005; Munz and Werding, 2005; Barrett and Bergin, 2009), although some authors consider the impact to be limited (Serrano *et al.*, 2011, Chusseau *et al.*, 2007). The vulnerable position of immigrants (Borjas, 1994, 1995; Aleksynska and Tritah, 2011) may however result in increasing risks of poverty, a lower work intensity and increased income inequality (Lelkes en Zólyomi, 2011, Casarico en Devillanova, 2003; Moore en Pacey, 2002; Kemnitz, 2003; Chassamboulli and Palivos, 2010), which in time could increase the pressure on the minimum provisions in the first pillar of the pension system in Belgium. In comparative perspective within the OECD, Belgium unfortunately does not do very well in this respect. Even though, most immigrants enter the country during their active phase of life, the employment rate of immigrants is low (Corluy and Verbist, 2010) and also their income position is lower than that of non-immigrants. Over time this results in weak labour market careers, which could translate into a lower pension benefit after retirement.

MIDAS might be used to assess the possible impact of immigration on the prospective adequacy of first-pillar pension benefits. However, as will become clear, the modelling in MIDAS of migration in all its ramifications is unfinished. We now have introduced modules that simulate exit and entry of foreigners (people born outside Belgium) and of Belgian-born persons. The numbers of immigrants and emigrants are aligned on the population projections regarding migration, that are yearly produced by the FPB. Furthermore, we distinguish 2nd generation immigrants, one or both parents of whom is an immigrant, from other Belgian-born persons. However, the simulation of migration is incomplete in that immigrant

status is in general not included in the behavioral equations regarding demographic (including household formation) and labour-market states and transitions. This implies that many results presented below should not be taken as projected consequences of immigration, but as implications of the way the introduction of migration works out in the current, MIDAS model. This document is therefore a kind of half-way report.

This document is structured as follows. In the next chapter we present the methodology used for simulating migration, starting with our definition of immigrant. In chapter 3 we discuss the input data of the model, which are of two kinds: first, administrative micro-data from a large sample drawn in 2011; second, the projections of immigration and emigration that are published each year by the FPB in the context of the population projections (Vandresse, 2015, 2019). In chapter 4 we present the results of the model including migration, using the version of MIDAS and the MALTESE projections for 2019. Chapter 5 concludes.

2. Simulating immigration: the methodology

2.1. The definition of immigrant

Before turning to the methodology, it is relevant to define what, in the context of this study, is meant by “immigrant”. There are differences in the literature as to how immigrants are defined. The OECD (2011, 384) makes a broad distinction between those countries (including European ones) where immigrants are “foreign residents” and others, mainly the settlement countries such as the US, Australia, Canada and New Zealand, where immigrants are the foreign-born. In the context of this study, we follow the latter definition. In the register data that is the starting dataset of 2011, key variables like nationality and country of birth are available. This allows us to define the stock of immigrants as those that have been born outside Belgium, irrespective of their nationality. This is in line with the definition used by OECD (2015, 40) and Peeters *et al.* (2017, 51). As regards socio-economic characteristics and consequences, being born outside Belgium is more relevant than nationality.

Defining immigrants using the “origin criterion” also has the advantage that the naturalisation process that changes the nationality of the immigrant does not affect his or her immigration status (OECD/European Union, 2015, 16). Nationality may be affected substantially by changes in the conditions for naturalisation (Peeters *et al.*, 2017, 23). Naturalisation does not change in a direct way the socio-economic state of the person (*idem*, 51), although it may have indirect effects. In addition to the foreign-born, the immigration status in MIDAS also takes into account whether or not at least one of the parents is an immigrant. So the immigration status variable of an individual can have the following entries:

1. Natives: born in Belgium
2. 2nd generation-once: born in Belgium, one parent foreign-born, one parent born in Belgium
3. 2nd generation-twice: born in Belgium, both parents foreign-born¹
4. Immigrant: not born in Belgium

The second-generation individuals are not immigrants in a strict sense but have what the OECD calls an “immigration background” (OECD/European Union, 2015, Table 31.1, 235). Peeters *et al.* (2017, 60) use the terminology “first-generation migrants”, for those that immigrated into Belgium, and natives and “later-generation migrations” for the others. The motivation for distinguishing second-generation migrants from natives is that research has shown that the labour market position of the latter group differs in important ways from that of natives (Peeters *et al.*, 2017).

¹ Yet, the immigration status of parents is derived from their first nationality (as we lack a variable for country of birth of parents), assuming that most foreign-born immigrants will not have Belgian nationality at entry. The starting data include a variable indicating the first nationality of parents of all sample persons, though this has the value “unknown” for 33% of the fathers of sample persons, and 30% of the mothers. The proportion of “unknown” increases strongly with age; we have assumed that these parents were in fact born in Belgium. This implies that the proportions of 2nd generation immigrants are underestimated, more so at higher ages.

This definition of immigrant creates some problems, however. The various immigration and emigration projections of the FPB (see below, section 3.2), that are the starting point for our model, distinguish between immigrating Belgians and other immigrants, and the flow of immigrants therefore includes persons with the Belgian nationality at entry, who presumably were born in Belgium, emigrated later and who now return. In section 3.2 we discuss the implications of this discrepancy for our projections of immigration. Secondly, there arises an unfortunate terminological ambiguity, for which there is no easy solution, and which we do not resolve. So, in the context of migration flows in a particular year, ‘immigration’ and ‘immigrants’ refer to all who take up residence in Belgium in that year, and were not living in Belgium in the previous year, irrespective of their place of birth. In the context of the composition of the population (the ‘stock’), the word ‘immigrants’ refers to people born outside Belgium. Where there is ambiguity, the terms are clarified by mentioning that the stock-immigrants are ‘foreign-born’, or that the flow-immigrants include returning Belgians.

2.2. Simulating immigration: general approach

The population projections of the Planning Bureau provide the number of people immigrating and emigrating each year. For MIDAS, the main problem is to impute the characteristics of immigrants. The only realistic way is to copy those from recent immigrants, i.e. to “clone” those immigrants. In fact, it is technically straightforward to simulate individuals to immigrate as singles: one just makes a clone of a selected earlier immigrant to represent the new immigrant. But this approach produces biased simulations of household size and other characteristics because many individuals – especially children – enter or leave the country not alone, but as part of a household. So this technique of individual migration is not satisfactory since most poverty and social measures are based on household characteristics and income. We therefore propose to follow the “donor approach” (Duleep and Dowha, 2008). This involves using the *pageant* algorithm (Chénard, 2000, O’Donoghue *et al.*, 2010) to sample from a dataset of individuals grouped in households, while aligning to aggregate FPB projections up to 2070 on the individual level. In the case of immigration, in each year of simulation the donor sample is the group of immigrants already in the simulation dataset. This cloning procedure of individuals subject to alignment has been developed in the framework of the model building toolbox LIAM II and a technical description can be found in Dekkers (2015). In the case of emigration, a household is selected from the sample available in the model at any future point in time. The *pageant* algorithm ensures that the selection of households for emigration or for cloning as immigrants is done in such a way that the members of the selected households correspond to the – externally given – number of immigrating and emigrating individuals by age and gender. A more detailed discussion is left for section 2.3 below.

The simulation in any year $t=2012..2070$ starts with the then available dataset, that is either the starting dataset of 2011 (if $t=2012$) or the results of the simulations between 2012 and $t-1$. In the first step, the selection of households eligible for emigration is done based on the required numbers of emigrant individuals by age and gender. This selection is done separately for Belgians and the foreign-born. If a household is selected for emigration, so are its members, which are then removed from the dataset. In the next step, households are selected as donors for immigration using the required numbers of immigrant individuals, by age and gender, and again separately for Belgians nationals and foreign-born persons. If a household is selected as a donor for an immigrating household, then its members are cloned to become the immigrants. The next section presents a more detailed discussion of the simulation and

selection of emigrants and immigrants. In the discussion of the immigration procedure, we use the term *donors* for households or persons who are already in the simulation sample and reserve the term *clones* for their copies who are added to the sample.

2.3. Selecting households and individuals for migration

The first step is the selection of households as possible *candidates* as for emigration. We distinguish between Belgian-born emigrants and foreign-born emigrants. The pool of potential foreign-born emigrants is limited to households made up of foreign immigrants, which do not include children born in Belgium or natives. This criterion is not very selective: in the starting data 60% of all immigrants live in households where all members are foreign-born; moreover, this pool of immigrant-only households is constantly replenished by new immigrants. Households with natives and/or 2nd generation immigrants, but no immigrants, constitute the pool for Belgian emigrants. These selection rules imply that households with both immigrants and non-immigrants are never selected for emigration, which is of course unrealistic. However, given the pageant algorithm used, it would be impossible or very complicated to align with the externally given emigration data in terms of individuals if those households could emigrate. Within each pool, households are ranked by size, with large households coming first. In the pool for foreign-born emigrants, some preference is also given to those who have immigrated more recently. In the pool for Belgian emigrants, 2nd generation immigrants-once count for two and 2nd generation immigrants-twice count for three in the calculation of household size; in this way preference is given to households with an immigrant background.

The pool for donors of foreign-born new immigrants is made up from all households which are composed of immigrants and/or 2nd generation immigrants, but no natives. This pool is defined rather widely (not just including immigrants, but also people, mainly children, born in Belgium), in order to have a sufficient number of candidates for immigration. Households with only natives constitute the pool of donors for Belgian immigrants. These selection rules imply that households with both natives and (2nd generation) immigrants are never selected as donors, which is a drawback of this procedure. However, given the pageant algorithm used, it would be impossible or very complicated to align with the externally given immigration data in terms of individuals if clones of those households could immigrate. Within each pool, households are ranked by size, with large households coming first. From each of these pools, households are selected for emigration using the *pageant* algorithm, in such a way that the target numbers of emigrating individuals by age and gender are achieved as closely as possible.

The second step uses the *pageant* algorithm (Chénard, 2000) to select donors from the candidate households, in such a way that the total number of immigrants by gender, age and whether they are Belgian-born or foreign-born matches the numbers specified by the population projections. If the numbers do not match exactly, undershooting errors are taken to the next period. At the end of this step, the households that need to be cloned to represent foreign and Belgian immigrating households are known. Note that the largest households are taken first. As a result, immigrating households will on average consist of more individuals than households residing in Belgium (given background variables age and gender). This also will limit the risk of important undershootings when a large household is selected while most of the target numbers of emigrating individuals by age and gender have been reached. However, this

will push up the average household size of immigrating households relative to households residing in Belgium.

The third step is the actual cloning, which copies the donor (original) households and the individuals within them into the cloned (offspring) households and individuals. While the values of most variables are retained, some need to be adapted. A new identification number is assigned to each new household and each cloned individual. The immigration status of the new individuals is set to (foreign-born) “immigrant” if the donor household was from the pool of immigrant households, and to “native” if it was from the pool of Belgian households, and the year of immigration is set equal to the simulation period. The nationality and country of birth of the new individuals are copied from the original individuals. The relationships between the individuals (partners, parents - children) within each cloned household are supposed to remain the same as they were in the donating household. The relevant identification numbers are adapted accordingly. Retaining the variable values of the donating individuals will maintain existing correlations of relevant variables (for example levels of education of partners, age of partners, etc.) in the immigrating sub-samples. However, some variables pertaining to (i) the working state and (ii) the first-pillar pension benefit at retirement must be adapted. As for the working state, we assume that foreign immigrants either enter into education or are out of the labour market with no social security benefit (i.e. “other non-active). We thus assume that they start looking for a job only after arrival (if they are of working age) and are not entitled to an unemployment benefit or a disability benefit. As for (ii), assuming that foreign immigrants have not previously been in Belgium, whereas Belgian (returning) immigrants have, a number of variables relevant for calculating the pension benefit at retirement and the eligibility conditions for retirement are set to zero for foreign immigrants only. So, the labour market history of the donating individual is removed for foreign immigrants but not for Belgian (returning) immigrants. This will of course overestimate the pension claim of Belgian immigrants in Belgium, but we hope that this covers the underestimation caused by the fact that we cannot simulate the pension claim that the re-immigrating Belgian individuals may have gathered while being abroad. Thus, we implicitly assume here that the re-immigrants have earned a pension claim during the years that they were abroad that is about equal to the Belgian pension claim that they could have gathered if they had stayed in Belgium.

3. Input data

The data that are used for the projection of migration in MIDAS are of two kinds: on the one hand the starting dataset consisting of administrative variables for a large sample of individuals, and on the other hand the demographic projections of migration. We discuss these below in turn.

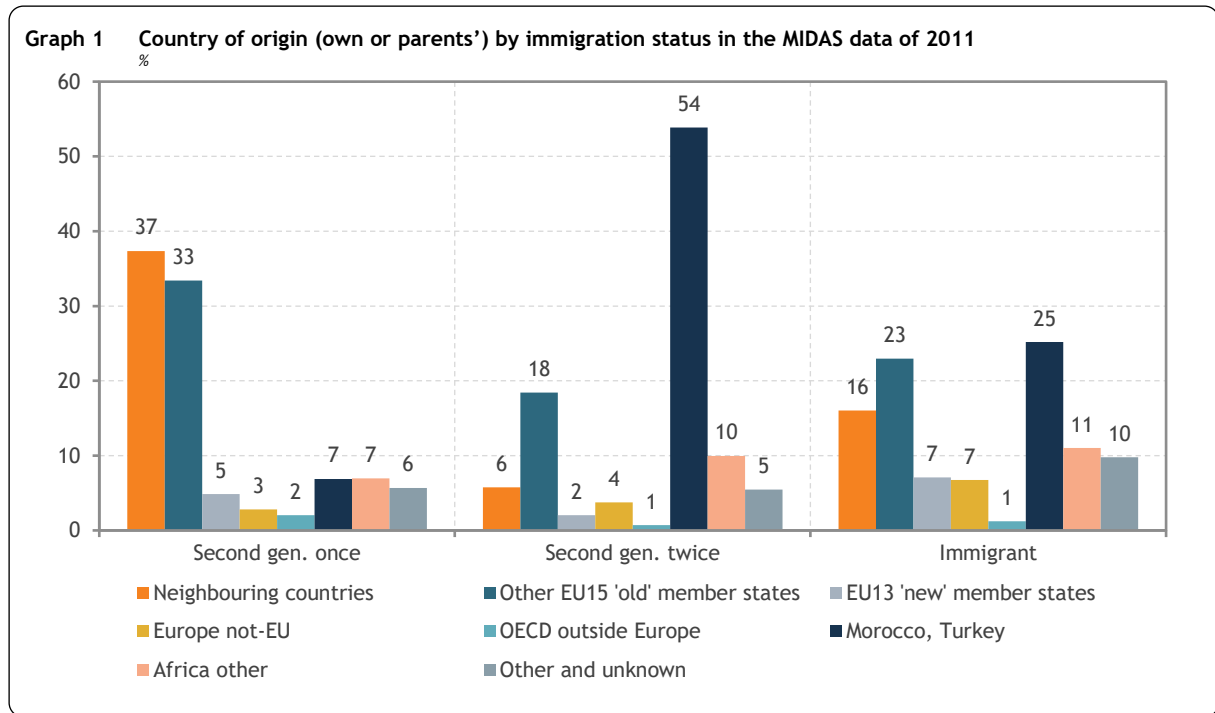
3.1. Migration characteristics in the starting dataset

The current starting dataset of MIDAS is a compound of register data from the Crossroads Bank Labour Market and Social Security, combined with Census data and register data from the IPCAL data base containing the data of the Ministry of Finances on the taxable income per taxpayer. For a description of the sampling procedure, the selection of the observations that were actually used, and comparisons of the sample with external data on key variables, we refer to a technical report (Dekkers et al., 2019). Note that the samples were of similar size in the three regions of Belgium (so the Brussels Capital Region was strongly oversampled); for results on the Belgian level weights are used.

In this section we focus on the characteristics of immigrants in the MIDAS data, compared with native Belgians. We use the four-category immigrant status variable introduced above. This variable distinguishes between immigrants – those that were born outside Belgium, irrespective of their (current) nationality – 2nd generation immigrants, and native Belgians. 2nd generation immigrants are all born in Belgium, but the “once” have one parent that is an immigrant, and one parent that was born in Belgium, while the “twice” 2nd generation immigrants are those born in Belgium whose parents are both immigrants. Thus, taking up the Belgian nationality through naturalisation does not change one’s immigration status.

Graph 1 shows the country of origin of immigrants. For 2nd generation immigrants this is the country of first nationality of the parent who immigrated; in case both parents immigrated and for both the first nationality is known, that of the mother was used preferentially. For 2nd generation migrants-once, the immigrated parent came mostly from the neighbouring countries (DE, FR, LU, NL) or from other EU15 countries.² The parents of 2nd generation migrants-twice came predominantly from Morocco or Turkey. Unsurprisingly, the countries mentioned are also the most important among immigrants (those born outside Belgium), with smaller groups coming from other European countries and Africa.

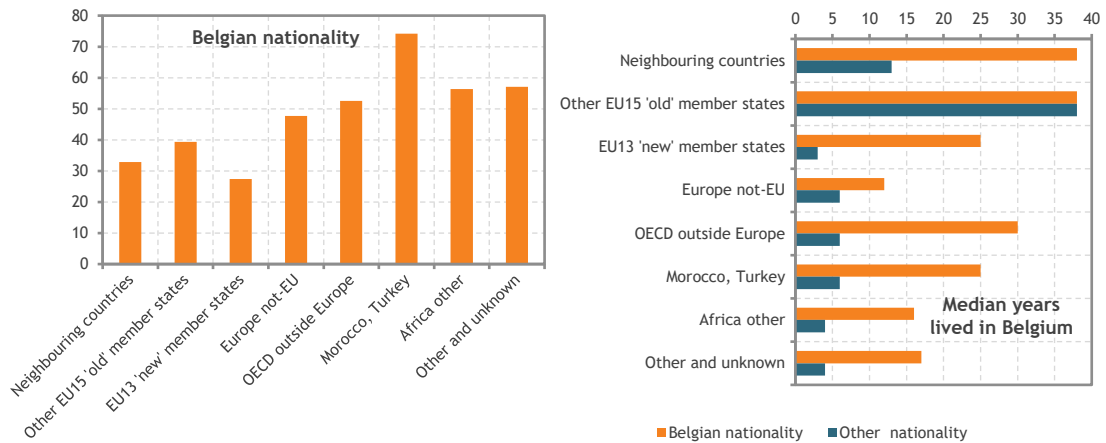
² The EU15 countries (excluding Belgium) Austria, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, Netherlands, Portugal, Spain, Sweden, United Kingdom. These are the countries that were in the EU prior to the accession of ten candidate countries on 1 May 2004 (OECD, 2007). Since 2004 there have been 13 new countries added to the European Union - Bulgaria, Croatia, Cyprus, Czech Republic, Estonia, Hungary, Latvia, Lithuania, Malta, Poland, Romania, Slovakia and Slovenia. See FPS Employment, Labour and Social Dialogue *et al.* (2013, page 26) for a more detailed discussion.



Virtually all people born in Belgium, including 2nd generation migrants, have the Belgian nationality. About half of all foreign-born migrants have acquired the Belgian nationality. The proportion varies somewhat by country of birth, as Graph 2, left graph, shows. It is more than 70% among migrants originating from Morocco or Turkey, and less than 40% among people from the EU, for whom naturalization brings few advantages except some voting rights.

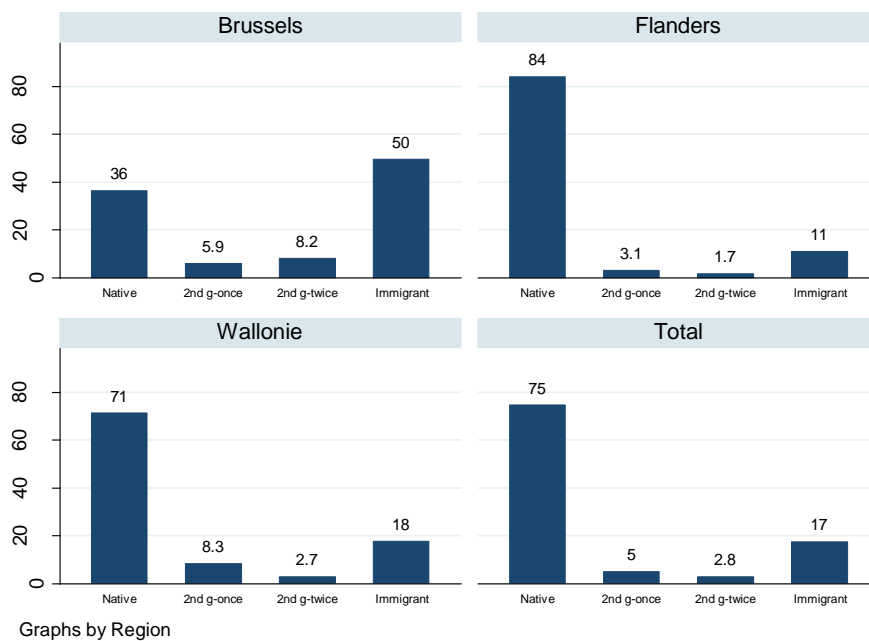
The right graph of Graph 2 shows that the number of years that the foreign-born have already lived in Belgium varies a lot by country of birth and by whether they have acquired the Belgian nationality. Except for immigrants from neighbouring countries and other EU15 countries, the median number of years lived in Belgium is always 15 or more years if the immigrants has acquired the Belgian nationality, and 6 or less if not. This is of course partly a selection effect: it generally takes a number of years to get naturalized. The high median years of residence in Belgium of foreign-born people with Belgian nationality suggests that these persons seldom re-emigrate.

Graph 2 Nationality and length of stay in Belgium of by country of birth of foreign-born
%, median number of years

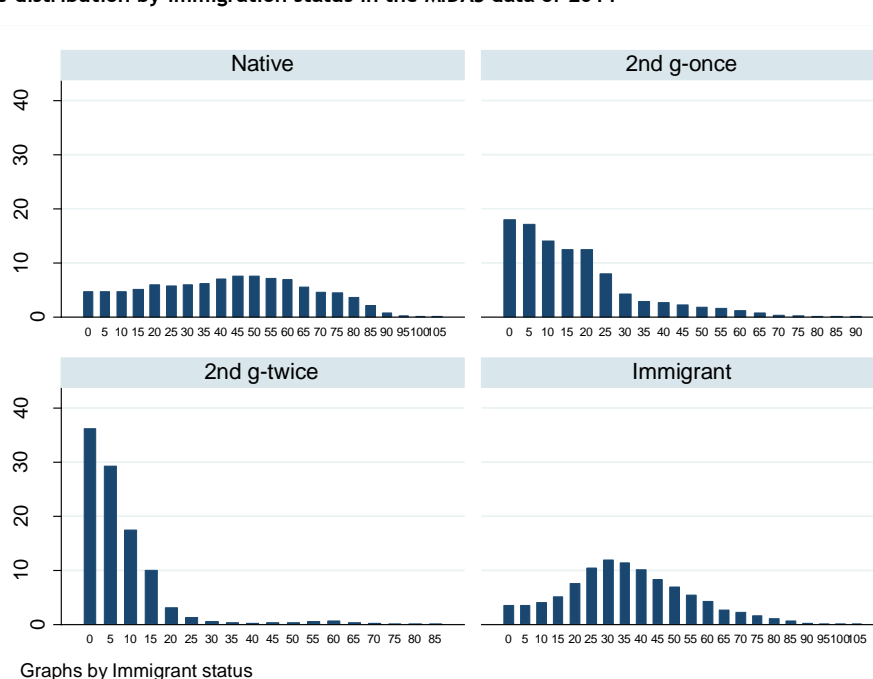


We now present a number of graphs to bring out differences between the four migrant status groups distinguished above. First, Graph 3 shows the prevalence of immigration status within the regions of Belgium, as well as in Belgium as a whole. In Flanders and certainly Wallonia, the native Belgians are very much in the majority. This is in contrast with the Brussels area, where half of the population consists of immigrants (either of Belgian or foreign nationality). This is probably also why the group of 2nd generation immigrants—twice (born in Belgium, with both parents not born in Belgium) is proportionally larger in Brussels than in Flanders and Wallonia.

Graph 3 Prevalence of immigration status by region in the MIDAS data of 2011
%



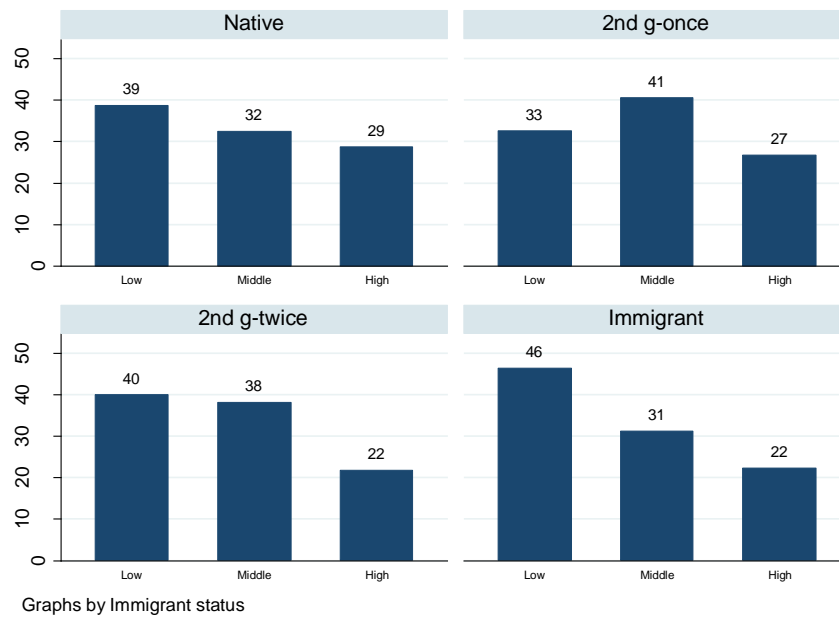
Graph 4 Age distribution by immigration status in the MIDAS data of 2011



Graph 4 shows the age distribution of the four immigration status categories. Clearly the group of Belgian natives in the North-West panel is the oldest (Peeters *et al.*, 2017, 65). Its modus is in the 50-year age group. Also, it is more equally spread over the age categories than the other groups, that have a more outspoken profile. The group of immigrants (South-East panel) are more often younger, between 20 and 45. But the two groups of 2nd generation immigrants have a different distribution altogether. They are much younger; most of them are in their teens and early twenties and, especially among the 2nd generation immigrants-twice, there are almost no individuals in the higher age groups. The preponderance of children and young adults in these groups is largely the result of the fact that immigration has increased strongly since 1990³; Belgian-born children of immigrants after that date could not be more than 21 in 2011. However, the virtual absence of older persons among the 2nd generation immigrants is at least partly a data-artefact: as mentioned above, for a large proportion of the sample, the first nationality of the parents is unknown (33% for the father and 30% for the mother); this proportion increases to more than half among the 50plus. When constructing the migration status variable, we have assumed for all these cases that the parent's first nationality was in fact Belgian. The proportion of 2nd generation migrants (of both kinds) is therefore underestimated.

³ See <https://statbel.fgov.be/nl/themas/bevolking/migraties#figures>

Graph 5 Prevalence of educational attainment level by immigration state in the MIDAS data of 2011 (only persons 20+ not in full-time education)

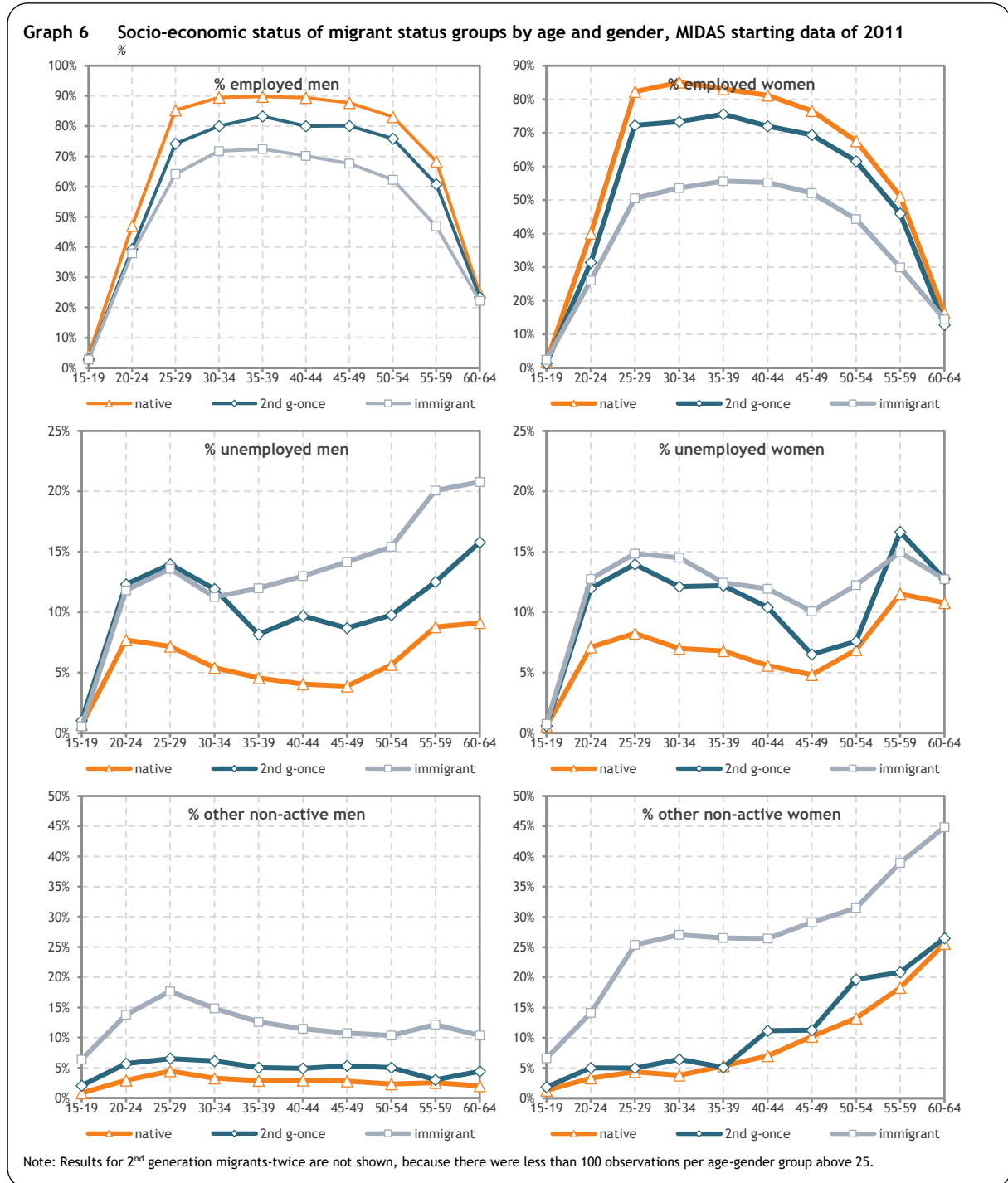


Graph 5 shows the educational attainment levels within the immigrant categories. The educational attainment is grouped into three levels. The lowest is ISCED lower or equal to 2, which is having at most a lower secondary education level. The highest is ISCED of 5 or more, which is having at least a tertiary educational attainment level, up to a doctorate. The group with the “middle” educational attainment level therefore includes those with ISCED equal to 3 (higher secondary education) and 4 (post-secondary non-tertiary education, which is rare in Belgium). As we do not model the educational career, Graph 5 shows results only for the individuals of 20 years and older and who are no longer in full-time schooling.

The results first show that the proportion of people with a low educational attainment level is considerably higher among immigrants than among native Belgians, while the reverse is true for the proportion with a high educational attainment level. The prevalence of educational attainment level of 2nd generation immigrants-twice is comparable to those of the immigrants themselves, though the group with middle educational attainment level is a bit higher at the expense of the group with the lowest level. This is even more clearly the case for the group of 2nd generation immigrants-once: the proportional size of the group with the middle level of education is by far the largest, but the size of the group with the highest educational attainment level is comparable to that of the native population. When considering these differences between the education levels of immigrant groups, one should of course take into account the age composition of these groups, as revealed in Graph 4. The youngest cohorts tend to have received more schooling than the older ones.

Graph 6 shows the socio-economic status by age and gender of the migrant status groups. (The 2nd generation migrants-twice have been left out, because this group is nearly non-existent among the 25plus.) Migrant men are much less often employed than native men (gap of 20 pp. around age 40), and the difference is even larger among women (more than 25 pp.). Children of couples of which one partner is

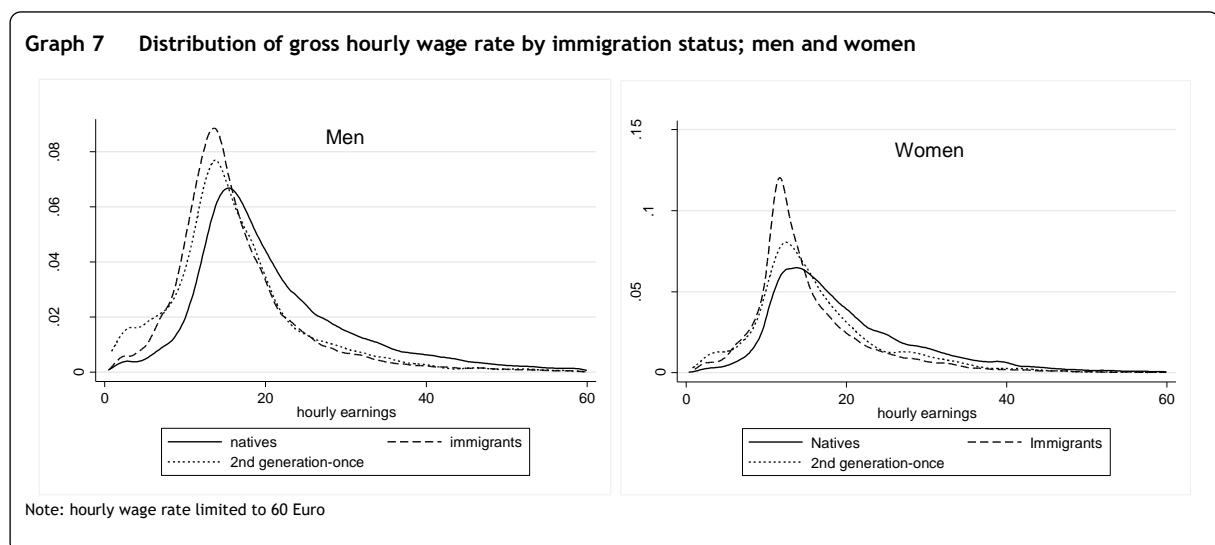
an immigrant (2nd generation-once) are in between, though for women their curve is closer to that of natives than that of migrants. Disaggregating employment by sector (not shown in graph) reveals that, unsurprisingly, hardly any migrant is a civil servant; otherwise there are no sharp differences between the migrant status groups.

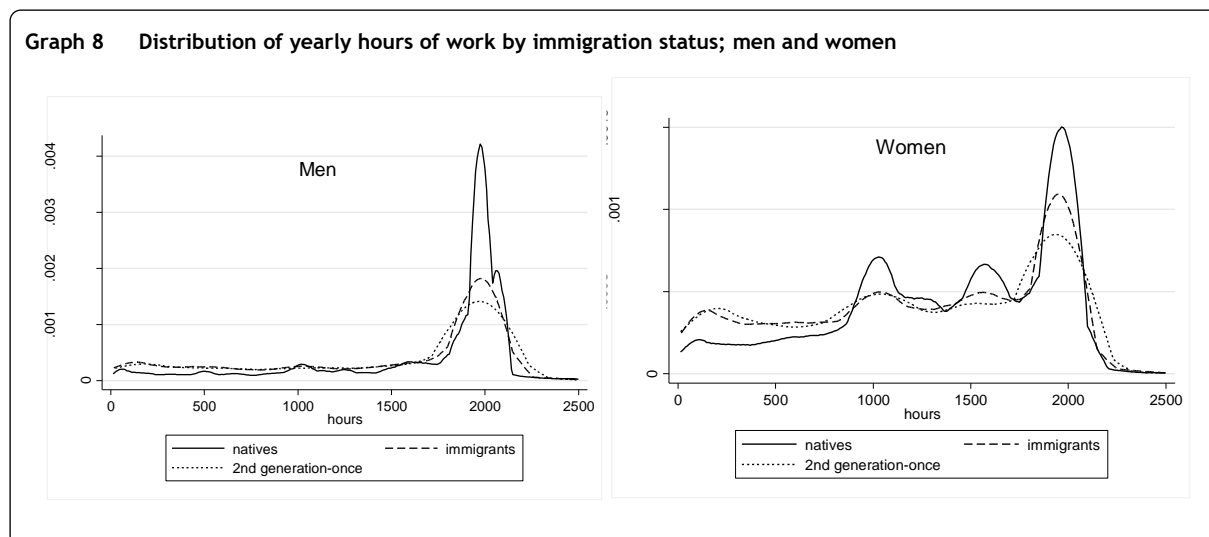


Unemployment is far higher for migrants of both genders, and also for 2nd generation migrants, than it is for natives. Up to the age of 34 for men, and to 44 for women, there is no or little difference in the unemployment rate between the foreign-born and the children of couples of which one partner is a migrant (2nd generation-once). Not shown in Graph 12 is that disability occurs more often among

migrant men over 45 than native men in the same age group; the difference increases to more than 5 pp. among those age 60-64. (Among women differences in disability rates between migrant groups are limited.) By contrast, many more natives in the age group 60-64 are retired (or in CELS) than migrants; the difference is 23 pp. for men and 20 pp. for women. Migrants are much more likely than natives to be in the state of “other non-active”, which implies that they have no work and no social security income (they might have a social assistance benefit); note that students are not included in this category. Among migrant men over 20 the percentage is in no age group below 10 percent, and always four times the rate among native men. The differences are even more glaring for women, amounting to around 20 pp. in all age groups of 25 and over.

Finally, we consider the earnings and hours worked by immigration status of people in the starting dataset. Graphs 7 and 8 make the distinction by gender and by immigrant status. To facilitate comparison the results are smoothed by expressing them as Kernel density estimates. Graph 15 shows that for men and to a lesser extent also for women the distribution of hourly wages of migrants, and also of 2nd generation migrants-once, is shifted to the left, compared to natives, and less positively skewed. (Results for 2nd generation migrants-twice are not shown, because of the low number of observations.) Among men, the median hourly wage of migrants, 2nd generation migrants-once and natives are respectively 14.8, 15.0 and 18.5 Euro, for women these medians are 13.3, 14.6 and 17.4 Euro (cf. Peeters *et al.*, 2017, 109). The curve for 2nd generation migrants is rather close to that of 1st generation migrants, except that it is a bit less concentrated at the median. Considering these curves, it must be taken into account that immigrants are on average younger than natives, and 2nd generation immigrants are much younger than both immigrants and natives. For the hours worked per year (among those that work; shown in Graph 16), differences are less outspoken, though the results in the right panel of this Graph suggests that foreign women, if they work, more often than native women work low or very low hours (Peeters *et al.*, 2017, 118). The differences reflect both part-time work, and periods out of work during the year; the finding that the curve of migrant women shows smaller spikes around the hours corresponding to half-time and 4/5 regimes suggests that the latter factor may be more important for them.





3.2. The migration projections

The immigration and emigration projections that underlie this document are described in Vandresse *et al.*, (2017; 2018; 2019) and a methodological paper (Vandresse, 2015). In the simulations to be discussed in the next chapter, the 2018-2070 projections (Vandresse *et al.*, 2019) are taken as the point of departure.

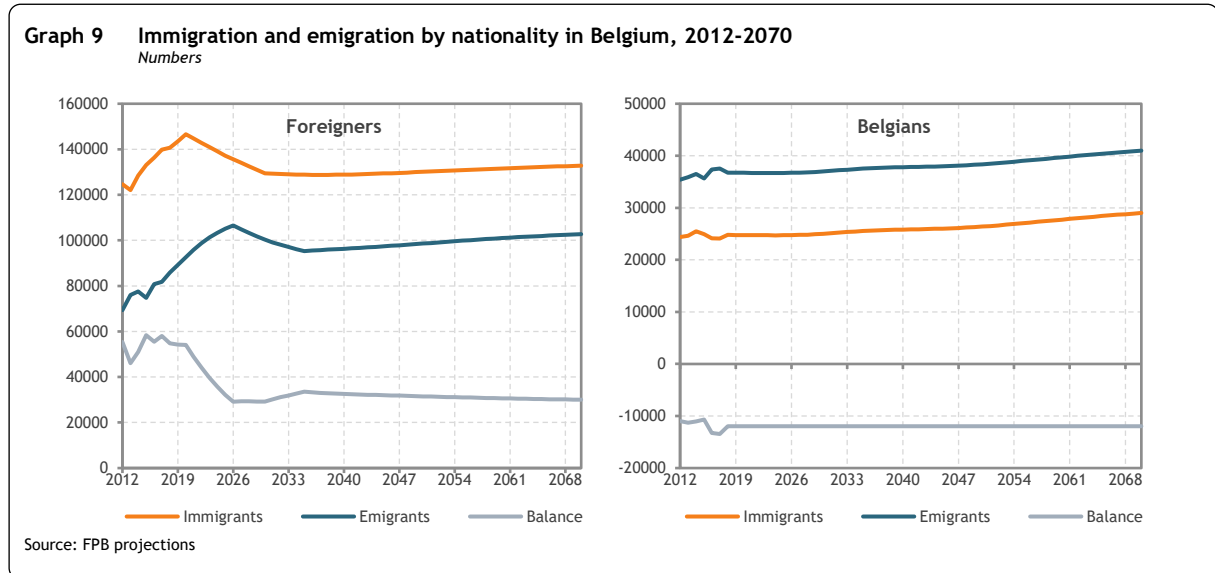
The projections assume that the international immigration of foreigners into Belgium would stabilise to the levels observed at the end of the 2000s. This is based on various hypotheses, among which

1. Immigration from non-EU countries would in the long run stabilise at a comparatively high level. The general driving factor is the population growth rates in these so-called “third countries”. Particularly important are i) the gradual inclusion of refugees that entered the country in the peak year 2015 and ii) the population growth in the countries where many refugees come from (Vandresse *et al.*, 2018, 11).
2. The immigration from “old” EU countries would remain relatively stable, but immigration from “new member states” would decline, as the economies in these countries would catch up to the rest of the EU (*idem*).
3. The composition of future immigrants by gender and age (and region) will be equal to the observed distributions over the period 2015 up to 2017.

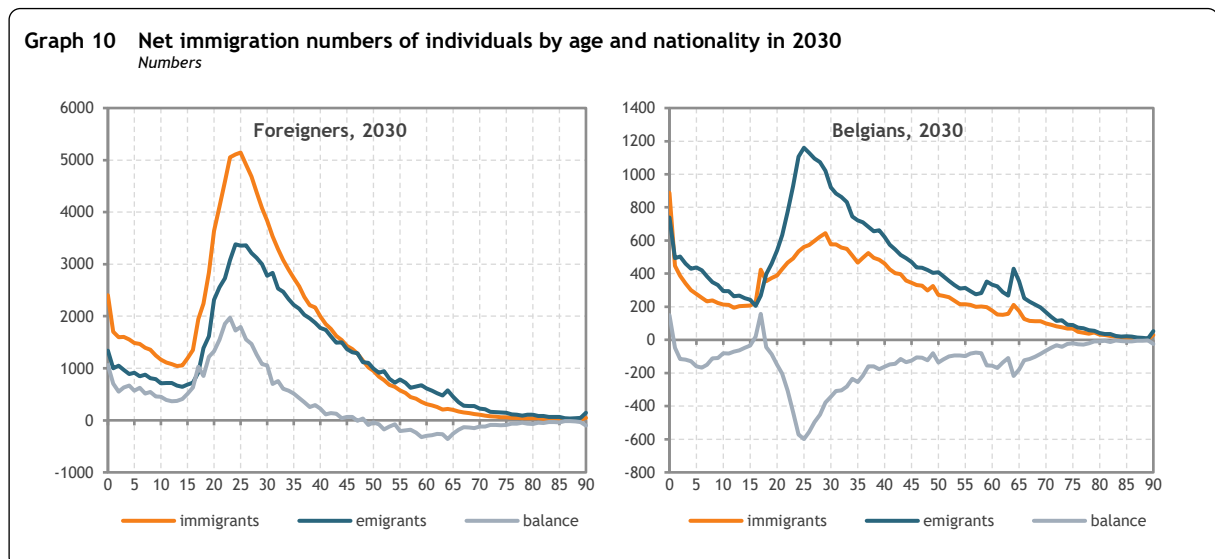
In these projections, individuals are distinguished by gender, nationality (Belgians, others) and age. Graph 9 shows the projections of the total numbers for immigration, emigration and their balance (net migration), for foreigners and Belgians. See Vandresse, 2015 and Vandresse *et al.*, 2018b, for an extensive discussion. More foreigners enter Belgium than leave Belgium. There is a slight overrepresentation of women in net immigration of foreigners (not shown in the graph); foreign men immigrate in greater numbers than women, but they re-emigrate also more often. The hypothesis is that net immigration of foreigners gradually decreases over the projection period and reaches a floor around 2040. This decreasing trend is the balance of a stabilizing number of immigrating foreigners, and an increasing number of emigrating foreigners, driven by the growing stock and constant observed emigration rates. For

Belgians, the development is different. Net immigration is not only considerably smaller, but, more importantly, it is negative, meaning that more Belgians leave the country than enter the country.

All in all, the baseline 2019 simulations assume, in the long run, a slow increase to about 160,000 immigrations per year in 2070, an equally slow increase of emigration to about 140,000 persons per year in 2070, and hence a (after 2037) nearly constant positive balance of about 20,000 persons per year (Vandresse *et al.*, 2019, 7).



Graph 10 shows the immigration, emigration and their balance, of foreigners and Belgians by age in 2030. Apart from changes in the numbers, as discussed above, the age distribution of projected emigration and immigration remains virtually unchanged through all projection years.



Both immigration and emigration of foreigners are for the largest part acts of people in their prime active age. Foreigners that enter the country, but also that leave Belgium again, are predominantly in their 20s or 30s, sometimes with their children. From about age 50, emigration of foreigners exceeds immigration, though the numbers for both are relatively limited. The age pattern of migration of

Belgians is to some extent the mirror image of net immigration of foreigners, with most emigration when people are in their 20s or early 30s; the largest number of Belgian immigrants occurs at slightly higher ages. We observe curious spikes of immigration at ages 0 and 17, and of emigration at 0 and 64.

The attentive reader may have noticed that the projections of aggregate immigration and emigration discussed in this section are split up in terms of nationality, while in the starting data we have defined immigrants and emigrants in terms of country of birth. Unfortunately, migration projections in terms of country of birth were not available to us, while, as explained above, for socio-economic outcomes that are important in the MIDAS projections, country of birth is much more relevant than nationality. However, we do not believe this apparent discrepancy is a very serious problem. Let us first consider immigration. As mentioned above, our starting data show that in the population there are very few foreign nationals who were born in Belgium, so we can assume that all immigrants of foreign nationality are also foreign-born. Immigrants with Belgian nationality who are foreign-born are possible, but would involve scenarios (such as children of Belgian ex-pats coming back, or immigration followed by naturalization followed by emigration followed by re-immigration) that cannot be very frequent. So we can safely assume that for immigrants, nationality (foreign or Belgian) and country of birth (foreign or Belgium) virtually coincide. The situation is a bit less straightforward for emigrants. It is likely that a disproportionately high number of emigrants with Belgian nationality were born in a foreign country. So we probably underestimate the emigration rate of the foreign-born to some extent. However, there are reasons or circumstances mitigating this. First, naturalization may be indicative of a commitment to Belgium, so among the foreign-born those with Belgian nationality are less likely to emigrate than those who retained their original nationality. Secondly, we do not simulate naturalization in MIDAS, so over the projection period, an increasing proportion of the foreign-born have foreign nationality.

4. Results

Below we present the results of the MIDAS projections including migration. These results are based on the version of MIDAS used for the Ageing Commission's Report of 2019; also all alignment tables are from that year. These projections, and the graphs based on them, were made before the 2020 demographic and economic projections were available (which are quite similar to the 2019 ones), and therefore also before the revised projections taking account of consequences of the Corona epidemic were produced. The latter are of course quite different from the 2019 projections, in particular for the medium-term. Since we are concerned here mainly with the long-term, a time-consuming update did not seem necessary. Also, as mentioned in the Introduction, these projections are intended to be illustrative of the implications of including migration in MIDAS; they are by no means final.

We first look at some indicators of the quality of procedure to simulate migration in Section 4.1. Section 4.2 is about the projections of the population, its age structure, and the evolution by immigrant status. In section 4.3, we consider education, socio-economic position, earnings and pensions. Results for the distribution of equivalent household income, inequality and poverty are presented in section 4.4. Where relevant, we compare the results of MIDAS simulations with migration to those without migration, using otherwise the same model and alignment tables.

4.1. Indicators about the procedure to simulate migration

In this section we look at the results from a methodological point of view, and present some indicators about the way the simulation procedure worked. As described above, immigration is simulated by *cloning* (copying) households and individuals in the simulated sample. However, over the simulation years not all households and individuals are cloned, given the selection and the algorithm described in chapter 2, some get cloned much more often than others. Table 1 provides indications of the distribution of clones of individuals by the original donor. This table was constructed in the following way. First, there are chains of clones, where individuals that were cloned in some year are the donor of another clone in later years, and so on. The latest clone is still indirectly a clone of the original donor, and these are assigned to the original donor in the table below (even if the original donor had disappeared, through death or emigration, from the simulation sample in the year the latest clone was created). Secondly, as in the migration procedure, we distinguish between foreign-born immigrants and Belgian-born ones. For the former, the donors are foreign-born immigrants and 2nd generation immigrants. For the latter only natives can be donors. Thirdly, for some clones, the original donor is not an individual from the starting data, but someone born in a later simulation year.

The last column of Table 1 shows that for 42% of all clones, the original donor was a (2nd generation) immigrant in the starting sample; while for 8% it was a 2nd generation immigrant born later. Note that these clones are necessarily foreign-born immigrants. By contrast, natives in the starting sample were the donor for only 9% of all clones, while natives born in later years were the donor for 41% of all clones. This imbalance between immigrants and natives comes about because for a Belgian-born immigrant aged below 10 in 2021 (the 10th simulation year), the donor cannot have been present in the starting data and so must have been born later. The original donor for a foreign-born immigrant of the same age in

the same year can be a 2nd generation immigrant born later, but is more likely to be an immigrant in the starting data (through a chain involving one or more intermediate donors). As the simulation proceeds through subsequent years, the age limit below which this mechanism operates shifts upwards, of course.

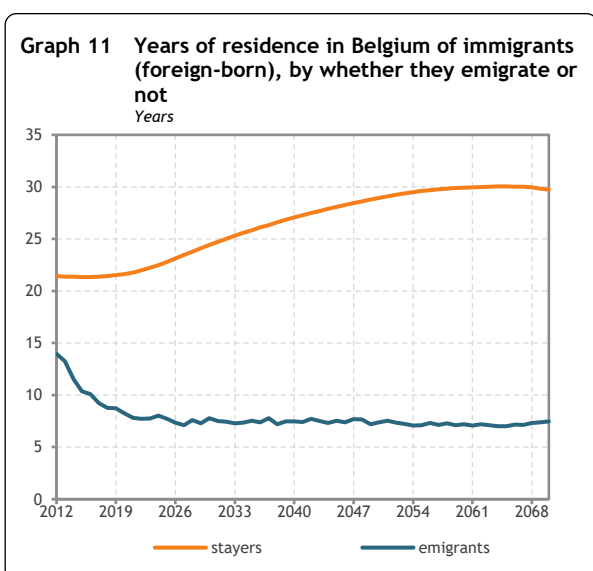
Columns 2-5 in Table 1 show the share of clones within each group by percentile of the original donors, when these were ranked by the number clones that they have, directly or indirectly, originated. So the third row, column two, shows that 0.1% of all individuals in the pool for foreign-born immigrants in the starting sample were the original donors for no less than 4.3% of all clones from this group. Similarly, 1% of all individuals in this group were the original donors for 17.9% of all clones in this group, and 25% of the individuals were the original donors for 100% of clones, implying that at least 75% of individuals in this group did not generate any clones. The distribution is even more skewed in the group of natives in the starting sample, and a bit less so in the donor groups born later. Overall it is clear that a relatively small number of individuals were the donors for most of the clones. We discuss the implications of this finding at the end of this section.

Table 1 Distribution of clones over original donors

Original donor group	Share by percentile				Share in total number of clones
	0.10%	1%	10%	25%	
Starting sample - immigrants	4.3%	17.9%	71.4%	100.0%	42.2%
Starting sample - natives	5.9%	39.1%	100.0%	100.0%	8.9%
Born later - immigrants	0.7%	7.8%	42.9%	100.0%	7.7%
Born later - natives	3.1%	10.9%	41.4%	65.7%	41.3%

Note: "immigrants" are persons included in the "pool" of donors for foreign-born immigrants, i.e. immigrants and 2nd generation immigrants. "natives" are people born in Belgium from Belgian-born parents.

Graph 11 shows the years of residence of the foreign-born, distinguishing between those who emigrate and those who stay, in a particular year. Even though this is not explicitly taken into account in the selection of emigrants, in any year the foreign-born who emigrate have resided in Belgium for a much shorter period of time than the foreign-born who stay in Belgium. The most plausible reason for this is



that we select for emigration only foreign-born people that live in households with no natives or 2nd generation immigrants; and these are mostly those who have immigrated fairly recently, and therefore had less opportunity to mix with Belgian-born people. Over time, the difference becomes even bigger, probably because in the simulation the foreign-born immigrants mix more quickly with Belgian-born persons than they did in reality before 2012. In any case, given the similarity of the age profiles of immigrating and emigrating foreign-born people, it seems plausible that those who immigrated recently are those most likely to re-emigrate, while the longer they reside in Belgium, the

more likely they are to stay for another year. Over time, the average number of years that all foreign-born immigrants have resided in Belgium increases. This reflects the large numbers of foreign-born that have immigrated into Belgium in recent decades, and will accumulate years of residence in future, and the projection of less immigration in the coming decades.

Some of the results reported in this section are a bit worrying: a relatively small number of individuals were the donors for most of the clones; and this small number might be a selective sample of all possible donors for immigration clones. However, limited adaptations in the selection procedure (i.e. random selection of households within pools, instead of ranking by household size) meant that the alignment targets for immigration and emigration by age and gender would not be achieved. As a correct distribution of the simulation sample by age and gender is the backbone of long-term dynamic microsimulation of populations, achieving those targets is an indispensable requirement. It is not clear whether the procedure can be much improved without leaving the *pageant* algorithm, which has become the accepted method for simulation of immigration in the field of dynamic microsimulation.

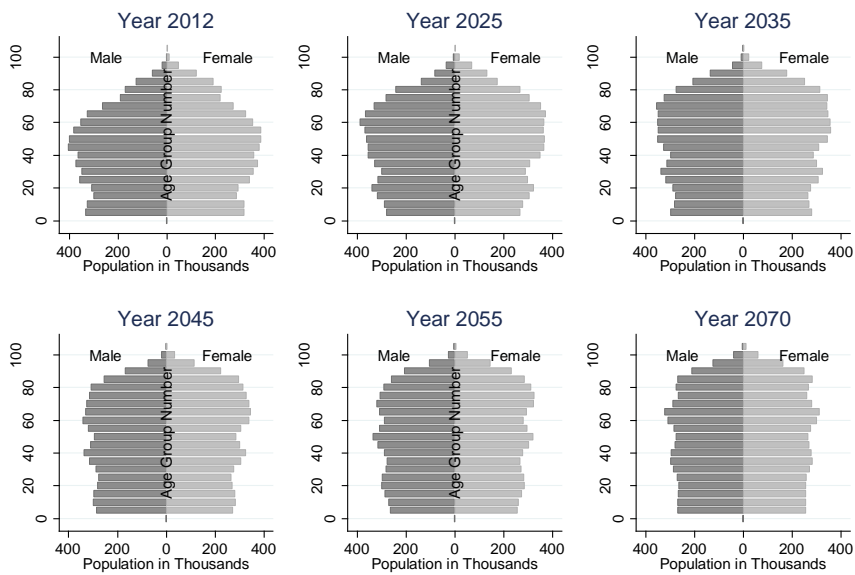
At the same time, the importance of possible selectivity should not be overstated. Apart from age and gender, the main characteristics that are copied from donors to immigrating clones refer to the educational level and to the household structure (partners, parents and children). For foreign-born immigrants, all variables about labour market characteristics, incomes and pension build-up are set at default values, independent of the donors.

4.2. Population and migrant status

The simulation results produced by MIDAS are based on the sample of register data, which was expanded (weighted up) using frequency weights (see Dekkers and Cumpston, 2011; 2012 for a discussion). Note that MIDAS models the main determinants of population size – birth, death, immigration and emigration – but the results are not calibrated to or corrected by the population projection itself. Without immigration and emigration, total population size in MIDAS would be a severe underestimate, relative to the projected population size according to the population projections. When including immigration and emigration, the fit of the simulated population with the demographic projections is nearly perfect.

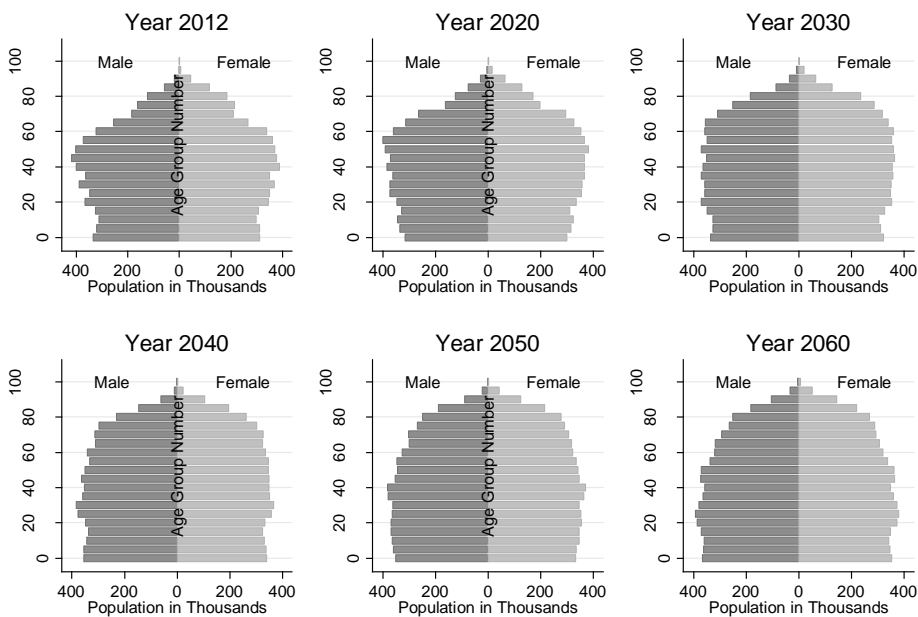
Next we compare how the population by age and gender changes when migration is being taken into account. This is important as migration changes the ageing pattern in Belgium. We do this by comparing the development of the well-known age pyramid for various years, both with and without migration. These are presented in Graph 12 (without migration) and Graph 13 (with migration).

Graph 12 Population Pyramid for Belgium, various years; simulations without migration
 '000s



Source: MIDAS without Migration

Graph 13 Population Pyramid for Belgium, various years; simulations with migration
 '000s

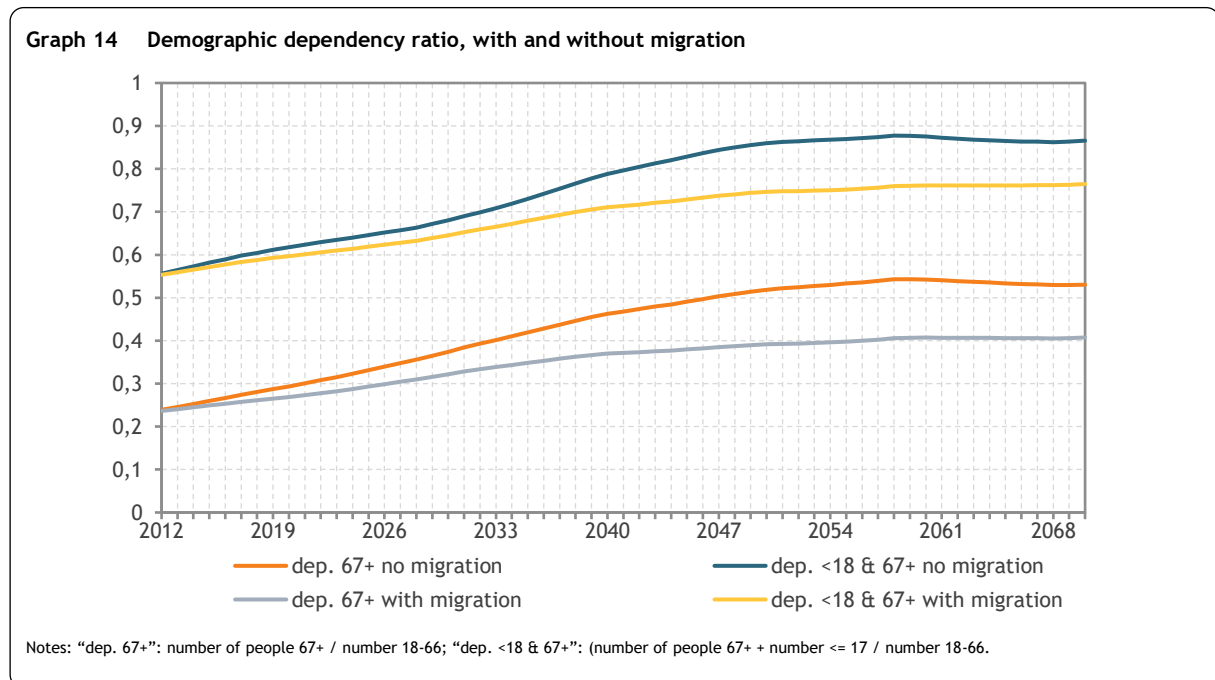


Source: MIDAS with Migration

For obvious reasons, the pyramids are almost the same in the first simulation year 2012. But in later years, the age pyramids without migration are “mushrooming”, as the large baby boom generation gradually moves upwards as time goes by, and is replaced by smaller cohorts, due to the below-replacement fertility rate. By contrast, including the net immigration in Belgium clearly broadens the bottom half of the pyramids in Graph 13 compared to Graph 12, as net immigration adds mostly to the younger

active age groups; see Graph 10 above. The impacts of the low fertility rate and the large size of the baby-boom generation are therefore reduced.

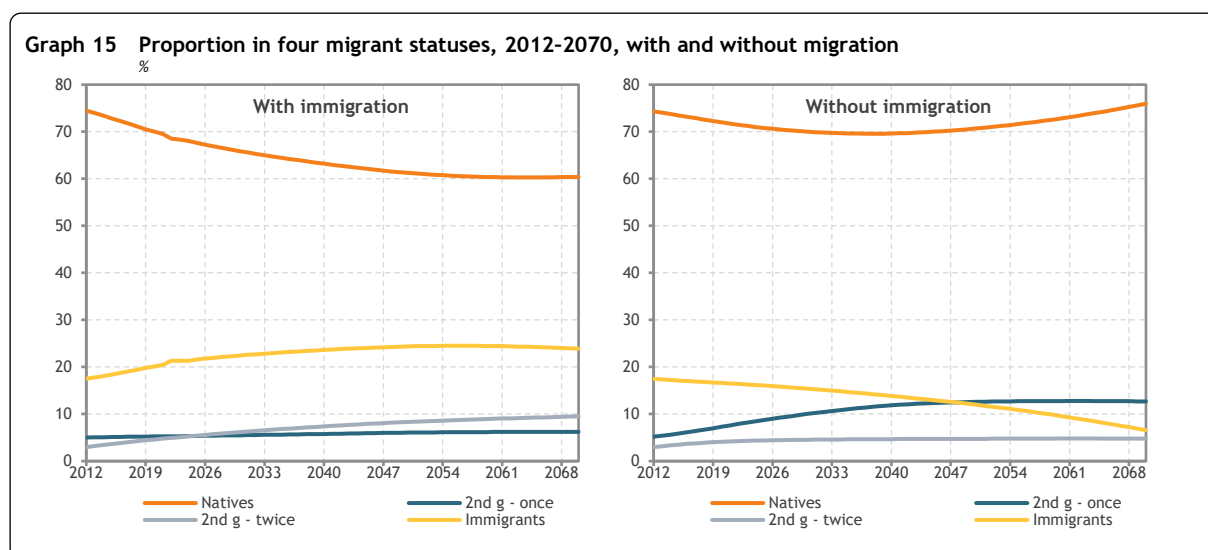
Another way of summarising the impact of migration is on demographic dependency ratios. These are shown in Graph 14. The first ratio, labelled “dep. 67+” represents the number of people of 67 and older divided by the number at working ages, 18-66 (taking into account the increases in the statutory retirement age in 2025 and 2030) (Vandresse et al., 2019). The second ratio, labelled “dep. <18 & 67+” represents the number of people of 67 and older plus the number of children below 18, divided by the number at working ages, 18-66.



