

Schattingen onrechtmatig in Nederland verblijvende vreemdelingen 2017 – 2018

Peter G.M. van der Heijden
Maarten J.L.F. Cruyff
Godfried B.M. Engbersen
Ger H.C. van Gils
Utrecht, oktober 2020

© 2020; Wetenschappelijk Onderzoek- en Documentatiecentrum. Auteursrechten voorbehouden. Niets uit dit rapport mag worden verveelvoudigd en/of openbaar gemaakt door middel van druk, fotokopie, microfilm, digitale verwerking of anderszins, zonder voorafgaande schriftelijke toestemming van het WODC.

Inhoudsopgave:

Voorwoord	
Samenvatting	i
Summary	<i>i</i>
1. Inleiding, probleemstelling, onderzoeksvragen	1
1.2. Probleemstelling en onderzoeksvragen	2
1.3. Leeswijzer	3
2. Schattingsmethoden	4
2.1. Het principe van de vangst-hervangstmethode met één bestand	4
2.2. Assumpties van het Poissonmodel	6
2.3. Het gehanteerde schattingsmodel	7
2.4. Vangst-hervangstschattingen met meer dan één bestand ¹	8
3. Data	12
3.1. Poisson schatting met PSH-V	12
3.2. Databestanden Dual Systems Estimation	14
4. Schattingen	17
4.1. Omvangschattingen met Poisson	17
4.2. Trendanalyse Poisson schattingen	18
4.3. Enige achtergronden voor de populatieschattingen 1997 – 2018	21
4.4. Populatieschatting DSE, 2017 – 2018	31
4.5. Poisson schatting en DSE 2017 – 2018 vergeleken	32
4.6. Slotbeschouwing en conclusie	38
Bijlage 1: Parameterschattingen INEE, 2017 - 2018	
Bijlage 2: Gegevens afhandeling vreemdelingenzaken	
Bijlage 3: Probabilistisch koppelen bestanden ten behoeve van schatting onrechtmatig verblijvende vreemdelingen in Nederland	
Bijlage 4: Schattingen PEW Research Center voor EU-lidstaten 2017	
Bijlage 5: De begeleidingscommissie	

¹ UNODC research based on contributions by Jan J. M. van Dijk (University of Tilburg, Netherlands), Maarten Cruyff and Peter G. M. van der Heijden (University of Utrecht, Department of Psychology, Netherlands).

Voorwoord

Met het onderhavige rapport wordt een nieuwe schatting toegevoegd aan een inmiddels lange reeks van schattingen van de populatie onrechtmatig in Nederland verblijvende vreemdelingen. Het primaire doel van deze reeks is het volgen van de ontwikkeling van de omvang van de betreffende populatie. Om het inzicht in de schattingen te vergroten, is er deze keer ook een nieuwe benadering en schattingsmethode aan het onderzoek toegevoegd. Dat betekende onder meer onderzoek en gebruik van meer bestanden dan bij vorige onderzoeken.

Het onderzoek nam een aanvang aan het einde van 2018. Het kwam daarmee onder de regels van de AVG te vallen. Het voldoen aan deze nieuwe regels heeft, mede vanwege het voor de wet complicerende feit dat meerdere bestanden met persoonsgegevens aan elkaar gekoppeld moesten worden, de nodige tijd gevraagd. Het is in belangrijke mate aan de welwillende, vaak geduldige en vasthoudende medewerking van onze contactpersonen bij de betrokken organisaties te danken dat wij erin zijn geslaagd de wettelijke hordes te nemen en het onderzoek, inclusief koppeling van verschillende bestanden, zij het met de nodige vertraging, nu af te ronden. Het gaat om medewerkers van de Internationale Organisatie voor Migratie (IOM), het Ministerie van Justitie en Veiligheid, Directoraat-Generaal Migratie, het CAK en CBS. Verder danken we natuurlijk de leden van de begeleidingscommissie voor hun kritische en ook constructieve bijdragen aan het onderzoek en rapport. Bijlage 5 vermeldt de leden van de begeleidingscommissie.

De onderzoekers,

Utrecht, 9 oktober 2020

Samenvatting

Achtergronden en probleemstelling

Het ministerie van Justitie en Veiligheid heeft ten behoeve van het vreemdelingenbeleid behoefte aan een beeld van de ontwikkeling van het aantal onrechtmatig in Nederland verblijvende vreemdelingen. Sinds 2000 worden er daarom schattingen gemaakt van het aantal vreemdelingen in Nederland zonder verblijfsrecht. Met het beschikbaar komen van de nieuwe schatting is er een reeks van schattingen verkregen voor een 20-jarige periode van 1997 tot medio 2018. De schattingen worden gemaakt met een vangst-hervangst methode. Deze groep methoden biedt een aantal duidelijke voordelen, zoals betrekkelijk lage kosten, een adequate theoretisch-statistische onderbouwing, expliciete assumpties en duidelijkheid over gevolgen van schending van die assumpties. De specifieke methode waarmee de 20-jarige reeks is gemaakt, het afgeknotte Poisson regressiemodel, is in een internationale vergelijking aangemerkt als een methode die voor het schatten van een populatieomvang voldoet aan wetenschappelijke kwaliteitseisen.

De probleemstelling voor het nieuwe onderzoek luidt als volgt:

Wat was in de periode medio 2017-medio 2018 naar schatting de omvang van de populatie van onrechtmatig in Nederland verblijvende vreemdelingen? Wat is de trendmatige ontwikkeling van de geschatte omvang van het aantal vreemdelingen zonder rechtmatig verblijf in Nederland, wanneer de nieuwe schatting met de vorige schattingen wordt vergeleken?

Welke zekerheid kan aan de schatting worden verbonden?

Het belangrijkste doel van de schattingen is het kunnen volgen van de ontwikkeling van de omvang en samenstelling van de populatie vreemdelingen zonder verblijfsrecht. De gedachte was, en blijft, dat de mate waarin de schattingsmethoden realistische veronderstellingen maken over de gebruikte data niet goed is te bepalen, maar dat we er van uit gaan dat een eventuele schending van de veronderstellingen over de jaren heen vergelijkbaar is. Hoewel de puntschattingen niet exact hoeven te kloppen, is dan wel de gesignaleerde ontwikkeling als valide te beschouwen.

In dit onderzoek is aan de reeks schattingen een nieuwe schatting voor een nieuw jaar toegevoegd. Daarnaast zijn de twijfels en de discussie over de ontwikkeling van de schattingen sinds deze vanaf het jaar medio 2005 – medio 2006 een scherpe daling heeft ingezet, aanleiding geweest voor twee aanvullingen. In de eerste plaats is voor het nieuwe jaar ook een schatting met een tweede methode gemaakt. Het betreft een schatting met een variant van vangst-hervangstmethoden die gebruik maakt van meerdere bestanden. Het doel van het maken van een tweede schatting is het verkrijgen van meer inzicht in de validiteit en het verhogen van de kwaliteit van de schattingen.

Een tweede belangrijke toevoeging van dit onderzoek is een verkenning van achtergronden die de snelle daling van de schattingen sinds 2005 kunnen verklaren, of, bescheidener, plausibel kunnen maken.

Methoden en data

De vangst-hervangstmethode is een veelgebruikte methode, eigenlijk een groep van methoden, voor het schatten van de omvang van ‘verborgen populaties’. Dit zijn populaties waarvan de leden niet gemakkelijk zijn te observeren en niet (allen) zijn geregistreerd. Standaard schattingstechnieken die gebaseerd zijn op steekproeftrekking kunnen dan niet worden gebruikt, omdat een steekproefkader ontbreekt of zeer gebrekkig is. Met één van de methoden, de Poissonregressie, is de hele reeks schattingen vanaf 1997 gemaakt. Daarbij wordt gebruik gemaakt van politiebesteden met gegevens over aan- en staandehoudingen van vreemdelingen zonder rechtmatig verblijf. Sinds 2005 is dat het bestand Politie Suite Handhaving Vreemdelingen (PSHV). De schatting is gebaseerd op het aantal mensen dat geregistreerden in het bestand in een jaar zijn aan- of staandegehouden.

De tweede methode, die in dit onderzoek voor het eerst wordt gebruikt, is een vangst-hervangstmethode voor meer dan één bestand. Ook dit is inmiddels een bekende methode om de omvang van een verborgen menselijke populatie te schatten met behulp van twee onafhankelijke registratiesystemen (of ‘registers’) die een deel van de populatie registreren. Door de gegevens van individuen in twee (of meer) registers aan elkaar te koppelen, kan het aantal individuen worden geschat dat in geen van beide (of meerdere) registers is opgenomen. Drie bestanden waren beschikbaar, te weten PSHV, een bestand van de Internationale Organisatie voor Migratie (IOM) en een bestand van het CAK. Dit laatste bleek helaas niet bruikbaar vanwege het ontbreken van voldoende identificerende variabelen om een koppeling te maken met andere bestanden. De schattingen zijn daarom gemaakt met twee bestanden, een zogenaamde Dual Systems Estimate (DSE).

Poisson Schattingen 2017 – 2018

De schatting van de totale populatie onrechtmatig in Nederland verblijvende vreemdelingen voor het jaar lopend van 1 juli 2017 tot 1 juli 2018 is 22.711. Het betrouwbaarheidsinterval geeft aan dat de werkelijke omvang met 95% kans tussen de 18.370 en 27.051 ligt (tabel i).

Ten opzichte van de schatting voor het jaar medio 2012 – medio 2013 ((35.530), is er sprake van een substantiële daling, maar het verschil is niet significant.

Tabel i. Poisson omvangschattingen INEE, IEE, Europeanen en totaal voor juli 2017 – juli 2018

Groep	Geobserveerd	Geschatte populatie	95% betrouwbaarheidsinterval
INEE	1.603	12.451	(9.501, 15.400)
IEE	498	3.707	(2.139, 5.276)
EUROPEANEN	852	6.553	(3.781, 9.324)
Totaal	2.953	22.711	(18.370, 27.051)

Ontwikkeling van de schattingen

De omvangschattingen van het totaal aantal onrechtmatig in Nederland verblijvende vreemdelingen na 2002 vertonen een daling die na 2003 en vooral na 2009 versnelt. We signaleren een aantal ontwikkelingen in beleid, toezicht, asiel- en migratiestromen die het aannemelijk maken dat het aantal

onrechtmatig in Nederland verblijvende vreemdelingen de afgelopen 10 à 15 jaar inderdaad aanzienlijk is afgenomen.

- Al vanaf ongeveer het midden van de jaren '90 werd onrechtmatig verblijf van vreemdelingen in Nederland door diverse maatregelen gaandeweg onaantrekkelijk(er) gemaakt. Voor die tijd was het voor een onrechtmatig in Nederland verblijvende vreemdeling nog betrekkelijk eenvoudig om regulier te werken en de bijbehorende voordelen te genieten. Door diverse maatregelen, zoals de Wet op de Identificatieplicht 1994 en de Koppelingswet uit 1998 die tot doel heeft vreemdelingen zonder rechtmatige verblijfstitel uit te sluiten van publieke voorzieningen, werd het voor een vreemdeling zonder verblijfstitel allengs moeilijker om een zelfstandig bestaan op te bouwen.
- De groei van de asielstromen ten gevolge van oorlogsgeweld in Midden- en Oost-Europa en delen van Azië en Afrika eveneens vanaf ongeveer midden jaren '90, maakte een eventueel effect van deze maatregelen mogelijk ongedaan en moeilijker zichtbaar. De instroom van asielzoekers kan, met enige vertraging, hebben gezorgd voor de toename van de populatie vreemdelingen zonder verblijfsrecht in de jaren rond de millenniumwisseling. Er tekende zich toen een duidelijk verschil af tussen het (groeiende) aantal asielverzoeken en het (beperkte) aantal toegekende verblijfsvergunningen. Na een afname en stabilisatie van beide aantallen werden in de jaren 2014 – 2017 juist relatief veel asielverzoeken toegekend. Dit kan een reden zijn waarom de toename van asielverzoeken in deze jaren geen toename in het geschatte aantal vreemdelingen zonder verblijfsrecht te zien geeft.
- Er zijn aanwijzingen dat de uitbreiding van de toezichts- en opsporingsbevoegdheden door de Vreemdelingenwet 2000 en de reorganisatie van de vreemdelingenpolitie in de eerste jaren van het nieuwe millennium het aantal onrechtmatig in het land verblijvende vreemdelingen heeft terug gebracht en daarnaast onrechtmatig verblijf onaantrekkelijker heeft gemaakt en aldus heeft bijgedragen aan een aanhoudende daling van het aantal onrechtmatig in Nederland verblijvende vreemdelingen.
- De veranderingen die met de nieuwe Vreemdelingenwet in gang werden gezet hebben ook geleid tot meer aandacht voor persoonsidentificatie in het vreemdelingentoezicht. In zoverre deze ontwikkelingen een betere identificatie van vreemdelingen mogelijk hebben gemaakt, zijn de schattingen voor eerdere jaren een overschatting en zijn de actuele schattingen accurater.
- Het meest duidelijk effect op de omvang van de populatie gaat uit van de uitbreidingen van de EU in 2004 en 2007. De schattingen van het aantal onrechtmatig in Nederland verblijvende Europeanen vóór deze uitbreidingen variëren van 65.000 tot 107.000, terwijl de schattingen van de laatste jaren op ongeveer 5.000 à 6.000 uitkomen.

DSE

Tabel ii laat zien dat de schatting met DSE aanmerkelijk hoger uitvalt dan de Poisson schatting. Het geschatte aantal onrechtmatig in Nederland verblijvende vreemdelingen voor het jaar medio 2017 – medio 2018 bedraagt ruim 58.000. Het betrouwbaarheidsinterval laat zien dat het werkelijke aantal volgens DSE met 95% waarschijnlijkheid ligt tussen de 47.500 en ruim 74.000.

Tabel ii: DSE aantallen observaties en omvangschattingen 2017/2018

		IOM		
		1	0	tot
PSHV	1	70	2.831	2.901
	0	1.332	53.870	55.202
	tot	1.402	56.701	58.103

Poisson en DSE vergeleken

De populaties geschat met de Poissonmethode en met DSE verschillen aanzienlijk in omvang en in samenstelling wat betreft geslacht, leeftijd en nationaliteit. De schattingen met beide methoden overlappen elkaar voor een deel, maar lijken voor een deel ook andere segmenten van de populatie beter te schatten. Op grond van schending van de assumpties van de methoden en van de mate waarin deze schendingen al of niet kunnen worden ondervangen, concluderen we dat de Poisson methode eerder een onderschatting oplevert van de ware populatieomvang en de DSE eerder een óverschating. De conclusie is daarom, dat de ware omvang van de populatie onrechtmatig in Nederland verblijvende vreemdelingen in het jaar medio 2017 – medio 2018 waarschijnlijk tussen de ca. (bijna) 23.000 en ca. 58.000 ligt.

Summary

Backgrounds and problem definition

For the purpose of immigration policy, the Ministry of Justice and Security needs a picture of the development of the number of foreign nationals residing unlawfully in the Netherlands. Since 2000, therefore, estimates have been made of the number of foreigners in the Netherlands without a right of residence. With a new estimate available, a series of estimates has been obtained for a 20-year period ranging from 1997 to mid-2018. The estimates are made with a capture-recapture method. This group of methods offers a number of clear advantages, such as relatively low costs, an adequate theoretical-statistical foundation, explicit assumptions and clarity about the consequences of violating these assumptions. The specific method with which the 20-year series was made, the truncated Poisson regression model, has been identified in an international comparison as a method that meets scientific quality requirements for estimating population sizes.

The problem definition for the new research is as follows:

What was the estimated size of the population of foreigners residing unlawfully in the Netherlands in the period mid-2017 to mid-2018? What is the trend of the estimated size of the number of foreign nationals without lawful residence in the Netherlands when the new estimate is compared with the previous estimates? What certainty can be attached to the estimate?

The main purpose of the estimates is to be able to follow the development of the size and composition of the population of foreign nationals lacking right of residence. The idea was, and remains, that although the extent to which the estimation methods make realistic assumptions about the data used is not easy to determine, we assume that a possible violation of the assumptions is comparable over the years. Although the point estimates might not be exactly correct, the development which has been observed can then be regarded as valid.

In this study an estimate with the truncated Poisson regression model for a new year has been added to the set of estimates. Further, the doubts and the discussion about the development of the estimates since they started to decline sharply from the year mid 2005 - mid 2006 have led to two additions. Firstly, an estimate for the new year was also made using a second method. This is an estimate with a variant of capture-recapture methods that uses multiple files. The purpose of making a second estimate is to gain more insight into the validity and to increase the quality of the estimates.

A second important addition to this line of research is an exploration of backgrounds that may explain, or, more modestly, make plausible the rapid decline in estimates since 2005.

Methods and data

The capture-recapture method is a widely used method, actually a group of methods, for estimating the size of 'hidden populations'. These are populations whose members are not easy to observe and are not (all) registered. Standard estimation techniques based on sampling can then not be used, because a sampling frame is missing or very deficient. With one of the methods, the Poisson regression, a series of estimates has been made since 1997. This involves the use of police files containing data on foreign nationals without lawful residence that are arrested or held up. Since 2005 this has been the Police Suite Handhaving Vreemdelingen (PSHV) file. The estimate is based on the number of times that persons registered in the file have been arrested or detained in a year.

The second method, used for the first time in this type of research, is a capture-recapture method for more than one file. This is also a well-known method to estimate the size of a hidden human population using two independent registration systems (or 'registers') that register part of the population. By linking the data of individuals in two (or more) registers, it is possible to estimate the number of individuals not included in either (or several) registers. Three files were available, namely PSHV, a file of the International Organization for Migration (IOM) and a file of the CAK. Unfortunately, the latter proved to be unusable due to the lack of sufficient identifying variables to make a good link with other files. The estimates were therefore made with two files, a so-called Dual Systems Estimate (DSE).

Poisson Estimates 2017 - 2018

The estimate of the total population of foreign nationals unlawfully residing in the Netherlands for the year running from 1 July 2017 to 1 July 2018 is 22,711. The confidence interval indicates that the actual size with a 95% probability lies between 18,370 and 27,051 (Table i).

Compared to the estimate for the year mid-2012 - mid-2013 (35.530), there is a substantial decrease, but the difference is not statistically significant.

Table i. Poisson size estimates INEE, IEE, Europeans and total for July 2017 - July 2018

Group	Observed	Estimated population	95% confidence interval
INEE	1.603	12.451	(9.501, 15.400)
IEE	498	3.707	(2.139, 5.276)
EUROPEANS	852	6.553	(3.781, 9.324)
Total	2.953	22.711	(18.370, 27.051)

Development of the estimates

The estimates of the total number of foreigners unlawfully residing in the Netherlands after 2002 show a decrease that accelerates after 2003 and especially after 2009. We observe a number of developments in policy, supervision, asylum and migration flows that make it plausible that the number

of foreign nationals residing unlawfully in the Netherlands has indeed decreased considerably over the past 10 to 15 years.

- Since approximately the middle of the 1990s, various measures have gradually made unlawful residence of foreign nationals in the Netherlands unattractive. Before that time, it was still relatively easy for a foreign national residing unlawfully in the Netherlands to work regularly and enjoy the associated benefits. As a result of various measures, such as the Compulsory Identification Act 1994 and the Linking Act of 1998, which aims to exclude foreign nationals without a legal residence permit from public facilities, it gradually became more difficult for a foreign national without a residence permit to build up an independent existence.
- The growth in asylum influx as a result of war violence in Central and Eastern Europe and parts of Asia and Africa, also from around the mid-1990s onwards, possibly made any effect of these measures undone and more difficult to observe. The influx of asylum seekers may, with some delay, have led to an increase in the population of foreigners without right of residence in the years around the turn of the millennium. A clear difference then emerged between the (growing) number of asylum applications and the (limited) number of residence permits granted. After a decrease and stabilization of both numbers, a relatively large number of asylum applications were granted in the years 2014 - 2017. This may be a reason why the increase in asylum applications in these years did not lead to an increase in the estimated number of foreign nationals without residence permits.
- There are indications that the extension of the supervisory and investigative powers by the Foreigners Act 2000 and the reorganisation of the Aliens Police in the first years of the new millennium has reduced the number of foreign nationals residing unlawfully in the country and has also made unlawful residence more unattractive, thus contributing to a continuous decrease in the number of foreigners unlawfully residing in the Netherlands.
- The changes initiated by the new Foreigners Act have also led to more attention for personal identification in the supervision of foreigners. To the extent that these developments have made better identification of foreigners possible, the estimates for previous years are overestimates and the current estimates are more accurate.
- The most obvious effect on the size of the population comes from the enlargements of the EU in 2004 and 2007. Estimates of the number of Europeans unlawfully residing in the Netherlands before these expansions vary from 65,000 to 107,000, while the estimates for the last few years are approximately 5,000 to 6,000.

DSE

Table ii shows that the estimate with DSE is significantly higher than the Poisson estimate. The estimated number of foreign nationals residing unlawfully in the Netherlands for the year mid-2017 - mid-2018 is more than 58,000. The 95 % confidence interval shows that the actual number according to DSE is likely to lie between 47,500 and over 74,000.

Table ii: DSE number of observations and size estimates 2017/2018

		IOM		
		1	0	tot
PSHV	1	70	2.831	2.901
	0	1.332	53.870	55.202
	tot	1.402	56.701	58.103

Poisson and DSE compared

The populations estimated with the Poisson method and with DSE differ considerably in size and composition in terms of gender, age and nationality. The estimates of both methods overlap to some extent, but also seem to be doing better at estimating different segments of the population. On the basis of violations of the assumptions of the methods and of the extent to which these violations can or cannot be overcome, we conclude that the Poisson method rather provides an underestimate of the true population size and the DSE rather an overestimate. The conclusion is, therefore, that the true size of the population of foreign nationals unlawfully residing in the Netherlands in the year mid-2017 - mid-2018 probably lies between approx. (almost) 23,000 and approx. 58,000.

1. Inleiding, probleemstelling, onderzoeksvragen

1.1. Achtergrond

Het ministerie van Justitie en Veiligheid heeft ten behoeve van het vreemdelingenbeleid behoefte aan een beeld van de ontwikkeling van het aantal onrechtmatig in Nederland verblijvende vreemdelingen. Daartoe worden periodiek schattingen van deze populatie gemaakt. De laatste schatting is in 2015 gemaakt voor het jaar lopende van 1 juli 2012 tot en met 30 juni 2013. Eerder zijn schattingen gemaakt voor de periode 1997 - 2003 en voor 2005. Opnieuw is er nu behoefte aan een actualisering van de cijfers.

