

Spijtminimalisatie: een nieuwe modelvorm en een toepassing op vertrektijdstipkeuze-data

Caspar G. Chorus*, Gerard C. de Jong**

** Sectie Transport en Logistiek, TU Delft, Postbus 5015, 2600 GA, Delft*

*** ITS, University of Leeds, Leeds, LS2 9JT, UK; Significance, Koninginnegracht 23, 2514 AB Den Haag, Nederland; Centre for Transport Studies, VTI/KTH, Postbus 55685, SR-102 15, Stockholm, Zweden; NEA, Postbus 276, 2700 AG Zoetermeer, Nederland*

Summary

This paper presents a new discrete choice-model, rooted in the paradigm of Random Regret Minimization (RRM). In contrast with earlier RRM-models, this new model formulation is estimable using conventional discrete choice-software. It is shown how the new RRM-based Multinomial Logit (MNL) model circumvents the IIA-property associated with its Random Utility Maximization (RUM)-based counterpart. An empirical application based on a stated departure time-choice dataset is provided, which shows how the new RRM-based (Mixed) MNL models perform well – also when compared to RUM-based (Mixed) MNL models.

Keywords: Random Regret Minimization; Mixed Logit; IIA-property; Departure time choice

Samenvatting

Dit artikel presenteert een nieuw discrete-keuzemodel, geworteld in het 'Random Regret Minimization'-paradigma (RRM). In tegenstelling tot eerdere RRM-modellen kan de nieuwe modelvorm geschat worden met behulp van standaard softwarepakketten. We laten zien hoe het spijtgebaseerde Multinomiale Logit (MNL)-model de zogenaamde IIA-eigenschap van nutsgebaseerde MNL-modellen omzeilt. Vervolgens presenteren we een empirische toepassing op een 'Stated Choice'-dataset betreffende vertrektijdstipkeuzes. Schattingsresultaten tonen goede prestaties van het RRM-gebaseerde (Mixed) MNL-model, ook wanneer het vergeleken wordt met zijn nutsgebaseerde tegenhangers.

Keywords: Random Regret Minimization; Mixed Logit; IIA-eigenschap; Vertrektijdstipkeuze

1. Inleiding

De minimalisatie van geanticipeerde spijt is met succes toegepast ter verklaring van keuzegedrag, in vakgebieden zo uiteenlopend als marketing (Simonson, 1992; Zeelenberg & Pieters, 2009), microeconomie (Loomes & Sugden, 1982; Sarver, 2008), psychologie (Zeelenberg, 1999; Connolly, 2005), besliskunde (Savage, 1954; Bell, 1982) en ook mobiliteitsonderzoek (Chorus et al., 2006a, b). Kort gezegd is spijt wat men voelt wanneer een niet-gekozen alternatief beter presteert dan een gekozen alternatief, en spijtgebaseerde keuzetheorieën en –modellen veronderstellen dat de wens om geanticipeerde spijt te minimaliseren (naast, of in plaats van, de wens om geanticipeerd nut te maximaliseren) centraal staat wanneer mensen keuzes maken. Recentelijk is de notie van spijtminimalisatie vertaald in een generiek discrete-keuzemodel: dit zogenaamde *Random Regret Minimization* (RRM)-model (Chorus et al., 2008a, b) is ontwikkeld voor de econometrische analyse van geobserveerde keuzes en vormt een spijtgebaseerde tegenhanger van de Multinomiale Logit (MNL) modellen (McFadden, 1974) de Mixed MNL modellen (McFadden & Train, 2000), die beiden gebaseerd zijn op *Random Utility Maximization* (RUM)-veronderstellingen.

Een belangrijk nadeel van de eerste generatie RRM-modellen is dat hun modelstructuur een discontinue Likelihoodfunctie genereert, wat problemen kan opleveren bij het optimalisatieproces dat onderdeel uitmaakt van de modelschatting¹. Mede hierdoor kunnen deze RRM-modellen niet geschat worden met behulp van standaard softwarepakketten voor discrete-keuzemodellen (in plaats daarvan is handgeschreven code nodig), wat hun gebruiksvriendelijkheid fors vermindert.