This first ratio starts off at about 0.24, meaning that there is almost one older individual for four individuals at active age. In the version of the model without migration, it continuously increases to 0.53 in 2070, more than double the level in 2012. Taking account of net migration, this increase would be reduced, and this demographic dependency ratio would end up at 0.41 in 2070, i.e. about 23% less than in the scenario without migration. The second demographic dependency indicator “dep. <18 & 67+” includes the number of children, and its level is therefore much higher than that of the first indicator. On the other hand, its increase is less dramatic: 72% without migration and 38% with migration. Again, taking account of migration limits the increase of demographic dependency. However, it must be kept in mind that in the simulation of fertility no distinction is made between migrants and natives; the same fertility rates by age are applied to all kinds of women. If in fact migrant women have higher fertility, the actual impact of migration on demographic dependency might be different from what is projected here, although the direction of the difference is ambiguous (more children would mean a higher “dep. <18 & 67+” dependency ratio at first, but after those children become 18, the effect might be the reverse). These results are in agreement with many studies that show that net migration can soften the impacts of demographic ageing (Gil Alonso, 2009; Vasquez Grenno, 2009; Casarico and Devillanova, 2003; Dustmann *et al.*, 2010; Chojnicki *et al.*, 2005).

We now turn to the composition of the population in terms of immigration status. The proportion of immigrants would increase from almost 18% in 2012 to almost 32% in the mid-2030's. After that date, immigration of foreigners is balanced by emigration and death. Especially the proportion of 2nd generation-once immigrants, i.e. those that have been born in Belgium but who have one parent who is an immigrant, increases considerably, from 5% to about 14% in 2070. This is at least partly the result of the increasing stock of immigrants, who are in the reproductive years of their life, but also because at this stage, the matching routine for marriage or cohabiting does not consider immigration status. Hence the model at this point ignores any homogamy among immigrants and natives, which probably results in the proportion of 2nd generations-once immigrants being too high in the simulation. For the same reason, and also because the fertility probabilities do not take into account the immigration status of the mother, it is possible that the increase of the proportion of "2nd generation immigrants-twice" – those born in Belgium but with both parents being immigrants – is somewhat understated. On the other hand, as indicated above, the proportion of 2nd generation immigrants in the starting data is underestimated, and this bias disappears gradually in the projection. The size of the remaining category, native Belgians whose parents are both native Belgians too, would decrease from about 75% in 2012 to about 63% in the early 2050's.

Without migration the proportion of immigrants would obviously decline, as the right panel of Graph 15 shows. Perhaps more surprisingly, the proportions of 2nd generation immigrants would still increase, and in 2070 would end up at levels not much less than in the simulation with migration. As we saw above, current 2nd generation immigrants are quite young, and even in 2070 death has not much impact on this group. Also, even without immigration after 2012, these groups are replenished by the children of the many immigrant women who are in or will get into their fertile ages during the projection period.



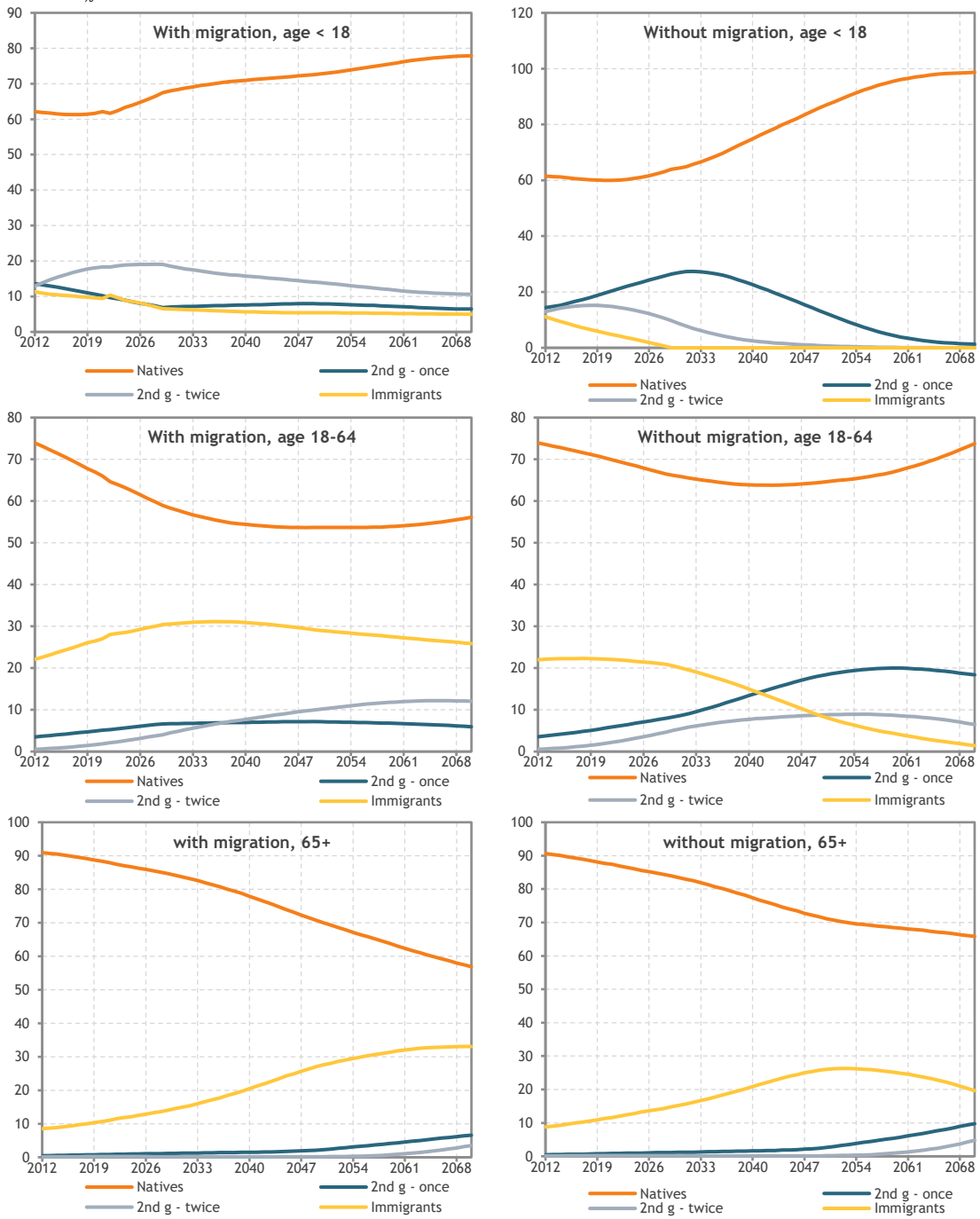
The importance of the movement through the ages of various cohorts, and of the large current stock of young 2nd generation immigrants, is also illustrated in Graph 16, which shows the population shares of the four migrant groups by three age categories. The top right panel indicates that even without migration after 2012, for the reasons mentioned above, among children the proportion of 2nd generation-once immigrants will increase up to the mid-2030s. And even with migration, this proportion will decline after those years. Obviously, without migration all child immigrants will have become adults in 2029.

With migration, the proportion of immigrants among children will decline – because of the projection that immigration of foreigners will be less in the coming years than it was in the recent past – and reaches a stable level of 5% after 2035. The decline of 2nd generation immigrants-twice – and the increase of 2nd generation immigrants-once is partly due to the lack of homogamy by immigrant status in the partnering module of MIDAS, and partly to the fact that they have a higher chance of being selected for emigration.

Among the persons at working age, the percentage of immigrants will reach a maximum of 28% in the mid-2030s, and after that decline a bit to 19% in 2070. Without immigration, the proportion of immigrants in this group starts to decline only at the end of the 2020s; up to that time the number of immigrants that reach the age of 65 is compensated by children becoming adults. The proportion of 2nd generation immigrants among the 18 to 64 rises continuously. Rather surprisingly, this proportion increases *as much* in the scenario *without* migration. After 2060 the projected absolute number of 2nd generation immigrants is smaller in the simulation without migration, but this smaller number still constitutes about the same percentage of a smaller total population. But before that year, there are in fact fewer 2nd generation immigrants 18-64 in the migration scenario than in the no-migration one. The reason is that when there is migration, those persons can and do emigrate, which is not possible in the no-migration scenario. (And, by definition, 2nd generation immigrants cannot immigrate as such, but can only be born in Belgium.)

The immigration wave of the past decades reaches the 65+ only from the 2040s on; the proportion of immigrants (in the migration scenario) among the older population reaches its maximum of 27% in the 2060s. It is striking that up to the mid-2050s there is hardly any difference with the no-migration scenario. The proportions of 2nd generation immigrants become significant only towards the end of the projection period; and they are higher in the no-migration scenario than in the migration one, for the same reasons as explained with reference to the 18-64 group.

Graph 16 Proportion in four migrant statuses, 2012-2070, with and without migration, by age group



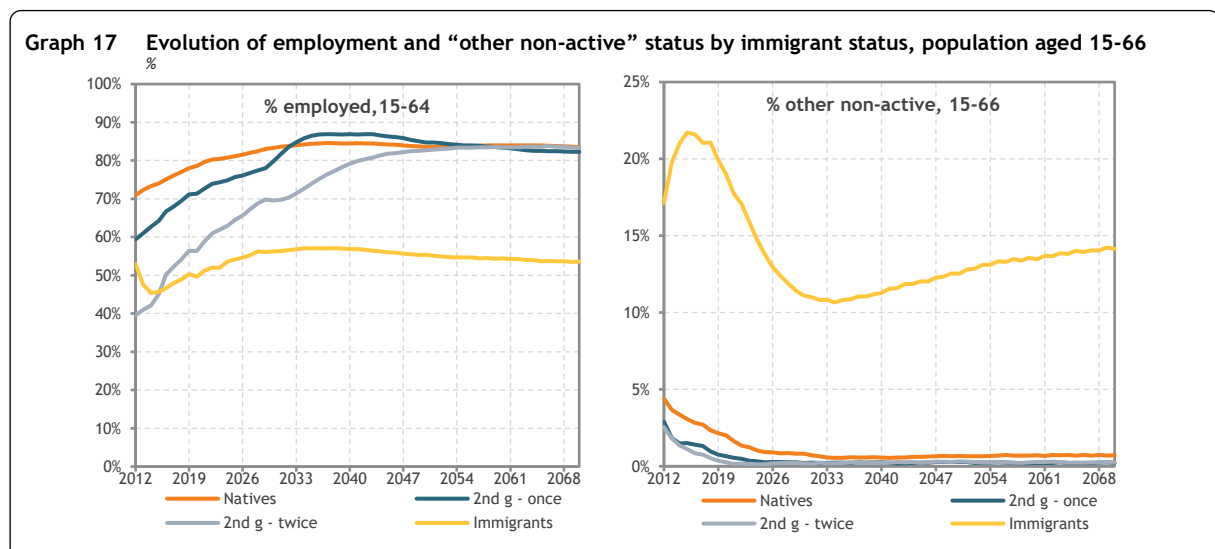
4.3. Socio-economic position, earnings and pensions by immigrant status

In this section we consider the socio-economic position, earnings and pensions by immigrant status. For a discussion of the educational attainment levels of the various immigration groups, we refer to Table 1 of Annex 3 of this Working Paper. We assume that the observed differences in educational attainment between immigrant groups will persist in the future. In particular, that of immigrants would remain lower than that of the native population.

We start this section by considering the prevalence of labour market states by immigrant status. The labour market module of MIDAS is basically a decision tree in which individual “risk” profiles, represented by the results of logistic regressions are combined with alignment proportions by age group and gender combination, as explained above. The immigration status (immigrant vs. non-immigrant) is included in the logistic equations. In addition, differences between immigrants and native Belgians should be the results of differences in characteristics of the two groups which are included in these logistic equations, viz. educational attainment level, age, and gender.

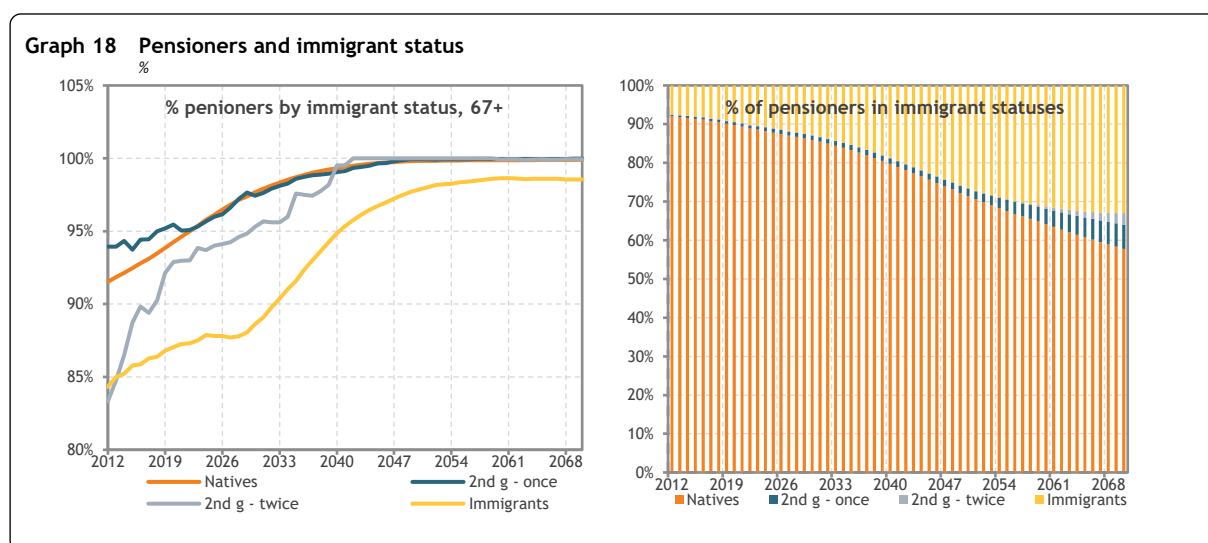
Above we have seen that the starting data show that in 2011, immigrants compared to natives were less likely to be employed and to be retired, and more likely to be unemployed and to be in the state of ‘other non-active’. Does this remain the case during the projection period? Graph 17, left panel, indicates that the employment rates of 2nd generation immigrants converge towards those of natives, as those children of immigrants age. Employment among immigrants remains lower than that of natives, and the employment gap between them and natives even increases. We have to keep in mind that the stock of immigrants is constantly replenished through immigration and emigration; in MIDAS new foreign-born immigrants at first are assigned the status of “other non-active” and may take some years to find employment.

Unemployment rates by immigrant status converge in about 10 years to nearly the same level for all immigrant groups (not shown in graph). The proportions in retirement (including CELS) by immigrant status among the 15-66 also converge, although this takes much longer, and is more a matter of the natives converging to the low rate of immigrants. Finally, the proportion of migrants who are “other non-active” first rises, then declines and after the mid-2030s gradually increases again. This is partly the result of the general evolution of the number of people who are ‘other non-active’, though it seems that this situation becomes more concentrated among immigrants. Among 2nd generation immigrants, the proportion of people in this status first drop, and then converges to that of natives.

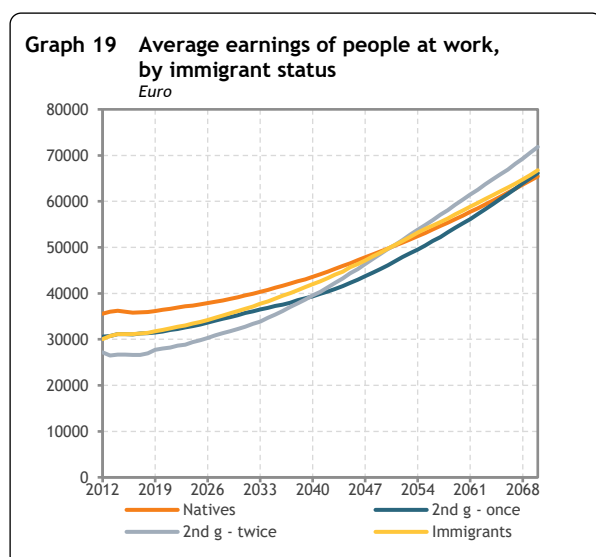


Graph 18, left graph, shows the evolution of the percentage of pensioners by immigrant status among the 67plus. Pensioners are defined as those who are retired and receive a legal pension. Among natives and 2nd generation immigrants, the proportion receiving a pension increases to virtually 100% by the

mid-2040s. The proportion of pensioners among immigrants in this age group also increases strongly, but they always lag behind natives, and the proportion never reaches 100%. Due to the large number of persons who are “other non-active” among immigrants, some may have never worked or received a social security benefit that entails pension rights; others may have immigrated after or close to the statutory retirement age. The right graph of Graph 18 shows that the proportion of immigrants among pensioners will increase drastically: from 8% in 2012 to 26% in 2070. This reflects the demographic evolution discussed above. The proportion is very high towards the end of the projection period, because the large influx of immigrants in the first two decades of the 21st century is going into retirement around that year. On the other hand, the children of those immigrants will start to retire in large numbers only after 2070.



Graph 19 shows average earnings by immigrant status. The average earnings of immigrants start about 16% lower than those of native Belgians. Those of 2nd generation immigrants-twice are even lower, as



this group is quite young in 2012. Over time, the average earnings of all groups converge towards each other. To understand these results, we should explain briefly how earnings are simulated in MIDAS. This is done in four steps (see the Working Paper, section 6.2). First, it is determined who works part-time, and if so, for how many hours, using the proportions of part-time workers by age and gender observed in 2011. Second, for persons who did not change employment status, hourly earnings are updated from those of last year, taking account of the average wage increase, as projected by MALTESE. Third, for new employees (including civil servants), gross earnings per hour are

calculated from a regression equation, including as explanatory variables age, gender, the educational attainment level, being partnered, having children and hours worked in the previous year (if any), and adding a stochastic factor. For the newly self-employed, hourly earnings are copied from a random

person who was already self-employed. Fourth, hours and hourly earnings are multiplied to get total gross earnings.⁴ In the current context, the crucial aspect of this procedure is that immigrant status plays no role in the simulation of earnings of new employees and self-employed persons. This explains why the earnings by immigration status converge in the long run. The fact that for people who do not change employment status, earnings are derived from those of last year, explains why this convergence process takes a long time.

The comparison of the total gross individual state pension (including retirement pensions, survival pensions and the Guaranteed Income for the Elderly) by immigrant status is shown in Graph 20. The pensions of immigrants are lower than those of native Belgians (cf. Peeters *et al.*, 2017, 81); in the starting year, the difference is 26% for men and 22% for women. These differences are only slightly reduced throughout the projection period. As foreign immigrants can start their career only after they have found work in Belgium, their career lengths are shorter than those of natives, implying lower retirement pensions. Pensions from foreign countries are imputed for immigrants from the old EU15 countries only, see the Working Paper, section 7.5.



⁴ See Dekkers *et al.* (2010b) for a more detailed discussion of this 'monetary alignment'.

5. Conclusions

The dynamic microsimulation model MIDAS aims to simulate the prospective adequacy of social security provisions, specifically pensions. This document describes the introduction of immigration in MIDAS. In previous versions of MIDAS, the starting data included immigrants who were in the Belgian population at the moment of sampling. However, future immigration and emigration were not modelled. We now have introduced modules that simulate exit and entry of foreigners (people born outside Belgium) and of Belgian-born persons. The numbers of immigrants and emigrants are aligned on the population projections regarding migration, that are yearly produced by the FPB. Furthermore, we distinguish 2nd generation immigrants, one or both parents of whom is an immigrant, from other Belgian-born persons.

There are some worries about the methodology: a relatively small number of individuals were the donors for most of the clones; and this small number might be a selective sample of all possible donors for immigration clones. This is an issue that we should look into in the future.

The simulation of immigration involves that immigrant status is included in the behavioral equations regarding demographic and labour-market states and transitions. As foreigners have a high rate of re-emigration, the stock of immigrants is constantly replenished, and this is one reason they keep lagging behind natives in terms of employment, earnings and pensions.

The main results of the projection as regards immigrants are as follows.

1. Including immigration and emigration makes sure the MIDAS population projections nearly exactly match the FPB demographic population projections.
2. The simulations indicate that the proportion of immigrants among children will be halved by 2030, and after the mid-2030s this proportion will start to fall among the population at active age. By contrast, among the 65plus, the share of immigrants will increase continuously to about 30% in 2070. These results are implied by the FPB projections of immigration, emigration, death and (to some extent) fertility, and do not depend on the modelling of migration in MIDAS.
3. The educational attainment levels of immigrants is and would remain for a long time considerably lower than that of natives.
4. Currently, at active age, immigrants, and to a lesser extent also children of immigrants, are less often employed, and more often unemployed than natives. A substantial number of immigrants, but not of 2nd generation immigrants, are in the state of “other non-active”, so having no work nor a social security benefit. Many of these differences disappear after some years in the simulation. However, immigrants would continue to have lower employment rates, and are increasingly often “other non-active”.

5. The proportions of pensioners among the 67+ would converge to 100% in all immigrant groups, except for foreign-born immigrants, which continue to lag behind. In 2070 immigrants would constitute about 26% of all pensioners.
6. Average earnings of immigrants start up lower than those of natives, and continue to be lower.
7. Especially the longer career of natives causes their average pension benefit to exceed that of immigrants; over time the gap would become only slightly smaller, but not disappear completely.

References

- Aleksynska, Mariya, Ahmed Tritah, 2011, Occupation-Education Mismatch of Immigrant Workers in Europe: Context and Policies. CEPII Document de Travail No 2011 – 16 July, Centre d'Etude Prospectives et d'Informations Internationales.
- Barrett, Alan, and Adele Bergin, 2009, Estimating the Impact of Immigration in Ireland. *Nordic Journal of Political Economy*, Volume 35, article 2.
- Borjas, George J., 1994, The Economics of Immigration. *Journal of Economic Literature*, 32, December 1994, 1667-1737
- Borjas, George J., 1994, The Economic benefits of Immigration. *Journal of Economic Perspectives*, 9(2), spring 1995, 3-22.
- Casarico, Alessandra, Carlo Devillanova, 2003, Social security and migration with endogenous skill upgrading. *Journal of Public Economics* 87 (2003) 773–797.
- Chassamboulli, Andri, and Theodore Palivos, 2010, "Give me your Tired, your Poor," so I can Prosper: Immigration in Search Equilibrium. Department of Economics, University Of Cyprus, Working Paper 2010-12
- Chènard, Denis, 2000, Individual alignment and group processing: an application to migration processes in DYNACAN. in Mitton, L., Sutherland, H. and Weeks, M. (eds), *Microsimulation Modelling for Policy Analysis*, Cambridge University Press, Cambridge.
- Chojnicki, Xavier, Frédéric Docquier, Lionel Ragot, 2005, Should the U.S. Have Locked the Heaven's Door? Reassessing the Benefits of the Postwar Immigration. Discussion Paper Series, Institute for the Study of Labor, July 2005.
- Chusseau, Nathalie and Dumont, Michel and Hellier, Joël and Rayp, Glenn and Willemé, Peter, Wage Inequality and Immigration: Western-Europe in the Sixties (June 1, 2007). Available at SSRN: <https://ssrn.com/abstract=2276463> [16/06/2018] or <http://dx.doi.org/10.2139/ssrn.2276463> [16/06/2018]
- Corluy, Vincent, and Ninke Mussche, 2014, Participatie van Niet-Europese Migranten op de Belgische Arbeidsmarkt. *Migratiemaatschappij*, Chapter 9, pp 83-104. Brussels: Kruispunt Migratie en Integratie.
- Corluy, Vincent, and Gerlinde Verbist, 2010, Inkomen en diversiteit: onderzoek naar de inkomenspositie van migranten in België. Berichten Centrum Voor Sociaal Beleid Herman Deleeck, Universiteit Antwerpen, Mei 2010
- Corluy, Vincent, and Gerlinde Verbist, 2014, *Can Education Bridge the Gap? Education and the Employment Position of Immigrants in Belgium*. University of Antwerp, Centre For Social Policy, ImPROvE Discussion Papers No 14/02, March 2014
- Corluy, Vincent, and Ive Marx, 2014, Sociale Zekerheid en Migratie in België: een paradoxale situatie van inclusie en armoede. *Tijdschrift voor Welzijnswerk*, 38(339), p. 9-16.

- Dekkers, Gijs, R. Desmet, G. De Vil, 2010 - The long-term adequacy of the Belgian public pension system: An analysis based on the MIDAS model, Working Paper 10-10, May 2010, Brussels: Federaal Planbureau.
http://www.plan.be/publications/publication_det.php?lang=nl&TM=30&IS=63&KeyPub=946.
 [09/04/2018]
- Dekkers, Gijs, Hermann Buslei, Maria Cozzolino, Raphael Desmet, Johannes Geyer, Dirk Hofmann, Michele Raitano, Viktor Steiner, Paola Tanda, Simone Tedeschi, Frédéric Verschuere, 2010b - What are the consequences of the European AWG-projections on the adequacy of pensions? An application of the dynamic micro simulation model MIDAS for Belgium, Germany and Italy. in O'Donoghue, C. (ed.) Life Cycle Microsimulation Modelling, Chapter 12, pp. 230-253, Lambert Academic Press.
- Dekkers, Gijs, and Richard Cumpston, 2011 - On weights in dynamic-ageing microsimulation models. Paper presented at the 3rd General Conference of the International Microsimulation Association: "Microsimulation and Policy design", Statistics Sweden, Stockholm, Sweden, June 8th to 10th, 2011.
https://www.researchgate.net/publication/324248357_On_weights_in_dynamic-ageing_microsimulation_models [09/04/2018]
- Dekkers, Gijs, and Richard Cumpston, 2012, On weights in dynamic microsimulation modelling, International Journal of Microsimulation, Vol. 5(2), Fall 2012, pp. 59-65.
http://www.microsimulation.org/IJM/V5_2/4_IJM_5_2_Dekkers_Cumpston.pdf [10/7/2015]
- Dekkers, Gijs, 2013a, *The imputation of the level of education, nationality and country of birth in the "MIMOSIS 2001" dataset using the SILC*. Deliverable project MIMBEL. Mimeo Federal Planning Bureau January 2013.
- Dekkers, Gijs, 2013b, *Do birds of a foreign feather flock together? The estimation of the marriage market reduced form equations of MIDAS_BE using the SILC*. Deliverable project MIMBEL. Mimeo Federal Planning Bureau May 2013.
- Dekkers, Gijs, 2014, *O brother, where are thou from? The nationality of immigrants in Belgium*. Deliverable project MIMBEL. Mimeo Federal Planning Bureau January 2014.
- Dekkers, Gijs, 2014b, *The estimation of the labour market reduced form equations of MIDAS_BE using the SILC*. Mimeo project MIMBEL, Version November 18th, 2014.
- Dekkers, Gijs, 2014c, *Immigration in MIDAS*, MIMBEL Deliverable October 2014
- Dekkers, Gijs, Raphaël Desmet, Nicole Fasquelle, Saskia Weemaes, 2015 - The social and budgetary impacts of recent social security reform in Belgium. In Ioana Salagean, Catalina Lomos & Anne Hartung, "The young and the elderly at risk: Individual outcomes and contemporary policy challenges in European societies", Intersentia. ISBN 978-1-78068-343-0. Chapter 6, pp. 129-158.
- Dekkers, Gijs, 2015. On the Modelling of Immigration and Emigration using LIAM2. Federal Planning Bureau, Note LIAM2 – 11155. November 2015. https://www.researchgate.net/publication/286869730_On_the_modelling_of_immigration_and_emigration_using_LIAM2
- Dekkers, Gijs, Bart Capéau, André Decoster, 2015. *The adequacy of public pensions among immigrants in Belgium*. Report on the MIMBEL project. October 2015.
- Dekkers, Gijs, and Raphaël Desmet, 2011, "Inégalité et pauvreté chez les personnes âgées: une projection à long terme" Reflets & Perspectives de la Vie Economique», Inégalités et Pauvreté , n° 2011-4 (December 2011), pp. 155-168.

- Dekkers, Gijs, Seiichi Inagaki and Raphaël Desmet, 2012, Dynamic Microsimulation Modeling for Policy Support: An Application to Belgium and Possibilities for Japan. Review of Socionetwork Strategies, 6(2), pp 31-47.
- Dekkers, Gijs, 2013, Love it or Leave It. naturalisation and emigration following immigration in Belgium. Version September 20th, 2013. Mimeo Federal Planning Bureau. Project MIMBEL
- Dekkers, Gijs, Raphaël Desmet, Tomas Pettersson, Ádám Rézmovits, Olle Sundberg, and Krisztián Tóth, 2015(b), *On using dynamic microsimulation models to assess the consequences of the AWG projections and hypotheses on pension adequacy: Simulation results for Belgium, Sweden and Hungary*. Paper presented to the SPC Working Group on Ageing Issues (SPC-WG-AGE), Brussels, March 27th, 2015.
- Dekkers, Gijs, Raphaël Desmet and Karel Van den Bosch (2019), *De bijdrage van MIDAS aan het Jaarlijkse verslag voor 2019 van de Studiecommissie voor de Vergrijzing. Technisch rapport*. Mimeo, Juli 2019.
- Denton, Frank T., Christine H. Feaver, Byron G. Spencer, 1997, Immigration, Labour Force, and the Age Structure Of The Population. IESOP Research Paper No. 24, November 1997.
- Docquier, Frédéric, Çağlar Özden and Giovanni Peri, 2010, The Wage Effects of Immigration and Emigration. Institut de Recherches Économques et Sociales de l'Université Catholique de Louvain. Discussion Paper 2010-44
- Drinkwater, Stephen, Paul Levine, Emanuela Lotti and Joseph Pearlman, 2002, The Economic Impact of Migration: A Survey. University of Surrey, Discussion Papers in Economics, DP 01/03. Also published in *International Journal of Economic Development*, Vol. 5 No. 2
- Dustmann, Christian, Tommaso Frattini And Caroline Halls, 2010, Assessing the Fiscal Costs and Benefits of A8 Migration to the UK, *Fiscal Studies*, vol. 31, no. 1, pp. 1–41 (2010) 0143-5671
- European Commission/Directorate-General Employment, social affairs and equal opportunities, 2010, *Joint Report on Social Protection and Social Inclusion 2010*, Luxembourg: Publications Office of the European Union, 2010.
- European Commission, 2018, *Pension Adequacy Report 2018*. Brussels: European Commission/Directorate-General for Employment, Social Affairs and Inclusion.
- Eurostat, 2015, Eurostat Statistics Explained - Material deprivation and low work intensity statistics.
http://ec.europa.eu/eurostat/statistics-explained/index.php/Material_deprivation_and_low_work_intensity_statistics [09/10/2015]
- Eurostat, 2018, *Glossary: Equivalised disposable income*. Retrieved from the Eurostat website: http://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:Equivalised_disposable_income [16/5/2018]
- Eurostat, 2018b, *Glossary: Gini coefficient*. Retrieved from the Eurostat website: http://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:Gini_coefficient [16/5/2018]
- Eurostat, 2018c, *Gini coefficient of equivalised disposable income*. Retrieved from the Eurostat website: http://ec.europa.eu/eurostat/web/products-datasets/product?code=sdg_10_40 [16/5/2018]
- Eurostat, 2018d, *Glossary: At-risk-of-poverty rate*. Retrieved from the Eurostat website: http://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:At-risk-of-poverty_rate

- Federal Planning Bureau, 2017, *Economic Policy Committee's Ageing Working Group. Belgium: Country Fiche 2017*. Note Federal Planning Bureau.
- Federal Public Service Employment, Labour and Social Dialogue and Centre for Equal Opportunities and Opposition to Racism, 2013, *Socio-Economische Monitoring 2013*. Brussels: Federale Overheidsdienst Werkgelegenheid, Arbeid en Sociaal Overleg.
- Gil Alsonso, Fernando, 2009, Can the rising pension burden in Europe be mitigated by immigration? Modelling the effects of selected demographic and socio-economic factors on ageing in the European Union, 2008-2050. *Vienna Yearbook of Population Research 2009*, pp. 123-147.
- Can the rising pension burden in Europe... (PDF Download Available). Available from: https://www.researchgate.net/publication/227639369_Can_the_rising_pension_burden_in_Europe_be_mitigated_by_immigration_Modelling_the_effects_of_selected_demographic_and_socio-economic_factors_on_ageing_in_the_European_Union_2008-2050 [accessed Jun 01 2018].
- High Council of Finances, 2017, *Jaarlijks verslag van de Studiecommissie voor de Vergrijzing*. Paper retrieved from the website of the Federal Planning Bureau: <https://www.plan.be/publications/publication.php?lang=nl> [16/5/2018]
- Holzmann, R. and R. Hinz, 2005, *Old-Age Income Support in the Twenty-first Century: An International Perspective on Pension Systems and Reform*, Web Version of 18 February 2005, World Bank.
- Kemnitz, Alexander, 2003, *Immigration, Unemployment and Pensions*. *The Scandinavian Journal of Economics*, 105(1), March 2003, pp. 31-47.
- Lelkes, Orsolya, Eszter Zólyomi, 2011, *Poverty and Social Exclusion of Migrants in the European Union*, Policy Brief March 2011, European Centre for Social Welfare Policy and Research, Vienna.
- Markaki, Yvonne, 2014, *Do labour market conditions shape immigrant-native gaps in employment outcomes? A comparison of 19 European countries*. Institute for Social and Economic Research, University of Essex, ISER Working Paper Series 2014-41.
- Moore, Eric, Michel Pacey, 2003, *Changing Income Inequality and Immigration in Canada, 1980-1995*. *Canadian Public Policy - Analyse de Politiques*, 29(1), pp. 33-52.
- Munz, Sonja, and Martin Werding, 2005, *Public pensions and international migration: some clarifications and illustrative results*. *Journal of Pension Economics and Finance*, 4 (2), pp. 181-207.
- O'Donoghue, Cathal, Howard Redway, John Lennon, 2010, *Simulating Migration in the Pensim2 Dynamic Microsimulation Model*. *International Journal of Microsimulation*, 3(2), pp. 65-79.
- OECD, 2007, *Glossary of Statistical Terms*. <https://stats.oecd.org/glossary/detail.asp?ID=6805> [18/4/2018].
- OECD, 2008, *Jobs for Immigrants. Volume 2 : Labour Market Integration in Belgium, France, the Netherlands and Portugal*. Paris: OECD.
- OECD, 2011, *International Migration Outlook: SOPEMI 2011*, OECD Publishing, http://dx.doi.org/10.1787/migr_outlook-2011-en [01/06/2018] <http://www.oecd.org/migration/internationalmigrationoutlook2011.htm> [01/06/2018]

- OECD, 2018, *What are Equivalence Scales?* Paper retrieved from the OECD website <http://www.oecd.org/eco/growth/OECD-Note-EquivalenceScales.pdf> [16/05/2018]
- OECD/European Union (2015), *Indicators of Immigrant Integration 2015: Settling In*, OECD Publishing, Paris. <http://dx.doi.org/10.1787/9789264234024-en>
- Peeters, Hans, Karel Neels, Nele Havermans, Dimitri Mortelmans, Nina Donvil, David De Wachter, Lise Cordeel, 2017, *De Kleur van Vergrijzing. Over pensioenen van migranten*. Tielt: Lannoo.
- Serrano, Felipe, Begoña Eguía, Jesús Ferreiro, 2011, Public pensions' sustainability and population ageing: Is immigration the solution? *International Labour Review*, 150(1-2), June 2011. pp. 63-79.
- STATBEL/Statistics Belgium, 2018, Open Data. TF_CENSUS_2011_HC11_L = Census 2011 - Leeftijd, Arbeidsmarktsituatie, Land van staatsburgerschap, Verblijfplaats, Economische sector, Geslacht, Beroepsstatuut. http://statbel.fgov.be/sites/default/files/files/opendata/Census%202011%20-%20Leeftijd%2C%20Arbeidsmarktsituatie%2C%20Land%20van%20staatsburgerschap%2C%20Verblijfplaats%2C%20Economische%20sector%2C%20Geslacht%2C%20Beroepsstatuut/TF_CENSUS_2011_HC11_L.xlsx [17/4/2018]
- STATBEL/Statistics Belgium, 2018b, *Census 2011 België: Glossarium* http://census2011.fgov.be/info/gloss2_nl.html [17/4/2018]
- Vandresse, M., 2014, *Een methode voor huishoudensprojecties: het HPRM-model*, Working Paper 9-14, Brussel, Federaal Planbureau
- Vandresse, Marie, 2015, *Une modélisation de l'évolution future de la migration internationale pour la Belgique*. Federal Planning Bureau, Working Paper 2-15, March 2015. http://www.plan.be/admin/uploaded/201503201138360.WP_1502_10948.pdf [10/07/2015]
- Vandresse, Marie, Johan Duyk, Jean-Marc Paul, 2017, *Demografische vooruitzichten 2016-2060 - Bevolking en huishoudens*. Paper retrieved from the website of the Federal Planning Bureau https://www.plan.be/admin/uploaded/201703070757590.FOR_POP1660_11440_N.pdf [17/5/2018]
- Vandresse, Marie, Johan Duyk, Jean-Marc Paul, 2018, *Perspectives démographiques 2016-2060 – analyses de sensibilité, scénarios alternatifs et effets budgétaires et sociaux*. Paper retrieved from the website of the Federal Planning Bureau https://www.plan.be/admin/uploaded/201802120922470.WP_1801_11620.pdf [17/5/2018]
- Vandresse, Marie, Johan Duyk, Jean-Marc Paul, 2018b, *Demografische vooruitzichten 2017-2070 - Bevolking en huishoudens*. Paper retrieved from the website of the Federal Planning Bureau https://www.plan.be/admin/uploaded/201802221219560.FOR_POP1770_11621_N.pdf
- Vandresse, Marie, Johan Duyk, Jean-Marc Paul, 2019, *Demografische vooruitzichten 2018-2070 - Bevolking en huishoudens*. Paper retrieved from the website of the Federal Planning Bureau https://www.plan.be/uploaded/documents/201901240958450.FOR_POP1870_11813_N.pdf
- Vasquez Grenno, Javier, 2009, *Spanish Pension System: Population Aging and Immigration Policy (Preliminary)*. Working Paper Departament d'Economia Aplicada. Universitat Autoonoma de Barcelona. April 2009.

MIDAS: Modelling van demografische variabelen

Annex 3 bij 'MIDAS 2.0: Revisie van een
dynamisch microsimulatiemodel'

Januari 2023

Gijs Dekkers, gd@plan.be,
Raphaël Desmet, rd@plan.be,
Karel Van den Bosch, kvdb@plan.be

Inhoudstafel

1. Inleiding	1
2. Onderwijsniveau	2
3. Sterfte	4
4. Samenstelling van de huishoudens	5
4.1. Inleiding	5
4.2. Beschrijving van de procedures	8
4.2.1. Overzicht van de procedures	8
4.2.2. Leave_the_nest	12
4.2.3. First_time_mothers	14
4.2.4. One_more_child	15
4.2.5. Baby_born	18
4.2.6. Divorce	19
4.2.7. Partnership	21
4.2.8. Marriage	23
4.2.9. Orphan_adoption	24
4.2.10. Caring_daughter	24
4.2.11. Pseudo_lipro	25
4.3. Resultaten	26
4.3.1. Projectie van de bevolking naar leeftijd en geslacht	26
4.3.2. Projectie van de bevolking naar LIPRO-positie	27
4.3.3. Andere aspecten van de projectie van demografische variabelen	33
5. Conclusie	38
Literatuur	39

Lijst van tabellen

Tabel 1	Verdeling onderwijsniveau van nieuwe individuen in simulatie naargelang geslacht en migratiestatus	3
Tabel 2	Logistische regressie kans op overlijden in 2012	4
Tabel 3	LIPRO-positie: vergelijking Datawarehouse variabele met variabele geconstrueerd in MIDAS ..	7

Tabel 4	Overzicht procedures inzake positie in het huishouden	10
Tabel 5	Aligneringstabel voor aantal kinderen per vrouw, naargelang regio en gezinssituatie	15
Tabel 6	Aligneringstabel voor scheiding van gehuwden en uiteengaan van samenwonenden	20
Tabel 7	Leeftijd man naargelang leeftijd vrouw binnen koppels volgens startdata	34
Tabel 8	Leeftijd man naargelang leeftijd vrouw in jaar van koppeling	35

Lijst van figuren

Figuur 1	Finale onderwijsniveau van vrouwen en mannen 16+	3
Figuur 2	Mogelijke transities tussen lipro-posities met aanduiding van procedures	11
Figuur 3	Vruchtbaarheid van vrouwen per leeftijd, gesimuleerd in MIDAS en volgens bevolkingsprojecties	19
Figuur 4	Aantal geadopteerde kinderen per jaar	24
Figuur 5	Frequentie van hereniging van bejaarde ouders met (surrogaat-) dochter of zoon, naar geslacht	25
Figuur 6	Totale bevolking volgens MIDAS simulatie en bevolkingsprojectie	26
Figuur 7	Samenstelling van de bevolking naar brede leeftijdsgroep, per geslacht, volgens MIDAS	27
Figuur 8	Proportie kinderen naar leeftijdsgroep, bij koppels en in éénoudergezinnen, volgens MIDAS en volgens de bevolkingsprojecties	28
Figuur 9	Proportie alleenstaanden per leeftijdsgroep en geslacht, volgens MIDAS en volgens de bevolkingsprojecties	29
Figuur 10	Proportie koppels zonder kinderen (gehuwd of samenwonend) per leeftijdsgroep en geslacht, volgens MIDAS en volgens de bevolkingsprojecties	30
Figuur 11	Proportie gehuwden, onder vrouwen 15-64 met zonder kinderen, volgens MIDAS en bevolkingsprojectie	30
Figuur 12	Proportie koppels met kinderen (gehuwd of samenwonend) per leeftijdsgroep en geslacht, volgens MIDAS en volgens de bevolkingsprojecties.	31
Figuur 13	Proportie alleenstaande ouders per leeftijdsgroep en geslacht, volgens MIDAS en volgens de bevolkingsprojecties	32
Figuur 14	Proportie 'overigen' per leeftijdsgroep en geslacht, volgens MIDAS en volgens de bevolkingsprojecties	33
Figuur 15	Aantal kinderen in huishouden per vrouw met kinderen (16-64), naargelang partnersituatie, volgens MIDAS	34
Figuur 16	Onderwijsniveau man naargelang onderwijsniveau vrouw in startdata en gesimuleerd	35
Figuur 17	Migrant-status man naargelang migrant-status vrouw in startdata en gesimuleerd	36
Figuur 18	Proportie met diploma hoger onderwijs naargelang lipro-positie	37

Federaal Planbureau

Het Federaal Planbureau (FPB) is een instelling van openbaar nut die beleidsrelevante studies en vooruitzichten maakt over economische, socio-economische en milieuvraagstukken. Daarnaast bestudeert het de integratie van die vraagstukken in een context van duurzame ontwikkeling. Het stelt zijn wetenschappelijke expertise onder meer ter beschikking van de regering, het Parlement, de sociale gesprekspartners, nationale en internationale instellingen.

De werkzaamheden van het FPB worden steeds gekenmerkt door een onafhankelijke benadering, transparantie en aandacht voor het algemeen welzijn. De kwaliteit van de gegevens, een wetenschappelijke methodologie en de empirische geldigheid van de analyses staan daarbij centraal. Tot slot zorgt het FPB voor een ruime verspreiding van de resultaten van zijn werkzaamheden en draagt zo bij tot het democratisch debat.

Het Federaal Planbureau is EMAS en Ecodynamische Onderneming (drie sterren) gecertificeerd voor zijn milieubeheer.

<https://www.plan.be>

e-mail: contact@plan.be

Overname wordt toegestaan, behalve voor handelsdoeleinden, mits bronvermelding.

Verantwoordelijke uitgever: Saskia Weemaes

Wettelijk Depot: D/2023/7433/3

Federaal Planbureau

Belliardstraat 14-18, 1040 Brussel

tel.: +32-2-5077311

e-mail: contact@plan.be

<https://www.plan.be>

Modellering van demografische variabelen in MIDAS

Annex 3 bij 'MIDAS 2.0: Revisie van een dynamisch microsimulatiemodel'

Januari 2023

[Gijs Dekkers, gd@plan.be](mailto:gd@plan.be), [Raphaël Desmet, rd@plan.be](mailto:rd@plan.be), [Karel Van den Bosch, kvdb@plan.be](mailto:kvdb@plan.be)

Abstract – Deze annex presenteert de modules inzake demografische processen in MIDAS, inclusief gezinsvorming, met meer technisch detail dan mogelijk is in de Working Paper.

1. Inleiding

Dit document presenteert de modellering in MIDAS van de demografische variabelen, gebeurtenissen en evoluties: loop van de bevolking, gezinsvorming en onderwijsniveau. Deze modellering is volledig herzien en sterk uitgebreid ten opzichte van Dekkers et al. (2010), met het doel de resultaten van MIDAS te aligneren aan de projectie van de positie in het huishouden (LIPRO-positie), die jaarlijks door het Planbureau uitgevoerd en gepubliceerd wordt. Hieronder bespreken we eerst de modellering en de resultaten voor het onderwijsniveau, omdat dit een bepalende factor is voor de andere variabelen. Vervolgens komt de sterfte aan de orde. Geboorte is een integraal onderdeel van de gezinsvorming, en wordt daarom in dat kader behandeld. De modellering van de LIPRO-positie – equivalent met de gezinsvorming – vormt de hoofdmoot van dit document.

Deze nieuwe modellering sluit aan bij onze betrachting om de resultaten van MIDAS zoveel als mogelijk te laten aansluiten bij de demografische en sociaal-economische projecties die het FPB produceert, met name in het kader van MALTESE. De belangrijkste techniek die wij hiervoor gebruiken is de ‘alignering met rangorde’. Zie sectie 2 van het WP voor een niet-technische bespreking van alignering.

2. Onderwijsniveau

Het onderwijsniveau is een bepalende factor op de arbeidsmarkt, en ook voor demografische transitie. De modellering hiervan is dezelfde gebleven als in de vorige versie van MIDAS, zie Dekkers et al. (2010). In de startdata is een variabele met het onderwijsniveau (hoogst behaalde diploma) opgenomen, afkomstig uit de administratieve Census 2011. De oorspronkelijke ISCED-waarden van de variabele in de startdata zijn gehercodeerd tot 3 categorieën:

1. ISCED 0-2 (geen onderwijs, alleen lager onderwijs, lager secundair onderwijs)
2. ISCED 3-4 (hoger secundair onderwijs)
3. ISCED 5-6 (hoger onderwijs)

Het onderwijsniveau is alleen gegeven voor personen van 16 jaar en ouder. Daarenboven heeft 14 percent van de 16+ers een ontbrekende waarde op deze variabele. Voor de meeste van deze was het onderwijsniveau onbekend, terwijl een beperkter aantal nog niet in België woonde ten tijde van de Census. De ontbrekende waarden werden geïmputeerd, op basis van verdelingen uit de Arbeidskrachtenenquête.

Veel personen tussen 16 en 25 jaar hebben een waarde op deze variabele, maar zijn nog schoolgaand of studierend, zodat het in 2011 hoogst behaalde diploma waarschijnlijk niet het onderwijsniveau is dat zij uiteindelijk zullen behalen. Aan deze personen, evenals de personen beneden 16 jaar wordt op willekeurige wijze een onderwijsniveau toegekend, volgens de verdeling getoond in Tabel 1. Deze komt overeen met de verdeling van het onderwijsniveau van de 25-34 jarigen in de startdata. Hierbij geldt de beperking dat studerenden van 22 jaar en ouder altijd ISCED 5-6 behalen, en studerenden tussen 19 en 21 jaar minstens ISCED 3-4.

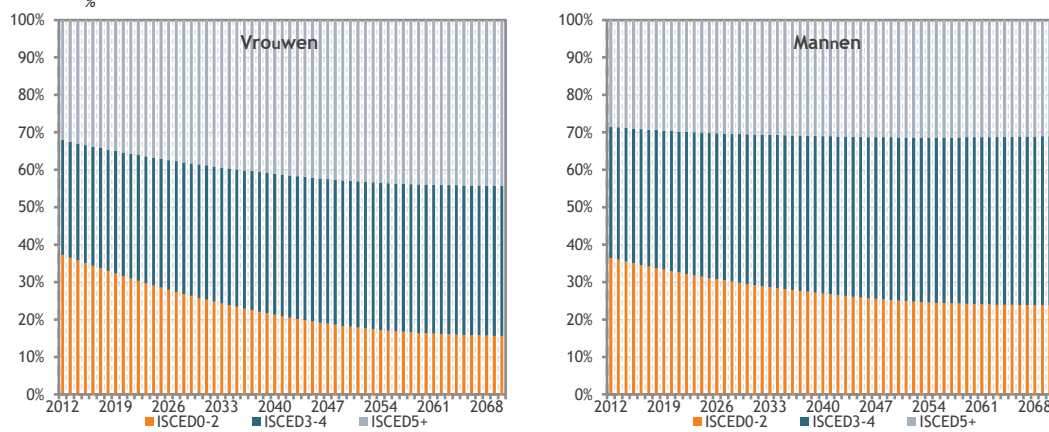
In de simulatie wordt in ieder projectiejaar aan iedereen die nog geen onderwijsniveau heeft (in feite de pasgeborenen) op willekeurige wijze een onderwijsniveau toegekend, volgens de verdeling getoond in Tabel 1.¹ Dit onderwijsniveau behoudt het individu gedurende zijn gehele leven. Voor de doeleinden van MIDAS is het niet relevant om de diverse stadia in de onderwijs carrière te simuleren; enkele uitzonderingen op deze regel worden hieronder vermeld. Behalve met immigratiestatus wordt geen rekening gehouden met andere achtergrondvariabelen, omdat onderwijsniveau aan het begin staat van de causale keten die de arbeidsmarktpositie en de inkomens bepaalt. De implicatie van deze methode is dat de verdeling van het onderwijsniveau van de 25-34 jarigen in alle projectiejaren dezelfde is, terwijl het onderwijsniveau van de bevolking als geheel geleidelijk stijgt, naarmate oudere cohorten worden vervangen door jongere met een gemiddeld hoger onderwijsniveau, zie Figuur 1. Het onderwijsniveau stijgt sterker bij vrouwen dan bij mannen, en is steeds hoger bij de vrouwen dan bij de mannen. Dit weerspiegelt de huidige situatie van 25-34 jarigen, zie Tabel 1.

¹ Voor nieuwe immigranten wordt het onderwijsniveau gekopieerd van de "donor" van hun gegevens. Zie Annex 2: Immigration, sectie 2.2.

Tabel 1 Verdeling onderwijsniveau van nieuwe individuen in simulatie naargelang geslacht en migratiestatus

Geslacht	Migratie-status	ISCED0-2	ISCED3-4	ISCED5+	Totaal
Vrouwen	autochtoon	10,0%	37,0%	53,1%	100,0%
	2de gen. éénzijdig	14,7%	43,0%	42,3%	100,0%
	2de gen. tweezijdig	20,6%	51,8%	27,6%	100,0%
	immigrant	35,5%	29,8%	34,8%	100,0%
Mannen	autochtoon	17,5%	44,9%	37,7%	100,0%
	2de gen. éénzijdig	27,2%	46,3%	26,5%	100,0%
	2de gen. tweezijdig	31,2%	51,1%	17,7%	100,0%
	immigrant	43,0%	32,1%	25,0%	100,0%

Bron: MIDAS startdata

Figuur 1 Finale onderwijsniveau van vrouwen en mannen 16+

Bron: MIDAS

Noot: Van personen die nog studeren is het onderwijsniveau weergegeven dat zij uiteindelijk zullen bereiken

3. Sterfte

De sterfte in MIDAS is gealigneerd aan het aantal overlijdens naargelang leeftijd en geslacht, volgens de bevolkingsprojecties. Een vernieuwing ten opzichte van de vorige versie van MIDAS (Dekkers et al. 2010), is dat de kans op overlijden daarnaast ook wordt bepaald door het onderwijsniveau en de partnerstatus (al of niet een partner hebben), aan de hand van een logistische vergelijking. De coëfficiënten in deze vergelijking zijn geschat met behulp van de startdata, waarin een variabele voorkomt die aangeeft of een persoon die op 31/12/2011 aanwezig was in de steekproef, in de loop van 2012 is overleden. In de selectieve steekproef (zie Annex1: Steekproef en data), verder beperkt tot 65+ers, is het aantal overlijdens 4 776 (5,1 %) op een totaal van 94 108 personen. Tabel 2 toont de resultaten van de schatting van de logistische vergelijking. Deze zijn weinig verrassend: gegeven de leeftijd, hebben vrouwen een kleinere kans op overlijden dan mannen, evenals personen met partner vergeleken met alleenstaande ouderen (hoewel het verschil vermindert met de leeftijd); hoe hoger geschoold, hoe lager de kans op overlijden op een bepaalde leeftijd.

Tabel 2 Logistische regressie kans op overlijden in 2012

	Schatting coëfficiënt	Standaard fout
Leeftijd	0,0903***	(0,00514)
Vrouw	-2,522***	(0,372)
Interactie leeftijd * vrouw	0,0231***	(0,00458)
ISCED niveau 0-2	Referentie categorie	
ISCED niveau 3-4	-0,142***	(0,0461)
ISCED niveau 5-6	-0,317***	(0,0536)
Onderwijsniveau missing	0,136***	(0,0448)
Met partner	-1,565***	(0,509)
Interactie leeftijd * met partner	0,0145**	(0,00665)
Leeftijd (gecentreerd op 75) - kwadraat	0,000704***	(0,000266)
Interactie leeftijd-kwadraat * met partner	0,000602	(0,000511)
Constante	-9500***	(0,403)
Aantal observaties	94108	

4. Samenstelling van de huishoudens

4.1. Inleiding

De samenstelling van de huishoudens wordt bestudeerd aan de hand van de LIPRO-typologie (een indicator van de positie van individuen binnen hun huishouden), de omvang van de huishoudens en het aantal kinderen per huishouden. De LIPRO-positie wordt afgeleid van de relatie van een individu tot de referentiepersoon van het gezin, en de burgerlijke staat. Het Federaal Planbureau produceert ieder jaar projecties van de LIPRO-positie (Vandresse, 2014), en de MIDAS-resultaten worden gealigneerd aan de projecties daterend van mei 2021 (na calibrering van de laatste, zie onder). Ook kijken we naar de evolutie van het aantal kinderen per vrouw, de homogamie binnen koppels, en de samenhang tussen de LIPRO-positie en het onderwijsniveau.

In de versie van MIDAS beschreven in Dekkers et al. (2010) was de simulatie van huishoudvorming beperkt tot sterfte, geboorte, huwelijk en samenwonen, scheiding en het verlaten van het ouderlijk huis. Bij de eerste twee processen werd rekening gehouden met mortaliteit en vruchtbaarheid naargelang leeftijd. Voor de koppeling van partners, en voor de kans op scheiding, waren gedragsvergelijkingen uitgewerkt. Personen verlieten het ouderlijk huis bij huwelijk of samenwoning, of als zij 24 jaar werden. De resulterende gezinssituaties waren volledig het resultaat van de dynamiek in het model. In latere versies van MIDAS was de proportie samenwonende of gehuwde vrouwen al gealigneerd aan de projectie van de LIPRO-posities.

De LIPRO-posities zijn de volgende (zie Vandresse, 2014, voor meer informatie en referenties naar de literatuur)². Een 'kind' wordt in de LIPRO-typologie gedefinieerd door de relatie tot de referentiepersoon van het huishouden, en staat los van de leeftijd of de arbeidsmarktstatus. Niet alle kinderen zijn dus afhankelijke kinderen zoals gedefinieerd in sociale zekerheids- en fiscale regelingen. Om de LIPRO-typologie te kunnen simuleren, moesten twee vereenvoudigingen in deze typologie worden aangebracht; deze zijn aangeduid tussen vierkante haken. De belangrijkste is dat personen in collectieve huishoudens worden samengevoegd met alleenstaanden. Tussen (ronde) haakjes worden de gangbare afkortingen van deze posities aangegeven, die ook hieronder zullen gebruikt worden.

1. Alleenstaande (SING);
2. gehuwde zonder kind (MAR0);
3. gehuwde met kind (MAR+);
4. kind bij gehuwd paar (CMAR);
5. ongehuwd samenwonende zonder kind (UNM0);
6. ongehuwd samenwonende met kind(eren) (UNM+);
7. kind bij ongehuwd paar (CUNM);

² Zie ook <https://www.ksz-bcss.fgov.be/nl/dwh/variabledetail/rijks-en-ksz-register/Variables/lipro-positie#attachement-da66fcaa-83c9-49ee-dd61-a3d08e2183df> voor de implementatie van de LIPRO-typologie in het Datawarehouse op basis van gegevens uit het Rijksregister.

8. hoofd eenoudergezin (H1PA);
9. kind in een eenoudergezin 'C1PA);
10. andere inwonende; [samengevoegd met overige personen] (NFRP)
11. overige personen (OTHR);
12. wonend in een collectief huishouden [samengevoegd met alleenstaande] (COLL)

Door de combinatie van aligering en zorgvuldige modellering in diverse processen, die hieronder in detail zullen worden beschreven, zijn we in staat om deze exogene LIPRO projecties na te bootsen. Toch komen de MIDAS resultaten niet perfect overeen met de FPB projecties van de LIPRO-positie. Hiervoor zijn twee redenen:

- de LIPRO-posities van gerelateerde individuen staan niet los van elkaar: wanneer ouders van kinderen huwen, scheiden of gaan samenwonen, verandert bijvoorbeeld ook de LIPRO-positie van de kinderen in het huishouden. Dit maakt het lastig om binnen alle leeftijd-gender categorieën de verdeling van de LIPRO-positie volledig te reproduceren.
- de FPB projecties van LIPRO-positie per leeftijdscategorie staan los van elkaar, hetgeen in latere projectiejaren soms leidt tot merkwaardige verschillen tussen opeenvolgende leeftijdsgroepen, die in een dynamisch microsimulatiemodel soms moeilijk te volgen zijn. Een voorbeeld is het profiel van het percentage alleenstaande ouders bij oudere (65plus) vrouwen naargelang haar leeftijd. In 2020 is dit 3,9 % bij de 65-69 jarigen, 3,6 % bij de 70-74 jarigen, en het stijgt dan tot 6,2 % bij de 80-85 jarigen. In 2070 zijn deze percentages respectievelijk 3,0 %, 1,4 % en 6,6 %, dus een daling bij de jongere ouderen en een stijging bij de oudere ouderen. De transities in LIPRO-positie die nodig zijn om deze percentages te reproduceren komen soms in conflict met andere aligeringdoelen, of zijn weinig realistisch. Een ander voorbeeld, dat in sectie 4.3.2 aan de orde komt, zijn de percentages alleenstaanden en koppels bij de 80plussers.

Anmaak van data en aligeringstabellen

Bij de voorbereiding van de startdata voor MIDAS werden een aantal huishoudens opgesplitst (zie Annex1: Steekproef en data). Sommige huishoudens bevatten meer dan één familiekeren, dat wil zeggen meer dan één koppel, of een ouder-kind paar waarvan de ouder niet de referentiepersoon of zijn of haar partner is. Omdat dit problemen zou veroorzaken bij de simulatie van toekomstige veranderingen in de huishoudensamenstelling, werden deze huishoudens opgesplitst. Een alternatief zou zijn om deze huishoudens volledig te verwijderen, maar dit een groter verlies van informatie betekenen. Daarenboven hebben de personen in deze huishoudens een specifiek profiel. Er werden 2773 huishoudens opgesplitst (1 % van het totaal), hetgeen resulteerde in 2873 extra huishoudens (sommige originele huishoudens telden meer dan twee familiekeren), die 7288 personen bevatten. Hierdoor nam het aantal grote huishoudens (6 of meer personen) af van 7,2 % tot 5,1 %. De LIPRO-positie werd aangepast aan de nieuwe huishoudensamenstelling.³

³ Bijvoorbeeld: een multi-generatieel gezin met grootmoeder, dochter en kleinkind werd opgesplitst tot een alleenstaande vrouw en een éénoudergezin. Origineel had de grootmoeder de positie 'hoofd eenoudergezin', de dochter de positie 'kind in een eenoudergezin', en het kleinkind de positie 'overige personen'. Na opsplitsing is de grootmoeder alleenstaande, de dochter hoofd eenoudergezin en het kleinkind een kind.

Deze opsplitsing van een beperkt aantal huishoudens impliceert een belangrijke vereenvoudiging van de structuren die huishoudens mogelijk kunnen aannemen. Alle personen kunnen verdeeld worden onder vier categorieën: referentiepersoon, partner, kind en overig. Kinderen en 'overigen' kunnen geen partner en ook geen kinderen hebben. Vooral dit laatste vereenvoudigt sterk de situaties waarmee in de simulatie rekening moet worden gehouden.

Zoals gezegd werd in het startjaar de LIPRO-positie gereconstrueerd, en dit op basis van informatie die ook voor alle volgende projectiejaren beschikbaar is. Hierdoor wordt de LIPRO-positie op een consistente manier bepaald doorheen de projectie. De aldus bepaalde LIPRO-positie komt echter niet volledig overeen met de LIPRO-positie zoals vastgelegd in de Datawarehouse gegevens.⁴ Tabel 3 toont de verschillen en de overeenkomsten. De overeenkomst is zeer sterk (minstens 95 %) voor alle posities, met uitzondering van de 'andere inwonenden' (NFRP) en de 'overige personen' (OTHR), waarvan een groot deel door MIDAS aan andere LIPRO-posities toebedeeld wordt, met name 'kind bij gehuwd koppel (CMAR)' en 'kind bij alleenstaande ouder (C1PA)'. Dit is het resultaat van de opsplitsing van een aantal huishoudens. Alle personen in collectieve huishoudens (COLL) worden als alleenstaande beschouwd (SING). De vermindering van het aantal 'andere inwonenden' en 'overige personen', die als restcategorieën dienen, is voor de modellering in MIDAS een voordeel. Door de reconstructie wijzigen ook de marginale aantallen per LIPRO-positie, vergelijk de kolom 'totaal' met de rij '% rijgewijs'.