De eerdere schattingen zijn alle gemaakt met een variant van vangst-hervangst methoden, het zogenaamde afgeknotte Poissonmodel. Met behulp van het afgeknotte Poissonmodel is het mogelijk om het aantal personen uit de populatie dat niet is geobserveerd te schatten op basis van het aantal personen uit de populatie dat minimaal één keer is geobserveerd. De methode biedt een aantal belangrijke voordelen, zoals betrekkelijk lage kosten, een adequate theoretisch-statistische onderbouwing, expliciete assumpties en duidelijkheid over gevolgen van schending van die assumpties. De methode is dan ook in een internationale vergelijking aangemerkt als een die tegemoetkomt aan wetenschappelijke kwaliteitseisen².

Niettemin is er ook een aantal problemen bij het gebruik van de methoden voor het volgen van de ontwikkeling van de aantallen onrechtmatig in Nederland verblijvende vreemdelingen. De schattingen hebben vaak betrekkelijk brede betrouwbaarheidsintervallen wat afbreuk doet aan hun waarde als beleidsinformatie. Bovendien geeft de ontwikkeling van de geschatte aantallen een trendbreuk te zien, met name sinds het jaar 2009, die op basis van de kwantitatieve informatie alleen moeilijk plausibel is te maken. Verder is ten aanzien van de onderzoeken waarin de schattingen zijn gemaakt opgemerkt dat er een gebrek is aan (kwalitatieve) contextualisering bij de huidige methode, hetgeen interpretatie van de schattingen bemoeilijkt.

Naar aanleiding van de problemen met de schattingen in de meest recente rapportages heeft het WODC een brede consultatie van deskundigen op het gebied van vreemdelingenbeleid, populatieschattingen en statistiek georganiseerd. Mede op basis van de discussies tijdens deze consultatie is besloten om voor een nieuwe schatting meerdere methoden en databronnen naast elkaar te gebruiken en te proberen op deze manier een betere schatting te krijgen.

Vervolgens heeft het WODC een vooronderzoek laten uitvoeren naar methoden en databronnen waarmee aanvullend op de bestaande praktijk schattingen kunnen worden gemaakt van het aantal onrechtmatig in Nederland verblijvende vreemdelingen³. De conclusie van het onderzoek was dat er mogelijkheden bestonden om schattingen te maken met de multipliemethode en met de vangst-hervangstmethode voor meerdere bestanden. De bestanden die voor deze schattingen in

² Dita Vogel and Vesela Kovacheva, *Classification report: Quality assessment of estimates on stocks of irregular migrants*, Hamburg Institute of International Economics. Database on Irregular Migration., Working Paper No. 1/2008, <http://irregular-migration.hwwi.net/>

³ J. Snippe, R. Mennes, Vooronderzoek data en methoden illegalschatting, WODC, ministerie van Justitie en Veiligheid, Den Haag, september 2018

aanmerking komen zijn registraties van het CAK⁴, de Inspectie Sociale Zaken en Werkgelegenheid (SZW) en de Internationale Organisatie voor Migratie (IOM) en van hulporganisaties voor vreemdelingen, voor zover die toestemming geven voor het gebruik van hun data.

1.2. Probleemstelling en onderzoeksvragen

De probleemstelling voor het nieuwe onderzoek luidt als volgt:

Wat was in de periode medio 2017-medio 2018 naar schatting de omvang van de populatie van onrechtmatig in Nederland verblijvende vreemdelingen? Wat is de trendmatige ontwikkeling van de geschatte omvang van het aantal vreemdelingen zonder rechtmatig verblijf in Nederland, wanneer de nieuwe schatting met de vorige schattingen wordt vergeleken?

Welke zekerheid kan aan de schatting worden verbonden?

Een inschatting van de mate van zekerheid van de nieuwe schatting voor het jaar medio 2017 – medio 2018, wordt verkregen door triangulatie, het maken van schattingen voor dezelfde periode met andere methoden en databronnen en het volgens een systematiek vergelijken van de resulterende schattingen. Er wordt een triangulatie uitgevoerd van de nieuwe populatieschatting gemaakt met het afgeknotte Poisson regressiemodel en schattingen gemaakt met de alternatieve methoden die in het vooronderzoek naar voren kwamen.

De probleemstelling resulteert in de volgende onderzoeksvragen:

1. Wat is het geschatte aantal vreemdelingen dat in de periode medio 2017-medio 2018⁵ onrechtmatig in Nederland verblijft wanneer gebruik gemaakt wordt van het afgeknotte Poisson-regressiemodel en de data en definities waarmee ook de eerdere schattingen zijn uitgevoerd?

2. Wat is de trendmatige ontwikkeling in het geschatte aantal vreemdelingen dat onrechtmatig in Nederland verblijft?

3. Wat is het geschatte aantal vreemdelingen dat in dezelfde periode als bij vraag 1 onrechtmatig in Nederland verblijft wanneer gebruik gemaakt wordt van de multiplier methode die gebruik maakt van het databestand CAK.

4. Wat is het geschatte aantal vreemdelingen dat in dezelfde periode als bij vraag 1 onrechtmatig in Nederland verblijft wanneer gebruik gemaakt wordt van een vangst-hervangst methode die gebruik maakt van de overlap tussen databronnen van CAK, de Inspectie Sociale Zaken en Werkgelegenheid (SZW) en IOM (Internationale Organisatie voor Migratie) en hulporganisaties die toestemming verlenen voor het gebruik van hun data.

⁴ Oorspronkelijk een afkorting van Centraal Administratie Kantoor. Bij de overgang naar een publiekrechtelijke status is in 2015 deze naam geschrapt waardoor CAK een betekenisloze afkorting is geworden.

⁵ Bij de keuze van de schattingsperiode dient te worden uitgegaan van het meest recente jaar waarvoor betrouwbare data beschikbaar zijn voor de schattingen uit de vragen 1, 3 en 4. Deze schattingen bestrijken bij voorkeur dezelfde periode.

5. Hoe verhouden de schattingen van de aantallen onrechtmatig in Nederland verblijvende vreemdelingen volgens de drie verschillende methoden zich tot elkaar?

6. Aan welke schattingsmethode kan de grootste waarde worden toegekend en waarom is dat zo?

1.3. Leeswijzer

Hoofdstuk 2 geeft een beschrijving van de gebruikte schattingsmethoden. De beschrijving van de Poissonregressie komt overeen met beschrijvingen in eerdere rapporten. Om dit rapport als een zelfstandig en een lopend verhaal leesbaar te maken, nemen we dit deel hier toch weer op in de tekst in de paragrafen 2.1 – 2.3. Vervolgens volgt in paragraaf 2.4 een korte beschrijving van de methode die nieuw is in dit onderzoek, de vangst-hervangst schattingsmethode die gebruik maakt van meerdere bestanden. We gebruiken de benaming ‘Dual Systems Estimation’, afgekort tot DSE, voor schattingen met 2 bestanden. De naam Multiple Systems Estimation of MSE wordt gebruikt voor schattingen met meer dan 2 bestanden,

In hoofdstuk 3 gaan we in op de bestanden waarmee de verschillende schattingen zijn gemaakt: de Politie Suite Handhaving Vreemdelingen (PSHV), waarmee al sinds 2001 Poisson schattingen worden gemaakt in paragraaf 3.1 en het bestand van de International Organization for Migration (IOM) dat gekoppeld aan PSHV de basis vormt voor de Dual Systems Estimation in paragraaf 3.2. In deze paragraaf bespreken we ook het voor schattingen niet bruikbaar gebleken CAK-bestand. Hierbij komt ook de reden aan de orde waarom er geen multiplierschatting en geen Multiple Systems Estimation is gemaakt.

In hoofdstuk 4 behandelen we de verschillende schattingen, zowel de Poisson schatting als de DSE. De nieuwe Poisson schatting voor het jaar lopend van 1 juli 2017 tot en met 30 juni 2018 presenteren we in paragraaf 4.1. In paragraaf 4.2 bespreken we de ontwikkeling van de Poisson schattingen sinds 1997, inclusief de meest recente toevoeging voor 2017 – 2018. In paragraaf 4.3 gaan we in op een aantal belangrijke achtergronden waarmee we de ontwikkeling van de schattingen kunnen duiden. In paragraaf 4.4 presenteren we vervolgens de schatting voor het jaar 2017 – 2018 met de nieuwe DSE-methode. In paragraaf 4.5 vergelijken we de resultaten van de schattingen met de twee methoden, Poissonregressie en DSE, bespreken we de mate waarin aan de veronderstellingen van beide methoden wordt voldaan en welke consequenties dat heeft voor de hoogte van de schattingen. Het hoofdstuk sluit in paragraaf 4.5 af met een slotbeschouwing en conclusie over de ontwikkeling van de schattingen sinds 1997 en over schattingen met de twee methoden.

2. Schattingsmethoden

In dit hoofdstuk bespreken we de methoden waarmee de schattingen in dit rapport zijn gemaakt. Het gaat om twee verschillende varianten van vangst-hervangst methoden. De vangst-hervangstmethode is een veelgebruikte methode, eigenlijk een groep van methoden, voor het schatten van de omvang van ‘verborgen populaties’. Dit zijn populaties waarvan de leden niet gemakkelijk zijn te observeren en niet (allen) zijn geregistreerd. Standaard schattingstechnieken die gebaseerd zijn op steekproef-trekking kunnen dan niet worden gebruikt, omdat een steekproefkader ontbreekt of zeer gebrekkig is.

In de paragrafen 2.1 tot en met 2.3 bespreken we de vangst-hervangstmethode met één bestand. We benoemen de assumpties van de methode (2.2) en we gaan in op de toepassing van deze methode op de populatie onrechtmatig verblijvende vreemdelingen (2.3). In een nieuwe paragraaf (2.4) beschrijven we kort de vangst-hervangst methode om met meerdere bestanden schattingen te maken.

2.1. *Het principe van de vangst-hervangstmethode met één bestand*

Stel dat een register bestaat uit observaties van personen uit een populatie, zoals aanhoudingen door de politie van dronken chauffeurs of bezoeken van druggebruikers aan een afkickklinik. Als bepaalde personen niet in het register voorkomen terwijl zij wel tot de populatie behoren, dan geeft het register een incompleet beeld van de populatie. Het doel is te komen tot een schatting van het aantal personen dat niet in het register voorkomt, maar dat wel deel uitmaakt van de te schatten populatie. In dit onderzoek wordt deze methode toegepast op een bestand waarin contacten van de politie en de Koninklijke Marechaussee met (onrechtmatig verblijvende) vreemdelingen worden geregistreerd.

Het afgeknotte Poissonmodel

De vangst-hervangstmethode schat het aantal personen uit de populatie dat *nooit* is geobserveerd op basis van het aantal personen uit de populatie dat *minimaal één keer* is geobserveerd. Het basismodel hiervoor is het *afgeknotte Poissonmodel*. Het model werkt als volgt. Door voor een bepaalde periode, bijvoorbeeld een periode van één jaar, te tellen hoeveel personen één keer, hoeveel twee keer, hoeveel drie keer, enzovoort zijn geobserveerd, wordt een bepaalde frequentieverdeling verkregen. Deze zou er bijvoorbeeld als volgt uit kunnen zien:

aantal observaties	0	1	2	3	4	5	6	...
aantal personen	?	2.226	115	17	2	0	1	

In dit voorbeeld zijn er 2.226 personen één keer geobserveerd, 115 personen twee keer geobserveerd, enzovoort. Deze verdeling is *afgeknot* omdat de frequentie voor nul observaties ontbreekt; dit zijn immers de onrechtmatig verblijvende vreemdelingen die niet staande zijn gehouden en waarvan we het aantal niet weten.

We veronderstellen nu dat de geobserveerde frequenties een bepaalde theoretische kansverdeling volgen, namelijk de *Poissonverdeling*. Deze verdeling specificeert de kansen dat een individu uit een bepaalde populatie nul keer, een keer, twee keer, drie keer, enz. wordt

geobserveerd. De vorm van deze verdeling wordt bepaald door de *Poissonparameter* λ ('*lambda*'). Deze parameter is te interpreteren als de 'zichtbaarheid' van de populatie; hoe groter de Poissonparameter, des te beter 'zichtbaar' de leden van de populatie zijn. Ter illustratie zijn hieronder twee Poissonverdelingen met Poissonparameters van respectievelijk 0.5 en 1.0 getoond.

aantal observaties	0	1	2	3	4	5	6	...
kans ($\lambda = 0.5$)	.607	.303	.076	.013	.002	.000	.000	
kans ($\lambda = 1.0$)	.368	.368	.184	.061	.015	.003	.001	

We zien hier dat, bij een Poissonparameter van 0.5, de kans om niet geobserveerd te worden .607 is, de kans om één keer geobserveerd te worden .303, de kans om twee keer geobserveerd te worden .076, enzovoort. Deze kansen tellen op tot 1. Dat betekent bijvoorbeeld, dat bij een Poissonparameter van 1 de kans om niet geobserveerd te worden .368 is, en de kans om wel geobserveerd te worden dus $1 - .368 = .632$. Deze kans van .632 wordt ook verkregen door de kansen op één, twee, drie enzovoort observaties op te tellen.

Als de verdeling is afgeknot en de nulfrequentie niet kan worden geobserveerd, zoals in onze registers het geval is, dan kan hiervoor worden gecorrigeerd. Voor dezelfde twee Poissonparameters zien de kansverdelingen er dan als volgt uit:

aantal observaties	0	1	2	3	4	5	6	...
kans ($\lambda = 0.5$)	-	0.771	0.193	0.033	0.005	0.000	0.000	
kans ($\lambda = 1.0$)	-	0.582	0.291	0.097	0.024	0.005	0.002	

We zien hier de kansen om een aantal keer geobserveerd te worden *gegeven dat men is geobserveerd*. Deze verdeling waarin de kans op nul observaties buiten beschouwing wordt gelaten, noemt men een *afgeknotte Poissonverdeling*. Voor beide verdelingen tellen de kansen weer op tot weer op tot 1.

Op basis van de geobserveerde frequentieverdeling van het aantal personen dat één keer, twee keer, etc. is geobserveerd kan met behulp van het *afgeknotte Poissonmodel* de Poissonparameter van de populatie worden geschat. Deze schatting wordt verkregen door de Poissonparameter te berekenen waarvoor de theoretische kansverdeling op één observatie, twee observaties, etc. zo goed mogelijk correspondeert met de geobserveerde frequentieverdeling.

Met behulp van de geschatte Poissonparameter kan vervolgens de kans op nul observaties worden berekend. Stel dat de geschatte Poissonparameter gelijk is aan 0.5, dan is de kans om niet geobserveerd te worden 0.607 (zie tabel). Dit betekent dat van elke 10 vreemdelingen in de populatie we er grofweg 6 niet hebben gezien (en dus 4 wel hebben gezien). De totale populatieomvang wordt dan berekend door bij elke geobserveerde persoon $1 \frac{1}{2}$ ($= 6/4$) niet-geobserveerde personen op te tellen.

2.2. Assumpties van het Poissonmodel

Hieronder bespreken de drie belangrijkste assumpties van het hierboven beschreven model.

Homogeniteit

Het model gaat ervan uit dat alle populatielieden dezelfde (homogene) Poissonparameter hebben. In werkelijkheid is dat zelden het geval; sommige personen zijn vaak 'zichtbaarder' dan anderen, en hebben daarom een grotere Poissonparameter. Het hierboven beschreven model is in zo'n geval niet in staat om de geobserveerde frequenties goed te reproduceren, en geeft een onderschatting van de werkelijke populatieomvang (zie Van der Heijden et al. 2003a, 2003b)⁶. Dit probleem is (deels) op te lossen door covariaten in het model op te nemen (zoals geslacht en leeftijd), waardoor voor iedere categorie personen een afzonderlijke Poissonparameter kan worden geschat. Op deze manier wordt voor heterogeniteit die het gevolg is van deze variabelen gecorrigeerd. De vraag of de covariaten de heterogeniteit in de Poissonparameters voldoende verklaren kan worden beantwoord met de zogenaamde overdispersietoets. Een significante uitkomst van de overdispersietoets geeft aan dat er nog meer heterogeniteit is dan de heterogeniteit die het gevolg is van de variabelen die in het model zijn opgenomen. Als dit het geval is en de informatie ontbreekt om alle relevante variabelen in het model op te nemen, dan valt de populatieschatting te laag uit⁷.

Gesloten populatie

De methode veronderstelt een 'gesloten populatie', een populatie die gedurende de periode van observatie (of registratie) van, bijvoorbeeld een jaar, niet in omvang verandert. In een gesloten populatie vindt geen migratie van individuen plaats. Als dat wel het geval is, is dat een schending van het Poissonmodel. Als mensen korter aanwezig zijn dan de observatieperiode van een jaar, dan wordt het aantal "1"-en (eenmalige aanhoudingen) t.o.v. het aantal "2"-en (personen die in het jaar een tweede keer zijn aangehouden) groter. In dit geval wordt de schatting te hoog. Schendingen van de assumptie van een gesloten populatie leiden dus tot een overschatting van de populatieomvang.

Onafhankelijkheid van observaties

Volgens het model mag de Poissonparameter van een persoon niet veranderen als gevolg van het wel (of juist niet) geobserveerd worden. Schendingen van deze assumptie treden bijvoorbeeld op wanneer door een observatie de kans op een volgende observatie toeneemt (herkenning door dezelfde politieagent) of afneemt (vanwege bewuste gedragsverandering). Het toenemen van observatiekansen ten gevolge van eerdere observaties leidt tot een onderschatting van de populatieomvang en het afnemen van observatiekansen tot een overschatting. De Poissonparameter mag overigens wel door de tijd heen veranderen als die veranderingen niet het gevolg zijn van het al dan niet plaatsvinden van observatie.

⁶ Van der Heijden, P.G.M., Bustami, R., Cruyff, M.J.L.F., Engbersen, G., & van Houwelingen, H.C. (2003a). Point and interval estimation of the truncated Poisson regression model. *Statistical Modelling*, 3, 305-322.; Van der Heijden, P.G.M., Cruyff, M.J.L.F., & van Houwelingen, H.C. (2003b). Estimating the size of a criminal population from police registrations using the truncated Poisson regression model. *Statistica Neerlandica*, 57, 289-304.

⁷ Van der Heijden, P.G.M., Bustami, R., Cruyff, M.J.L.F., Engbersen, G.B.M. & van Houwelingen, H. (2003). Point and interval estimation of the truncated Poisson regression model. *Statistical Modelling*, 3, (pp. 305-322) (18 p.).

2.3. *Het gehanteerde schattingsmodel*

Om zo goed mogelijk te anticiperen op mogelijke schendingen van de hierboven genoemde assumpties is een regressiemodel gespecificeerd, en is een stapsgewijze schattingsprocedure toegepast. Beide worden hieronder in meer detail besproken.

Regressiemodel

Om de verschillen in 'zichtbaarheid' van leden van de populatie, de verschillen in 'pakkansen', in een model te vangen nemen we informatie over zogenaamde 'covariaten', kenmerken van de geregistreerde personen, mee in het model. Voor de schatting zijn de covariaten geslacht, leeftijd (jonger dan 40 versus 40 of ouder), nationaliteit (7 groepen), reden van aanhouding (verdacht van illegaliteit versus andere redenen) en regio van staandehouding (de vier grote steden en de rest van Nederland) opgenomen. In dit regressiemodel is de relatie tussen de logaritme van de Poissonparameter (λ) en de covariaten lineair:

$$\log \lambda = b_0 + b_1 \text{Geslacht} + b_2 \text{Leeftijd} + b_3 \text{Nationaliteit} + b_4 \text{Regio} + b_5 \text{Reden} \quad (1)$$

Het model zoals hier gepresenteerd is een vereenvoudiging van het werkelijke model, omdat voor de covariaten met meerdere categorieën (voor Nationaliteit en Regio) ook meerdere parameters worden geschat (zie bijlage 1). Dit model maakt het ook mogelijk om de omvang van een willekeurige subpopulatie te schatten (bijvoorbeeld mannen jonger dan 40 die zijn staande- en/of aangehouden in de regio Amsterdam).

In dit model kan maar één waarde voor de covariaat per persoon worden ingevoerd. De covariaten Regio en Reden per aan- of staandehouding kunnen in werkelijkheid echter meerdere waarden per persoon hebben (iemand kan bijvoorbeeld de 1^{ste} keer in Amsterdam en de 2^{de} keer in Den Haag zijn staandegehouden). In zulke gevallen is steeds de waarde op het moment van de 1^{ste} aan- of staandehouding genomen.

Stapsgewijze schattingsprocedure

Onrechtmatig in het land verblijvende vreemdelingen zijn migranten en vormen daarom geen gesloten populatie. Om zo goed mogelijk aan de assumptie van een gesloten populatie te voldoen, is de schatting in drie stappen uitgevoerd. Voor het schatten van de parameters van het regressiemodel (formule 1) zijn alleen onrechtmatig verblijvende vreemdelingen gebruikt die met een grote waarschijnlijkheid (i) een vaste verblijfplaats in Nederland hebben en (ii) gedurende de gehele observatieperiode in Nederland aanwezig zijn geweest. Daarom is een onderscheid gemaakt tussen:

- EUR: uit in Europa gelegen landen afkomstige onrechtmatig verblijvende vreemdelingen
- IEE: effectief uitgezette onrechtmatig verblijvende vreemdelingen
- INEE: niet effectief uitgezette onrechtmatig verblijvende vreemdelingen

Voor de groep EUR geldt dat de kans op een vaste verblijfplaats in Nederland minder groot is dan bij de andere twee groepen. Vanwege de geografische nabijheid en de betrekkelijk lage kosten van migratie is het risico groter dat het om pendelaars of éénmalige bezoekers gaat. Voor de IEE geldt dat

hun verblijf in Nederland wegens de uitzetting niet de gehele observatieperiode beslaat. Omdat de INEE de grootste kans hebben om gedurende de hele observatieperiode in Nederland aanwezig te zijn geweest, zijn de parameterschattingen van model (1) dus alleen op deze groep gebaseerd.

We gaan als volgt te werk:

Stap 1. De parameters van model (1) worden geschat op basis van de INEE, en op basis daarvan wordt een schatting gemaakt van het totaal aantal INEE. Dit betreft de niet effectief uitgezette vreemdelingen.

Stap 2. De parameterschattingen verkregen in stap 1 worden gebruikt om de totale populatie IEE te schatten. Deze populatie kan worden geïnterpreteerd als de effectief uitzetbare/ uitgezette vreemdelingen.

Stap 3. Voor de EUR is geen Poissonparameter geschat, omdat deze groep op de covariaten sterk kan afwijken van de INEE en IEE. De schatting voor deze groep is verkregen door de ratio niet-geobserveerd/geobserveerd voor de INEE plus IEE te berekenen, en deze ratio als ophoogfactor toe te passen op het aantal geobserveerde EUR. Deze populatieschatting kan worden geïnterpreteerd als het aantal EUR dat op een willekeurig moment in Nederland aanwezig is (de individuele leden in de groep kunnen door de tijd heen verschillen).

De schatting van de totale populatie onrechtmatig in Nederland verblijvende vreemdelingen wordt verkregen door de schattingen uit deze drie stappen bij elkaar op te tellen.