Onlangs is een nieuwe RRM-modelvariant afgeleid (Chorus, 2010) die deze beperking niet kent: de bijbehorende Likelihoodfunctie is continu, en het model kan mede hierdoor geschat worden op bestaande softwarepakketten zoals BIOGEME, ALOGIT, NLOGIT. In dit paper presenteren we dit model (sectie 2), en bespreken we hoe het verschilt van zijn nutsgebaseerde tegenhanger (sectie 3). In sectie 4 presenteren we een empirische modelschatting (en een vergelijking tussen RRM- en RUM-modellen), gebruik makend van een dataset betreffende vertrektijdstipkeuzes van automobilisten (de Jong et al., 2003). De schatting van RRM-modellen op deze data is nog niet eerder gepubliceerd. Sectie 5 presenteert conclusies.

2. Een nieuw RRM-model

We veronderstellen de volgende keuzesituatie: een reiziger kiest tussen J reisopties (zoals routes, vertrektijdstippen), die elk gedefinieerd zijn in termen van M attributen x_m . We richten ons in de rest van deze paragraaf op het voorspellen van de kans dat de reiziger alternatief i kiest uit deze keuzeset. Het nieuwe RRM-model veronderstelt dat de reiziger probeert geanticipeerde random spijt te minimaliseren, en dat –vanuit het gezichtspunt van de onderzoeker– de random spijt die geassocieerd

¹ De afgeleide van de Likelihoodfunctie, die berekend moet worden als onderdeel van het optimalisatieproces, bestaat immers niet op die punten waar de functie discontinue is.

wordt met alternatief i bestaat uit de som van een iid storingsterm ε_i en 'systematische spijt' R_i . De afkorting 'iid' staat voor '*independent and identically distributed*'; dit houdt in dat de waarde van de storingsterm die bij het ene alternatief hoort, onafhankelijk is van de waarde van de storingsterm die bij elk van de andere alternatieven hoort, en dat alle storingstermen uit dezelfde verdeling getrokken zijn. Meer concreet impliceert deze iid-veronderstelling dat de storingstermen zogenaamde 'witte ruis' vormen. In lijn met McFadden's nutsgebaseerde MNL-model, veronderstellen we dat ε_i een zogenaamde 'Extreme Value Type I' verdeling volgt. Deze verdeling lijkt op de normaalverdeling, maar heeft dikkere 'staarten'.

Systematische spijt is gedefinieerd als de som van de zogenaamde bilaterale spijt die geassocieerd wordt met de vergelijking tussen i en elke andere keuzeoptie: $R_i = \sum_{j \neq i} R_{i \leftrightarrow j}$. Als

zodanig veronderstelt dit spijtmodel dat elk alternatief dat beter presteert dan i een bijdrage levert aan de spijt die de reiziger anticipeert en wil vermijden. De bilaterale spijt die geassocieerd wordt met de vergelijking tussen i en alternatief j bestaat uit de som van de spijt die geassocieerd wordt met de vergelijking tussen de twee keuzeopties in termen van elk attribuut: $R_{i \leftrightarrow j} = \sum_{m=1..M} R_{i \leftrightarrow j}^m$. Tenslotte, de

spijt die geassocieerd wordt met de vergelijking tussen i en j in termen van attribuut m (in formulevorm: $R_{i \leftrightarrow j}^m$) is ofwel gelijk aan 0 (dit is het geval wanneer i beter presteert dan j in termen van m), ofwel gelijk aan het gewogen verschil in attribuutwaarde. Het gewicht β_m is een parameter (waarvan grootte en teken geschat worden) en representeert het belang van attribuut m voor de reiziger (en of hogere waarden van m worden geprefereerd over lagere of andersom). Twee iid storingstermen ν_{0m}, ν_{xm} (weer met 'Extreme Value Type I'-verdeling) worden toegevoegd, vanuit de gedachte dat de onderzoeker niet in staat is om bilaterale spijt op attribuutniveau foutloos te bepalen:

$$R_{i \leftrightarrow j}^m = \max \left\{ 0 + \nu_{0m}, \beta_m \cdot (x_{jm} - x_{im}) + \nu_{xm} \right\}.$$

