

# Time series modelling in repeatedly conducted sample surveys

## Citation for published version (APA):

Balabay, O. (2016). *Time series modelling in repeatedly conducted sample surveys*. Datawyse / Universitaire Pers Maastricht. <https://doi.org/10.26481/dis.20160511ob>

## Document status and date:

Published: 01/01/2016

## DOI:

[10.26481/dis.20160511ob](https://doi.org/10.26481/dis.20160511ob)

## Document Version:

Publisher's PDF, also known as Version of record

## Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

## General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.umlib.nl/taverne-license](http://www.umlib.nl/taverne-license)

## Take down policy

If you believe that this document breaches copyright please contact us at:

[repository@maastrichtuniversity.nl](mailto:repository@maastrichtuniversity.nl)

providing details and we will investigate your claim.

# Nederlandse samenvatting

Kort samengevat is het doel van dit proefschrift het verbeteren van de kwaliteit van officiële statistieken. Kwaliteit van statistische informatie heeft meerdere facetten. Dit werk richt zich op twee van de door EUROSTAT gedefinieerde kwaliteitskenmerken: precisie van geproduceerde cijfers en vergelijkbaarheid daarvan door de tijd heen.

Statistische bureaus voeren steekproefonderzoeken uit om sociaal-economische ontwikkelingen in onze samenleving zo nauwkeurig mogelijk te beschrijven. In een steekproef onderzoek wordt aan de hand van een kansmechanisme een klein deel van de doelpopulatie geselecteerd. Aan de hand van de in de steekproef getrokken elementen worden vervolgens schattingen gemaakt voor de onbekende populatie grootheden. Omdat slechts een fractie van de doelpopulatie wordt waargenomen ontstaat er altijd een afwijking tussen de schatting en de echte (maar onbekende) waarde van de populatiegrootte. Deze onzekerheid wordt gemeten met de zogenaamde variantie van de steekproefschattingen. Deze variantie meet in hoeverre puntschattingen over alle mogelijke kanssteekproeven variëren, die volgens het steekproefontwerp zouden kunnen worden getrokken. Gebruikelijke schattingsmethodieken zijn voornamelijk gebaseerd op het kansmechanisme waarmee de steekproef uit de doelpopulatie is getrokken. Statistische modellen spelen hierbij geen of slechts in zeer beperkte mate een rol. Deze methoden worden daarom vaak design-gebaseerde schattingsmethoden genoemd. Het voordeel van deze methoden is dat ze niet gevoelig zijn voor veronderstellingen van een gekozen statistisch model. Bij steekproeven van voldoende omvang resulteren deze schattingsmethoden in uitkomsten die in verwachting overeen komen met de echte populatiewaarden. Dit soort eigenschappen heeft er in het verleden voor gezorgd dat het gebruik van design-gebaseerde schattingsmethoden bij statistische bureaus erg populair is ge-

worden.

Het nadeel van design-gebaseerde schattingsmethoden is dat bij kleine steekproeven de varianties van de schatters onacceptabel groot worden. Dit probleem ontstaat indien schattingen voor gedetailleerde uitsplitsingen moeten worden gemaakt, bijvoorbeeld een uitsplitsing van een variabele naar de afzonderlijke gemeenten. Deelpopulaties waarvoor schattingen gemaakt moeten worden, worden vaak kortweg aangeduid als domeinen. Indien de steekproefomvang van dergelijke domeinen te klein is om met design-gebaseerde methoden voldoende betrouwbare schattingen te maken, spreekt men ook wel over kleine domeinen. Soms komt het voor dat een domein helemaal geen steekproefmassa krijgt, terwijl er toch een schatting voor is vereist. Een triviale manier om de betrouwbaarheid van deze cijfers te verbeteren is het verhogen van de steekproefomvang in de domeinen waarvoor schattingen gemaakt moeten worden. Meestal is dit niet realistisch omdat dit resulteert in onacceptabel hoge kosten voor nationale statistische bureau's en in een te hoge enquêtedruk voor respondenten. Het probleem van kleine steekproefomvang gaf in de afgelopen decennia aanleiding tot een nieuw onderzoeksgebied – small area estimation. Dit refereert naar een klasse van schattingsmethoden waarbij, via een statistisch model, de effectieve steekproefomvang van een afzonderlijk domein wordt vergroot met steekproefinformatie waargenomen in andere domeinen of voorgaande waarnemingsperioden. Bij dergelijke schattingsmethoden spelen statistische modellen een prominente rol en worden daarom ook wel model-gebaseerde schattingsmethoden genoemd.