Tabel 3 LIPRO-positie: vergelijking Datawarehouse variabele met variabele geconstrueerd in MIDAS

MIDAS-variabele	Datawarehouse variabele												totaal
	SING	MARO	MAR+	CMAR	UNMO	UNM+	CUNM	H1PA	C1PA	NFRP	OTHR	COLL	
SING	100,0%	0,9%	0,3%	0,2%	1,3%	0,9%	0,0%	2,5%	1,1%	0,4%	2,1%	100,0%	15,6%
MARO	0,0%	97,8%	1,1%	0,1%	1,8%	0,1%	0,0%	0,0%	0,3%	8,9%	3,3%	0,0%	17,7%
MAR+	0,0%	1,0%	96,7%	0,3%	0,1%	2,0%	0,0%	0,0%	0,6%	5,5%	0,1%	0,0%	19,5%
CMAR	0,0%	0,0%	0,0%	95,5%	0,0%	0,0%	2,1%	0,0%	0,0%	26,7%	0,1%	0,0%	18,7%
UNMO	0,0%	0,3%	0,0%	0,0%	95,9%	0,4%	0,0%	0,0%	0,0%	0,5%	4,3%	0,0%	4,6%
UNM+	0,0%	0,0%	0,5%	0,0%	0,9%	96,1%	0,1%	0,2%	0,0%	0,9%	0,3%	0,0%	5,4%
CUNM	0,0%	0,0%	0,0%	0,5%	0,0%	0,0%	96,3%	0,0%	0,2%	3,3%	0,3%	0,0%	4,6%
H1PA	0,0%	0,0%	1,4%	0,5%	0,0%	0,6%	0,3%	97,2%	1,2%	0,2%	16,5%	0,0%	4,8%
C1PA	0,0%	0,0%	0,0%	2,7%	0,0%	0,0%	1,1%	0,0%	95,2%	19,2%	20,9%	0,0%	7,5%
OTHR	0,0%	0,0%	0,0%	0,1%	0,0%	0,0%	0,0%	0,1%	1,3%	34,3%	52,2%	0,0%	1,5%
totaal	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%
totaal (n)	367428	462712	519577	500853	122930	143293	120712	109676	174218	43587	40437	29389	2634812
% rijgewijs	13,9%	17,6%	19,7%	19,0%	4,7%	5,4%	4,6%	4,2%	6,6%	1,7%	1,5%	1,1%	100,0%

De FPB-projectie van de LIPRO-posities bestaat uit de absolute aantallen per jaar, geslacht, leeftijdscategorie van 5 jaar breed (0-4, 5-9, ... 90+) en LIPRO-positie. Deze data werden omgezet in een aligeringstabel die per jaar, geslacht en leeftijdscategorie de procentuele verdeling over de (vereenvoudigde) LIPRO-posities bevat. Omdat deze verdeling niet volledig overkwam met verdeling in de MIDAS startdata van 2011,⁵ werd de projectie van de LIPRO-posities gehercalibreerd (per geslacht en leeftijdscategorie), zodat deze voor 2011 overeen kwam met de MIDAS-startdata. Deze (omvangrijke)

⁴ Om deze reden heet de betreffende variabele binnen MIDAS 'pseudo_lipro'.

⁵ Overigens kwamen deze ook niet perfect overeen met de geobserveerde verdeling in de MIDAS startdata van 2011. Wellicht is dit te wijten aan de selectie van huishoudens uit de startdata, zie Annex1: Steekproef en data. .

aligneringstabel wordt in alle procedures gebruikt inzake gezinsvorming en positie in het huishouden; per procedure wordt de gepaste selectie gemaakt.

4.2. Beschrijving van de procedures

4.2.1. Overzicht van de procedures

De samenstelling van huishoudens in projectie wordt bepaald door de situatie in de startdata en de demografische transitie die personen ondergaan. Hier geven we allereerst een overzicht van de diverse procedures en hun onderlinge samenhang, waarna een gedetailleerde bespreking van de afzonderlijke procedures volgt. De volgorde hieronder is die waarin deze procedures worden uitgevoerd.

- *leave_the_nest* : deze procedure bepaalt welke kinderen wanneer het ouderlijk huishouden verlaten. Zij krijgen dan de LIPRO-positie van alleenstaande. (In dezelfde periode kunnen volgende procedures deze positie nog veranderen.) Daarnaast wijzigt deze procedure de LIPRO-positie van sommige kinderen in die van 'overige personen'. De procedure is gealigneerd op de proporties kinderen (CMAR, CUNM, C1PA) en overige personen (OTHR)⁶.
- *first_time_mother* : deze procedure zorgt voor de transitie van vrouwen zonder kinderen (gehuwd, samenwonend of alleenstaand; MAR0, UNM0, SING) naar de corresponderende LIPRO-posities van vrouwen met kinderen (MAR+, UNM+, H1PA), met andere woorden, de transitie van vrouw zonder kind naar moeder. Dit is gealigneerd op de proporties vrouwen met kinderen. De positie van eventuele partners verandert mee.
- *one_more_child*: deze procedure identificeert de moeders die tweede of volgende kinderen krijgen. Alleen gehuwde of samenwonende moeders komen hiervoor in aanmerking. De LIPRO-positie van deze vrouwen verandert hierbij niet. Deze procedure is gealigneerd op het aantal geboorten volgens de bevolkingsprojectie, na aftrek van de geboorten van eerste kinderen volgens *first_time_mother*. Er wordt naar gestreefd de verdeling van de het aantal kinderen per moeder (1, 2, 3, 4+) in de startdata zoveel mogelijk te reproduceren.
- *baby_born*: in deze procedure worden de nieuwe personen die er volgens *first_time_mother* en *one_more_child* moeten bijkomen, effectief gesimuleerd, en wordt aan alle individuele variabelen van deze personen een waarde gegeven.
- *divorce*: deze procedure regelt de scheiding van gehuwde personen en het uiteengaan van samenwonende koppels. De proporties gehuwde en samenwonende koppels die uiteengaan zijn afgeleid uit de startdata. Eventuele kinderen worden meestal aan de moeder toegewezen en soms aan de vader.
- *partnership*: deze procedure regelt enerzijds hoeveel en welke mannen en vrouwen een nieuw koppel gaan vormen. Dit is gealigneerd aan de proporties alleenstaanden met of zonder kinderen (SING, H1PA) bij de vrouwen. Anderzijds bepaalt *partnership* welke vrouw aan welke man gekoppeld wordt.

⁶ Wanneer gesproken wordt over proporties gaat het steeds over proporties per geslacht en leeftijdscategorie van 5 jaar breed.

Hierbij wordt de homogamie naargelang leeftijd, onderwijsniveau en immigratiestatus zoals geobserveerd in de startdata zoveel mogelijk gereproduceerd.⁷

- *marriage* : deze procedure bepaalt welke koppels huwen. Koppels kunnen onmiddellijk na koppeling huwen, of in een later jaar. Koppels die niet huwen, wonen samen. Er wordt geen verschil gemaakt tussen geregistreerd en ongeregistreerd samenwonen.

Samen met sterfte regelen bovenstaande procedures de belangrijkste gezinstransities in de levensloop van mensen in België. Daarnaast zijn er nog twee procedures die een eerder marginale rol in het geheel vervullen.

- *caring_daughter*: uit de startdata bleek dat er onder hoogbejaarden, en vooral onder de vrouwen, een relatief groot aantal de status ‘alleenstaande ouder’ of ‘overig’ had. Het gaat wellicht om personen die zorg ontvangen van inwonende kinderen of andere familieleden. Deze procedure regelt de nodige transities om deze situaties te reproduceren.
- *orphan_adoption*: in de startdata kwamen enkele alleenstaande minderjarigen voor, en in de projectie ontstaat deze situatie soms wanneer een alleenstaande ouder sterft. Omdat alleenstaande minderjarigen voor de simulatie van inkomens problematisch zijn, worden deze situaties opgeheven door de betreffende minderjarigen aan andere huishoudens toe te wijzen;

Tabel 3 presenteert de verschillende procedures, met vermelding van de leeftijdsgroepen die daaraan onderhevig zijn. Figuur 2 geeft een overzicht van de mogelijke transities, met vermelding van de procedure die die transitie simuleert. Daarna worden deze procedures in meer detail besproken.

⁷ Zoals blijkt uit deze omschrijving worden in MIDAS alleen heteroseksuele koppels gevormd. Wegens het – vooralsnog – kleine aantal homoseksuele koppels was een modellering van deze situaties niet mogelijk en niet zinvol. In de startdata komen enkele koppels voor van uitsluitend mannen of uitsluitend vrouwen. Deze blijven in projectie behouden, en in de programmering wordt met hun situatie rekening gehouden, zodat zij bijvoorbeeld niet aan een andere vrouw of man gekoppeld worden.

Tabel 4 Overzicht procedures inzake positie in het huishouden

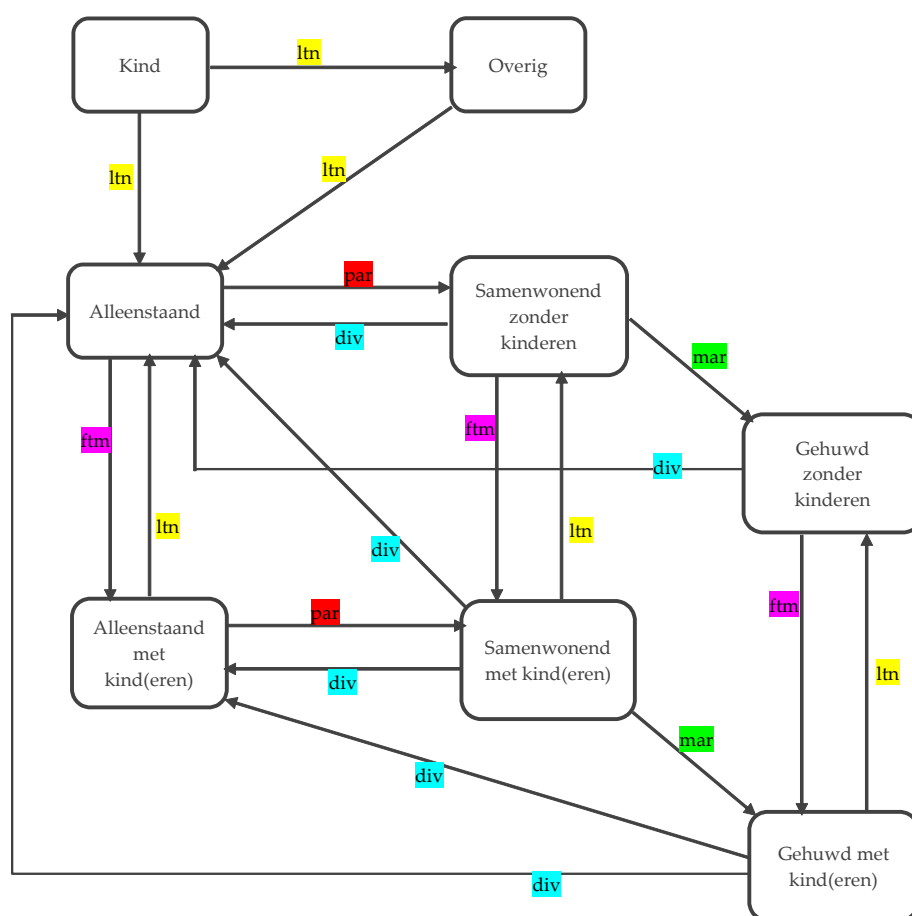
Naam*	Transities van --> naar** (LIPRO positie)	Alignering op:**	Leeftijdsgroep	Opmerkingen**
leave_the_nest	kinderen [3,6,9] --> alleenstaande [1]	% kinderen [3,6,9]	15-44	Daarnaast wordt rekening gehouden met aanwezigheid kinderen bij vrouwen in koppel van 40 jaar en ouder.
	kinderen [3,6,9] --> overig [10]	% overig [10]	15-44	
	overig [10] --> alleenstaande [1]	% overig [10]	15+	
first_time_mothers	vrouwen (alleenstaand of in koppel) [1,2,5] --> vrouw met kind [3,6,8]	% vrouwen met kinderen [3,6,8]	18-40	
one_more_child	-	Aantal geboorten - geboorten bij first_time_mothers	19-44	Alleen bij vrouwen met partner en kinderen [3,6].
baby_born	-		0	Situatie moeder bepaalt LIPRO-positie kind [4, 7 of 9]
divorce	gehuwd of samenwonend [2,3,5,6] --> Alleenstaand, al dan niet met kind [1,8]	% scheidingen per jaar	15+	LIPRO-positie eventuele kinderen verandert analoog [4,7] --> [9]
partnership	Alleenstaand, al dan niet met kind [1,8] --> gehuwd of samenwonend [2,3,5,6]	% alleenstaand, al of niet met kinderen [1,8]	15+	LIPRO-positie eventuele kinderen verandert analoog [9] --> [4,7]
marriage	samenwonend [5,6] --> gehuwd [3,4]	% gehuwd [3,4]	15+	LIPRO-positie eventuele kinderen verandert analoog [7] --> [4]
caring_daughter	alleenstaand [1] --> alleenstaand met kind of overig [8,10]	% alleenstaand met kind [8], % overig [10]	70+	Voegt ouderen toe aan huishouden van één van de "kinderen"
orphan_adoption	alleenstaand of overig [8,10] --> kind bij koppel [4,7]	-	< 18	Om huishoudens met alleen minderjarigen te vermijden, na overlijden ouder(s)
pseudo_lipro	-	-	Allen	Bepaalt nieuwe LIPRO-positie voor alle leden huishouden na ieder van bovenstaande procedures

* het toevoegsel "_procedure" of "_process" is weggelaten

** de cijfers tussen vierkante haken geven de LIPRO-posities aan; zie boven in de tekst

Figuur 2 geeft een overzicht van de mogelijke transitie, met vermelding van de procedure die die transitie simuleert. Beginnend bij de positie 'kind' (CMAR, CUNM of C1PA) simuleert *leave_the_nest* de transitie naar de posities 'overig' en 'alleenstaande'. Alleenstaanden kunnen alleenstaande ouders worden via *first_time_mother*, of samenwonend zonder kinderen via *partnership*. Die laatste procedure regelt ook de transitie van alleenstaande ouder naar samenwonend met kinderen. Via *marriage* kunnen samenwonenden met of zonder kinderen veranderen in gehuwden met of zonder kinderen. *Divorce* splitst huishoudens op: samenwonenden of gehuwden zonder kinderen worden twee alleenstaanden, koppels met kinderen worden verdeeld in één alleenstaande en één alleenstaande ouder. Tenslotte heeft de procedure *leave_the_nest* ook implicaties voor de ouder(s): wanneer het laatste kind het ouderlijk huis verlaat, verandert hun LIPRO-positie van gehuwd, of samenwonende of alleenstaande ouder (MAR+, UNM+ of HIPA) in respectievelijk gehuwde, samenwonende, of alleenstaande zonder kinderen (MAR0, UNM0 of SING). Niet vermeld in Figuur 2 is sterfte. Hierdoor wordt de overlevende partner een alleenstaande, of een alleenstaande ouder als er kinderen in het huishouden zijn. Wanneer de laatste ouder overlijdt wijzigt de status van het inwonende volwassen kind in alleenstaande, of in 'overig' wanneer er meerdere inwonende kinderen zijn.

Figuur 2 Mogelijke transities tussen lipro-posities met aanduiding van procedures



Noot: **ltn**: leave_the_nest; **ftm**: first_time_mother; **par**: partnership; **mar**: marriage; **div**: divorce.

De positie 'kind' is als zodanig geen lipro-positie. Naargelang de situatie van de ouder(s) omvat deze de posities 4, 7 en 9. Omwille van de leesbaarheid van het schema zijn deze posities niet afzonderlijk vermeld.

De procedures *one_more_child* en *baby_born* zijn niet vermeld, omdat deze geen transities tussen lipro-posities simuleren.

De procedures *caring_daughter* en *orphan_adoption* zijn niet vermeld, omdat deze een marginale rol vervullen, en om het schema niet overvol te maken.

Hoewel het schema vele pijlen telt, zijn er ook transities uitgesloten.

- Vanuit de posities ‘kind’ en ‘overig’ kan men alleen overgaan naar de positie ‘alleenstaand’; van daaruit kunnen andere LIPRO-posities bereikt worden.
- Het is niet mogelijk om terug te keren naar de posities ‘kind’ en ‘overig’. (Per uitzondering kan dit wel in de procedure *caring_daughter*, die echter weinig transities teweegbrengt.)
- Het is niet mogelijk om vanuit de posities ‘alleenstaand’, met of zonder kinderen, direct over te gaan naar de posities ‘gehuwd’, met of zonder kinderen. Om deze transitie te maken, moet men passeren via de positie ‘samenwonend’, dit kan echter wel in dezelfde simulatieperiode.

4.2.2. Leave_the_nest

De procedure *leave_the_nest* simuleert de transities van de positie ‘kind’ naar ‘alleenstaande’, ‘kind’ naar ‘overig’, en ‘overig’ naar ‘alleenstaande’. ‘Kinderen’ zijn iedereen, onafhankelijk van de leeftijd, die eigen kind, stiefkind of pleegkind is van de referentiepersoon of van zijn of haar partner. De positie ‘overig’ omvat iedereen die geen partner is van de referentiepersoon, en geen kind van de referentiepersoon of zijn of haar partner. Zoals hierboven aangegeven, combineert deze de originele LIPRO-categorieën ‘andere inwonenden’ en ‘overigen’. Inspectie van de variabele ‘relatie met referentiepersoon’ in de startdata toonde aan dat de meesten familieleden zijn zoals ouders, schoonzonen en -dochters, kleinkinderen, ooms en tantes, al is er ook een grote groep personen die niet gerelateerd zijn. Het percentage ‘overigen’ is zeer laag onder minderjarigen, neemt toe tussen 20 en 30 jaar, en blijft onder de 30plussers ongeveer stabiel. Het is hoger bij mannen dan bij vrouwen. Het is een erg heterogene categorie; het onderwijsniveau en de tewerkstellingsgraad liggen wat lager dan bij hun leeftijdsgenoten, maar de verschillen zijn niet groot. ‘Overig’ is een residuele categorie die als zodanig moeilijk te simuleren is.

Voor personen die een – al of niet gesimuleerde – Inkomensvervangende Tegemoetkoming (IVT, uitgekeerd door de FOD-SZ) wordt aangenomen dat zij in de status ‘kind’ of ‘overig’ blijven.⁸

De procedure *leave_the_nest* beschrijft in de eerste plaats het verlaten van het ouderlijk huis door kinderen van 15 jaar of ouder. Deze kans is onder meer afhankelijk van het geslacht, de leeftijd, opleidingsniveau, regio, arbeidsmarktstatus en migratiestatus. De betreffende kinderen vormen vervolgens een eigen huishouden en krijgen de LIPRO-positie ‘alleenstaande’. Daarnaast wordt voor sommige kinderen de LIPRO-positie veranderd in ‘overig’ (OTHR), zonder dat zij van huishouden veranderen (zie onder voor toelichting bij dit laatste aspect). Deze ‘overigen’ kunnen in een latere periode alsnog het huishouden verlaten om alleenstaande te worden. Wanneer een persoon het laatste kind is dat zelfstandig gaat wonen – of verandert van ‘kind’ in ‘overig’ – dan wijzigt ook de LIPRO-positie van de ouder(s) van gehuwd, samenwonend of alleenstaand mét kinderen naar de corresponderende positie zonder kinderen. Met deze bijkomende transities wordt zo veel mogelijk rekening gehouden bij de selectie van de kinderen die alleenstaand of ‘overig’ worden.

De procedure *leave_the_nest* omvat de volgende stappen:

⁸ We simuleren alleen de personen die tussen hun 21^{ste} levensjaar en hun pensionering steeds deze uitkering ontvangen, wegens invaliditeit vanaf de geboorte of de jeugd.

1. Transitie van sommige 'kinderen' naar 'overigen', en transitie van sommige 'overigen' naar 'alleenstaanden'.
2. Bepaling kans op verlaten ouderlijk huishouden voor kinderen
3. Identificatie van kinderen die bij hun moeder moeten blijven wonen, of die die juist moeten achterlaten
4. Bepaling aligneringsdoelstellingen en transitie van sommige 'kinderen' naar 'alleenstaanden'

Hieronder beschrijven we deze stappen in meer detail.

1. Transities van en naar 'overig'

De transitie van 'kind' naar 'overig' is beperkt tot de 15-44 jarigen, en is gealigneerd op het percentage 'overigen'⁹. Hierbij wordt rekening gehouden met het hogere percentage 'overigen' in Brussel en onder immigranten, maar voor de rest is de selectie van 'overigen' willekeurig. Bij deze transitie blijven deze personen in hetzelfde huishouden. Dit laatste komt uiteraard niet overeen met transities van 'kind' naar 'overig' in de realiteit. Het is echter niet duidelijk hoe het nieuwe huishouden van 'overigen' geselecteerd zou moeten worden. De motivatie voor de niet-verhuizing is dat het profiel van de bestemmingshuishoudens van 'overigen' waarschijnlijk overeenkomt met hun huishouden van origine, zodat het voor de resultaten niet uitmaakt of we deze 'overigen' laten verhuizen tussen huishoudens, of in hun oorspronkelijke huishoudens laten blijven.

Wanneer in de loop der tijd in sommige gender-leeftijdscategorieën het gesimuleerde percentage 'overigen' hoger wordt dan het percentage volgens de aligneringstabel, worden sommige 'overigen' veranderd in 'alleenstaanden'.

2. Bepaling kans op verlaten ouderlijk huishouden voor kinderen

De transitie van 'kind' naar 'alleenstaande' is gealigneerd per gender-leeftijdsgroep, en ook per originele LIPRO-positie (kind bij gehuwde ouders, kind bij samenwonende ouders, kind bij alleenstaande ouder), en wordt daarnaast bepaald door een groot aantal variabelen door middel van een logistische regressievergelijking. Omdat we in de startdata geen transities observeren tussen 'kinderen' en 'alleenstaanden', is deze vergelijking geschat op de cross-sectionele gegevens met als afhankelijke variabele de LIPRO-positie 'kind' dan wel alle andere posities (niet alleen 'alleenstaande', daar de positie 'alleenstaande' in feite dient als de poort naar alle andere LIPRO-posities, behalve 'overig'). De variabelen in de vergelijking zijn: leeftijd (ook in kwadraat en derdemacht), gender, onderwijsniveau (ook interactietermen met leeftijd en gender), regio, immigrant en socio-economische status (in opleiding, werknemer of ambtenaar, zelfstandige, werkloze, invalide, gepensioneerde, andere niet-actieve). Met dit laatste is hier rekening gehouden, omdat uiteraard jongvolwassenen die niet meer in opleiding zijn vaker al zelfstandig wonen dan anderen; er bleek in dit opzicht weinig verschil te zijn tussen werkenden en personen met een uitkering. De volledige resultaten van de logistische vergelijking zijn verkrijgbaar op aanvraag.

⁹ Volgens de aligneringstabel, en deze percentages zijn altijd per gender-leeftijdsgroep van 5 jaar breed. Om de tekst niet te belasten wordt dit niet steeds vermeld.

3. Identificatie van kinderen die bij hun moeder moeten blijven wonen, of die die juist moeten achterlaten

Zoals gezegd, kan de procedure *leave_the_nest* ook de LIPRO-positie van de ouder(s) wijzigen. Hierdoor werden soms (dwz., in sommige jaren voor sommige leeftijdscategorieën) de aligneringspercentages voor de posities ‘gehuwd met kinderen’, ‘samenwonend met kinderen’ en ‘alleenstaande met kinderen’ voor vrouwen van 40 jaar en ouder niet behaald, of juist overschreden. Om dit te vermijden wordt de transitie van het laatste inwonende kind naar alleenstaande geblokkeerd (als het gesimuleerde percentage lager was dan het aligneringspercentage), of juist geforceerd (in het andere geval).

4. Bepaling aligneringsdoelstellingen en transitie van sommige ‘kinderen’ naar ‘alleenstaanden’

De doelstelling voor het aantal kinderen dat de transitie naar alleenstaande moet ondergaan wordt berekend uit het verschil tussen het feitelijk percentage kinderen (na aftrek van degenen die de transitie naar ‘overig’ hebben gemaakt) en het aligneringspercentage (met minimum 0). Dit gebeurt afzonderlijk voor kinderen bij gehuwde of samenwonende koppels en kinderen bij alleenstaande ouders. Aan de personen die de transitie ondergaan wordt een nieuw huishouden toegewezen, en zij krijgen de LIPRO-positie ‘alleenstaande’.

4.2.3. *First_time_mothers*

De procedure *first_time_mothers* simuleert de transities van de posities ‘alleenstaande’, ‘samenwonend zonder kinderen’ en ‘gehuwd zonder kinderen’ naar de corresponderende posities mét kinderen. Met andere woorden, deze procedure beschrijft de transitie dat vrouwen die nog geen moeder zijn, voor de eerste keer een kind krijgen. Ze is gealigneerd op de percentages van de LIPRO-posities met kinderen, en beperkt tot vrouwen tussen 18 en 40 jaar. Bijna alle vrouwen krijgen hun eerste kind na hun 17de en vóór hun 40ste verjaardag; deze beperking vergemakkelijkt de simulatie, omdat haar kinderen het ouderlijk huishouden beginnen te verlaten vanaf 40 jaar. Impliciet in bovenstaande beschrijving is dat vrouwen met de LIPRO-positie ‘kind’ of ‘overig’ a priori door het model niet in aanmerking worden genomen voor het krijgen van kinderen. De LIPRO-positie van de eventuele partners van de *first_time_mothers* verandert analoog.

De rangorde van mogelijke *first-time-mothers* wordt bepaald aan de hand van een logistische vergelijking, waarin de volgende variabelen voorkomen: leeftijd (ook kwadratisch), gehuwd, samenwonend of alleenstaand (ook interactietermen met leeftijd), onderwijsniveau (ook interactietermen met leeftijd) en regio. Deze vergelijking is geschat op basis van de cross-sectionele startdata, binnen de substeekproef van vrouwen zonder kinderen of met een kind van één jaar en geen andere kinderen, met als afhankelijke variabele het hebben van een kind van 0 jaar. Deze manier om vrouwen die voor de eerste keer een kind krijgen te selecteren is niet sluitend, omdat vrouwen een eerder kind kunnen hebben verloren door overlijden van het kind, of door scheiding en toewijzing aan de vader; we nemen aan dat het aantal van zulke gevallen te verwaarlozen is. De volledige resultaten van de logistische vergelijking zijn verkrijgbaar op aanvraag.

4.2.4. One_more_child

De procedure *one_more_child* simuleert de geboorte van kinderen bij vrouwen die al één of meer kinderen hebben, en maximaal 44 jaar oud zijn. Alleen samenwonende of gehuwde moeders kunnen een bijkomend kind krijgen. Deze beperkte vereenvoudiging ten opzichte van de realiteit vergemakkelijkt de modellering aanzienlijk. Door deze procedure wijzigt de LIPRO-positie van deze moeders niet, noch die van andere gezinsleden. Het aantal moeders dat een bijkomend kind krijgt wordt gealigneerd op het aantal geboorten volgens de bevolkingsprojecties per gewest, na aftrek van de geboorten bij *first_time_mothers*.

De rangorde van moeders die mogelijk een bijkomend kind krijgen wordt bepaald aan de hand van een logistische vergelijking, waarin de volgende variabelen voorkomen: leeftijd (ook kwadratisch), aantal kinderen (1, 2, 3 of 4+), gehuwd, samenwonend of alleenstaand (ook interactietermen met leeftijd en met aantal kinderen), onderwijsniveau (ook interactietermen met leeftijd), vreemdeling (ook interactietermen met aantal kinderen), en regio. Deze vergelijking is geschat op basis van de cross-sectionele startdata, binnen de substeekproef van vrouwen met minstens één kind, met als afhankelijke variabele het hebben van een kind van 0 jaar.

Om bijkomende geboorten te vermijden bij vrouwen waarvan sommige kinderen haar huishouden al verlaten hebben, en terwille van groter realisme, wordt de kans dat een moeder een volgend kind krijgt enkel bepaald voor vrouwen waarvan het jongste kind jonger is dan 7 jaar, omdat een groter leeftijdsverschil tussen twee opeenvolgende kinderen in de startdataset vrijwel niet voorkomt.

De verdeling van kinderen naargelang het aantal kinderen per vrouw is belangrijk voor de inkomens- en welvaartspositie van de betreffende huishoudens. Er is naar gestreefd om zoveel mogelijk de verdeling naar aantal kinderen uit de startdata te behouden, en dit afzonderlijk voor gehuwde of samenwonende koppels en voor alleenstaande ouders. Hiertoe is uit de startdata een aligneringstabel aangemaakt met de verdeling van vrouwen met kinderen naargelang aantal kinderen, per gewest (Tabel 5), die voor alle projectiejaren wordt toegepast. Ter vergelijking wordt ook de verdeling voor alleenstaande moeders vermeld, die minder kinderen hebben dan moeders levend in koppelverband.

Tabel 5 Aligneringstabel voor aantal kinderen per vrouw, naargelang regio en gezinssituatie
%

Regio	Partner situatie	Aantal kinderen				Totaal
		1	2	3	4+	
Brussel	Alleenstaande vrouw	48,1%	32,9%	13,2%	5,7%	100%
Brussel	Vrouw in koppel	34,6%	36,4%	18,1%	10,8%	100%
Vlaanderen	Alleenstaande vrouw	49,9%	34,8%	11,7%	3,6%	100%
Vlaanderen	Vrouw in koppel	33,2%	46,2%	15,6%	5,0%	100%
Wallonië	Alleenstaande vrouw	47,5%	34,2%	13,4%	4,9%	100%
Wallonië	Vrouw in koppel	34,2%	42,5%	16,3%	7,0%	100%

Het voorgaande betekent dat het totaal aantal bijkomende kinderen in de procedure *one_more_child* idealiter aan twee voorwaarden zou moeten voldoen: enerzijds moet het aantal geboorten volgens de bevolkingsprojecties gerealiseerd worden, anderzijds moet dit aantal toelaten de verdeling van het aantal kinderen per vrouw in overeenstemming te brengen met de verdeling volgens Tabel 5. Het is eerder uitzonderlijk dat aan deze twee vereisten tegelijk exact voldaan kan worden. De eerste vereiste krijgt altijd prioriteit, omdat de correcte verdeling van de steekproef naargelang leeftijd en geslacht het

fundament van dynamische microsimulatie vormt. Informeel uitgelegd wordt de volgende procedure gevolgd. Als het aantal geboorten groter is dan het aantal dat nodig is om de verdeling van het aantal kinderen per vrouw volgens Tabel 5 te reproduceren (tweede vereiste), dan wordt het overschot van geboorten willekeurig gespreid over de vrouwen die in aanmerking komen voor een bijkomend kind. In de omgekeerde situatie worden de aantallen geboorten, die nodig zouden zijn om aan de tweede vereiste te voldoen, proportioneel verminderd. Een meer formele uitleg is te vinden in het tekstkader 1.

Kader 1 Procedure voor de toewijzing van bijkomende kinderen

Laten we het aantal te creëren babies volgens de aligeringstabel B noemen. Het aantal babies dat zou toelaten aan de tweede vereiste (spreiding aantal kinderen per vrouw) te voldoen duiden we aan met R, dat als volgt berekend wordt:

$$R = M * \sum_{k=1}^{k=3} (4 - k)(p_k^s - p_k^t)$$

waar M het totaal aantal moeders is, p_k^s de proportie moeders met k kinderen in de startsituatie, en p_k^t dezelfde proportie volgens de aligeringstabel (de target). Als $R > B$ dan worden de aantallen moeders die van één kind naar twee gaan, van twee naar drie en van drie naar 4 gaan proportioneel verlaagd met een calibratiefactor $c = B / R$. Als $R < B$ dan worden eerst de aantallen toegewezen om de aligeringspercentages te bereiken, en wordt de rest toevalsgewijs gespreid over alle moeders.

Bij de toekenning van de bijkomende kinderen naargelang het aantal kinderen worden 9 stappen doorlopen:

1. Bepaal het totaal aantal te simuleren babies, zoals hierboven beschreven: B.
2. Gegeven de huidige verdeling van het aantal kinderen, bepaal het aantal kinderen dat nodig is om deze verdeling in overeenstemming te brengen met de verdeling in Tabel 5: R. Normaal gesproken is dit een positief aantal, omdat de procedure `first_time_mothers` resulteert in een relatief te groot aantal moeders met één kind.
3. Bepaal een calibratiefactor door het aantal te simuleren babies volgens stap 1 te delen door het aantal kinderen bepaald in stap 2: $c = B / R$, met een maximum van 1.
4. Bepaal het aantal moeders met één kind dat een bijkomend kind moet krijgen, om de proportie vrouwen met één kind te verlagen tot het percentage volgens Tabel 4, en vermenigvuldig dit aantal met de calibratiefactor bepaald in stap 3: $Q_1 = c * M * (p_1^s - p_1^t)$.
5. Selecteer de moeders met één kind die effectief een bijkomend kind krijgen, op basis van het aantal bepaald in stap 4, en de score berekend met de logistische vergelijking.
6. Stappen 7 en 8 worden gezet voor k (aantal kinderen) achtereenvolgens gelijk aan 2 en 3:
7. Bepaal het aantal moeders met k kinderen dat een bijkomend kind moet krijgen, om de proportie vrouwen met k kinderen te verlagen tot het percentage volgens Tabel 5, en vermenigvuldig dit aantal met de calibratiefactor bepaald in stap 3; hierbij wordt rekening gehouden met het aantal moeders dat in de voorgaande stap een bijkomend kind heeft gekregen: $Q_k = R_{k-1} + c * M * (p_k^s - p_k^t)$.
8. Selecteer de moeders met k kinderen die effectief een bijkomend kind krijgen, op basis van het aantal bepaald in stap 7, en de score berekend met de logistische vergelijking.
9. Als $B > R$, dan worden de resterende $B - R$ kinderen verdeeld over in aanmerking komende moeders die in de voorgaande stappen nog geen kind hebben gekregen, volgens de score berekend met de logistische vergelijking.

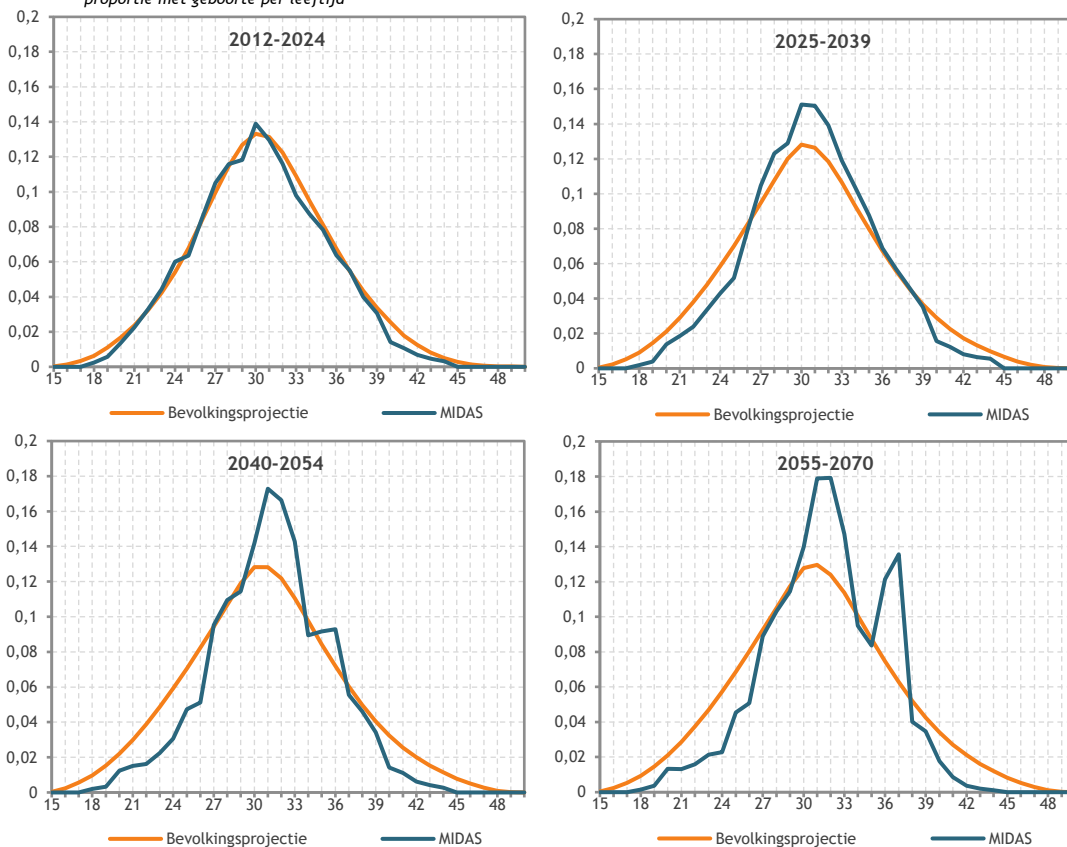
4.2.5. Baby_born

De procedure *baby_born* creëert de nieuwe personen, bij de moeders die geselecteerd zijn in de procedures *first_time_mothers* en *one_more_child*. Het geslacht wordt willekeurig gekozen volgens een verdeling 51,1 % jongens en 48,9 % meisjes; dit is de verdeling van 0-jarigen in de startdata. Andere variabelen worden op de gepaste waarde voor 0-jarigen gezet. Links met de moeder en de eventuele partner van de moeder als vader worden vastgelegd. (Kinderen bij alleenstaande moeders krijgen geen vader toegekend.)

De beschrijving van de procedures voor geboorten maakt duidelijk dat hierbij geen rekening wordt gehouden met de projectie van de vruchtbaarheid die in het kader van de bevolkingsprojecties gemaakt worden. Figuur 3 vergelijkt de vruchtbaarheid naargelang leeftijd vrouw gesimuleerd in MIDAS met die volgens de bevolkingsprojectie, waarbij de volledige projectieperiode is verdeeld in vier tijdvakken. (MIDAS laat geen geboorten toe bij vrouwen jonger dan 18 jaar of ouder dan 44 jaar, omdat dit weinig voorkomt, en dit de simulaties lastiger zou maken.) In het eerste tijdvak is de overeenkomst zeer goed. In de volgende tijdvakken is de geprojecteerde vruchtbaarheid volgens MIDAS lager dan volgens de bevolkingsprojecties bij vrouwen jonger dan 27 en ouder dan 37, terwijl voor de tussenliggende leeftijden de MIDAS-projectie een relatieve piek laat zien.

Deze afwijkingen ontstaan uiteraard door de vereisten van de aligeringstabellen inzake LIPRO-positie, die niet altijd consistent lijken met de projecties van de vruchtbaarheid. Zo daalt de vruchtbaarheid van 20-24-jarigen van 4,3 % in 2012 tot een minimum van 2,8 % in 2019, om dan in projectie weer te stijgen tot 4,0 % in 2030, en daarna praktisch constant te blijven (3,8 % in 2070). Volgens de projectie van de LIPRO-positie daalt het percentage moeders in dezelfde leeftijdsgroep eveneens in de periode 2012-2019 van 36,3 % tot 28,5 %, maar blijft daarna dalen – aan een langzamer tempo – tot 20,8 % in 2070. Deze laatste evolutie duwt de geboorten in MIDAS naar hogere leeftijden van de moeder, wat zorgt voor de piek tussen 25 en 35 jaar. In MIDAS staat de projectie van de inkomensverdeling en armoederisico's centraal, en deze worden gemeten op het niveau van het huishouden. Daarom is de overeenkomst met externe projecties van de samenstelling van huishoudens qua leeftijd en aantal kinderen belangrijker dan de exacte reproductie van de geprojecteerde vruchtbaarheid naar leeftijd.

Figuur 3 Vruchtbaarheid van vrouwen per leeftijd, gesimuleerd in MIDAS en volgens bevolkingsprojecties
proportie met geboorte per leeftijd



4.2.6. Divorce

De procedure *divorce* simuleert scheiding, inclusief feitelijke scheiding, van zowel gehuwde als geregistreerd samenwonende als niet-geregistreerd samenwonende koppels. Koppels zonder kinderen vallen uiteen in alleenstaanden; koppels met kinderen in éénoudergezinnen en alleenstaanden. De procedure is opgezet vanuit de vrouwen: de partners scheiden mee. Scheiding is gealigneerd aan de proportie vrouwen die per jaar scheiden of uiteengaan per leeftijdscategorie, zoals geobserveerd in de startdata.

De procedure *divorce* omvat de volgende stappen:

1. Bepaling kans op (feitelijke) scheiding voor vrouwen.
2. Identificatie van vrouwen die scheiden aan de hand van aligneringsdoelstellingen voor alleenstaanden en éénoudergezinnen; hun partners scheiden mee.
3. Voor koppels met kinderen: bepaling wie de kinderen krijgt
4. Aanpassing variabelen.

De kans op scheiding wordt afzonderlijk berekend voor gehuwden en samenwonenden, ook wordt onderscheid gemaakt tussen personen beneden 65 jaar en ouderen. Tabel 6 toont de aligneringstabel

voor de vrouwen tussen 15 en 64 jaar, die berekend is op basis van veranderingen in de burgerlijke staat tussen 2011 en 2012 in de startdata. Gegeven deze algemene proporties per leeftijdsgroep wordt de kans op scheiding verder gedifferentieerd aan de hand van logistische vergelijkingen, die geschat zijn op basis van de startdata. De startdata bevatten de burgerlijke staat zowel in 2011 als in 2012, zodat transitie van gehuwd naar gescheiden, en van (geregistreerd) samenwonend naar alleenstaand of gescheiden geïdentificeerd konden worden.¹⁰ De vergelijkingen werden afzonderlijk geschat voor deze twee transitie. Voor scheiding van gehuwden bevat deze de variabelen: leeftijd (lineair en als derde macht), het verschil in leeftijd tussen partners (met afzonderlijke variabelen voor een positief en een negatief verschil), onderwijsniveau, regio, aanwezigheid kind van 0 of 1 jaar, aantal kinderen beneden 6 jaar, aanwezigheid kinderen tussen 6 en 24 jaar, aanwezigheid boven 24 jaar en interactietermen voor de combinatie aanwezigheid kinderen tussen 6 en 24 jaar en onderwijsniveau. Voor scheiding van samenwonenden is de set van variabelen beperkter: leeftijd (lineair, gekwadraterd en als derde macht), het verschil in leeftijd tussen partners, onderwijsniveau, regio, aantal kinderen beneden 6 jaar, en aanwezigheid kinderen tussen 6 en 24 jaar. Voor de ouderen (65plus) wordt een uniforme kans van scheiding aangehouden van 0,3 % voor gehuwden, en van 4,3 % voor samenwonenden. Op basis van de aligeringstabel en de berekende kans op scheiding per vrouw, worden de vrouwen geïdentificeerd die gaan scheiden of uit elkaar gaan. Hun partners scheiden uiteraard mee.

Tabel 6 Aligeringstabel voor scheiding van gehuwden en uiteengaan van samenwonenden

	15	20	25	30	35	40	45	50	55	60
Scheiding (huwelijk)	0,6%	2,5%	2,4%	2,4%	2,4%	2,2%	1,8%	1,2%	0,7%	0,4%
Uiteengaan (samenwonen)	12,1%	12,1%	4,3%	4,3%	4,3%	4,3%	4,3%	4,3%	4,3%	4,3%

Noot: cijfers tonen het percentage gehuwde, resp. samenwonende vrouwen in jaar t dat in jaar t+1 gescheiden is, resp. is uiteengegaan.

In geval van scheiding met kinderen, worden alle kinderen aan één partner toegewezen (co-ouderschap wordt niet gesimuleerd). Het aantal vaders die hun kinderen behouden wordt afgeleid van het verschil tussen het aantal alleenstaande vaders met kinderen volgens de de aligeringstabel, en het aantal al aanwezige vaders in éénoudergezinnen. De selectie van deze vaders is verder geheel willekeurig.

De ouder die de kinderen behoudt blijft in hetzelfde huishouden leven, terwijl de andere partner naar een ander huishouden verhuist (dwz. een andere huishoudennummer krijgt toegewezen). In geval van kinderloze koppels verhuist per conventie altijd de man.

Tenslotte worden diverse variabelen aangepast aan de nieuwe situatie: variabelen met betrekking tot de partners, de burgerlijke staat, en eventueel worden voor kinderen de links met de vader, c. q. de moeder doorgeknipt.

¹⁰ Deze transitie op basis waarvan de vergelijkingen zijn geschat vallen niet volledig samen met de transitie die gesimuleerd worden. Deze laatste omvat ook het uiteengaan van niet-geregistreerd feitelijk samenwonende partners, en niet alleen de geregistreerd samenwonenden. Daarnaast valt de burgerlijke staat niet volledig samen met het al of niet samenwonen met een partner: sommige gescheiden personen wonen samen met een partner, bijvoorbeeld. We brengen in herinnering dat de kansen die met deze vergelijking worden berekend niet rechtstreeks de koppelvorming bepalen, maar alleen de rangorde vastleggen bij de selectie van personen voor koppelvorming. De genoemde inconsistenties lijken voor dat doel geen doorslaggevend probleem.

Er is geprobeerd om deze aligeringstabel af te leiden uit statistieken gepubliceerd door Statbel, maar dit is niet gelukt.

4.2.7. Partnership

De procedure *partnership* simuleert de koppelvorming tussen vrouwen en mannen. Deze procedure bepaalt voor elke niet gehuwde of samenwonende man en vrouw of ze in aanmerking komen voor huwelijk of samenwonen, en – indien dat zo is – koppelt beschikbare mannen en vrouwen aan elkaar door middel van een koppelingsproces (zie Dekkers et al., 2010, voor een nadere bespreking). Gekoppelde personen gaan samenwonen in hetzelfde huishouden als partners (samen met hun eventuele inwonende kinderen), en kunnen eventueel trouwen in hetzelfde jaar of later (procedure *marriage*).

De koppelvorming is gealigneerd aan het percentage vrouwelijke alleenstaanden, met of zonder kinderen, per leeftijdscategorie. Dit bepaalt het aantal te koppelen vrouwen, en dit aantal bepaalt op zijn beurt het aantal te koppelen mannen. Bij de selectie van de mannen wordt zoveel als mogelijk rekening gehouden met de aligneringstabel, maar deze kan niet altijd volledig gerespecteerd worden. Omdat kinderen alleen uit vrouwen geboren kunnen worden, en de partnersituatie een belangrijke factor is bij haar vruchtbaarheid, wordt bij de alignering van de koppelvorming prioriteit gegeven aan de vrouwen.

De procedure *partnership* omvat de volgende stappen:

1. Bepaling kans op koppelvorming voor vrouwen en mannen.
2. Identificatie van vrouwen die gekoppeld worden aan de hand van aligneringsdoelstellingen voor alleenstaanden en éénuoudergezinnen; identificatie van mannen die gekoppeld worden
3. Koppeling (*matching*) van mannen en vrouwen.
4. Aanpassing variabelen, waaronder linken van kinderen aan stiefouders.

Stap 1 Bepaling kans op koppelvorming voor vrouwen en mannen

De berekening van de kans op koppelvorming gebeurt afzonderlijk voor personen die nog nooit hebben samengewoond of getrouwd zijn geweest, en zij die gescheiden zijn. Deze kans wordt bepaald aan de hand van logistische vergelijkingen, die geschat zijn op de startdata. De startdata bevatten de burgerlijke staat zowel in 2011 als in 2012, zodat twee soorten transities geïdentificeerd konden worden: van alleenstaand naar gehuwd of geregistreerd samenwonend, en van gescheiden of verweerd naar gehuwd of geregistreerd samenwonend.¹¹ Voor de eerste transitie bevat de vergelijking de volgende variabelen: leeftijd (lineair, kwadratisch en als derde macht), geslacht, onderwijsniveau, regio, het aantal kinderen beneden 6 jaar, aanwezigheid kinderen tussen 6 en 24 jaar, en interactietermen tussen enerzijds geslacht en anderzijds leeftijd, het aantal kinderen beneden 6 jaar en de aanwezigheid van kinderen tussen 6 en 24 jaar. Daarenboven wordt aan vrouwen die juist moeder zijn geworden (*first_time_mother*), een zeer hoge kans op koppelvorming toegekend. Uit de data bleek dat

¹¹ Deze transities op basis waarvan de vergelijkingen zijn geschat vallen niet volledig samen met de transities die gesimuleerd worden. Deze laatste omvat ook de transitie naar niet-geregistreerd feitelijk samenwonen. Daarnaast valt de burgerlijke staat niet volledig samen met het al of niet samenwonen met een partner: sommige gescheiden personen wonen samen met een partner, bijvoorbeeld. We brengen in herinnering dat de kansen die met deze vergelijking worden berekend niet rechte lijnen de koppelvorming bepalen, maar alleen de rangorde vastleggen bij de selectie van personen voor koppelvorming. De genoemde inconsistenties lijken voor dat doel geen doorslaggevend probleem.

koppelvorming vaak in hetzelfde jaar gebeurt waarin een vrouw voor de eerste keer een kind krijgt. Zonder deze toevoeging kwamen er in de simulatieresultaten te veel alleenstaande moeders voor, vergeleken met de observaties.

Voor de kans op het vinden van een nieuwe partner voor personen die gescheiden of verweduwd zijn, werd een vergelijking met de volgende variabelen gebruikt: leeftijd (lineair), geslacht, onderwijsniveau, het aantal kinderen beneden 6 jaar, aanwezigheid kinderen tussen 6 en 24 jaar, en interactietermen tussen enerzijds geslacht en anderzijds leeftijd, het aantal kinderen beneden 6 jaar en de aanwezigheid van kinderen tussen 6 en 24 jaar.

Stap 2: Identificatie van vrouwen en mannen die gekoppeld worden

Het aantal vrouwen met kinderen die een (nieuwe) partner moeten krijgen wordt afgeleid uit het verschil tussen het aantal éénoudergezinnen volgens de aligneringstabel, en het aantal éénoudergezinnen dat al aanwezig is. Analoog wordt het aantal vrouwen zonder kinderen die een (nieuwe) partner moeten krijgen afgeleid uit het verschil tussen het aantal alleenstaanden volgens de aligneringstabel, en het aantal alleenstaande vrouwen dat al aanwezig is. Op basis van deze aantallen en de berekende kans op koppelvorming per vrouw, worden de vrouwen bepaald die een partner zullen krijgen.

De mannen die een (nieuwe) partner moeten krijgen worden in eerste instantie op dezelfde manier geselecteerd als de vrouwen. Als dit aantal groter is dan het aantal te koppelen vrouwen zullen sommige van deze mannen toch niet gekoppeld worden. Als er minder te koppelen mannen dan vrouwen zijn, dan wordt een extra aantal mannen geselecteerd, zodat alle te koppelen vrouwen inderdaad gekoppeld kunnen worden. Deze mannen worden willekeurig geselecteerd uit alle mannen van 16 tot 90 jaar zonder partner die geen kind of 'overig' zijn en geen kinderen hebben, en dit afzonderlijk voor oudere mannen en mannen beneden 65 jaar. Dit betekent dat voor de mannen de aligneringsaantallen niet altijd gerespecteerd worden, omdat, zoals boven aangegeven, prioriteit wordt gegeven aan de alignering voor de vrouwen.

Stap 3: Koppeling (*matching*) van mannen en vrouwen.

De simulatie van de koppeling van vrouwen en mannen op individueel niveau (de *matching*) heeft als voornaamste doel de huidige patronen van leeftijdsverschillen tussen partners, en van homogamie naargelang onderwijsniveau en migratiestatus te reproduceren – beide zijn belangrijke variabelen voor de arbeidsmarktstatus. Om de coëfficiënten te bepalen van een *matching*-vergelijking die dit zou bewerkstelligen, werd de volgende analyse uitgevoerd. Uit de startdata werden toevalsgewijs 2000 vrouwen geselecteerd tussen 18 en 59 jaar, met een partner tussen 15 en 74 jaar. Binnen deze steekproef werden alle partners aan alle geselecteerde vrouwen gekoppeld (zodat een databestand van 4M gevallen resulteerde). Voor iedere vrouw werd de dichotome afhankelijke variabele op 1 gesteld voor de werkelijke partner, en op 0 voor alle anderen. De onafhankelijke variabelen waren de leeftijd van de vrouw, het leeftijdsverschil met de man (als dit positief was), het leeftijdsverschil met de man (als dit negatief was), interactietermen tussen de leeftijd van de vrouw en de variabelen inzake leeftijdsverschil, iedere combinatie van het onderwijsniveau van de vrouw (drie categorieën, zie hoofdstuk 2) met dat

van de man, en iedere combinatie van de migratiestatus van de vrouw met dat van de man. Hierbij was migratiestatus tot twee categorieën gereduceerd: enerzijds geboren in België en minstens één ouder geboren in België (dus inclusief ‘gemengde’ tweede-generatie migranten), en anderzijds geboren buiten België of beide ouders geboren buiten België (dus inclusief ‘pure’ tweede-generatie migranten). Deze specificatie laat toe dat het leeftijdsverschil tussen mannen en vrouwen een ander effect op de partnerkeuze heeft, naargelang dit positief of negatief is.

Op basis van deze vergelijking wordt een score berekend. Voor iedere vrouw wordt de (overblijvende) man genomen met de hoogste score. De te koppelen vrouwen worden hierbij overlopen van jong naar oud. Aan de vergelijking om de score te berekenen werden nog twee variabelen toegevoegd, die aangeven dat de vrouw geen kinderen heeft, en de man wel (met sterk negatieve coëfficiënt), of juist andersom (met sterk positieve coëfficiënt). Hierdoor wordt voorkomen dat vrouwen zonder kinderen worden gekoppeld aan een man met kinderen, terwijl mannen zonder kinderen voorrang krijgen bij de koppeling van vrouwen met kinderen. Deze ingreep is nuttig om de alignerings percentages voor gehuwde of samenwonende vrouwen en mannen met kinderen te behalen.

Stap 4: Aanpassing variabelen

Tenslotte worden voor de gekoppelde vrouwen en mannen de variabelen inzake partnerstatus, burgerlijke staat en duur van de relatie aangepast. Kinderen van gekoppelde personen worden aan hun nieuwe vader of moeder toegewezen als stiefkinderen. Aan de nieuwe partners, hun eventuele kinderen, en ook aan ‘overigen’ die eventueel in hetzelfde huishouden leefden, wordt een nieuwe huishouden toegewezen.

4.2.8. Marriage

Koppels die samenwonen kunnen huwen, ofwel in het jaar dat zij gekoppeld worden, ofwel in de jaren daarna. Het initiatief voor een huwelijk gaat in MIDAS uit van de vrouw. Om complicaties in de berekening van overlevingspensioenen te vermijden, kunnen weduwen en weduwnaars niet opnieuw trouwen (zij kunnen wel ongehuwd samenwonen). Koppels die niet huwen, blijven ongehuwd samenwonen. Er wordt geen verschil gemaakt tussen geregistreerd en ongeregistreerd samenwonen.

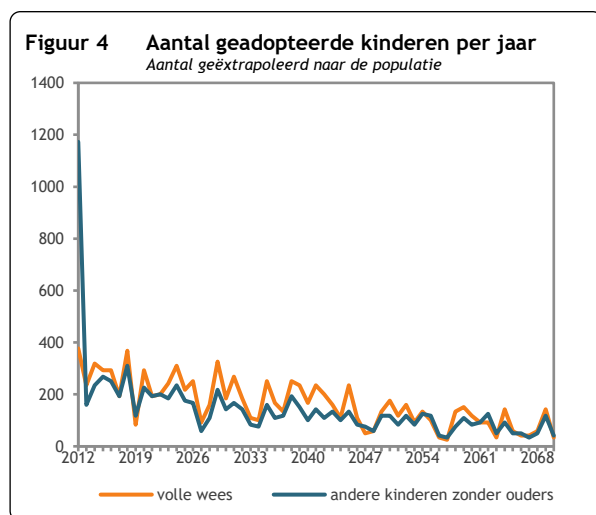
Het aantal vrouwen dat gaat trouwen wordt afgeleid uit het verschil tussen het percentage volgens de aligneringstabel en het percentage in de uitgangssituatie (nadat de procedures *divorce* en *partnership* hebben gelopen), en dit afzonderlijk voor vrouwen met en zonder kinderen. Een score die de kans op huwen aangeeft wordt berekend met een logistische vergelijking, die geschat is op basis van de startdata. Afhankelijke variabele was een variabele die aangaf of een vrouw gehuwd was, binnen de steekproef van gehuwde en samenwonende vrouwen jonger dan 40 jaar.¹² Onafhankelijke variabelen waren regio, opleidingsniveau en aantal kinderen.

¹² Een logistische regressie met als afhankelijke variabele de transitie van samenwonend naar gehuwd leverde geen significante resultaten op.

Als een vrouw wordt geselecteerd om te trouwen, dan trouwt uiteraard ook haar partner. Voor beiden worden de variabelen die partnerstatus en burgerlijke staat beschrijven, in dat geval aangepast, evenals de LIPRO-categorie van henzelf en hun eventuele kinderen.

4.2.9. Orphan_adoption

Op enig moment in de simulatie kunnen er door het overlijden van alleenstaande ouders huishoudens ontstaan die enkel uit minderjarige kinderen bestaan. Daarnaast komen er enkele huishoudens met



alleen minderjarigen voor in de startdata, en ontstaan deze ook in de verdere simulaties. De procedure *Orphan_adoption* voorkomt deze situaties door minderjarigen zonder ouders op willekeurige wijze toe te wijzen aan ('laten adopteren door') moeders met eigen kinderen die samenwonen of gehuwd zijn. Figuur 4 toont dat het (na 2012) om een zeer klein aantal gevallen gaat, zodat deze procedure andere verdelingen niet verstoort.

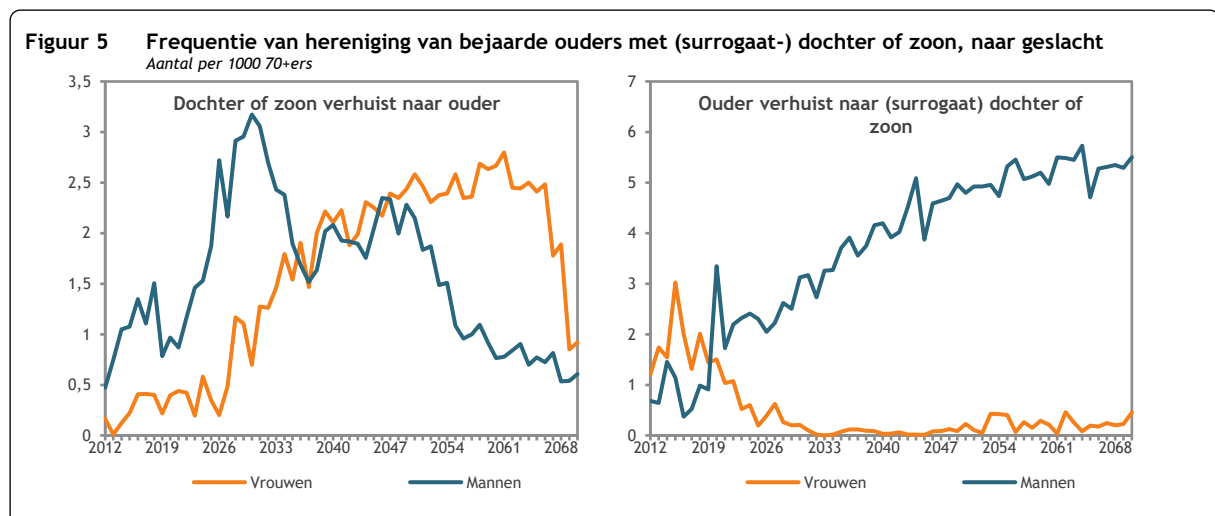
4.2.10. Caring_daughter

Tot dusver in het model konden enkel kinderen die op enig moment niét uit het huis van hun ouders vertrekken, door de procedure *leave_the_nest* de status 'overig' krijgen. Maar uit de startdata bleek dat er eveneens onder hoogbejaarden, en vooral onder de vrouwen, een niet verwaarloosbaar aantal de status 'overig' had. Dit blijft ook in de toekomst het geval volgens de LIPRO-projecties die onze modellering als uitgangspunt neemt. De proportie stijgt met de leeftijd: van 3,4 % bij de 70-74 jarige vrouwen tot 7,6 % bij de vrouwen van 90 en ouder (voor de mannen zijn deze percentages 1,9 % en 4,0 %). Daarnaast vinden we bij de oudere leeftijdsgroepen ook een aantal éénoudergezinnen, dat eveneens toeneemt met de leeftijd: van 4,0 % bij de 70-74 jarige vrouwen tot 6,7 % bij vrouwen tussen 85 en 89 jaar, en 5,2 % bij de vrouwen van 90 en ouder (voor de mannen zijn deze percentages 1,3 %, 3,6 % en 4,2 %). Wellicht gaat het in beide situaties vooral om ouderen die wegens zorgbehoefte bij een dochter, zoon of andere familieleden zijn gaan wonen, ofwel, om dezelfde reden, kinderen die terug bij hun moeder of vader zijn gaan wonen. Een logistische regressie met als afhankelijke variabele de transitie van samenwonend naar gehuwd leverde geen significante resultaten op.

Om het aantal éénoudergezinnen bij hoogbejaarden te laten overeenkomen met de aligeringpercentages, selecteren we alleenstaande ouderen met een volwassen alleenstaande dochter of zoon (die dus een eigen, eenpersoonshuishouden hebben). Een aantal van deze dochters en zonen worden herenigd in hetzelfde huishouden met hun moeder of vader, zodat het percentage éénoudergezinnen bij de hoogbejaarden volgens de aligeringstabel wordt bereikt. In geval van meerdere alleenstaande dochters of zonen van dezelfde ouder, wordt voorrang gegeven aan dochters

en het oudste kind. In geval twee ouders – beide alleenstaand na scheiding – in aanmerking komen voor dezelfde dochter of zoon, wordt voorrang gegeven aan de moeder. Merk op dat alleen ouders en dochters of zonen herenigd kunnen worden als zij in de startdata in 2011 nog in hetzelfde huishouden woonden, zodat zij gedurende de projectieperiode (vanaf 2012) het ouderlijk huis hebben verlaten. Voor de groep die in de startdata al onafhankelijk van hun ouders wonen, is de familiale relatie namelijk onbekend. Dit betekent dat deze procedure vooral werkzaam is in de tweede helft van de projectieperiode. Door deze procedure verandert de LIPRO-positie van de dochter of zoon van alleenstaande naar kind. Figuur 5 (links) toont dat dit een weinig voorkomende transitie is, die na het midden van de 2030s vooral vrouwen betreft.

De omgekeerde beweging, waarbij een ouder in het huishouden van de dochter of zoon gaat wonen, komt vaker voor (zie Figuur 5 rechts), en betreft vooral mannen. Deze transitie wordt gealigneerd aan het percentage ‘overigen’ onder de 70plussers. Ook hiervoor komen alleen alleenstaande ouderen in aanmerking. Als er geen eigen dochter of zoon beschikbaar of bekend is, dan wordt een willekeurige andere vrouw of man tussen 40 en 64 jaar geselecteerd. De koppeling van deze ‘surrogaat’-dochters of -zonen aan ‘hulpbehoevende’ ouderen gebeurt op basis van een score berekend uit het leeftijdsverschil, die zijn maximum bereikt bij een verschil van 28 jaar. De ouder krijgt in het huishouden van de (surrogaat-) dochter of zoon de LIPRO-positie ‘overig’, terwijl de LIPRO-posities van personen in het ontvangende huishouden niet veranderen. De verschillende trends, inclusief die naar geslacht, worden gestuurd door de aligneringstabellen.



4.2.11. Pseudo_lipro

De procedures *dead* en *orphan_adoption* vormen samen één blok van procedures, evenals *first_time_mother* en *baby_born*; en ook *one_more_child* en *baby_born*. Bij het begin van de demografische module, na de procedure *dead*, en na iedere procedure, of blok van procedures, wordt de procedure *pseudo_lipro* uitgevoerd, die de LIPRO-positie van ieder individu bepaalt, zodat deze correct is voor de volgende stap in het proces. Deze procedure vervult zo als het ware de functie van boekhouder van de gezinssituatie. De keuze om de LIPRO-positie opnieuw te bepalen na iedere demografische procedure, en dit in een afzonderlijke procedure, is ingegeven door het feit dat verandering in de situatie van één lid van het huishouden gevolgen kan hebben voor de posities van andere leden van het huishouden.