Keuzekansen voor elk van de J keuzeopties worden nu verkregen door te integreren over de kansdichtheidsfuncties van alle storingstermen (wiens waarden we immers niet kennen). Wanneer we beginnen bij storingstermen ν , kan gebruik worden gemaakt van het feit (Ben-Akiva & Lerman, 1985) dat het verwachte maximum van twee termen die elk bestaan uit de som van een systematisch gedeelte en een iid-Extreme Value Type I storingsterm, de vorm aanneemt van een zogenaamde Logsum. Dit resulteert in de volgende formulering van de verwachte bilaterale spijt betreffende alternatieven i en j in termen van attribuut m :

$$\int_{\nu} \left[R_{i \leftrightarrow j}^m \cdot f(\nu) \right] d\nu = \ln \left(1 + \exp \left[\beta_m \cdot (x_{jm} - x_{im}) \right] \right).$$

De resulterende systematische spijt die geassocieerd wordt met alternatief i kan nu worden geformuleerd als $\tilde{R}_i = \sum_{j \neq i} \sum_{m=1..M} \ln \left(1 + \exp \left[\beta_m \cdot (x_{jm} - x_{im}) \right] \right)$, waar de ' \sim ' staat voor het feit dat

storingstermen ν zijn 'uitgeïntegreerd'. De kans dat de reiziger keuzeoptie i kiest, is nu gelijk aan de

kans dat de random spijt die geassocieerd wordt met i kleiner is dan de random spijt van de andere alternatieven. Uit het feit dat de ε -storingstermen verondersteld worden de iid-Extreme Value Type I verdeling te volgen, volgt dat deze kans gegeven wordt door een zogenaamde logit-formulering:

$$P_i = \exp(-\tilde{R}_i) / \sum_{j=1..J} \exp(-\tilde{R}_j). \text{ Indien nodig, kunnen constanten worden toegevoegd voorafgaand}$$

aan de systematische spijt-term.

3. Een theoretische vergelijking met het nutsgebaseerde MNL-model

De lezer die bekend is met het nutsgebaseerde MNL-model ziet direct de overeenkomst tussen de keuzekans-formule voor het RRM-model en die van het nutsgebaseerde MNL-model: deze laatste

wordt gegeven door de formule $P_i = \exp(V_i) / \sum_{j=1..J} \exp(V_j)$, waar $V_i = \sum_{m=1..M} \beta_m x_{im}$. Naast de

gelijkenis in termen van keuzekansen, valt op dat elke parameter in het nutsgebaseerde MNL-model een tegenhanger heeft in het spijtgebaseerde MNL-model, en dat alle parameters hetzelfde teken (zouden moeten) hebben in RUM- en RRM-modellen. In het licht van deze gelijkenissen, is het interessant om een cruciaal verschil tussen de twee modellen te belichten².

Dit verschil zit hem in de wijze waarop de twee modellen veronderstellen dat de keuzekansratio van twee alternatieven beïnvloed wordt door de aanwezigheid en kenmerken van de overige alternatieven in de keuzeset. Het nutsgebaseerde MNL-model veronderstelt dat deze ratio, bijvoorbeeld die van de keuzekansen voor i en j , *niet* afhangt van de aanwezigheid en kenmerken van de overige alternatieven in de keuzeset. Deze zogenaamde IIA-eigenschap ('Independence from Irrelevant Alternatives') volgt uit het feit dat de noemers van de keuzekansen die voor verschillende keuzeopties geproduceerd worden door nutsgebaseerde MNL-modellen, wegvallen wanneer men keuzekansen door elkaar deelt. Doorredenerend zijn er twee onderliggende oorzaken aan te wijzen voor het optreden van deze eigenschap, die gedragsmatig zeer onrealistische substitutiepatronen kan produceren.