Naast steekproeffouten, is er altijd sprake van (systematische) meetfouten bij de dataverzameling. Deze hangen af van de manier waarop respondenten worden benaderd en de manier waarop vragenlijsten worden ontworpen. Een herontwerp van de enquête resulteert daarom vaak in een verandering van deze systematische fouten en in een sprong (breuk) in de trend van de tijdreeks. Als gevolg daarvan, zijn cijfers vóór en na een herontwerp niet meer vergelijkbaar.

Een manier om beide problemen op te lossen is om tijdreeksen, waargenomen via herhaaldelijk uitgevoerde kansteekproeven, te modelleren met behulp van tijdreeksmodellen. In dit proefschrift zijn zijn structurele (state space) tijdreeksmodellen en multilevel tijdreeksmodellen gehanteerd. Hoofdstuk 2 illustreert hoe univariate (d.w.z. één tijdreeksmodel per afzonderlijk domein) en multivariate (ge-

zamenlijk modeleren van meerdere tijdreeksen) modellen kunnen worden gebruikt in een officiële statistische toepassing (namelijk, Enquête Wegvervoer van Nederland (EWN)) om variantie te reduceren en trendbreuken te schatten. De EWN toepassing laat zien dat het grootste deel van variantiereductie afkomstig is van steekproefinformatie uit voorgaande perioden (in univariate modellen). Een extra variantiereductie wordt bereikt door gebruik te maken van steekproefinformatie uit andere domeinen, door tijdreeksen van verschillende domeinen te modeleren in een multivariaat model. Bij de kleinste domeinen van de EWN worden standaardfouten tot wel 70 of 80 procent gereduceerd. Bij de grote domeinen en bij de reeks op nationaal niveau wordt een reductie van 40-60 procent bereikt.

In hoofdstuk 3 wordt nader ingegaan op de verschillende manieren waarop tijdreeksen op nationaal niveau kunnen worden geschat. Het uitgangspunt hierbij is dat met behulp van een tijdreeksmodel preciezere schattingen moeten worden verkregen maar dat deze schattingen consistent moeten zijn met de som van onderliggende domeinen. Een mogelijkheid is om de reeks op nationaal niveau aan het model toe te voegen of met deze reeks een andere domeinreeks te vervangen, omdat reeksen op een hoger aggregatieniveau een betere signaal-ruisverhouding hebben. Beargumenterd wordt dat dit niet efficiënt is, vooral als het weggelaten domein trendbreuken heeft. De reden daarvoor is dat, bij het afleiden van model-schattingen voor het (kleine) weggelaten domein, verwaarloosbare misschattingen op geaggregeerd niveau tot enorme misschattingen op domeinniveau kunnen leiden. Voor de EWN wordt aanbevolen om schattingen op nationaal niveau af te leiden uit het multivariaat model voor de onderliggende domeinen.

State-space modellen worden vaak geanalyseerd met het Kalman filter. Een probleem bij het toepassen van de standaard Kalman filter recursie is dat de bijbehorende standaardfouten geen rekening houden met het feit dat de onbekende hyperparameters in het state-space model zijn vervangen door hun maximum likelihood schattingen. In de praktijk wordt de onzekerheid rond deze hyperparameter schattingen vaak genegeerd, wat resulteert in een onderschatting (of een negatieve vertekening) in de standaardfouten van de kleinedomeinschattingen op basis van het state-space model. Het gebruik van te optimistische betrouwbaarheidsintervallen bij belangrijke sociaal-economische indicatoren, zoals bijvoorbeeld werkloosheid, bemoeilijkt een correcte interpretatie van de resultaten voor bijvoorbeeld beleidsmakers. Het Centraal Bureau voor de Statistiek (CBS) maakt gebruik van

een state-space model voor het maken van officiële maandcijfers over de werkloze en werkzame beroepsbevolking. Standaardfouten zijn gebaseerd op de standaard Kalman filter recursie. In hoofdstuk 4 wordt onderzocht hoe groot de onderschatting is in de onzekerheid van maandcijfers over de werkloze beroepsbevolking in de Nederlandse Enquête Beroepsbevolking (EBB) ten gevolge van het vervangen van de onbekende hyperparameters door hun maximum likelihood schatters. Daartoe is een simulatie studie opgezet waarbij het tijdreeksmodel van het CBS gebruikt wordt als het data genererend proces.