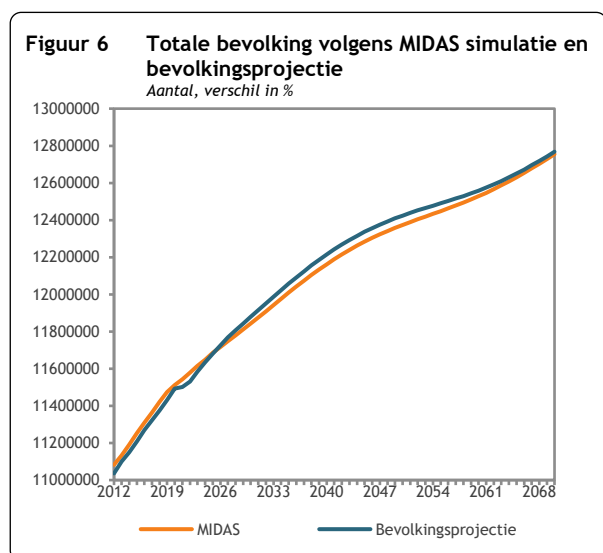
De LIPRO-positie wordt geüpdated op basis van het aantal personen in het huishouden, de variabele die aangeeft of een persoon een 'kind' is, de partner-situatie (geen partner, gehuwde partner, samenwonende partner), en de variabelen die de links tussen ouders en kinderen vastleggen.

4.3. Resultaten

In deze sectie tonen we allereerst de MIDAS projecties van de gehele bevolking. Vervolgens komen de projecties van de LIPRO-positie aan de orde. We vergeleken steeds met de bevolkingsprojecties, per brede leeftijdsgroep (0-14, 15-29, 30-49, 50-64, 65-79, 80+) en geslacht. Tenslotte bestuderen we de gesimuleerde homogamie van koppels naar leeftijd, onderwijsniveau en migrant-status, en de gradiënten van onderwijsniveau naargelang LIPRO-positie.

4.3.1. Projectie van de bevolking naar leeftijd en geslacht

Figuur 6 toont dat de gesimuleerde totale bevolking volgens MIDAS praktisch gelijk is aan die volgens

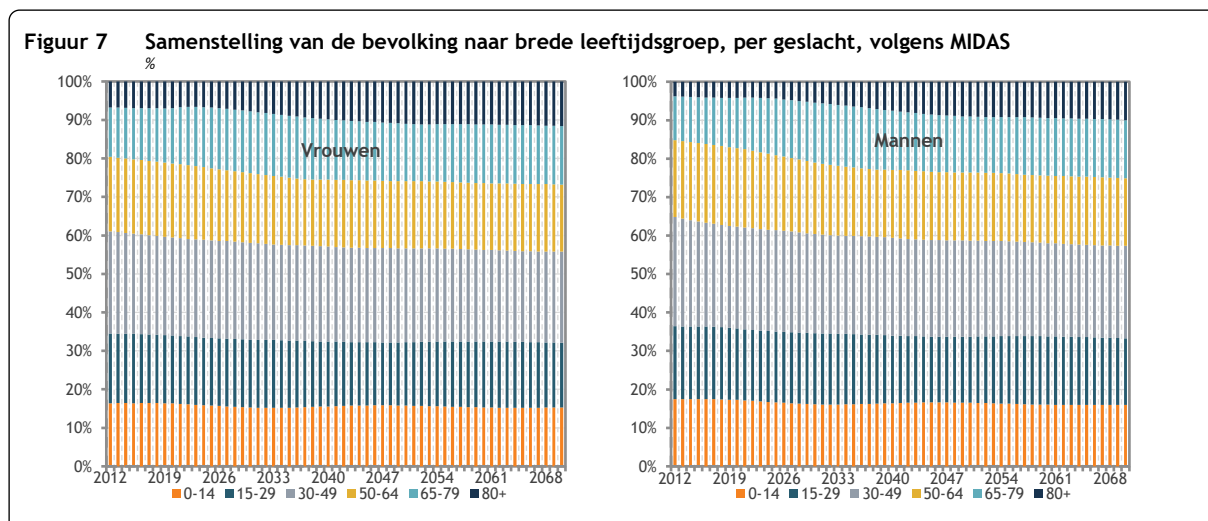


de bevolkingsprojectie. Het maximale verschil is 0,4 %. Geboorte, sterfte en migratie in MIDAS worden gealigneerd aan de bevolkingsprojectie, zodat de kleine verschillen te wijten zijn aan steekproefafwijkingen.¹³

De gesimuleerde samenstelling van de bevolking (Figuur 7) toont de vergrijzing: tot rond 2030 blijft de proportie 65-79 jarigen stijgen, terwijl vanaf dan de proportie 80ers sterk toeneemt tot het midden van de jaren 2050. Het percentage van de bevolking tussen 30 en 49 jaar, en tussen 50 en 64 jaar daalt, terwijl de proportie jongeren (0-14 en 15-29 jaar) nauwelijks verandert. Deze proporties

zijn eveneens erg gelijkaardig aan die welke volgen uit de bevolkingsprojecties; het maximale verschil (in absolute waarde) over alle projectiejaren en leeftijdsgroepen is 0,7 procentpunt.

¹³ Niet zozeer in de originele steekproef uit het Datawarehouse, maar in de selectieve en geëxpandeerde steekproef die in MIDAS wordt gebruikt. Omdat hiervoor de gewichten naar gehele getallen moeten worden afgerond, ontstaan kleine afwijkingen in de leeftijdsopbouw. Zie Annex 1 voor meer details.

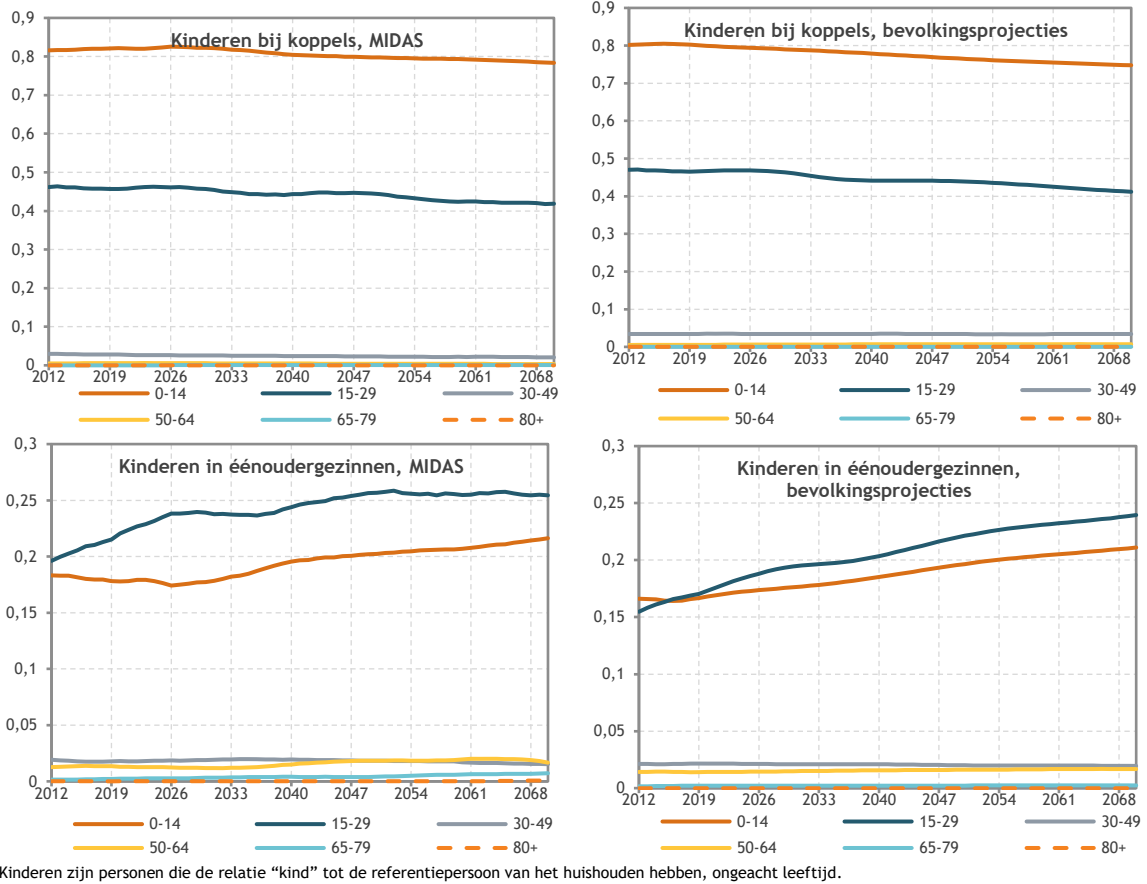


4.3.2. Projectie van de bevolking naar LIPRO-positie

We bespreken achtereenvolgens de proporties van de bevolking in de verschillende LIPRO-posities: kinderen bij (gehuwde of samenwonende) koppels, kinderen in éénoudergezinnen, alleenstaanden, (samenwonende of gehuwde) koppels zonder kinderen, (samenwonende of gehuwde) koppels met kinderen, alleenstaande ouders, en overigen.

Figuur 8 toont dat bij de 0-14 jarigen (in deze leeftijdsgroep heeft praktisch iedereen de status 'kind'), de proportie die bij twee ouders leeft (inclusief eventuele stiefouders) langzaam daalt van 81 % in 2011 tot 76 % in 2070, terwijl tegelijk de proportie van kinderen in éénoudergezinnen stijgt van 20 % naar 24 %. Omdat in MIDAS een aantal 'overigen' zijn geherdefinieerd als 'kinderen bij éénoudergezinnen' (zie sectie 4.1) liggen de betreffende percentages constant ongeveer 2 à 3 procentpunt hoger dan volgens de bevolkingsprojecties, maar qua trend is er nauwelijks verschil. Een gelijkaardige evolutie zien we voor de groep van 16 tot 29 jaar, hoewel in deze groep ruim één op drie het ouderlijk huis al heeft verlaten. Het verschil tussen de percentages van kinderen in éénoudergezinnen tussen MIDAS en de bevolkingsprojecties is steeds ongeveer 5 procentpunt. Bij de 30-49 jarigen woont in 2011 nog ongeveer 5 % bij moeder en/of vader, dalend naar 3 % in 2070. Bij de 50-64 jarigen is dit percentage geslonken tot ongeveer 2 %, en de grote meerderheid van hen vormt een éénoudergezin. Onder 65+ers komt de LIPRO-positie van 'kind' nauwelijks voor, hoewel dit in MIDAS een beperkt maar groeiend fenomeen is bij de 65-79 jarigen, dat bijna 1 % bereikt in 2070. Rond 65 % van de personen van 30 jaar en ouder in de positie 'kind' is een man.

Figuur 8 Proportie kinderen naar leeftijdsgroep, bij koppels en in éénoudergezinnen, volgens MIDAS en volgens de bevolkingsprojecties

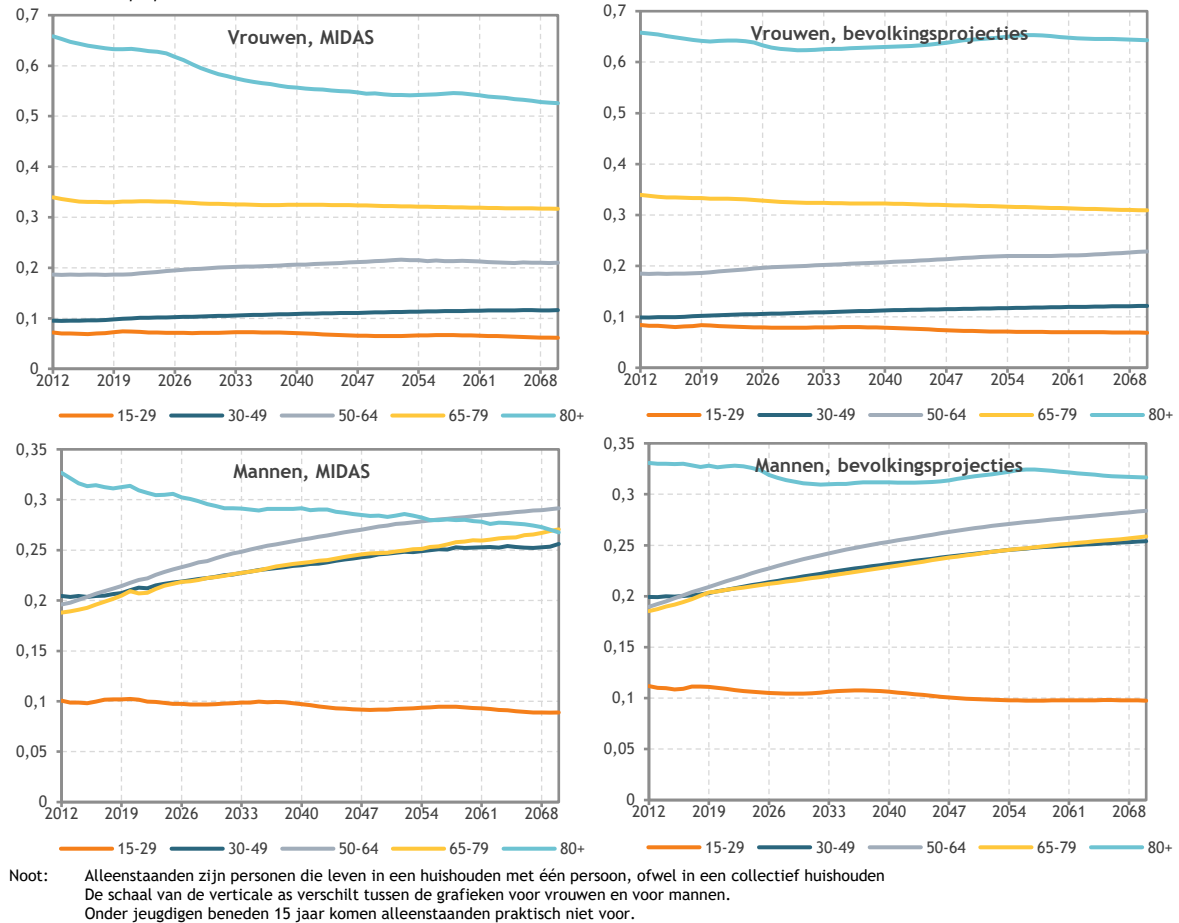


Kinderen zijn personen die de relatie "kind" tot de referentiepersoon van het huishouden hebben, ongeacht leeftijd.

Voor de vrouwen en mannen beneden 80 jaar worden de proporties alleenstaanden volgens de bevolkingsprojectie door MIDAS perfect gereproduceerd (Figuur 9), vanzelfsprekend dankzij de aligering. Deze proportie daalt licht bij de 15-29 jarigen en de 65-79 jarigen, en stijgt licht in de leeftijdsgroepen tussen 30-49 en 50-64 jaar. Voor de 80+ers projecteert MIDAS een daling van de proporties alleenstaanden, zowel bij vrouwen als bij mannen, terwijl deze volgens de bevolkingsprojecties zouden fluctueren, maar per saldo in 2070 nauwelijks lager zouden zijn dan in 2011. Zoals we hieronder zullen zien, stijgt in MIDAS tegelijk de proportie gehuwde koppels onder de 80+ers. De reden hiervoor is de stijgende levensverwachting, bij mannen meer dan bij vrouwen, waardoor er minder weduwen en weduwnaars zullen zijn. Om bij de hoogbejaarden de proporties alleenstaanden gelijk te houden met de bevolkingsprojecties zouden we moeten aannemen dat deze leeftijdsgroep veel vaker zou scheiden dan nu het geval is (ongehuwd samenwonen komt bij hen nauwelijks voor). Dit lijkt weinig realistisch.¹⁴ Het kan eigenaardig, of zelfs ongeloofwaardig, lijken dat tegen 2070 bij mannen de proportie alleenstaanden onder 80plussers lager is dan in de leeftijdsgroep 65-79 jaar. De stijgende proportie alleenstaanden bij de laatsten doet zich echter vooral voor bij de 65-69 jarigen. Tussen 75-79 jarige en 80-84 jarige mannen is er na 2060 nauwelijks verschil in het percentage alleenstaanden: beide 22 %.

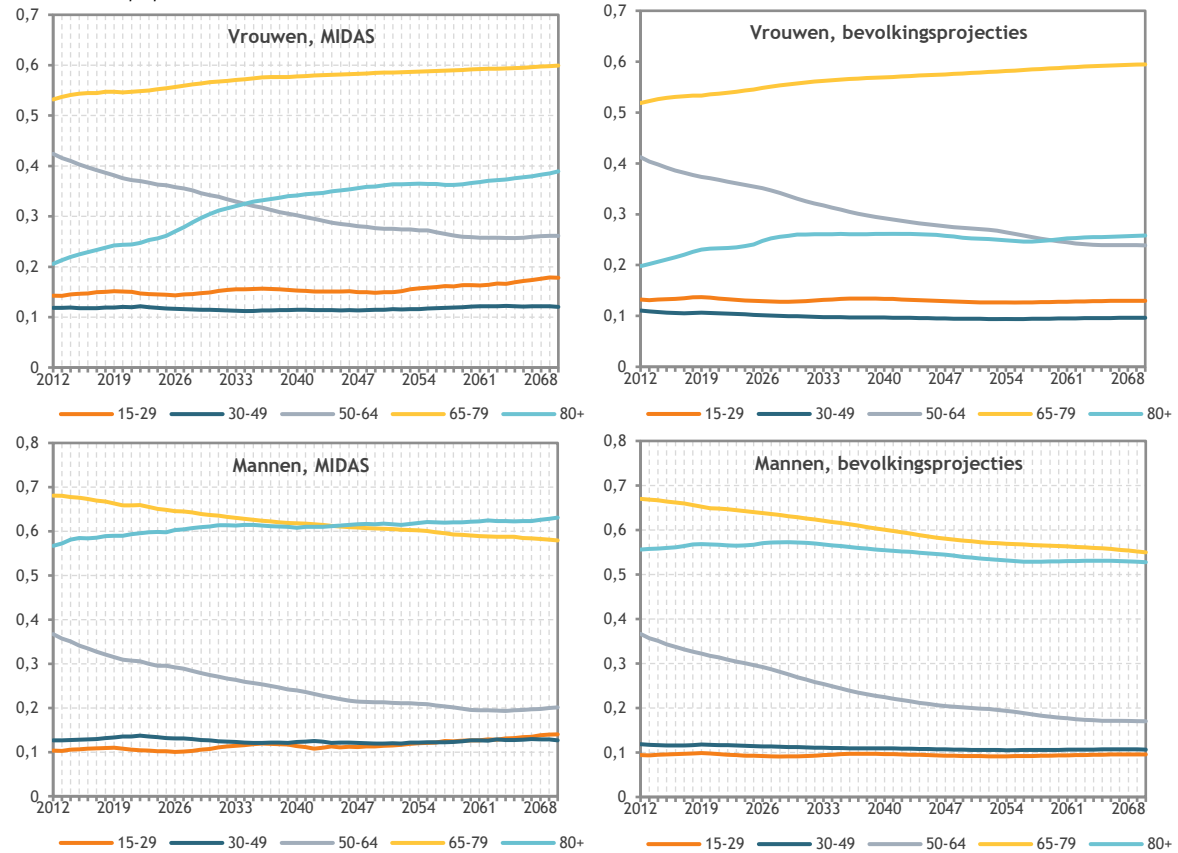
¹⁴ Zoals in sectie 3 uiteengezet, is de kans op sterfte lager voor personen in koppels dan voor alleenstaanden. Verwijdering van dit effect doet de proporties alleenstaanden onder 80+ers echter nauwelijks stijgen.

Figuur 9 Proportie alleenstaanden per leeftijdsgroep en geslacht, volgens MIDAS en volgens de bevolkingsprojecties



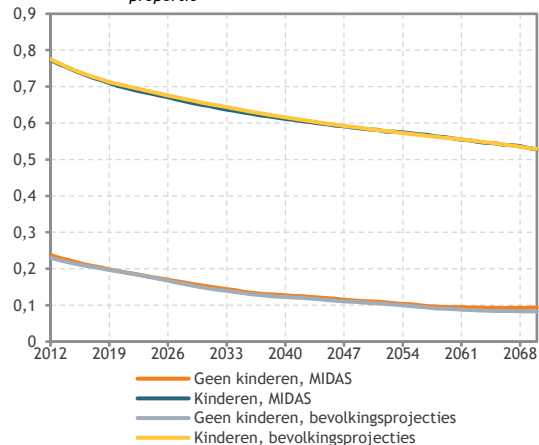
Ook voor de koppels zonder kinderen (gehuwd of samenwonend; Figuur 10) is de overeenkomst tussen de MIDAS simulaties en de bevolkingsprojecties zeer sterk voor de leeftijdsgroepen tot 80 jaar, zowel voor vrouwen als voor mannen. Voor de 80+ers duidt de MIDAS projectie echter op een stijging van de proporties mensen levend in koppelverband, vooral bij de vrouwen. Volgens de bevolkingsprojecties is deze stijging afwezig bij de mannen, en veel minder sterk en beperkt tot de periode 2011-2030 bij de vrouwen. Deze verschillen bij de 80ers zijn het spiegelbeeld van die voor de proporties alleenstaanden in dezelfde leeftijdsgroep, en ontstaan om dezelfde reden: de stijging van de levensverwachting, vooral voor mannen.

Figuur 10 Proportie koppels zonder kinderen (gehuwd of samenwonend) per leeftijdsgroep en geslacht, volgens MIDAS en volgens de bevolkingsprojecties



Voor de simulatie door MIDAS van de proporties koppels mét kinderen geldt hetzelfde als voor die van de koppels zonder kinderen: dankzij de aligeringen komen deze sterk overeen met de

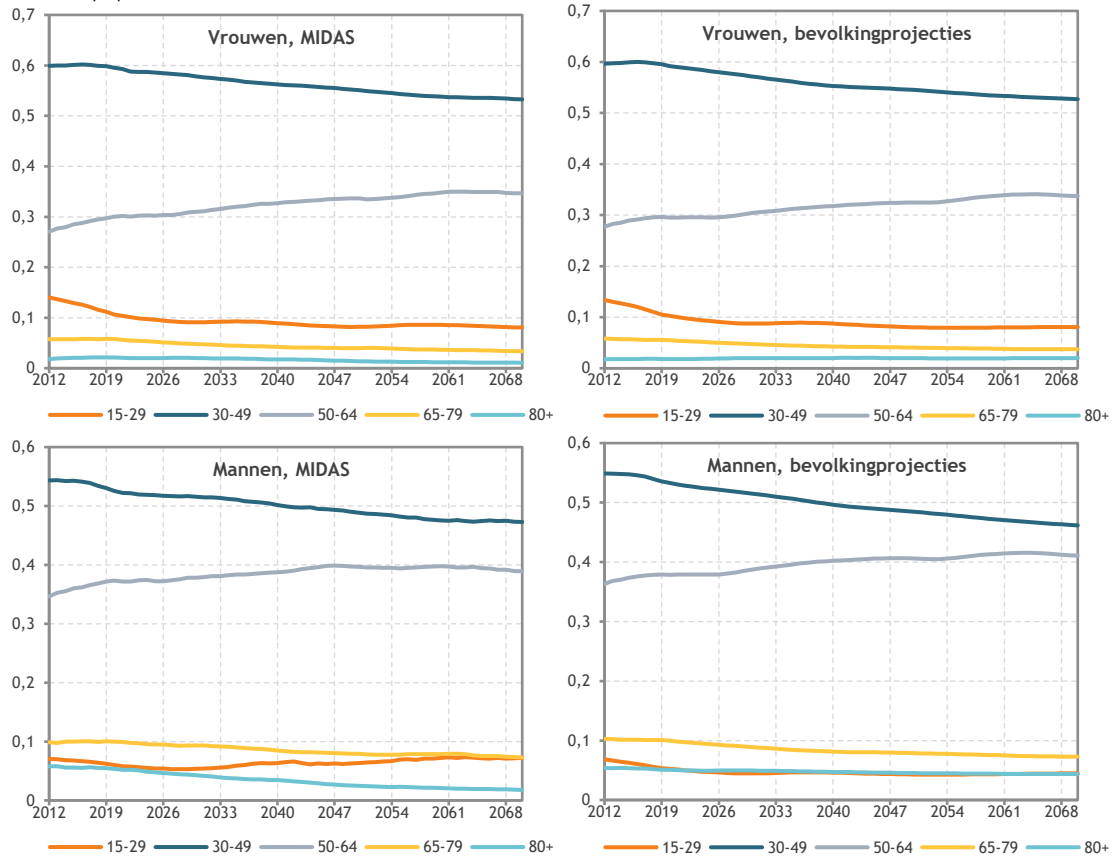
Figuur 11 Proportie gehuwden, onder vrouwen 15-64 met zonder kinderen, volgens MIDAS en bevolkingsprojectie



bevolkingsprojectie, volledig voor vrouwen, iets minder goed voor de mannen (Figuur 12). Onder de 30-49 jarige vrouwen daalt deze proportie, terwijl deze juist stijgt bij de 50-64 jarigen, mogelijk omdat vrouwen hun kinderen op latere leeftijd gaan krijgen. Merk op dat ook onder ouderen (65+, en zelfs 80+) er een beperkt aantal koppels met kinderen voorkomen. Bij de mannen zien we in de tweede helft van de projectieperiode verschillen ontstaan tussen de MIDAS projectie en de bevolkingsprojecties, die echter beperkt blijven tot maximaal 3 procentpunt. Dit is te wijten aan het feit dat in MIDAS het krijgen van kinderen gealigneerd is op de LIPRO-positie van vrouwen.

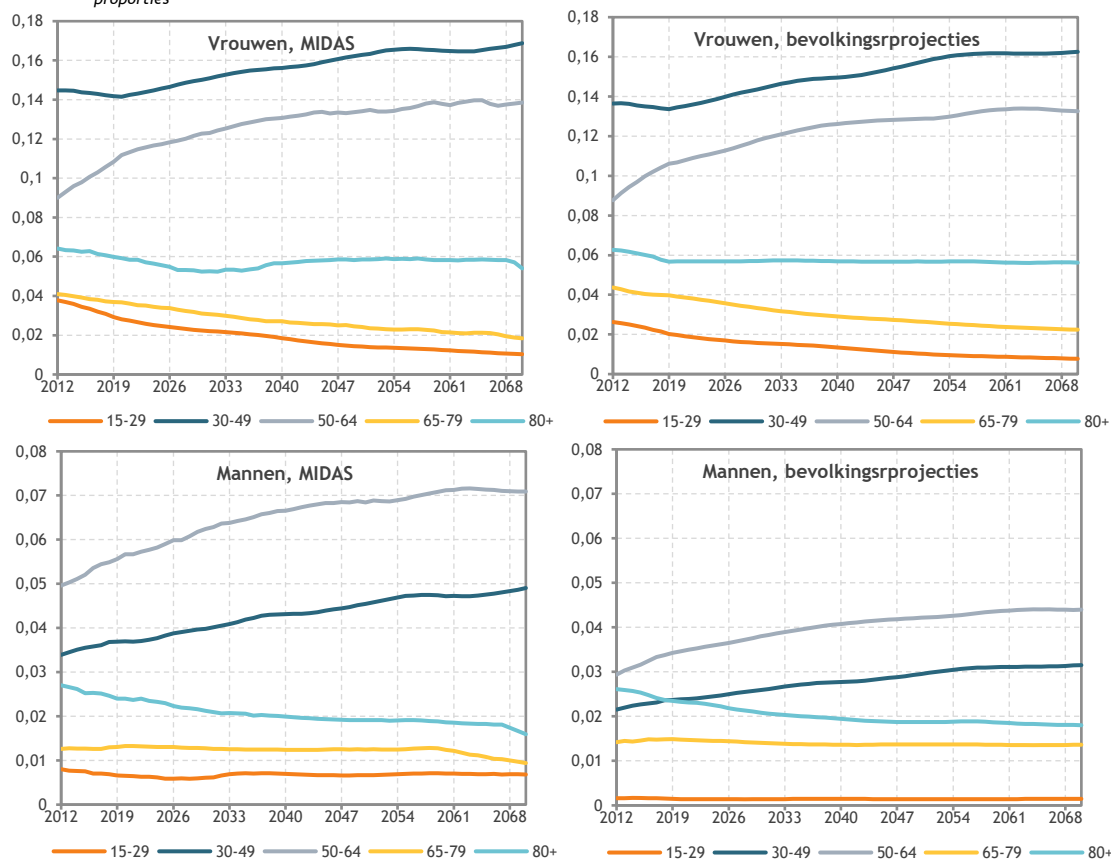
Figuur 11 toont dat het percentage gehuwden onder de vrouwen op actieve leeftijd sterk daalt gedurende de projectieperiode, zowel bij de vrouwen met als zonder kinderen. De MIDAS simulatie volgt hierin nauwkeurig de bevolkingsprojectie.

Figuur 12 Proportie koppels met kinderen (gehuwd of samenwonend) per leeftijdsgroep en geslacht, volgens MIDAS en volgens de bevolkingsprojecties.



De proporties alleenstaande ouders geprojecteerd door MIDAS liggen steeds hoger dan de proporties volgens de bevolkingsprojecties, vanwege de herdefiniëring van de LIPRO-positie in MIDAS (zie sectie 4.1). De geprojecteerde trends komen echter zeer goed overeen. Dit speelt vooral bij de mannen, ook al blijven de proporties alleenstaande vaders altijd relatief laag. Bij de vrouwen zien we een daling van de proportie alleenstaande ouders bij de jongeren (15-29 jaar) en bij de jongere ouderen (65-79 jaar), terwijl deze stijgt bij de leeftijdsgroepen 30-49 en 50-64 jaar. Merk de niet verwaarloosbare proportie alleenstaande ouders op bij de 80+ers (6 % bij de vrouwen, rond 2 % bij de mannen), dat duidelijk hoger is dan bij de 65-79 jarigen. Zoals hierboven al gesuggereerd, gaat het hierbij wellicht om ouders die vanwege zorgbehoeften bij hun kinderen zijn gaan wonen, of, omgekeerd, kinderen die bij hun ouders zijn ingetrokken.

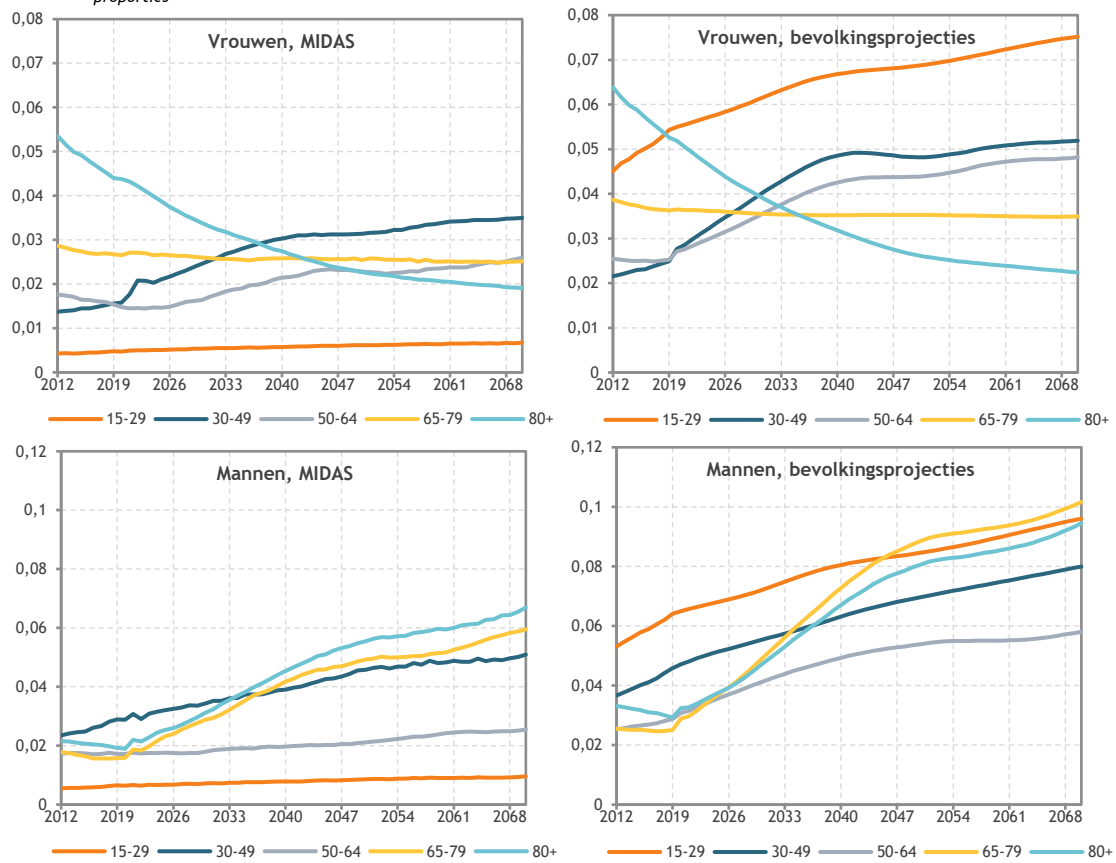
Figuur 13 Proportie alleenstaande ouders per leeftijdsgroep en geslacht, volgens MIDAS en volgens de bevolkingsprojecties



Noot: De definitie van alleenstaande ouder omvat iedere persoon zonder partner die in hetzelfde huishouden leeft als één of meer (stief)kinderen, ongeacht de leeftijd van die kinderen.

Zoals in sectie 4.1 uitgelegd, impliceert de herdefiniëring van de LIPRO-positie in MIDAS dat de proporties ‘overigen’ veel lager zijn dan volgens de bevolkingsprojecties. Figuur 14 toont dat dit zo blijft gedurende de gehele projectieperiode. De trends van deze proporties volgens de MIDAS simulaties komen wel sterk overeen met die volgens de bevolkingsprojecties. Opvallend is de halvering tussen 2011 en 2050 van de proportie ‘overigen’ onder de vrouwen van 80 jaar en ouder, terwijl deze proportie bij haar mannelijke leeftijdsgenoten juist verdrievoudigt gedurende de projectieperiode. Bij de mannen is er overigens een stijging van de proportie ‘overigen’ in alle leeftijdsgroepen, bij de vrouwen alleen beneden 65 jaar.

Figuur 14 Proportie ‘overigen’ per leeftijdsgroep en geslacht, volgens MIDAS en volgens de bevolkingsprojecties



Noot: "Overigen" zijn personen die niet vallen onder één van de voorgaande lipro-posities. Zie sectie 4.1.

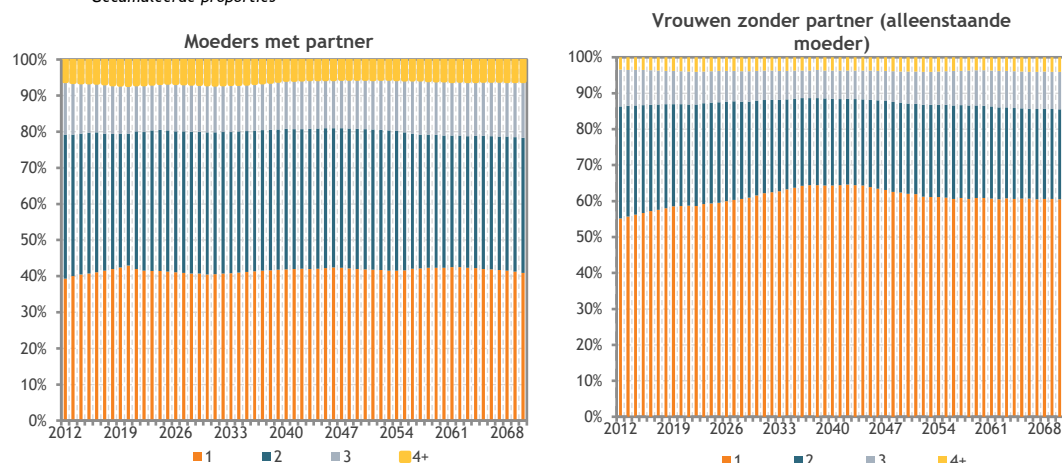
4.3.3. Andere aspecten van de projectie van demografische variabelen

Aantal kinderen per vrouw

Zoals uiteengezet in sectie 4.2.5 proberen we de verdeling van het aantal kinderen per vrouw zoveel mogelijk gelijk te houden aan die in de startdata, binnen de beperkingen opgelegd door enerzijds het aantal geboorten per jaar, en anderzijds het aantal moeders bepaald door de LIPRO-posities. Zoals blijkt uit Figuur 15 lukt dit redelijk goed voor vrouwen in koppels (twee-oudergezinnen), en iets minder voor de alleenstaande moeders (éénoudergezinnen), bij wie de proporties met één kind stijgt tot 2040. De proportie alleenstaande moeders met drie of meer kinderen is steeds lager dan voor gehuwde of samenwonende vrouwen.

Figuur 15 Aantal kinderen in huishouden per vrouw met kinderen (16-64), naargelang partnersituatie, volgens MIDAS

Gecumuleerde proporties



Homogamie

Zoals uiteengezet in sectie 4.2.7 is getracht om de homogamie naargelang leeftijd, onderwijsniveau en migrant-status binnen koppels (gehuwd en samenwonend) zoals vastgesteld in de startdata, in de gesimuleerde koppels te reproduceren. Tabel 7 toont de leeftijd van de man naargelang de leeftijd van de vrouw van jonge koppels in de startdata. Tabel 8 is de analoge tabel voor gesimuleerde koppels, in het jaar van de koppeling. Zoals de onderste rij van Tabel 8 aangeeft, vindt ruim de helft van alle koppelingen plaats voordat de vrouw 30 jaar wordt. In vergelijking met de startdata, is binnen deze groep de leeftijd van de man teveel geconcentreerd in dezelfde leeftijdscategorie als die van de vrouw, en zijn er te weinig koppels waar de man in een hogere leeftijdscategorie zit. Bij de vrouwen boven de 30 lijkt dan weer het omgekeerde te gebeuren: zij worden te weinig aan mannen in dezelfde leeftijdscategorie gekoppeld, en wat te vaak aan oudere mannen. Toch lijkt globaal de leeftijdsverdeling van de man naargelang die van de vrouw plausibel: de grootste percentages worden gevonden op de diagonaal, en in de cellen daar juist boven, en relatief weinig vrouwen zijn gekoppeld aan een jongere man.

Tabel 7 Leeftijd man naargelang leeftijd vrouw binnen koppels volgens startdata

Leeftijd man	Leeftijd vrouw			Totaal
	15-24	25-29	30-34	
15-24	29,5	2,8	0,4	5,7
25-29	45,1	42,0	6,9	26,0
30-34	16,0	37,0	42,4	36,4
35-39	5,2	11,7	33,7	21,1
40-44	2,3	3,9	11,0	7,0
45+	1,9	2,6	5,7	3,9
Totaal	100,0	100,0	100,0	100,0
n (unweighted)	4 187	10, 604	12 995	27 786

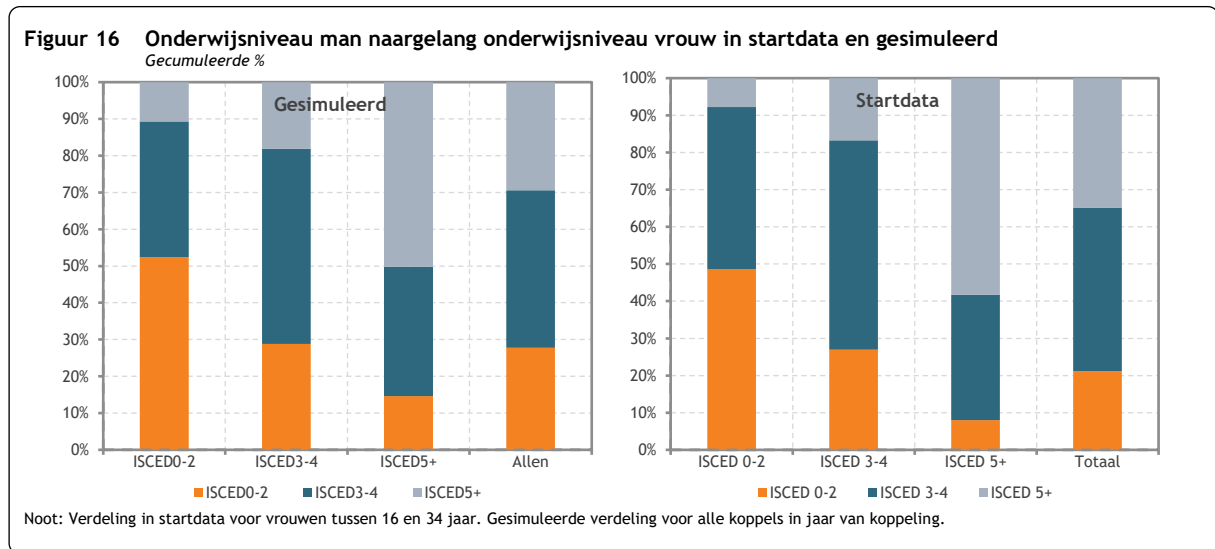
Bron: MIDAS

Tabel 8 Leeftijd man naargelang leeftijd vrouw in jaar van koppeling

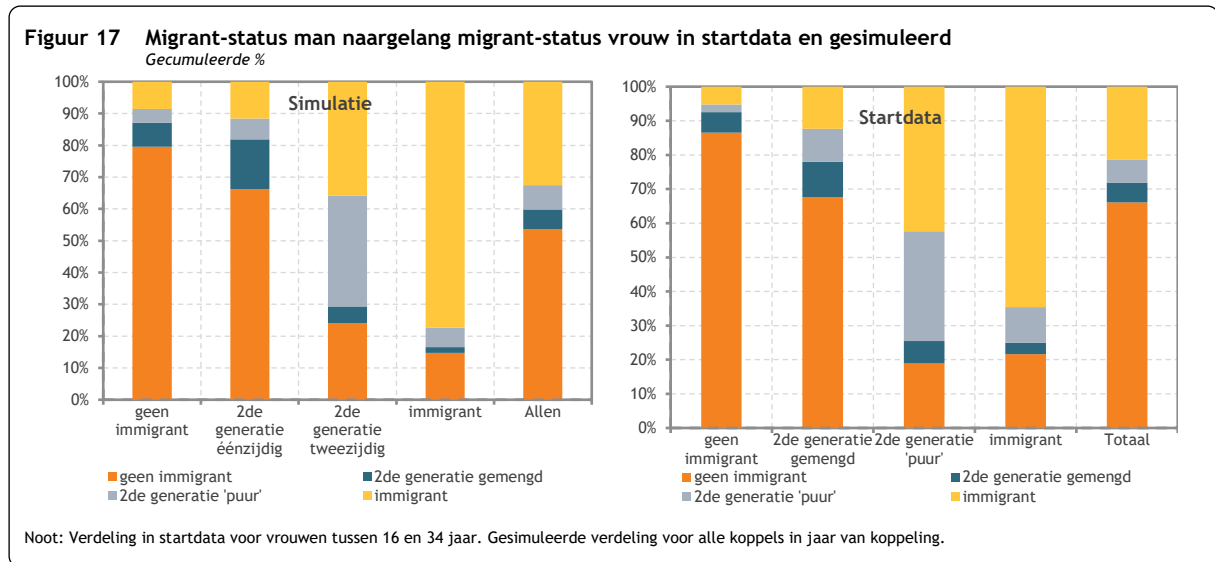
Leeftijd man	Leeftijd vrouw													
	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80+
15-19	66,3	5,6	1,9	0,6	0,6	0,2	0,3	0,8	0,3	0,1	0,0	0,0	2,0	0,6
20-24	26,6	36,1	3,1	0,5	0,1	0,0	0,1	0,1	0,0	0,0	0,0	0,0	0,0	0,0
25-29	6,6	53,7	58,0	4,3	0,6	0,2	0,4	0,2	0,0	0,2	0,0	0,0	0,3	1,2
30-34	0,5	4,1	29,8	35,8	3,3	0,4	0,4	0,4	0,2	0,0	0,0	0,0	0,2	3,7
35-39	0,1	0,5	6,8	50,8	46,2	5,0	1,7	0,8	0,4	0,1	0,0	0,0	0,0	0,0
40-44	0,0	0,1	0,4	7,1	37,3	38,1	5,7	0,5	0,1	0,1	0,0	0,0	0,0	0,0
45-49	0,0	0,0	0,1	0,8	9,6	39,6	31,0	1,5	0,3	0,0	0,0	0,0	0,0	0,6
50-54	0,0	0,0	0,0	0,1	2,0	12,8	40,5	23,0	2,5	1,0	0,1	0,1	0,0	1,9
55-59	0,0	0,0	0,0	0,1	0,4	2,9	17,0	48,5	34,3	7,5	2,0	0,8	1,9	1,9
60-64	0,0	0,0	0,0	0,0	0,1	0,5	2,7	20,2	46,6	52,4	10,8	3,6	4,1	1,9
65-69	0,0	0,0	0,0	0,0	0,0	0,1	0,2	3,9	13,8	33,6	56,8	22,0	10,6	1,9
70-74	0,0	0,0	0,0	0,0	0,0	0,0	0,1	0,2	1,1	4,2	11,3	11,2	2,6	4,3
75-79	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,3	0,8	14,5	44,5	34,2	4,3
80-84	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,1	3,8	14,2	23,2	22,2
85-89	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,1	0,5	2,6	8,5	30,3
90+	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,2	1,1	12,3	25,3
Totaal		100	100	100	100	100	100	100	100	100	100	100	100	100
% of all matches	2,4	14,9	34,8	11,8	7,5	3,2	4,7	6,9	2,4	0,8	5,0	2,8	2,8	0,0

Bron: MIDAS

Figuur 16 laat zien dat de MIDAS simulatie de homogamie qua onderwijsniveau van jonge koppels in 2011 goed reproduceert. De homogamie bij vrouwen met hoger onderwijs (ISCED 5+) wordt enigszins onderschat.



De migrant-status van gesimuleerde koppels komt minder goed overeen met die van jonge koppels in 2011. In de MIDAS simulatie worden vrouwen die niet in België geboren zijn (immigrant) verhoudingsgewijs te vaak gekoppeld aan mannen die eveneens emigranten zijn (Figuur 17). Hierbij moet rekening worden gehouden met het feit dat het ‘aanbod’ van mannelijke immigranten gedurende de projectieperiode groter is dan in 2011, zoals blijkt uit de kolom ‘totaal’.

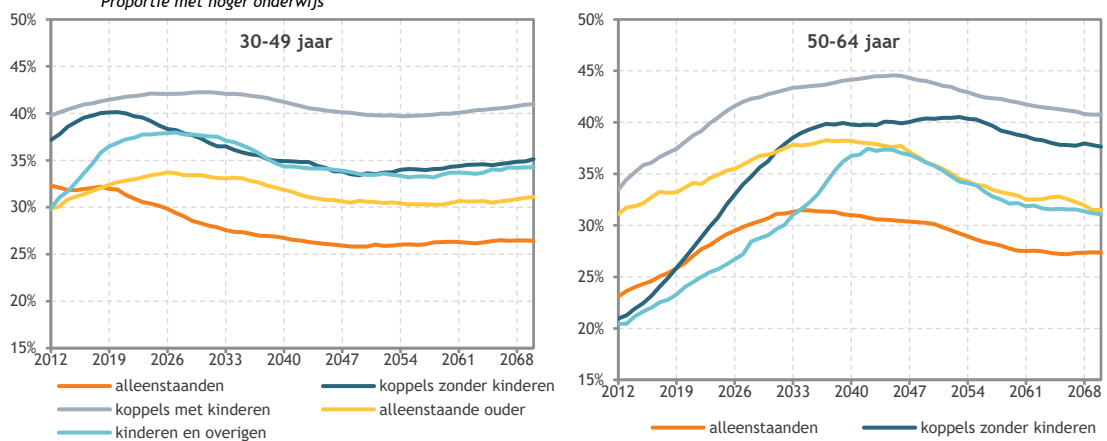


Gradiënten naargelang onderwijsniveau

Onderwijsniveau is een belangrijke determinant van tewerkstelling en het arbeidsinkomen. Om deze reden is geprobeerd de samenhang tussen de LIPRO-positie en het onderwijsniveau zoals vastgesteld in de startdata te behouden. Vooral voor alleenstaanden en éénoudergezinnen hangt het armoederisico sterk samen met het onderwijsniveau van de enige kostwinner. Om de grafieken overzichtelijk te houden beschouwen we de proportie met een diploma hoger onderwijs (op de huidige arbeidsmarkt is dit de belangrijkste scheidingslijn) naargelang LIPRO positie, waarin de samenwonende en gehuwde koppels zijn samengevoegd tot twee categorieën (met of zonder kinderen), evenals alle kinderen en ‘overigen’.

Figuur 18 laat zien dat onder de 30-49 jarigen in 2012, alleenstaande ouders, kinderen en ‘overigen’, en alleenstaanden minder vaak een diploma hoger onderwijs hadden dan personen die leefden met een partner, al dan niet met kinderen. In de loop van de projectieperiode, bij globaal stijgende proporties personen met hoger onderwijs, blijven deze verschillen bestaan, met uitzondering van de kleine en heterogene categorie van kinderen en ‘overigen’. Het patroon is gelijkaardig voor de 50-64 jarigen, hoewel in die leeftijdsgroep het aandeel van de personen met hoger onderwijs sterker stijgt. Ongeveer de helft van de koppels zonder kinderen in deze leeftijdsgroep zijn in feite 60 jaar of ouder. De cohortsgewijze stijging van het onderwijsniveau bereikt de personen in koppels zonder kinderen daarom later dan die in andere LIPRO-posities, hetgeen de sterke stijging tussen 2012 en 2035 van de proportie met hoger onderwijs in deze groep verklaart. Dankzij het feit dat onderwijsniveau als variabele is opgenomen in de scores die de kans op koppelvorming (positief) en scheiding (negatief) bepalen, blijven dus de gradiënten qua onderwijsniveau tussen de LIPRO-posities globaal behouden.

Figuur 18 Proportie met diploma hoger onderwijs naargelang lipro-positie
Proportie met hoger onderwijs



5. Conclusie

Dit document presenteert de modellering in MIDAS van de demografische variabelen, gebeurtenissen en ontwikkelingen: geboorte, sterfte, gezinsvorming en onderwijsniveau. Deze modellering is volledig herzien en sterk uitgebreid ten opzichte van Dekkers et al. (2010), met het doel de resultaten van MIDAS te aligneren aan de projectie van de positie in het huishouden (LIPRO-positie), die jaarlijks door het Planbureau uitgevoerd en gepubliceerd wordt. Beperkte aanpassingen aan de LIPRO-categorieën waren nodig om deze op basis van de beschikbare gegevens te kunnen projecteren. De geprojecteerde trends in de LIPRO-posities per geslacht en leeftijdscategorie komen bijna steeds sterk overeen met de genoemde gepubliceerde projecties. Waar er afwijkingen zijn, zijn deze verklaarbaar en verdedigbaar.

Daarnaast reproduceert de modellering van koppelvorming van vrouwen en mannen vrij adequaat de homogamie naargelang leeftijd, onderwijsniveau en migratiestatus die geobserveerd wordt in de startdata. Hetzelfde geldt – voor zover mogelijk binnen de vereisten van de alignering van de LIPRO – voor de verdeling van het aantal kinderen per vrouw, en voor de verschillen in onderwijsniveau tussen LIPRO-posities.

Literatuur

Dekkers, Gijs, Raphaël Desmet, en Greet De Vil, 2010 - *The long-term adequacy of the Belgian public pension system: An analysis based on the MIDAS model*, Working Paper 10-10, Mei 2010, Brussel: Federaal Planbureau.

http://www.plan.be/publications/publication_det.php?lang=nl&TM=30&IS=63&KeyPub=946.

Li, Jinjing, Cathal O'Donoghue en Gijs Dekkers, 2014 - Dynamic Microsimulation. In: Cathal O'Donoghue (Ed). *Handbook of Microsimulation*. Emerald Insight. Contributions in Economic Analysis series.

MIDAS: The modelling of the population by region

Annex 4 bij 'MIDAS 2.0: Revisie van een
dynamisch microsimulatiemodel'

January 2023

Gijs Dekkers, gd@plan.be

Table of contents

- 1. Introduction 1
- 2. Structure of the modules 4
- 3. Output 8
- 4. References..... 11
- 5. Appendix regressions 12

List of tables

- Table 1 Key figures for the simulation error of the regions 8
- Table 2 Regression results for Flanders..... 12
- Table 3 Regression results for the Walloon area 12
- Table 4 Regression results for Brussels 13

List of graphs

- Graph 1 Size of the 3 regions of Belgium; demographic projections..... 2
- Graph 2 Proportional size of the regions; demographic projections versus simulation MIDAS without alignment 2
- Graph 3 Input for the alignment tables of region..... 5
- Graph 4 The alignment tables of region 5
- Graph 5 Flow chart of the modelling of regions in MIDAS 6
- Graph 6 Surface plots of the simulation error to age group and period 9

Federal Planning Bureau

The Federal Planning Bureau (FPB) is a public agency that carries out, in support of political decision-making, forecasts and studies on economic, social-economic and environmental policy issues and examines their integration into a context of sustainable development. It shares its expertise with the government, parliament, social partners, national and international institutions.

The FPB adopts an approach characterised by independence, transparency and the pursuit of the general interest. It uses high-quality data, scientific methods and empirical validation of analyses. The FPB publishes the results of its studies and, in this way, contributes to the democratic debate.

The Federal Planning Bureau is EMAS-certified and was awarded the Ecodynamic enterprise label (three stars) for its environmental policy.

<https://www.plan.be>

e-mail: contact@plan.be

With acknowledgement of the source, reproduction of all or part of the publication is authorized, except for commercial purposes.

Responsible publisher: Saskia Weemaes

Legal deposit: D/2023/7433/3

Federal Planning Bureau

Rue Belliard – Belliardstraat 14-18, 1040 Brussels

phone: +32-2-5077311

e-mail: contact@plan.be

<https://www.plan.be>

MIDAS: The modelling of the population by region

Annex 4 bij ‘MIDAS 2.0: Revisie van een dynamisch microsimulatiemodel’

January 2023

Gijs Dekkers, gd@plan.be

Abstract - This annex presents and discusses a new module for modelling the population distribution by region in MIDAS, by moving households between regions. The author would like to thank Alix Damman for his help in developing the necessary Python/LArray code.

1. Introduction

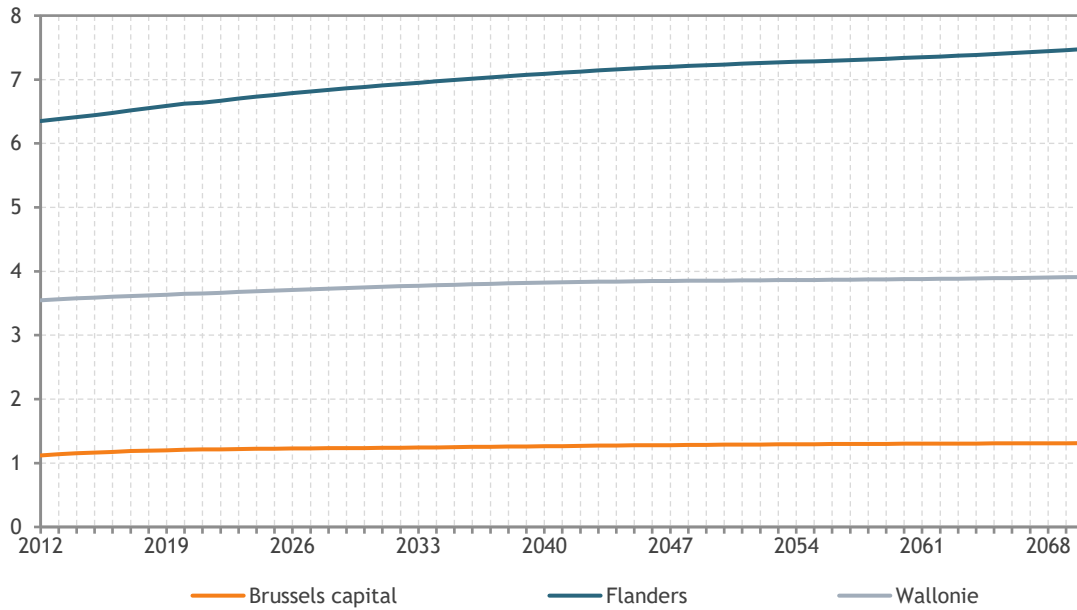
Belgium is a federal state that consists of communities and regions – apart from- and in-between the federal level, provinces, and local governments (Belgian Federal Government, 2020). The communities and regions are federal entities, are complementary and operate on the same hierarchical level.

The communities are related to language and culture. There are 3 communities: the French-speaking community, the Flemish community, and the German-speaking community. They decide on matters such as culture, education, language, and health care (Flemish Government, 2020). So, for example, the French speaking community includes all the residents of Wallonia as well as the French-speaking inhabitants of the bi-lingual region of Brussels-capital. The Flemish speaking community includes the inhabitants of Flanders and the Flemish-speaking inhabitants of Brussels, and -finally- the inhabitants German-speaking municipalities near the German border.

There are three regions: the Flemish region, the Walloon area and the Brussels-capital region. The competence of the regions is bound to their territory, and are related to socio-economic policy, including economy, housing, transport, spatial planning, etcetera (op cit.).

The Federal Planning Bureau produces demographic projections specific to the various entities (Belgium as a whole, the communities, the regions, the provinces and district ('arrondissement') (see Federal Planning Bureau, 2020). Currently, the model MIDAS does not differentiate the simulated population to regions. This is problematic because the economic developments between the regions are different, and because region is an important factor in the behavioural equations and earnings equations. Furthermore, there are important trends in transitions between the various regions, that affect at any point in time the distribution of age and gender in the regions. For example, the last years the Brussels-capital region has seen emigration of especially young households to Flanders (Federal Planning Bureau, 2020b, 13). The below Graph 1 shows the numbers of individuals in the three regions in Belgium, and their projected development.

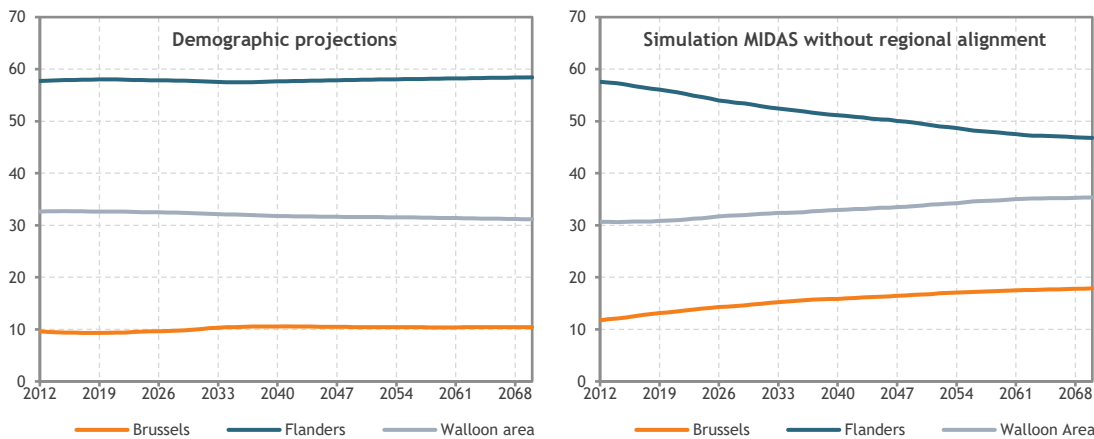
Graph 1 Size of the 3 regions of Belgium; demographic projections
millions of people



Source: FPB demographic projections 2019-2070 update COVID-19. Loop van de bevolking per gewest. <https://www.plan.be/databases/data-35-nl-bevolkingsvooruitzichten+2019+2070+update+covid+19> [09/06/2020].

The left panel of Graph 2 shows the proportional size of the three regions in the demographic projections. In the version of MIDAS including migration and excluding regional alignment, the projected development of the proportional sizes of the regions are shown in the right panel of Graph 2.

Graph 2 Proportional size of the regions; demographic projections versus simulation MIDAS without alignment



Source: MIDAS; FPB demographic projections 2019-2070 update COVID-19.

Clearly the two projections are different in their development. The proportional size of Brussels and, to a lesser extent, the Walloon area is the result of the comparably younger population than in Flanders. This implies that more women are getting children in Brussels and Walloon area. Furthermore, the migration module uses the pageant algorithm to “clone” selected existing immigrants as projected immigrants (Dekkers et al., 2020). During this process, the region of the ‘donor’ is kept for the clone. The

proportion of migrants is higher in Brussels than in the other regions, which means that most donors live in Brussels, and most clones therefore live in Brussels too.

Clearly this discrepancy in the proportional sizes of the regions is problematic if the region is used as an argument either in modelling or in the behavioural equations. The purpose of this document is to explain the way that MIDAS aligns to the demographic projections by region, age, and gender.

2. Structure of the modules

It is known from the demographic projections which percentage of any age-gender combination is in which region in any given year between 2012 and 2070. The input file has the form shown in Graph 3.

The purpose of this document is to explain the way that MIDAS aligns to the demographic projections to region, age and gender. In principle, one could expect this adaptation to be straightforward: just move individuals of a designated age and gender from one region to another. However, the variable “region” is a household-level variable because we assume that those individuals that share a household all live in the same region. Thus, moving individuals from one region to another would mean breaking up households, which is not what we want. Furthermore, we have 3 regions which can either have too many or too few persons in them. This means that there are 8 combinations possible, and this for each combination of age group and gender.

The first problem is resolved by using the Chénard algorithm (Chénard, 2000; Dekkers, 2015), which allows us to select households to meet targets on the level of the individuals. The second problem is resolved by working in two steps. In the first step, households and individuals from regions with a surplus are “moved” to an intermediate region (with value 0). In the second step, the individuals and households in this intermediate region are distributed over regions with shortages. This is done separately for all combinations of age group and gender, of course.

Thus, this approach is an example of using the Chénard algorithm, taking it beyond the usual application in the modelling of migration (see Dekkers, 2015). But there is another specificity, namely that the confrontation between the actual data and the simulation results in any simulation period is one on the level of the individual, while the actual application of the Chénard algorithm is applied on the level of the household. We therefore need a way for the tables that are to be used for the transitions to the intermediate region and from the intermediate region back to the actual regions to be made “on the fly” i.e. each period, and on the level of the individual. And then they need to be used on the level of the household. This we do using temporary alignment tables that the model creates itself.

The analysis is done for men and women and for the variable “*agegroup_civilstate*”. The latter groups all individuals younger than 90 in 5-year age groups (0-4, 5-9, 10-14, ...) and takes all of 90 and older together in one group. If in the rest of this text we refer to age group, we mean the above variable.

Next, the demographic projections that are the starting point of the alignment are provided by an Excel worksheet, called “*popregnl.xlsx*”, and have the form shown in Graph 3. These are numbers of individuals by age and gender. The Python/LArray program “*create_aggregate_popregnl.py*”¹ derives the age groups, and transforms this data in the appropriate form, which is for each combination of year, age group and gender. This output takes the form of an Excel/csv file, called “*popregnl_aggregate.csv*”, and is shown in Graph 4.

¹ Developed by Alix Damman.