Een eerste oorzaak is de veronderstelde iid-verdeling van storingstermen (die impliceert dat storingstermen enkel zogenaamde 'witte ruis' representeren). Deze veronderstelling kan worden opgeheven door zogenaamde Nested Logit of Mixed Logit-modellen te formuleren, die in staat zijn correlaties tussen de storingstermen te creëren. Een tweede oorzaak is dat de systematische nutten in conventionele (zogenaamde 'linear-additive') MNL-modellen enkel afhangen van de prestaties van de desbetreffende keuzeoptie; de prestaties van alternatieven hebben geen effect op het nut dat wordt toegekend aan een keuzeoptie (ze hebben uiteraard wel effect op de resulterende *keuzekansen* voor die keuzeoptie). Deze tweede oorzaak wordt niet verholpen door het gebruik van Nested Logit of Mixed Logit modellen, die de IIA-eigenschap, en de in sommige gevallen resulterende onrealistische substitutiepatronen, daardoor maar ten dele kunnen elimineren.

² Zie Chorus (2010) voor een uitgebreidere studie van de verschillen tussen de RRM- en RUM- benaderingen.

Dat het spijtgebaseerde MNL-model geen IIA veronderstelt, is direct te zien door de simpelweg de RRM-keuzekansen van keuzeopties i en j door elkaar te delen (in de context van een keuzeset die bestaat uit opties i, j en k):

$$\frac{P_i}{P_j} = \frac{\exp(-\tilde{R}_i) / (\exp(-\tilde{R}_i) + \exp(-\tilde{R}_j) + \exp(-\tilde{R}_k))}{\exp(-\tilde{R}_j) / (\exp(-\tilde{R}_i) + \exp(-\tilde{R}_j) + \exp(-\tilde{R}_k))} = \frac{\exp(-\tilde{R}_i)}{\exp(-\tilde{R}_j)} =$$

$$\frac{\exp\left(-\sum_{m=1..M} \ln\left(1 + \exp\left[\beta_m \cdot (x_{jm} - x_{im})\right]\right)\right) - \sum_{m=1..M} \ln\left(1 + \exp\left[\beta_m \cdot (x_{km} - x_{im})\right]\right)}{\exp\left(-\sum_{m=1..M} \ln\left(1 + \exp\left[\beta_m \cdot (x_{im} - x_{jm})\right]\right)\right) - \sum_{m=1..M} \ln\left(1 + \exp\left[\beta_m \cdot (x_{km} - x_{jm})\right]\right)}$$

Het is duidelijk dat de attributen van alternatief k (weergegeven als x_{km}) opduiken in deze ratio. Hierdoor is k in het spijtgebaseerde MNL-model geen 'Irrelevant Alternative', zoals het dat wel zou zijn in een nutsgebaseerd MNL-model. In andere woorden: het spijtgebaseerde MNL-model – ook al veronderstelt het dezelfde 'witte ruis'-verdeling voor zijn storingstermen als zijn nutsgebaseerde tegenhanger – voorspelt dat de relatieve populariteit van twee keuzeopties een functie is van hun kenmerken, én van de aanwezigheid van alternatieven (en hun kenmerken). Als zodanig vertoont het spijtgebaseerde MNL-model de IIA-eigenschap niet, ook al worden storingstermen verondersteld iid te zijn. Door hiernaast ook correlaties te veronderstellen tussen storingstermen (en op die manier spijtgebaseerde Mixed Logit-modellen te creëren), kan ook de tweede onderliggende oorzaak van de IIA-eigenschap worden geëlimineerd.

Concluderend kan gesteld worden dat het RRM-model, ongeacht de veronderstelde verdeling van storingstermen en in tegenstelling tot zijn nutsgebaseerde tegenhanger, veronderstelt dat de relatieve populariteit van elke twee alternatieven in een keuzeset afhangt van de samenstelling van die keuzeset in zijn geheel. Als zodanig is de RRM-benadering consistent met een stroom aan literatuur in de marketingwetenschappen (zie bij voorbeeld Simonson (1989), Wernerfelt (1995), Kivetz et al. (2004), Steenburgh (2008) en de referenties in deze artikelen).

