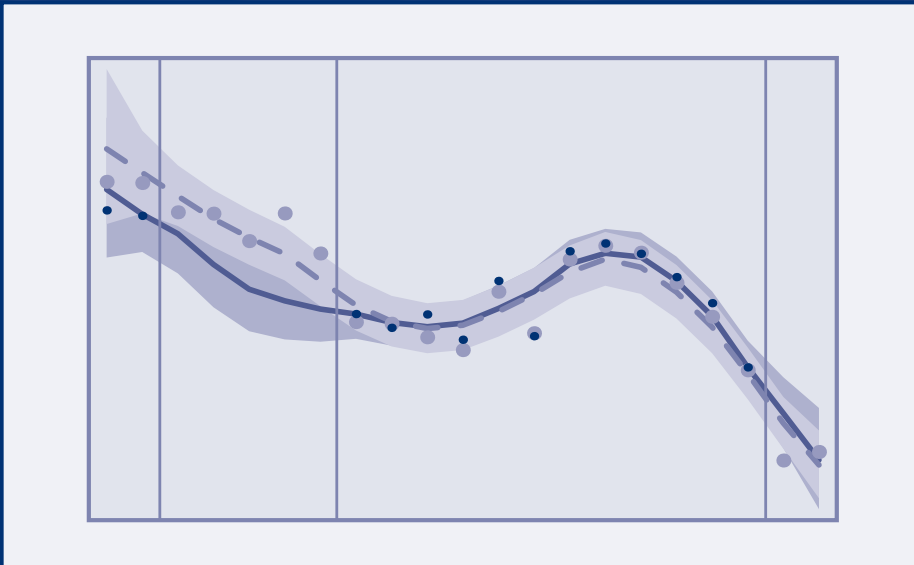


# Time series analysis in road safety research using state space methods



## Samenvatting

### Tijdreeksanalyse in verkeersveiligheidsonderzoek met behulp van state space methodologie

In dit proefschrift wordt een aantal studies gepresenteerd waarin tijdreeksanalyse wordt toegepast op geaggregeerde verkeersveiligheidsgegevens, waaronder aantallen verkeersongevallen, aantallen verkeersdoden en aantallen zwaar-gewonden.

Veel onderzoek is en wordt verricht naar hoe de verkeersveiligheid kan worden verbeterd. Daarbij wordt vaak geprobeerd verbanden te leggen tussen veranderingen in aantallen verkeersongevallen of verkeersslachtoffers aan de ene kant en aan de andere kant bijvoorbeeld factoren zoals expositie (een maat voor de hoeveelheid verkeer), beleid, het rijden onder invloed van alcohol, snelheidsgedrag en infrastructurele maatregelen. Een belangrijk doel hiervan is op te sporen welke factoren gecontroleerd kunnen worden *en* een positief effect op de verkeersveiligheid hebben, zodat langs die weg de verkeersveiligheid verbeterd kan worden.

Sommige van deze factoren, zoals regelgeving, wetgeving en beleid kunnen rechtstreeks worden waargenomen: de datum van implementatie staat vast, hoewel dat niet hoeft te gelden voor de naleving daarvan. Andere factoren kunnen alleen in theorie rechtstreeks worden waargenomen. In de praktijk zou hun rechtstreekse meting zeer moeilijk of zeer kostbaar zijn. Een voorbeeld van dergelijke factoren is de hoeveelheid reizigerskilometers. In theorie zou voor ieder individu iedere dag de hoeveelheid reizigerskilometers kunnen worden vastgesteld. In de praktijk wordt dit gegeven met behulp van enquêtes geschat, hetgeen ten koste gaat van de nauwkeurigheid. Een ander voorbeeld is het percentage bestuurders dat onder invloed van alcohol aan het verkeer deelneemt. Dit wordt geschat met behulp van steekproefsgewijze alcoholcontroles. Tenslotte is een aantal factoren nog moeilijker waar te nemen zoals de ervaring van bestuurders. Dergelijke verschillen in waarneming hebben hun weerslag op de verschillen in nauwkeurigheid van de gegevens. Als onder gebruikte gegevens (grote) verschillen in nauwkeurigheid voorkomen kan dat statistische analyse nadelig beïnvloeden als dat genegeerd wordt.