Graph 3 Input for the alignment tables of region

The screenshot shows an Excel spreadsheet with the following content:

- Row 1:** Bevolking per gewest en leeftijd, op 1 januari - Mannen en vrouwen
- Row 2:** Bron : 1991-2019 : waarnemingen, Statbel; 2020-2021 : vooruitzichten, FPB en Statbel
- Row 3:** Copyright: Federaal Planbureau; FOD Economie - Statbel
- Row 4:** Headers for years: 1991, 1992, 1993, 1994, 1995, 1996, 1997, 1998, 1999, 2000, 2001, 2002, 2003, 2004, 2005, 2006
- Row 5:** 0 jaar: 12799, 12677, 12808, 12382, 12464, 12404, 12588, 12606, 12786, 13044, 13572, 14263, 13940, 14609, 15139, 15496
- Row 6:** 1 jaar: 12230, 12513, 12371, 12493, 12207, 12135, 12261, 12363, 12507, 12639, 12898, 13364, 14110, 13739, 14360, 15012
- Row 7:** 2 jaar: 11973, 11888, 12192, 12049, 12238, 11843, 11904, 11967, 12092, 12391, 12426, 12775, 13189, 13816, 13416, 14118
- Row 8:** 3 jaar: 11457, 11652, 11591, 11905, 11869, 11911, 11596, 11709, 11735, 11949, 12254, 12393, 12683, 12882, 13506, 13275
- Row 9:** 4 jaar: 11204, 11203, 11411, 11416, 11758, 11633, 11752, 11461, 11518, 11596, 11813, 12179, 12345, 12505, 12644, 13266
- Row 10:** 5 jaar: 10620, 10956, 11096, 11218, 11329, 11556, 11515, 11655, 11320, 11488, 11500, 11827, 12107, 12179, 12294, 12573
- Row 11:** 6 jaar: 10889, 10431, 10778, 10929, 11119, 11138, 11407, 11417, 11537, 11345, 11343, 11557, 11802, 11964, 12016, 12236
- Row 12:** 7 jaar: 10626, 10708, 10293, 10608, 10878, 10968, 11074, 11373, 11259, 11480, 11272, 11361, 11512, 11719, 11856, 11967
- Row 13:** 8 jaar: 10770, 10438, 10588, 10176, 10526, 10768, 10878, 11008, 11263, 11218, 11363, 11391, 11360, 11401, 11593, 11859
- Row 14:** 9 jaar: 11175, 10656, 10287, 10483, 10131, 10371, 10699, 10840, 10883, 11274, 11222, 11421, 11411, 11284, 11310, 11597
- Row 15:** 10 jaar: 11121, 11016, 10541, 10235, 10404, 10060, 10306, 10640, 10758, 10891, 11211, 11237, 11398, 11384, 11282, 11343
- Row 16:** 11 jaar: 10804, 10934, 10924, 10467, 10207, 10300, 10059, 10283, 10604, 10789, 10865, 11245, 11266, 11332, 11272, 11278
- Row 17:** 12 jaar: 10681, 10658, 10843, 10845, 10391, 10121, 10298, 9992, 10262, 10583, 10795, 10976, 11261, 11230, 11276, 11302
- Row 18:** 13 jaar: 10477, 10521, 10587, 10766, 10781, 10339, 10101, 10309, 9963, 10326, 10586, 10906, 11008, 11216, 11221, 11324
- Row 19:** 14 jaar: 10243, 10308, 10452, 10531, 10761, 10725, 10341, 10082, 10293, 10001, 10346, 10672, 11009, 11045, 11205, 11289
- Row 20:** 15 jaar: 10203, 10171, 10244, 10425, 10511, 10695, 10660, 10337, 10117, 10337, 10043, 10494, 10747, 11049, 11129, 11254
- Row 21:** 16 jaar: 10275, 10102, 10143, 10233, 10467, 10469, 10749, 10629, 10354, 10167, 10383, 10177, 10618, 10805, 11059, 11202
- Row 22:** 17 jaar: 10861, 10221, 10126, 10163, 10295, 10522, 10551, 10783, 10715, 10478, 10218, 10527, 10324, 10752, 10944, 11177

Source: popregnl.xlsx

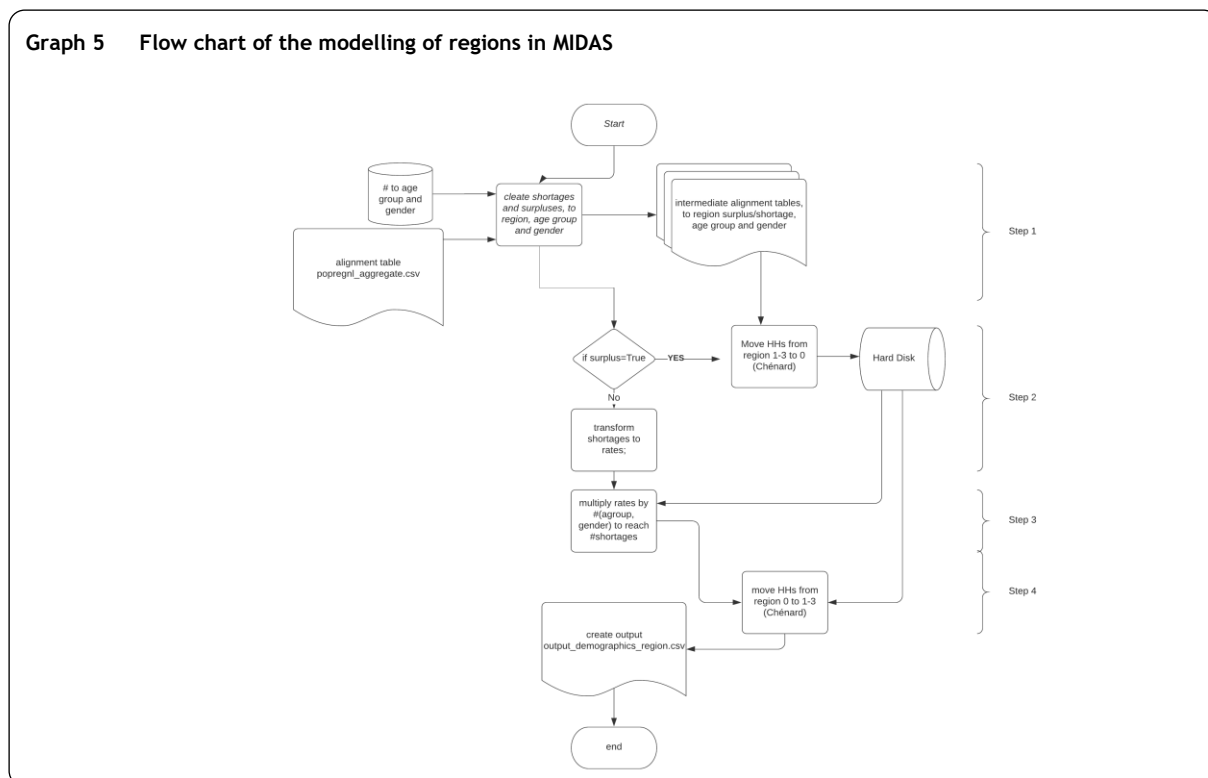
Graph 4 The alignment tables of region

The screenshot shows an Excel spreadsheet with the following content:

- Row 1:** agegroup, period, gender, region
- Row 2:** Headers for regions: 1, 2, 3
- Row 3:** 0, 1991, FALSE, 9.894694, 55.77823, 34.32708
- Row 4:** 0, 1991, TRUE, 9.88244, 55.8205, 34.29706
- Row 5:** 0, 1992, FALSE, 9.805293, 55.78268, 34.41203
- Row 6:** 0, 1992, TRUE, 9.752928, 55.8838, 34.36328
- Row 7:** 0, 1993, FALSE, 9.762769, 55.91956, 34.31767
- Row 8:** 0, 1993, TRUE, 9.694008, 56.00506, 34.30093
- Row 9:** 0, 1994, FALSE, 9.742548, 56.22793, 34.02952
- Row 10:** 0, 1994, TRUE, 9.673068, 56.26288, 34.06405
- Row 11:** 0, 1995, FALSE, 9.814423, 56.42228, 33.76329
- Row 12:** 0, 1995, TRUE, 9.869214, 56.38311, 33.74768
- Row 13:** 0, 1996, FALSE, 9.875698, 56.52931, 33.59499
- Row 14:** 0, 1996, TRUE, 9.958904, 56.46474, 33.57635
- Row 15:** 0, 1997, FALSE, 10.06941, 56.48686, 33.44373
- Row 16:** 0, 1997, TRUE, 10.16443, 56.31336, 33.52222
- Row 17:** 0, 1998, FALSE, 10.18792, 56.35494, 33.45714
- Row 18:** 0, 1998, TRUE, 10.35921, 56.15897, 33.48182
- Row 19:** 0, 1999, FALSE, 10.43466, 55.91077, 33.65457
- Row 20:** 0, 1999, TRUE, 10.50652, 55.92608, 33.5674
- Row 21:** 0, 2000, FALSE, 10.68182, 55.57137, 33.74681
- Row 22:** 0, 2000, TRUE, 10.66883, 55.5424, 33.78877
- Row 23:** 0, 2001, FALSE, 10.89412, 55.10158, 34.0143
- Row 24:** 0, 2001, TRUE, 10.92976, 54.99119, 34.07904
- Row 25:** 0, 2002, FALSE, 11.26177, 54.56326, 34.17496
- Row 26:** 0, 2002, TRUE, 11.30574, 54.52532, 34.16893
- Row 27:** 0, 2003, FALSE, 11.60274, 54.13815, 34.25911
- Row 28:** 0, 2003, TRUE, 11.54231, 54.15443, 34.30325
- Row 29:** 0, 2004, FALSE, 11.87825, 53.85669, 34.26506

Source: popregnl_aggregate.csv

Next, we proceed in four separate steps which coincide with four processes. These are called upon in each period separately, so the period will not be mentioned in the below discussion. These are also shown in the next flowchart.

Graph 5 Flow chart of the modelling of regions in MIDAS

Note that the steps are separated by the fact that they apply to different entity levels. Steps 1 and 3 apply to the level of the individuals in the model, whereas steps 2 and 4 apply to the level of the household.

Step 1 (individual level): procedure “region_process_persons”

In this step, the proportions for each region are derived for each combination of age group and gender, as produced by the model in the previous period, and confronted with the exogenous demographic projections. The difference between the two percentages allows to determine whether any particular age group-gender-region combination is in surplus or shortage, if the proportion in the dataset at time t ($t=2012-2070$) is higher or is lower than the proportion in the demographic projection. This module creates for each region two csv files, one containing the surpluses (for the combination of age group and gender where these exist) and one containing the shortages. These files are named Brussels_surplus.csv, Flanders_surplus.csv and wallonie_surplus.csv; and Brussels_shortage.csv, Flanders_shortage.csv and wallonie_shortage.csv.

Step 2 (household level): procedure “region_process”

In the second step, the Chénard algorithm is used to select households using the input on surpluses on the individual levels from Brussels, Flanders and the Walloon area. The households selected move from their respective region to the intermediate region with value 0.

Next, the shortages are standardized so that they sum to 1. This will allow to distribute the total surplus (i.e. all individuals in region 0) to the regions with shortages in the next step.

Step 3 (individual level): procedure “region_process_persons_2”

The third step is an intermediate step. In order to assess the numbers of individuals that are in the intermediate region state 0 and that need to be “distributed back” to the other regions, we need the standardized shortage rates that were derived in the previous step to be multiplied by the numbers per age group and gender category. The latter is gathered in step 3².

Step 4 (household level): procedure “region_process_2”

In the fourth and final step, the individual information by age group and gender is gathered. The numbers of individuals per age group and gender are derived for each region by multiplying the shortage rates per region by the number of individuals in region 0. Next the Chénard algorithm is applied again to move households from the intermediate region 0 to the receiving regions. This is done for Brussels and Flanders. The last region, the Walloon area, finally acts as a balance entry and receives the remaining households in the intermediate region.

² Note that, in order for the result of the “groupby()” command to get the correct form to be combinable with the other arrays, it must be multiplied by a unit matrix of the same dimensions.

3. Output

The output of this series of modules is presented in “*output_demographics_region.csv*”. This contains the proportions of individuals in each region, by gender and age group. This means that the regions add up to 100% within each age group. In order to validate these results, the original input dataset can be transformed using the Python/LArray program “*Aggregate_popregn_comparable_output.py*”³. This results in a view of the data, which can be copied into “*output_demographics_region.csv*” and compared to the input alignment tables, “*popregnl_aggregate.csv*”.

So now we take as simulation error $\text{abs}(\text{sim_prop} - \text{align_prop})$; or the absolute value of the difference between the simulated proportion and alignment rate. This simulation error is available for each region, and all arguments of the alignment tables of the the regions: being gender (male, female), age group (0, 5, ..., 105), and period (2012, ..., 2070), region (Flanders, Walloon area, Brussels). Clearly this is too much to study in detail, so we need to use statistical analyses and graphs. First of all, the below table shows some key results for the simulation error of flanders (e_{Fl}), the Walloon area (e_{Wa}) and Brussels (e_{Br}).

Table 1 Key figures for the simulation error of the regions

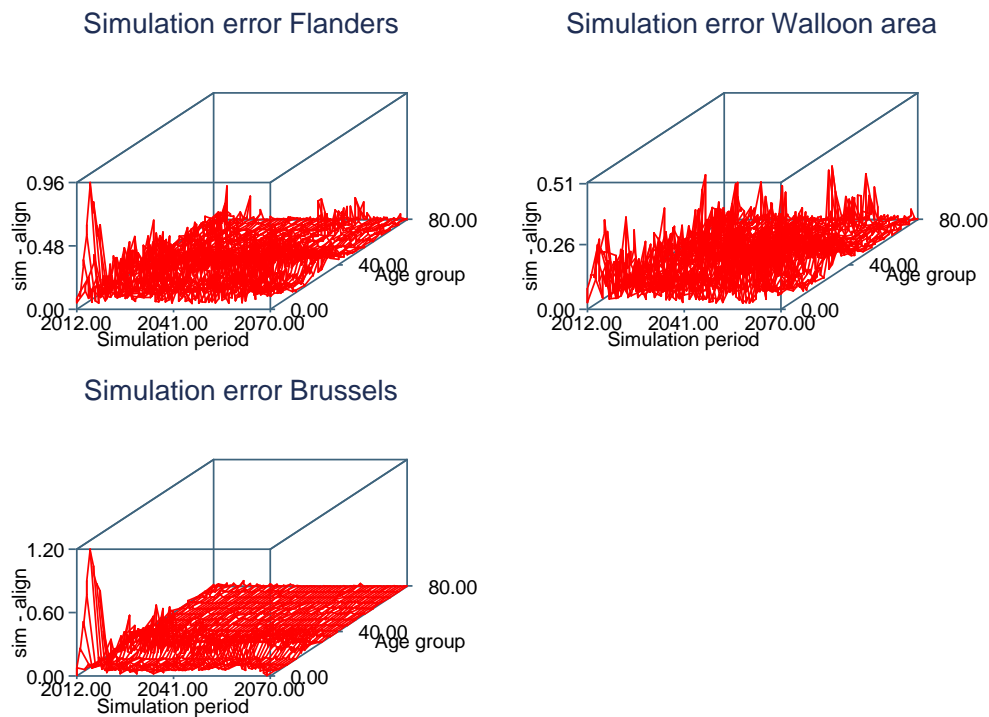
	MEAN	MEDIAN	P90	P10	MIN	MAX
Flanders	0.067	0.028	0.171	0.003	0.000	0.962
Walloon area	0.052	0.026	0.137	0.003	0.000	0.513
Brussels	0.032	0.004	0.09	0.000	0.000	1.198

Note: all values are in percentage points of the absolute difference between the simulated proport and the target alignment proportion.

In all three regions, the errors are small. The next Graph 6 shows the surface plots of the simulation error for the three regions, using age group and period as the x and y axis.

³ Developed by Alix Damman.

Graph 6 Surface plots of the simulation error to age group and period



Source: MIDAS.

First of all, this graph confirms that the simulation errors are small or even absent for the vast majority of combinations of age group and period. Secondly, the above Graph 6 suggests that simulation results are largest at the origin of the axis; i.e. among younger age groups and in the first decades of simulation. It is however difficult to see whether the angle of the surface is more towards the simulation period (i.e. towards the east) than towards the age group (i.e. towards the north).

To see this, we did a simple dominance analysis (Luchman, 2013; Budescu, 1993) on the simulation errors in Flanders, the Walloon area and Brussels, using age group, period, and gender as arguments. Dominance is defined as a pairwise relationship that can be tested for all pairs of variables included in the model. One variable is more important than its competitor if its predictive ability exceeds the others in all subset regressions (Budescu, 1993). (The results are available on request). The conclusions are the same for the three regions, namely that age group is dominant over period and gender, while period is dominant over gender. Furthermore, the results in Appendix show the results of regressing the simulation error on age group, period, and gender in the three regions. In all three regions, the relation between the simulation errors and age group and period are of a higher order. Furthermore, the errors decrease immediately with age, only to stabilize for the age groups of 35 and older, where the errors increase with period until the mid- or late-2020s, after which a strong decrease sets in.

Hence, the development of the simulation errors shown in Graph 6 appear for the largest part related to the age group of the simulated individuals. The reason for this is straightforward and is that the proposed modelling of the regions uses the Chénard algorithm to transfer households between regions to match the alignment tables expressed for individuals. Households that consist of one or more individuals of the same age group therefore are ideal, because a transfer of such a household will not result

in an error in other age groups. Thus, the more the members of a specific age group are found in one household with other age groups (but not the other way around), the more they will be prone to simulation errors. Typically, this is the case with underage children, as they will usually live with at least one individual of other age groups. Furthermore, households with children typically are larger than households without children. This explains why the simulation errors are larger among the younger age groups, despite the fact that the number of births is aligned to the population projections by region.

4. References

- Belgian Federal Government, 2020, *Belgium, a federal state*. Retrieved from the Belgium.be portal. https://www.belgium.be/en/about_belgium/government/federale_staat [09/06/2020]
- Budescu, D. V. (1993). Dominance analysis: A new approach to the problem of relative importance of predictors in multiple regression. *Psychological Bulletin*, 114(3), 542-551. doi:<http://dx.doi.org.kuleuven.ezproxy.kuleuven.be/10.1037/0033-2909.114.3.542>.
- Chènard, Denis, 2000, Individual alignment and group processing: an application to migration processes in DYNACAN. in Mitton, L., Sutherland, H. and Weeks, M. (eds), *Microsimulation Modelling for Policy Analysis*, Cambridge University Press, Cambridge.
- Dekkers, Gijs, 2015. *On the Modelling of Immigration and Emigration using LIAM2*. Federal Planning Bureau, Note LIAM2 – 11155. November 2015. https://www.researchgate.net/publication/286869730_On_the_modelling_of_immigration_and_emigration_using_LIAM2
- Dekkers, Gijs, Raphaël Desmet and Karel Van den Bosch, 2020, *Immigration in MIDAS*. Mimeo Federal Planning Bureau, August 2020.
- Flemish Government, 2020, *Structure of Belgium*. Retrieved from the website of the Flemish Parliament. <https://www.flemishparliament.eu/about-the-flemish-parliament/structure-belgium> [09/06/2020]
- Federal Planning Bureau, 2020, *Bevolking per Gewest en Leeftijd, on January 1st*. Retrieved from the website of the Federal Planning Bureau. <https://www.plan.be/databases/demog/pop/popcomnl.xlsx> [11/05/2020]
- Federal Planning Bureau, 2020b, *Demografische Vooruitzichten 2019-2070. Bevolking en Huishoudens. Vooruitzichten* Federal Planning Bureau. Retrieved from the website of the Federal Planning Bureau. https://www.plan.be/uploaded/documents/202003030902490.FOR_POP1970_12071_N.pdf [09/06/2020]
- Luchman, 2013, Joseph N. *DOMIN: Stata module to conduct dominance analysis*, Statistical Software Components S457629, Boston College Department of Economics, revised 04 May 2015.

5. Appendix regressions

Table 2 Regression results for Flanders

Variables	(1) res1 errfl	(2) res2 errfl	(3) res3 errfl	(4) res4 errfl
agegroup	-0.00779*** (0.000493)	-0.0318*** (0.00179)	-0.0764*** (0.00405)	-0.0764*** (0.00402)
agegroup2		0.000300*** (2.16e-05)	0.00174*** (0.000120)	0.00174*** (0.000119)
agegroup3			-1.20e-05*** (9.82e-07)	-1.20e-05*** (9.76e-07)
period	-0.00582*** (0.000708)	-0.00582*** (0.000677)	-0.00582*** (0.000653)	0.00698*** (0.00256)
period2				-0.000221*** (4.26e-05)
gender	0.0702*** (0.0241)	0.0702*** (0.0230)	0.0702*** (0.0222)	0.0702*** (0.0221)
Constant	0.671*** (0.0332)	0.970*** (0.0384)	1.222*** (0.0424)	1.100*** (0.0482)
Observations	2,006	2,006	2,006	2,006
R-squared	0.140	0.216	0.270	0.280

Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Table 3 Regression results for the Walloon area

Variables	(1) res1 errwa	(2) res2 errwa	(3) res3 errwa	(4) res4 errwa	(5) res5 errwa
agegroup	-0.00329*** (0.000256)	-0.0124*** (0.000952)	-0.0285*** (0.00220)	-0.0285*** (0.00219)	-0.0285*** (0.00217)
agegroup2		0.000114*** (1.15e-05)	0.000631*** (6.49e-05)	0.000631*** (6.47e-05)	0.000631*** (6.42e-05)
agegroup3			-4.30e-06*** (5.33e-07)	-4.30e-06*** (5.31e-07)	-4.30e-06*** (5.26e-07)
period	-0.00239*** (0.000369)	-0.00239*** (0.000360)	-0.00239*** (0.000354)	0.00303** (0.00139)	0.0213*** (0.00337)
period2				-9.34e-05*** (2.32e-05)	-0.000886*** (0.000136)
period3					9.11e-06*** (1.54e-06)
gender	0.0766*** (0.0126)	0.0766*** (0.0123)	0.0766*** (0.0121)	0.0766*** (0.0120)	0.0766*** (0.0119)
Constant	0.297*** (0.0173)	0.411*** (0.0204)	0.501*** (0.0230)	0.450*** (0.0262)	0.366*** (0.0297)
Observations	2,006	2,006	2,006	2,006	2,006
R-squared	0.108	0.151	0.177	0.184	0.198

Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Table 4 Regression results for Brussels

Variables	(1) res1 errbr	(2) res2 errbr	(3) res3 errbr	(4) res4 errbr	(5) res5 errbr
agegroup	-0.00520*** (0.000289)	-0.0190*** (0.00105)	-0.0420*** (0.00240)	-0.0420*** (0.00239)	-0.0420*** (0.00234)
agegroup2		0.000173*** (1.27e-05)	0.000913*** (7.11e-05)	0.000913*** (7.07e-05)	0.000913*** (6.93e-05)
agegroup3			-6.17e-06*** (5.83e-07)	-6.17e-06*** (5.80e-07)	-6.17e-06*** (5.68e-07)
period	-0.00425*** (0.000416)	-0.00425*** (0.000398)	-0.00425*** (0.000388)	0.00220 (0.00152)	0.0333*** (0.00364)
period2				-0.000111*** (2.54e-05)	-0.00146*** (0.000147)
period3					1.55e-05*** (1.66e-06)
gender	0.0316** (0.0142)	0.0316** (0.0136)	0.0316** (0.0132)	0.0316** (0.0131)	0.0316** (0.0129)
Constant	0.463*** (0.0195)	0.636*** (0.0226)	0.766*** (0.0252)	0.704*** (0.0287)	0.561*** (0.0320)
Observations	2,006	2,006	2,006	2,006	2,006
R-squared	0.178	0.247	0.287	0.294	0.323

Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

MIDAS: Simulation of the GRAPA/IGO, THAB/APA, the IVT/ARR and Career Interruption Benefits

Annex 5 bij 'MIDAS 2.0: Revisie van een
dynamisch microsimulatiemodel'

January 2023

Karel Van den Bosch, kvdb@plan.be

Table of contents

- 1. Introduction 1
- 2. Simulation of the GRAPA/IGO 3
- 3. Simulation of the THAB/APA 12
- 4. Simulation of the IVT/ARR 19
- 5. Simulation of career interruption benefits 22
- 6. Conclusion 27
- Appendix..... 28

List of tables

Table 1	Logit regression of being an actual GRAPA/IGO beneficiary, within the group of old people with simulated entitlement	7
Table 2	Logistic regression of receiving the THAB/APA, 2011	13
Table 3	Alignment table: proportion of people eligible for the THAB/APA, by age and partner status, 2011/2012	13
Table 4	Alignment table: distribution of people eligible for the THAB/APA over THAB/APA levels, by age and partner status, 2011/2012	15
Table 5	Transition table of THAB/APA levels from one year (t) to the next year (t+1), 2012/2020	16
Table 6	THAB/APA maximum amounts by level, and income thresholds in 2011	17
Table 7	Logistic regression of transition into the ARR/IVT between 2011 and 2012	20
Table 8	Regression of career interruption benefit on background variables, 2011	24

List of graphs

Graph 1	Proportion of actual GRAPA/IGO beneficiaries among older people with simulated entitlement, by simulated amount (as a percentage of maximum amount)	8
Graph 2	Scatterplot of simulated GRAPA/IGO amounts with observed benefits	9
Graph 3	Results of the GRAPA/IGO projection: simulated entitlement and selected beneficiaries	10
Graph 4	Results of the GRAPA/IGO projection: selected beneficiaries as a proportion of all older people, by gender and age category	10
Graph 5	Average amount of GRAPA/IGO for selected beneficiaries, overall and by beneficiary status	11
Graph 6	Distribution of THAB/APA amounts among people receiving the THAB/APA and below the THAB/APA thresholds, 2011 and 2012	15
Graph 7	Projected evolution of THAB/APA receivers, 2012-2070	18
Graph 8	Projected evolution of THAB/APA amounts, 2012-2070	18
Graph 9	Proportion dependent on ARR/IVT, by gender and age group, in 2011	19
Graph 10	Projected evolution of ARR/IVT beneficiaries and amounts, 2012-2070	21
Graph 11	Career interruption benefit beneficiaries among part-time employees, by gender, age category and part-time work proportion, 2011-2012	23
Graph 12	Projected Career Interruption Benefits, 2012-2070	26

Federal Planning Bureau

The Federal Planning Bureau (FPB) is a public agency that carries out, in support of political decision-making, forecasts and studies on economic, social-economic and environmental policy issues and examines their integration into a context of sustainable development. It shares its expertise with the government, parliament, social partners, national and international institutions.

The FPB adopts an approach characterised by independence, transparency and the pursuit of the general interest. It uses high-quality data, scientific methods and empirical validation of analyses. The FPB publishes the results of its studies and, in this way, contributes to the democratic debate.

The Federal Planning Bureau is EMAS-certified and was awarded the Ecodynamic enterprise label (three stars) for its environmental policy.

<https://www.plan.be>

e-mail: contact@plan.be

With acknowledgement of the source, reproduction of all or part of the publication is authorized, except for commercial purposes.

Responsible publisher: Saskia Weemaes

Legal deposit: D/2023/7433/3

Federal Planning Bureau

Rue Belliard – Belliardstraat 14-18, 1040 Brussels

phone: +32-2-5077311

e-mail: contact@plan.be

<https://www.plan.be>

MIDAS: Simulation of the GRAPA/IGO, THAB/APA, the IVT/ARR and Career Interruption Benefits

Annex 5 bij ‘MIDAS 2.0: Revisie van een dynamisch microsimulatiemodel’

January 2023

Karel Van den Bosch, kvdb@plan.be

Abstract - This annex presents the projection of four income components in MIDAS: the “Garantie de Revenus aux Personnes Âgées” (GRAPA) / “Inkomensgarantie voor Ouderen” (IGO), the “Tegemoetkoming voor hulp aan bejaarden” (THAB) / “Allocation pour l'aide aux personnes âgées” (APA), the “Inkomensvervangende tegemoetkoming” (IVT) / “Allocation de remplacement de revenus” (ARR), and the career interruption benefits paid out by the RVA / ONEM.

1. Introduction

This document presents the projection of projection of four income components: the “Garantie de Revenus aux Personnes Âgées” (GRAPA) / “Inkomensgarantie voor Ouderen” (IGO), the “Tegemoetkoming voor hulp aan bejaarden” (THAB) / “Allocation pour l'aide aux personnes âgées” (APA), the “Inkomensvervangende tegemoetkoming” (IVT) / “Allocation de remplacement de revenus” (ARR), and the career interruption benefits paid out by the RVA / ONEM.

The “Garantie de Revenus aux Personnes Âgées” (GRAPA) / “Inkomensgarantie voor Ouderen” (IGO) is a means-tested scheme that provides income to older persons (over standard pensionable age) who have insufficient resources. It is a country-wide scheme that is administered by the Federal Pension Service. All older Belgians and EU-citizens are eligible, as well as non-EU citizens who meet the residence criterion of having lived five years in Belgium or another EU country, or who have contributed to a Belgian pension.¹ The maximum amounts are higher, and the means-test is in some respects more lenient, than is the case for the general minimum income scheme for all residents of Belgium. This income component was already simulated in MIDAS, but the module has been thoroughly revised. The GRAPA/IGO is for many beneficiaries a relatively limited supplement to other incomes, in particular pensions. In this respect, it is unlike the other two benefits that constitute the floor of the Belgian income protection system (the Revenu d'Intégration / Leefloon and the ARR/IVT, see below). In 2020, only 18% of all GRAPA/IGO beneficiaries received no Belgian statutory pension from the employees, self-employed or civil servants schemes; and 33% received less than 250 Euro per month.²

The second income component is the “Tegemoetkoming voor hulp aan bejaarden” (THAB) / “Allocation pour l'aide aux personnes âgées” (APA). This a cost-compensation scheme, aimed at people aged 65 or over and experiencing difficulties with daily activities (ADL). The latter condition (often referred to as dependency) is assessed by a doctor, and graded into five categories, increasing in severity of dependency. Like the GRAPA/IGO, it is means-tested against the income of the person and their partner. It used to be administered by the Federal Public Service Social Security (SPF-SS / FOD-SZ), but from 1/1/2017 on this benefit is regulated by the three regions and the German-speaking community³. Up to today, as far as we have been able to ascertain, this separation has not resulted in a divergence of conditions or benefit levels⁴, and, without information on the future plans of the governments involved, we assume that this will remain the case in the future. The number of users has increased from 115,000 in 2005 to more than 154,000 in 2014. Simulating the THAB/APA is important, because for many older people with low pensions it is a significant source of income.

¹ See <https://www.sfpd.fgov.be/fr/droit-a-la-pension/grapa>

² Calculated from statistics published by the Federal Pension Service, see Jaarlijkse statistiek van de uitkeringsgerechtigden 2020 (fgov.be), p. 213 and p. 216.

³ At the time of writing this report, the Brussels, Flemish and Walloon regions have also taken over the administration of this benefit, while the SPF-SS/FOD-SZ continues to do this for the German-speaking community. See <https://handicap.belgium.be>

⁴ However, the system in Flanders received another name, which is “Zorgbudget voor ouderen met een zorgnood” (Care Budget for Elderly Persons with a Care Need) (<https://handicap.belgium.be/docs/nl/tegemoetkomingen-en-andere-maatregelen.pdf>, page 6).

The third income component is the “Inkomensvervangende tegemoetkoming” (IVT) / “Allocation de remplacement de revenus” (ARR), which is an allowance that is intended to support persons younger than 65 who are disabled and are not entitled to a social insurance invalidity benefit, mostly because they have never been able to work. It is an income-replacement scheme. The benefit is means-tested against the income of the person and their partner. Like the “leefloon” / “revenu d’intégration” and the GRAPA/IGO, it is part of the social safety net in Belgium. It is often supplemented by the “Integratietegemoetkoming” (IT) / “Allocation d’intégration” (AI), which is an allowance that is intended to compensate for the additional costs that a person with a disability has in order to participate in social life. Both benefits are administered by the SPF-SS / FOD-SZ. Eligibility for the ARR/IVT and the AI/IT is limited to persons below 65. However, people who received these benefits before they were 65 can continue to receive them afterwards, though only after exploiting all other entitlements, in particular the means-tested “Guaranteed Income for Older Persons” IGO/GRAPA and the above-mentioned THAB/APA. We simulate only the ARR/IVT, and not the AI/IT. This is because the latter generally involves limited amounts, and would be hard to model.

The fourth income component discussed in this report are career break benefits. Public and private sector workers have the option of temporarily interrupting their career or reducing their working hours. For civil servants this is referred to as a career break, while in the private sector, it is named time credit. The modalities of both leave systems differ. In addition, depending on their personal situation, employees can be eligible for a number of special leave arrangements: parental leave, leave for medical assistance, leave for long-term care (“mantelzorg”) and leave for palliative care. These leaves can be full-time, but are mostly part-time, for 20% or 50%. The RVA/ONEM pays a lump-sum benefit to partially compensate the loss in earnings. The benefit does not depend on the (previous) earnings, but varies by kind of leave, family situation and seniority⁵.

This note is structured as follows: the next three sections describe how these income benefits are modelled in MIDAS⁶. Then the resulting changes to the MIDAS projections are presented. The final section concludes.

⁵ See Dekkers, Gijs, and Karel Van den Bosch, 2020, 'Results of the Standard Simulations for Belgium', Mimeo Federal Planning Bureau. Deliverable of the MIGAPE project. Version 27/04/2020. Taken from the MIGAPE website http://www.migape.eu/pubs/20200423_BE_MIGAPE_report_WP2.pdf [18/06/2020], and the references to the Belgian legislation on career breaks in that report.

⁶ The description of the various systems today hinges upon the pivotal age of 65. The THAB/APA is restricted to the population of 65 or older, while the ARR/IVT/AI/IT is restricted to the population younger than 65. However, the Standard Pensionable Age in the Belgian pension system, currently 65, is decided to increase to 66 in 2025 and 67 in 2030. The simulations presented in this report implicitly assume that the age-thresholds for the THAB/APA and ARR/IVT/AI/IT follow this development.

2. Simulation of the GRAPA/IGO

Rules of the benefit

As noted in the introduction, the “Garantie de Revenus aux Personnes Âgées” (GRAPA) / “Inkomensgarantie voor Ouderen” (IGO) is a means-tested scheme that provides income to older persons (over standard pensionable age) who have insufficient resources. The maximum amount has two levels: the ‘basic amount’ for people living with a partner (married or cohabiting) and the ‘enhanced basic amount’ for single older persons. Older persons living with parents and/or (grand)children⁷ are considered to be singles. This is also the case for older people in care institutions, though not for monks or nuns living in convents or similar communities. If older persons share the same household with other relatives (e.g. brothers or sisters) or non-relatives, the ‘basic amount’ is the maximum.

The means test takes account only of the economic resources of the older person themselves and of their partner (married or legally cohabiting), but not of other persons living in the same household.⁸ Various income components and other economic resources are taken into account in different ways in the means test, as follows:

1. 90% of all first-pillar pensions (gross), including foreign pensions. In case of couples, the sum of the pensions is divided by two. In case of dependent children for whom a child allowance is received, the denominator is increased by the number of those children (for single older persons, the denominator is one plus the number of those children).
2. 75% of gross earnings of employees and civil servants, and 100% of the earnings of self-employed people (net of professional costs and social security contributions, but before taxes). There is an exemption of 5,000 Euros for this income component.
3. 100% of other social security benefits, 2nd and 3rd pillar pensions, and some private transfers.
4. Imputed income from financial capital (saving accounts, financial investments etc.): 0% of the first 6,200 Euro, 4% of the bracket 6,200-18,600 Euro, and 10% of capital exceeding 18,600 Euro.
5. Imputed income from real estate, including the own home: 100% of the non-indexed ‘cadastral income’. From this amount 743.68 Euro is subtracted, and the remainder is multiplied by three.⁹

From the sum of components 2 to 5 (so excluding the first-pillar pensions) 625 Euro is subtracted in case of cohabiting older persons (who can be entitled to the ‘basic amount’), and 1,000 Euro in case of single older persons (who are assigned the ‘enhanced basic amount’). If the resulting total means (including

⁷ This applies both to children in the sense of persons for whom the older persons is the parent or the guardian, and also to persons who depend on the older person for their subsistence, and for whom the latter receives family allowances. Those dependent persons are almost always, but not necessarily, also a child-in the first sense.

⁸ So non-relatives and relatives other than parents and (grand)children etc. count as cohabitators when setting the maximum amount (basic or enhanced basic), but their incomes are not included in the means test.

⁹ The “cadastral income” is an imputed rental value of a property, used for tax purposes. It was estimated originally in 1975 and has not been adjusted since, except after renovations. New buildings are assessed in terms of their supposed values in 1975. For tax purposes, the cadastral income is indexed with a factor that is adapted each year. The multiplication factor of three is presumably another way to take account of the increase in property values since 1975.

the pension component) are less than the 'basic amount' c.q. the 'enhanced basic amount', the difference is paid out as the GRAPA/IGO.

Older couples who are entitled to the GRAPA/IGO will most often each get the 'basic amount', reduced by the means test. However, if only one partner is 65 or over, only that partner will obtain the GRAPA/IGO.

Simulation of the benefit: introduction

The starting data do not contain information on financial capital, nor on the 'cadastral income' from real estate. Moreover, homeownership is not projected. This implies that in the projection we cannot take account of components 4 and 5. Especially the lack of information on homeownership is a serious omission, since about 70% of older people with low incomes are homeowners.¹⁰ In addition, some older persons who would meet the means-test may not apply for the GRAPA/IGO, i.e. this benefit may suffer from non-take-up. In fact, preliminary analysis shows that only about a quarter of those who meet the means-test as it is implemented in MIDAS actually are GRAPA/IGO beneficiaries. So if in the projection of the poverty risk we would assign the GRAPA/IGO to all older persons who meet this test, this would lead to very unrealistic results. We have to estimate the number and profile of those who are not GRAPA/IGO beneficiaries despite meeting the means test as implemented in MIDAS.

So the simulation of the GRAPA/IGO proceeds in three steps. First, the rules as described above are applied, including the identification of possible cohabitants, and the application of the means test as far as that is possible with the projected variables. In the second step, a proportion of those who meet the means-test are selected as actual GRAPA/IGO beneficiaries, using a logit equation that is estimated on the starting data. Finally, the amount received by beneficiaries is determined.

Simulation of the benefit: implementing the rules

As explained above, the status of being single (which determines whether the 'basic amount' or the 'enhanced basic amount' is applicable) depends on the relationships between the older person and other household members. This is determined on the basis of the lipro position (see the report on the revised demographic modules¹¹): if he or she had the lipro positions of 'single' (i.e. living in a one-person household)¹², 'head of a one-parent family' or 'child', the person is regarded as single. In case the older person was in the lipro position of 'other', they were assigned the status of single if there were no other older persons in the household with the same lipro position, and not single otherwise. This procedure produced the best agreement between the simulated and observed GRAPA/IGO amounts.

The means-test is programmed according to the rules described above, except that we have no information on income components 4 and 5 (imputed income from financial capital and real estate), and also not on 2nd and 3rd pillar pensions. In the calculation of the income component 1 (statutory pensions),

¹⁰ De Vil, G. and Van den Bosch, K. (2013), De evolutie van de armoede bij ouderen nader bekeken, Working Paper 09-13, Brussels, Federaal Planbureau, p. 29.

¹¹ This document is in progress.

¹² As explained in the report quoted, this includes people in collective households.

dependent children were only counted for older persons in lipro-positions ‘married with children’, ‘cohabiting with children’ or ‘head of a one-parent family’.

Application of this simulation procedure on the starting data resulted in an estimate of 18.8% of all older persons being entitled to the GRAPA/IGO. Only 5.7% of older persons are actual GRAPA/IGO beneficiaries; within the group of simulated beneficiaries, only 27.5% actually receive it. On the other hand, there is a small group of older persons for which the simulation indicated no entitlement, while they do receive a GRAPA/IGO amount (8.6% of all actual GRAPA/IGO beneficiaries; 0.5% of all older persons). The data for this group were inspected in some detail, in order to discover possible errors in the simulation procedure. No such errors were discovered. Some persons in this group had turned 65 in 2011, and so had incomes from other sources than pensions. Many had married partners with a substantial pension, or another social security benefit (if they were below 65). In most cases in this group (55%, 76% for singles), the GRAPA/IGO was small: less than 1,000 Euro. It seems probable that this discrepancy between the simulation results and the observations is due to timing issues. The reference period for the official assessment of GRAPA/IGO entitlement is the last month¹³, while it is the calendar year in the simulation. So presumably, where the simulation indicated GRAPA/IGO entitlement, while no GRAPA/IGO benefit was observed, incomes and/or partners had changed recently.

Simulation of the benefit: selection of beneficiaries

The selection of beneficiaries from those who seemed entitled from the simulation is based on a logistic equation, estimated in the starting data set. Using this equation, a score is calculated, and cases with a score above a certain threshold are selected. For the period 2012-2019 the proportion of beneficiaries is aligned on the proportion of actual receivers as reported by the Federal Pension Service. The details of this procedure are as follows.

In the logistic equation, the dependent variable is being a GRAPA/IGO beneficiary, and the estimation sample is comprised of those for whom entitlement was simulated. Separate equations were estimated for singles and not-singles. The variables in the equation had to be limited to those that were projected. We considered:

- Age in brackets of five years, up to 90+
- Gender
- Lipro position
- Education level: up to lower secondary, higher secondary, higher education
- Immigrant status: 1. Belgian born and both parents Belgian born, 2. Belgian born with one parent foreign-born, 3. Belgian born with two parents foreign-born, 4. foreign born
- Region: Brussels, Flanders, Wallonia
- Number of dependent children: 0, 1, 2+
- Receiving a pension from the self-employed scheme (including a survivor pension)

¹³ Mail from Arne Cornelis, FOD-SZ on 10/08/2021. The reference period is not mentioned in any official document we have been able to consult.

- Receiving a pension from the civil servant scheme (including a survivor pension)
- The simulated GRAPA/IGO amount as a proportion of the maximum possible amount, given the household situation (single or not single): linear, squared (after subtracting 0.5 to avoid multicollinearity), and a special category for when this proportion was exactly 1 (i.e. the maximum amount was received; see below).

Age and gender turned out to be not significant in both equations. Immigrant status was collapsed to two categories: people born in Belgium with at least one parent born in Belgium, versus people who were not born in Belgium or with two parents who were not born in Belgium. The lipro position was reduced to two categories (which were only statistically significant in the equation for singles): being a single parent, and having the lipro position 'other'. The results reported in Table 1 are somewhat difficult to interpret, as they can represent the balance of two effects with opposite directions: on the one hand the impact of income components that cannot be simulated, and on the other hand 'true' non take up, i.e. older people who are actually entitled to the GRAPA/IGO but not using it. E.g., those with a higher education levels are more likely to be home owners (and have other assets), but may also be better informed about their rights. The results suggests that the first factor predominates. Given simulated entitlement, first- and second-generation immigrants have a greater probability to receive a GRAPA/IGO benefit than people born in Belgium with a Belgian parent (given that entitlement is simulated), probably because they are much less likely to be homeowners.¹⁴ Older persons who receive a pension from the scheme for the self-employed are much more likely to be homeowners than those with pensions from other schemes, and consequently have a lower probability to be a GRAPA/IGO beneficiary. The reason for the negative effect of receiving a civil servant pension is not clear, within the estimation sample (unlike among older people generally) they are not more likely to be home owners.

In the logistic equation, the simulated amount is entered as a proportion of the maximum possible benefit. This has the advantage (over the amount in Euro) that this can be used throughout the projection period. The particular form, with a squared component and a binary variable indicating that the simulated amount is exactly equal to the maximum, is motivated by the non-linear relation of this variable with the probability of being a GRAPA/IGO beneficiary, as shown in Graph 1. The higher the simulated amount, the higher the proportion of beneficiaries, up to about 60% of the maximum amount, after which it is stable or slightly declining. However, the proportion of beneficiaries is considerably lower again when the simulated amount is exactly equal to the maximum amount (i.e. 100%), which implies that the beneficiary and also their partner, if present, has no pension income, nor any other income that is registered in our data. As this is a special (and in this context an interesting) group, their characteristics have been studied in some detail; see the Appendix. At any level of the simulated GRAPA/IGO amount, singles are more likely to be actual beneficiaries, probably because they are much less likely to be homeowners (53%, vs. 83% for non-singles).

In the simulation, the probability that is calculated from applying these logit equations is set to 1 for any person who received the GRAPA/IGO in the previous year. Once people are GRAPA/IGO beneficiaries, and as long as they remain entitled, there is no reason why this would stop.

¹⁴ Separate logit regressions showed that being a homeowner has a strong negative impact on the probability of being a GRAPA/IGO beneficiary. Furthermore, foreigners are much less likely to be homeowners.

Table 1 Logit regression of being an actual GRAPA/IGO beneficiary, within the group of old people with simulated entitlement

Variables	Single		Not single (men only)	
	Coefficients	St. errors	Coefficients	St. errors
Education level lower secondary	ref. cat.		ref. cat.	
Education level higher secondary	-0.300***	(0.0634)	-0.141	(0.114)
Education level higher	-0.587***	(0.0846)	-0.572***	(0.160)
Foreigner (1)	0.359***	(0.0642)	0.782***	(0.126)
Region Brussels	ref. cat.		ref. cat.	
Region Flanders	0.230***	(0.0615)	-0.140	(0.112)
Region Wallonia	0.201***	(0.0588)	-0.174	(0.116)
Simulated amount (2) linear	2.213***	(0.104)	2.298***	(0.221)
Simulated amount (2) squared	-7.013***	(0.352)	-3.098***	(0.636)
Simulated amount (2) is exactly 1	-0.898***	(0.124)	-0.890***	(0.263)
Number of dependent children 0	ref. cat.		ref. cat.	
Number of dependent children 1	-0.380**	(0.166)	-0.931***	(0.193)
Number of dependent children 2+	-0.523***	(0.154)	-0.740***	(0.220)
Receiving a pension from the scheme for the self-employed	-0.264***	(0.0545)	-0.410***	(0.117)
Receiving a pension from the scheme for civil servants	-1.346***	(0.231)	-0.183	(0.487)
Lipro position not single parent nor 'other'	ref. cat.		ref. cat.	
Lipro position single parent	-0.516***	(0.0838)		
Lipro position 'other'	-0.910***	(0.108)		
Education level partner lower secondary			ref. cat.	
Education level partner higher secondary			-0.511***	(0.122)
Education partner level higher			-0.920***	(0.189)
Constant	0.0682	(0.0864)	-1.291***	(0.180)
Observations	9,896		4,908	
Log-likelihood	-5,580		-2,045	
degrees of freedom model	14		14	
Pseudo-R ²	0.170		0.142	

*** p<0.01, ** p<0.05, * p<0.1. Ref. cat.: Reference category

Notes: (1) both partners foreign in the not single equation

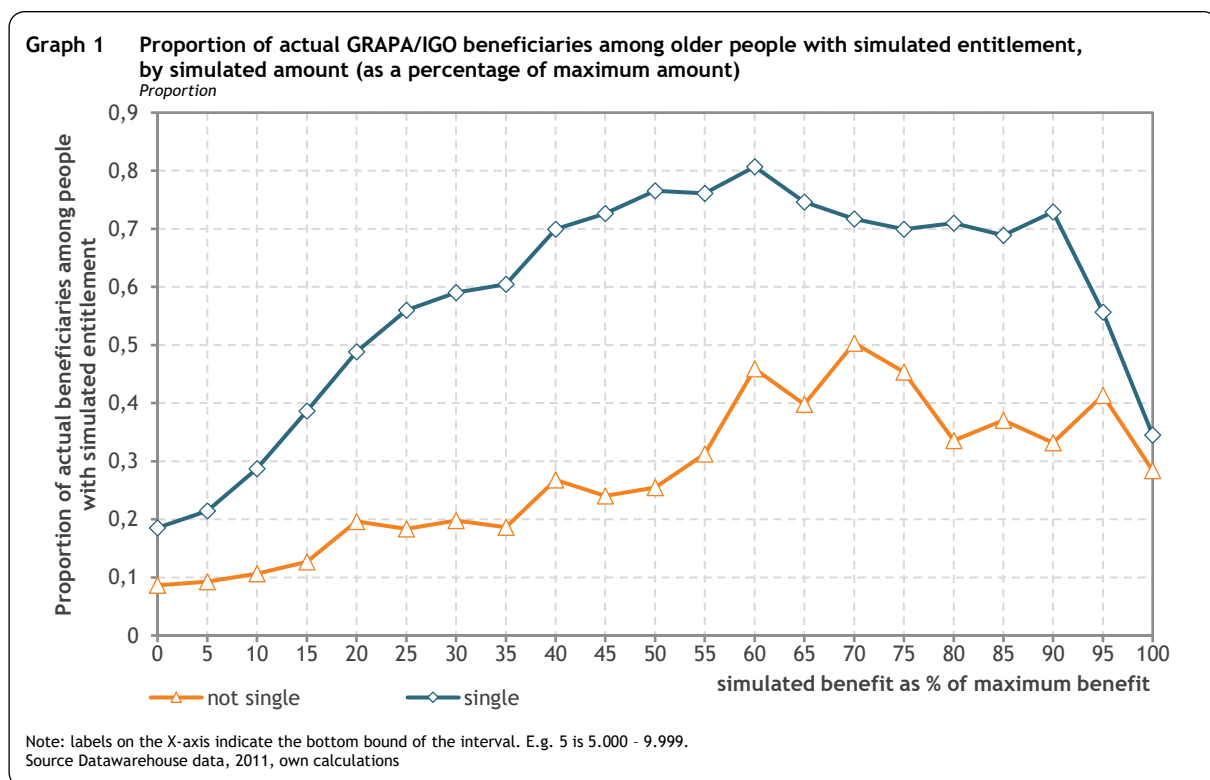
(2) Simulated amount as a proportion of the maximum amount applicable. Before taking the square, 0.5 is subtracted to avoid multicollinearity

The last step in the procedure to determine who among those for whom entitlement is simulated will be a GRAPA/IGO beneficiary, is to either set a threshold for the score itself, or take a certain proportion or number of those for whom entitlement was simulated. A hybrid procedure was chosen. For the period 2012-2019, for which we have observations (i.e. statistics from the Federal Pension Service¹⁵), the number of actual receivers (by gender) was aligned on the observed number. In 2019, score thresholds (by gender) were set in such a way that in that year the simulated proportion of all older people receiving the GRAPA/IGO by gender was the same (5.7% among women, 3.9% among men) as was observed according to the FPS statistics. These score thresholds were then applied for all years between 2020 and

¹⁵ <https://www.sfpd.fgov.be/nl/kenniscentrum/statistieken/statistieken-werknemers#jaarlijks>. The numbers of women and men receiving either the GRAPA/IGO or the "Gewaarborg inkomen voor Bejaarden" / "Revenu Garanti Aux Personnes Âgées" (the latter measure is fading out) were added up. Since the published numbers refer to the 1st January of each year, the average of the number for the year itself and the next year was used as an estimate of the number during the year.

the end of the projection period. We chose this option for the long-term, instead of keeping constant the proportion of receivers among older people for whom entitlement is calculated. There is no particular reason why that proportion should remain constant over time, while it is plausible that if the probabilities to become GRAPA/IGO beneficiary increase for some reason (e.g. more single people), the proportion actually getting it will also increase.

An important constraint is that if a couple is entitled to the benefit, the partners will jointly apply for it, so either both of them or none will receive it. To ensure joint beneficiary status in the simulation, we first select men, using the alignment proportions until 2019 and the score threshold afterwards. The wives of selected men are also selected as beneficiaries. For the years 2012-2019, the remaining number of women needed to reach the alignment proportion are then selected from single women. After 2019, the score threshold was applied to single women as described above.

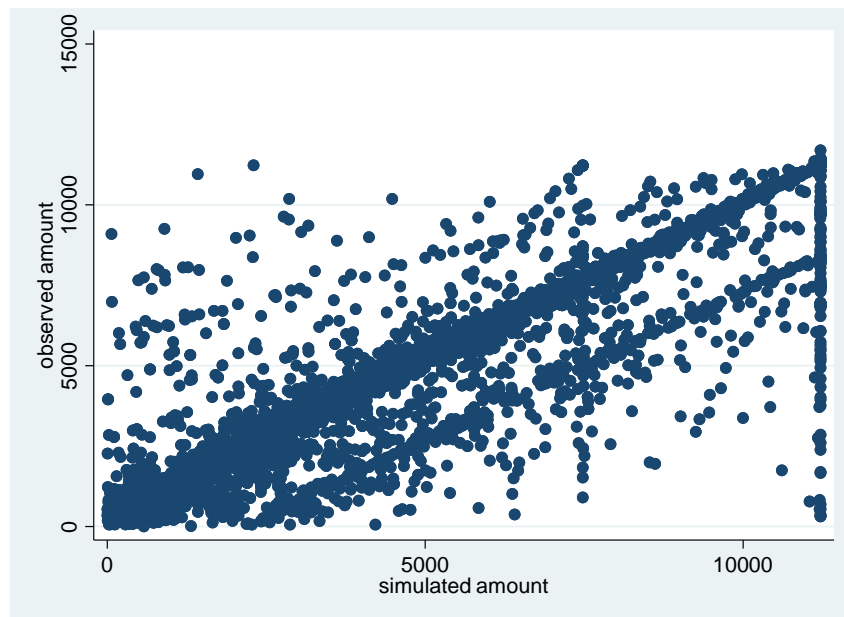


Simulation of the amount

For the selected beneficiaries, the benefit is set equal to the amount simulated as described above. A simulation on the starting data showed that the simulated amounts were, generally, quite close to the observed amounts. The average difference is only 77 Euro and the median difference -162 Euro (cf. the average observed benefit is 4,201 Euro). 80% of simulated amounts are within a bracket of -1,000 to +1,000 Euro of the observed amount for the same older person. The scatterplot in Graph 2 also shows the strong correlation, which is 0.903. There are two deviations of the pattern of concentration around the diagonal. First, for a limited group the simulated amount is between 2,000 and 3,000 Euro higher than the observed amount. Almost all of these are in the benefit category of singles, according to the simulation. Many of them are in a home. Possibly, their situation has recently changed from non-single to single. Second, for a smaller group the simulated amount is the maximum (for singles) while the

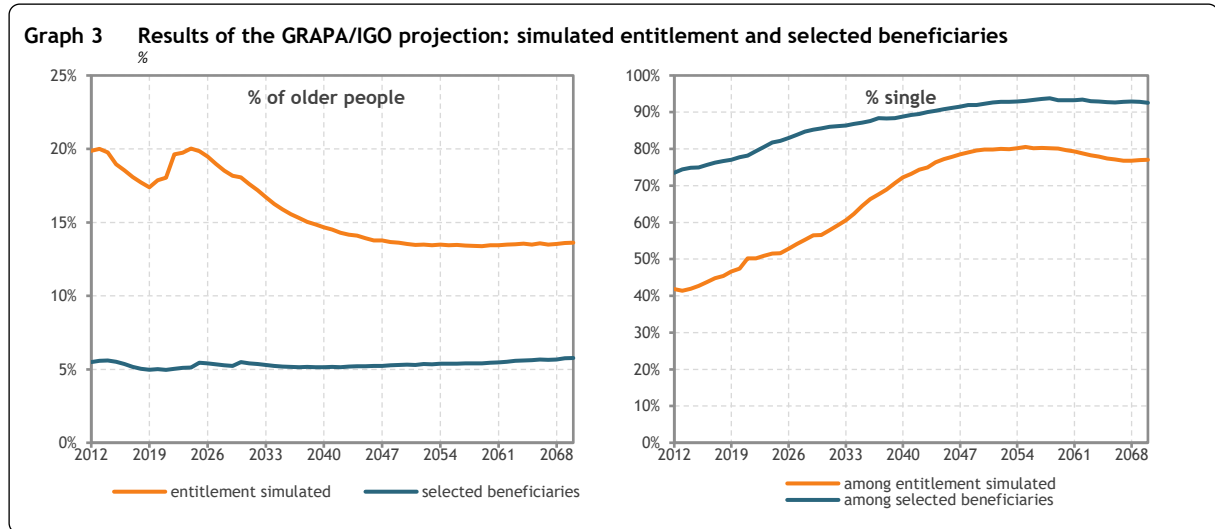
observed amount is much smaller. Recent changes in household situation or incomes are again the most probable reason.

Graph 2 Scatterplot of simulated GRAPA/IGO amounts with observed benefits

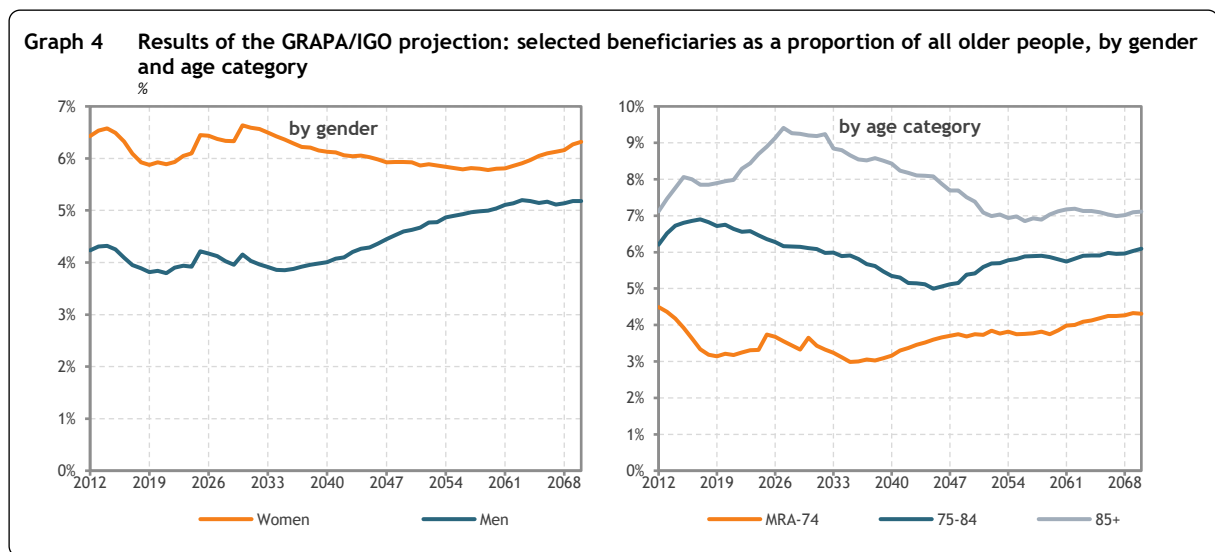


Projection results

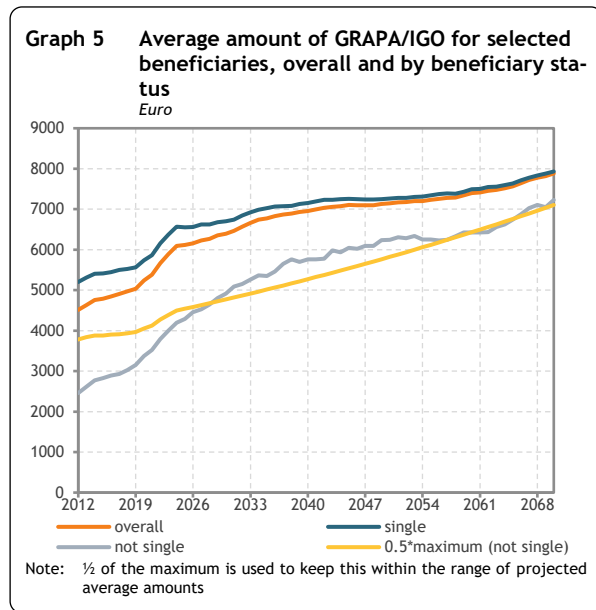
Graph 3, left graph, shows that the proportion of older persons for whom GRAPA/IGO entitlement is simulated at first declines somewhat, but then rises steeply from 2020 on to nearly 21% in 2025, as a consequence of the special increases of the maximum GRAPA/IGO amounts in this period. After 2025, this proportion gradually declines until the early 2050s and thereafter remains stable. This is the result of, on the one hand, the increase in the pensions of older people, while, on the other hand, the maximum GRAPA/IGO amounts also keep rising by 1% per year. The proportion of older people who are selected to be GRAPA/IGO beneficiaries falls slightly from 5.5% in 2012 to 5.0% in 2019, in line with observations as explained above. After that, it rises again to a maximum of 5.5% in 2030, and thereafter declines gradually to 5.0% in 2070. Overall, therefore, there is little variation in this proportion. Between 2020 and 2025 the number of actual beneficiaries does not follow the strongly increasing number of people for whom entitlement is simulated, because most of the growth in the latter group are older persons who would get only a small GRAPA/IGO amount on top of their retirement or survivor pension, and therefore have a small probability of being selected as a beneficiary. The long-term results imply that selected beneficiaries rise somewhat as a proportion of those for whom entitlement was simulated, from 29% in 2012 to a maximum of 35% in the 2040s and 2050s. The graph on the right in Graph 3 shows the reason for this: the proportion of older persons in the benefit category of 'single' increases up to 2050, both among those with simulated entitlement and among selected beneficiaries. The proportion of GRAPA/IGO beneficiaries among those with simulated entitlement is much higher in the 'single' category: 41% compared to 15% for non-singles in 2011. Though in both categories the trend of this proportion is downward, the difference between the singles and non-singles declines only slightly throughout the projection period (36% vs. 13% in 2070).



Graph 4 shows that among women the proportion of selected GRAPA/IGO beneficiaries, after a drop until 2020, at first rises until the mid-2040s, and then decreases somewhat, while it is virtually stable for men. Among the 85plus, this proportion increases up to the mid-2040s, and then declines again, while the trend is downward for those aged 75-84, and fluctuating for older people below 75. This is related to the increasing proportion of single persons among the 85plus. Note that at the same time, the 85plus will become much more numerous, so that the share of selected GRAPA/IGO beneficiaries in this age group grows from 17% in 2012 to 44% in the early 2050s.



The amounts of the GRAPA/IGO benefit increase over time, as the maxima rise by 1% per year following the hypothesis used by the Study Commission on Ageing.¹⁶ However, overall and in particular for beneficiaries in the 'single' category, the average amounts increase less than those maxima. The reason for this is that average retirement and survivor pensions received by the 'single' beneficiaries grow more quickly than the maximum GRAPA/IGO benefit. Over the period 2012-2070, the increase of the former is 112%, that of the latter 91%. Within the smaller and declining category of non-single beneficiaries, the reverse is true: their pensions increase more slowly, and therefore their average GRAPA/IGO benefit rises more quickly, and approaches that of 'single' beneficiaries.



¹⁶ Between 2012 and 2020 the maxima follow the actual amounts (which in most years rose by less than 1%), while for the period 2020-2024 additional increases were legislated.

3. Simulation of the THAB/APA

The simulation of the THAB/APA proceeds in three steps: first, we determine who satisfies the dependency condition for the THAB/APA (below this is referred to as 'eligibility'). Second, a severity-of-dependency category is assigned. Third, the amount of the benefit is calculated, taking into account the means-test.

Who is eligible for the THAB/APA?

The eligibility for the THAB/APA is assessed by a medical doctor and graded by severity of dependency. This poses a problem for the simulation of the eligibility, because we do not have any data on disability itself, nor on morbidity (chronic diseases are a major cause of disabilities). Hence the simulation of who is eligible for the THAB/APA at any simulation year between 2012 and 2070 is aligned on proportions by age category and having a partner or not. These proportions were determined on the subsample of persons aged 65+ in the starting data for 2011 and 2012, with incomes below the THAB/APA income thresholds (see below for the income definition and the thresholds). For these persons, eligibility (in terms of disability) implies receiving the THAB/APA, and vice-versa, which is not necessarily true for persons with incomes above the threshold. This procedure implies that we assume that, now and in the future, the probabilities of eligibility are constant within the relevant income range (up to the THAB/APA income thresholds).

The selection of the dimensions on which to align was based on the results of a logistic regression of receiving a THAB/APA on various background variables, using the starting dataset (Table 2). They showed that age category and having a partner or not most affected the probability of receiving the THAB/APA. Although women are more likely to receive the THAB/APA than men, the statistical significance of this gender-difference disappeared when controlling for age and partner status. Table 3 presents the alignment proportions. These are kept unchanged throughout the projection period, because we have no data to estimate possible trends in these proportions. This means that the overall proportion of THAB/APA-recipients is solely a function of changes of the distribution of age and the proportion of singles among the active population. Hence, these constant proportions imply an increasing absolute number of people who are eligible, given the ageing of the population.