4. Het nieuwe RRM-model toegepast: vertrektijdstipkeuzes

Deze sectie presenteert een empirische toepassing van het in de vorige sectie gepresenteerde RRM-model. We doen dit in de vorm van een vergelijking met RUM-modellen. Echter, het doel van de empirische analyses die hier gepresenteerd worden is niet om de twee modelparadigma's te vergelijken, of om verschillen in modelfit en schattingsresultaten te interpreteren in termen van de verschillende veronderstellingen over keuzegedrag die aan beide modellen ten grondslag liggen. Hiervoor is een veel uitgebreidere analyse nodig, die meerdere databronnen omvat. Ook heeft deze

sectie niet tot doel om inhoudelijke inzichten te genereren aangaande de determinanten van vertrektijdstopkeuzes. Zie hiervoor een zeer uitgebreid verslag van empirische analyses, uitgevoerd op dezelfde data, zoals gerapporteerd in de Jong et al. (2003). We willen in deze sectie enkel aantonen dat het RRM-model succesvol kan worden toegepast op empirische keuzedata, en dat toepassing van het model resulteert in intuïtieve schattingsresultaten.

De dataset is gebaseerd op een zogenaamd Stated Choice-experiment dat werd uitgevoerd voor het Ministerie van Verkeer en Waterstaat. Hierin werd forenzen die de auto gebruiken voor hun woon-werkverkeer³ gevraagd te kiezen tussen een aantal vertrektijdstoppen betreffende de rit van en naar werk: men kon ofwel op het voor hen reguliere tijdstip kiezen, ofwel substantieel eerder of later (om de files te vermijden). Elk van de drie opties is gedefinieerd in termen van onder andere de reistijd heen, de reistijd terug, de reistijd in de file (heen, en terug), en de tijd die overbleef om te werken. In het licht van de functie van de hier gepresenteerde empirische analyses (zoals hierboven uiteengezet), gaan we hier niet verder in op de opzet van het experiment, en op zaken als de representativiteit van de steekproef. Zie de Jong et al. (2003) voor een gedetailleerde bespreking van deze en gerelateerde aspecten. Voor onze analyses gebruiken we een deel van de beschikbare data: het betreft automobilisten, die hadden aangegeven dat gebruik maken van het OV voor hen geen optie is. Met andere woorden: hun keuzeset bestond enkel uit de drie gepresenteerde vertrektijdstopkeuze-opties. Deze sub-dataset bestond uit 883 waarnemingen (keuzes). Alle hierna beschreven modellen zijn geschat met behulp van de gratis discrete-keuze software BIOGEME (Bierlaire, 2003, 2008).

4.1. MNL-modellen

Tabel 1 toont de schattingsresultaten voor zowel de nuts- als spijtgebaseerde MNL-modellen. Om overschatting van de t -waarden, veroorzaakt door de panel-structuur van de data (voor elke respondent zijn 8 keuzes geobserveerd), te voorkomen zijn enkel de zogenaamde robuuste t -waarden gerapporteerd.

Beginnend bij de modelstatistieken: beide modellen bereiken een redelijke modelfit, wat gereflecteerd wordt door een rho-square van tegen de 0.18. De Loglikelihood van het RRM-model is marginaal groter dan die van zijn nutsgebaseerde tegenhanger, wat aangeeft dat het spijtminimalisatiemodel iets beter past bij de data. Het verschil is echter geenszins substantieel. In termen van parameters geven de beide modellen ook op hoofdlijnen dezelfde inzichten (NB: gegeven dat het RRM-model parameters schat voor verschillen in attribuutwaardes, terwijl het RUM-model parameters schat voor absolute attribuutwaarde, is de verwachting niet dat beide parametersets *in absolute zin* op elkaar lijken). Het blijkt dat, zoals men zou verwachten, er *ceteris paribus* een voorkeur is voor vertrekken op het reguliere tijdstip (in plaats van vroeger of later). Een model met twee constanten (voor vroeger, en voor later vertrekken) gaf voor elk van de modellen geen vergroting van de modelfit, en is niet gerapporteerd. De parameters voor reistijd zijn niet significant (zowel voor