Een andere complicerende factor voor (tijdreeks)analyse in verkeersveiligheids-onderzoek is dat er geen unieke maat voor de verkeersveiligheid beschikbaar

is. Meestal wordt verkeersveiligheid gemeten in termen van het aantal ongevallen *of* het aantal slachtoffers. Hoewel in de praktijk de situatie gecompliceerder ligt, kan worden gesteld dat sommige maatregelen voor de verkeersveiligheid vooral van invloed zullen zijn op het ontstaan van ongevallen, en andere maatregelen vooral op de afloop ervan (maar zo eenvoudig ligt het niet). Wanneer een onderzoek wordt uitgevoerd naar het effect van een maatregel die het ontstaan van ongevallen zou moeten beïnvloeden, zou de ontwikkeling van het aantal van een relevant type ongevallen moeten worden bestudeerd. Aan de andere kant, als het effect van een maatregel wordt onderzocht die in hoofdzaak gevolgen zou hebben voor de ernst van ongevallen, zou de ontwikkeling van het aantal slachtoffers kunnen worden bestudeerd. De voorkeur heeft een analyse uit te voeren op zowel de ontwikkeling van het aantal slachtoffers van een type waarop de maatregel van invloed geacht wordt alsook, ter vergelijking, de ontwikkeling van het aantal slachtoffers van een type waarvan niet wordt verwacht dat de maatregel daar invloed op heeft. Behalve als gevolg van de maatregel moeten beide ontwikkelingen vergelijkbaar zijn. Indien een vermindering van het aantal slachtoffers is aangetoond is het vervolgens van belang vast te stellen dat het aantal slachtoffers is verlaagd omdat ernst van ongevallen is afgenomen, en niet omdat het aantal ongevallen is afgenomen. Daarnaast worden idealiter geen veranderingen gevonden in de aantallen van het type slachtoffer dat niet geacht wordt beïnvloed te worden door de maatregel.

Het is echter niet waarschijnlijk dat in de praktijk aan beide voorwaarden kan worden voldaan wanneer het aantal slachtoffers over een langere periode wordt onderzocht. De mogelijkheid bestaat dat in de waargenomen periode andere factoren van invloed zijn geweest op de verkeersveiligheid. Het is mogelijk dat de invloed van deze factoren op zich moeten worden gemodelleerd. Daarnaast hoeft het niet zo te zijn dat de invloed van deze factoren volledig onafhankelijk van de maatregel is. Om in dergelijke gevallen een verantwoorde statistische analyse uit te voeren kan het zinvol zijn een simultaan model te specificeren dat de gezamenlijke afhankelijke variabelen, met name ongevallen en slachtoffers omvat.

In dit proefschrift wordt een nieuwe benadering voor op verkeersveiligheid georiënteerde tijdreeksanalyse van de ontwikkelingen van geaggregeerde verkeersveiligheidsgegevens gepresenteerd, met als doel de mogelijkheden en betrouwbaarheid te verbeteren ten opzichte van gewoonlijk gebruikte alternatieven. De benadering is gebaseerd op zogenaamde multivariate heteroscedastische structurele tijdreeksmodellen en vermindert veel van de problemen van de gebruikelijke tijdreeksmodellen gebruikt voor verkeersveiligheidsanalyse,

en alle van de bovengenoemde problemen. De momenteel veel gebruikte modellen kunnen deze problemen slechts ten dele verminderen.

De gelijktijdige combinatie van drie fundamentele aspecten van de aanpak beschreven in dit proefschrift maken verbeteringen van de tijdreeks modellen in het verkeersveiligheidsonderzoek mogelijk. Deze drie aspecten zijn:

- Het gebruik van structurele componenten. De tijdreeksmodellen zijn opgebouwd met behulp van interpreteerbare structurele componenten, die in principe de expositie, het risico en eventueel de ernst representeren, en zonodig andere invloeden representeren zoals de registratiegraad, het percentage gebruik van autogordels, het percentage van de bestuurders met meer dan het wettelijke maximum voor de bloed alcohol concentratie, het snelheidsgedrag van bestuurders. De structurele componenten kunnen ook seizoensgebonden trends en patronen hebben. De voordelen van het gebruik van interpreteerbare componenten worden duidelijk als een individuele component kan worden gerelateerd aan meer dan één afhankelijke variabele. Zo zal het gordeldraagpercentage voor meerdere typen ongevallen hetzelfde kunnen zijn (geldt niet altijd). Bovendien stellen de structurele componenten in deze modellen de onderzoeker in staat onderscheid te maken tussen effecten die de verkeersveiligheid of belangrijke onderdelen daarvan beïnvloeden en effecten die hoe we de verkeersveiligheid waarnemen beïnvloeden. Hierbij kan worden gedacht aan de afnemende volledigheid van de registratie van ongevallen: ze gebeuren wel, maar we zien ze niet meer. Daarnaast stelt deze benadering de onderzoeker op een relatief transparante manier in staat effecten van verklarende variabelen op de relevante componenten te bestuderen in plaats van op het aantal ongevallen of slachtoffers: bijvoorbeeld ten behoeve van onderzoeksvragen als “heeft politie toezicht effect op het percentage autogordelgebruik, en daarmee effect op de verkeersveiligheid of zit het toch anders?”
- Meerdere afhankelijke variabelen: zowel aantallen ongevallen als aantallen slachtoffers kunnen worden opgenomen in een model, zonodig apart voor meerdere typen ongevallen. Verklarende variabelen kunnen op de traditionele manier worden opgenomen. Verklarende variabelen die met onzekerheid gemeten zijn, zoals bijvoorbeeld gegevens verkregen uit steekproeven of resultaten van ander onderzoek, kunnen in een model worden opgenomen waarbij een structurele component gebruikt wordt als schatting van de werkelijke waarde (de waarde zonder observatiefout, hetgeen zinvol kan zijn als verondersteld kan worden dat de

werkelijke waarde invloed heeft op de verkeersveiligheid of andere componenten, en niet de geobserveerde waarde). Een voorbeeld is het percentage autogordelgebruik. Gegevens hiervoor kunnen zijn verkregen uit relatief kleine onderzoeken, en niet noodzakelijkerwijs beschikbaar voor alle tijdstippen. Als kan worden verondersteld dat het percentage gebruik van de autogordel ongeveer constant is, kan een structurele component worden gebruikt die het gemiddelde van alle waarnemingen is. Deze structurele component kan vervolgens worden gebruikt voor alle tijdstippen, inclusief tijdstippen waarvoor geen eigen onderzoeksresultaat beschikbaar is. Als *niet* kan worden verondersteld dat het percentage gebruik van de autogordel constant is, kan de structurele component daarop aangepast worden, zodat nog steeds de observaties van alle tijdstippen bruikbaar zijn (uiteeraard wordt de onzekerheid over het percentage autogordelgebruik op een bepaald tijdstip dan (veel) groter).

- Niet identieke structuur van waarnemingsfouten: de nauwkeurigheid kan variëren in de tijd. Een voorbeeld is de OVG/MON mobiliteitsenquête, die behoorlijk in omvang is veranderd. Een extreem voorbeeld is het geheel ontbreken van gegevens. Niet alle enquêtes of onderzoeken worden ieder jaar uitgevoerd.

De modellen gebaseerd op de aanpak beschreven in dit proefschrift kunnen omstandigheden hanteren waarin variabelen verschillen in nauwkeurigheid. Bijvoorbeeld, de hoeveelheid verkeer in auto's kan nauwkeuriger worden bepaald dan de hoeveelheid verkeer op motorfietsen (de relatieve fout voor de totale hoeveelheid verkeer in auto's is ongeveer 2 tot 3% in de mobiliteitsenquête van 2003, terwijl de relatieve fout voor het totale verkeer van motorfietsen bijna 30% is in deze mobiliteitsenquête van 2003). Het is de combinatie van deze eigenschappen in een model en het gebruik van gedeelde structurele componenten dat het model aantrekkelijk maakt voor tijdreeksanalyse in verkeersveiligheidsonderzoek. Gezien het feit dat in verkeersveiligheidsonderzoek over het algemeen een langere periode wordt bestudeerd, waar de omstandigheden kunnen veranderen in de loop van de tijd, kunnen de volgende praktische voordelen worden genoemd:

- Omdat verschillende afhankelijke variabelen tegelijk in samenhang kunnen worden geanalyseerd en het feit dat geen identieke structuur van waarnemingsfouten wordt geëist, kunnen de modellen worden gebruikt om rekening te houden met covariantie tussen de afhankelijke variabelen. Het is bekend dat het aantal slachtoffers in een periode (mede) af-

hankelijk is van het aantal ongevallen. Het is niet per se opvallend dat een jaar met meer ongevallen ook meer slachtoffers heeft.