Table 2 Logistic regression of receiving the THAB/APA, 2011

Variables	Coefficients	St. errors
age 65-69	Ref. cat.	
age 70-74	-0.0156	(0.0508)
age 75-79	0.246***	(0.0481)
age 80-84	0.451***	(0.0471)
age 85-89	0.744***	(0.0483)
age 90-94	0.858***	(0.0574)
age 95+	1.067***	(0.0839)
has partner	-1.259***	(0.0903)
age 70-74 and has partner	0.0968	(0.122)
age 75-79 and has partner	0.178	(0.115)
age 80-84 and has partner	0.288**	(0.117)
age 85-89 and has partner	0.0662	(0.136)
age 90-94 and has partner	0.0901	(0.213)
age 95+ and has partner	0.0865	(0.608)
Women	Ref. cat.	
Men	-0.00586	(0.0299)
Partner receives THAB	1.515***	(0.0794)
Constant	-1.417***	(0.0415)
Observations	37,570	
log-likelihood	-18,943	
df model	15	
Pseudo-R ²	0.0492	

Notes: Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1
 Estimation sample limited to people aged 65+ and below THAB/APA income thresholds

The logistic regression reported in Table 2 also shows that older persons of which the partner receives the THAB/APA are far more likely to receive the THAB/APA themselves than others. We take account of this finding in the determination of eligibility in the following way. First of all, THAB/APA eligibility is simulated for males within couples (given the age category). In a second step, THAB/APA eligibility is determined for women within couples (given the age category), while increasing the odds to be selected by 4.55 if her partner was simulated to be eligible in the first step. Finally, there is a *soft take*¹⁷ of persons who were eligible in the previous year, which in practice means that nearly anyone who starts being eligible, remains in that situation until death.

Table 3 Alignment table: proportion of people eligible for the THAB/APA, by age and partner status, 2011/2012

Age	No partner	With partner	Total
65-69	0.198	0.070	0.155
70-74	0.203	0.076	0.159
75-79	0.256	0.112	0.209
80-84	0.304	0.173	0.271
85-89	0.374	0.205	0.348
90-94	0.429	0.217	0.412
95+	0.483	0.233	0.478
Total	0.287	0.115	0.241

Note: Estimated on subsample of people below THAB/APA income thresholds
 Source: Datawarehouse data

A complication when estimating the alignment proportions was that in the MIDAS startdata of 2011, we cannot distinguish between income from ARR/IVT, AI/IT and THAB/APA, because all three benefits

¹⁷ Alignment works by assigning each case a score, and then selecting the cases with the highest scores, such that the target proportion or number is reached. The score can be based a substantive variable, a random number, or a mix of these. A *soft take* of certain cases means that those cases are assigned a relatively very high score, such that they are nearly always selected before any others. However, if the number or proportion of those cases in the risk set exceeds the target number or proportion, not all of those cases are selected (this is what makes the *take soft*). See the LIAM2 manual, page 32 (<https://liam2.plan.be/download/LIAM2UserGuide-0.12.0.pdf>)

are included in one variable *uitkering_bb_fod_sz*. The statistics published by the FOD-SZ indicate that in 2014 there were 29,090 ARR/IVT and/or AI/IT receivers of 65 years or older. This is about 1.46% of the 65+ population in 2014. The fact that most users of ARR/IVT and/or AI/IT over 65 receive a pension, as well as the limited amounts of ARR/IVT and/or AI/IT received, suggest that it concerns almost always the AI/IT (a cost-compensation scheme, like the THAB/APA¹⁸), and not the ARR/IVT (an income replacement scheme). Given the limited number of AI/IT beneficiaries and the small amounts concerned, we assume that all amounts included in *uitkering_bb_fod_sz* received by people aged 65 and over is THAB/APA.¹⁹ Therefore, the simulation procedure of the THAB/APA could perhaps be described more accurately as the simulation of the THAB/APA jointly with the AI/IT for those over 65, as it covers implicitly also the latter. This is not ideal, but since the AI/IT is not simulated for the 65+, it does not produce double-counting. Also, the eligibility criteria in terms of dependency, and the amounts are fairly similar. The means-test is rather different, though.²⁰ For brevity, we keep referring to this procedure as the THAB/APA simulation, and the resulting beneficiaries and amounts as THAB/APA beneficiaries and amounts.

Which dependency category of the THAB/APA?

Given eligibility, the second step is that a severity-of-dependency category is assigned. The simulation of the five dependency categories of the THAB/APA is also aligned on proportions by age category and having a partner or not. These proportions were determined by looking at the amounts received within the subsample of persons aged 65+ in the starting data for 2011 and 2012, with incomes below the THAB/APA income thresholds (see below for the income definition and the thresholds). These persons should receive the maximum amounts, so that there should be a one-to-one correspondence between amounts received and the five dependency categories. This means that the unknown classification of recipients in five dependency categories should manifest itself in ‘spikes’ of the observed distribution of the THAB/APA benefit. Graph 6, showing the distribution of the THAB/APA amounts, indeed reveals prominent spikes corresponding to the maximum amounts for the five dependency levels in 2011 and 2012.²¹ The ‘noise’ around those spikes is presumably the result of means-testing for incomes that are not in the Datawarehouse files, of persons receiving the benefit only part of the year, or of THAB/APA receivers changing from one category to another sometime during the year. This noise is filtered out to assign every individual in the dataset to one dependency level on the basis of the spike that their THAB/APA amount is closest to. This gives the proportions of the dependency levels, which are then the basis of the alignment table (Table 4) that show the proportion of the sample across THAB/APA levels, given partner status and age category. The level of dependency increases with age. At younger ages, those with a partner tend to be more severely dependent, but this effect disappears at higher ages.

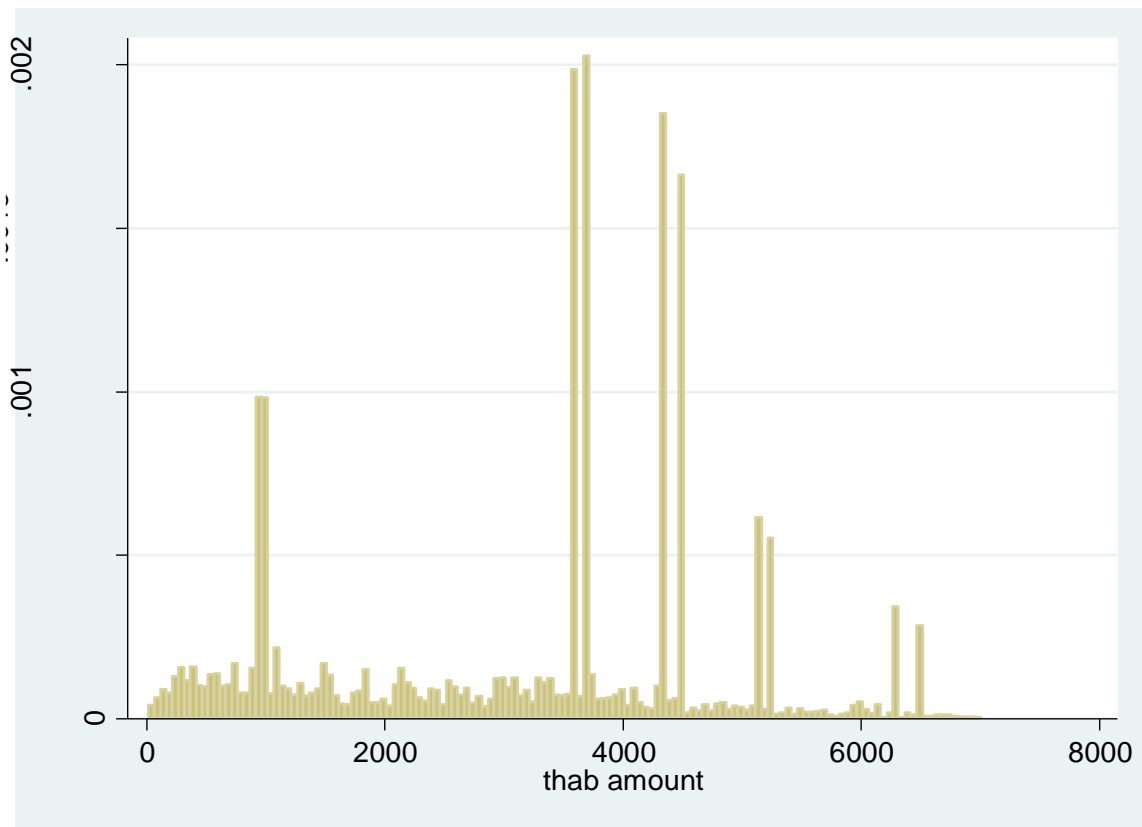
¹⁸ The AI/IT does take into account the taxable income of the recipient, but 1) it includes only taxable income of the person itself and not his or her partner, and 2) this income is only used to lower the maximum amount. The actual amount, given this maximum, is however dependent on a detailed assessment by a medical doctor.

¹⁹ A few persons with benefits over € 6,500 are excluded, as this cannot be THAB/APA, but has to be ARR/IVT and possibly AI/IT.

²⁰ Zie: “Nota Pensioencommissie 2020-2040: Inkomensgarantie voor ouderen: beschrijving van de bestaande regeling en situering binnen het geheel van sociale bijstandsuitkeringen”, 11 december 2013, Technische nota, voor een gedetailleerd overzicht.

²¹ However, because the maximum amounts were adapted to the price index twice both during 2011 and 2012, they are in fact averages across the months of those years.

Graph 6 Distribution of THAB/APA amounts among people receiving the THAB/APA and below the THAB/APA thresholds, 2011 and 2012
Proportions



Note: Range limited to 0-7000
Source: Datawarehouse data

Table 4 Alignment table: distribution of people eligible for the THAB/APA over THAB/APA levels, by age and partner status, 2011/2012
Percentages

Partner status	Age category	THAB/APA level					Total
		1	2	3	4	5	
No partner	65-69	42.4	33.2	18.3	5.1	1.0	100
No partner	70-74	34.4	33.3	17.3	9.2	5.9	100
No partner	75-79	22.3	41.2	26.8	6.6	3.1	100
No partner	80-84	17.6	38.6	31.2	9.0	3.6	100
No partner	85-89	9.6	34.2	39.5	9.9	6.7	100
No partner	90-94	5.9	35.0	36.8	15.6	6.7	100
No partner	95+	2.1	22.0	40.6	19.0	16.4	100
Partner	65-69	30.4	38.1	22.5	9.0	0.0	100
Partner	70-74	32.0	23.4	26.9	14.0	3.8	100
Partner	75-79	14.3	41.3	28.3	9.1	7.0	100
Partner	80-84	16.6	36.8	28.6	11.4	6.6	100
Partner	85-89	21.2	25.3	34.9	15.6	3.0	100
Partner	90-94	2.6	38.9	33.3	25.1	0.0	100
Partner	95+	0.0	0.0	26.4	73.6	0.0	100
Total		16.5	35.8	31.8	10.3	5.5	100

Note: Estimated on subsample below THAB/APA income thresholds
Source: Datawarehouse data

In the simulation, the THAB/APA level is determined with alignment, in steps going from the highest (5) to the one-but lowest level (2). In the simulation of any of these levels, a *soft take* assigns a preferential selection possibility to those persons that were already on that level, or a higher one, in the

previous period. Those that are eligible and not assigned to levels 2 to 5 go into the lowest level 1. Given that the THAB/APA level increases with age, and that those at high age have a relatively high probability to die, this means that the THAB/APA level functions virtually as a ratchet, with many more moves up than down, as shown in Table 5²².

Table 5 Transition table of THAB/APA levels from one year (t) to the next year (t+1), 2012/2020
Percentages

THAB/APA level year t+1	THAB/APA level year t					
	Not eligible	1	2	3	4	5
Not eligible	94.9	23.1	0.4	0.0	0.0	0.0
1	2.1	27.6	4.6	1.3	0.4	0.0
2	0.6	43.4	62.6	2.7	0.6	0.0
3	0.0	0.1	21.9	73.9	2.7	0.1
4	0.0	0.0	0.0	9.4	74.4	2.3
5	0.0	0.0	0.0	0.0	8.7	82.2
Died	2.4	5.9	10.5	12.7	13.1	15.4
Total	100	100	100	100	100	100

Source: MIDAS simulations (20% sample)

In the initial year 2011, the THAB/APA level was derived from the observed THAB/APA amounts and a simulation of the means test (see below). As shown above, this is rather straightforward for those who are below the thresholds, so that the amounts they receive are equal to the maximum amounts and not reduced by the means test. For those whose incomes are above the thresholds, but still are THAB/APA beneficiaries, an attempt was made to reconstruct the maximum possible amount (before the means-test was applied), but this did not produce credible results. Instead, the THAB/APA levels of people with incomes above the THAB/APA threshold were set in such a way that the overall distribution of THAB/APA levels conformed closely to the administrative figures for 2011.

Simulation of the THAB/APA amount

Given the THAB/APA level, the THAB/APA amount can be simulated. The income thresholds and maximum amounts are presented in Table 6. The means test proceeds by adding up the incomes of the potential beneficiary, as well as those of their partner. Pensions are included only for 90% (though the IGO/GRAPA for 100%). The means test also considers the imputed income from having a home, as well as sales and gifts of property during the last 10 years; but we ignore these as we have no data available on these sources of income. From the total income thus calculated, the appropriate income threshold is deducted. The remainder (if greater than 0) is deducted from the maximum amount; the amount is of course 0 when the remainder exceeds the maximum amount. The system distinguishes 3 categories of potential recipients, each with their own income threshold. Category A are those that are married or cohabiting²³, and both have an income. Category B are singles, and category C includes those that are married or cohabiting, and only one partner has an income. In case that both partners in a couple are

²² In the starting data we do not observe such a pattern when comparing the THAB/APA level of individuals between 2011 and 2012. Instead, they either stay in the same category, or they move outside any category. Presumably, this pattern comes about because no or very few persons change category on 1 January, but at some other point during the year, so that the amount received during a calendar year is not equal to any of the maximum amounts, but falls within the 'noise'.

²³ Relatives in the 1st, 2nd or 3rd degree are excluded from this.

eligible for the THAB/APA, threshold category C applies, and the remainder income is split equally across both partners.²⁴

Table 6 THAB/APA maximum amounts by level, and income thresholds in 2011

	Amount in €
Maximum amount level 1	935
Maximum amount level 2	3,575
Maximum amount level 3	4,355
Maximum amount level 4	5,125
Maximum amount level 5	6,295
Income threshold cat. A (partnered, both partners receive income)	12,000
Income threshold cat. B (single)	12,000
Income threshold cat. C (partnered, only one partner receives income)	14,995

Source: FOD SZ, <https://handicap.belgium.be/nl/contact/publicaties/index.htm>

The THAB/APA being a cost-compensation scheme, the amounts have been adapted only to prices since 2011. In accordance with the medium-term assumptions in HERMES and the long-term assumptions in MALTESE, it is assumed that this policy will continue until the end of the simulation period. As MIDAS runs in prices of 2011, the 2011 amounts are used throughout the simulation period. The treatment of the thresholds is different. These thresholds were increased somewhat in real terms between 2011 and 2016. Moreover, a simulation revealed that keeping them constant in real terms throughout the simulation period would mean that by 2070 hardly any beneficiaries were left. This is due to the real increases in pension in general and in minimum pensions and the IGO/GRAPA in particular, so that the means test excluded nearly all older people. In fact, over time the IGO/GRAPA maximum amounts would start to exceed the THAB/APA thresholds. The virtual disappearance of the scheme seems inconsistent with the general assumption of unchanged policy that is a guideline for all projections. Therefore, and somewhat arbitrarily, the thresholds are kept constant in real terms until 2025, and increased by 1% in every later year, in line with the assumptions made for the uprating of the IGO/GRAPA and the minimum pensions.

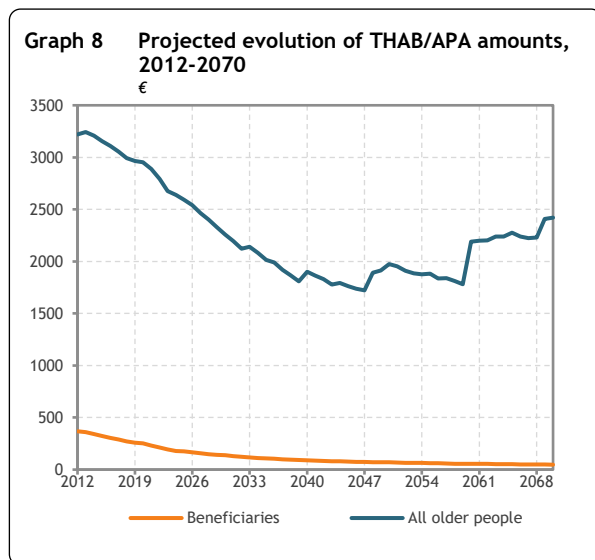
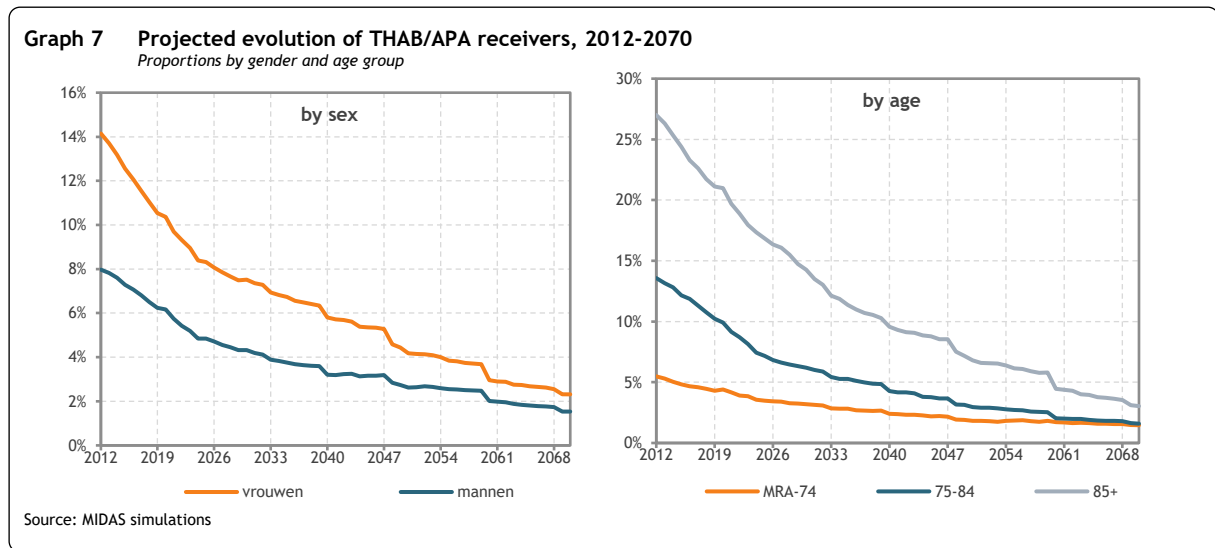
The THAB/APA is not liable for tax, so it is ignored in the tax module, and simply added to net income.

Projection results

Graph 7 shows the simulated proportion of THAB/APA receivers, by gender and by age category. Proportions are higher among women than among men, and increase with age. The proportions for both women and men first drop rather steeply until 2025, and then increase until about 2050, after which it declines again. For the older people below 75 it remains stable throughout the projection period, though. Since, by construction, the rates of dependency remain constant by age and partner status, the changes in the proportions of THAB/APA receivers must be due in large part to developments in the income situation of those age groups in that period. The drop during the first part of the projection period until 2025 can be attributed to the fact that the thresholds remain constant in real terms while pensions, and in particular minimum pensions and the GRAPA/IGO improve during this period. The subsequent increase until around 2050 is partly due to the strong growth in the number of the oldest old (85+) in that period, among whom the rate of dependency is highest. Among the 85plus themselves, the proportion

²⁴ The latter step is derived from the 'THAB/APA simulator', which can be found on <https://handicap.belgium.be/nl/mijn-rechten/tegemotkoming-hulp-bejaarden.htm>

of THAB/APA beneficiaries rises because at first the number of new beneficiaries is higher than the outflow through death. Once the age distribution in this group reaches equilibrium, deaths balance new entries.



Graph 8 shows that the average THAB/APA amount received by beneficiaries declines somewhat until 2020, and more steeply between 2020 and 2026, as increasing pensions have a downward effect on the amounts due to the means test. The additional increases of the GRAPA/IGO between 2021 and 2025 play a particularly important role. The average benefit declines more slowly during the remainder of the projection period. As a percentage of the beneficiaries' pension, it drops from 28% in 2012 to 19% in 2025, and further to 8% in 2070. Recall that the THAB/APA amounts are kept constant in real terms. Averaged over all people above retirement age, this benefit also declines rather quickly until 2025 and more gradually after that year. As a percentage of the average pension, there is continuous decline from 3.1% in 2012 to only 0.5% in 2070, indicating that the role of the THAB/APA in the incomes of the older population will become smaller over time.

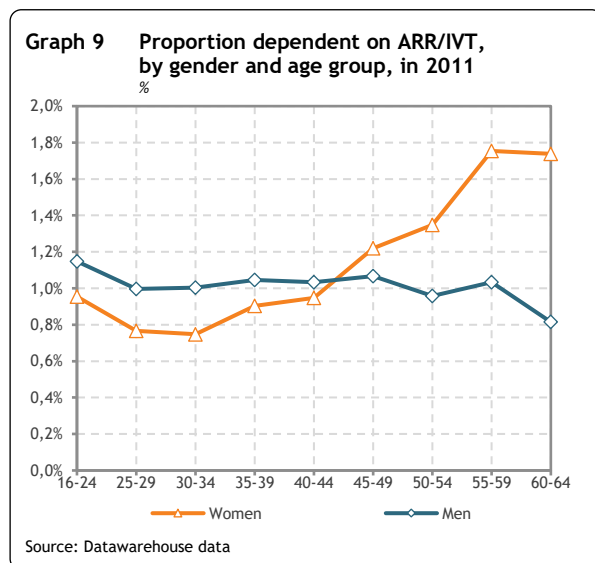
As a percentage of the average pension, there is continuous decline from 3.1% in 2012 to only 0.5% in 2070, indicating that the role of the THAB/APA in the incomes of the older population will become smaller over time.

4. Simulation of the IVT/ARR

Simulation of the beneficiaries

As explained in the introduction of this note, the “Inkomensvervangende tegemoetkoming” (IVT) / “Allocation de remplacement de revenu” (ARR) is an allowance that is intended to support persons younger than 65 who are disabled and are not entitled to a social insurance invalidity benefit, mostly because they have never been able to work. It is means-tested against the income of the person and their partner. Mainly due to lack of data, the simulation of the IVT /ARR remains rather elementary. The objective is to simulate the number and profile of those for whom the ARR/IVT is the only source of income. In the starting data, these are identified by code 38 in the ‘nomenclature’ of the socio-economic position. Being dependent on the ARR/IVT as indicated by this code in the ‘nomenclature’ is a rather stable situation. Of the persons who were in this status in 2011, 96% stayed in this state in 2012 (unless they reached the age of 65).²⁵ Conversely, 92% of those dependent on the ARR/IVT in 2012 were already in this state in 2011. Of the remaining 8% who entered ARR/IVT, 30% were dependent children in 2011 and 62% were in the state of “other non-active”. There are very few transitions between ARR/IVT dependency and work or social security benefit receipt.²⁶

Graph 9 shows the proportions dependent on ARR/IVT, by age category and gender. These proportions are used as alignment targets during the whole of the projection period. The proportions of the group



60-64 are also used for the 65 and 66-year olds in the years after the retirement age is extended in 2025 and 2030.

In the simulation, there are two ways to enter ARR/IVT status. For some people, entry is determined at age 15, before persons can start working or receive a benefit, or can have a partner. The only variable taken into account is the simulated eventual education level (i.e. the education level that will be attained at 25²⁷). At any simulation period, the number of these new entrants in the current cohort of 15-year olds is aligned to the proportions of ARR/IVT receivers among those aged 25-49 observed in the starting dataset, by education level. These proportions are as follows: at most lower secondary education: 2.5%; higher secondary education: 0.5%; higher education: 0.05%. These persons are assumed to remain ARR/IVT receivers until they retire.

observed in the starting dataset, by education level. These proportions are as follows: at most lower secondary education: 2.5%; higher secondary education: 0.5%; higher education: 0.05%. These persons are assumed to remain ARR/IVT receivers until they retire.

²⁵ 95% of those who stopped being dependent on the ARR/IVT became ‘other non-active’.

²⁶ Note that this conclusion applies only to those for whom ARR/IVT is the only source of income. For some ARR/IVT beneficiaries, the ARR/IVT is a complement to other income sources.

²⁷ In MIDAS, for individuals who have not left education at the start of the simulation, the highest education level that they will eventually attain is determined at that start. Similarly, for new births, the highest education level they will obtain eventually is set at birth. The separate steps in the schooling career are not modelled. See the section on education in the report on the demographic modules.

Other people can enter ARR/IVT status later in life. A score is calculated to determine who is most likely to become ARR/IVT beneficiary, using a logit equation of the transition into this state that included education level (people with higher education are much less likely to enter ARR/IVT status) and LIPRO-position (people who are single or live with their parents are more likely to become ARR/IVT beneficiary), see Table 7. Given the transition patterns reported above, only people who are dependent children in education or who have the status of ‘other non-active’ can become ARR/IVT beneficiary. These ARR/IVT beneficiaries can exit this status until age 30, although there is a ‘soft take’ of people dependent on the ARR/IVT in the previous year. After age 30, there is no exit from ARR/IVT until retirement.

Furthermore, it is assumed that ARR/IVT receivers do not marry or cohabit (in the starting data only 15% are part of a couple) and that they do not get children (only 13% in the starting data). So their LIPRO-positions are limited to single, child and other. However, a small number of people with a partner and/or children become ARR/IVT beneficiaries.

Table 7 Logistic regression of transition into the ARR/IVT between 2011 and 2012

	Coefficients	St. errors	p.
education_level			
ISCED 0-2 (lower secondary or less)	ref. cat.		
ISCED 3-4 (higher secondary)	-0.07	0.21	0.734
ISCED 5-6 (higher education)	-1.15	0.35	0.001
LIPRO position is child	1.24	0.41	0.002
LIPRO position is single	1.95	0.19	0.000
constant	-5.57	0.16	0.000

Note: estimation sample limited to persons between 30 and 65 who were “other inactive” in 2011.

Simulation of ARR/IVT amounts

The maximum possible ARR/IVT benefit depends on the household and care situation of the beneficiary. Three categories are used.²⁸ The amounts mentioned below were valid on 1/1/2021:

- Category A: living together with relatives (up to the 3rd degree): € 7,879.96
- Category B: single, or living in a special institution (even when keeping the domicile with relatives): € 11,819.95
- Category C: living with a non-relative (assumed partner) or a dependent child (or living in an institution, while keeping the domicile with a non-relative): € 15,973.93

In accordance with the standard assumptions about lump-sum benefits used by the Study Commission on Ageing, the amounts mentioned above are increased each year by 1% in real terms. Between 2012 and 2020 the actual amounts, deflated by the CPI, were used.

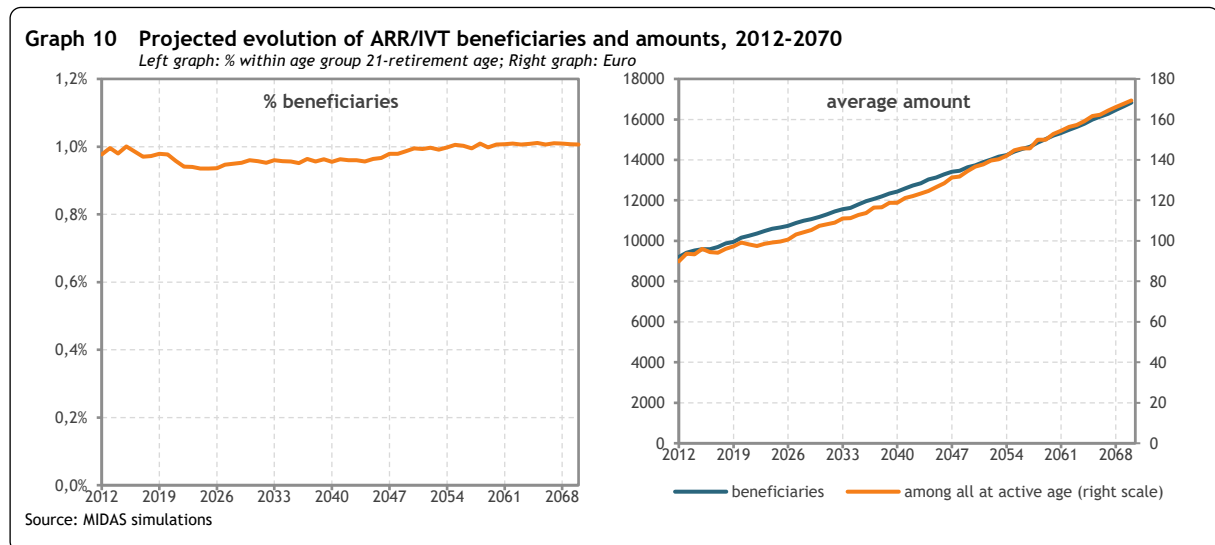
The benefit is means-tested against the taxable income of the fiscal unit of the ARR/IVT receiver. This implies that only the own income and that of the possible partner is taken into account, but not that of the parents or other relatives in the household. Given the assumptions made in the projection about transitions into and out of ARR/IVT status, which imply that beneficiaries have no other income, the means test only impacts the small and declining number of ARR/IVT beneficiaries with a partner.

²⁸ <https://handicap.belgium.be/nl/mijn-dossier/gezinscategorie.htm>

Strikingly, in the starting data of 2011, the average and median benefits received from the FOD-SZ by ARR/IVT-beneficiaries who had lipo-positions ‘child’ or ‘other’ were over € 12,000, which is much higher than the ARR/IVT maxima for category A in that year. This is possible if they had a large “Integratietegemoetkoming” (IT) / “Allocation d’intégration” (AI) in addition of the ARR/IVT benefit, and/or were in fact in benefit category B (the latter is possible if they were in an institution while keeping the domicile with their parents or other relatives). In order not to deviate too much from these amounts in the projection, we assume that all ARR/IVT beneficiaries, irrespective of their household position, receive the amount corresponding category B, except those with a partner, who are in category C.

Projection results

Graph 10 shows that the simulated proportion of ARR/IVT beneficiaries among all persons between 21 and retirement age is nearly constant across the projection period at about 1% of this group. Also, by age group (not shown in graph) the proportions follow a flat trajectory between 2012 and 2070. The average amount received by beneficiaries increases by about 1% per year, in accordance with the standard assumption about lump-sum benefits used by the Study Commission on Ageing, as explained above. Between 2012 and 2020 the average actual amount, deflated by the CPI, is used. Averaged over all persons in the age range 21 - retirement age, the amount increases at the same rate, from € 94 in 2012 to € 167 in 2070.



5. Simulation of career interruption benefits

Simulation of beneficiaries

There are many kinds of career interruptions (as mentioned in the Introduction), each with their own conditions, maximum durations and target groups. These leaves can be full-time, but are more often part-time, for 20% or 50%. The RVA / ONEM pays a lump-sum benefit to partially compensate for the loss in earnings. The benefit does not depend on the previous earnings, but varies by kind of leave, family situation and seniority. We refer to the RVA / ONEM website for an overview and details: <https://www.rva.be/nl/burgers/loopbaanonderbreking-tijdskrediet-en-thematische-verloven>.

Given the variety of career interruption benefits (CIBs), it seemed difficult to try to simulate all these arrangements and the associated benefits according to the legislation. Therefore, the simulated amounts were derived from the observed benefits in the starting data, as described below.

In a first step, the usual procedure with alignment using a table derived from the starting dataset and a probability score through a logistic regression was used in the projection to identify persons taking up a CIB. In the starting data, persons receiving CIBs from the RVA/ONEM could be identified on the basis of the variable 'fiche7'.²⁹ Quarterly data were aggregated to a yearly value by using the mode across quarters (and the minimum code in case of ties). In fact, in case a CIB was received in one quarter, it tended to be received in all quarters. For a large majority, the number of days of the benefit was 312, i.e. it was received during the whole year.

Of all CIB beneficiaries, 80% were employed part-time, while 10% were in another socio-economic position during most of the year (unemployed, disabled etc.), and another 10% were employed full-time (among the latter, the average number of days of CIB was only 121, which suggest short interruptions). Also, Datawarehouse data³⁰ show that few people take full-time career interruption (only 11 939 in the first quarter of 2019, while 228 956 wage-earners took part-time career interruptions.) For this reason, the simulation of the CIB is limited to part-time employees, including civil servants. Also, for part-time employees, the calculation of earnings takes is based on the limited number of hours, and therefore simulation of the CIB constitutes a clear improvement of the projected incomes of these persons, as it comes on top of simulated earnings.³¹ The simulation of CIBs takes as given the part-time status and number of working hours of possible beneficiaries, as they are simulated by the labour market modules.

For the population of part-time employees, an alignment table was constructed from the starting dataset with proportions distinguished by gender, age bracket (5-year intervals) and part-time work proportion. The latter was derived by expressing the reported number of work hours during the year as a proportion of full-time hours (set at 1976), and bracketing this into five categories: 90% or more (full-time or nearly full-time)³²; 75-90% (about 4/5); 60-75% (between half-time and 4/5); 40-60% (about half-

²⁹ Codes 200 to 252 indicate a career interruption benefit.

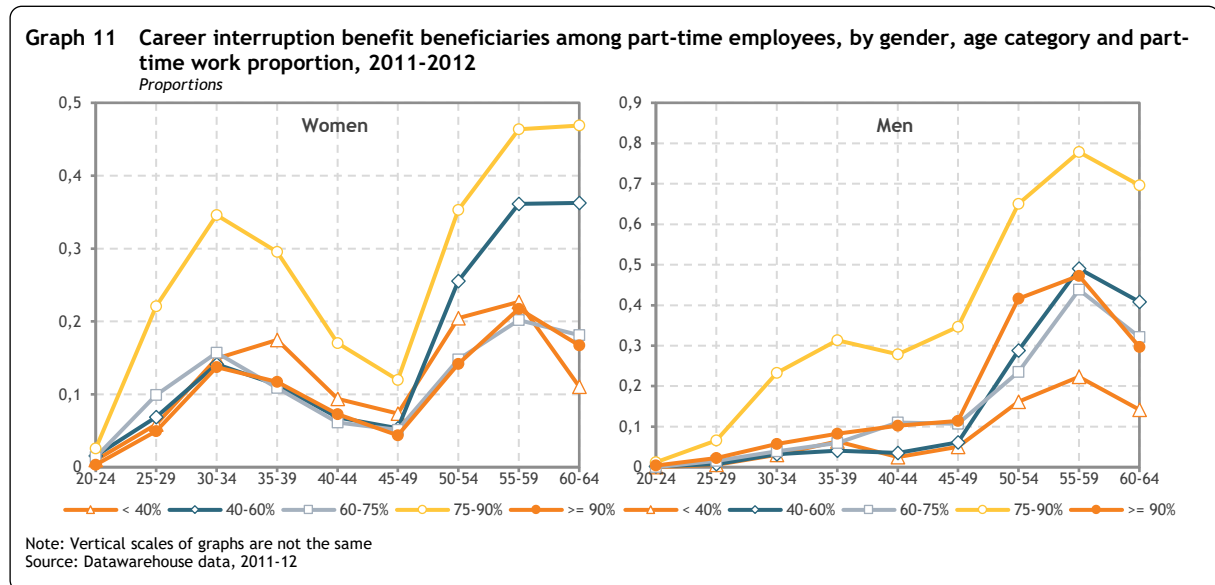
³⁰ On-line global numbers:

https://dwh-live.bcss.fgov.be/nl/dwh/dwh_page/content/websites/datawarehouse/menu/webtoepassing-globale-cijfers.html

³¹ For other categories (full-time employed persons, unemployed and disabled persons) the calculation of earnings or benefits is based on the assumption that this situation applies full-year and full-time.

³² Some people who are registered as working part-time still work around 1976 hours or even more.

time) and less than 40% (less than half-time). Among part-time workers, 4/5s and half-time are the most popular arrangements, together accounting for 59% of the total group. Graph 11 shows that the age pattern of CIB is clearly different for men and women. Among men, the highest incidence is around age 60, while among women relatively high proportions are also observed at the ages of child-birth and child-care. For both genders, CIB is more often received when people work around 4/5th of FT work than at other proportions. When considering these figures, it has to be taken into account that the prevalence of part-time work is much lower among men than women: 16% against 52%. In future years when the standard age of retirement will be increased to 66 and 67, the alignment proportions of the age bracket 60-64 are also used for ages 65 and 66.



The score, which determines the ranking of persons to be selected for a CIB, given the alignment table, was calculated with separate logit equations for each gender, that included the following variables: age category, whether the person had a CIB in the previous year (interacted with the age category), the workstate and part-time status of the partner, the number of children aged 0-3 years (interacted with having a partner), the number of children aged 4-6 years, education level and the part-time work proportion in the five categories mentioned above. There is a strong cross-year correlation in CIB: 68% of the part-time working CIB receivers in 2012 were in the same status in 2011. This correlation increases with age. For women, the presence of young children is also very important.

The total number of years that people can receive a CIB before age 60 is limited to 5.³³ From age 60 on, employees can enter the system of “Tijdscrediet – landingsbaan”³⁴ and stay in this until they retire, irrespective of how much career break they have taken up earlier in their lives. For those who were receiving a CIB in the starting year, the starting number of years was set in a random way at 1, 2 or 3 with a probability of 1/3 for each. For people below 30, it was fixed at 1 year, though.

³³ This is a bit arbitrary. “Tijdscrediet-algemeen” is limited to 51 months, irrespective of the reduction of working time (1/5 or 1/2). The duration of “Loopbaanonderbreking met motief” (which can be taken after or before “Tijdscrediet” depends on the reason and is inversely proportional to the proportion. For the most common “motief” (reason) it is 8 months if half-time and 20 months if 1/5.

³⁴ Literally ‘landing strip’. This is a system supposedly making it easier to continue working, instead of retiring early.

Simulation of amounts

As mentioned above, due to the variety of CIB options, it was not attempted to simulate the CIB amounts according to the regulations. Instead, these were derived from the observed amounts in the starting data. Some exploratory analysis led to the regression equation reported in Table 8. The most important variables were age group (below age 50, or 50 or over) and the part-time work proportion in five categories. Gender had a small impact. In the projection, the outcome of this regression equation was updated to future years by 1% per year in real terms, in accordance with the standard assumption about lump-sum benefits used by the Study Commission on Ageing. The implicit assumption of this procedure is that in future years, the profile of beneficiaries in terms of arrangement used will remain the same, by age group and part-time work proportion.

Table 8 Regression of career interruption benefit on background variables, 2011

	Coefficients	St. errors	t	P>t
age < 50	ref. cat.			
age >= 50	1,519	70	21.80	0.000
part-time prop. cat.				
< 40%	ref. cat.			
40-60%	-88	57	-1.54	0.123
60-75%	-831	64	-12.93	0.000
75-90%	-1,029	51	-20.07	0.000
>= 90%	-1,076	70	-15.44	0.000
age>= 50 * part-time prop. cat.				
40-60%	753	80	9.42	0.000
60-75%	-14	94	-0.15	0.881
75-90%	-463	75	-6.20	0.000
>= 90%	-255	95	-2.68	0.007
couple	ref. cat.			
1.single	178	25	7.03	0.000
employee private sector	ref. cat.			
public sector (inc. civil servants)	-560	20	-27.41	0.000
constant	2,682	49	54.53	0.000

Notes: $R^2 = 0.608$; $n = 9,400$, $df = 11$; ref.cat: reference category; estimation sample limited to people with CIB who work part-time

Source: Datawarehouse data

Persons with CIB build up pension rights for the working days that are foregone. These are based on the last year of earnings before the reduction of working time. In case of "tijdskrediet-landingsbaan" (which is assumed for all career breaks from age 60 on), after two years in case of halftime career interruption and five year in case of 1/5 interruption, pension credits are based on the "limited last earnings"

with the minimum wage as a ceiling. The proportion of working time to which the CIB applies is set at 1/5 if people are working part-time for at least 60%, and 1/2 otherwise.

Projection results

Graph 12 shows the projection of Career Interruption Benefits (CIBs). The proportion of beneficiaries among part-timers (and also more generally among people at active age) is virtually stable around 18%. The average amount received by beneficiaries increases by about 1% per year, in accordance with the standard assumption about lump-sum benefits used by the Study Commission on Ageing. Averaged over all persons at active age, the trend is quite similar, given the nearly constant proportion of beneficiaries.

The projected assimilated earnings, from which pension entitlements are calculated, rise from € 10,700 to € 20,300 in 2070 for CIB beneficiaries, the corresponding figures averaged over all persons at active age are € 360 and € 771. The increase is slow at first, but then accelerates until about 2040, following the assumed trend in the average wage increase. The bottom graph in Graph 11 shows the resulting increase in pension rights at the Standard Pensionable Age (i.e. the total earnings, including assimilated earnings, over the career, adjusted for ceilings and minima). Only increases that are obtained after 2011 are registered, which implies that until the mid-2050s they cover only part of the career. This is why the curves become stable only after that time. Pension rights linked to CIBs obtained before 2012 are in fact included in total (assimilated) earnings, but for those years we cannot distinguish those rights from other ones.

Graph 12 Projected Career Interruption Benefits, 2012-2070
 Left graph: proportions. Right graph: Euros, bottom graph: % change



Source: MIDAS simulations
 Note: active age = 18 - SPA (Standard Pensionable age)

6. Conclusion

This note presents the projection of four relatively small income components: the “Garantie de Revenus aux Personnes Âgées” (GRAPA) / “Inkomensgarantie voor Ouderen” (IGO), the “Tegemoetkoming voor hulp aan bejaarden” (THAB) / “Allocation pour l'aide aux personnes âgées” (APA), the “Inkomensvervangende tegemoetkoming” (IVT) / “Allocation de remplacement de revenus” (ARR), and the career interruption benefits paid out by the RVA / ONEM. The projection of the GRAPA/IGO has been thoroughly revised, while the other components appear for the first time in MIDAS.

The maximum amounts of the GRAPA/IGO will rise strongly between 2020 and 2024, and will continue to increase by 1% in real terms throughout the projection period. Nevertheless, the projected proportion GRAPA/IGO beneficiaries among older people rises only moderately until about 2030, and declines gradually afterwards. The main reason for the limited variation in the proportion of GRAPA/IGO beneficiaries across years is that the improvement of (minimum) retirement and survivor pensions keeps pace with the rises of the GRAPA/IGO. The GRAPA/IGO has a substantial and somewhat increasing downward impact on the poverty risk of the older population: it amounts to 3.0 pp in 2012, and the impact rises to a maximum of 3.7 pp around 2030.

The income thresholds used for the means-test of the THAB/APA are assumed to rise by 1% per year in real terms (after 2025), in line with the IGO and the minimum pensions. The proportion receiving the THAB/APA among the older population first declines, then increases because of the growth in number of the oldest old, and then stabilizes. Because the THAB/APA benefit amounts are updated only to prices, they become less important as an income source for older people. The impact of the THAB/APA on the poverty risks of the population above standard pensionable age is substantial at first, reducing it by 2.2 percentage-points, but declines afterwards to about 0.5 percentage point by 2070.

The thresholds and benefit amounts for the ARR/IVT and the career interruption benefits are assumed to increase in real terms following the assumptions of the Study Committee on Ageing. The proportion of people at active age receiving the ARR/IVT is constant around 1%, while the amounts received are increased in real terms by 1% per year. Between 3% and 4% of the active age population obtain a career interruption benefit, which also increases in real terms by 1% per year. The ARR/IVT and the career interruption benefits have quite small impacts on poverty risks.

Appendix

In the starting data, some older persons (65plus) have no retirement pension, nor any survivor pension. Since these cases with zero pension are obviously important in an analysis and projection of income inequality and the risk of poverty among the older population, it is useful to take a closer look at them.

In the starting data, referring to 2011, 12.1% of all people over 65³⁵ have no retirement pension, nor any survivor pension. A large majority in this group (86.5%) have a partner with a pension; almost all of these are women. Presumably, these are people with no or small pension rights, whose partner receives a family pension. A small group has a partner aged less than 65.

We continue with the much smaller group (1.5% of the population 66plus) who do not have a pension of their own, and who either do not have a partner (90.0% of this group), or have a partner who is also over 65 and lacks a pension (10.0%). About 4% of this group are still at work, mostly as employees, even though they are 65 or over. They may have postponed applying for their pension, or their pension may have been reduced to zero because of high earnings.³⁶ The remainder have apparently not build up any pension rights within the Belgian public system. A first reason for this can be that they immigrated after their active life. In fact, of this group, 58% are immigrants (i.e. were born outside Belgium), and the majority of them (52%) entered Belgium only after retirement age, and another 30% after age 54.

A second reason can be that they never had any earnings or assimilated earnings in Belgium, even though they lived all or most of their life in this country. We have retrospective data on the earlier socio-economic position (“nomenclature”) of people, but these go back to 2003 only, which implies that we can assess this position at age 64 only for those aged 72 or younger, who form 37% of the total group. Within this – selective – group (from which immigrants after age 60 are also excluded), 11% were working (as employee or self-employed), 16% were retired, 14% were on social assistance, 17% were receiving the IVT/ARR, while another 16% had the status of dependent child for which family allowances were paid. The latter status is only given to persons over 25 if they are disabled (and living with a parent or guardian). It is probable that during active age they were never able to work, and so could not obtain earnings or social insurance benefits at any point in their lives. As explained above, this is likely to be also true for a majority of the IVT/ARR beneficiaries. Finally, 25% were labelled “other non-active”, which implies that they had no (registered) earnings and no social benefit of any kind in that year.

Even though only a few people in this group have partners, this does not mean that most are living as singles, in the sense of one-person households. That is only the case for 23% of them. 42% (60% among immigrants) are “others”, i.e. unrelated to the household reference person, or related to them in another way than as a partner or a child. A large proportion (23%, 50% among non-immigrants) are living in an institution.

In summary, the primary reason why individuals or couples over 65 do not receive any pension is that they immigrated after or close to retirement age. A second important reason seems to be that they were

³⁵ We take this age limit, because the age is registered at the end of 2011, the income reference year. So people who are 65 were in 2011 retired for only part of the year.

³⁶ Cumulation of a pension and earnings was limited in 2011.

never able to work before 65 due to disability. The large proportion living in an institution or with relatives or non-relatives also suggests this.

Finally, one would expect that take-up of the IGO would be high in this group. Nearly all are eligible, except for a few individuals who immigrated very recently (note that the current residence requirement of 5 years for non-EU citizens was not yet in effect in 2011). In fact, it is far from complete: overall only 34%. It is particularly low among immigrants from outside the EU (6%), somewhat higher among EU-citizens (21%), while it reaches 52% for the Belgian-born in this group. Pensions from foreign sources (in so far as these are registered by the Federal Pension Service) are not the reason for this low take-up: only 11% of EU-citizens receive such pensions, 8% of the Belgian-born, and less than 1% of the non-EU citizens. Controlling for immigrant status, the living situation does not play a role. Take-up is also about half for people in collective households.

It is not clear whether this incomplete take-up is due to income sources that are not registered in our data, or is true non-take-up of people who are in fact entitled to the GRAPA/IGO. Analyses by De Vil and Van den Bosch using EU-SILC data suggest that the 'cadastral income' from the own home is very seldom sufficient to lift those with very low pensions to the level of the GRAPA/IGO thresholds. On the other hand, a few people in this group have quite substantial financial capital and might be rentiers.³⁷ As regards 'true' non-take-up, it is suggestive that take-up in this group is considerably lower than among those with positive but very low pensions (where it is close to 60%, see above). The fact that the latter are in contact with Federal Pension Service, while the former may be unknown to that agency, could be a significant factor.

³⁷ De Vil, G. and Van den Bosch, K. (2013), De evolutie van de armoede bij ouderen nader bekeken, Working Paper 09-13, Brussels, Federaal Planbureau, p. 29-32.



**Federaal
Planbureau**

Economische analyses en vooruitzichten

WORKING PAPER 2-23 - Annex 6

MIDAS: Geregionaliseerde kinderbijslag

Annex 6 bij 'MIDAS 2.0: Revisie van een
dynamisch microsimulatiemodel'

Januari 2023

Gijs Dekkers, gd@plan.be

Belliardstraat 14-18
1040 Brussel

e-mail: contact@plan.be
<https://www.plan.be>

Inhoudstafel

1. Inleiding	1
2. De hervorming van de stelsels van kinderbijslag	2
2.1. Het stelsel vóór de hervorming	2
2.2. De verschillende stelsels na de hervorming	3
2.2.1. Vlaanderen	3
2.2.2. Wallonië	7
2.2.3. Brussel	9
2.2.4. Indexering	10
3. Resultaten	11
Referenties	14

Lijst van tabellen

Tabel 1	Bedrag sociale toeslag per kind in Brussel	10
---------	--	----

Lijst van figuren

Figuur 1	Gemiddelde uitkering groeipakket/kinderbijslag; België en naar regio	11
Figuur 2	Gemiddelde bedragen per kind naar basisbedrag en toeslagen, per gewest	12

Federaal Planbureau

Het Federaal Planbureau (FPB) is een instelling van openbaar nut die beleidsrelevante studies en vooruitzichten maakt over economische, socio-economische en milieuvraagstukken. Daarnaast bestudeert het de integratie van die vraagstukken in een context van duurzame ontwikkeling. Het stelt zijn wetenschappelijke expertise onder meer ter beschikking van de regering, het Parlement, de sociale gesprekspartners, nationale en internationale instellingen.

De werkzaamheden van het FPB worden steeds gekenmerkt door een onafhankelijke benadering, transparantie en aandacht voor het algemeen welzijn. De kwaliteit van de gegevens, een wetenschappelijke methodologie en de empirische geldigheid van de analyses staan daarbij centraal. Tot slot zorgt het FPB voor een ruime verspreiding van de resultaten van zijn werkzaamheden en draagt zo bij tot het democratisch debat.

Het Federaal Planbureau is EMAS en Ecodynamische Onderneming (drie sterren) gecertificeerd voor zijn milieubeheer.

<https://www.plan.be>

e-mail: contact@plan.be

Overname wordt toegestaan, behalve voor handelsdoeleinden, mits bronvermelding.

Verantwoordelijke uitgever: Saskia Weemaes

Wettelijk Depot: D/2023/7433/3

Federaal Planbureau

Belliardstraat 14-18, 1040 Brussel

tel.: +32-2-5077311

e-mail: contact@plan.be

<https://www.plan.be>

MIDAS: Geregionaliseerde kinderbijslag

Annex 6 bij 'MIDAS 2.0: Revisie van een dynamisch microsimulatiemodel'

Januari 2023

Gijs Dekkers, gd@plan.be

Abstract – Deze annex presenteert de revisie van de module voor kinderbijslagen, naar aanleiding van de regionalisering van deze uitkering¹.

¹ Hartelijk dank aan Hendrik Nevejan voor zijn commentaren op een eerdere versie van deze tekst.

1. Inleiding

Deze tekst beschrijft de nieuwe regionale modules voor de simulering van de kinderbijslag in MIDAS. De zesde staatshervorming had onder meer tot gevolg dat de federale bevoegdheid voor kinderbijslag naar Brussel en de Gemeenschappen² ging, waarbij de Franse gemeenschap deze bevoegdheid overdroeg naar het Waalse Gewest. Het Vlaamse Groeipakket was vanaf 2019 van toepassing, terwijl de stelsels voor kinderbijslag van Brussel en Wallonië vanaf 2020 in werking traden³.

De vorige versie van het model MIDAS bevatte de federale kinderbijslag zoals deze voor de hervorming van toepassing was. De nieuwe versie van het model maakt een onderscheid naar regio's waardoor het nu mogelijk is om de afzonderlijke stelsels van kinderbijslag of Groeipakket te simuleren.

In deze appendix zullen deze verschillende stelsels worden beschreven. Hierbij is het vooral de bedoeling om te bespreken hoe ze in het model MIDAS zijn gemodelleerd. Aangezien MIDAS een dynamisch model is, met als uiteindelijke bedoeling de projectie van het armoederisico, is het niet mogelijk om regelingen of voorwaarden op te nemen die niet in het model onderscheiden worden. Kinderen die in het buitenland verblijven vanwege de detachering van hun ouders of opvoeders, kunnen bijvoorbeeld maximaal één jaar de kinderbijslag behouden. Maar MIDAS modelleert geen detachering, en dus negeren we deze regel.

Wanneer de regels in vereenvoudigd vorm worden toegepast, wordt dit eveneens expliciet gemaakt. Bijvoorbeeld, in Vlaanderen zijn in principe beide ouders de begunstigden van het Groeipakket, en zij bepalen aan wie het bedrag wordt uitbetaald. In MIDAS wordt daarentegen verondersteld dat de begunstigde de oudste is van de ouders in welk gezin de kinderen wonen⁴. In geval van scheiding worden in het model de kinderen aan het huishouden van ofwel de moeder, ofwel de vader toegevoegd (er is geen co-ouderschap in MIDAS), waardoor die de rechthebbende op de kinderbijslag-uitkeringen wordt.

Als laatste nog de opmerking dat de kinderbijslagen zowel bestaan uit eenmalige bedragen (bijvoorbeeld de geboortepremie) als maandbedragen. Aangezien MIDAS een jaarmodel is, worden alle maandbedragen omgezet naar jaarbedragen.

² Binnen de simulaties van de laatste versie van MIDAS wordt een onderscheid gemaakt naar de Gewesten, en niet de Gemeenschappen. Hierdoor is het niet mogelijk om regelingen afzonderlijk voor de Duitstalige Gemeenschap te simuleren.

³ In Wallonië vond deze hervorming in twee fasen plaats: in 2019 werden de sociale toeslagen losgekoppeld van het socio-professioneel statuut, waardoor de werkende armen ook recht kregen op sociale toeslagen, terwijl het nieuwe stelsel in 2020 werd ingevoerd. Ook in het Vlaams Gewest was er voor de hervorming reeds een gedeeltelijke afwijking, omdat er in 2017 een indexesprong werd doorgevoerd. In de overige gewesten was dit niet het geval. Deze invoering in twee fasen wordt in MIDAS echter genegeerd.

⁴ Merk overigens op dat deze keuze voor het gezinshoofd geen consequenties heeft voor het bedrag van de kinderbijslag na de hervorming. Daar waar vroeger de toeslag afhankelijk was van de socioprofessionele categorie van de beide ouders, is het gezinsinkomen nu de belangrijkste factor.

2. De hervorming van de stelsels van kinderbijslag

In dit hoofdstuk worden de verschillen tussen de diverse stelsels besproken, te beginnen met het stelsel zoals het vóór de regionalisering op het federale niveau bestond. Elk van deze stelsels bestaat uit een universeel gedeelte, al dan niet gekoppeld aan de leeftijd en/of rangorde van het kind. Daarnaast is er een selectief gedeelte, waarbij deze selectie ofwel gebaseerd is op een inkomensgrens (waardoor huishoudens met een laag inkomen dus een hogere uitkering genieten) en/of de kenmerken van het kind (ondersteuningsbehoefte/aandoening) of ouder (ziekte, invaliditeit, alleenstaand, ...). Voor de hervormde, regionale, stelsels zijn er overgangsmatregelen voorzien, waarbij in hoofdzaak een verschil wordt gemaakt tussen kinderen die voor of na de invoering van het regionale stelsel geboren zijn. Al deze kenmerken zullen worden beschreven zoals ze in het model worden toegepast, en daar waar om redenen van vereenvoudiging wordt afgeweken van de regelgeving wordt dit zo veel mogelijk beschreven. Ook zullen de verschillende overgangsmatregelen worden beschreven. Voor een gedetailleerde vergelijking van de verschillende bedragen wordt verwezen naar Nevejan et al. (2021), Tabel 1, bladzijde 15 (voor de hervorming), en Tabel 2, bladzijde 17 (na de hervorming).

Om recht te hebben op kinderbijslag (voor de hervorming) was een noodzakelijke voorwaarde dat men i) in België werkt als werknemer of zelfstandige, of ii) dat een recht ontstaat omdat er gewerkt is. Door de hervorming viel deze voorwaarde gekoppeld aan de arbeidspositie van de ouder weg, en bleef enkel de voorwaarde over dat men in de betreffende regio moet wonen.

Voor het Vlaamse Groeipakket is bijvoorbeeld de voorwaarde dat men in Vlaanderen moet wonen, en ofwel de Belgische nationaliteit moet hebben of over een verblijfsvergunning beschikken. Deze worden in MIDAS onderscheiden. Daarnaast zijn er nog enkele gelijkgestelde situaties die in MIDAS niet worden onderscheiden, zoals bijvoorbeeld een niet-begeleide minderjarige.

Verder zijn er nog enkele voorwaarden betreffende het kind zelf; dit betreft een leeftijdsvoorwaarde (jonger dan 18 zonder bijkomende voorwaarden, of jonger dan 21 in geval van een handicap, of jonger dan 25 in geval van een student of stagiair). Deze worden in het model onderscheiden. Ook mag het kind niet meer uren werken dan een bepaalde bovengrens. In Vlaanderen is dit bijvoorbeeld 80 uur per maand in een normale job, of 475 uren per jaar in een studentenjob. In Wallonië is dit 240 uur per kwartaal, met uitzondering van het derde kwartaal (zomervakantie) waarvoor deze bovengrens niet wordt toegepast, tenminste indien het kind na de zomervakantie studeert. In het model wordt verondersteld dat een minderjarige of een student niet werkt, dus deze situatie kan zich in het model niet voordoen.

2.1. Het stelsel vóór de hervorming

Vóór de hervorming bestond het stelsel uit een éénmalige geboortepremie, afhankelijk van de rangorde binnen het gezin, en een maandelijks universele uitkering, die steeg met de leeftijdscategorie (0-5, 6-11, 12-17 en 18-24) en met de rangorde van het kind (eerste kind het laagst, gevolgd door het tweede kind; en de hoogste uitkering voor het derde en overige kinderen). Hierdoor ontvingen grotere gezinnen per kind meer kinderbijslag dan kleinere gezinnen (Nevejan et al, 2021, 15). Onder bepaalde

voorwaarden betreffende het socio-professioneel statuut en het inkomen van de ouders werd de gewone kinderbijslag verhoogd, nl. indien het bruto belastbare inkomen van de ouder en partner niet hoger was dan een vastgesteld plafond, én indien de rechthebbende (meestal de ouders van het kind) getroffen was door langdurige ziekte of ongeval, invaliditeit of handicap; gepensioneerd of langdurig werkloos is, of wanneer het kind van de rechthebbende een lichamelijke of geestelijke aandoening had. Indien dit allemaal niet het geval is, kon de universele uitkering worden verhoogd met een toeslag indien de ouder alleenstaande is of een laag inkomen heeft.

2.2. De verschillende stelsels na de hervorming

In deze paragraaf worden de verschillende hervormde stelsels kort besproken. Voor een gedetailleerde vergelijking van de verschillende bedragen wordt verwezen naar Tabel 2, bladzijde 17 van Nevejan et al. (2021). Het stelsel voor Vlaanderen, herdoopt tot Groeipakket, trad in werking vanaf 1 januari 2019 voor de bevolking van Vlaanderen. In Brussel en Wallonië kwamen de nieuwe stelsels in 2020 in voege. In alle drie de regio's wordt een onderscheid gemaakt tussen kinderen die reeds rechtgevend waren voor de datum van invoering, en kinderen die daarna rechtgevend zijn geworden – in de meeste gevallen omdat ze na dat tijdstip geboren zijn. Hierdoor ontstaat er dus een gemengd stelsel, waarbij binnen één gezin verschillende kinderen aan verschillende stelsels onderhevig kunnen zijn. Volgens Nevejan et al. (2021, 16) zijn deze hervormingen structureel (dat wil zeggen meer fundamenteel dan de aanpassing van de parameters), en volgens een patroon wat in de drie regio's vrij gelijkaardig is: zo werden de bedragen van het universeel gedeelte niet langer ranggebonden en minder leeftijdsgebonden; de koppeling van het selectief gedeelte werd verfijnd, onder meer door de invoering van een tweede – hogere – inkomensgrens. Ook werd de koppeling met de gezinsgrootte versterkt. Toch blijven er belangrijke verschillen bestaan, zoals uit onderstaande bespreking mag blijken.

2.2.1. Vlaanderen

In tegenstelling tot het oude stelsel, waar de geboortepremie afhankelijk was van de rang van het kind, ontvangt de rechthebbende bij de geboorte van elk kind eenzelfde startbedrag, in 2022 1 167,33⁵ euro.

Daarbovenop komen de maandelijks basisbedragen van de gezinsbijslag. Zoals gezegd wordt er hierbij in Vlaanderen een verschil gemaakt tussen kinderen geboren vóór en na 31/12/2018, dat wil zeggen indien het gezin op 1/1/2019 reeds in Vlaanderen woonde. Indien een gezin na die datum vanuit een ander gewest naar Vlaanderen verhuist, dan wordt op dezelfde manier het onderscheid gemaakt. Maar indien het gezin na die datum vanuit het buitenland naar Vlaanderen verhuist, dan vallen alle kinderen onder het nieuwe stelsel.

⁵ Alle bedragen genoemd in de beschrijving van de stelsels voor Vlaanderen en Brussel zijn afkomstig uit jaargang 2022 van het Praktijkboek Sociale Zekerheid. Zie eveneens myfamily.be (2022) en Vlaanderen.be (2022; 2022(b)).

a. Kinderen geboren voor 2019

Basisbedrag

In Vlaanderen zijn de op het moment van invoering van het stelsel reeds aanwezige kinderen onderhevig aan de regels zoals die reeds bestonden, waarbij er echter wel aanpassingen zijn gebeurd om een betere aansluiting bij de nieuwe regels te bekomen, en het nadeel van het oude stelsel voor kleinere gezinnen wat te verminderen. Het basisbedrag van het groeipakket is afhankelijk van de rangorde van het kind, maar in tegenstelling tot het oude systeem krijgen de oudere kinderen hogere bedragen dan de jongere kinderen. Voor het jongste kind is dit €97,73 per maand; voor het tweede kind €180,83 per maand, en voor de oudere kinderen is dit per kind €259,49 per maand⁶. Bovenop het basisbedrag komt in het oude stelsel een leeftijdsbijslag voor alle kinderen ouder dan 5. Net als bij het basisbedrag nemen de leeftijdstoelagen toe met de leeftijdsgroep. Maar naast het onderscheid naar leeftijdsgroep wordt een lagere uitkering toegekend aan het eerste kind dan aan de overige kinderen. Voor het eerste kind neemt de leeftijdsbijslag toe van €16,36 (voor een kind van 6 t/m 11 jaar), via €24,92 (voor een kind tussen 12 en 18) naar €28,72 voor een kind ouder dan 18. Voor de overige kinderen in het gezin is deze leeftijdsbijslag gegeven de leeftijdscategorie twee keer zo hoog (of zelfs 2,2 keer zo hoog voor de oudste categorie), namelijk €32,63 (voor een kind van 6 t/m 11 jaar), €49,86 (voor een kind tussen 12 en 18) en €63,40 voor een kind ouder dan 18. Dit onderscheid geldt enkel wanneer het gezin geen recht heeft op de sociale toeslag⁷. Als dat wél zo is, dan wordt ook voor het eerste kind de hogere uitkering toegekend.

Toelagen

Bovenop het basisbedrag komen de sociale toeslag en enkele zorgtoelagen die weliswaar bestaan voor alle rechtgevend kinderen, maar die specifieke kenmerken hebben voor kinderen geboren voor 2019, zoals de wezentoeslag. Deze regelingen zullen dus hier en in de volgende paragraaf worden besproken. De toelagen waarvan dezelfde regels voor kinderen zowel voor als vanaf 2019 van toepassing zijn (schoolbonus, zorgtoeslag voor specifieke ondersteuningsbehoefte, en de kleutertoeslag) zullen naderhand apart worden besproken.