Het blijkt dat de vertekening in de standaardfout bij deze toepassing niet groot is en verwaarloosbaar wordt naarmate de reeks langer wordt. De simulatie laat ook zien hoe het model beter gespecificeerd kan worden. Tenslotte, geeft de simulatie in hoofdstuk 4 nieuwe inzichten in de verschillende methoden die in de literatuur bekend zijn om rekening te houden met de onzekerheid van maximum likelihood schattingen voor de hyperparameters in state space modellen.

Zoals hierboven vermeld, zijn multilevel tijdreeksmodellen een alternatieve optie voor herhaaldelijk uitgevoerde enquêtes. De hiërarchische (volledige) Bayesiaanse (HB) aanpak, die vrij eenvoudig in het kader van multilevel modellen is te implementeren, is een generieke alternatieve aanpak om rekening te houden met alle onzekerheid, ook rond hyperparameters. In hoofdstuk 5 worden de state space en multilevel tijdreekstechnieken vergeleken op basis van het Onderzoek Verplaatsingen in Nederland (OVIN). Dit onderzoek heeft last van methodebreuken ten gevolge van veranderingen in de onderliggende enquête en kleine steekproefomvang. Het doel van het onderzoek beschreven in Hoofdstuk 5 is om na te gaan of beide methoden vergelijkbare uitkomsten leveren, en om de negatieve vertekening in de gemiddelde standaardfouten in state space modellen te kwantificeren. Het blijkt dat punt- en variantieschattingen afkomstig uit beide tijdreekstechnieken vergelijkbaar zijn. Verschillen verschijnen eerder in de geschatte varianties dan in de puntschattingen en komen alleen in kleinschalige domeinen of in domeinen met beperkt flexibele stochastische trends voor, omdat de variantie hyperparameters van de trends daar dicht bij de grens van de parameter ruimte liggen. Beide methoden bieden een behoorlijke reductie in design-gebaseerde standaardfouten. Voor het OVIN worden standaardfouten op nationaal niveau met circa 30 procent gereduceerd, uitgaande van het HB multilevel model. Op provinciaal niveau is de reductie circa 50 procent. Om dezelfde reductie in de standaardfouten van

de design-gebaseerde schattingsmethoden te bewerkstelligen zou de steekproefomvang meer dan verdubbeld moeten worden op nationaal niveau en meer dan vier keer zo groot moeten worden op provinciaal niveau. In het geval van OViN kan de negatieve vertekening in standaardfouten oplopen tot 8 procent op provinciaal niveau en tot meer dan 10 procent op nationaal niveau indien de standaard Kalman filter recursies gebruikt worden.

In state space modellen en multilevel modellen worden standaardfouten van de design-gebaseerde schattingen als prior informatie in het model gebruikt. Vaak zijn deze erg onbetrouwbaar en moeten daarom gesmoothd worden over de domeinen en over de tijd. In hoofdstuk 5 wordt een nieuwe manier voorgesteld om met ontbrekende en/of onbetrouwbare design-gebaseerde variantieschattingen om te gaan.

Vrijwel elke enquête die herhaaldelijk wordt uitgevoerd kan profiteren van tijdreeksmodellen beschreven in dit proefschrift. Deze beschreven technieken maken het mogelijk om de precisie van design-gebaseerde schattingen te verbeteren. Daarnaast kan rekening worden gehouden met systematische effecten op de uitkomsten die ontstaan door het veranderingen in het onderliggende enquêteproces. Dit laatste voorkomt dat cijfers voor en na een herontwerp niet meer vergelijkbaar zijn. Het gebruik van dergelijke model-gebaseerde schattingsmethode is in de productie van officiële statistieken nog vrij beperkt. Twee voorbeelden van nationale statistische bureau's die state space tijdreeksmodellen gebruiken in de productie van officiële cijfers zijn het Bureau of Labor Statistics in de Verenigde Staten en het CBS in Nederland. Hopelijk maakt dit proefschrift meer officiële statistici bewust van de potentie van state space en multilevel tijdreeksmodellen en biedt het richtlijnen voor enquête beoefenaars.