Het zou wel opvallend zijn als een jaar meer ongevallen heeft *en* minder slachtoffers, of minder ongevallen en meer slachtoffers. Terwijl de absolute verschillen even groot zijn (een zelfde aantal meer of minder ongevallen of slachtoffers), is een jaar met meer ongevallen en meer slachtoffers of minder ongevallen en minder slachtoffers minder onwaarschijnlijk dan een jaar met meer ongevallen en minder slachtoffers, of minder ongevallen en meer slachtoffers.

Met andere woorden: als er in een jaar meer ongevallen zijn gebeurd, hoeft dat niet opvallend te zijn. Als er in een jaar minder slachtoffers zijn gevallen hoeft dat ook niet opvallend te zijn. Maar als in datzelfde jaar zich meer ongevallen hebben voorgedaan terwijl er minder slachtoffers zijn gevallen, kan dat best opvallend zijn.

Met deze afhankelijkheid moet rekening worden gehouden in statistische toepassingen. Ook de fout in de hoeveelheid verkeer geschat op basis van de enquêtes kan samenhangen. De samenhang tussen aantallen ongevallen en slachtoffers is gebruikt in beide toepassingen van hoofdstuk 3. De fout in de hoeveelheid verkeer van de enquêtes wordt eveneens gebruikt in de toepassingen van hoofdstuk 3, de Nederlandse toepassing in hoofdstuk 5 en de aan verkeersveiligheid gerelateerde toepassing in hoofdstuk 7.

- Het gebruik van structurele componenten stelt de onderzoeker in staat een beperkte mate van verificatie van de resultaten uit te voeren. De ontwikkeling van structurele componenten over de tijd kan worden vergeleken met secundaire informatie, hoewel als veel secundaire informatie beschikbaar is, het waarschijnlijk beter is om die informatie in het model op te nemen. Een succesvolle toepassing van zo'n verificatie is te vinden in hoofdstuk 6, waar de vorm van de ontwikkeling van de hoeveelheid gemotoriseerd verkeer buiten de bebouwde kom vergeleken is met een schatting daarvan op basis van de lengte van het wegennet en de verkeersintensiteit.
- Structurele componenten kunnen gemeenschappelijk worden gebruikt voor de specificatie van vele afhankelijke variabelen. In de verkeersveiligheidstoepassing van hoofdstuk 7 worden weersomstandigheden (duur van de neerslag) gemeten aan de hand van tien weerstations, gerelateerd aan één structurele component die de gemiddelde duur van de neerslag in Nederland voorstelt. Als het weer waargenomen door de stations onderling overeenkomt, zal de waarde van de structurele component nauw-

keurig bekend zijn. Aan de andere kant, als de weerstations een verschillend weerbeeld geven, zal de waarde van de structurele component minder nauwkeurig bekend zijn. Op deze wijze kunnen observaties waarbij de gemiddelde hoeveelheid neerslag minder goed bekend is, minder invloed hebben op het eindresultaat dan observaties waarbij de gemiddelde hoeveelheid neerslag wel goed bekend is, zodat eventuele fouten hopelijk minder gevolgen hebben. De mogelijkheid componenten te delen wordt ook gebruikt in hoofdstuk 3, waarin een structurele component die het aantal slachtoffers per ongeval voorstelt wordt gedeeld door de afhankelijke variabelen die “het aantal door de politie geregistreerde slachtoffers” en “een schatting van het werkelijke aantal slachtoffers” voorstellen. Deze component wordt gebruikt voor het verbeteren van de schatting van het werkelijke aantal ongevallen, die niet kan worden afgeleid uit ziekenhuisgegevens.