³ In hetzelfde project zijn ook data verzameld voor andere reismotieven en voor treinreizigers; die worden in deze paper niet gebruikt.

de heen- als voor de terugweg), wat ligt aan het feit dat de bijbehorende attributen weinig variatie kenden in het keuze-experiment. Ook de parameter voor resulterende werktijd is voor geen van de modellen significant. Wel significant zijn de negatieve parameters voor reistijd in de file; beide modellen geven aan dat reistijd in de file op de weg naar werk vervelender wordt gevonden dan op de terugweg.

Tabel 1: RRM-MNL versus RUM-MNL

Attribuut	RRM		RUM	
	beta	t-waarde	beta	t-waarde
Constante (regulier vertrektijdstip)	1.48	16.11	1.47	16.02
Reistijd heen (minuten)	-0.00376	-1.27	-0.00579	-1.48
Filetijd heen (minuten)	-0.0264	-4.62	-0.0377	-4.67
Reistijd terug (minuten)	-0.00206	-0.37	-0.00318	-0.36
Filetijd terug (minuten)	-0.0168	-2.35	-0.0258	-2.20
Resulterende werktijd (minuten)	0.000366	0.36	0.000624	0.41
Eind-Loglikelihood	-796.306		-797.435	
Aangepaste rho-square ⁴	0.173		0.172	

4.2. *Mixed MNL-modellen met panelstructuur*

Zoals hierboven aangegeven, is in het schatten van de MNL-modellen deels rekening gehouden met de panelstructuur van de data (het feit dat elke respondent meerdere observaties gegenereerd heeft), door middel van het gebruik van robuuste *t*-waarden. Een aanvullende en meer directe manier om met de panelstructuur om te gaan, is om te veronderstellen dat een deel van de ruis (die gemodelleerd wordt door middel van de storingstermen) niet varieert tussen verschillende waarnemingen voor dezelfde respondent. Deze aanvullende veronderstelling impliceert dat het individueel-specifieke deel van de ruis wordt gemodelleerd door één of meerdere normaalverdeelde additionele storingstermen, waarvan de standaarddeviaties geschat worden; deze representeren de mate van variatie in de smaak van respondenten betreffende de bij de parameters horende alternatieven en hun attributen. Deze smaak varieert binnen de populatie van reizigers, maar niet tussen de keuzes die een bepaalde reiziger maakt. Het resultaat van een dergelijke uitbreiding is een zogenaamd panel-data Mixed Logit-model (zie Train (2003) voor een uitgebreide bespreking van dit modeltype).

Specifiek hebben we er in dit geval voor gekozen om te veronderstellen dat de intrinsieke waardering van reizigers voor het vertrekken op het reguliere tijdstip (in plaats van uit te wijken naar

⁴ Deze is gecorrigeerd voor het aantal geschatte parameters.

vroegere of latere tijdstippen), varieert binnen de reizigerspopulatie, maar niet binnen de 8 keuzes die elke reiziger maakt in het experiment. Met andere woorden, we schatten de standaarddeviatie voor de constante uit het MNL-model⁵. De significantie en grootte van deze standaarddeviatie is een indicatie van de mate waarin reizigers verschillen in hun voorkeur voor het op de reguliere tijd vertrekken. Omdat we slechts 1 van de parameters een individueel-specifieke component meegeven, blijven we overigens robuuste *t*-waarden rapporteren. Het resultaat is te zien in tabel 2: de verschillen met het geschatte MNL-model in termen van parameterwaarden en –significanties zijn klein, en dezelfde inhoudelijke inzichten komen bovendien. De significante en substantiële standaarddeviatie voor de constante geeft aan dat er substantiële heterogeniteit is in de reizigerspopulatie betreffende de voorkeur voor het aanhouden van hun reguliere vertrektijd: voor de ene reiziger is dit een zeer belangrijk punt, voor de andere maakt het weinig uit.