Detentie

De assumptie van een gesloten populatie wordt ook geschonden doordat - onrechtmatig verblijvende vreemdelingen na een staande- of aanhouding enige tijd in detentie verblijven en gedurende deze periode niet opnieuw aan - of staande kunnen worden gehouden voor onrechtmatig verblijf. Als de detentietijden een relatief groot deel van de observatieperiode beslaan, kan het negeren ervan leiden tot een substantiële overschatting van de populatieomvang. Voor het verdisconteren van detentietijden is echter een uitbreiding van het afgeknotte Poissonmodel noodzakelijk, aangezien het model zoals geformuleerd in (1) hiertoe niet de mogelijkheid biedt.

Een methode om rekening te houden met zaken als detentie, het zogenaamde 'Recurrent Events Model (REM)' is verder ontwikkeld en uitgetoetst, met goede resultaten. Echter bleek eveneens dat geschikte data over detentie ontbreken⁸.

2.4. Vangst-hervangstschattingen met meer dan één bestand⁹

Vangst-hervangstschattingen met meer dan één bestand zijn inmiddels een bekende methode om de omvang van een verborgen menselijke populatie te schatten met behulp van twee onafhankelijke registratiesystemen (of 'registers') die een deel van de populatie registreren. Door de gegevens van individuen in twee registers aan elkaar te koppelen, kan het aantal individuen worden geschat dat in

⁸ Maarten Cruyff, Simulatie Recurrent Events Model, 24 mei 2013

⁹ UNODC research based on contributions by Jan J. M. van Dijk (University of Tilburg, Netherlands), Maarten Cruyff and Peter G. M. van der Heijden (University of Utrecht, Department of Psychology, Netherlands).

geen van beide registers is opgenomen. Bijvoorbeeld, met twee registers A en B geeft koppeling een telling van individuen in A maar niet in B, een telling van individuen in B maar niet in A, en een telling van individuen in zowel A als B. De tellingen vormen een kruistabel (zie voorbeeld), aangeduid met $A \times B$, waarbij de variabele met het label A staat voor 'opname in register A', waarbij onderscheid wordt gemaakt tussen de categorieën 'ja' en 'nee', en hetzelfde geldt voor variabele B ten aanzien van register B. De statistische opgave is om de waarde in de cel 'nee, nee' te schatten, d.w.z. het aantal personen dat in geen van beide registers voorkomt (de onbekende, want niet geobserveerde, frequentie n_{00} in voorbeeldtabel). Een schatting van de totale populatiegrootte wordt dan verkregen door de geschatte telling van personen gemist in beide bestanden op te tellen bij de tellingen van de personen die in ten minste één van de registers worden aangetroffen (dus op te tellen bij $n_{11}+n_{10}+n_{01}$).

		Bestand B	
		1	0
Bestand A	1	n_{11}	n_{10}
	0	n_{01}	n_{00}

Een fictief voorbeeld kan de methode verder verduidelijken. Stel we willen het aantal vissen in een vijver schatten. We vangen een aantal vissen (zeg 100), markeren ze en gooien ze vervolgens terug in de vijver. Enige tijd later nemen we een nieuwe vangst van zeg, 80, vissen uit dezelfde vijver en we tellen hoeveel van de tweede vangst al gemerkt als zijnde al gevangen in de eerste vangst. Als de overlap tussen de twee vangsten nul is, of zeer klein, suggereert dit dat de niet geobserveerde, gemiste deel van populatie van vissen in de vijver veel groter is dan het totaal aantal vissen dat in de twee vangsten is waargenomen. Als de overlap aanzienlijk is, bijvoorbeeld 50, suggereert dit dat het gemiste deel van de populatie betrekkelijk klein is. Hoe groter de overlap, hoe kleiner de vispopulatie. Als van de 80 vissen in de tweede vangst, er 20 al zijn gemerkt, betekent dit dat de gemerkte vissen in de vijver een kans van één op vier hebben gehad om te worden gevangen. Ervan uitgaande dat de niet-gemerkte vissen een vergelijkbare kans hebben om te worden gevangen, kan het totale aantal vissen in de vijver worden geschat op 400, waarvan 240 niet-gemerkte. De twee afzonderlijke steekproeven uit dezelfde vijver ('populatie') kunnen we beschouwen als twee afzonderlijke bestanden of registers en de gemerkte vissen in de tweede steekproef als de overlap tussen de twee bestanden.

De vangst-vangstmethode met meer dan één bestand is met succes toegepast om de omvang van verborgen menselijke populaties te schatten. Met behulp van een dergelijke capture-recapture analyse zijn bijvoorbeeld schattingen van het aantal daklozen in Nederland gemaakt¹⁰.

Veronderstellingen van de methode

Om optimale schattingen te maken met vangst-hervangst methoden met twee (of meer) registraties moet aan een aantal aannames of eisen worden voldaan. In de praktijk blijkt het doorgaans niet mogelijk om aan alle aannames en eisen te voldoen. Wanneer de aannames worden geschonden heeft dit consequenties voor de schatting van de omvang. Hieronder behandelen we kort de belangrijkste aannames, plus de gevolgen die schending van de assumpties (kunnen) hebben.

Onafhankelijkheid registraties

De voorwaarde waaraan bij lijsten van personen het moeilijkste is te voldoen, is waarschijnlijk de voorwaarde dat de kans op inclusie in het ene register statistisch onafhankelijk is van de kans op inclusie in het andere register. In het geval van persoonsregistraties wordt meestal niet voldaan aan deze aanname van onafhankelijkheid. Het is bijvoorbeeld mogelijk dat de betreffende organisaties in voorkomende gevallen naar elkaar door verwijzen of gegevens uitwisselen. In dat geval is er sprake van positieve afhankelijkheid. Aangezien een dergelijke positieve afhankelijkheid de overlap tussen de twee lijsten vergroot, wordt het aantal niet-geregistreerden in de populatie onderschat. Ook een negatieve afhankelijkheid is mogelijk. In dat geval vermindert opname in het ene register de kans op registratie in het andere register, wat leidt tot een overschatting. Dit kan gebeuren als de instantie die het ene register beheert, het minder nodig vindt een persoon te registreren indien de persoon al is geregistreerd door de instantie die het andere register beheert.

Een veelbelovende aanpak om de voorwaarde van onafhankelijkheid te versoepelen is het opnemen van een derde register, of meerdere extra registers, en het analyseren van de drieweg-, of meerwags-kruistabellen. Als er in het onderzoek gebruik gemaakt kan worden van drie of meer bestanden, kan de aanname van onafhankelijkheid van registraties worden versoepeld. Bij gebruik van drie bestanden mogen wel paarsgewijze interacties voorkomen, zolang er maar geen drieweg-interactie tussen de registraties is. Bij 4 bestanden zijn drievoudige interacties acceptabel, maar viervoudige niet, etc.

Homogene kansen op registratie

Een andere belangrijke veronderstelling betreft homogeniteit van insluitkansen: de kans om voor te komen in een register moet voor iedereen gelijk zijn. Over het algemeen is dit geen realistische aanname. Daarom is het van belang covariaten in het model op te nemen. Het gaat om variabelen die samenhang vertonen met registratie in bestand I en/of bestand II, zoals leeftijd en geslacht. Door bijvoorbeeld een interactie-effect tussen registratie in bestand I en geslacht in het model op te nemen, wordt het mogelijk dat voor vrouwen een andere kans op registratie in bestand I wordt geschat dan voor mannen. Dit wordt ook wel aangeduid met de term geobserveerde heterogeniteit. Met covariaten kunnen meerdere interactie-effecten in het model worden opgenomen; het doel van de analyse

¹⁰ Coumans, A.M., Cruyff, M., Van der Heijden, P.G.M. *et al.* Estimating Homelessness in the Netherlands Using a Capture-Recapture Approach. *Soc Indic Res* **130**, 189–212 (2017). <https://doi.org/10.1007/s11205-015-1171-7>

is dan het meest spaarzame model (een model met zo min mogelijk interactie-effecten) te vinden dat goed bij de geobserveerde data past.

Als er bij een schatting met twee bestanden voor beide bestanden niet geobserveerde heterogeniteit bestaat van de kans om in het bestand te komen, waarbij de samenhang bij beide bestanden hetzelfde werkt, d.w.z. van dezelfde variabelen afhankelijk is, bijvoorbeeld voor het ene bestand is geslacht van belang en voor het andere bestand leeftijd en leeftijd en geslacht zijn gerelateerd, dan is de schatting onder de aanname van onafhankelijkheid te laag¹¹.

Anders dan voor de Poisson regressie is er voor de DSE geen test beschikbaar om te toetsen voor (resterende, net beschreven heterogeniteit)

Perfecte koppelingen

Tot slot geldt de veronderstelling dat een perfecte koppeling tussen de gebruikte registers tot stand kan worden gebracht. Er moet geen sprake zijn van onterechte en gemiste koppelingen. Dat wil zeggen dat er geen records uit de twee bestanden die op één en dezelfde persoon betrekking hebben ongekoppeld mogen blijven, en andersom geen records uit de twee bestanden gekoppeld mogen worden die op twee verschillende personen betrekking hebben.

Het missen van koppelingen leidt tot een overschatting van de populatie en het ten onrechte koppelen van records tot een onderschatting. Bij een geringe overlap tussen bestanden kan een klein aantal gemiste of onterechte koppelingen een merkbare invloed hebben op de hoogte van de schattingen.

Over de koppeling van de gebruikte bestanden is meer te lezen in paragraaf 3.2 en bijlage 3.

¹¹ International Working Group on Disease Monitoring and Forecasting (1995). Capture-recapture and multiple-record systems estimation: history and theoretical development. *American Journal of Epidemiology*, 142, 1047-1058.

3. DATA

In dit hoofdstuk geven we informatie over de bestanden waarmee de verschillende schattingen van onrechtmatig in Nederland verblijvende vreemdelingen¹² zijn gemaakt: de Politie Suite Handhaving Vreemdelingen (PSHV) in paragraaf 3.1 en het bestand van de International Organization for Migration (IOM) dat gekoppeld aan PSHV de basis vormt voor de Dual Systems Estimation in paragraaf 3.2. In paragraaf 3.2 bespreken we ook het voor schattingen niet bruikbaar gebleken CAK-bestand en komt aan de orde waarom er geen multiplierschatting en geen Multiple Systems Estimation is gemaakt.

3.1. Poisson schatting met PSH-V

De gegevens over aanhoudingen van onrechtmatig in Nederland verblijvende vreemdelingen zijn verkregen uit PSH-V (Politie- Suite Handhaving Vreemdelingen). PSH-V is het landelijke bedrijfsproces-sensysteem dat de Afdeling Vreemdelingenpolitie, Identificatie en Mensenhandel (AVIM) gebruikt voor de uitoefening van haar taken in het kader van de Vreemdelingenwet 2000. In hoofdlijnen bestaan deze taken uit:

- Het uitvoeren van identiteitsonderzoeken bij vreemdelingen die een asielaanvraag indienen;
- Het toezien op de naleving van de aan vreemdelingen opgelegde meldingsplichten;
- Het staandehouden, ophouden of overnemen van andere diensten en zo nodig in vreemdelingenbewaring stellen van vreemdelingen die niet rechtmatig in Nederland verblijven;
- Het vaststellen van de identiteit en nationaliteit van deze vreemdelingen zonder rechtmatig verblijf;
- Hen ter verwijdering uit Nederland overdragen aan de Dienst Terugkeer en Vertrek (DT&V);
- Het naar aanleiding van strafrechtelijke gedragingen doen van voorstellen tot intrekking van de verblijfsvergunning en tot ongewenstverklaring.

Het systeem bevat dus alle vreemdelingen die naar aanleiding van een van bovenstaande taken contact hebben (gehad) met de AVIM. Ten behoeve van de schattingen is gebruik gemaakt van gegevens over de staandehouding of overname van de basispolitiezorg van vreemdelingen die niet rechtmatig in Nederland verblijven.

Er zijn gegevens uit PSH-V opgevraagd betreffende staandehoudingen en overnames van de basispolitiezorg uit de periode 1 juli 2017 tot en met 30 juni 2018. Hierbij is tevens informatie gevraagd over persoonskenmerken als het geslacht, de leeftijd, het land van geboorte, de nationaliteit, verblijfstatus van betrokkene, als ook de reden voor staandehouding of aanhouding, de regio waar de vreemdeling is staande- of aangehouden.

In PSH-V zijn van 1 juli 2017 tot en met 30 juni 2018 3.232 staandehoudingen en aanhoudingen van onrechtmatig in het land verblijvende vreemdelingen geregistreerd. Het betreft 2.953 vreemdelingen

¹² Net als in de rapporten over eerdere jaren spreken we van ‘onrechtmatig verblijf’ als de vreemdeling niet (meer) in het bezit is van een geldige verblijfsvergunning en bijgevolg geen toestemming heeft om in Nederland te verblijven. De betreffende vreemdeling dient Nederland op eigen initiatief te verlaten en kan worden staande gehouden en uitgezet worden. Conform deze juridische definitie hebben we onrechtmatig verblijvende vreemdelingen administratief (volgens PSHV en BVV) gedefinieerd als de vreemdelingen die ‘geen rechtmatige verblijfstitel (meer)’ hebben en ‘verwijderbaar’ zijn.

die op de datum van hun aanhouding of staandehouding geen rechtmatige verblijfstitel hadden en dus onrechtmatig in Nederland verbleven. Een aantal vreemdelingen is dus vaker dan één keer aan- of staandegehouden.

De gegevens uit PSH-V zijn gekoppeld aan gegevens uit de Basis Voorziening Vreemdelingen (BVV) om na te gaan of de vreemdeling in kwestie is uitgezet of dat anderszins een vertrek uit Nederland is geregistreerd. Op basis van deze informatie is een overzicht gemaakt van de wijze van verwijdering of vertrek (in de periode 1 juli 2017 tot en met 30 juni 2018) van vreemdelingen uit Nederland. Het betreft vreemdelingen waarvoor, na aanhouding en (eventuele) detentie, een 'bericht verwijdering' is geregistreerd. Het gemaakte onderscheid is van belang om rekening te houden met de mate van 'openheid' of 'geslotenheid' van de populatie. Het belang van deze kwestie en de wijze waarop er bij de schattingen mee rekening wordt gehouden wordt toegelicht in paragraaf 2.3, de subparagraaf over de stapsgewijze schattingsmethode, en in paragraaf 4.1 over de resultaten van de schattingen.

De volgende indeling is aangehouden (zie tabel 1):

- Vreemdelingen waarvoor een zogenaamd 'zelfstandig vertrek via de Internationale Organisatie voor Migratie', al of niet vanuit vreemdelingenbewaring) of een 'gedwongen vertrek' (o.a. uitzetting, uitzetting vanuit strafrecht en vertrek onder toezicht MTV) is geregistreerd, krijgen in de tabel de kwalificatie 'aantoonbaar vertrek'. Deze groep wordt in de gebruikte schattingsprocedure beschouwd als zijnde 'effectief uitgezet' (Illegals Effectively Expelled, IEE), tenzij zij afkomstig zijn uit een Europees land.
- Aan vreemdelingen waarvoor een 'zelfstandig vertrek zonder toezicht' is geregistreerd (aanzegging Nederland te verlaten) is in de tabel de gelijknamige kwalificatie toegekend. Deze groep wordt bij de schattingen beschouwd als onderdeel van de niet (effectief) uitgezette onrechtmatig verblijvende vreemdelingen (Illegals Not Effectively Expelled, INEE, tenzij zij afkomstig zijn uit een Europees land), een relatief 'gesloten' populatie.
- Tot slot is er een groep waarvoor (nog) geen vertrek of aanzegging daartoe geregistreerd is, de groep 'onbekend'. Ook deze zijn bij de schattingen beschouwd als zijnde niet (effectief) uitgezet (tenzij zij afkomstig zijn uit een Europees land).
- Een derde categorie in de opbouw van de schatting naast IEE en INEE is EUR voor de geregistreerde vreemdelingen uit een Europees land. Onrechtmatig verblijvende vreemdelingen uit Europese landen beschouwen we immers als lid van een zeer mobiele en 'open' populatie. Het gaat daarbij om landen behorend tot geografisch Europa. Van belang is immers de geografische en mogelijk, culturele, afstand en niet de juridische status.

In voorkomende gevallen wordt in de BVV een bericht verwijdering en een bijbehorend vertrek niet als laatste stap in de afhandeling van een zaak geregistreerd. In die gevallen wordt (ten onrechte) een opheffing van de inbewaringstelling ten behoeve van uitzetting als laatste geregistreerd. Om deze vertekening te ondervangen hebben we steeds wanneer er voor een geregistreerde vreemdeling een bericht verwijdering is aangetekend, dit bericht geselecteerd en beschouwd als een afsluiting van de zaak. Dit leidt tot de volgende resultaten (voor een toelichting op de verwerking van de gegevens over verwijdering en vertrek zie bijlage 2).

Tabel 1. Inhoud 'Bericht Verwijdering', volgens BVV voor onrechtmatig verblijvende vreemdelingen geregistreerd in PSHV juli 2017 – juli 2018 (aantallen zaken)

	n	%
Zelfstandig vertrek zonder toezicht	334	10%
Aantoonbaar vertrek	986	31%
Onbekend	1.903	59%
Totaal	3.223	100%

3.2. Databestanden Dual Systems Estimation

Het vooronderzoek noemt een aantal databronnen dat (in principe) in aanmerking komt voor toepassing van de vangst-hervangst methode met meerdere bestanden. Het gaat om (naast het politieb Bestand) het bestand met onverzekerbare vreemdelingen van het CAK, een bestand van de Inspectie SZW met illegaal tewerkgestelde vreemdelingen en bestanden van een aantal hulporganisaties voor onrechtmatig verblijvende vreemdelingen en (ex-)asielzoekers.

Het bestand van de Inspectie SZW is niet geschikt om op te nemen in een schatting met meerdere bestanden samen met het politieb Bestand. Het biedt geen uitbreiding of aanvulling voor de politieregistratie. Illegaal tewerkgestelden worden ook door de politie geregistreerd. De politie en Inspectie doen gezamenlijk de controles op illegale arbeid. De registraties vallen hierdoor (vrijwel) samen en voldoen niet aan de eis van onafhankelijkheid.

Gezien de verwachte complicaties en weerstanden bij het verkrijgen van toestemming voor het gebruik van gegevens, is al vroeg in het onderzoek besloten om alleen gegevens op te vragen bij de 2 landelijk opererende organisaties die bovendien op betrekkelijk onafhankelijke werkterreinen actief zijn, te weten CAK en IOM-Nederland. Beide organisaties bleken bereid aan het onderzoek mee te werken en gegevens te leveren.

CAK-bestand

Zoals beschreven in bijlage 3 zijn de CAK gegevens niet geschikt voor het maken van een schatting met een multipliemethode of met DSE.

Het CAK beschikt over een databestand met declaraties van zorgverleners die medisch noodzakelijke zorg hebben verleend aan onverzekerbare vreemdelingen. Medisch noodzakelijke zorg is alle zorg die een arts of verloskundige noodzakelijk acht. Onverzekerbare vreemdelingen zijn vreemdelingen die onrechtmatig in Nederland verblijven, maar ook vreemdelingen voor wie een aanvraag voor niet asiel gerelateerd verblijf (meestal gezinshereniging) bij de IND nog in behandeling is.

Het databestand van het CAK bevat initialen, geslacht, geboortejaar en nationaliteit van de onverzekerbare vreemdelingen. Het bestand kent een aantal beperkingen waardoor (probabilistisch) koppelen aan een ander bestand en het maken van een multiplier schatting met de gegevens niet mogelijk is. De gegevens in het bestand zijn te beperkt en te weinig onderscheidend om eenduidig unieke personen te identificeren. Dat betekent ook dat er te weinig informatie beschikbaar is om vreemdelingen met rechtmatig en met niet rechtmatig verblijf in het bestand te onderscheiden.

Bestand IOM

IOM (international Organization for Migration) is het migratieagentschap van de Verenigde Naties. Sinds 1991 heeft IOM een kantoor in Nederland. IOM ondersteunt migranten die vrijwillig terugkeren naar het land van herkomst of doormigreren naar een derde land waar permanent verblijf is gewaarborgd (hervestiging). De ondersteuning kan bestaan uit voorlichting en advies, het verzorgen van reisdocumenten, bekostiging van de terugreis en steun bij herintegratie in het land van herkomst. Het gaat om vreemdelingen met verschillende achtergronden: uitgeprocedeerde asielzoekers, vreemdelingen waarvan de verblijfstitel is verlopen of vreemdelingen die zonder verblijfstitel en zonder contact met de autoriteiten het land zijn binnengekomen. De volgende categorieën worden dienovereenkomstig onderscheiden: ex-asielzoeker, vergunning verlopen, ongedocumenteerd. IOM zoekt actief contact met vreemdelingen die mogelijk een wens tot terugkeer hebben. Dat gebeurt via een aantal kantoren verspreid in het land en via het netwerk van instellingen waarmee vreemdelingen in Nederland contact hebben. Er zijn verspreid over het land, 5 IOM-kantoren, ca. 50 spreekuurlocaties, bijvoorbeeld bij COA-locaties, vreemdelingendetentiecentra, etc. en er zijn 20 ambulante consulenten.

De ondersteuning van vrijwillige terugkeer is niet beschikbaar voor burgers uit EU-lidstaten of staten uit de Europese economische ruimte (EER)¹³, de Europese 'ministaten of landen uit de top-35 voor de hoogte van inkomens. Ingezetenen van EU-landen komen alleen voor repatriëring via IOM in aanmerking indien ze zich in een zeer kwetsbare situatie bevinden zoals bijvoorbeeld (mogelijke) slachtoffers van mensenhandel.

Van de eerste, informele contacten met kandidaten voor terugkeer, waarbij uitleg over IOM en de mogelijkheden voor ondersteuning worden gegeven, worden geen persoonsgegevens geregistreerd; alleen de totale aantallen zijn bekend. Wanneer een aanvraag tot terugkeer wordt ingediend, worden persoonsgegevens genoteerd en wordt een vreemdelingennummer aangemaakt door de IND, of, indien al beschikbaar, opgevraagd bij de IND.

Als een aanvraag is ingediend en wordt gehonoreerd, wil dat nog niet zeggen dat er daadwerkelijk wordt vertrokken. Het kan zijn dat vertrek niet mogelijk is, (bijvoorbeeld vanwege het ontbreken van geldige reisdocumenten en de mogelijkheid deze alsnog aan te vragen), of betrokkene ziet alsnog af van vertrek. Een deel van de aanvragers verdwijnt daarmee uit beeld. Daarvan is niet duidelijk of ze op een andere manier het land hebben verlaten of dat ze nog steeds ('ongedocumenteerd' in de IOM terminologie) in Nederland verblijven.