- De aanpak maakt het mogelijk andere verdelingen dan de normale voor de waarnemingsfouten in de modellen te gebruiken. Deze verdelingen kunnen bovendien gecombineerd worden en per tijdstip verschillen. In de verkeersveiligheidstoepassing van hoofdstuk 7 wordt een model voor weer en verkeersveiligheid ontwikkeld met behulp van zowel meerdere normaal verdeelde waarnemingsfouten als Poisson verdeelde afhankelijke variabelen.
- De aanpak maakt het mogelijk verbanden te leggen tussen de zogenaamde innovaties van structurele componenten. Innovaties zijn schokken in de ontwikkeling van de componenten. Wanneer de innovaties van twee componenten gecorreleerd zijn, dan hebben de ontwikkelingen iets gemeenschappelijks. Dit is van belang wanneer structurele componenten verschijnselen beschrijven die elkaar kunnen beïnvloeden, zoals expositie en risico. Een ander voorbeeld is de ontwikkeling van de ernst van ongevallen en ongevalsrisico. De ontwikkeling van de ernst van ongevallen is van invloed op het ontstaan van ongevallen die een bepaalde ernst overstijgen. Als we in staat zijn de gemiddelde ongevals ernst zoveel te verminderen dat er bijna geen doden in het verkeer meer te betreuren zijn, dan zullen er ook bijna geen dodelijke ongevallen meer gebeuren. Een zelfde effect doet zich voor ten gevolge van veranderingen in de bezettingsgraad van auto's. Minder inzittenden betekent minder kans dat er iemand zo zwaar gewond raakt dat hij of zij in het ziekenhuis moet worden opgenomen of zelfs omkomt. Gemiddeld genomen zal het lijken alsof er dus minder ongevallen gebeuren die zo erg zijn dat er ziekenhuisgewonden of doden vallen.

In de praktijk worden altijd min of meer ernstige ongevallen bestudeerd. In dit proefschrift zijn dat dodelijke ongevallen en ongevallen met doden en of ziekenhuisgewonden. Praktisch betekent dit dat valt te verwachten dat met een afname van de gemiddelde ongevalsernst, ook het aantal ongevallen zal afnemen.

Het feit dat de expositie van invloed kan zijn op risico's wordt gebruikt in bijna alle toepassingen in dit proefschrift, terwijl het feit dat de ernst van ongevallen van invloed kan zijn op het ontstaan van ernstige ongevallen wordt gebruikt in de eerste toepassing van hoofdstuk 3.

In de verkeersveiligheidstoepassingen in dit proefschrift is de expositie altijd in één op één verhouding verondersteld met de omvang van het verkeer, gemeten in voertuigkilometers. Deze beperking is echter niet fundamenteel. De aanpak kan ook worden toegepast op modellen die van een niet-lineaire relatie uitgaan tussen expositie en voertuigkilometers, alsmede van een andere maat voor het verkeersvolume dan voertuigkilometers. Bovendien zijn de dynamische relaties gebruikt in dit proefschrift afgeleid van lokale lineaire trendmodellen. Hoewel geleidelijke ontwikkelingen over het algemeen op bevredigende wijze benaderd lijken te kunnen worden door lokale lineaire trendmodellen, hoeft dit niet altijd het geval te zijn. Met name het model voor de ontwikkeling van de fractie van de hoeveelheid van het verkeer met neerslag in hoofdstuk 7 zou kunnen worden verbeterd.

Voor de beschrijving van ontwikkelingen over de tijd is bij de modellen in dit proefschrift in alle gevallen uitgegaan uit van lineaire dynamische relaties met normaal verdeelde toevalsfluctuaties. Dit is een aanname die niet in alle gevallen hoeft op te gaan, hoewel de benadering met behulp van lokaal lineaire trendmodellen in de praktijk goed lijkt te functioneren. De effectiviteit ervan wordt in dit proefschrift empirisch aangetoond, alsmede van de voorgestelde nieuwe methode van tijdreeksanalyse voor verkeersveiligheidsonderzoek. Het is de bedoeling de ontwikkeling van de methodiek verder voort te zetten in hogere dimensies en in meer gedetailleerde modellen voor de verkeersveiligheid. In dit proefschrift zijn de belangrijkste bijdragen van de nieuwe aanpak gerapporteerd.