Tabel 2: RRM-Panel Mixed MNL versus RUM-Panel Mixed MNL

Attribuut	RRM		RUM	
	beta	t-waarde	beta	t-waarde
Constante (regulier vertrektijdstip)	1.79	8.19	1.77	8.14
Reistijd heen (minuten)	-0.00604	-1.44	-0.00846	-1.59
Filetijd heen (minuten)	-0.0244	-2.48	-0.0336	-2.52
Reistijd terug (minuten)	-0.00566	-0.82	-0.00799	-0.76
Filetijd terug (minuten)	-0.0185	-1.68	-0.0291	-1.66
Resulterende werktijd (minuten)	0.00109	0.03	0.000171	0.06
Standaarddeviatie (constante)	-1.87	-7.98	-1.87	-7.95
Eind-Loglikelihood	-712.074		-713.081	
Aangepaste rho-square	0.259		0.258	

In termen van modelfit valt op dat het modelleren van deze ongeobserveerde heterogeniteit een zeer aanzienlijke verbetering in fit met zich meebrengt. Met andere woorden: een model dat deze heterogeniteit serieus neemt, past de data veel beter dan een MNL-model dat veronderstelt dat alle storingstermen enkel witte ruis representeren (en dat de voorkeur voor het reguliere vertrektijdstip niet structureel varieert binnen de reizigerspopulatie). Tot slot: ook hier weer genereert het RRM-model een wat kleinere Loglikelihood-waarde dan zijn nutsgebaseerde tegenhanger: spijtminimalisatie blijkt de gemaakte keuzes iets beter te kunnen verklaren dan nutsmaximalisatie.

Een belangrijkere conclusie – in het licht van de doelstelling zoals aangegeven bovenaan deze sectie – is dat het RRM-model blijkbaar zonder problemen is toe te passen met behulp van

⁵ Omdat deze door middel van simulatie wordt geschat, kan de geschatte standaarddeviatie negatief zijn. Dit min-teken heeft geen betekenis, en de standaarddeviatie dient uiteraard als positief te worden geïnterpreteerd.

standaard-softwarepakketten, zowel in zijn MNL- als Mixed-MNL-vorm, en resulteert in intuïtieve modeluitkomsten.

5. Conclusies

Dit artikel presenteert een nieuw discrete keuzemodel, geënt op het recentelijk geïntroduceerde paradigma van *Random Regret Minimization*. Het model is gebouwd op de veronderstelling, dat menselijk keuzegedrag gebaseerd is op de wens om spijt achteraf te voorkomen. Het ontwikkelde model is even spaarzaam met parameters als conventionele nutsgebaseerde keuzemodellen, het heeft een gesloten vorm-formulering voor keuzekansen en kan met behulp van bestaande software geschat worden. Een belangrijk verschil met conventionele nutsgebaseerde modellen zoals het lineair-additieve MNL-model, is dat het RRM-model veronderstelt dat de relatieve populariteit van twee keuzeopties afhankelijk is van de aanwezigheid en kwaliteit van alternatieven in de keuzeset. Met andere woorden, in tegenstelling tot nutsgebaseerde MNL-modellen vertoont het RRM-MNL model geen IIA-eigenschap. Empirische toepassing van het RRM-model (in MNL-vorm en panel-Mixed Logit vorm) op vertrektijdstipkeuze-data laat zien dat het intuïtieve parameters genereert. Een vergelijking met nutsgebaseerde modellen suggereert dat het spijtgebaseerde model de data iets beter verklaart.