Van degenen die een aanvraag hebben ingediend zijn o.m. de volgende gegevens bekend:

- Namen
- Geslacht
- Geboortedatum
- Vreemdelingennummer
- Status/ verblijfstitel: ex-asielzoeker, vergunning verlopen, ongedocumenteerd
- Nationaliteit

¹³ European Economic Area (EEA)

Koppeling IOM- en PSHV-bestand

Het IOM bestand is 'probabilistisch gekoppeld' aan het PSHV-bestand. Om een vangst-hervangst schatting te doen met meerdere bestanden is het nodig om de records in de verschillende bestanden aan elkaar te koppelen. Idealiter gebeurt deze koppeling middels een unieke identificerende code, maar als dit niet mogelijk is kan dit ook op basis van persoonskenmerken gebeuren, zoals bijvoorbeeld naam of initialen, geboortedatum, adres, geslacht of een combinatie van dit soort kenmerken. De mate van overeenkomst in die kenmerken bepaalt dan de waarschijnlijkheid dat records uit verschillende bestanden op dezelfde persoon en/ of zaak betrekking hebben. Meer informatie over probabilistisch koppelen en de manier waarop het IOM bestand en PSHV zijn gekoppeld is te vinden in bijlage 3.

Het geleverde IOM bestand heeft betrekking op de jaren 2017 en 2018 en bevat 5.606 records. Als we selecteren op records die tussen 1 juli 2017 en 30 juni 2018 zijn ontstaan blijven er 2.281 records over. Daarvan zijn 134 een dubbele registratie (van hetzelfde vreemdelingsnummer). Het aantal unieke records in het IOM bestand is dus 2.147. Daarvan zitten 730 personen nog in een asielprocedure en verblijven dus nog rechtmatig in Nederland en 19 zijn ingezetenen van een land binnen de EU (waarvan 4 met asielstatus). Resteren nog 1.402 personen.

4. Schattingen

In dit hoofdstuk presenteren we de schattingen van de omvang van de populatie onrechtmatig in Nederland verblijvende vreemdelingen in de periode 1 juli 2017 tot en met 30 juni 2018. In paragraaf 4.1 komen eerst de schattingen met de Poissonregressie aan de orde. In paragraaf 4.2 bespreken we de ontwikkeling van de serie schattingen van de populatie onrechtmatig in Nederland verblijvende vreemdelingen sinds 1997. In paragraaf 4.3 passeren ontwikkelingen de revue die mogelijk hebben bijgedragen aan de geschatte ontwikkeling van de populatie. In paragraaf 4.4 presenteren we de schattingen met de nieuwe methode DSE. Paragraaf 5 vergelijkt de uitkomsten van de schattingen met de twee methoden, bespreekt de mate waarin aan de veronderstellingen van beide methoden wordt voldaan en welke consequenties dat heeft voor de hoogte van de schattingen. In paragraaf 6 geven we een slotbeschouwing en conclusie over de ontwikkeling van de schattingen sinds 1997 en over schattingen met de twee methoden.

4.1. Omvangschattingen met Poisson

De Poisson omvangschattingen voor de periode 2017 - 2018 zijn, net als in de eerdere onderzoeken sinds 2005, uitgevoerd volgens de stapsgewijze procedure zoals besproken in paragraaf 2.3.

Schattingen 2017 – 2018

Tabel 2 toont de frequentieverdeling van het aantal staande- en aanhoudingen voor de groep INEE in PSH-V geregistreerd voor het jaar lopend van 1 juli 2017 tot en met 30 juni 2018 (voor een overzicht van de parameterschattingen verwijzen we naar bijlage 1).

Tabel 2. Geobserveerde en geschatte aantallen personen voor de INEE in juli 2017 – juli 2018

Aantal staandehoudingen	Geobserveerd	Geschat
0	-	10.848
1	1.485	1.473
2	101	122
3	14	8
4	2	0,5
5	1	0,02
Totaal	1.603	12.451

Het model fit redelijk goed en de overdispersietest is significant ($p = 0.0$). Dat wil zeggen dat er sprake is van heterogeniteit van pakkansen die niet geheel in de covariaten wordt gevangen. Op het belang van deze vaststelling komen we in paragraaf 4.5 over de vergelijking van de Poisson schatting en de DSE terug.

Het aantal niet-geobserveerde INEE (f_0) is geschat op 10.848, wat opgeteld bij de 1.603 geobserveerde INEE de totale populatieschatting geeft van 12.451. Tabel 3 geeft de geobserveerde aantallen en omvangschattingen, inclusief betrouwbaarheidsinterval, voor de IEE en EUR. In vergelijking met 2009 zijn er meer vreemdelingen behorend tot de categorie 'EUR' (Europeanen; zie tabel 4).

Tabel 3. Poisson omvangschattingen INEE, IEE, Europeanen en totaal voor juli 2017 – juli 2018

Groep	Geobserveerd	Geschatte populatie	95% betrouwbaarheidsinterval
INEE	1.603	12.451	(9.501, 15.400)
IEE	498	3.707	(2.139, 5.276)
EUR	852	6.553	(3.781, 9.324)
Totaal	2.953	22.711	(18.370, 27.051)

De omvangschatting van de IEE bedraagt 3.707. Voor de INEE en IEE tezamen is de verhouding geschat/geobserveerd bij benadering 7,7. Op basis van deze verhouding is de populatie EUR op 6.553 geschat. De schatting van de totale populatie onrechtmatig in Nederland verblijvende vreemdelingen komt hiermee op 22.711. Het betrouwbaarheidsinterval geeft aan dat de werkelijke omvang met 95% kans tussen de 18.370 en 27.051 ligt. Ten opzichte van de schatting voor het jaar medio 2012 – medio 2013 is er sprake van een substantiële daling, maar het verschil is niet significant¹⁴.

4.2. Trendanalyse Poisson schattingen

In deze paragraaf vergelijken we de Poisson omvangschatting voor het jaar van 1 juli 2017 tot en met 30 juni 2018 met dezelfde schattingen van de voorgaande jaren. Tabel 4 geeft de geobserveerde en geschatte aantallen voor de jaren 1997-2003, 2005 - 2006, 2009, 2012 – 2013 en 2017 - 2018. De gegevens voor de jaren 1997-2003 zijn afkomstig uit Leerkes et al (2004), en die van de overige jaren uit rapporten van Van der Heijden et al¹⁵.

¹⁴ Het verschil met 95% BTI is: $N_{diff} = 12819$, $min_{95} = -554$, $max_{95} = 26329$; het BTI bevat '0', dus het verschil is NIET significant.

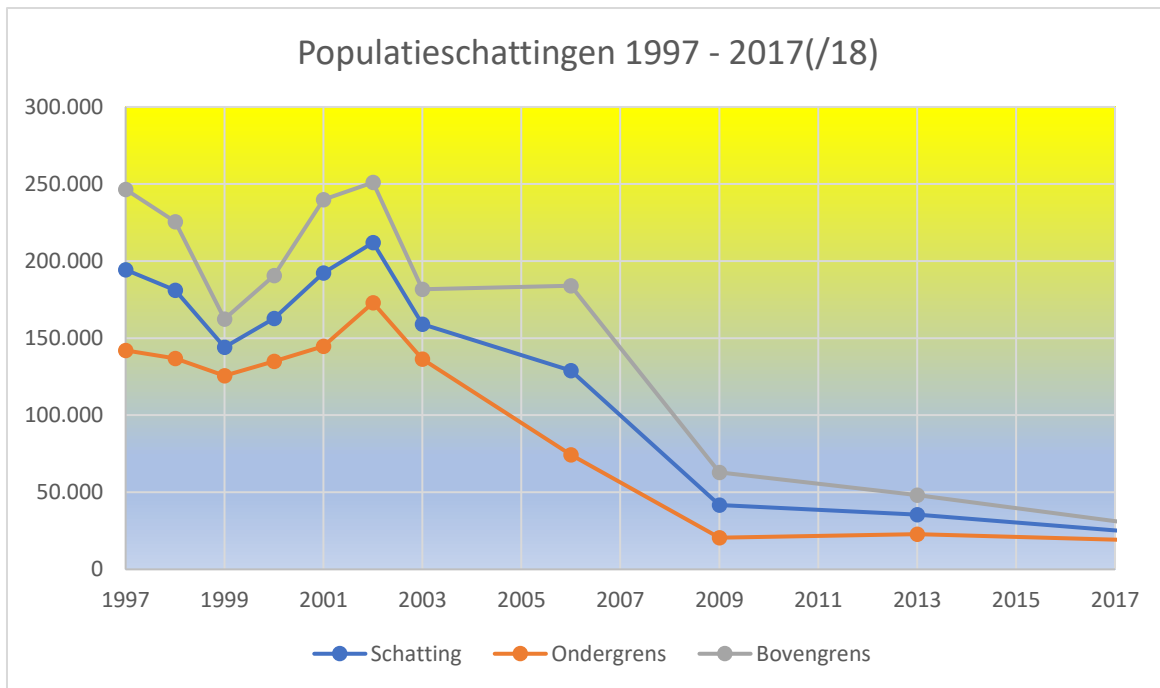
¹⁵ Leerkes, A. M. van San, G. Engbersen, M. Cruyff en P.G.M. van der Heijden (2004). *Wijken voor illegalen: Over ruimtelijke spreiding, huisvesting en leefbaarheid*. Sdu Uitgevers, 2004. O.a. Van der Heijden, P.G.M., M.J.L.F. Cruyff en G.H.C. van Gils (2015). *Schattingen illegaal in Nederland verblijvende vreemdelingen, 2012-2013*. In opdracht van het Ministerie van Justitie. Utrecht, Universiteit Utrecht, IOPS

Tabel 4. Aantallen geobserveerde en geschatte onrechtmatig verblijvende vreemdelingen in Nederland, 1997 – 2017/18(Poisson)

	Geobserveerd		Geschat	
	IEE+INEE	Europa	IEE+INEE+Europa	95% betrouwbaarheidsinterval
1997	8.618	5.523	194.304	(142.113, 246.495)
1998	7.708	5.189	181.198	(136.943, 225.453)
1999	7.335	5.381	144.081	(125.657, 162.505)
2000	7.827	5.686	162.788	(134.911, 190.665)
2001	8.287	6.018	192.373	(144.788, 239.958)
2002	9.345	8.513	211.990	(172.965, 251.015)
2003	7.747	6.520	159.077	(136.499, 181.655)
2005/06	5.795	2.694	128.907	(74.320, 183.912)
2009	2.696	213	41.835	(20.654, 63.015)
2012/13	1.948	511	35.530	(22.881, 48.179)
2017/18	2.101	852	22.711	(18.370, 27.051)

De tabel laat een duidelijke ontwikkeling zien in het geobserveerde aantal onrechtmatig in Nederland verblijvende vreemdelingen. Het geobserveerde aantal onrechtmatig in Nederland verblijvende vreemdelingen neemt na 2002 duidelijk af. Het aantal geobserveerde Europeanen dat onrechtmatig in Nederland verbleef neemt tussen 2005 en 2009 scherp af. De omvangschattingen van het totaal aantal onrechtmatig in Nederland verblijvende vreemdelingen na 2002 vertonen een daling die na 2003 versnelt. Verder valt op dat de betrouwbaarheidsintervallen voor 2012-2013 en 2017-2018 in vergelijking met de schattingen voor 2005-2006 en 2009 weer smaller worden ten opzichte van de puntschatting.

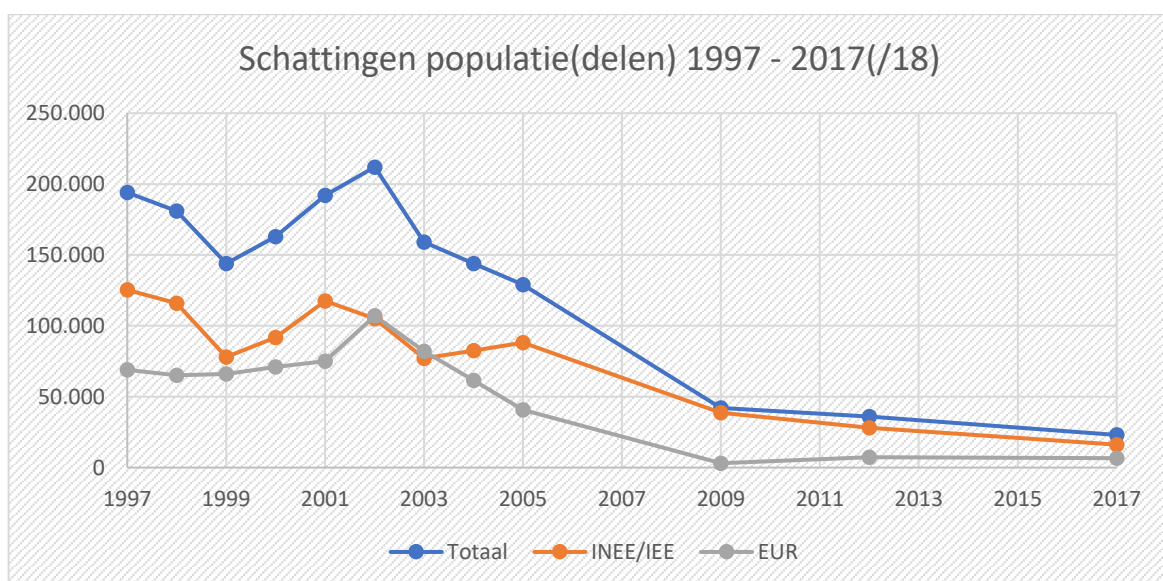
Om een beter inzicht in de ontwikkelingen te krijgen, zijn de getallen voor de totale populatie en voor beide groepen afzonderlijk in 2 grafieken weergegeven.



Grafiek 1: populatieschattingen met betrouwbaarheidsintervallen (ondergrens, bovengrens) 1997 – 2017(/18)

Grafiek 1 geeft de puntschattingen en de betrouwbaarheidsintervallen voor de gehele periode weer. De bolletjes in de grafiek geven de schattingen aan. De verbindende lijnen weerspiegelen geen reële waarden.

Grafiek 2 laat het verloop van populatieomvang voor de Europeanen en niet-Europeanen afzonderlijk zien. Daaruit wordt duidelijk, dat de daling in belangrijke mate is toe te schrijven aan de afname van het aantal onrechtmatig in Nederland verblijvende vreemdelingen afkomstig uit Europa, met name in de periode 2002 - 2009.



Grafiek 2: geschatte populatieomvang INEE, IEE en Europeanen, 1997 – 2017(/18)

4.3. Duiding van enige achtergronden voor de populatieschattingen 1997 – 2018

In de tweede helft van de jaren '90, de eerste periode waarvoor we schattingen hebben, daalt het aantal geschatte onrechtmatig verblijvende vreemdelingen van ruim 194.000 in 1997 tot 144.000 in 1999. Vanaf 2000 tot 2002 stijgen de geschatte aantallen weer: van bijna 163.000 in 2000 tot bijna 212.000 in 2002 (9% hoger dan in 1997). Na 2002 zet een daling in die tot op heden aanhoudt en na 2005 versnelt (zie tabel 5 met indexcijfers).

Tabel 5. Ontwikkeling aantallen geregistreerden en Poisson schattingen 1997 – 2017/18
(index, 1997 = 1)

	Geregistreerden	Schattingen
1997	1,00	1,00
1998	0,91	0,93
1999	0,90	0,74
2000	0,96	0,84
2001	1,01	0,99
2002	1,26	1,09
2003	1,01	0,82
2005/06	0,60	0,66
2009	0,21	0,22
2012/13	0,17	0,18
2017/18	0,21	0,12

Over de periode waarvoor schattingen nu beschikbaar zijn en vooral sinds 2000 heeft zich een aantal belangrijke veranderingen voltrokken die voor onrechtmatig verblijf in Nederland relevant (kunnen) zijn. Het zijn ontwikkelingen die het aannemelijk maken dat de omvang van de populatie onrechtmatig verblijvende vreemdelingen in Nederland vanaf kort na het begin van de nieuwe eeuw aanzienlijk is afgenomen. Er hebben zich belangrijke beleidsontwikkelingen voorgedaan die invloed hebben gehad op de aard en omvang van irreguliere migratie. Die ontwikkelingen hebben enerzijds te maken met Europees beleid en de uitbreiding van de EU, en anderzijds met een reeks van nationale beleidsmaatregelen. Die maatregelen hebben betrekking op legalisering van categorieën onrechtmatig in Nederland verblijvende vreemdelingen, op uitsluiting van de arbeidsmarkt en publieke voorzieningen, op betere opsporingsmethoden, etc.

EU-uitbreidingen

Een eerste en al genoemde factor met een duidelijke invloed op de omvang en samenstelling van de populatie onrechtmatig in Nederland verblijvende vreemdelingen zijn de achtereenvolgende

uitbreidingen van de EU¹⁶. Burgers uit Midden- en Oost-Europa hadden een belangrijk aandeel in de aanzwellende migratiestromen tijdens de jaren '90 en de eerste jaren na 2000. Het ging om zowel asielmigratie, bijvoorbeeld als gevolg van de burgeroorlogen in voormalig Joegoslavië, als om arbeidsmigratie. Een deel van deze migratie was of resulteerde in irreguliere migratie. Door de uitbreidingen van de EU werd het verblijf van aanzienlijke groepen migranten uit Oost-Europa die onrechtmatig in Nederland verbleven, gelegaliseerd. Bovendien betreft het een groep waarvoor het voorheen betrekkelijk eenvoudig was om op een toeristenvisum naar Nederland te komen en voor kortere of langere tijd zonder rechtmatige titel te blijven, maar die sindsdien als EU-ingezetene rechtmatig in Nederland kan verblijven.

De impact van de EU-uitbreidingen op de populatie onrechtmatig in Nederland verblijvende vreemdelingen is niet goed te bepalen. Maar de ontwikkeling van de herkomst van geregistreerde onrechtmatig verblijvende vreemdelingen geeft een indruk. Over de periode 1997 – 2000 maakten Europeanen (uit West- en Oost-Europa) ruim 40% uit van het aantal onrechtmatig in Nederland verblijvende vreemdelingen. Voor periode 1997 – 2003 werden schattingen gemaakt voor het aantal onrechtmatig in het land verblijvende (west- en oost-) Europeanen variërend van 65.000 tot 107.000. Voor het jaar 2005 – 2006 werd het aantal onrechtmatig in Nederland verblijvende Europeanen op 41.000 geschat. In 2009, na de belangrijkste EU-uitbreidingen van 2004 en 2007, was dat aantal teruggelopen tot ruim 4.600 en in 2017/ 18 is het weer iets opgelopen tot een aantal van ruim 6.000. De uitbreiding van de EU vanaf 2004 zou dus verantwoordelijk kunnen zijn voor een reductie van zo'n 40 à 60.000 onrechtmatig in Nederland verblijvende vreemdelingen in de jaren sinds ongeveer 2003.

Generaal-pardon

Een soortgelijk effect als van de legalisering van EU-burgers, zij het beperkter in omvang, ging uit van het Generaal-pardon uit 2007. De officiële naam van de regeling uit 2007 luidde '*Regeling afwikkeling nalatenschap oude Vreemdelingenwet*'. Op grond van de regeling werd een verblijfsvergunning verstrekt aan de vreemdeling die voor 1 april 2001 een asielaanvraag had ingediend, sindsdien Nederland niet had verlaten en waarvoor geen van de gestelde uitsluitingsgronden van toepassing was. Het pardon gold tevens voor gezinsleden en in Nederland geboren kinderen van de vreemdeling. De regeling betrof zowel asielzoekers in procedure, uitgeprocedeerde asielzoekers, als vreemdelingen die op irreguliere wijze het land waren binnengekomen¹⁷.

Ontwikkeling asielmigratie

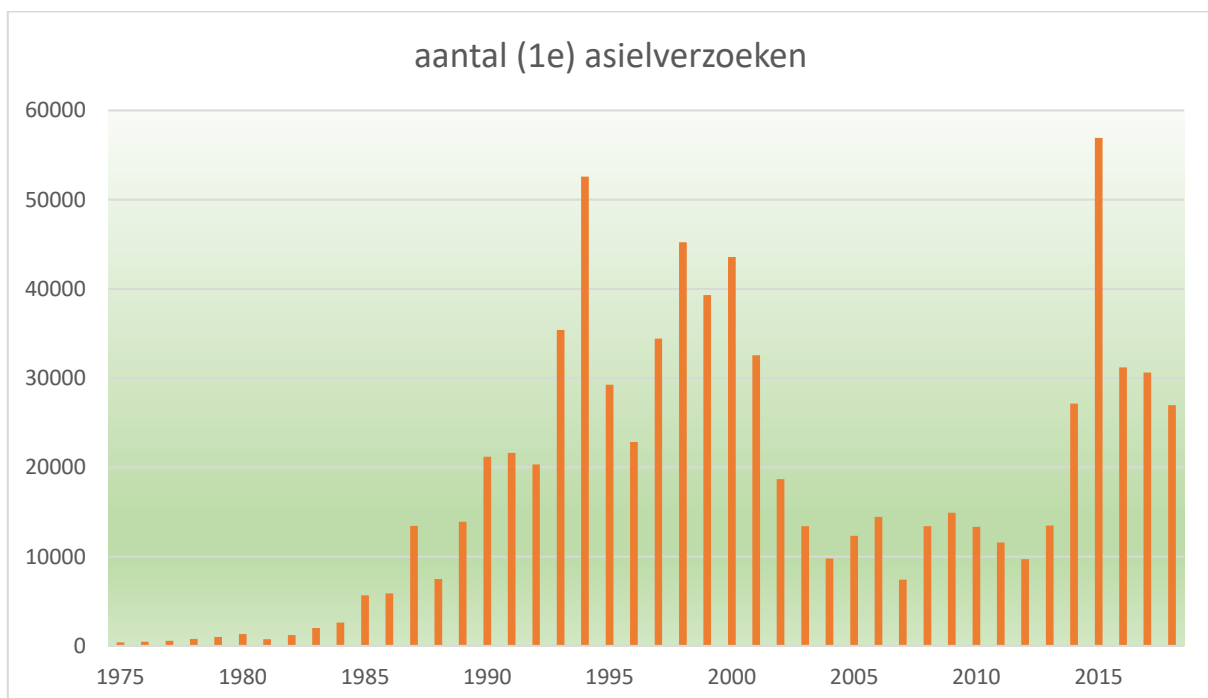
De zogenaamde asielmigratie kan leiden tot onrechtmatig verblijf van vreemdelingen van wie de asielaanvraag is afgewezen. Het precieze aandeel van voormalig asielzoekers in de onrechtmatig verblijvende populatie is onbekend. Er zijn aanwijzingen dat dit aandeel substantieel is. In de periode 1997 – 2000 namen staandegehouden onrechtmatig verblijvende vreemdelingen afkomstig uit landen waar veel asielzoekers vandaan komen tezamen ruim veertig procent van de staandehoudingen

¹⁶ 1 mei 2004: Estland, Griekscyprus, Hongarije, Letland, Litouwen, Malta, Polen, Slovenië, Slowakije en Tsjechië treden toe tot de Europese Unie; 2007: Bulgarije en Roemenië; 1 juli 2013: Kroatië.