De wezentoeslag voor kinderen geboren voor 2019 blijft dezelfde als voor de hervorming. Dus hebben dezen recht op een verhoogde kinderbijslag van €375,41. Een aanvullende voorwaarde daarbij is dat de overgebleven ouder geen gezin mag vormen met een persoon die geen bloed- of aanverwant is. Dit wordt in het model weerspiegeld door de aanvullende voorwaarde dat de overgebleven ouder alleenstaande moet zijn⁸.

⁶ Hierbij wordt de situatie op 1/1/2019 'vastgeklekt', waardoor deze bedragen niet veranderen als de rang van kinderen wijzigt. Het model heeft om technische redenen deze bepaling niet volledig overgenomen. De rangorde wordt in het model afzonderlijk bepaald voor kinderen geboren na 1/1/2019 en daarna. Dit impliceert dat als er in een gezin kinderen worden geboren na 1/1/2019, dit geen invloed heeft op de kinderbijslag toegekend aan kinderen die op 1/1/2019 al in het gezin leefden (in overeenstemming met de wetgeving). De rangorde van de kinderen geboren vóór 1/1/2019 verandert in het model wel als de kinderen de status van rechtgevend kind niet verlaten in volgorde van de leeftijd (dus bijvoorbeeld als een jonger kind eerder stopt met zijn studies als een ouder kind).

⁷ Of indien het kind zelf recht geeft op een bijkomende toeslag voor kinderen met een aandoening of wezenbijslag, maar deze mogelijkheden worden door het model niet onderscheiden.

⁸ We hebben dubbelstellingen opgelost door het basisbedrag voor kinderen van voor 2019 enkel te berekenen voor niet-wezen of wezen waarvan de overgebleven ouder géén alleenstaande is.

Voor gezinnen met een gezinsinkomen lager dan €32 238,01 per jaar⁹ is er een sociale toeslag voorzien.¹⁰ Voor kinderen geboren voor 2019 is het bedrag van de toeslag gelijk aan €49,75 voor het jongste kind, €30,83 voor het tweede kind, en €5,41 voor de oudste kinderen (zie de rangorde bij de vaststelling van het basisbedrag). Met andere woorden, de sociale toeslag is hoger voor kinderen waarvan het (leeftijdsonafhankelijke) basisbedrag lager is.

b. Kinderen geboren vanaf 2019

Waar het basisbedrag voor de kinderen geboren voor 2019 afhankelijk was van de rangorde van het kind, en vervuldigd werden met een leeftijdsgebonden component, is dit in het stelsel voor kinderen geboren vanaf 2019 niet meer het geval. Elk kind heeft recht op éénzelfde maandbedrag van €169,79 per maand.

Daarbovenop komen enkele zorgtoeslagen, zoals een wezentoeslag en de sociale toeslag. Hierbij wordt een verschil gemaakt tussen halve en volle wezen, en het bedrag van de *wezentoeslag* is 50 of 100 % van het basisbedrag. De *sociale toeslag* gebruikt voor kinderen geboren vanaf 2019 dezelfde inkomensgrens (namelijk €32 238,01) en -berekening als voor kinderen geboren voor 2019. Maar er is nog een tweede, en hogere, inkomensgrens, namelijk €63 672,48, voor diegenen die weliswaar boven de eerste inkomensgrens zitten, maar toch nog een laag inkomen hebben, en waarvan het derde of volgende kind na 2019 wordt geboren. Een tweede verschil is dat, waar de toeslagen voor de oudere kinderen afhankelijk waren van dezelfde rangorde als het basisbedrag, de hoogte van deze toeslag voor de jongere kinderen afhankelijk is van de gezinsgrootte. Indien het gezin één of twee kinderen heeft en een inkomen onder de eerste ondergrens, dan is het bedrag per kind geboren vanaf 2019 €53,06 per maand. Indien dit gezin meer dan twee kinderen heeft, dan is dit bedrag hoger, namelijk €84,89 per maand. Maar indien een gezin met meer dan twee kinderen een inkomen boven de eerste grens, maar onder de tweede grens heeft, en het derde of volgende kind is geboren vanaf 2019, dan bedraagt de sociale toeslag €63,67 per maand.

c. Regelingen die voor alle kinderen hetzelfde zijn

Dit betreft de zorgtoeslag voor kinderen jonger dan 21 met een specifieke ondersteuningsbehoefte, de schoolbonus en de kleutertoeslag. In de eerste regeling wordt het basisbedrag voor een kind met een specifieke ondersteuningsbehoefte aangevuld met een zorgtoeslag. Hiertoe worden de gevolgen van een ondersteuningsbehoefte door een door het Agentschap Opgroeien erkende arts beoordeeld op drie

⁹ Het gezinsinkomen wordt berekend als de som van de beroepsinkomsten (in loonverband, of als zelfstandige – waarbij dit bedrag in het laatste geval verhoogd wordt met 25%), uitkeringen in het kader van ziekteverzekering en werkloosheid, pensioenen, inkomensvervangende tegemoetkoming voor personen met een handicap, leefloon. Daarnaast wordt in het Vlaams Gewest ook het Kadastraal Inkomen in vreemd gebruik of gebruikt voor beroepsdoeleinden (kortom, alle KI buiten de hoofdverblijfplaats) opgenomen. Dit bedrag wordt bepaald voor de inkomsten van twee jaar geleden. Indien de samenstelling van het gezin in de tussentijd is veranderd, dan wordt dit gezinsinkomen gereconstrueerd. Het inkomen gaat uit van het inkomen op het aanslagbiljet, dus van twee jaar geleden. Indien in het jaar zelf de gezinsgrootte of -samenstelling wijzigt, dan wordt deze evaluatie herhaald. Verder is er nog een “alarmprocedure” die automatisch of via aanvraag de toeslag toekent indien de ouder een leefloon, inkomensvervangende tegemoetkoming of IGO ontvangt, of indien het huidige inkomen aantoonbaar onder de geldende grenzen ligt. In ons model wordt dit allemaal vereenvoudigd door uit te gaan van het (aangepaste) gezinsinkomen in het vorige kalenderjaar. Aangezien het KI buiten het bestek van MIDAS valt, wordt ook deze genegeerd.

¹⁰ Daarnaast worden onderhoudsgelden en KI van onroerende goederen die niet de hoofdverblijfplaats zijn of voor beroepsdoeleinden worden gebruikt, opgenomen. Maar aangezien deze niet worden gemodelleerd, worden deze genegeerd.

dimensies: 1) lichamelijke of geestelijke ongeschiktheid; 2) beoordeling van de gevolgen op de activiteit en participatie van het kind, en 3) beoordeling van de gevolgen op de familiale omgeving van het kind. Het resultaat van deze beoordeling is een score van maximaal 36 punten, die een gewogen som is van de score op de drie dimensies. Het bedrag van deze zorgtoeslag is dan afhankelijk van het aantal punten, en de verdeling van deze punten over de drie pijlers.

Deze regeling is zeer complex en zeer moeilijk in een model als MIDAS na te bootsen. Dus we hebben een sterke vereenvoudiging gemaakt, waarbij we enkel sterk afhankelijke kinderen onderscheiden, die de toeslag uit de categorie “tussen 6 en 8 punten waarvan minder dan 4 voor de eerste pijler” ontvangen¹¹. De selectie van deze kinderen bij de geboorte is willekeurig, maar ze blijven minstens tot hun 18^e verjaardag in deze situatie. Deze selectie sluit aan bij de modellering van de IVT: deze kinderen die geselecteerd worden voor de zorgtoeslag worden vanaf hun 18^e verjaardag rechthebbende voor de IVT. Als gevolg hiervan wordt deze zorgtoeslag een ‘voorportaal’ van de IVT. Bij de selectie van kinderen met een zorgtoeslag wordt rekening gehouden met de aantallen en kenmerken van de ontvangers van de IVT tussen 18 en 30 jaar.

De tweede regeling die geen onderscheid maakt tussen kinderen geboren voor – of vanaf 2019 is de schoolbonus. Deze bonus wordt jaarlijks uitbetaald, en stijgt met de leeftijd van het rechthebbende kind¹². Deze schoolbonus is met €21,23 het laagst voor kinderen van 0 tot en met 3, stijgt dan via €37,41 en €53,06 voor kinderen van 4 tot en met 11 en 12 tot en met 16, maar het hoogste bedrag van €63,67 voor rechthebbende kinderen ouder dan 16.

De derde regeling waarbij er geen onderscheid naar geboortjaar wordt gemaakt, is de kleutertoeslag. Deze heeft geen equivalent in de regeling kinderbijslag voor 2019, en wordt uitbetaald in het jaar dat het rechthebbende kind 3 of 4 jaar oud wordt (vanaf 2021; in de jaren voordien was dit enkel op 3 jaar). Alle kinderen die de Belgische nationaliteit hebben of legaal in Vlaanderen verblijven zijn rechtgevend, indien zij ingeschreven zijn in een door de Vlaamse Gemeenschap erkende, gefinancierde of gesubsidieerde onderwijsinstelling voor gewoon of buitengewoon kleuteronderwijs. Vanaf 2020 daalde de minimumleeftijd voor de leerplicht van 6 naar 5 jaar. Echter, vrijwel alle kinderen van 3 jaar zijn regelmatig aanwezig in een kleuterklas, dus wij veronderstellen dat deze regeling op alle kleuters van toepassing is¹³.

¹¹ Bij deze keuze volgen we de keuze gemaakt in het model EXPEDITION. FamiStat, het statistiekportaal van Famifed (zie <https://stat.famifed.be/demographic/index.php?lang=nl>) rapporteert de resultaten van tellingen van de kinderbijslag die tweejaarlijks plaatsvond (op 30 juni en 31 december) tot 2018. Uit resultaten van de telling van 30 juni 2018 (resultaten 1/7/2018) blijkt dat het percentage rechthebbende in deze één-na-laagste categorie ‘6-8 punten over de 3 pijlers en minder dan 4 punten in de 1ste pijler’ schommelt tussen de 51 % in het Vlaams Gewest en 42 % in het Waals Gewest. Op de tweede plaats staat de categorie ‘9-11 punten over de 3 pijlers en minder dan 4 punten in de 1ste pijler’, waarvan het percentage schommelt tussen 16 % in Brussel en 13 % in het Vlaams Gewest.

¹² Tot 2021 waren deze leeftijdsgroepen 0-4; 5-11; 12-17; en 18-24, maar in 2022 is dit veranderd naar 0-3; 4-11; 12-16; en 17-24.

¹³ Volgens het Vlaams Ministerie Onderwijs en Vorming was van de 3-jarigen 98,6 % in het schooljaar 2016-2017 ingeschreven in de eerste kleuterklas. Voor de 2de en 3de kleuterklas (op 4- en 5-jarige leeftijd) zijn deze percentages 98,9 % en 98,8 %. Bron: Vlaams Ministerie van Onderwijs en Vorming (2022).

2.2.2. Wallonië

Bij de beschrijving van het stelsel in Wallonië worden de cijfers van 1 augustus 2022 (index 118,36 t.o.v. 2013) gebruikt. Het regionale stelsel van kinderbijslag was vanaf 2020 van toepassing. Hierbij werd er een verschil gemaakt tussen kinderen geboren vóór en vanaf 2020¹⁴.

Net als in Vlaanderen is de geboortepremie onafhankelijk van de rang van het kind, en bedraagt €1 238,82.

a. Kinderen geboren voor 2020

Het basisbedrag van het groeipakket is afhankelijk van de rangorde van het kind, maar in lijn met het oude systeem, en in tegenstelling tot Vlaanderen, krijgen de oudere kinderen lagere bedragen dan de jongere kinderen. Voor het oudste kind is dit €107,89 per maand; voor het tweede kind €199,64 per maand, en voor de jongste kinderen is dit per kind €298,07 per maand. Bovenop het basisbedrag komt in het oude stelsel een leeftijdsbijslag voor alle kinderen ouder dan 5. Net als in het oude stelsel, en zoals in Vlaanderen, nemen de leeftijdstoelagen toe met de leeftijdsgroep. En naast het onderscheid naar leeftijdsgroep wordt eveneens een lagere uitkering toegekend aan het eerste kind dan aan de overige kinderen. Voor het eerste kind neemt de leeftijdsbijslag toe van €18,80 (voor een kind van 6 t/m 11 jaar), via €28,62 (voor een kind tussen 12 en 18) naar €32,99 voor een kind ouder dan 18. Voor de overige kinderen in het gezin is deze leeftijdsbijslag gegeven de leeftijdscategorie ook weer twee keer zo hoog (of zelfs 2.2 keer zo hoog voor de oudste categorie), namelijk €37,48 (voor een kind van 6 t/m 11 jaar), €57,27 (voor een kind tussen 12 en 18) en €72,82 voor een kind ouder dan 18. Deze verhoogde bedragen zijn eveneens als leeftijdsbijslag van toepassing voor alle kinderen (dus ook het eerste kind) in geval dat het gezin voor toeslagen in aanmerking komt (zie later).

Bovenop het maandelijks basisbedrag en de maandelijks leeftijdsbijslagen komen jaarlijkse leeftijdsafhankelijke toeslagen, de *prime de rentrée scolaire*, die in augustus van elk jaar worden uitbetaald. In de basisbedragen zijn deze €23,91 voor een kind jonger dan 6; €51,39 voor een kind van 6 tot en met 11; €71,70 voor een kind van 12 tot en met 17, en €95,61 voor een rechthebbend kind van 18 of ouder. Deze bedragen worden met ongeveer 37 % verhoogd indien het gezin voor toeslagen in aanmerking komt, en zijn dan respectievelijk €32,99; €70,01; €98,02; en €131,95.

Net als in Vlaanderen is het recht op verhoogde kinderbijslagen (de toeslagen) sinds 1 januari 2019 niet meer afhankelijk van de socio-professionele status, maar van het bruto-inkomen. Deze toeslagen volgen dezelfde structuur als het basisbedrag van de kinderbijslag, met dit verschil dat het bedrag van de toeslag *afneemt* met het aantal kinderen. Indien een gezin een inkomen heeft van minder dan €31 603,68¹⁵, dan bedragen deze maandelijks toeslagen €54,92 voor het oudste kind; €34,05 voor het tweede kind, en per kind €5,98 voor de jongste kinderen, of €27,45 voor de jongste kinderen indien het een eenoudergezin betreft.

¹⁴ Bij de beschrijving van het stelsel in Wallonië worden de cijfers van 1 augustus 2022 (index 118,36) gebruikt. Zie AVIQ (2019a; 2019b). Hierdoor is de vergelijking met de cijfers voor Vlaanderen en Brussel (overgenomen uit editie 2022 van het Praktijkboek Sociale Zekerheid) niet reeksreeks mogelijk. Zie eveneens Courtois (2022).

¹⁵ Zie voetnoot 8. Dit is het inkomen over twee kalenderjaren geleden (dus 2020 voor een recht van 1/7/2022 tot 30/6/2023). In ons model wordt dit allemaal vereenvoudigd door uit te gaan van het (aangepaste) gezinsinkomen in vorige kalenderjaar.

Een aanvullende toeslag is er voor kinderen uit gezinnen waarvan het inkomen lager is dan de inkomensgrens die we reeds noemden, en waarvan de ontvangende volwassene in een situatie van langdurige ziekte of invaliditeit is, waardoor hij of zij niet kan werken. In dit geval zijn de verhoogde kinderbijslagen voor het tweede en/of jongste kinderen hetzelfde als de gewone toeslag (ook ingeval van een éénoudergezin, maar de toeslag voor het eerste kind wordt €118,17 per maand.

Als laatste toeslag krijgen wezen een uitkering van €414,47. Dit is een uitkering die een minimum oplegt aan de gewone kinderbijslag en eventuele toeslagen, en ze dus in de meeste gevallen vervangt.

b. Kinderen geboren vanaf 2020

Waar het basisbedrag voor de kinderen geboren voor 2020 afhankelijk is van de rangorde van het kind, en vervuldigd werden met een leeftijdsgebonden component, is dit in het stelsel voor kinderen geboren vanaf 2020 niet meer het geval, net zoals dat in Vlaanderen het geval is. Maar waar elk kind in Vlaanderen recht heeft op éénzelfde maandbedrag, wordt er in Wallonië een verschil gemaakt tussen kinderen jonger dan 18 (€174,65) en 18 of ouder (€185,82). Daarbovenop komt een leeftijdsgebonden schoolpremie, waarbij in tegenstelling tot het stelsel van oudere kinderen het jaarlijks bedrag toeneemt met de leeftijdscategorie. Deze bedragen zijn €22,52 voor kinderen tot en met 4 jaar, €33,79 voor kinderen van 5 tot en met 10; €56,31 van 11 tot en met 16 en €90,10 voor kinderen vanaf 17 jaar oud.

Daarbovenop komen een aantal supplementen. Waar er voor de kinderen geboren voor 2020 één enkele ondergrens was, worden er nu twee ondergrenzen vastgesteld. De eerste ondergrens is dezelfde als van toepassing is op kinderen geboren voor 2020, namelijk €31 603,68. Daarboven komt een tweede grens van €51 000. De eerste toeslag is die voor gezinnen met een laag inkomen: deze is per maand €61,94 indien het inkomen lager is dan de eerste grens, en €28,16 indien het inkomen lager is dan de tweede grens.

Een aanvullende toeslag is er voor kinderen uit gezinnen waarvan het inkomen lager is dan de eerste inkomensgrens, en waarvan de ontvangende volwassene in een situatie van langdurige ziekte of invaliditeit is, waardoor hij of zij niet kan werken. In MIDAS wordt deze voorwaarde geïmplementeerd als dat minimaal één van de ouders niet werkt en een invaliditeitsuitkering ontvangt. Voor kinderen geboren voor 2020 was dit een aanvullende uitkering afhankelijk van de rangorde van het kind, variërend tussen €118,17 voor het eerste kind en €5,98 voor het derde of latere kind. Voor de kinderen geboren vanaf 2020 is dit één bedrag, namelijk €11,26, wat voor alle kinderen hetzelfde is. Maar daarnaast, en dat is nieuw, is er een aanvullende toeslag van €73,20, die niet afhankelijk is van een inkomensgrens, en die wordt toegekend wanneer er minimaal één volwassene IVT (of THAB) ontvangt.

Daarenboven zijn er twee toeslagen. Voor de gezinnen waarvan het hoofd een alleenstaande ouder is zijn de bedragen van de toeslag €22,52 (indien het inkomen lager is dan de eerste inkomensgrens), respectievelijk €11,26 (indien het inkomen hoger is dan de eerste, maar lager dan de tweede inkomensgrens) per kind. En voor grote gezinnen (minimaal 3 kinderen) is deze toeslag €39,42, respectievelijk €22,52 per kind.

Daarnaast is er een toeslag voor (half) wezen, die afhankelijk is van of ze 18 of ouder zijn – deze toeslag bedraagt de helft van de basisuitkering (dus €87,28 voor kinderen tot 18, en €92,91 voor kinderen vanaf

die leeftijd). Indien een kind volledige wees is, dan wordt de gezinsbijslag *verhoogd* tot €394,17 per maand.

c. Alle kinderen

Daarbovenop is er nog een maandelijkse toeslag voor kinderen jonger dan 21 die zorg- of hulpbehoevend zijn ("*atteints d'une affection*"). Deze regeling is van toepassing op kinderen zowel voor als vanaf 2019 geboren. Net zoals het geval was voor het stelsel in Vlaanderen veronderstellen we hier de één-na-laagste uitkering, wat aanleiding geeft tot een bedrag van €630,76. Dit wordt gemodelleerd voor alle rechtgevende kinderen die op latere leeftijd in aanmerking zullen komen voor een InkomensVervangende Tegemoetkoming (IVT).

2.2.3. Brussel

Het stelsel van kinderbijslag in Brussel werd ingevoerd vanaf 2020. In principe is dit nieuwe stelsel van toepassing voor alle kinderen. Maar als overgangsmaatregel is voorzien dat, voor gezinnen die in 2019 in Brussel woonden, de kinderen die voor 2020 geboren (of geregistreerd) zijn, in aanmerking komen voor het oude basisbedrag (basisbedrag plus leeftijdstoelagen, afhankelijk van de rang van het kind) en de oude toeslagen, indien de resulterende uitkering groter is dan het basisbedrag onder het nieuwe stelsel. Vanaf 2025 zouden dan voor de berekening van de basisbedragen dezelfde bedragen gelden als voor de kinderen geboren vanaf 2020. Los hiervan blijven eveneens de oude sociale toeslagen (afhankelijk van de socio-economische status en de rangorde van het kind) van toepassing, indien voordeliger voor de rechthebbende, maar enkel in geval van de nieuwe inkomensvoorwaarde (nl. dat het inkomen lager is dan €33 923,83). De nu volgende bespreking wordt beperkt tot het nieuwe stelsel (zie, onder meer, KidsLife (2022)).

Het basisbedrag van de kinderbijslag is afhankelijk van of er één of meerdere kinderen in het gezin zijn. Voor gezinnen met één kind is dit bedrag €162,36 per maand, en dus onafhankelijk van de leeftijd van het kind. Voor gezinnen met meerdere kinderen wordt hetzelfde bedrag genomen indien het kind jonger is dan 12. Voor kinderen vanaf 12 en tot en met 24 neemt het toe tot €173,18, en voor rechthebbende (dus studerende) kinderen boven die leeftijd is het basisbedrag €184,01. Het basisbedrag wordt aangevuld met een wezenbijslag, die gelijk is aan 50 of 100 % van het basisbedrag voor halve en volle wezen.

Bij de modellering van de maandelijkse toeslag voor kinderen jonger dan 21 die zorg- of hulpbehoevend zijn veronderstellen we opnieuw de één-na-laagste uitkering voor alle rechthebbende kinderen die op latere leeftijd in aanmerking zullen komen voor een InkomensVervangende Tegemoetkoming (IVT).

Daarnaast is er een sociale toeslag, die, zoals in de andere deelstaten, enkel nog gebaseerd is op het gezinsinkomen. De grens hierbij is €33 923,83. Daarnaast is er, net zoals in het stelsel in Wallonië, een tweede grens op €49 244,28. Verder is de toeslag afhankelijk van het aantal kinderen, de leeftijd van die kinderen. Deze toeslag wordt samengevat in de volgende tabel.

Tabel 1 Bedrag sociale toeslag per kind in Brussel

Aantal kinderen	Inkomen lager dan €33 923,83				Inkomen lager dan €49 244,28
	0 tot en met 11		12 tot en met 24		
	Éénouder	Niet éénouder	Éénouder	Niet éénouder	
1	€43,30		€54,12		-
2	€86,38	€75,77	€97,20	€86,59	€27,06
≥ 3	€140,28	€119,06	€151,11	€129,89	€77,93

Als laatste wordt in juli van elk jaar de kinderbijslag verhoogd met een leeftijdsgebonden schoolpremie. Net als in Vlaanderen en Wallonië is deze met €21,65 het laagst voor een kind jonger dan 6; en neemt toe via €32,47 voor een kind van 6 tot en met 11; en €54,12 voor een kind van 12 tot en met 17, naar €86,59 voor een rechthebbend kind van 18 of ouder, indien dit kind hoger onderwijs volgt. Indien dit niet het geval is, dan blijft de schooltoeslag voor de 18+'er even hoog als voor de kinderen van 12 tot en met 17. In tegenstelling tot Wallonië, maar in overeenstemming met Vlaanderen, worden deze schooltoelagen niet verhoogd wanneer het gezin voor toeslagen in aanmerking komt.

2.2.4. Indexering

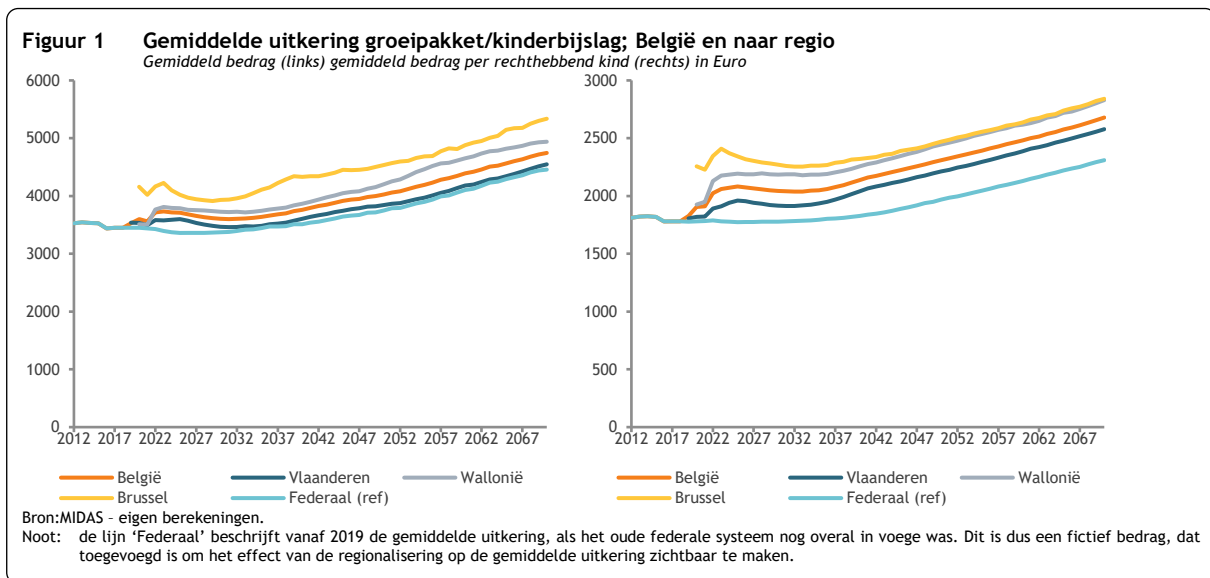
Een belangrijke beslissing in de modellering van elke socialezekerheidsregeling is de ontwikkeling van de parameters doorheen de tijd. Hierbij kunnen we niet terugvallen op de parameters en aannames zoals ze in de berekening van de welvaartsenveloppe worden gebruikt, omdat de geregionaliseerde kinderbijslag volgens de programmawet van 29 december 2014 hierin niet zijn opgenomen.

Tot en met 2022 (en 2023 voor het Waals Gewest) volgen wij de historische ontwikkeling van de verschillende parameters. In projectie hebben wij ervoor gekozen om in de reële herwaardering de keuzes gebruikt in het meest recente rapport van de Studiecommissie voor de Vergrijzing (Hoge Raad voor Financiën, 2022, 27) te volgen. Deze worden als volgt omschreven: "Volgens het principe van een projectie bij ongewijzigde wetgeving op middellange termijn wordt de kinderbijslag niet in reële termen geherwaardeerd tot 2027. Op lange termijn wordt het principe van een projectie bij ongewijzigd beleid nageleefd: de kinderbijslag wordt vanaf 2028 jaarlijks geherwaardeerd, op basis van het gemiddelde verschil dat de afgelopen dertig jaar werd opgetekend tussen de groei van het gemiddelde loon (0,9 %) en de groei van het gemiddelde bedrag van die uitkeringen (0,2 %), of een loskoppeling van 0,7 procentpunt". Dus bij een op lange termijn veronderstelde reële loongroei van 1,5 % per jaar zouden de uitkeringen met 0,8 % per jaar geherwaardeerd worden.¹⁶ Deze herwaardering geldt ook voor de inkomensgrenzen zoals die gebruikt worden voor de sociale toeslagen.

¹⁶ Deze veronderstelde herwaardering wordt op dezelfde manier voor alle gewesten toegepast, hoewel de historische evolutie van de parameters in recente jaren duidelijk verschilt.

3. Resultaten

Het linker-luik van Figuur 1 hieronder laat het gemiddelde bedrag van de kinderbijslag van alle rechthebbende huishoudens zien. Het rechterluik is dit het gemiddelde van het bedrag per rechthebbend kind. De lijn 'België' beschrijft het gemiddelde (of het gemiddelde per kind) voor het land als geheel, dus als gewogen samenstelling van de resultaten van de drie afzonderlijke stelsels.

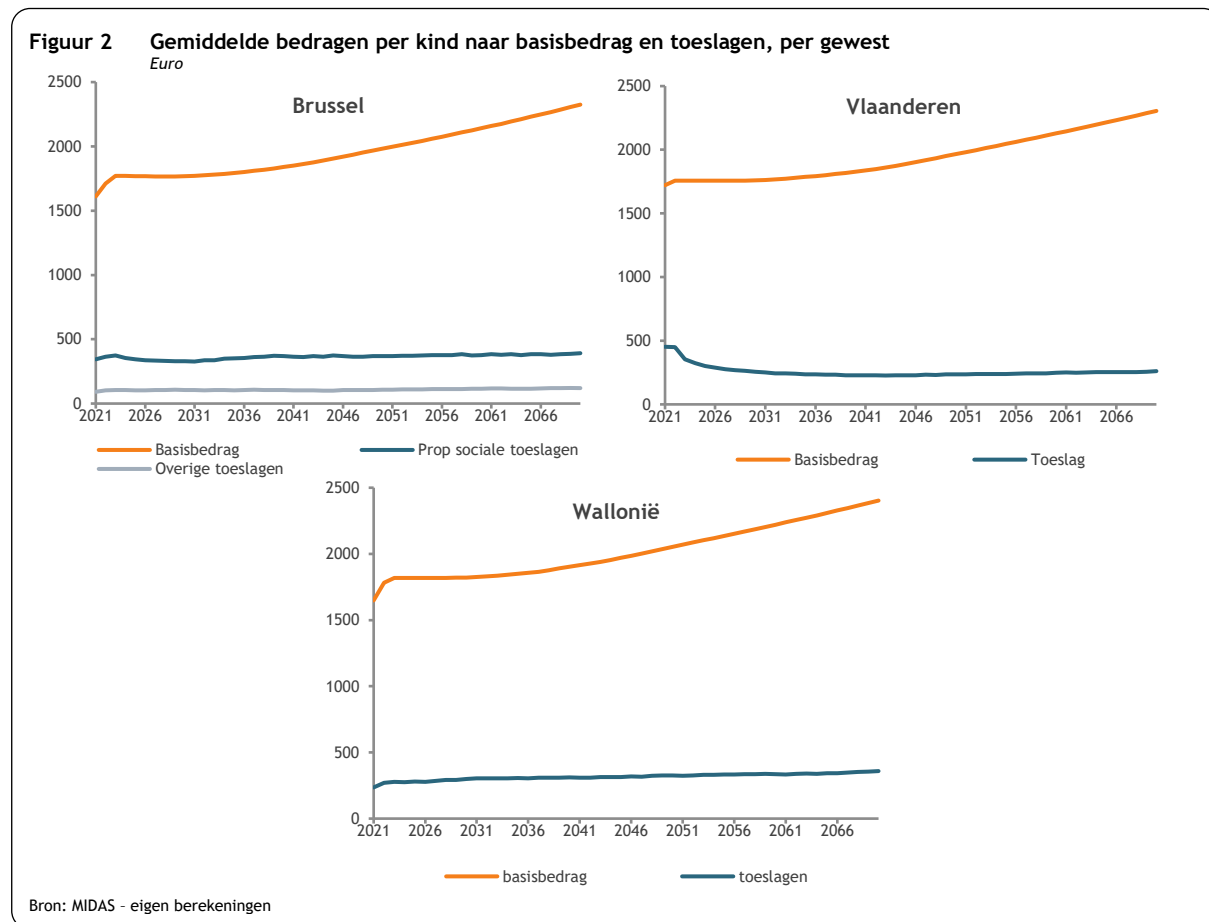


Vergeleken met het oude stelsel liggen de gemiddelde uitkeringen hoger in de nieuwe stelsels. De uitkeringen in het nieuwe stelsel liggen verhoudingsgewijs vrij hoog in Brussel (linkergrafiek), maar dat komt omdat er in Brussel verhoudingsgewijs meer kinderen per gezin zijn; wanneer de gemiddelde uitkering wordt gedeeld door het aantal rechtgevendende kinderen, dan is het verschil minder groot. De 'piek' van de gemiddelde bedragen in het begin van de simulatieperiode, vooral in het Brussels en Waals Gewest, en minder uitgesproken in Vlaanderen, is het gevolg van de vrij hoge groei van de geobserveerde parameters tot 2023, onmiddellijk gevolgd door de veronderstelde volledige afwezigheid van herwaardering tot en met 2027.

In het oude stelsel was het basisbedrag afhankelijk van de rangorde van het kind, en vervuldigd met een leeftijdsgebonden component. De hoogte van de sociale toeslag was afhankelijk van de socio-economische categorie van de rechthebbende, en het aantal kinderen in het gezin. In de drie landsdelen is dit niet langer het geval. Elk kind in Vlaanderen heeft recht op éénzelfde maandbedrag; en ook de hoogte van de sociale toeslag per kind is niet afhankelijk van de rangorde van de kinderen. In Wallonië zijn het basisbedrag en de leeftijdsgebonden uitkering eveneens onafhankelijk van de rangorde van de kinderen. In Brussel is het basisbedrag van de kinderbijslag afhankelijk per kind minimaal even hoog wanneer er meerdere kinderen in het gezin zijn. Daarnaast is er een sociale toeslag, die, behalve van het gezinsinkomen, afhankelijk is van het aantal kinderen, en de leeftijd van die kinderen. Alles bij elkaar kunnen we dus verwachten dat, ten opzichte van elkaar, het Groeipakket in Vlaanderen en de kinderbijslag in het nieuwe systeem van Wallonië verhoudingsgewijs neutraler zullen zijn ten opzichte

van het aantal kinderen, terwijl het nieuwe systeem in Brussel iets meer in het voordeel van grotere gezinnen zal zijn.

In de drie landsdelen bestaat de kinderbijslag of groeipakket uit een basisbedrag en een of meerdere toeslagen. De onderstaande figuren geven het gemiddelde basisbedrag per kind en de gemiddelde



In Brussel wordt een onderscheid gemaakt naar de sociale toeslag en overige toeslagen, waarbij die laatste de geboortepremie, het invaliditeitssupplement, de schoolbonus en de wezenuitkering omvatten.

In alle drie de regio's is het basisbedrag veruit het meest belangrijke onderdeel van de kinderbijslag of het Groeipakket. Over de hele simulatieperiode genomen is het gemiddelde basisbedrag per kind tussen de 4 en 8,8 keer hoger dan het gemiddelde aanvullende bedrag per kind in Vlaanderen; tussen 6 en 7 in Wallonië, en tussen 3,6 en 4,5 in Brussel. De conclusie van Nevejan et al. (2021, 18) dat het universele gedeelte van de kinderbijslag verhoudingsgewijs laag ligt in Brussel en hoog in Vlaanderen, worden dus bevestigd.

Merk tenslotte op dat het belang van het universele gedeelte voor de ontvangende huishoudens toeneemt over de tijd. Vooral in Wallonië is de toename aan het begin van de simulatieperiode opvallend, en deze komt door de forse stijging (8 % in lopende prijzen tussen 2021 en 2022) van de beide bedragen voor kinderen tussen 0 en 17 en 18 en 24. In Vlaanderen starten de bedragen iets hoger, maar stijgen ze minder in de eerste jaren van simulatie (dus gebaseerd op de geobserveerde parameters), waardoor de gemiddelde bedragen per kind in de drie gewesten op de lange termijn vergelijkbaar zijn.

Daarnaast bestaat de toeslag voor het grootste gedeelte uit de sociale toeslagen, die afhankelijk is van of het gezinsinkomen lager is dan een (of twee) inkomensgrenzen. Deze grenzen groeien trager dan de lonen, waardoor deze inkomensgrenzen verhoudingsgewijs verstrengen en het aantal ontvangende gezinnen van een sociale toeslag afneemt.

Referenties

- AVIQ, 2019, AVIQKid - Wallonie, Familles, Sante, Handicap. Historique des Montants. Overgenomen van de AVIQ website <https://aviqkid.aviq.be/Pages/montants.aspx> [24/10/2022]
- AVIQ, 2019b, AVIQKid - Wallonie, Familles, Sante, Handicap. Montants généraux au 1er août 2022. Overgenomen van de AVIQ website. <https://aviqkid.aviq.be/%2FDocuments%2Fassets%2Fmontants%2FBar%C3%A8me%20consolid%C3%A9%2020220801.pdf> [24/10/2022]
- Courtois, P. (2022). Allocations familiales en Wallonie – Toutes les aides reçues. Overgenomen van de website Mes Aides Financières. <https://mes-aides-financieres.be/allocations-familiales/wallonie/> [12/12/2022].
- Hoge Raad voor Financiën, 2022. Jaarlijks Verslag van de Studiecommissie voor de Vergrijzing. Overgenomen van de website van het Federaal Planbureau: https://www.plan.be/publications/publication-2263-nl-studiecommissie_voor_de_vergrijzing_jaarlijks_verslag_2022 [17/11/2022]
- KidsLife.be, 2022. Kinderbijslag in België. Overgenomen van de website KidsLife.be: <https://www.kidslife.be/nl/kinderbijslag> [12/12/2022].
- MyFamily.be, 2022. Van Kinderbijslag naar Groeipakket. Overgenomen van Myfamily.be. <https://www.myfamily.be/groeipakket/> [12/12/2022]
- Nevejan, Hendrik, Guy van Camp en Dieter Vandelannoote, 2021. Regionale kinderbijslaghervormingen – Een impactanalyse met het microsimulatiemodel EXPEDITION. Federaal Planbureau Working Paper 04-21. Brussel: Federaal Planbureau. Overgenomen van de website van het Federaal Planbureau https://www.plan.be/publications/publication-2120-nl-regionale_kinderbijslaghervormingen_een_impactanalyse_met_het_microsimulatiemodel_expedition [06/01/2023]
- Vlaams Ministerie van Onderwijs en Vorming, 2022. Vlaanderen is Onderwijs en Vorming: Kleuteronderwijs. Overgenomen van de website van het Vlaams Ministerie van Onderwijs en Vorming. <https://onderwijs.vlaanderen.be/nl/kleuteronderwijs#cijfers> [24/10/2022]
- Vlaanderen.be, 2022. Groeipakket: Bedragen voor kinderen geboren voor 2019 (overgangsmatregelen). Overgenomen van de website van de Vlaamse overheid. <https://www.groeipakket.be/bedragen/voor-kinderen-geboren-voor-2019> [12/12/2022]
- Vlaanderen.be, 2022(b). Groeipakket: bedragen. Overgenomen van de website van de Vlaamse overheid. <https://www.groeipakket.be/bedragen> [12/12/2022]

MIDAS: Validering en kalibratie van de gesimuleerde sociale bijdragen en belastingen

Annex 7 bij 'MIDAS 2.0: Revisie van een
dynamisch microsimulatiemodel'

Januari 2023

Karel Van den Bosch, kvdb@plan.be

Inhoudstafel

1. Inleiding	1
2. Sociale bijdragen	3
2.1. Lonen en salarissen	3
2.2. Inkomsten van zelfstandigen	4
2.3. Uitkeringen wegens ziekte en invaliditeit	4
2.4. Pensioenen	5
3. Het gezamenlijk belastbaar inkomen (GBI)	7
4. De betaalde belasting	9
4.1. Volledige selectieve steekproef	9
4.2. Steekproef waarin het verschil tussen gesimuleerd en geobserveerd GBI beperkt is, of zelfs 0	10
4.3. De bijzondere bijdrage aan de sociale zekerheid	11
5. Het netto-inkomen	13
5.1. Globale verschillen	13
5.2. Afwijking tussen simulatie en observatie naargelang niveau inkomen	15
6. Kalibratie	18
6.1. Methode	18
6.2. Resultaten	19
6.3. Implementatie	23
6.4. Beoordeling	23
7. Conclusie	25

Lijst van tabellen

Tabel 1	Variabelen uit IPCAL in MIDAS start-dataset gebruikt voor kalibratie	1
Tabel 2	Resultaten voor simulatie sociale bijdragen op lonen en salarissen	3
Tabel 3	Resultaten voor simulatie sociale bijdragen op uitkeringen wegens ziekte en invaliditeit	5
Tabel 4	Resultaten voor simulatie sociale bijdragen op pensioenen	5
Tabel 5	Resultaten voor simulatie gezamenlijk belastbaar inkomen	7
Tabel 6	Resultaten voor simulatie belasting staat	9

Tabel 7	Resultaten voor simulatie belasting staat wanneer het absolute verschil tussen gesimuleerd en geobserveerd GBI kleiner is dan 1 000 euro, c.q. 0 is; alleen degenen voor wie de belasting > 0	10
Tabel 8	Resultaten voor de gesimuleerde bijzondere bijdrage aan de sociale zekerheid	12
Tabel 9	Resultaten voor simulatie netto-inkomen, naargelang het verschil tussen gesimuleerd en geobserveerd GBI	13
Tabel 10	Kengetallen voor verschil tussen gesimuleerd en geobserveerd netto-inkomen, naargelang sociaal statuut	14
Tabel 11	Regressie van verschil tussen gesimuleerd en geobserveerd netto inkomen, per sociaal statuut-groep	20
Tabel 12	Kengetallen voor de kwaliteit van de kalibratie	21
Tabel 13	Vergelijking van gesimuleerd en gekalibreerd netto-inkomen met het geobserveerd netto-inkomen, naargelang sociaal statuut	24

Lijst van figuren

Figuur 1	Vershil tussen gesimuleerde en geobserveerde sociale bijdragen op lonen en salarissen	4
Figuur 2	Vershil tussen gesimuleerde en geobserveerde sociale bijdragen op uitkeringen wegens ziekte en invaliditeit	5
Figuur 3	Vershil tussen gesimuleerde en geobserveerde sociale bijdragen op wettelijke pensioenen ..	6
Figuur 4	Vershil tussen gesimuleerd en geobserveerd GBI	8
Figuur 5	Vershil tussen gesimuleerde en geobserveerde belasting staat	9
Figuur 6	Vershil tussen gesimuleerde en geobserveerde belasting staat, wanneer het absolute verschil tussen gesimuleerd en geobserveerd GBI kleiner is dan 1 000 euro, c.q. 0 is	10
Figuur 7	De gesimuleerde bijzondere bijdrage aan de sociale zekerheid, naargelang gesimuleerd belastbaar inkomen	12
Figuur 8	Vershil tussen gesimuleerd en geobserveerd netto inkomen	14
Figuur 9	Gemiddeld verschil gesimuleerd netto inkomen - geobserveerd netto inkomen, naargelang verschillende inkomensconcepten	16
Figuur 10	Kengetallen van het GBI zonder en met afzonderlijk belastbare inkomsten naargelang categorie van beroepsinkomsten, werknemers met bruto-inkomsten > 80 000 euro	17
Figuur 11	Mediaan van verschil tussen gesimuleerd netto inkomen - geobserveerd netto inkomen, gesimuleerd (vóór kalibratie) en na kalibratie, naargelang bruto inkomen, per sociaal statuut groep	21
Figuur 12	Mediaan van verschil tussen direct gekalibreerd netto inkomen - geobserveerd netto inkomen, naargelang bruto inkomen, werknemers	23

Federaal Planbureau

Het Federaal Planbureau (FPB) is een instelling van openbaar nut die beleidsrelevante studies en vooruitzichten maakt over economische, socio-economische en milieuvraagstukken. Daarnaast bestudeert het de integratie van die vraagstukken in een context van duurzame ontwikkeling. Het stelt zijn wetenschappelijke expertise onder meer ter beschikking van de regering, het Parlement, de sociale gesprekspartners, nationale en internationale instellingen.

De werkzaamheden van het FPB worden steeds gekenmerkt door een onafhankelijke benadering, transparantie en aandacht voor het algemeen welzijn. De kwaliteit van de gegevens, een wetenschappelijke methodologie en de empirische geldigheid van de analyses staan daarbij centraal. Tot slot zorgt het FPB voor een ruime verspreiding van de resultaten van zijn werkzaamheden en draagt zo bij tot het democratisch debat.

Het Federaal Planbureau is EMAS en Ecodynamische Onderneming (drie sterren) gecertificeerd voor zijn milieubeheer.

<https://www.plan.be>

e-mail: contact@plan.be

Overname wordt toegestaan, behalve voor handelsdoeleinden, mits bronvermelding.

Verantwoordelijke uitgever: Saskia Weemaes

Wettelijk Depot: D/2023/7433/3

Federaal Planbureau

Belliardstraat 14-18, 1040 Brussel

tel.: +32-2-5077311

e-mail: contact@plan.be

<https://www.plan.be>

MIDAS: Validering en kalibratie van de gesimuleerde sociale bijdragen en belastingen

Annex 7 bij 'MIDAS 2.0: Revisie van een dynamisch microsimulatiemodel'

Januari 2023

Karel Van den Bosch, kvdb@plan.be

Abstract – Deze annex presenteert een poging tot validering van de gesimuleerde sociale bijdragen en belastingen, en stelt een kalibratie voor om systematische afwijkingen te verwijderen. ¹

¹ Met dank aan Guy Van Camp voor nuttige opmerkingen.

1. Inleiding

Een vergelijking met IPCAL gegevens en met EU-SILC toonde aan dat in MIDAS de belastingen worden overschat, en het netto-inkomen dus wordt onderschat. De voornaamste reden is wellicht dat een aantal aftrekposten en belastingverminderingen niet kunnen worden gesimuleerd door gebrek aan relevante gegevens. Als oplossing wordt gekozen voor een kalibratie van de het gesimuleerde netto-inkomen op het netto inkomen dat kan worden afgeleid van de IPCAL (en andere) gegevens in de startdata. In dit document wordt de kalibratie uitgewerkt.

Een kalibratie is mogelijk omdat we in de start-dataset beschikken over enkele variabelen uit IPCAL (een databestand over de definitieve belastingaanslag van de FOD Financiën) voor de jaren 2011 en 2012; de belangrijkste zijn vermeld in Tabel 1.

Tabel 1 Variabelen uit IPCAL in MIDAS start-dataset gebruikt voor kalibratie

A8299	Belasting staat
A8300	Belasting staat (gezin)
A7555	Gezamenlijk belastbaar inkomen
A7556	Gezamenlijk belastbaar inkomen gezin
A7557	Totaal van de afzonderlijk belastbare inkomsten

Om de kalibratie zo nauwkeurig mogelijk te maken, is gekozen voor de volgende aanpak. De bruto-inkomens voor het jaar 2012 (bron: het Datawarehouse) worden ingevoerd in de belastingmodule van MIDAS, met de parameters voor het jaar 2012. Het resultaat wordt vergeleken met de variabele “belasting staat” volgens IPCAL. Omdat in deze module ook de sociale bijdragen worden gesimuleerd, maken we allereerst een vergelijking van deze gesimuleerde variabele met de sociale bijdragen die afgeleid kunnen worden uit de Datawarehouse variabelen. Deze bron bevat voor diverse inkomensbronnen zowel het bruto-bedrag, als het bruto belastbaar bedrag. Het verschil zijn de (gewone) sociale bijdragen. (Met uitzondering van de zelfstandigen, zie onder.) Om de inkomensbedragen in overeenstemming te brengen met de parameters voor de bijdragen en belastingen, zijn alle inkomens verminderd met 2,8 %. MIDAS werkt immers met parameters in constante prijzen van 2011; de gemiddelde afwijking van de parameters in MIDAS van hun werkelijke waarde is 2,8 %.²

Deze aanpak heeft ook het belangrijke voordeel dat het ons in staat stelt de belastingmodule in MIDAS te valideren. De vergelijking van gesimuleerde met geobserveerde variabelen bracht inderdaad enkele tekortkomingen aan het licht, die hersteld konden worden.

Speciale vermelding verdient de bijzondere bijdrage voor de sociale zekerheid. Dit is geen deel van de sociale bijdragen, maar hoort bij de personenbelasting. Deze werd tot nu toe niet gesimuleerd, en we beschikken niet over observaties voor deze variabele. Tegelijk is dit – in tegenstelling tot vele andere bijzondere bijdragen – een maatregel waaraan alle werknemers onderworpen zijn, en gemakkelijk om te simuleren. Kwantitatief is deze bijdrage niet onbelangrijk: de totale opbrengst was 1 199 miljoen euro

² Deze werkwijze is niet de optimale. Ten eerste zou het beter zijn om de genoemde inkomensnoties uit de basisgegevens van het DWH AM&SB zelf samen te stellen. Ten tweede geniet het uiteraard de voorkeur om de parameters voor 2011 toe te passen op de gegevens voor 2011. Wegens tijdsgebrek was dit niet haalbaar.

in 2016, ofwel 2,3 % van de totale personenbelasting.³ Een eigenaardigheid van deze maatregel is dat de inkomensdrempels en bedragen niet geïndexeerd worden.

Een belangrijke methodologische keuze is de volgende. De simulatie in MIDAS van alle inkomens is gebaseerd op de assumptie dat iedere persoon per jaar slechts één sociaal statuut kan hebben en slechts één type inkomen kan ontvangen. De belastingmodule sluit daarbij aan. Om deze reden is de vergelijking beperkt tot de personen in de steekproef voor wie deze assumptie in feite opgaat. Bij de gepensioneerden nemen we alleen de gevallen op in de vergelijking die geen andere dan wettelijke pensioenen hebben. In de steekproef van 553 722 individuen hebben 389 207 personen een inkomen, en hiervan worden er 232 003 (59,6 %) geselecteerd voor de vergelijking. Om de verstoring van het beeld door extreme waarden te vermijden, zijn gevallen met een belastbaar inkomen van meer dan 100 000 euro (hoger dan het 99^{ste} percentiel) meestal uitgesloten.

Dit document is opgezet als volgt. We vergelijken achtereenvolgens simulaties met observaties voor de sociale bijdragen, het gezamenlijk belastbaar inkomen, de betaalde belasting en het netto-inkomen (secties 2, 3, 4 en 5). Dit netto-inkomen is gedefinieerd als de geobserveerde belastbare bruto inkomens, minus sociale bijdragen en belastingen inclusief de bijzondere bijdrage voor de sociale zekerheid. Het omvat dus niet niet-belastbare inkomens, zoals kinderbijslag en bijstandsuitkeringen.⁴ Sectie 6 behandelt de eigenlijke kalibratie en sectie 7 bevat algemene slotopmerkingen.

³ Fiscaal Memento 2018 (V01), p. 14

⁴ Het geobserveerde netto-inkomen komt als zodanig niet voor in de startdata, maar wordt afgeleid uit het totaal bruto inkomen en de sociale bijdragen en belastingen.

2. Sociale bijdragen

In MIDAS worden vier inkomenscomponenten onderscheiden waarop sociale bijdragen geheven kunnen worden: lonen en salarissen, inkomsten van zelfstandigen, uitkeringen wegens ziekte of invaliditeit (inclusief uitkeringen wegens beroepsziekte of arbeidsongeval) en pensioenen. Er zijn geen sociale bijdragen op kinderbijslag, bijstandsuitkeringen en werkloosheidsuitkeringen. Voor de bruggepensioneerden hebben we alleen informatie over hun uitkering door de RVA (=werkloosheidsuitkering), maar niet over de toeslag door de werkgever. Hieronder gaan we kort in op de sociale bijdragen per inkomenscomponent. De sociale bijdragen volgens de DWH AM&SB gegevens worden berekend als het verschil tussen het bruto inkomen en het bruto belastbaar inkomen.

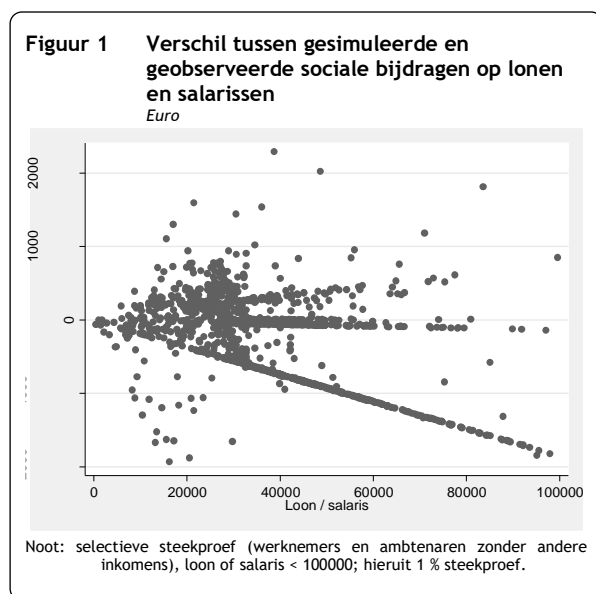
2.1. Lonen en salarissen

In de simulatie van sociale bijdragen op lonen en salarissen gaan we uit van de algemene bijdragevoeten (0,1105 voor ambtenaren en 0,1307 voor alle andere werknemers), en houden we rekening met de werkbonus. Tabel 2 geeft een aantal kengetallen voor de betreffende variabelen in de selectieve steekproef (werknemers en ambtenaren zonder andere inkomens dan loon of salaris).

Tabel 2 Resultaten voor simulatie sociale bijdragen op lonen en salarissen

	Gemiddelde	10 ^{de} percentiel	25 ^{ste} percentiel	Mediaan	75 ^{ste} percentiel	90 ^{ste} percentiel
<i>Alle werknemers</i>						
Sociale bijdragen volgens Datawarehouse	4 840	1 613	3 080	4 508	6 141	8 375
Sociale bijdragen gesimuleerd in MIDAS	4 579	1 748	3 207	4 325	5 688	7 563
Vershil (op individueel niveau	-261	-1 025	-696	-78	104	373
<i>Alle werknemers behalve ambtenaren</i>						
Sociale bijdragen volgens Datawarehouse	4 858	1 380	2 779	4 489	6 393	8 783
Sociale bijdragen gesimuleerd in MIDAS	4 523	1 480	2 931	4 250	5 763	7 790
Vershil (op individueel niveau	-335	-1 105	-786	-342	112	403
<i>Ambtenaren</i>						
Sociale bijdragen volgens Datawarehouse	4 775	3 070	3 721	4 537	5 519	6 937
Sociale bijdragen gesimuleerd in MIDAS	4 791	3 102	3 727	4 556	5 492	6 893
Vershil (op individueel niveau	17	-96	-69	-53	1	307

De sociale bijdragen worden iets onderschat (gemiddeld met 5,4 %), en dit kennelijk vooral voor de hoge bedragen; voor de ambtenaren zijn de verschillen kleiner dan voor andere werknemers.



Figuur 1 is een scatterplot van het verschil tussen gesimuleerde en geobserveerde sociale bijdragen, naargelang het brutoloon of salaris. Deze vertoont een merkwaardig patroon: voor een groot aantal personen, en dit voor alle inkomensniveaus, zijn de sociale bijdragen onderschat met exact 1,85 % van het loon of salaris; dit veroorzaakt de diagonale streep van 0 naar rechtsonder in de grafiek. Verdere analyse liet zien dat deze onderschatting zich alleen voordoet bij bedienden in de privésector. Het strakke patroon duidt op een sociale bijdrage die niet gesimuleerd wordt; het is echter niet duidelijk welke bijdrage dit zou kunnen zijn.⁵ Om deze reden is deze afwijking niet gecorrigeerd.

2.2. Inkomsten van zelfstandigen

Het inkomen van zelfstandigen in de startdata is het netto belastbare inkomen, dus *na* afhouding van sociale zekerheidsbijdragen, en ook na aftrek van beroepskosten.⁶ Het beroepsinkomen van zelfstandigen in MIDAS is van deze variabele afgeleid; er is geen poging gedaan het bruto inkomen vóór deze aftrekken te imputeren.

2.3. Uitkeringen wegens ziekte en invaliditeit

Deze variabele omvat in MIDAS de uitkeringen wegens ziekte (door de mutualiteiten, geregistreerd door het NIC), invaliditeit (RIZIV), arbeidsongevallen (FAO) en beroepsziekten (FBZ). Vanwege de assumptie van éénzelfde sociaal statuut gedurende het gehele jaar, nemen we aan dat het in simulatie steeds gaat over een invaliditeitsuitkering, waarop maximum 3,5 % sociale bijdragen betaald moeten worden. Omdat de uitkering niet beneden een drempel mag zakken (die verschilt naargelang gezinssituatie), worden op de meeste invaliditeitsuitkeringen geen sociale bijdragen betaald, zoals Tabel 3 laat zien; de afwijking is in absolute zin klein, behalve voor het 90^{ste} percentiel.

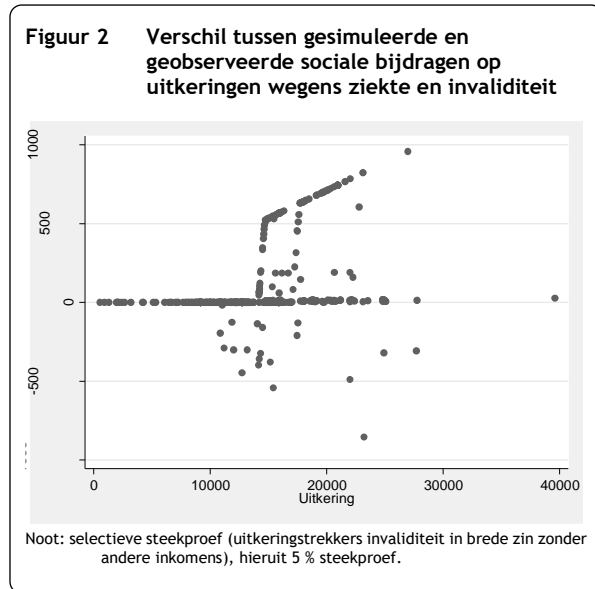
⁵ Het kan (bijvoorbeeld) niet gaan om de inhouding op het dubbel vakantiegeld in de privésector, omdat het dubbel vakantiegeld is opgenomen in het brutoloon volgens de Datawarehouse definitie. Consultatie van DVDL leverde geen uitsluitsel op.

⁶ De omschrijving van deze variabele in het Datawarehouse: "Het inkomen uit zelfstandige arbeid dat is opgenomen in deze variabele bestaat uit het nettobedrag van het beroepsinkomen. Het gaat hierbij over de bruto beroepsinkomen verminderd met de beroepskosten en eventueel met de beroepsverliezen. De sociale bijdragen (met inbegrip van de werkingskosten van de sociale verzekeringsfondsen en regularisaties van de startende zelfstandigen) zijn beroepskosten en zijn reeds afgehouden van dit inkomen. Dit inkomen is een element dat in een later stadium van het fiscaal proces wordt opgenomen in het globaal belastbaar inkomen, dat als belastbare basis dient voor de berekening van de personenbelasting. Het inkomen van de zelfstandigen is enkel beschikbaar op dit niveau."

Tabel 3 Resultaten voor simulatie sociale bijdragen op uitkeringen wegens ziekte en invaliditeit
Euro

	Gemiddelde	10 ^{de} percentiel	25 ^{ste} percentiel	Mediaan	75 ^{ste} percentiel	90 ^{ste} percentiel
Sociale bijdragen volgens Datawarehouse	114	0	0	0	0	573
Sociale bijdragen gesimuleerd in MIDAS	157	0	0	0	41	662
Vershil (op individueel niveau	42	0	0	0	0	189

Noot: Selectieve steekproef (uitkeringstrekkers invaliditeit in brede zin zonder andere inkomens), zowel werknemers als zelfstandigen



Figuur 2 toont dat de afwijkingen een duidelijk patroon volgen: voor een groot deel van de uitkeringen hoger dan 14 170 euro is de afwijking exact 3,55 % van de uitkering. Kennelijk betalen deze uitkeringsgerechtigden geen sociale bijdragen. De afwijkingen doet zich zowel voor bij werknemers als bij zelfstandigen. Wellicht gaat het hier om personen die zich nog in het 1^{ste} jaar van de arbeidsongeschiktheid bevinden, wanneer er nog geen sociale bijdragen verschuldigd zijn. Echter, ook bij personen die uitsluitend een invaliditeitsuitkering in strikte zin (van het RIZIV) ontvangen, zien we dit patroon verschijnen. Dit is nader te onderzoeken.

2.4. Pensioenen

Deze variabele omvat in MIDAS de wettelijke pensioenen wegens ouderdom of verweduwing. Op deze pensioenen is een ZIV-bijdrage verschuldigd van 3.55%, en ook een solidariteitsbijdrage van 2 %⁷; beide vanaf bepaalde minima die afhangen van de gezinstoestand (zie <https://www.sfpd.fgov.be/nl/betaling/brut-net#inhoudingen>). In de startdata beschikken we over het bruto-belastbare pensioenbedrag voor alle pensioenen gezamenlijk, inclusief 2^{de} pijler en buitenlandse pensioenen. Om deze reden is de selectie van gevallen verstrengd tot degene die alleen een wettelijk Belgisch pensioen ontvangen, naast de vereiste dat zij geen andere inkomens ontvangen.

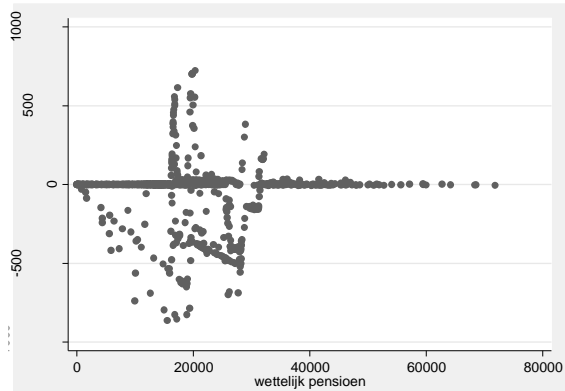
Tabel 4 Resultaten voor simulatie sociale bijdragen op pensioenen
€

	Gemiddelde	10 ^{de} percentiel	25 ^{ste} percentiel	Mediaan	75 ^{ste} percentiel	90 ^{ste} percentiel
Sociale bijdragen volgens Datawarehouse	589	0	0	107	884	1 778
Sociale bijdragen gesimuleerd in MIDAS	549	0	0	0	808	1 783
Vershil (op individueel niveau	-40	-333	0	0	2	24

Noot: Selectieve steekproef (uitkeringstrekkers wettelijk pensioen zonder andere inkomens of andere pensioenen)

⁷ In de originele *ssc_tax_process* was deze bijdrage per ongeluk op 20% gezet; dit is gecorrigeerd.

Figuur 3 Verschil tussen gesimuleerde en geobserveerde sociale bijdragen op wettelijke pensioenen
€



Noot: selectieve steekproef (uitkeringstrekkers wettelijk pensioen zonder andere inkomens of andere pensioenen), hieruit 2 % steekproef.

De verschillen tussen gesimuleerde en geobserveerde pensioenen zijn beperkt. Figuur 3 toont relatief veel positieve afwijkingen (gesimuleerd > geobserveerd) voor pensioenen rond 18 000 euro, en vooral negatieve voor pensioenen tussen 20 000 en 35 000 euro.

3. Het gezamenlijk belastbaar inkomen (GBI)

Allereerst maken we een vergelijking van het gezamenlijk belastbaar inkomen (GBI) volgens de IPCAL gegevens met de simulatie hiervan in MIDAS. Verschillen ontstaan ten eerste omdat we een aantal inkomsten niet kennen: onroerende inkomsten waaronder het kadastraal inkomen en diverse inkomsten; en ten tweede omdat gegevens ontbreken om een aantal aftrekken te simuleren, waaronder niet-forfaitaire beroepskosten en aftrekken voor enige woning en hypotheekinteressen. Ook hier wordt de selectieve steekproef gebruikt.

De IPCAL gegevens in de startdata bevatten, naast het GBI, ook een variabele met het totaal van de afzonderlijk belastbare inkomens. Deze omvatten een heteroog geheel van inkomsten, zowel beroepsinkomsten (onder meer opzeggingsvergoedingen en achterstallen) als andere inkomsten (bijv. toevallige winsten en baten, prijzen), die niet in het GBI worden opgenomen, maar wel worden belast, meestal aan de gemiddelde belastingvoet van de belastingplichtige. Sommige van de afzonderlijk belastbare beroepsinkomsten zijn opgenomen in de Datawarehouse variabele bruto lonen en salarissen, andere niet. De samenstelling van de afzonderlijk belastbare inkomens laat dus geen ondubbelzinnige conclusie toe of deze inkomens wel of niet moeten worden opgenomen in het belastbare inkomensbegrip waarmee het gesimuleerde belastbare inkomen zal worden vergeleken. De afzonderlijk belastbare inkomens zijn erg scheef verdeeld: slechts 21 % van degenen met $GBI > 0$ geniet deze inkomsten; binnen de groep van ontvangers bedragen de mediane afzonderlijke belastbare inkomsten slechts 640 euro, terwijl het gemiddelde 5 324 euro is. Initiële vergelijkingen toonden aan dat het gesimuleerde belastbare inkomen beter overeenkomt met het GBI zonder afzonderlijke belastbare inkomsten dan met het GBI inclusief deze inkomsten. Om deze reden is de analyse hieronder gebaseerd op het GBI zonder afzonderlijke belastbare inkomsten. Bij de vergelijking van de betaalde belastingen komen we echter terug op deze inkomenscomponent.