¹⁷ De evaluatie van het Generaal-Pardon vermeldt 28.000 vreemdelingen die op grond van het pardon een verblijfsvergunning kregen, maar noemt geen aantal voor het aantal ongedocumenteerden daaronder: L.J.J. Wijkhuijs, A.M. Galloway, M.H.C. Kromhout, I.C. van der Welle, M. Smit, Pardon? Evaluatie van de Regeling afwikkeling nalatenschap oude Vreemdelingenwet, WODC, [Cahier 2011-10](#).

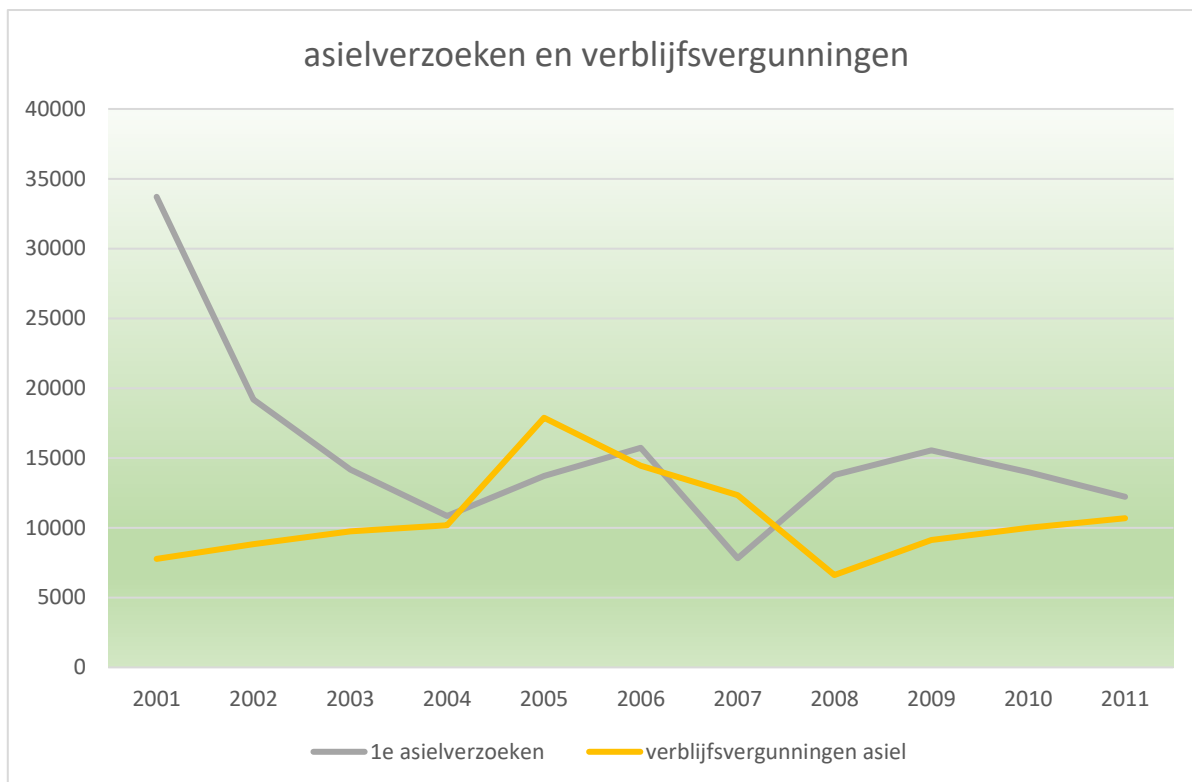
voor hun rekening. In 2009 had naar schatting 65% van de onrechtmatig in Nederland verblijvende vreemdelingen in de voorafgaande 4 jaar een asielprocedure doorlopen. Voor het jaar voorafgaand aan de staandehouding was dat percentage 29. Voor het jaar 2005 – 2006 was de schatting dat van 10% van alle onrechtmatig in Nederland verblijvende vreemdelingen de laatst geregistreerde staandehouding viel binnen het jaar na de einddatum van de laatste asielprocedure.

Na de Tweede Wereldoorlog is het asielrecht in verschillende internationale verdragen vastgelegd, waaronder het Geneefse vluchtelingenverdrag en het Europese Verdrag ter bescherming van de rechten van de mens en fundamentele vrijheden (EVRM). De aantallen bleven vele jarenlang klein en het betrof vooral mensen die persoonlijk politiek vervolgd werden. Als gevolg van oorlogsgeweld en toenemende geopolitieke spanningen in Midden- en Oost-Europa, Afrika en het Midden-Oosten werd de stroom asielaanvragen vanaf het midden van de jaren '80 massaler. Vanaf 1985 begon het aantal asielaanvragen merkbaar te stijgen en vanaf 1990 zette een versnelde groei in ten gevolge van toenemende politieke spanningen en oorlogen in de wereld, met name in voormalig Joegoslavië. In 1994 bereikte de toestroom een hoogtepunt. De jaren 1995 en 1996 zagen weer een daling waarna de periode van 1997 tot en met 2000 opnieuw groeiende aantallen liet zien. In 2001 zette een daling in waarna de aantallen asielaanvragen redelijk stabiel bleven tot aan 2014. In dat jaar nam het aantal asielzoekers opnieuw snel toe en bereikte een hoogtepunt in 2015. In 2014 kwamen vooral veel Eritreeërs naar Nederland door de aanhoudende onstabiele situatie in de Hoorn van Afrika. Ook kwamen in de jaren 2014 en 2015 veel Syrische asielzoekers naar Nederland. Na 2015 is er weer sprake van een daling, zij tot een niveau dat vooralsnog blijft boven de aantallen van het eerste decennium van de eeuw.



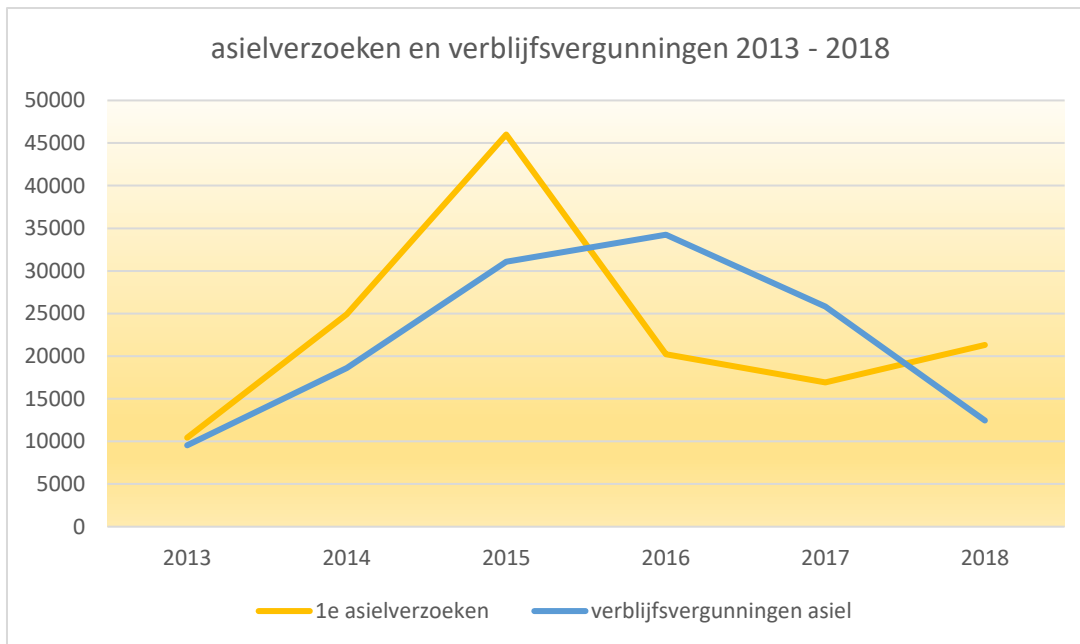
Grafiek 3: aantal (1^e) asielverzoeken, 1975 – 2018 (incl. nareizigers¹⁸). Bron: CBS

¹⁸ Gezinsleden die een asielzoeker binnen 3 maanden na het verkrijgen van een asielvergunning laat overkomen, hoeven sinds april 2014 geen asielverzoek meer in te dienen om in aanmerking te komen voor een verblijfsvergunning. Zij worden daarom niet langer als asielzoeker geteld. Om een breuk in de meerjarige asielreeksen te voorkomen, zijn ze in de grafiek meegeteld als asielzoekers.



Grafiek 4: Asielverzoeken (1^e) en verblijfsvergunningen (asiel) voor bepaalde tijd, 2001 – 2011. Bron: CBS

De daling van de populatieschattingen kan deels een gevolg zijn van de dalende aantallen asielverzoeken sinds 2000. De tijdelijke toename van de schattingen tussen globaal de jaren 2000 en 2002 kan te maken hebben met het aanzienlijk gat tussen de aantallen aanvragen en de aantallen toegekende verblijfsvergunningen (op grond van een asielverzoek) in die periode dat in grafiek 4 te zien is. De sterke toename van het aantal asielverzoeken in de jaren 2014 en 2015 is in de populatieschattingen voor 2017 – 2018 niet terug te zien. Grafiek 5 laat zien dat het aantal toegekende verblijfsvergunningen (op grond van een asielverzoek) in de jaren vrijwel gelijke tred hield met het aantal asielverzoeken en dat tussen medio 2015 en medio 2017 er zelfs sprake was van meer toegekende verblijfsvergunningen dan nieuwe asielverzoeken. Deze ontwikkelingen maken het aannemelijk dat niet grote aantallen uitgeprocedeerde of kansloze asielzoekers in die jaren zonder rechtmatige verblijfstitel in het land bleven.



Grafiek 5: Asielverzoeken (1^e) en verblijfsvergunningen (asiel) 2013 – 2018. Bron: CBS

Vreemdelingenbeleid

Als gevolg van aanscherpingen in het Nederlandse vreemdelingenbeleid (inclusief het asielbeleid) is Nederland in de afgelopen decennia in vergelijking met de periode daarvoor minder aantrekkelijk geworden voor onrechtmatig verblijvende vreemdelingen.

In de jaren zeventig en tachtig van de vorige eeuw was het voor onrechtmatig in Nederland verblijvende vreemdelingen vrij gemakkelijk om, met sofinummer¹⁹, werkzaam te zijn in bepaalde sectoren van de Nederlandse economie. Illegale arbeid werd in beperkte mate gecontroleerd en bestraft. Na een rapport over illegale tewerkstelling van de commissie Binnenlands Vreemdelingentoezicht, beter bekend onder de naam 'commissie Zeevalking' in 1991, tekende zich een omslag af. De voorstellen van de commissie vormden een belangrijke basis voor een nieuw vreemdelingenbeleid. De commissie wees op de noodzaak van 'modernisering' van het vreemdelingenbeleid door beter gebruik van moderne registratietechnieken, een efficiëntere werkwijze en samenwerking tussen verschillende opsporingsdiensten. Er werden wetten en andere maatregelen ingevoerd om onrechtmatig verblijf van vreemdelingen effectiever te bestrijden waaronder de Wet op de Identificatieplicht 1994, de Wet Voorkoming Schijnhuwelijken 1994, de koppeling van een sofinummer aan een geldige verblijfsstatus en de Koppelingswet uit 1998 die tot doel heeft vreemdelingen zonder rechtmatige verblijfstitel uit te sluiten van publieke voorzieningen. Het effect van deze wijzigingen was dat het voor vreemdelingen zonder rechtmatig verblijf moeilijker werd een zelfstandig bestaan op te bouwen²⁰.

De invoering van de Vreemdelingenwet 2000, beperkingen aan de asielprocedure en een hoger afwijzingspercentage in de procedure, lijken te hebben bijgedragen aan de daling van het Nederlandse aandeel in de asielinstroom naar Europa in de jaren na 2001. Vanaf 2003 werd geleidelijk aan ook een strikter terugkeerbeleid voor afgewezen asielzoekers ingevoerd. In 2007 werd bij het Ministerie

¹⁹ Het voormalige burgerservicenummer (BSN).

²⁰ Arjen Leerkes, Monica Varsanyi and Godfried Engbersen, Local Limits to Migration Control: Practices of Selective Migration Policing in a Restrictive National Policy Context, *Police Quarterly* 2012 15: 446 originally published online 6 August 2012, <http://pqx.sagepub.com/content/15/4/446>

van Justitie en Veiligheid de Dienst Terugkeer en Vertrek ingesteld voor de uitvoering van het terugkeerbeleid.

De nieuwe vreemdelingenwet vergrootte de mogelijkheden om onrechtmatig verblijvende vreemdelingen op te sporen. Onder de oude wetgeving konden vreemdelingen alleen worden staandegehouden wanneer er concrete aanwijzingen waren dat zij onrechtmatig in het land verbleven. Met de vreemdelingewet 2000 is een redelijk vermoeden voldoende. De meest ingrijpende wijzigingen ten aanzien van het operationeel vreemdelingentoezicht hebben betrekking op de bevoegdheid van politieambtenaren om personen staande te houden, over te brengen voor verhoor en op te houden voor onderzoek naar identiteit, nationaliteit en verblijfsrechtelijke positie (artikel 50) en de bevoegdheid om een woning te betreden zonder toestemming van de bewoner (artikel 53). De aanpassingen waren een reactie op een evaluatie van de oude vreemdelingenwet waarin was vastgesteld dat er vrijwel geen actief vreemdelingentoezicht op straat plaatsvond vanwege het als (te) strikt ervaren criterium van 'concrete aanwijzingen over onrechtmatig verblijf'.

Vreemdelingentoezicht

Na de invoering van de nieuwe vreemdelingenwet heeft ook een reorganisatie van de vreemdelingenpolitie plaats gevonden waarmee deze (grotendeels) is 'verlost' van de toelatingstaak. Er is daardoor formatie vrijgemaakt voor meer politieel toezicht met, in vergelijking met de fase vóór de reorganisatie, meer aandacht voor opsporing van migratiecriminaliteit en meer persoons- en objectgerichte controles²¹.

Meer opsporingsbevoegdheden, meer opsporing en intensiever toezicht hebben Nederland waarschijnlijk minder aantrekkelijk gemaakt voor onrechtmatig verblijf van vreemdelingen. Uit achtereenvolgende evaluaties ondernomen tot 2009 wordt niet geheel duidelijk in hoeverre en in welke richting de aangegeven ontwikkelingen de opsporingspraktijk daadwerkelijk en duurzaam hebben beïnvloed. Meerdere factoren (formatie, prioriteiten van lokale politiek en korpsleiding) zijn van invloed op het aantal staandehoudingen van vreemdelingen door de politie. In een van de evaluaties wordt geconstateerd dat de omschakeling van een administratief georiënteerde organisatie naar een executieve, operationele eenheid in 2006 nog niet was afgerond²². Volgens onderzoek van Leerkes (et. al., 2012) is het Nederlands vreemdelingentoezicht weinig gericht op actieve opsporing en bestrijding van onrechtmatig verblijf van vreemdelingen. Het beperkt zich daarentegen in hoofdzaak tot de bestrijding van overlast en criminaliteit van de zijde van onrechtmatig verblijvende vreemdelingen²³. Niettemin was er wel een kortstondig effect te bespeuren. In de periode 1998-2004 deed zich een redelijk grote toename van het totaal aantal staandehoudingen voor en vooral de toename van de staandehoudingen in 2002-2003 is opmerkelijk²⁴.

De gegevens over de periode na het jaar 2001 geven aan dat het aantal staandehoudingen op jaarbasis in vergelijking met eerdere jaren relatief fors toenam. Zo lag het aantal staandehoudingen in het

²¹ Evaluatie Vreemdelingenwet 2000, Terugkeerbeleid en Operationeel Vreemdelingentoezicht, Commissie Evaluatie Vreemdelingenwet 2000, WODC, Kiwa, BBSO, 2004; Paul Boekhoorn, Trees Speller, Vreemdelingenpolitie in transitie, Een nadere verkenning van het operationeel vreemdelingentoezicht, Nijmegen, BBSO, juli 2006.

²² BBSO, 2006; Vreemdelingentoezicht in ontwikkeling, Operationeel Vreemdelingentoezicht door de politie bij gewijzigde omstandigheden, Bureau Boekhoorn Sociaal-wetenschappelijk Onderzoek (BBSO), september 2009.

²³ Leerkes, et. al., 2012.

²⁴ Cijfers in: P.F.M. Boekhoorn, T.E.A.M. Speller, F. Kruijssen, Operationeel toezicht vreemdelingen, Evaluatie van de bevoegdheden in de Vreemdelingenwet 2000 voor het vreemdelingentoezicht door de politie, *Bureau Boekhoorn Sociaal-wetenschappelijk Onderzoek, Nijmegen, z.j., pp. 153 e.v.*

jaar 2002-2003 bijna 7800 hoger dan in het jaar 2000-2001. Na deze relatief sterke toename trad in de periode april 2003-april 2004 weer een stabilisatie op van het totaal aantal staandehoudingen van vreemdelingen.

De toename van het aantal staandehoudingen van vreemdelingen wordt toegeschreven aan het gebruik van twee instrumenten: de staandehoudingen op grond van de vreemdelingenwet en aanhoudingen op grond van het strafrecht. Een belangrijk deel van de toename van de staandehoudingen op grond van de vreemdelingenwet kan worden verklaard door een stijging van het aantal specifieke acties en van het aantal controles in het kader van de Wet Arbeid Vreemdelingen (WAV). Een verklaring voor de toename van het aantal staandehoudingen van vreemdelingen, in vervolg op de aanhouding op grond van het strafrecht, leek vooral te liggen in een toegenomen alertheid bij de politiekorpsen op de identiteit en verblijfsrechtelijke positie van de aangehouden persoon²⁵. Deze ontwikkeling is ook te herkennen in de ontwikkeling van de aantallen geregistreerden waarmee de schattingen zijn gemaakt (zie tabel 5). Deze nemen over de jaren 2000, 2001 en 2002 duidelijk toe. Vanaf 2003 zet een daling in. Het is mogelijk dat geïntensiveerd politietoezicht met de vreemdelingenwet en het strafrecht in de hand in deze jaren het aantal onrechtmatig in het land verblijvende vreemdelingen heeft teruggebracht en daarnaast onrechtmatig verblijf blijvend onaantrekkelijker heeft gemaakt en aldus heeft bijgedragen aan een aanhoudende daling van het aantal onrechtmatig in Nederland verblijvende vreemdelingen.

Identificatie en biometrie

Een belangrijk onderdeel van de reorganisatie van de vreemdelingenpolitie en later AVIM was het streven om meer te werken als een 'informatiegestuurde politie' (IGP). Identificatie kreeg meer aandacht in de organisatie en deskundigheidsbevordering. Medewerkers zijn en worden verder opgeleid tot specialisten documentherkenning. Verder worden 'identiteitsstraten' ingericht met een Progis-zuil, een apparaat dat vingerafdrukken van verdachten scant, foto's maakt en aan gezichtsherkenning doet.

Per 1 maart 2014 trad de wet biometrie vreemdelingenketen in werking. De wet bevat een wijziging van de vreemdelingenwet die voorziet in een uitbreiding van het gebruik van biometrie in de vreemdelingenketen. Met de wet is de afname en verwerking van een gezichtsopname en tien vingerafdrukken van in beginsel alle vreemdelingen in alle vreemdeling-rechtelijke procedures mogelijk gemaakt. Daarnaast kunnen deze biometrische gegevens centraal in de Basisvoorziening Vreemdelingen worden opgeslagen die door alle betrokken instanties kunnen worden geraadpleegd.

De invoering van de wet heeft geleid tot een sterke toename van het aantal geregistreerde vingerafdrukken en gezichtsopnames. Voor de inwerkingtreding van de wet werden tussen de 3.000 à 6.000 sets vingerafdrukken per kwartaal (gemiddeld: 4992) opgeslagen in de BVV. Na inwerkingtreding van de wet varieerde dit tussen 23.000 en 47.000 sets vingerafdrukken (gemiddeld: 32.997). Voor gezichtsopnames is een minder sterke percentuele toename waargenomen, namelijk gemiddeld van 129.786 naar 176.563 gezichtsopnames per kwartaal²⁶. Ook in EU verband is het gebruik van biometrische gegevens in het vreemdelingentoezicht aanmerkelijk toegenomen. Er is een 'Eurodac' informatiesysteem ingericht voor vingerafdrukken van asielzoekers en vreemdelingen die op een

²⁵ BBSO, 2004, p. 108.

²⁶ Evaluatie Wet Biometrie vreemdelingenketen, 2017

irreguliere wijze de EU binnentreden. Van elke asielzoeker binnen de EU worden nu als eerste stap in de asielpcedure vingerafdrukken gemaakt.

De identificatie bestaat, op alle plekken waar intake van vreemdelingen plaats vindt, uit:

- Pasfoto op studiokwaliteit
- Platte en gerolde vingerafdrukken
- Scan beschikbare documenten

Sinds 2014 staat er op elk politiebureau een zogenaamde ID-zuil. Deze is primair bedoeld voor strafrecht, maar wordt ook ingezet voor vreemdelingerecht. Identificerende, waaronder biometrische informatie is al 7 à 8 jaar beschikbaar voor de agent op straat via de zogenaamde MEOS-app op smartphones. Met de MEOS-app kan de politie snel en gemakkelijk via de smartphone een deel van het basispolitiewerk op straat afhandelen. Daarvoor hoeven politiemensen niet meer terug naar een politiebureau. De MEOS-app heeft verschillende functionaliteiten zoals de identificatie van personen en voertuigen, integrale bevraging en het schrijven van digitale bonnen. De verwachting is dat binnen afzienbare tijd de identiteitsvaststelling indien nodig ook op straat kan plaatsvinden.²⁷ Einde 2017 werkten zo'n 35.000 agenten met deze app²⁸.

De mate waarin biometrische informatie door de verschillende instanties in de vreemdelingenketen worden gebruikt, wordt betwist. De cijfers gepresenteerd in de evaluatie van de wet biometrie laten zien dat biometrische gegevens vooral worden gebruikt door IND en nauwelijks door de AVIM. In een reactie op de evaluatie doet de Minister van Justitie en Veiligheid de belangrijke constatering dat 'de koppeling van biometrie met een uniek V-nummer in de BVV betekent dat een vreemdeling niet meer onopgemerkt onder verschillende identiteiten in verschillende databestanden van ketenpartners kan voorkomen.'²⁹

Het correct identificeren en registreren van vreemdelingen die in het toezicht worden aangetroffen kan een sterke invloed hebben op de schattingen van de omvang van de populatie. De schattingen maken immers gebruik van het aantal malen dat een vreemdeling door de politie wordt aangetroffen (staande gehouden of aangehouden). Verbeterde identificatie, registratie en beschikbaarheid van gegevens over aangetroffen vreemdelingen leiden tot een groter aantal vreemdelingen dat wordt 'herkend' en dus meermalen wordt aangetroffen.

Naarmate de identificatie van door de politie staande- of aangehouden onrechtmatig in Nederland verblijvende vreemdelingen de afgelopen jaren is verbeterd, worden er minder vreemdelingen (ten onrechte) geregistreerd als voor de eerste keer staandegehouden en meer worden er herkend als eerder staandegehouden. Dat geeft in de tellingen waarmee de vangst-hervangstschattingen worden gemaakt als resultaat minder personen die 1 keer zijn staandegehouden en meer die meermalen zijn staandegehouden. In de mate waarin identificatie van vreemdelingen in de afgelopen jaren is

²⁷ Interviews medewerkers van de Afdeling Vreemdelingenpolitie, Identificatie en Mensenhandel (AVIM)

²⁸ Politie wint ICT-prijs, <https://www.politie.nl/nieuws/2017/november/1/00-de-politie-wint-belangrijke-ict-prijs.html>, 01-11-2017.

²⁹ Evaluatie van de Wet biometrie Vreemdelingenketen, WODC, 2017. De constatering van de minister wordt onderschreven door de geïnterviewde medewerkers van de Afdeling Vreemdelingenpolitie, Identificatie en Mensenhandel (AVIM).

verbeterd zijn oudere schattingen dus eerder een overschatting van de werkelijke omvang van de populatie en recentere schattingen een meer accurate schatting³⁰.

Samenstelling populatie en geregistreerden

Diverse waarnemers signaleren ontwikkelingen in de samenstelling van de door de politie geregistreerde 'steekproef' en de populatie onrechtmatig in Nederland verblijvende vreemdelingen. Een gemeenschappelijke observatie lijkt te zijn dat het aandeel van vreemdelingen die minder moeite doen om in het gastland een bestaan op te bouwen, toeneemt. Een aspect daarvan is dat de migrantenpopulatie volgens deze observaties gemiddeld genomen dynamischer, beter geoutilleerd en geïnformeerd en daardoor mobieler is geworden. Ze zijn minder gehecht aan één plek of land en zijn bereid om te 'shoppen', als het hier te lande te onaantrekkelijk wordt. In dat verband signaleren zeggelieden van de politie bijvoorbeeld dat het vaker voorkomt dan in het verleden, dat onrechtmatig verblijvende vreemdelingen die worden staandegehouden, op doorreis zijn.

Betrokkenen signaleren ook dat de nu aanwezige populatie voor een groter deel bestaat uit mensen die er geen been in zien om met enige regelmaat door de politie te worden opgepakt en vastgezet. Ze zijn volgens deze zeggelieden minder gericht op het opbouwen van een bestaan in Nederland en op het ontlopen van confrontaties met de politie of andere autoriteiten. Het gaat dan vaak om migranten afkomstig uit Noord-Afrika die niet uitzetbaar zijn of die hun uitzetting met succes weten aan te vechten. Eerder onderzoek heeft laten zien dat Noord-Afrikaanse jongemannen zonder vaste verblijfplaats na een eerste aanhouding voor een overtreding of strafbaar feit een duidelijk hogere kans op een nieuwe aanhouding hebben dan andere groepen. Na een aantal jaren loopt de kans dat ze opnieuw worden aangehouden op tot 80%³¹.

We hebben deze observaties niet onderzocht, maar we kunnen wel een aantal ontwikkelingen van kenmerken van de vreemdelingen signaleren die zijn aangehouden en geschat (zie tabel 6 - 8).

In de eerste plaats valt op dat de populatie lijkt te verouderen: het aandeel in de populatie van de groep 40+ stijgt van 16% in 2009 tot 24% in 2017/18. Deze verschuiving wijst mogelijk op een populatie die zich nog maar beperkt aanvult en vernieuwt.

Het aandeel in de populatie van personen afkomstig uit Noord-Afrika lijkt toe te nemen van 10% in 2005/ 2006 tot 19% in het meest recente jaar.

Er is ook sprake van een toename van het aandeel van de politiekorpsen Amsterdam en Rotterdam in het aantal geregistreerde vreemdelingen zonder verblijfsrecht. Het aandeel van Amsterdam groeit van 13% in de periode 1997 – 2000 tot 25% voor het jaar 2017/2018; voor Rotterdam zijn de percentages 19, respectievelijk 31. Deze groei gaat vooral ten koste van registraties in de rest van het land die afnemen van 55% aan het einde van de vorige eeuw tot 31% recent (tabel 7).

³⁰ Om een indruk te krijgen van het belang van deze ontwikkeling hebben we informatie over de aantallen samenvoegingen van identiteiten, o.a. op basis van biometrische gegevens, over de afgelopen jaren opgevraagd, maar nog niet ontvangen.

³¹ Arjen Leerkes, Monica Varsanyi and Godfried Engbersen, Local Limits to Migration Control: Practices of Selective Migration Policing in a Restrictive National Policy Context, *Police Quarterly* 2012 15: 446 originally published online 6 August 2012, <http://pqx.sagepub.com/content/15/4/446>

Tot slot valt op dat het aandeel van strafrechtelijke delicten en van overtredingen in het aantal staande- en aanhoudingen is toegenomen van gemiddeld 46% tussen 1997 en 2000 tot 78% voor het meest recente jaar 2017/18.

Tabel 6. Schattingen vreemdelingen zonder verblijfsrecht, 2009, 2012/13, 2017/18

		2009		2012/13		2017/18	
		<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
Geslacht	M	31.812	76	25.089	71	18.571	82
	V	10.023	24	10.441	29	4.140	18
		41.835	100%	35.530	100%	22.711	100%
Leeftijd	Tot 40 jaar	35.211	84	27.853	78	17.332	76
	Ouder dan 40	6.625	16	7.677	22	5.379	24
		41.835	100%	35.530	100%	22.711	100%
Nationaliteit	Turkije	6.786	16	2.472	7	1.053	5
	Noord-Afrika	3.993	10	3.431	10	4.278	19
	Rest Afrika	6.369	15	7.146	20	3.183	14
	Surinaams	3.179	8	708	2	464	2
	Aziatisch	6.368	15	10.045	28	4.174	18
	Amerika	753	2	2.932	8	2.673	12
	Europa	3.063	7	7.383	21	6.552	29
	Onbekend	11.324	27	1.413	4	334	1
	41.835	100%	35.530	100%	22.711	100%	

Tabel 7. Door politie geregistreerde vreemdelingen zonder verblijfsrecht, 1997 – 2000, 2009, 2012/13, 2017/18

		97 – 00	2005/06	2009		2012/13		2017/18	
		%	%	n	%	n	%	n	%
Regio	Amsterdam	13		383	9	442	11	727	25
	Haaglanden	9		286	7	300	11	252	9
	Rotterdam	19		284	7	297	22	905	31
	Utrecht	5		192	4	140	4	166	6
	Overig	55		3.185	74	1.280	52	903	31
Reden staande-houding	Strafrecht + overtredingen	46	38		39		44		78
	Vreemdelingenwet	54	62		61		56		22

Tabel 8. Schattingen kenmerken vreemdelingen zonder verblijfsrecht, 2009, 2012/13, 2017/18

		2009		2012/13		2017/18	
		n	%	n	%	n	%
Regio	Amsterdam	3.610	9%	3.979	11%	4.137	18%
	Haaglanden	3.188	8%	3.873	11%	2.460	11%
	Rotterdam	7.547	18%	7.775	22%	7.609	34%
	Utrecht	9.071	22%	1.352	4%	1.072	5%
	Overig	18.419	44%	18.551	52%	7.434	33%
		41.835	100%	35.530	100%	22.711	100%
Reden staande-houding	Strafrecht + overtredingen	17.273	41%	10.821	30%	17.377	77%
	Vreemdelingenwet	24.554	59%	24.709	70%	5.334	23%
		41.827	100%	35.530	100%	22.711	100%

4.4. Populatieschatting DSE, 2017 – 2018

In deze paragraaf bespreken we kort de nieuwe schatting van de populatie onrechtmatig in Nederland verblijvende vreemdelingen met de Dual Systems Estimation (DSE). Uit het gekoppelde bestand zijn uit het IOM-deel de personen verwijderd die nog in een asielprocedure zitten en die uit een EU-land afkomstig zijn. Ingezetenen van EU-landen komen alleen voor repatriëring via IOM in aanmerking indien ze (mogelijk) als slachtoffer van mensenhandel in Nederland zijn beland. Uit het PSHV-deel zijn ook de burgers van EU-landen verwijderd.

De schatting met DSE is gemaakt met gebruikmaking van de covariaten Geslacht, Leeftijd en Nationaliteit (ingedeeld in 7 categorieën/regio's). Uit een reeks mogelijke modellen is met de standaardmethode log-lineaire analyse het 'beste' model geselecteerd. Dat betekent dat is gezocht naar het meest spaarzame model met de beste 'fit', zijnde de beste overeenkomst, of het geringste verschil, tussen geobserveerde en geschatte frequenties. Naarmate meer (mogelijke) interacties tussen de gebruikte covariaten in het model worden opgenomen, verbetert in beginsel de fit, maar nemen ook de betrouwbaarheidsintervallen in omvang toe. De schatting wordt zuiverder, maar ook minder betrouwbaar³². Daarom wordt het model geselecteerd met de beste fit, bij een zo beperkt mogelijk aantal parameters of interacties tussen de covariaten.

Tabel 8 laat zien dat de schatting met DSE hoger uitvalt dan de Poisson schatting. Het geschatte aantal onrechtmatig in Nederland verblijvende vreemdelingen voor het jaar medio 2017 – medio 2018 bedraagt ruim 58.000. Het betrouwbaarheidsinterval laat zien dat het werkelijke aantal volgens DSE met 95% waarschijnlijkheid ligt tussen de 47.500 en ruim 74.000.

In de volgende paragraaf gaan we in het kader van een vergelijking van de schattingen met de 2 methoden nader in op DSE.

Tabel 8: DSE aantallen observaties en omvangschattingen 2017/2018

		IOM		
		1	0	tot
PSHV	1	70	2.831	2.901
	0	1.332	53.870	55.202
	tot	1.402	56.701	58.103

4.5. Poisson schatting en DSE 2017 – 2018 vergeleken

In deze paragraaf vergelijken we de schattingen met de twee methoden, Poisson en DSE, beoordelen we de mate waarin beide methoden voldoen aan de gemaakte veronderstellingen en we bespreken welke consequenties het al of niet voldoen aan die veronderstellingen heeft voor de hoogte en betrouwbaarheid van de schattingen.

Schattingen

Als we de uitkomsten van de Poisson en de DSE naast elkaar zetten, blijken deze ruim 35.000 personen (58.103 – 22.711) te verschillen. Beide schattingen vallen buiten het betrouwbaarheidsinterval

³² Naarmate een schatter gemiddeld genomen (over alle mogelijke steekproeven) de (onbekende) juiste populatiewaarde beter benadert, noemen we de schatter zuiverder.

van de andere schatting en het verschil tussen beide schattingen is dus niet aan het toeval toe te schrijven (tabel 9).

Tabel 9. Poisson schattingen en DSE 2017/2018

Groep	Geschatte populatie	95% betrouwbaarheidsinterval	
Poisson (PSHV)	22.711	18.370	27.051
DSE (PSHVxIOM)	58.103	47.524	74.306

Er zijn ook weinig overeenkomsten in de aantallen voor de achtergrondkenmerken die met beide methoden konden worden geschat (tabel 10). De populatie geschat met DSE bevat duidelijk meer vrouwen dan de populatie geschat met Poisson (35 versus 18%). De 'DSE-populatie' is gemiddeld genomen ook ouder dan de 'PSHV-populatie': 52% is ouder dan 40 jaar, versus 24% in het geval van de populatie geschat op basis van PSHV.

Ook de samenstelling van de populatie naar nationaliteiten is voor de twee schattingen nogal verschillend en vertoont weinig overeenkomsten. De aantallen geschat met de Poissonregressie voor vreemdelingen afkomstig uit Turkije en Noord-Afrika (respectievelijk 1.053 en 4.278) liggen binnen de betrouwbaarheidsintervallen van de schattingen voor de overeenkomstige groepen geschat met DSE. Een relatief groot deel van de vreemdelingen in DSE heeft een onbekende nationaliteit, maar voor zover de nationaliteit bekend is, zijn vreemdelingen afkomstig uit Azië beter vertegenwoordigd in de DSE dan in de Poisson schatting (32% vs 18%) en Noord-Afrikanen slechter (7% vs 19%). Vreemdelingen zonder rechtmatig verblijf afkomstig uit Amerika³³ zijn in de populatie geschat met politiegegevens iets sterker vertegenwoordigd dan in de populatie geschat met DSE (12% vs 8%).

³³ Noord- en Zuid-Amerika en Oceanië.

Tabel 10. Vergelijking DSE – Poisson schattingen onrechtmatig verblijvende vreemdelingen, punt-schatting, betrouwbaarheidsinterval (DSE), percentage van de populatie, 2017 - 2018

		Dual Systems Estimation (DSE)				Poisson	
		<i>n</i>	%	<i>min95</i>	<i>max95</i>		
Geslacht							
	m	37.836	65%	30.864	48.226	18.571	82%
	v	20.266	35%	16.166	26.794	4.140	18%
		58.102				22.711	
Leeftijd							
	40-	27.836	48%	22.692	35.646	17.331	76%
	40+	30.266	52%	24.440	39.948	5.379	24%
		58.102				22.710	
Nationaliteit							<i>Poisson</i>
	Turks	1.005	2%	616	1.507	1.053	5%
	Noord Afrika	4.051	7%	3.161	5.295	4.278	19%
	Overig Afrika	6.184	11%	4.752	8.206	3.183	14%
	Suriname	1.188	2%	757	1.728	464	2%
	Azië	18.496	32%	14.788	23.948	4.174	18%
	Amerika	4.396	8%	3.329	5.818	2.673	12%
	Europa	12.901	22%	10.332	16.985	6.552	29%
	Onbekend	9.882	17%	7.740	13.108	334	1%
		58.102				22.711	

Al met al zijn er weinig overeenkomsten in de aantallen die geschat zijn met de twee verschillende methoden. De schattingen met beide methoden overlappen elkaar voor een deel, maar lijken voor een deel ook andere segmenten van de populatie beter te schatten. Bij de DSE gaat het in vergelijking met de Poisson schatting om een ouder deel van de populatie met weinig politiecontacten, meer vrouwen, meer mensen afkomstig uit Azië en minder uit Noord-Afrika. De Poisson schatting bevat juist meer mensen in categorieën die ook sterk oververtegenwoordigd zijn in criminaliteit.

Veronderstellingen en (mogelijke) schending van de veronderstellingen

De schattingen zijn ook te beoordelen op de mate waarin aan de veronderstellingen van de gebruikte schattingsmethoden wordt voldaan en op wat een eventuele schending betekent voor de schattingen. Het gaat om drie belangrijke assumpties: de veronderstelling dat er geen niet-geobserveerde heterogeniteit is, de veronderstelling dat de populatie gesloten is en de veronderstelling dat opeenvolgende aanhoudingen onafhankelijk van elkaar plaats vinden. Voor DSE geldt de bijzondere veronderstelling dat de bestanden perfect te koppelen moeten zijn. Voor de reeks met Poisson schattingen is van belang dat we de ontwikkeling van de populatieschattingen over de jaren kunnen vertrouwen, zolang we mogen veronderstellen dat de mate en wijze waarop de assumpties van de methode worden geschonden niet ingrijpend verandert.

Geen niet-geobserveerde heterogeniteit

In een homogene Poissonverdeling wordt ervan uitgegaan dat alle leden van een populatie dezelfde (homogene) Poissonparameter of ‘pakkans’ hebben. In werkelijkheid is dat zelden het geval. Vaak zijn sommige personen ‘zichtbaarder’ dan anderen. Dan hebben ze een grotere Poissonparameter dan personen die minder goed zichtbaar zijn. Er is dan sprake van heterogeniteit van de Poissonparameter. Dit is problematisch, omdat, indien deze heterogeniteit wordt genegeerd, de schatting die de Poissonverdeling oplevert een onderschatting is van de werkelijk populatieomvang. Om deze reden maken we in dit onderzoek gebruik van het zogenaamde Poisson *regressie*model. Met dit model kan dit probleem deels worden opgelost doordat er covariaten in het model kunnen worden opgenomen (zoals geslacht en leeftijd), waardoor voor iedere persoon een Poissonparameter kan worden geschat die een functie is van deze covariaten. We spreken hier van geobserveerde heterogeniteit, omdat het gaat om heterogeniteit van de Poissonparameter voor zover die het resultaat is van de variabelen die we in het onderzoek hebben geobserveerd. Er kan nog meer heterogeniteit van de Poissonparameter zijn, en die duiden we aan met niet-geobserveerde heterogeniteit. Met een zogenaamde overdispersietest kan worden bepaald of de covariaten (de geobserveerde heterogeniteit) de totaal aanwezige heterogeniteit in de Poissonparameters voldoende weergeven. Een significant resultaat duidt erop dat er nog relevante covariaten ontbreken. In dat geval dient de populatieschatting dus geïnterpreteerd te worden als een *onderschatting* van de werkelijke populatiegrootte. Nu is er bij alle eerder uitgevoerde schattingen sprake van een significante overdispersietest. Zo ook bij de schatting voor 2017 – 2018. Dat betekent dat we de Poisson schatting kunnen beschouwen als een onderschatting van de werkelijke omvang van de populatie onrechtmatig in Nederland verblijvende vreemdelingen.

Voor de hele reeks van schattingen zijn steeds dezelfde kenmerken gebruikt om met de verschillen in pakkansen rekening te houden (geslacht, leeftijd, nationaliteit, regio, reden van staande- of aanhouding). Ook blijkt elke keer opnieuw dat er nog niet geobserveerde heterogeniteit is. We hebben geen aanwijzingen dat de aard en mate van heterogeniteit in de loop van jaren ingrijpend is veranderd.

Geobserveerde heterogeniteit bij DSE hadden we beter kunnen meenemen als de overlap tussen beide bestanden groter was geweest, maar die is vrij beperkt (70 records). Nu is gebruik gemaakt van de variabelen geslacht, leeftijd en nationaliteit. Bij niet-geobserveerde heterogeniteit van alleen de kans om in het politieb Bestand, of alleen in het IOM- Bestand te komen, is er geen probleem. Er is ook niets aan de hand met de schatting als er niet-geobserveerde heterogeniteit is voor het politieb Bestand, naast niet geobserveerde heterogeniteit om in de IOM te komen, maar de twee typen heterogeniteit niet gerelateerd zijn (bv, voor IOM is er heterogeniteit ten gevolge van een bepaalde

variabele en voor politie van een andere variabele, en deze twee variabelen zijn niet gerelateerd). Echter, als het zo is dat er niet geobserveerde heterogeniteit is, zowel om in het IOM-bestand als in politiebestand te komen, waarbij de samenhang bij beide bestanden hetzelfde werkt, dan is de gevonden schatting te laag.

Gesloten populatie

Het Poissonmodel veronderstelt voor iedere afzonderlijke schatting dat de populatie in de periode van de studie stabiel blijft. Personen die tijdelijk afwezig zijn uit de populatie zijn gedurende die periode niet 'pakbaar', of, in het geval van DSE niet 'koppelbaar' en hebben daardoor ook geen Poisson-parameter. Schendingen van de assumptie van een gesloten populatie leiden bij beide schattingsmethoden tot een overschatting van de populatieomvang. In het geval van de Poisson schattingen corrigeren we voor het open karakter van de populatie met een stapsgewijze procedure die rekening houdt met (eventuele veranderingen in) de mate van 'openheid' of 'geslotenheid' van delen van de geschatte populatie (INEE, IEE en EUR, zie hoofdstuk 2).

Vreemdelingen geregistreerd door IOM vertrekken na verloop van tijd met hulp van de organisatie uit Nederland. In die zin is de IOM-doelgroep te beschouwen als een open populatie. Na vertrek is er geen kans meer dat een registratie in het IOM-bestand nog gekoppeld wordt aan een registratie in PSHV. Dit leidt tot een overschatting van de populatie.

Onafhankelijkheid

Bij de Poisson schatting wordt aangenomen dat opeenvolgende aanhoudingen of observaties onafhankelijk zijn van elkaar. Een al dan niet aanhouding mag niet leiden tot een verandering in de kans om weer aangehouden te worden. Bijvoorbeeld, als iemand is aangehouden, dan mag dat er niet toe leiden dat iemand roekeloos wordt omdat hij heeft gemerkt dat de aanhouding niet tot uitzetting heeft geleid. Of, als iemand aangehouden is voor diefstal, dan mag dat niet leiden tot een verhoogde kans dat deze persoon weer wordt aangehouden, doordat de politie bij elk voorval 'the usual suspects' aanhoudt of in het vizier neemt.

We hebben geconstateerd dat er nu meer vreemdelingen op grond van overtredingen en delicten worden aangehouden dan op grond van de Vreemdelingenwet. Verder zijn er aanwijzingen dat de politie meer prioriteit legt op bestrijding van criminaliteit en overlast dan op handhaving van de Vreemdelingenwet. Deze ontwikkelingen kunnen het vermoeden oproepen dat de onafhankelijkheid van opeenvolgende aanhoudingen in het gedrang komt. Echter, in zoverre deze ontwikkelingen een verandering van de aard en samenstelling van de populatie weerspiegelen (meer vreemdelingen die delicten plegen en voor overlast zorgen of minder moeite doen om contacten met de politie te vermijden) is er géén sprake van schending van de veronderstelling. Ook de verlegging van prioriteiten door de politie hoeft geen schending te betekenen van deze veronderstelling. De Poisson schatting herstelt 'automatisch' de balans: als de politie de aandacht verlegt naar criminele en overlast veroorzakende vreemdelingen, wordt een groter deel van de vreemdelingen die op basis van het strafrecht en andere wetten dan de Vreemdelingenwet worden aangehouden, meerdere keren aangehouden. Anderzijds zal een groter deel van de vreemdelingen die op basis van de Vreemdelingenwet worden staandegehouden, maar één keer worden staandegehouden. Het resultaat is een kleinere bij-schatting van het aantal vreemdelingen dat delicten pleegt of overlast veroorzaken en een hogere bij-schatting van het aantal dat alleen op basis van de Vreemdelingenwet aangesproken kan worden.

In het geval de politie bij de bestrijding van overlast en criminaliteit van onrechtmatig verblijvende vreemdelingen zich duidelijk meer is gaan richten op het aanpakken van de ‘usual suspects’, de al bekende verdachten, dan heeft een eerdere aanhouding invloed op de kans van een volgende aanhouding (die wordt groter) en is er wèl sprake van een veranderende schending van de veronderstelling die invloed heeft op de ontwikkeling van de reeks schattingen.