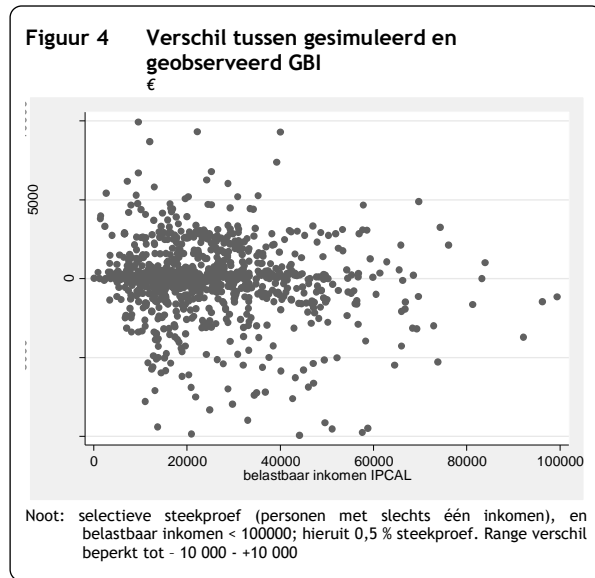
Voordat de vergelijking kon worden gemaakt, moest eerst een correctie worden uitgevoerd op de IPCAL variabele GBI. Om één of ander reden was deze variabele gelijk aan 0 voor een groot aantal gevallen – allen met een partner – die wel inkomen hadden; in deze gevallen was het GBI van het gezin (het fiscaal gezin) wellicht gelijk aan de som van belastbare inkomens van de twee partners. In alle andere gevallen was het GBI van het gezin gelijk aan het belastbare inkomen van de persoon. Voor de eerste soort gevallen werd het belastbaar inkomen geïmputeerd door – per individu – het verschil te nemen tussen het GBI van het gezin en het GBI van de partner.

Tabel 5 toont de resultaten. Hoewel er op individueel niveau er soms wel grote verschillen zijn, zijn de kengetallen (gemiddelden en percentielen) van het geobserveerde en het gesimuleerde GBI praktisch aan elkaar gelijk, en ook het gemiddelde verschil is erg klein.

Tabel 5 Resultaten voor simulatie gezamenlijk belastbaar inkomen
€

	Gemiddelde	10 ^{de} percentiel	25 ^{ste} percentiel	Mediaan	75 ^{ste} percentiel	90 ^{ste} percentiel
GBI volgens IPCAL	23 341	8 978	12 893	19 364	30 217	42 770
GBI berekend in MIDAS	23 381	9 131	12 723	19 303	30 344	42 650
Vershil (op individueel niveau)	39	-2 482	-621	-5	486	2 677

Noot: Selectieve steekproef (personen met slechts één inkomensbron, gepensioneerden met alleen een wettelijk pensioen); belastbaar inkomen in IPCAL < 100 000



Figuur 4 toont dat de afwijkingen geen duidelijk patroon vormen. Inderdaad is de correlatie van de afwijkingen met het geobserveerde GBI zeer klein (-0,05); met het gesimuleerde GBI is deze 0,36. De correlatie van het gesimuleerde GBI met het geobserveerde GBI bedraagt 0,91; de reden voor deze niet zo sterke correlatie moet gezocht worden in de ontbrekende gegevens voor een aantal inkomsten en aftrekposten die hierboven zijn genoemd.

4. De betaalde belasting

4.1. Volledige selectieve steekproef

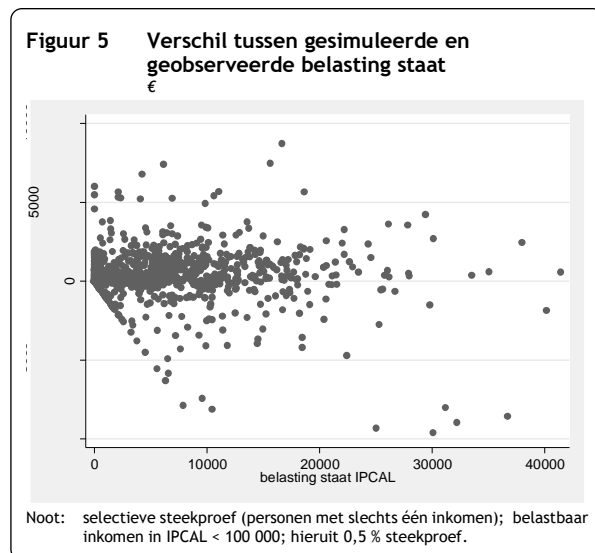
Gegeven de onvolkomen simulatie van het GBI, en omdat we daarenboven een aantal belastingverminderingen niet kunnen simuleren vanwege ontbrekende gegevens, vinden we uiteraard ook voor de belasting zelf afwijkingen tussen het geobserveerd en gesimuleerd bedrag. Het gaat hier om de variabele 'belasting staat', waar de gewestelijke en gemeentelijke opcentiemen nog niet zijn bijgeteld. We bestuderen eerst de afwijkingen voor de volledige selectieve steekproef, en vervolgens voor de substeekproeven waar het verschil tussen gesimuleerd en geobserveerd GBI beperkt is, of zelfs 0.

Tabel 6 Resultaten voor simulatie belasting staat
€

	Gemiddelde	10 ^{de} percentiel	25 ^{ste} percentiel	Mediaan	75 ^{ste} percentiel	90 ^{ste} percentiel
A. Alle belastingplichtigen						
Belasting staat volgens IPCAL	5 609	0	78	3 284	8 395	14 312
Belasting staat gesimuleerd in MIDAS	5 886	0	0	3 514	8 984	14 882
Vershil (op individueel niveau)	279	-749	-30	2	790	1 764
B. Belasting staat in IPCAL > 0						
Belasting staat in IPCAL	7 253	534	1 972	5 480	9 949	16 070
Belasting staat gesimuleerd in MIDAS	7 555	47	1 980	5 886	10 549	16 697
Vershil (op individueel niveau)	307	-1 083	-171	216	1 048	1 983

Noot: Selectieve steekproef (personen met slechts één inkomensbron); belastbaar inkomen in IPCAL < 100 000

Tabel 6 toont dat de gesimuleerde belasting de feitelijke belasting overschat met gemiddeld 5,0 % (4,2 % voor degenen die in feite belasting betalen); de afwijkingen stijgen met het bedrag van de belasting.



Voor 22,4 % van de selectieve steekproef wordt de belasting onderschat (het verschil is negatief); dit is waarschijnlijk te wijten aan niet-geobserveerde inkomensbronnen. De specifieke groep voor wie het gesimuleerde belastingbedrag 0 is, maar in feite wel belasting betaalt (7,4 %), bestaat voor het overgrote deel uit uitkeringstrekkers. Dankzij de belastingvermindering voor vervangingsinkomens zouden zij geen belasting betalen, als de uitkering hun enige inkomen was. Voor deze personen is bijna steeds het geobserveerde GBI (veel) hoger dan het gesimuleerde GBI, zodat zij kennelijk over andere inkomensbronnen beschikken. Daarnaast is er een beperkt aantal gevallen (voornamelijk gepensioneerden, daarnaast ook werklozen) die belasting betalen, ondanks een vrij laag belastbaar inkomen, ook volgens IPCAL. Deze gevallen doen in Figuur 5 de diagonale lijn vanaf het 0-punt naar rechtsonder ontstaan. Voor het overige duidt het patroon in Figuur 5 niet op systematische afwijkingen tussen gesimuleerde en geobserveerde belasting. Dit wordt bevestigd door

de correlaties: -0,027 en 0,335 tussen de afwijkingen en respectievelijk de geobserveerde en gesimuleerde belasting.

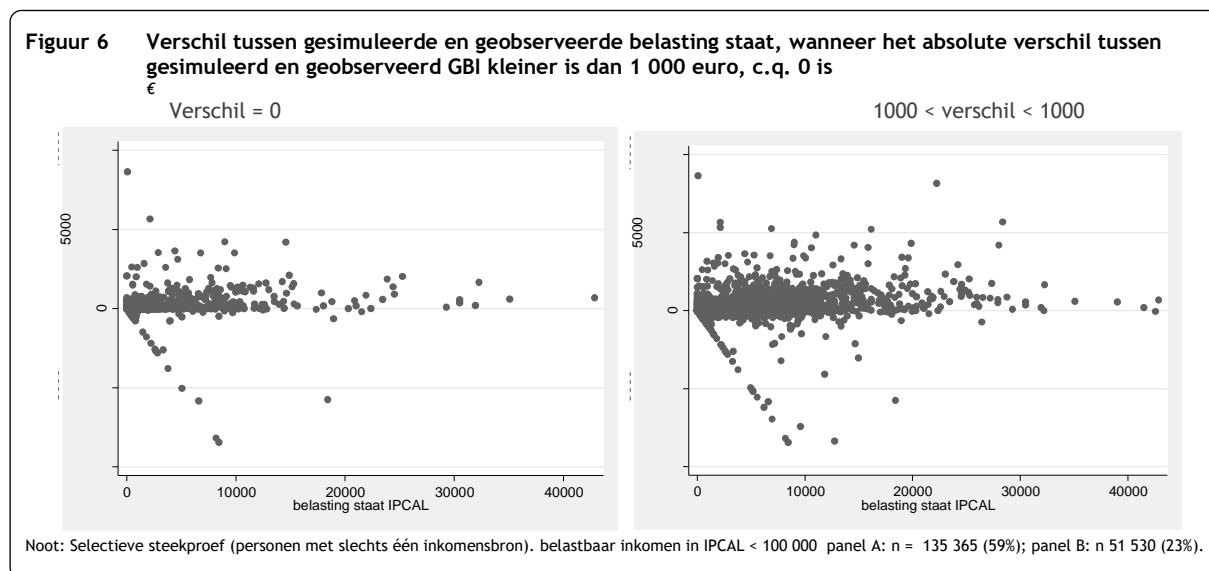
4.2. Steekproef waarin het verschil tussen gesimuleerd en geobserveerd GBI beperkt is, of zelfs 0

De hierboven getoonde afwijkingen tussen gesimuleerde en geobserveerde belasting staat zijn zowel het resultaat van onvolkomen simulatie van het GBI, als van de belasting, gegeven het GBI. In hoeverre is de simulatie van de belasting beter, als het verschil tussen gesimuleerd en geobserveerd GBI beperkt is, of zelfs 0? Tabel 7 en Figuur 6 tonen de resultaten wanneer dit laatste ligt tussen -1 000 en +1 000, en wanneer er geen verschil is.

Tabel 7 Resultaten voor simulatie belasting staat wanneer het absolute verschil tussen gesimuleerd en geobserveerd GBI kleiner is dan 1 000 euro, c.q. 0 is; alleen degenen voor wie de belasting > 0
€

	Gemiddelde	10 ^{de} percentiel	25 ^{ste} percentiel	Mediaan	75 ^{ste} percentiel	90 ^{ste} percentiel
A. $abs(GBI_{gesimuleerd} - GBI_{geobserveerd}) < 1000$						
Belasting staat volgens IPCAL	5 937	408	1 516	4 324	8 433	13 418
Belasting staat gesimuleerd in MIDAS	6 245	282	1 584	4 552	8 985	14 208
Vershil (op individueel niveau)	311	-252	-26	173	639	1 312
B. $abs(GBI_{gesimuleerd} - GBI_{geobserveerd}) = 0$						
Belasting staat volgens IPCAL	4 499	243	845	2 429	6 325	11 076
Belasting staat gesimuleerd in MIDAS	4 702	76	855	2 547	6 742	11 660
Vershil (op individueel niveau)	208	-175	-16	57	485	1 134

Noot: Selectieve steekproef (personen met slechts één inkomensbron), belastbaar inkomen in IPCAL < 100 000 en belasting staat > 0.
panel A: n = 94 529 (53 %); panel B: n = 27 846 (16%).
Percentages ten opzichte van totaal selectieve steekproef met belasting staat > 0.



Vergeleken met de volledige steekproef, daalt het gemiddelde verschil tussen gesimuleerde en geobserveerde belasting niet, na beperking tot gevallen waar het gesimuleerde GBI minder dan 1 000 euro (in absolute zin) verschilt van het geobserveerde GBI. En ook waar er geen verschil is voor GBI, is het gemiddelde verschil voor de belasting niet dramatisch lager. De spreiding is echter wel veel kleiner: de standaardafwijkingen van het verschil tussen gesimuleerde en geobserveerde belasting (onder gevallen waar de belasting staat > 0) zijn respectievelijk euro 1 919, euro 1 915 en euro 1 558, in de

volledige steekproef, de steekproef met beperkt verschil voor GBI, en de steekproef met geen verschil in GBI. Opvallend is verder dat, net als het gemiddelde verschil, het mediane verschil tussen gesimuleerde en geobserveerde belasting ongeveer gelijk is in de verschillende steekproeven; de vermindering van de spreiding komt vooral door een beperking van relatief grote afwijkingen. Een voor de hand liggende verklaring is dat de omvang van de meeste belastingverminderingen beperkt is, terwijl er geen limiet staat op de niet-geobserveerde inkomenscomponenten in het GBI. Bedacht moet verder worden dat de selecties op basis van het verschil in GBI niet neutraal is wat betreft het belastbaar inkomen: grote verschillen in GBI komen vooral voor bij hoge belastbare inkomens, hetgeen ook blijkt uit de lagere gemiddelde belastingbedragen in Tabel 7.

4.3. De bijzondere bijdrage aan de sociale zekerheid

Zoals in de inleiding vermeld, beschikken wij in de startdata niet over gegevens inzake de bijzondere bijdrage aan de sociale zekerheid. Bij de berekening van het netto-inkomen houden wij ter volledigheid toch rekening met deze bijdrage (zie onder). Om deze reden lijkt het raadzaam de omvang en spreiding van deze inkomensafdracht te tonen.

De bijzondere bijdrage wordt geheven bij werknemers, evenals bij de ambtenaren van de provincies en gemeenten en de gerechtigden op geïndexeerde socialezekerheidsuitkeringen. Niet onderworpen aan deze bijzondere bijdrage voor de sociale zekerheid zijn de fiscale gezinnen die uitsluitend bestaan uit zelfstandigen of uit personen die een rust- of overlevingspensioen genieten als zelfstandige, en de fiscale gezinnen die bestaan uit een zelfstandige wiens echtgeno(o)t(e) een minder dan halftijdse activiteit uitoefent als werknemer of daarmee gelijkgestelde. Het GBI van het fiscaal gezin, dus inclusief de inkomsten van niet-werknemers, vormt de berekeningsbasis. De bijdrage bedraagt 9 % van het GBI tussen 18 592 en 21 070 euro, 1 % van het GBI tussen 21 070 en 60 162 euro, en 0 % van het inkomen boven die grens. Hieruit volgt dat de maximumbijdrage 731 euro bedraagt. De inkomensdrempels worden niet geïndexeerd.

In de simulatie kunnen we geen rekening houden met het onderscheid tussen ambtenaren bij gemeenten en provincies en andere ambtenaren, zodat we aannemen dat alle ambtenaren niet onderworpen zijn.⁸ Ook weten wij niet wie een pensioen geniet als zelfstandige, zodat we dit element van de regelgeving negeren; in de praktijk blijft voor de grote meerderheid van de zelfstandige gepensioneerden het pensioen beneden de onderste inkomensgrens. De wetgeving specificeert niet hoe deze gezinsbijdrage moet verdeeld worden tussen twee partners binnen hetzelfde fiscale gezin die allebei werknemer zijn⁹. In de simulatie is ieders aandeel proportioneel aan het verschil tussen hun GBI en de helft van 21 070, ten opzichte van het bedrag waarmee het GBI van het gezin de onderste grens van 21 070 euro overschrijdt.

⁸ Dit betekent dat er ongelijkheid zou bestaan tussen ambtenaren waarvan de partner werknemer is en ambtenaren waarvan de partner ook ambtenaar is, omdat de bijzondere bijdrage wel geheven wordt op het salaris van de eerste (via de partner), maar niet op dat van de tweede.

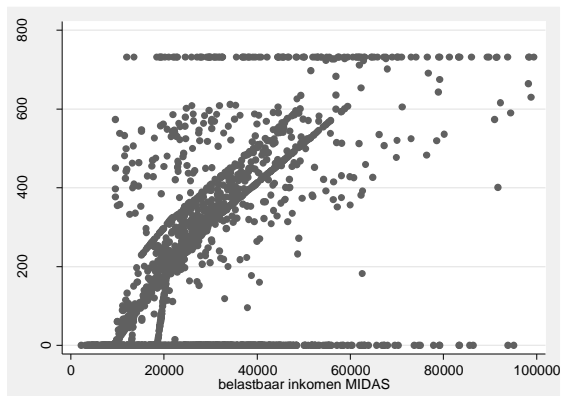
⁹ Voor zover wij hebben gezien.

Tabel 8 Resultaten voor de gesimuleerde bijzondere bijdrage aan de sociale zekerheid
€

	Gemiddelde	10 ^{de} percentiel	25 ^{ste} percentiel	Mediaan	75 ^{ste} percentiel	90 ^{ste} percentiel
Bijzondere bijdrage	158	0	0	33	284	437

Noot: Selectieve steekproef (personen met slechts één inkomensbron), belastbaar inkomen in IPCAL < 100 000

Figuur 7 De gesimuleerde bijzondere bijdrage aan de sociale zekerheid, naargelang gesimuleerd belastbaar inkomen
€



Noot: selectieve steekproef (personen met slechts één inkomen; belastbaar inkomen in IPCAL < 100 000; hieruit 2 % steekproef.

De gemiddelde bijzondere bijdrage blijft beperkt tot 158 euro; de variatie is gelimiteerd door het impliciete plafond. Figuur 7 toont het enigszins merkwaardige verloop van de individuele bijzondere bijdrage naargelang het belastbaar inkomen. Sommige werknemers met een vrij laag loon betalen toch een bijna maximale bijzondere bijdrage aan de sociale zekerheid, wanneer hun gezin een hoog belastbaar inkomen geniet dankzij een partner die ambtenaar of zelfstandige is met substantiële inkomsten.

5. Het netto-inkomen

5.1. Globale verschillen

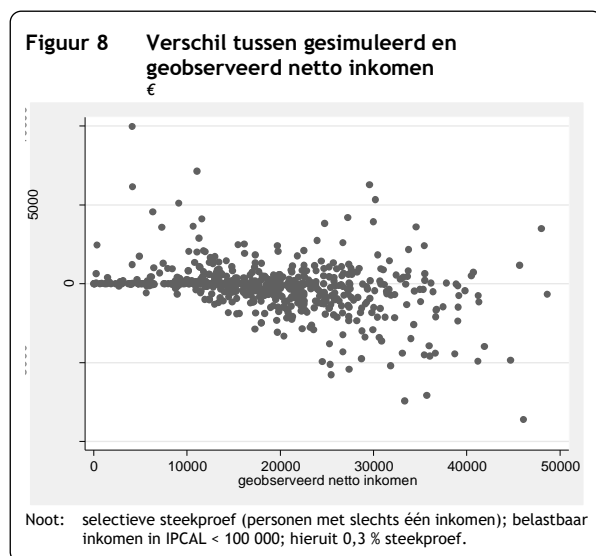
Een juiste schatting van het netto-inkomen is het uiteindelijke doel van de belastingmodule in MIDAS. Het geobserveerde netto-inkomen wordt berekend door de som van de inkomens te nemen die meegenomen worden in de berekening van het belastbaar inkomen, en door ons geobserveerd worden (lonen en salarissen, inkomsten van zelfstandigen, wettelijke pensioenen, uitkeringen van de RVA, uitkeringen wegens ziekte en invaliditeit), en daar de geobserveerde sociale bijdragen en belastingen, inclusief de bijzondere bijdrage, van af te trekken. De belastingen worden eerst vermeerderd met de gemiddelde gemeentelijke opcentiemen per regio. (N.B., dit netto-inkomen is niet gelijk aan het normale netto-inkomen; dat laatste omvat ook bijstandsuitkeringen en kinderbijslag.) Het uitgangspunt voor zowel het geobserveerde als gesimuleerde netto-inkomen is dus gelijk: het totaal van de genoemde bruto-inkomenscomponenten, die geobserveerd zijn in de startdata afkomstig van het Datawarehouse. Verschillen in netto-inkomen ontstaan door afwijkingen tussen de geobserveerde en gesimuleerde sociale bijdragen en belastingen.

Tabel 9 Resultaten voor simulatie netto-inkomen, naargelang het verschil tussen gesimuleerd en geobserveerd GBI
€

	Gemiddelde	10 ^{de} percentiel	25 ^{ste} percentiel	Mediaan	75 ^{ste} percentiel	90 ^{ste} percentiel
A. Alle belastingplichtigen						
Netto inkomen geobserveerd*	19 058	8 817	12 772	17 705	23 926	30 814
Netto inkomen gesimuleerd*	18 824	9 177	13 045	17 483	23 435	30 051
Vershil (op individueel niveau)	-231	-1 704	-684	0	157	984
B. $abs(GBI_{gesimuleerd} - GBI_{geobserveerd}) < 1000$						
Netto inkomen geobserveerd*	17 121	7 962	12 364	15 900	21 459	27 443
Netto inkomen gesimuleerd*	16 937	8 080	12 354	15 852	21 177	26 900
Vershil (op individueel niveau)	-181	-991	-370	0	22	385
C. $abs(GBI_{gesimuleerd} - GBI_{geobserveerd}) = 0$						
Netto inkomen geobserveerd*	14 702	5 832	11 007	13 860	17 570	23 358
Netto inkomen gesimuleerd*	14 567	6 037	11 006	13 956	17 340	22 894
Vershil (op individueel niveau)	-132	-729	-150	0	0	188

Noot: * Zie tekst voor definitie netto inkomen
Selectieve steekproef (personen met slechts één inkomensbron), belastbaar inkomen in IPCAL < 100 000.

Binnen de selectieve steekproef (panel A van Tabel 9) is het gemiddelde verschil tussen het geobserveerde en het netto inkomen slechts 231 euro; het laatste is 1,2 % lager dan het eerste. De onderschattingen situeren zich vooral bij de inkomens rond en boven de mediaan; lagere inkomens worden vaker overschat. Figuur 8 toont dit ook, en laat verder zien dat, zoals te verwachten, de spreiding van de afwijkingen toeneemt met het netto geobserveerde inkomen. Wanneer de steekproef verder wordt ingeperkt tot gevallen waar het verschil tussen het geobserveerde en het gesimuleerde belastbare inkomen (GBI) klein is, of zelfs 0 (panels B en C), bewegen de gemiddelde bedragen zich maar in zekere mate naar elkaar toe: de negatieve afwijkingen worden kleiner, maar de positieve eveneens, en de laatste verdwijnen zelfs bijna geheel (panel C; er blijft een klein aantal positieve afwijkingen op rond en boven het 90^{ste} percentiel). Waar er geen verschil is tussen geobserveerd en gesimuleerd GBI is kennelijk de gesimuleerde belasting bijna altijd hoger dan de feitelijke belasting,



vanwege het niet kunnen simuleren van een aantal belastingverminderingen. Hieruit kunnen we voorzichtig afleiden dat positieve afwijkingen waarschijnlijk vooral ontstaan door ontbrekende gegevens over inkomsten.

Bij opsplitsing van het verschil tussen gesimuleerd en geobserveerd netto-inkomen naargelang sociaal statuut (tabel 10) zien we dat er voor de grote meerderheid van de werklozen en invaliden dit verschil 0 is. (Het gemiddelde voor de werklozen wordt omhoog getrokken door enkele extreme waarden.) Voor de gepensioneerden is het gemiddelde en de mediane verschil ook beperkt,

maar is er zekere spreiding van de verschillen. Bij werknemers, ambtenaren en zelfstandigen zijn de afwijkingen aanzienlijk groter, zowel qua gemiddelde en mediaan, als qua spreiding.

Tabel 10 Kengetallen voor verschil tussen gesimuleerd en geobserveerd netto-inkomen, naargelang sociaal statuut
€

	Gemiddelde	10 ^{de} percentiel	25 ^{ste} percentiel	Mediaan	75 ^{ste} percentiel	90 ^{ste} percentiel
Werknemers (niet-ambtenaren)	-537	-2 147	-1 151	-356	242	983
Ambtenaren	-59	-2 171	-1 110	-244	661	2 118
Zelfstandigen	-881	-2 766	-1 482	-475	0	604
Werklozen	157	0	0	0	0	189
Invaliden	64	-564	0	0	0	356
Gepensioneerden	93	-510	-99	0	114	633

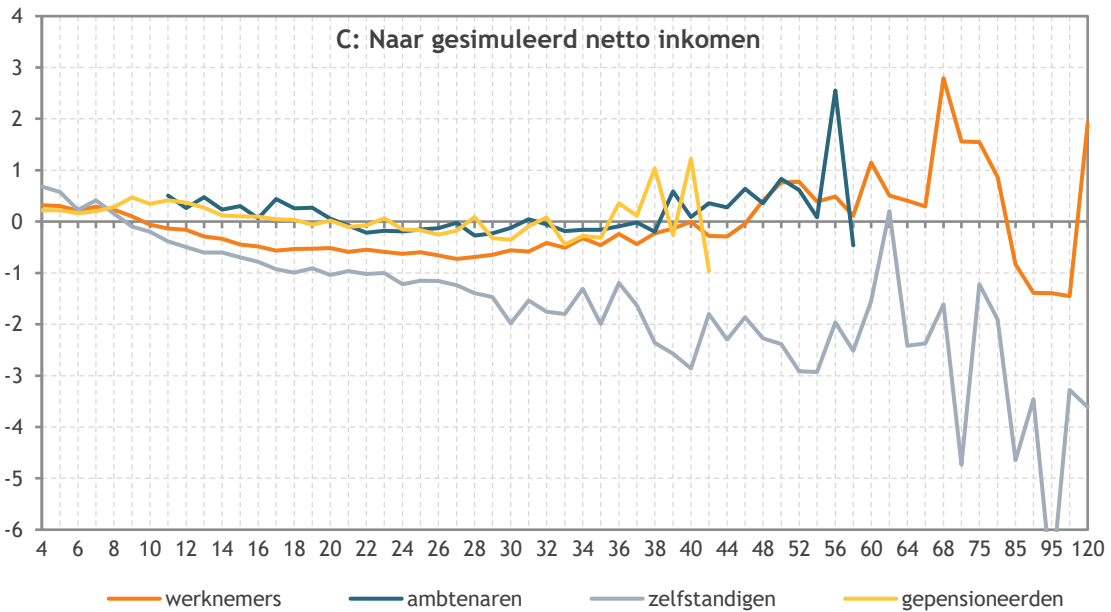
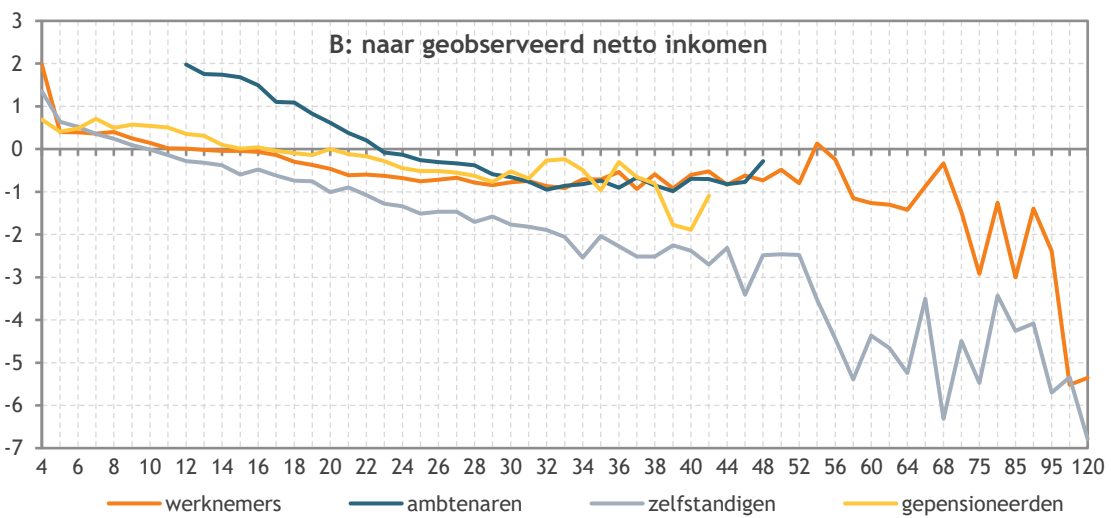
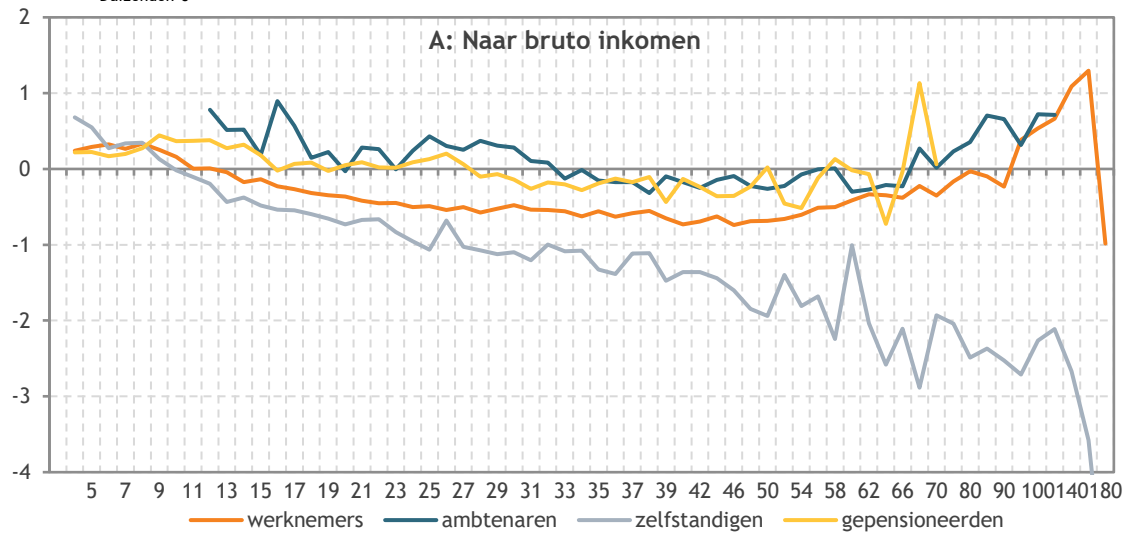
Noot: Selectieve steekproef (personen met slechts één inkomensbron), belastbaar inkomen in IPCAL < 100 000.

Gegeven deze vaststellingen, lijkt er voor werklozen en invaliden geen correctie nodig te zijn; de (beperkte) afwijkingen voor een klein percentage gevallen kunnen we negeren. We concentreren ons daarom hieronder op de werkenden en de gepensioneerden.

5.2. Afwijking tussen simulatie en observatie naargelang niveau inkomen

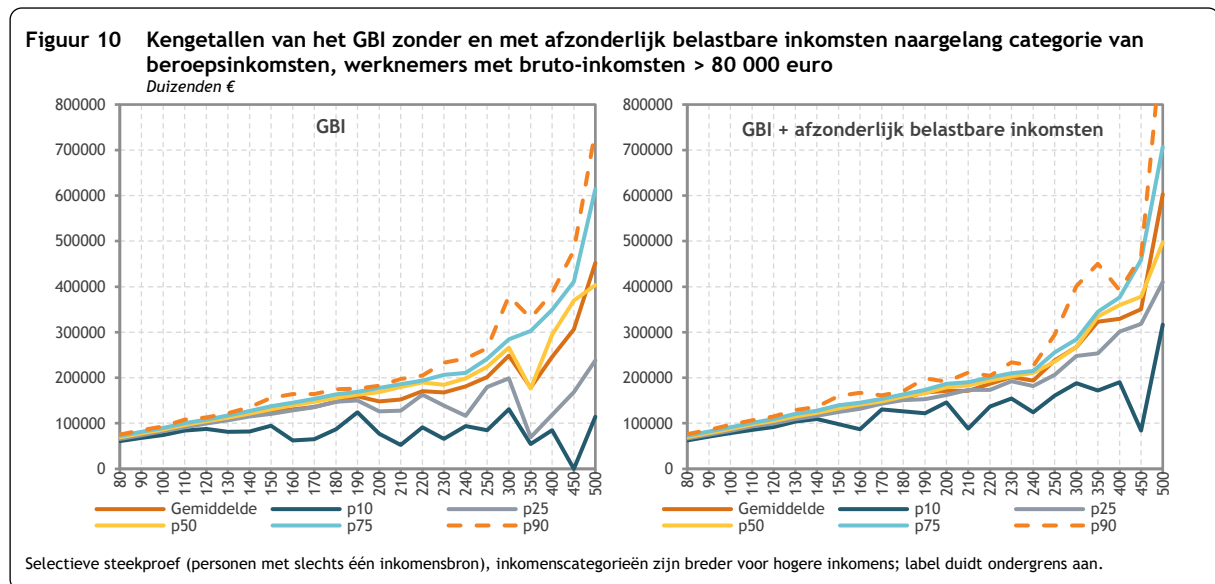
Belangrijk voor een correcte kalibratie is of en hoe de verschillen tussen gesimuleerd en geobserveerd netto-inkomen variëren naargelang het inkomensniveau. Figuur 9 toont deze verschillen naargelang categorieën van verschillende inkomensconcepten: het totaal bruto inkomen vóór sociale zekerheidsbijdragen, het geobserveerd netto inkomen (het netto inkomen na bijdragen en belastingen afgeleid van variabelen uit het Datawarehouse en IPCAL), en het gesimuleerd netto inkomen. Naargelang bruto-inkomen (grafiek A) daalt het verschil voor alle groepen van positief naar negatief als het bruto inkomen stijgt, tot ongeveer 60 000 euro. Daarna daalt het verder en dramatisch voor zelfstandigen, maar stijgt het weer voor werknemers en ambtenaren (bij de gepensioneerden zijn er nauwelijks gevallen boven deze inkomensgrens). De redenen voor deze stijging zijn echter verschillend: bij de ambtenaren met de hoogste salarissen wordt in simulatie het GBI meer onderschat dan bij ambtenaren met lagere salarissen (waarschijnlijk vanwege niet-geobserveerde componenten in het GBI), zodat voor de eersten de belastingen (licht) worden onderschat in plaats van overschat, en hierdoor komt het gesimuleerd netto inkomen dichterbij het geobserveerd netto inkomen te liggen. Bij de werknemers met de hoogste inkomens zijn kennelijk de sociale zekerheidsbijdragen in de simulatie onderschat; zie sectie 2.1. Het toenemende verschil bij de zelfstandigen is te wijten aan nog een ander fenomeen, zie onder.

Figuur 9 Gemiddeld verschil gesimuleerd netto inkomen - geobserveerd netto inkomen, naargelang verschillende inkomensconcepten
 Duizenden €



Noot: Selectieve steekproef (personen met slechts één inkomensbron), inkomenscategorieën zijn breder voor hogere inkomens; label duidt ondergrens aan. Minimum aantal observaties per grafiekpunt: 25

Een ander patroon zien we wanneer het verschil tussen gesimuleerd en geobserveerd netto-inkomen wordt uitgezet tegen het geobserveerd netto-inkomen (grafiek B). Voor werknemers blijft hier het verschil dalen naargelang stijgend inkomen, ook bij de hoogste inkomens. Figuur 10 (links) toont de directe reden hiervoor: in de hogere inkomensregionen van de werknemers blijven het 10^{de} en het 25^{ste} percentiel van het GBI achter bij het gemiddelde en de mediaan (p50). Wanneer we de afzonderlijk belastbare inkomsten bij het GBI optellen (rechtse grafiek), dan verlopen de curves voor het 10^{de} en het 25^{ste} percentiel duidelijk hoger, terwijl de mediaan en het gemiddelde nauwelijks verschuiven. De meest plausibele interpretatie is dat voor een belangrijke minderheid onder de personen met zeer hoge beroepsinkomsten, die laatste voor een groot deel bestaan uit uitzonderlijke, en afzonderlijk belastbare ontvangsten. Het kan bijvoorbeeld gaan om opzeggingsvergoedingen, die niet in het GBI worden opgenomen, maar wel worden belast, meestal aan de gemiddelde belastingvoet van de belastingplichtige, die uiteraard lager is dan de marginale belastingvoet. Ditzelfde fenomeen doet zich ook voor bij de zelfstandigen met zeer hoge inkomens.



Het patroon van het verschil tussen gesimuleerd en geobserveerd netto-inkomen naargelang het gesimuleerd netto-inkomen (Figuur 9 grafiek C), is weer anders, zonder de sterke daling bij de hoogste inkomens. Grote verschillen (het gesimuleerde netto-inkomen is veel lager dan het geobserveerde) doen de betreffende gevallen terecht komen in lagere categorieën van het gesimuleerde netto-inkomen. Omdat het gesimuleerde netto-inkomen mede het resultaat is van de afwijkingen, is deze grafiek moeilijk te interpreteren; ze wordt hier getoond als documentatie van de situatie die door de kalibratie zoveel als mogelijk rechtgetrokken moet worden.

6. Kalibratie

6.1. Methode

De kalibratie steunt op de volgende methodologische keuzes.

- We gebruiken de selectieve steekproef, dat wil zeggen, personen met slechts één inkomensbron. In de MIDAS simulatie hebben alle personen slechts één inkomensbron, en de belastingmodule is daarop toegesneden. Dit betekent uiteraard dat als ooit de assumptie van één inkomensbron per persoon per jaar verlaten zou worden, de kalibratie (evenals de belastingmodule) herzien moet worden.
- De kalibratie moet uiteraard gebruik maken van variabelen die in de simulatie bekend zijn.
- Uitgaande van deze variabelen wordt een schatting gemaakt van het verschil tussen geobserveerd en gesimuleerd netto-inkomen, en dit wordt van het gesimuleerd netto-inkomen afgetrokken om het gekalibreerd netto-inkomen te berekenen. Een alternatief zou zijn om rechtstreeks het gekalibreerd netto-inkomen te schatten, maar in dat geval zouden we de informatie die het gesimuleerd netto-inkomen bevat niet gebruiken.
- De kalibratie wordt uitgetest voor zowel de absolute afwijkingen, als voor de logaritme van de proportionele afwijkingen: $\log(\text{netto-inkomen-gesimuleerd} / \text{netto-inkomen-geobserveerd})$.
- We gebruiken kwantiel-regressie voor de mediaan, d.w.z. de te schatten maat is niet het conditioneel gemiddelde, maar de conditionele mediaan. De reden is dat de afwijkingen vaak een niet-normale verdeling hebben, met bijvoorbeeld een grote concentratie van observaties met waarde 0, en een beperkt aantal erg extreme waarden.
- Het succes van de kalibratie wordt beoordeeld aan de hand van twee maatstaven:
 - het gemiddelde kwadraat van de afwijkingen, c.q. de logaritme van de proportionele afwijkingen (Mean Square Deviation) Deze bevat niet alleen de variantie van de afwijkingen, maar ook de gemiddelde afwijking.
 - het verloop van de (proportionele) afwijkingen naargelang het totaal inkomen.
- De kalibratie geschiedt per beroepsgroep.
- Als onafhankelijke variabelen worden gebruikt:
 - totaal bruto inkomen (in duizenden euro's), en hiervan het kwadraat en de derde macht. Door middel van een dummy variabele voor bruto inkomens hoger dan 60 000 euro, en interactietermen van de genoemde variabelen met deze dummy variabele, wordt toegelaten dat het verband tussen de afhankelijke variabele en het totaal bruto-inkomen anders verloopt beneden en boven 60 000 euro.
 - de gesimuleerde socialezekerheidsbijdragen (in duizenden euro's), en hiervan het kwadraat.
 - de gesimuleerde belastingen (in duizenden euro's), en hiervan het kwadraat.
 - de leeftijd, en hiervan het kwadraat.

- een partner of niet
- het aantal kinderen ten laste, in categorieën, met 4 en meer als hoogste categorie.

6.2. Resultaten

De regressie van de absolute afwijkingen en die van de logaritme van de proportionele afwijkingen gaven erg gelijkaardige resultaten wat betreft de kwaliteit van de resulterende kalibratie, beoordeeld aan de hand van gemiddelde kwadraat van de afwijkingen verloop van de afwijkingen naargelang het totaal inkomen. Voor relatief lage inkomens produceerde de regressie van de absolute afwijkingen soms iets betere resultaten, zodat hiervoor gekozen is. Tabel 11 rapporteert de resultaten van deze quantiel-regressies. Niet alle variabelen bleken een significant effect te hebben in alle regressies.

Tabel 11 Regressie van verschil tussen gesimuleerd en geobserveerd netto inkomen, per sociaal statuut-groep
€

	Werknemers		Ambtenaren		Zelfstandigen		Gepensioneerden	
	Coëfficiënt	st. fout	Coëfficiënt	st. fout	Coëfficiënt	st. fout	Coëfficiënt	st. fout
Bruto inkomen	-0,0235***	(0,00318)	-0,0270***	(0,00172)	-0,0185***	(0,00250)	0,000896***	(0,000263)
Bruto inkomen kwadraat	-0,00172***	(0,000163)			0,000232***	(4,50e-05)	2,70e-05***	(8,97e-06)
Bruto inkomen derde macht			4,99e-06***	(1,12e-06)				
Bruto inkomen is hoger dan 60 000	-1,534***	(0,142)	-8,014***	(1,117)	-2,180***	(0,267)		
Bruto inkomen boven 60 000	0,0330***	(0,00248)	0,165***	(0,0246)	0,0600***	(0,00455)		
Bruto inkomen boven 60 000 kwadraat	-0,000269***	(2,89e-05)	-0,00148***	(0,000293)	-0,000985***	(5,39e-05)		
SZ bijdragen	-0,0677***	(0,0163)					-0,0770***	(0,00392)
SZ bijdragen kwadraat	0,116***	(0,00892)						
Belasting					-0,0778***	(0,00694)	0,00182***	(0,000638)
Belasting kwadraat					0,00294***	(0,000166)	-0,00109***	(5,32e-05)
Leeftijd	0,0155***	(0,000908)	0,0433***	(0,00180)	0,0142***	(0,00119)		
Leeftijd kwadraat	0,000530***	(3,93e-05)	0,00144***	(0,000104)	0,000575***	(6,25e-05)		
Met partner	-0,210***	(0,00968)	-0,349***	(0,0229)	-0,0895***	(0,0172)		
0 afhankelijke kinderen	Ref. cat							
1 afhankelijk kind	0,0341**	(0,0138)	0,0853***	(0,0301)	0,0408*	(0,0244)		
2 afhankelijke kinderen	-0,00808	(0,0144)	0,0753**	(0,0319)	0,0564**	(0,0247)		
3 afhankelijke kinderen	0,0280	(0,0231)	0,0213	(0,0529)	0,149***	(0,0378)		
4+ afhankelijke kinderen	0,206***	(0,0400)	-0,0293	(0,0975)	0,284***	(0,0645)		
Constante	-0,357***	(0,0564)	-1,139***	(0,111)	-0,633***	(0,0631)	0,000787	(0,00354)
Observaties	88 811		23 494		22 214		56 441	

Noot: Alle bedragen in duizenden euro's. Selectieve steekproef (personen met slechts één inkomensbron). *** p<0,01, ** p<0,05, * p<0,1

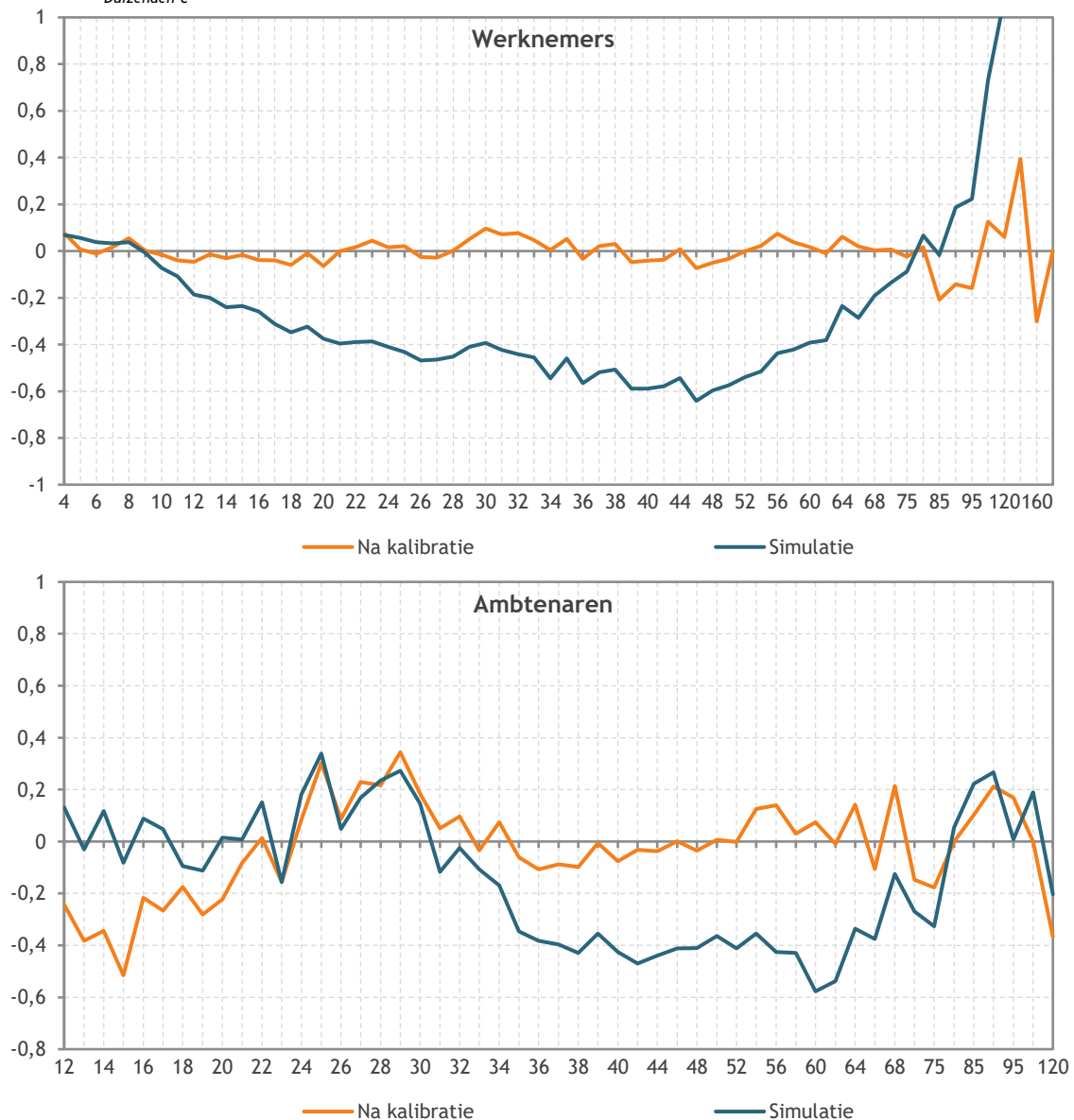
Op basis van de coëfficiënten in Tabel 9 kan het verschil tussen gesimuleerd en geobserveerd netto inkomen geschat worden. Het gesimuleerde netto inkomen wordt vervolgens gekalibreerd door dat geschatte verschil er vanaf te trekken. Tabel 10 toont de gemiddelde afwijking, en het gemiddeld kwadraat van de afwijkingen tussen, vóór en na kalibratie. Figuur 11 toont het verloop van de mediane afwijking naargelang het bruto inkomen, vóór (simulatie) en na kalibratie.

Tabel 12 Kengetallen voor de kwaliteit van de kalibratie

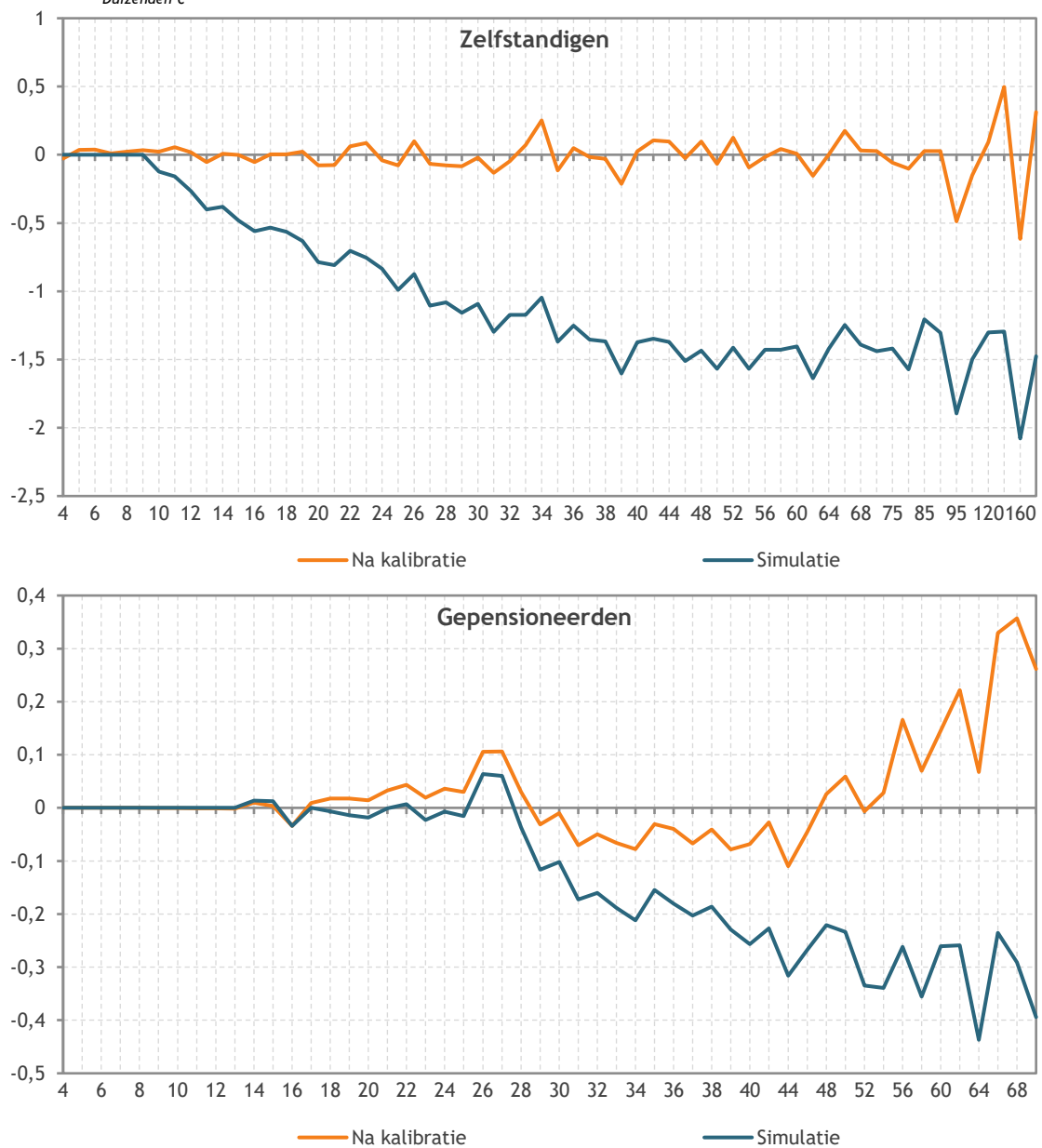
	Gemiddelde afwijking		Gemiddeld kwadraat van de afwijkingen	
	Vóór kalibratie	Na kalibratie	Vóór kalibratie	Na kalibratie
Werknemers	-0,439	-0,061	2,803	2,509
Ambtenaren	-0,110	0,159	3,962	3,856
Zelfstandigen	-0,706	0,013	4,046	3,108
Gepensioneerden	0,083	0,118	0,921	0,917

Noot: In duizenden euro's. Selectieve steekproef (personen met slechts één inkomensbron).

Figuur 11 Mediaan van verschil tussen gesimuleerd netto inkomen - geobserveerd netto inkomen, gesimuleerd (vóór kalibratie) en na kalibratie, naargelang bruto inkomen, per sociaal statuut groep
Duizenden €



Figuur 11 **vervolg**
Duizenden €



Noot: Selectieve steekproef (personen met slechts één inkomensbron), inkomenscategorieën zijn breder voor hogere inkomens; label duidt ondergrens aan. Minimum aantal observaties per grafiekpunt: 25

Voor werknemers wordt het gemiddelde verschil sterk gereduceerd door de kalibratie. De mediane afwijking is – na kalibratie – praktisch gelijk aan 0 voor alle bruto-inkomens (Figuur 11). De gemiddelde gekwadraterde afwijking daalt echter slechts met 10%. De variabelen in de regressie verklaren slechts een klein deel van de variatie in de afwijkingen tussen gesimuleerd en geobserveerd netto inkomen, zodat het overgrote deel van de afwijkingen ook na kalibratie blijft bestaan. De kalibratie zorgt ervoor dat de afwijkingen niet meer systematisch met het bruto inkomen samenhangen. Voor ambtenaren werkt de kalibratie minder goed: de gemiddelde afwijking is – in absolute waarde – groter ná dan vóór kalibratie, en de gekwadraterde afwijking daalt nauwelijks. Ook het verloop van het mediane verschil naargelang bruto inkomen is minder overtuigend, vooral voor inkomens beneden 22 000 euro (hierbij moet wel bedacht worden dat slechts 2,9 % van de ambtenaren in de steekproef een bruto inkomen

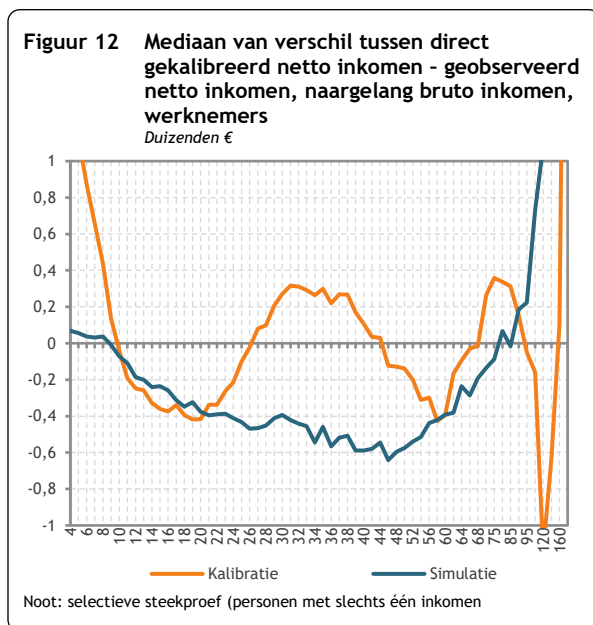
beneden die grens heeft). Voor de zelfstandigen neemt de kalibratie het overgrote deel van de systematische afwijking weg, d.w.z. de afwijking voor zover die samenhangt met het bruto inkomen; de gemiddelde gekwadraterde afwijking daalt met 23 %. Voor de gepensioneerden is de kalibratie geen succes: voor de bruto pensioenen tussen 20 000 en 30 000 (23 % van de gepensioneerden) leidt deze zelfs tot een vergroting van de afwijkingen. Andere specificaties van de regressievergelijking brachten hierin geen verbetering. Een vermindering van de verschillen is er pas voor de pensioenen boven 28 000 euro; deze groep omvat 15 % van alle gepensioneerden. Om deze reden is voor de gepensioneerden geen kalibratie toegepast.

6.3. Implementatie

De kalibratie is geïmplementeerd in MIDAS aan het einde van de module *ssc_tax_process*. Zoals hierboven aangegeven, worden alleen de netto-inkomens van werknemers, ambtenaren en zelfstandigen gekalibreerd; die van gepensioneerden en andere uitkeringstrekkers worden ongewijzigd overgenomen van de simulatie. Om de regressievergelijking te kunnen toepassen worden alle bedragen van inkomens, sociale bijdragen en belastingen eerst teruggerekend naar waarden voor 2012 aan de hand van de index van de productiviteit, die in MIDAS wordt gebruikt voor de aanpassing van de reële lonen, salarissen en inkomsten van zelfstandigen.¹⁰ De geschatte afwijking wordt eerst vermenigvuldigd met dezelfde index, en vervolgens van het gesimuleerde netto-inkomen afgetrokken.

6.4. Beoordeling

Een eerste vraag is of, gegeven de resultaten van de kalibratie, we het simulatiemodel van de belastingen en sociale bijdragen überhaupt wel nodig hebben. Kunnen we niet evengoed het netto inkomen via regressie afleiden uit het totaal inkomen en andere variabelen, zoals leeftijd? Het antwoord is: nee. Als test is een regressiemodel geschat (alleen voor werknemers) met dezelfde variabelen als



hierboven genoemd, maar met het geobserveerde netto inkomen zelf als afhankelijke variabele, in plaats van het verschil tussen gesimuleerd en geobserveerd netto inkomen. Op basis van de regressiecoëfficiënten kan dan het netto-inkomen geschat worden. In dat geval zijn de resultaten praktisch gelijk aan die welke hierboven getoond zijn. Echter, die regressievergelijking bevat variabelen die we slechts kennen dankzij de simulatie, met name de belastingen. Als we deze variabelen weglaten, is het resultaat duidelijk slechter dan bij simulatie gevolgd door kalibratie: het gemiddelde kwadraat van de afwijkingen is 3.25 in plaats van 2.51. Figuur 12 laat zien dat zulk een 'directe' kalibratie zonder voorafgaande

¹⁰ Global PRODGRCUM

simulatie niet duidelijk beter is dan een simulatie zonder kalibratie, en duidelijk slechter presteert dan simulatie gevolgd door kalibratie (vergelijk met Figuur 11, grafiek 'werknemers').

Tabel 13 vergelijkt het gesimuleerd en gekalibreerd netto inkomen met het geobserveerde bedrag. Na kalibratie is het gemiddeld negatieve verschil (zie Tabel 9) omgeslagen in een klein positief verschil. Het mediane verschil is 0; alle kengetallen liggen dicht bij elkaar. De spreiding van het verschil is na kalibratie nauwelijks kleiner geworden. Ook opgesplitst per beroepsgroep zijn de resultaten gelijkaardig: na kalibratie klein gemiddeld verschil; geen of verwaarloosbaar mediaan verschil, nauwelijks beperking van de spreiding van de afwijkingen.

Tabel 13 Vergelijking van gesimuleerd en gekalibreerd netto-inkomen met het geobserveerd netto-inkomen, naargelang sociaal statuut
€

	Gemiddelde	10 ^{de} percentiel	25 ^{ste} percentiel	Mediaan	75 ^{ste} percentiel	90 ^{ste} percentiel
<i>A. Allen met bruto inkomen > 0</i>						
Netto inkomen geobserveerd*	18 869	8 565	12 762	17 659	23 899	30 800
Netto inkomen gesimuleerd en gekalibreerd*	18 984	9 070	13 058	17 731	23 856	30 622
Vershil (op individueel niveau)	139	-1 209	-335	0	419	1 329
<i>B. Werknemers</i>						
Netto inkomen geobserveerd*	22 611	12 307	17 263	21 352	27 095	34 150
Netto inkomen gesimuleerd en gekalibreerd*	22 575	12 416	17 432	21 287	26 978	33 993
Vershil (op individueel niveau)	-15	-1 648	-681	5	638	1 401
<i>C. Ambtenaren</i>						
Netto inkomen geobserveerd*	26 555	18 373	21 638	25 713	30 155	36 008
Netto inkomen gesimuleerd en gekalibreerd*	26 828	19 076	21 988	25 931	30 105	36 001
Vershil (op individueel niveau)	278	-1 783	-787	15	899	2 410
<i>D. Zelfstandigen</i>						
Netto inkomen geobserveerd*	18 454	5 306	10 544	16 601	24 459	33 674
Netto inkomen gesimuleerd en gekalibreerd*	18 593	5 957	10 728	16 698	24 443	33 315
Vershil (op individueel niveau)	151	-1 704	-566	2	543	1 524
<i>E. Gepensioneerden **</i>						
Netto inkomen geobserveerd*	15 262	6 932	12 859	15 048	18 631	22 711
Netto inkomen gesimuleerd en gekalibreerd*	15 396	7 399	13 136	15 003	18 597	22 485
Vershil (op individueel niveau)	148	-506	-97	0	118	655
<i>F. Werklozen en invaliden **</i>						
Netto inkomen geobserveerd*	10 857	4 727	9 107	11 775	13 083	15 785
Netto inkomen gesimuleerd en gekalibreerd*	11 063	4 785	9 186	12 097	13 345	15 769
Vershil (op individueel niveau)	232	-9	0	0	0	293

Noot: * Zie tekst voor definitie netto inkomen

** Voor deze groepen is geen kalibratie uitgevoerd

Selectieve steekproef (personen met slechts één inkomensbron), totaal bruto inkomen < 120 000.

7. Conclusie

Doel van dit document was de rapportering van de kalibratie-oefening van de in MIDAS gesimuleerde belastingen aan de geobserveerde belasting (IPCAL) gegevens in de startdata. Motivering voor deze oefening was een kennelijke overschatting van de belastingen in MIDAS, omdat een aantal aftrekposten en belastingverminderingen niet kunnen worden gesimuleerd door gebrek aan relevante gegevens, zodat het netto-inkomen zou worden onderschat. De gekozen aanpak was een vergelijking op individueel niveau van de gesimuleerde en geobserveerde bedragen van de sociale bijdragen, de belastbare inkomens, de belastingen en het overblijvende netto-inkomen. Deze maakte het mogelijk om tegelijk de simulatieresultaten te valideren. Dit laatste is – achteraf gezien – wellicht belangrijker gebleken dan de kalibratie op zich. Daarnaast werd de simulatie van de bijzondere bijdrage voor sociale zekerheid toegevoegd aan de belastingmodule.

De oefening toonde namelijk aan dat het verschil kleiner was dan initieel gedacht, om twee redenen:

- Enkele correcties van de bijdragen-en-belasting module in MIDAS hadden tot gevolg dat de gesimuleerde sociale bijdragen kleiner werden.
- De beperking van de vergelijking tot de steekproef van personen voor wie een cruciale assumptie van MIDAS in feite opgaat, namelijk dat iedere persoon per jaar slechts één sociaal statuut kan hebben en slechts één type inkomen kan ontvangen.

Daardoor is de kloof die de kalibratie moet dichten in feite kleiner dan verwacht. Voor uitkeringstrekkers (gepensioneerden, werklozen en invaliden) bleek de kalibratie zelfs niet nodig. Wat betreft de ambtenaren kan men van mening verschillen of de resultaten verbeteren door de kalibratie. Voor werknemers en zelfstandigen brengt de kalibratie het mediane netto-inkomen duidelijk dichter bij de observatie, zowel algemeen als per bruto-inkomensniveau. De spreiding van de afwijking tussen simulatie en observatie wordt echter nauwelijks beperkt.

De conclusie dat de simulatie van bijdragen en belastingen in feite beter is dan vermoed, en dat de kalibratie “minder werk moet verrichten” dan verwacht, is uiteraard goed nieuws. Kalibratie is immers een erg ruwe techniek, en één die zich niet leent voor simulatie van verandering van beleid. Dat laatste is niet het doel van MIDAS, ook al vanwege de beperkingen in de simulatie van sociale bijdragen en belastingen die hierboven zijn vermeld. Het doel is de projectie van armoederisico's en inkomensongelijkheden op lange termijn, waarvoor het model wel geschikt is.