Bij de DSE is een belangrijke aanname dat de kans op het voorkomen in het politiebestand statistisch onafhankelijk is van de kans op het voorkomen in het IOM bestand. Als we de vier frequenties weer geven als, n_{11} , n_{01} , n_{10} en de schatting n_{00} , dan zijn de frequenties weer te geven in een 2x2 tabel:

$$\begin{array}{cc} n_{11} & n_{10} \\ n_{01} & n_{00} \end{array}$$

Bij onafhankelijkheid geldt dat $n_{11}/n_{10} = n_{01}/n_{00}$, dus $n_{00} = n_{10}n_{01}/n_{11}$. Als er negatieve afhankelijkheid is dan daalt n_{11} t.o.v. n_{10} en n_{01} . Als dan onafhankelijkheid wordt aangenomen, dan is de schatting voor n_{00} te hoog. Door de geringe overlap tussen de bestanden van de politie en IOM lijkt het erop dat, als men eenmaal contact heeft gehad met de politie (c.q. IOM), de kans op contact met de IOM (c.q. politie) lager wordt (negatieve afhankelijkheid). De schatting van (ruim) 58.000 is daardoor te hoog.

Bij de aanvraag voor ondersteuning stemmen IOM-cliënten in met automatische gegevensuitwisseling met het Ministerie van Justitie en Veiligheid. Die uitwisseling betreft gegevens over verblijfsrechtelijke status, asiel, beschikbaarheid van een vreemdelingsnummer ed. De gegevens die de vreemdeling verstrekt aan IOM worden echter niet gedeeld met de politie. Er is dus geen reden om aan te nemen dat er sprake is van positieve afhankelijkheid die tot overschatting van de populatie zou leiden.

Koppeling van bestanden

Bij DSE wordt verondersteld dat een perfecte koppeling tussen de gebruikte registers tot stand kan worden gebracht. Er mag geen sprake zijn van onterechte of gemiste koppelingen. Het IOM bestand en PSHV zijn probabilistisch aan elkaar gekoppeld om het aantal koppelingen te optimaliseren. De mogelijkheid bestaat echter nog steeds dat koppelingen zijn gemist. Als de overlap tussen bestanden gering is, zoals in het geval van het IOM-bestand en PSHV (70 records) kan een klein aantal gemiste koppelingen een merkbare invloed hebben op de hoogte van de schattingen. Dat wordt weerspiegeld door het grote betrouwbaarheidsinterval van de schatting. Het missen van koppelingen leidt tot een overschatting van de populatie (een verslag van de koppeling van IOM en PSHV is te vinden in bijlage 3).

Conclusie

De vergelijking van de schattingen met de twee methoden leidt tot de conclusie dat de beide schattingen voor een deel overlappen en voor een deel andere segmenten van de populatie in kaart brengen. De verschillen schuilen in de verhoudingen van de geslachten, de leeftijdsgroepen, nationaliteiten en de groepen met en zonder contacten met de politie. De populatie geschat met DSE heeft

betrekkelijk weinig politiecontacten, is ouder, vaker een vrouw, vaker afkomstig uit Azië en minder vaak uit Noord-Afrika dan de populatie geschat met Poisson. Beide schattingen mogen niet zomaar worden opgeteld, juist omdat ze deels overlappen. Hoe groot deze overlapping is, is niet goed te beoordelen. De beoordeling van de schendingen van de veronderstellingen van beide methoden leidt tot de conclusie dat de Poisson schatting waarschijnlijk een onderschatting is van de ware omvang van de populatie van onrechtmatig in Nederland verblijvende vreemdelingen in het jaar lopend van 1 juli 2017 tot en met 30 juni 2018 en de DSE-schatting waarschijnlijk een overschatting. De ware omvang van de populatie ligt dus waarschijnlijk tussen beide aantallen³⁴.

Voor de inmiddels 20-jarige reeks van Poisson schattingen is van belang dat met eventuele schending van een aantal van de veronderstellingen enigermate rekening kan worden gehouden (heterogeniteit van pakkansen, openheid van de populatie). Of er sprake is van verandering van schendingen van de veronderstellingen, kan niet worden nagaan op basis van de data die we hebben. Anderzijds hebben we ook geen aanwijzingen dat er van een dergelijke verandering sprake is. En dat is het standpunt dat we in dit rapport innemen: de puntschatting geproduceerd door het Poissonregressiemodel zit er mogelijk naast, maar wel op een vergelijkbare manier als bij eerdere schattingen.

4.6. Slotbeschouwing en conclusie

Achtergrond en doel van het onderzoek

Sinds 2000 worden er schattingen gemaakt van het aantal vreemdelingen in Nederland zonder verblijfsrecht. Er is een reeks van schattingen verkregen over een 20-jarige periode van 1997 tot en met het jaar lopend van 1 juli 2017 tot en met 30 juni 2018. De schattingen worden gemaakt met een vangst-hervangst methode. Deze methode biedt een aantal duidelijke voordelen en is in een internationale vergelijking aangemerkt als een methode die voor het schatten van een populatieomvang voldoet aan wetenschappelijke kwaliteitseisen.

Het belangrijkste doel van de schattingen is het kunnen volgen van de ontwikkeling van de omvang en samenstelling van de populatie vreemdelingen in Nederland zonder verblijfsrecht. De gedachte was, en blijft, dat de mate waarin de schattingsmethoden realistische veronderstellingen maken over de gebruikte data niet goed is te bepalen, maar dat we er van uit gaan dat een eventuele schending van de veronderstellingen over de jaren heen vergelijkbaar is.

Vanaf de schatting voor het jaar medio 2005 – medio 2006 geeft de reeks schattingen een scherpe daling te zien. Deze daling heeft met elke schatting opnieuw vragen en twijfels opgeroepen over de validiteit van de schattingen.

Dit nieuwe onderzoek voegt aan de reeks van schattingen een nieuwe schatting met het afgeknotte Poisson regressiemodel voor een nieuw jaar toe, te weten het jaar lopend van medio 2017 tot medio 2018. De twijfel en discussie over de validiteit van de schattingen vanwege de scherpe daling sinds

³⁴ We tekenen hierbij aan dat het Amerikaanse PEW Research Center voor het jaar 2017 een schatting heeft gemaakt met een tamelijk grofmazige methode, de zogenaamde ‘residuele methode’, gecombineerd met een ratiomethode onder de veronderstelling dat Nederland een vergelijkbare migratiegeschiedenis en vergelijkbaar migratiebeleid heeft als Duitsland, van maximaal ca. 90.000 onrechtmatig in Nederland verblijvende vreemdelingen. Bijlage 4 geeft meer informatie over de methode en de schatting.

2005, is aanleiding geweest voor twee toevoegingen aan het onderzoek. In de eerste plaats, is voor hetzelfde nieuwe jaar ook een schatting met een andere methode gemaakt. Het betreft een schatting met een variant van vangst-hervangstmethode die gebruik maakt van meerdere bestanden. In dit geval bleken 2 van de 3 beschikbare bestanden geschikt, te weten het sinds 2002 steeds gebruikte PSHV-bestand van de politie en een nieuw bestand van IOM. Het doel van het maken van een tweede schatting voor hetzelfde jaar met een andere methode is het verkrijgen van meer inzicht in de validiteit en het verhogen van de kwaliteit van de schattingen.

Een tweede belangrijke toevoeging van dit onderzoek is een verkenning van achtergronden die de snelle daling van de schattingen sinds 2005 kunnen verklaren, dan wel, bescheidener, plausibel kunnen maken.

Ontwikkeling van de geschatte populatie

We kunnen op basis van de genoemde verkenning van achtergronden geen sterke en exacte verbanden aantonen, maar we signaleren ontwikkelingen die het aannemelijk maken dat het aantal onrechtmatig in Nederland verblijvende vreemdelingen de afgelopen 10 à 15 jaar inderdaad aanzienlijk is afgenomen.

Al vanaf ongeveer het midden van de jaren '90 werd onrechtmatig verblijf van vreemdelingen in Nederland door diverse maatregelen gaandeweg onaantrekkelijk(er) gemaakt. Voor die tijd was het voor een onrechtmatig in Nederland verblijvende vreemdeling nog betrekkelijk eenvoudig om regulier te werken en de bijbehorende voordelen te genieten. Door diverse maatregelen werd het voor een vreemdeling zonder verblijfstitel allengs moeilijker om een zelfstandig bestaan op te bouwen. De toename van asielmigratie ten gevolge van politieke onrust in Midden- en Oost-Europa en delen van Azië en Afrika vanaf ongeveer dezelfde tijd, maakte een eventueel effect van deze maatregelen mogelijk ongedaan en moeilijker zichtbaar. De instroom van asielzoekers kan, met enige vertraging, hebben gezorgd voor de toename van de populatie vreemdelingen zonder verblijfsrecht in de jaren rond de millenniumwisseling. Er tekende zich in die jaren een duidelijk verschil af tussen het (groeiende) aantal asielverzoeken en het (beperkte) aantal toegekende verblijfsvergunningen. Dat verschil kan geleid hebben tot een toename van het aantal vreemdelingen dat na een afgewezen asielverzoek zonder verblijfstitel in het land bleef. Na een afname en stabilisatie van beide aantallen, deed zich in de jaren 2014 – 2017 bijna de omgekeerde ontwikkeling voor. Relatief veel asielverzoeken werden toegekend en ongeveer tussen medio 2015 en medio 2017 werden meer verblijfsvergunningen verleend dan asielverzoeken gedaan. Dit verschil kan een reden zijn waarom de toename van asielverzoeken in deze jaren geen toename in het geschatte aantal vreemdelingen zonder verblijfsrecht te zien geeft.

De uitbreiding van de toezichts- en opsporingsbevoegdheden door de Vreemdelingenwet 2000 en de reorganisatie van de vreemdelingenpolitie lijkt in de eerste helft van de jaren 2000 een kortstondig effect te hebben gehad op de aantallen onrechtmatig verblijvende vreemdelingen die werden staande- of aangehouden: het waren er duidelijk meer dan in eerdere jaren. Vanaf 2003 stabiliseerde het aantal aanhoudingen en ongeveer in 2004 zette een daling in. Het is mogelijk dat geïntensiveerd politietoezicht met de vreemdelingenwet en het strafrecht in de hand in deze jaren het aantal onrechtmatig in het land verblijvende vreemdelingen heeft teruggebracht en daarnaast onrechtmatig

verblijf onaantrekkelijker heeft gemaakt en aldus heeft bijgedragen aan een aanhoudende daling van het aantal onrechtmatig in Nederland verblijvende vreemdelingen.

De veranderingen die met de nieuwe Vreemdelingenwet in gang werden gezet hebben ook geleid tot meer aandacht voor persoonsidentificatie in het vreemdelingentoezicht. De deskundigheid en middelen voor identificatie werden sterk uitgebreid en er werd een prominente plaats gegeven aan biometrie. Een belangrijke ontwikkeling is dat identificatiegegevens inclusief biometrische gegevens eenvoudig toegankelijk werden gemaakt in de hele vreemdelingenketen, inclusief het basispolitiewerk op straat. In zoverre deze ontwikkelingen een betere identificatie van vreemdelingen mogelijk hebben gemaakt, zijn de schattingen voor eerdere jaren een overschatting en zijn de actuele schattingen accurater.

Het meest duidelijk effect op de omvang van de populatie gaat uit van de uitbreidingen van de EU in 2004 en 2007. De schattingen van het aantal onrechtmatig in Nederland verblijvende Europeanen uit West- en Oost-Europa vóór deze uitbreidingen variëren van 65.000 tot 107.000, terwijl de schattingen van de laatste jaren op 5.000 à 6.000 uitkomen.

Al met al denken we dat er reden is om te denken dat de beschreven ontwikkelingen in beleid, toezicht, asiel- en migratiestromen geleid hebben tot de substantiële afname van het aantal onrechtmatig in Nederland verblijvende vreemdelingen die door de reeks van schattingen wordt aangegeven. Verder is het van belang eraan te herinneren dat vooral de richting en kracht van de trend van belang zijn. De achtereenvolgende Poisson puntschattingen schatten de werkelijke populatieomvang mogelijk niet helemaal juist, gezien de geconstateerde schending van de veronderstelling is eerder aan te nemen dat er sprake is van (enige) onderschatting, maar als we kunnen aannemen dat deze schendingen van de veronderstellingen van het model ook in eerdere metingen voorkwamen, dan kunnen we nog wel steeds de ontwikkeling van de schattingen vertrouwen. Met (eventuele) schendingen van de veronderstellingen van de Poisson regressie kunnen we voor een deel rekening houden (heterogeniteit van pakkansen, open populatie). We hebben verder geen aanwijzingen dat de mate waarin de veronderstellingen worden geschonden in de loop van de tijd substantieel veranderd is. Onze conclusie is, dat de puntschatting er mogelijk naast zit, maar wel op een vergelijkbare manier als bij eerdere schattingen.

De schattingen voor het jaar medio 2017 – medio 2018

De populaties geschat met de Poissonmethode en met DSE verschillen aanzienlijk in omvang en vertonen weinig overeenkomsten in samenstelling. De schattingen overlappen elkaar voor een deel, maar brengen voor een ander deel juist ook andere delen van de populatie in kaart. De populatie geschat met DSE heeft betrekkelijk weinig politiecontacten, is ouder, vaker een vrouw, vaker afkomstig uit Azië en minder vaak uit Noord-Afrika dan de populatie geschat met Poisson. De twee schattingen mogen niet worden opgeteld, juist omdat ze deels overlappen. Hoe groot deze overlapping precies is, is niet goed te beoordelen. Daarom baseren we onze conclusies op de schending van de assumpties van de beide methoden en van de mate waarin deze schendingen al of niet kunnen worden onderzocht. Op die gronden concluderen we dat de Poisson methode waarschijnlijk een onderschatting oplevert van de ware populatieomvang en de DSE waarschijnlijk een overschatting. De ware omvang van de populatie onrechtmatig in Nederland verblijvende vreemdelingen in het jaar medio 2017 – medio 2018 ligt daarom waarschijnlijk tussen de puntschattingen van de Poisson methode en de DSE,

d.w.z. tussen de ca. 23.000 en ca. 58.000 (hierbij merken we op dat deze schattingen omgeven zijn door een betrouwbaarheidsinterval). Uit de hoogte van de Poissonschatting concluderen we dat de omvang van de populatie de laatste jaren grofweg stabiliseert (de DSE-schatting is voor eerdere jaren niet beschikbaar).

Bijlage 1: Parameterschattingen INEE, 2017 - 2018

	b	se	z-val	p-val
(Intercept)	-1.14	0.46	-2.47	0.01
Geslacht: man	0.00	-	-	-
Geslacht: vrouw	-0.73	0.37	-1.99	0.05
Leeftijd: 40+	0.00	-	-	-
Leeftijd: 40-	-0.03	0.19	-0.18	0.85
<u>Herkomst:</u>				
onbekend	0.00	-	-	-
Azië	-0.20	0.47	-0.44	0.66
Noord Afrika	-0.06	0.46	-0.13	0.90
Suriname	-0.28	0.60	-0.48	0.63
Afrika	-0.41	0.49	-0.85	0.40
Amerika	-0.77	0.55	-1.39	0.16
Turkije	-0.79	0.67	-1.19	0.23
<u>Regio:</u>				
Amsterdam	0.00	-	-	-
Den Haag	-0.65	0.33	-1.99	0.05
Rotterdam	-0.65	0.25	-2.61	0.01
Midden	-0.25	0.35	-0.71	0.48
Overig	-0.53	0.21	-2.45	0.01
<u>Reden aan/ staande-</u> <u>houding:</u>				
Delicten en overtredingen	0.00	-	-	-
Vreemdelingenwet	0.20	0.20	0.98	0.33

In de tabel staan de parameterschattingen van het Poissonmodel. Binnen elke covariaat is de parameter van de referentiecategorie op nul gezet, en de parameters van de overige categorieën geven het verschil aan met de referentiecategorie. Een positieve parameterschatting betekent dan dat de Poissonparameter van die categorie groter is dan die van de referentiecategorie, en dat dus ook de pakkans van die categorie groter is dan die van de referentiecategorie, en vice versa. Uit de tabel is dus bijvoorbeeld af te lezen dat vrouwen een kleinere Poissonparameter hebben dan mannen, en dat vrouwen dus een kleinere pakkans hebben. Dat betekent ook dat er per geobserveerde vrouw meer vrouwen zullen worden bijgeschat dan per geobserveerde man.

Bijlage 2: Gegevens afhandeling vreemdelingenzaken

Om in de Poissonregressie rekening te houden met de openheid van de populatie zijn de gegevens uit PSH-V gekoppeld aan gegevens uit de Basis Voorziening Vreemdelingen (BVV) om na te gaan of de vreemdeling in kwestie is uitgezet of dat anderszins een vertrek uit Nederland is geregistreerd. Op basis van deze informatie is een overzicht gemaakt van de wijze van verwijdering of vertrek (in de periode 1 juli 2017 tot en met 30 juni 2018) van vreemdelingen uit Nederland die in de periode zijn geregistreerd in PSHV. Het betreft vreemdelingen waarvoor, na aanhouding en (eventuele) detentie, een 'bericht verwijdering' is geregistreerd als laatste stap in de afhandeling van een zaak.

In voorkomende gevallen wordt in de BVV een bericht verwijdering en een bijbehorend vertrek niet als laatste stap in de afhandeling van een zaak geregistreerd. In die gevallen wordt (ten onrechte) bijvoorbeeld een opheffing van de inbewaringstelling ten behoeve van uitzetting als laatste stap geregistreerd. Om deze vertekening te ondervangen hebben we steeds wanneer er voor een geregistreerde vreemdeling een bericht verwijdering is aangetekend, dit bericht geselecteerd en beschouwd als een afsluiting van de zaak. Bovendien hebben we, om vertraagde registraties te ondervangen, de einddatum van de registratieperiode niet op 30 juni 2018 gesteld, maar op 7 juli.

De gegevens zijn in de volgende stappen verwerkt.

I. Het startbestand

In het geleverde BVV-bestand zijn de volgende selecties gemaakt met de volgende resultaten in aantallen records:

Tabel 2.1. Startbestand (BVV) en selecties

Selectie	Aantal records na selectie
1. Volledig bestand	15.617
2. Uitvoerdatum 01-07-2017 t/m 07-07-2018 (t/m 07/07 vanwege mogelijk vertraagde registratie)	7.097
3. act_activiteitsoortcode = 25 ('Bericht Verwijdering')	1.697
4. voor elke staandehouding laatste 'Bericht Verwijdering' in de genoemde periode	1.339
5. waarvan betrekking hebbend op staandehoudingen zonder rechtmatige verblijfstitel en verwijderbaar	1.320

II.

Voor de staandehoudingen in tabel 2.1 gelden de volgende aantallen voor de variabele 'referentiekenmerk' die de wijze van vertrek (of het gebrek aan evidentie voor vertrek) aangeeft:

Tabel 2.2. Referentiekenmerk (uit startbestand)

Referentiekenmerk	Aantallen staandehoudingen <i>n</i>
Aanzegging Nederland te verlaten	27
Opheffing IBS met aanzegging Nederland te verlaten	1
Overgave na controle MTV aan de landgrenzen	5
Uitzetting	885
Uitzetting vanuit strafrechttraject (conform de werkwijze van Vreemdelingen in de Strafrechtketen)	88
Vertrek onder toezicht MTV	8
Vertrek onder toezicht zelfmelder	51
Zelfstandig de woonruimte verlaten in of na de vertrektermijn van de procedure	202
Zelfstandig de woonruimte verlaten tijdens de procedure vÃ³r het ingaan van de vertrektermijn	31
Zelfstandig vertrek van een, bij controle op uitreis, illegaal gebleken vreemdeling	22
<i>Subtotaal informatie beschikbaar</i>	<i>1.320</i>
Onbekend	1.903
<i>Totaal</i>	<i>3.223</i>

III.

Omgezet in de volgende categorieën m.b.t. de aard van het 'vertrek':

Tabel 2.3. Vertaaltabel

	Gedwongen vertrek	Overig	Zelfstandig vertrek	Zelfstandig vertrek zonder toezicht
Aanzegging Nederland te verlaten				x
Ontruiming				x
Opheffing IBS met aanzegging Nederland te verlaten				x
Overgave na controle MTV aan landgrenzen	x			
Uitzetting	x			
Uitzetting vanuit strafrechttraject (conform VRIS-werkwijze)	x			
Vertrek onder toezicht MTV	x			
Vertrek onder toezicht van zelfmelder			x	
Zelfstandig de woonruimte verlaten in of na de vertrektermijn van de procedure				x
Zelfstandig de woonruimte verlaten tijdens de procedure vóór het ingaan van de vertrektermijn				x
Zelfstandig vertrek van een bij controle op uitreis illegaal gebleken vreemdeling			x	
NIET INGEVULD		x		

Dat geeft (tabel 2.4 is gelijk aan tabel 1 in het rapport):

Tabel 2.4. Inhoud 'Bericht Verwijdering', volgens BVV voor onrechtmatig verblijvende vreemdelingen geregistreerd in PSHV juli 2017 – juli 2018 (aantallen zaken)

	n	%
Zelfstandig vertrek zonder toezicht	334	10%
Aantoonbaar vertrek	986	31%
Onbekend	1.903	59%
Totaal	3.223	100%

IV. indeling INEE/ IEE/ EUR

Indeling tabel 4 wordt omgezet in de indeling in IEE, INEE en EUR met de volgende definities:

- o IEE: 'zelfstandig vertrek via de Internationale Organisatie voor Migratie', al of niet vanuit vreemdelingenbewaring) of een 'gedwongen vertrek' (o.a. uitzetting, uitzetting vanuit strafrecht en vertrek onder toezicht MTV) is geregistreerd, krijgen in de tabel de kwalificatie 'aantoonbaar vertrek'. Deze groep wordt in de gebruikte schattingsprocedure beschouwd als zijnde 'effectief uitgezet' (Illegal Effectively Expelled, IEE), tenzij zij afkomstig zijn uit een Europees land.

- INEE: 'zelfstandig vertrek zonder toezicht' / aanzegging Nederland te verlaten (tenzij zij afkomstig zijn uit een Europees land)
- INEE: (nog) geen vertrek of aanzegging daartoe geregistreerd is, de groep 'onbekend' (tenzij zij afkomstig zijn uit een Europees land).

Resultaat:

Tabel 2.5: INEE, IEE en EUR

	Personen	Staanhoudingen
INEE	1.603	1.742
IEE	498	578
EUR	852	903
Totaal	2.953	3.232

Bijlage 3: Probabilistisch koppelen bestanden ten behoeve van schatting onrechtmatig in Nederland verblijvende vreemdelingen

Daan Zult, CBS³⁵

Inleiding

Om een vangst-hervangst schatting te doen van het aantal onrechtmatig in Nederland verblijvende vreemdelingen is het nodig om verschillende bestanden met daarin records die betrekking hebben op deze vreemdelingen te koppelen. Idealiter gebeurt deze koppeling middels een unieke ID-code, maar als dit niet mogelijk is kan dit ook op basis van andere persoonskenmerken gebeuren, zoals bijvoorbeeld naam, adres, geslacht of een combinatie van dit soort kenmerken. Deze kenmerken worden koppelsleutels genoemd en deze kunnen worden gebruikt om probabilistisch te koppelen (zie hiervoor Fellegi en Sunter, 1969; Winkler, 1988; Jaro, 1989). Om goed probabilistisch te kunnen koppelen dienen de koppelsleutels voldoende onderscheidend en van voldoende kwaliteit te zijn. Dit betekent dat records aan de hand van hun koppelsleutels voldoende van elkaar moeten verschillen. Als in twee bestanden bijvoorbeeld alleen iemands geslacht bekend is dan is er nog onvoldoende informatie om te koppelen. Als daarbij in beide bestanden ook initialen, woonplaats en kennen, kunnen we iemand al met grotere kans correct koppelen. In verband met privacyoverwegingen merken we hierbij op dat in deze exercitie correct koppelen niet betekent dat iemand ook nauwkeurig geïdentificeerd wordt. Er wordt alleen bepaald of twee records bij dezelfde persoon behoren. Zonder dat bekend hoeft te zijn wie dit precies is. In hoeverre een combinatie van koppelsleutels voldoende onderscheidend is hangt ook af van de populatieomvang. Hoe groter de populatie, hoe waarschijnlijker dat twee records met dezelfde koppelsleutels toch verschillende personen zijn. Koppelsleutels moeten ook 'van voldoende kwaliteit' zijn, wat wil zeggen dat de koppelsleutels niet foutloos hoeven te zijn, maar ook niet te veel (ernstige) fouten mogen bevatten. Als in het ene bestand bijvoorbeeld een adres in het buitenland staat en in het andere een adres in Nederland, zal een probabilistisch koppelmodel dit als een teken opvatten dat het om twee verschillende personen gaat. Daarentegen, als in het ene bestand een kleine spelfout in het adres staat en in het andere bestand niet, dan zal een probabilistisch koppelmodel deze twee adressen als nagenoeg gelijk beschouwen. Tot slot is het belangrijk om op te merken dat een koppelsleutel in beide bestanden aanwezig moet zijn. De koppelsleutels zijn dus zo goed als de koppelsleutels in het minst rijke bestand.

Om een vangst-hervangst schatting van het aantal vreemdelingen te kunnen doen moeten ten minste twee maar liever meer bestanden gekoppeld worden. De bestanden die in deze studie zijn overwogen zijn afkomstig van het CAK, Internationale Organisatie voor Migratie (IOM) en de politie, de Politie suite Handhaving Vreemdelingen (PSHV). Vanuit privacyoverwegingen hebben deze organisaties er soms voor gekozen niet alle koppelsleutels volledig ter beschikking te stellen. De eerste vraag die we daarom in de datasectie zullen behandelen is of de bestanden nog genoeg informatie bevatten om nauwkeurig te kunnen koppelen. Maar eerst bespreken we de probabilistische koppel methode iets gedetailleerder.

³⁵ Behalve bij deze bijlage was het Centraal Bureau voor de Statistiek op geen enkele wijze betrokken bij het opstellen van de inhoud van dit rapport. De in deze bijlage weergegeven opvattingen zijn die van de auteur en komen niet noodzakelijk overeen met het beleid van Centraal Bureau voor de Statistiek.

Methode

Probabilistisch koppelen werkt over het algemeen als volgt. Er zijn twee bestanden met M en N records en beide hebben dezelfde K koppelsleutels. Deze koppelsleutels kun je per koppelsleutel met elkaar vergelijken en de mate van overeenkomst een score tussen 0 en 1 geven, waarbij twee identieke koppelsleutels een score van 1 krijgen en deze score loopt terug naar 0 als records helemaal anders zijn. Op deze manier krijgen records die door een spelfout (bijvoorbeeld ‘dorpstraat’ en ‘dropstraat’) toch een score dichtbij 1 maar records die totaal anders zijn (bijvoorbeeld ‘dorpstraat’ en ‘brink’) een score van 0. Vervolgens wordt ook van elke koppelsleutel bepaald hoe groot de kans a priori is dat ze identiek zijn. De kans dat bijvoorbeeld het geslacht van twee willekeurige personen identiek is, is over het algemeen groter dan een overeenkomst van iemands initialen. Een overeenkomst op de tweede koppelsleutel zou daarom zwaarder moeten wegen bij het bepalen of een paar een werkelijke koppeling is of niet. Op deze manier kan worden berekend wat de kans is dat twee koppelsleutels overeenkomen als twee records een koppeling zijn en wat de kans is dat twee koppelsleutels overeenkomen als twee records geen paar zijn. Dit zijn de zogenaamde m - en u -kansen van Fellegi en Sunter. Door deze twee kansen te combineren krijgen we een gewogen score waarbij geldt dat hoe hoger het gewicht, hoe groter de kans op een werkelijke koppeling. Op basis van deze gewogen scores kun je in de groep van $M \times N$ potentiële koppelingen van records, records identificeren die waarschijnlijker een koppeling vormen dan anderen. Over het algemeen geldt dat lage gewichten zeker niet gekoppeld kunnen worden en hoge gewichten zeker wel, maar daartussen zit een groep waar even goed naar gekeken moet worden. Door goed naar deze koppelingen te kijken kan worden besloten wanneer we ervan uitgaan wanneer iets waarschijnlijk een koppeling is en wanneer niet. Hieruit volgt een zogenaamde grenswaarde gewicht, waarboven alles een koppeling is en daaronder niet. De keuze voor deze grenswaarde is enigszins arbitrair en kan worden bepaald door de koppelingen met een waarde rond deze grenswaarde wat beter te bekijken. Tot slot, omdat beide bestanden in principe unieke records bevatten geldt bij het koppelen de restrictie dat een record niet vaker dan één keer gekoppeld mag worden, waardoor koppelingen van records met een hoge gewogen score toch af kunnen vallen ten faveure van een andere koppeling met nog hogere gewogen score. De programmatuur waarmee bovenstaande kan worden uitgevoerd is de R package *reclin* (van der Laan, 2018).

Data

In deze sectie zullen we een aantal relevante details over de drie databestanden die we ter beschikking hebben bespreken en aan de hand hiervan beoordelen of ze geschikt zijn om te gebruiken in een probabilistische koppeling. Wat voor alle bestanden geldt is dat ze de periode juli 2017 tot en met juni 2018 betreffen, de vangst-hervangst schatting zal dus ook deze periode betreffen.

CAK-bestand

Het CAK-bestand betreft onverzekerbare vreemdelingen die zorg hebben ontvangen van een zorginstelling. Het CAK zorgt ervoor dat dit financieel en administratief wordt afgehandeld. Het bestand kent echter een aantal beperkingen waardoor probabilistisch koppelen niet verstandig is. Ten eerste, het bestand is een stuk groter dan de andere twee bestanden. Dit komt mede doordat niet personen maar gebeurtenissen worden geregistreerd, waardoor een persoon die twee of meer keer

behandeld is ook twee of meer keer in het bestand kan voorkomen. Op basis van de beschikbare koppelsleutels (initialen (eerste letter voornaam en eerste letter achternaam), geslacht, geboortjaar en nationaliteit) zouden dan unieke personen kunnen worden geselecteerd, maar het is onduidelijk of dit goed gaat. Zo zien we bijvoorbeeld records met dezelfde koppelsleutels waardoor ze dezelfde persoon lijken, maar in andere ziekenhuizen geholpen zijn, waardoor je kunt twijfelen of dit wel echt dezelfde personen zijn. Wat het daarbij nog lastiger maakt is dat voor een deel van de records ook informatie (zoals initialen) ontbreekt, waarmee ook twijfel ontstaat over de kwaliteit van de koppelsleutels. Als een persoon bij een tweede ziekenhuisbezoek bijvoorbeeld andere initialen opgeeft dan blijft dat onopgemerkt. Voor de records met missende informatie blijft dermate weinig informatie over om te bepalen of zo'n record uniek is. Tot slot komt er nog bij dat zelfs als het lukt om unieke personen te identificeren, dan zijn de koppelsleutels vermoedelijk te weinig onderscheidend om nauwkeurig te kunnen koppelen aan andere bestanden. We beschouwen het CAK-bestand daarom ongeschikt om probabilistisch te koppelen.

IOM bestand

Het IOM bestand beslaat 2017 en 2018 en heeft 5606 records. Als we selecteren op records die tussen juli 2017 en juni 2018 zijn ontstaan blijven er 2281 records over. IOM ondersteunt migranten (ex-asielzoeker, vergunning verlopen, ongedocumenteerden) die vrijwillig terugkeren naar het land van herkomst of doormigreren naar een derde land waar permanent verblijf is gewaarborgd (hervestiging). De vreemdelingen in dit bestand zijn de personen waarvan de asielaanvraag is afgewezen of personen die nooit een asielaanvraag hebben gedaan. Ook in dit bestand zijn de personen geanonimiseerd, maar de koppelsleutels zijn toch iets completer. Dit zijn geslacht, land van herkomst, volledige geboortedatum, initialen (twee letters van de voornaam en twee letters van de achternaam). Deze koppelsleutels bevatten bovendien weinig missende informatie. Tot slot bevat dit bestand een zogenaamd vreemdelingsnummer, welke de rol van uniek identificatienummer heeft. De aanwezigheid van dit vreemdelingsnummer heeft als voordeel dat personen die dubbel voorkomen in het bestand kunnen worden geïdentificeerd, dit zijn er 134, het aantal unieke records in het IOM bestand is dus 2147. Het IOM bestand lijkt geschikt om probabilistisch te koppelen.

PSHV-bestand

Het PSHV beslaat juli 2017 tot en met juni 2018 bestand en heeft 4215 records. De records betreffen niet - geregistreerde vreemdelingen die in aanraking zijn geweest met de politie. Het PSHV-bestand bevat dezelfde koppelsleutels als het IOM bestand (geslacht, land van herkomst, volledige geboortedatum, initialen (twee letters van de voornaam en twee letters van de achternaam)) en ook bevatten ze weinig missende informatie. Daarbij bevat het PSHV-bestand ook het vreemdelingsnummer dat in het IOM bestand voorkomt. Aan de hand van dit vreemdelingsnummer blijkt dat er in het PSHV-bestand 378 dubbeltellingen zitten, het aantal unieke records is dus 3837. Het PSHV-bestand lijkt geschikt om probabilistisch te koppelen.

Conclusie met betrekking tot probabilistisch te koppelen

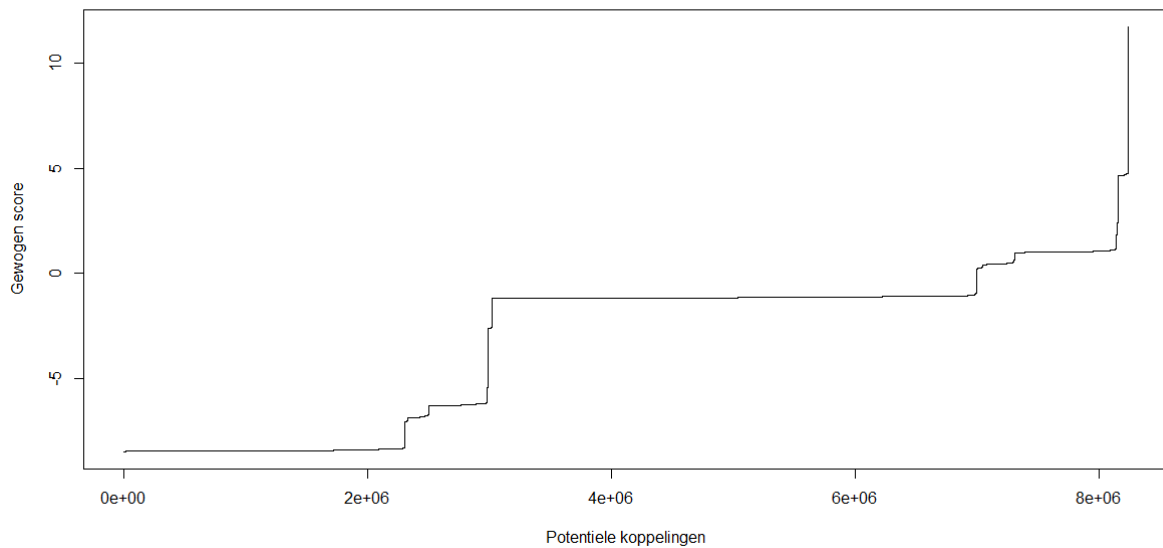
Op basis van de beschouwing van het CAK, IOM en PSHV-bestand komen we tot de conclusie dat het CAK-bestand ongeschikt is voor ons doel, maar het IOM en PSHV-bestand kunnen probabilistisch aan elkaar gekoppeld worden. Wel is de vraag of probabilistisch koppelen gezien de aanwezigheid van

een vreemdelingsnummer veel meerwaarde heeft, omdat hiermee in principe ook exacte koppeling mogelijk is.

Resultaten

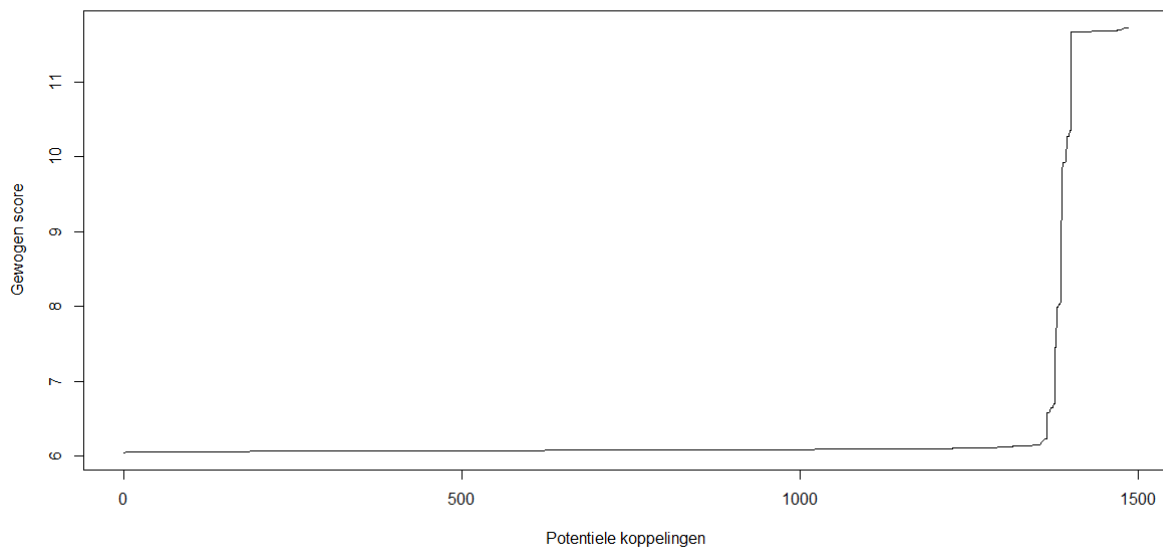
Om te beginnen kunnen we op basis van het vreemdelingsnummer 126 records koppelen. Gezien de omvang van de bestanden is dit aantal best klein, mogelijk worden er dus koppelingen gemist. Als we beide bestanden probabilistisch koppelen zien we wel dat de 126 records een hoge gewogen score krijgen. Maar, we vinden ook andere potentiële koppelingen met een verschillend vreemdelingsnummer die wel een hoge gewogen score krijgen. In figuur 1 zien we de gewogen scores van alle 8249550 potentiële koppelingen.

Figuur 1: Gewogen scores (gesorteerd) van probabilistische koppeling IOM en PSHV-bestand.



De 126 vreemdelingsnummer koppelingen vinden we in de staart rechts, maar daar vlakbij vinden we nog een groep met gewogen scores die oplopen van ongeveer 6 naar bijna 12. Als we daarop inzoomen krijgen we figuur 2.

Figuur 2: Gewogen scores (gesorteerd, groter dan 6) van probabilistische koppeling IOM en PSHV-bestand.



De lange platte lijn betreft rond de 1350 potentiële koppelingen en ligt ongeveer op een gewogen score van 6.15. Gezien de kwaliteit van het vreemdelingsnummer lijkt het onwaarschijnlijk dat er 1350 koppelingen zijn die aan de hand van het vreemdelingsnummer niet koppelen. De groep met een gewogen score > 6.15 lijkt dus een groep waar mogelijk gemiste koppelingen in zitten.

Voordat we nog verder inzoomen moeten we eerst een 1 op 1 koppeling afdwingen, zodat een record die erg op twee andere records lijkt niet aan beide gekoppeld wordt. Van de 126 records die op vreemdelingsnummer koppelen, hebben dan 108 koppelingen een gewogen score groter dan 6.15 en 18 koppelingen een gewogen score lager dan 6.15. Het lijkt echter onwaarschijnlijk dat twee records toevallig hetzelfde vreemdelingsnummer hebben, we beschouwen deze koppelingen daarom toch als correcte koppelingen. Daarbij hebben nog 21 potentiële koppelingen een gewogen score hoger dan 6.15, terwijl ze niet koppelen op basis van hun vreemdelingsnummer. Deze laatste groep van 21 koppelingen verdient dus een wat nadere beschouwing. Zonder op details in te gaan blijken er in deze groep 6 koppelingen te zijn waarvan we op basis van spelfouten en/of registratiefouten in de koppelsleutels sterk vermoeden dat het om hier om gemiste koppelingen gaat. Deze records hebben dus mogelijk onterecht een verschillend vreemdelingsnummer en zouden we dus als gemiste koppelingen moeten beschouwen. Tot slot is het goed om op te merken dat het missen van 6 van de 126 koppelingen op basis van vreemdelingsnummer weinig lijkt, maar in een vangst-hervangst model kan dit de schattingen behoorlijk beïnvloeden.

Literatuur

Fellegi, I. P. and Sunter, A. B. (1969). A Theory for Record Linkage. *Journal of the American Statistical Association*, 64, 1183 - 1210.

Jaro, M. (1989). Advances in Record Linkage Methodology as Applied to Matching the 1985 Test Census of Tampa, Florida. *Journal of American Statistical Association* 84: 414 - 420.

Van der Laan, J. (2018). reclin: Record Linkage Toolkit. R package version 0.1.1.

Winkler, W. E. (1988). Using the EM algorithm for weight computation in the Fellegi-Sunter model of record linkage. Section on Survey Research Methods, 667 - 671.

Bijlage 4: schattingen PEW Research Center voor EU-lidstaten 2017

Het Amerikaanse PEW Research center heeft voor het jaar 2017 schattingen gemaakt van het aantal 'unauthorized immigrants' in EU-landen. De instelling gebruikt daarvoor de zogenaamde 'residual method', of residuele schattingsmethode als basis. Dat is de methode waarbij de schatting wordt verkregen door vergelijking van bekende aantallen uit diverse bronnen (bijvoorbeeld volkstelling, bevolkingsregister, surveys)¹. Deze methode wordt in de Verenigde Staten al meer dan 15 jaar gebruikt voor de schattingen van de aantallen onrechtmatig verblijvende immigranten. Het basismodel is²:

$$\begin{array}{rcl} \text{Unauthorized} & & \text{Survey, Total} \\ \text{Immigrants} & = & \text{Foreign Born} - \\ \text{(U)} & & \text{(F)} \quad \quad \quad \text{Estimated Lawful} \\ & & & \text{Immigrant Population} \\ & & & \text{(L)} \end{array}$$

Pew Research Center (PRC) gebruikte deze methode om de omvang van de ongeautoriseerde immigrantenpopulatie in o.m. Duitsland te schatten. Schattingen voor een aantal andere landen, waaronder Nederland, worden vervolgens gemaakt met een combinatie van deze methode en de multiplier- of ratio-methode. Schattingen voor de totale niet-EU-EVA³-burgerbevolking voor Duitsland zijn gebaseerd op de cijfers van Eurostat, aangevuld met gegevens uit de Duitse enquête beroepsbevolking, een verplichte, jaarlijkse enquête op basis van adressen in het bevolkingsregister en actualiseringen van andere adressen in Duitsland, ongeacht de legale status van de ingezetenen. In 2017 haalde deze enquête een respons van 96%. De aanpak leidt tot een schatting van de onrechtmatig in Duitsland verblijvende immigrantenpopulatie van tussen 1,0 miljoen en 1,2 miljoen voor het jaar 2017, inclusief asielzoekers in procedure. Voor Nederland zijn niet dezelfde bronnen aanwezig. Daarom is het verhoudingsgetal tussen dit geschatte aantal onrechtmatig in Duitsland verblijvende vreemdelingen en het totale aantal in Duitsland verblijvende vreemdelingen afkomstig uit landen buiten de EU en EVA toegepast op datzelfde aantal vreemdelingen (van buiten de EU en EVA) in Nederland om te komen tot een schatting van de Nederlandse populatie 'unauthorized immigrants', oftewel onrechtmatig in Nederland verblijvende vreemdelingen, inclusief asielzoekers in procedure. Op deze manier komt PRC tot een schatting van minder dan 100.000 onrechtmatig in Nederland verblijvende vreemdelingen voor elk van de jaren 2014, 2015, 2016 en 2017 met in 2017 een aantal van 7.000 asielzoekers in procedure⁴.

¹ Zie ook: Sikkel, D., P.G.M. van der Heijden en G. van Gils (2006). Methoden voor omvangschattingen van verborgen populaties, met name illegalen. (WODC rapport 243, Onderzoek en beleid). Meppel: Boom Juridische Uitgevers.

² <https://www.pewresearch.org/hispanic/2016/09/20/methodology-10/>

³ Europese Unie/ Europese VrijhandelsAssociatie.

⁴ Phillip Connor, Jeffrey S. Passel, Europe's Unauthorized Immigrant Population Peaks in 2016, Then Levels Off, New Estimates find half life in Germany and the United Kingdom, PEW Research Center, November 13, 2019, pp. 29, 31-32, 43.

Bijlage 5: De begeleidingscommissie

De heer prof. dr. B.F.M. Bakker UU - Capaciteitsgroep Methodenleer en Statistiek VU - Faculteit der Sociale Wetenschappen (voorzitter)

De heer dr. J. de Boom, Risbo Wetenschappelijk Onderzoek- en Documentatiecentrum, EUR, Rotterdam

De heer dr. H.C.J. van der Veen, WODC

De heer dr. R. van Hest MD, MSc, PhD, GGD Groningen - Afdeling tuberculosebestrijding

De heer F. Topçu MSc, Ministerie van Justitie en Veiligheid - Directie Migratie Beleid

Mevrouw P. le Roy, Ministerie van Justitie en Veiligheid - Directie Migratie Beleid

Mevrouw dr. S. Scholten, Politie Eenheid Den Haag