Uiteraard is er veel meer werk nodig om een robuust inzicht te verkrijgen in de empirische potentie van RRM-modellen (in vergelijking met nutsgebaseerde modellen en alternatieve modelvormen). Zie Chorus (2010) voor een vergelijking tussen RRM-MNL en RUM-MNL op basis van vier andere datasets, die ook suggereert dat het RRM-model de potentie heeft om beter te presenteren dan het RUM-model. De sociaalpsychologische literatuur geeft mogelijke handvaten om gericht te onderzoeken in welke keuzecontexten de RRM-benadering een mogelijk alternatief is voor de nutgebaseerde benadering: Zeelenberg and Pieters (2007) concluderen, in een overzichtstudie over de rol van spijt als determinant van keuzegedrag, dat “regret is experienced when decisions are difficult and important and when the decision maker expects to learn the outcomes of both the chosen and rejected options quickly”, en dat “regret is anticipated when significant others in the decision maker’s social network view the decision as important”. Dit zou betekenen dat bij voorbeeld routekeuzes tijdens evacuaties (“decisions are difficult and important”), rijbaan-veranderingen (“the decision maker expects to learn the outcomes of both the chosen and rejected options quickly”) en autotypekeuzes (“significant others in the decision maker’s social network view the decision as important”) voor hand liggen als mogelijke studieobjecten.

Dankwoord

De data over vertrektijdkeuze die gebruikt zijn in de schattingen in dit paper zijn verzameld in een onderzoek voor Rijkswaterstaat (toen AVV, nu DVS) dat werd uitgevoerd door RAND Europe. Wij willen DVS bedanken voor de toestemming tot hergebruik van deze data.

Referenties

- Bell, D.E., 1982. Regret in decision making under uncertainty. *Operations Research*, 30(5), 961-981
- Ben-Akiva, M., Lerman, S.R., 1985. *Discrete choice analysis: theory and application to travel demand*. The MIT Press, Cambridge, Mass
- Bierlaire, M., 2003. BIOGEME: A free package for the estimation of discrete choice models, *Proceedings of the 3rd Swiss Transportation Research Conference*, Ascona, Switzerland.
- Bierlaire, M., 2008. *An introduction to BIOGEME Version 1.7*, biogeme.epfl.ch
- Chorus, C.G., Arentze, T.A., Molin, E.J.E., Timmermans, H.J.P., van Wee, G.P., 2006a. The value of travel information: Decision-strategy specific conceptualizations and numerical examples. *Transportation Research Part B*, 40(6), 504-519
- Chorus, C.G., Molin, E.J.E., van Wee, G.P., Arentze, T.A., Timmermans, H.J.P., 2006b. Responses to transit information among car-drivers: Regret-based models and simulations. *Transportation Planning & Technology*, 29(4), 249-271
- Chorus, C.G., Arentze, T.A., Timmermans, H.J.P., 2008a. A Random Regret Minimization model of travel choice. *Transportation Research Part B*, 42(1), 1-18
- Chorus, C.G., Arentze, T.A., Timmermans, H.J.P., 2008b. Het compromiseffect in mobiliteitskeuzes, en de potentie van spijtminimalisatie-modellen. *Tijdschrift Vervoerswetenschap*, 44(4), 136-141
- Chorus, C.G., 2010. A new model of Random Regret Minimization. *European Journal of Transport and Infrastructure Research*, 10(2), 181-196
- Connolly, T., Reb, J., 2005. Regret in cancer-related decisions. *Health Psychology*, 24(4), pp. 29-34
- Kivetz, R., Netzer, O., Srinivasan, V., 2004. Alternative models for capturing the compromise effect. *Journal of Marketing Research*, 41, 237-257
- de Jong, G., Daly, A., Pieters, M., Vellay, C., Bradley, M., Hofman, F., 2003. A model for time of day and mode choice using error components logit. *Transportation Research Part E*, 39, 245-268
- Loomes, G., Sugden, R., 1982. Regret-Theory: An alternative theory of rational choice under uncertainty. *The Economic Journal*, 92(368), 805-824
- McFadden, D., 1974. Conditional logit analysis of qualitative choice-behaviour. In Zarembka, P. (Eds.) *Frontiers in econometrics*, Academic Press, New York, 105-142
- McFadden, D., Train, K.E., 2000. Mixed MNL models for discrete response. *Journal of Applied Econometrics*, 15(5), 447-470
- Savage, Leonard J., 1954. *The Foundations of Statistics*. New York: Wiley
- Simonson, I., 1989. Choice based on reasons: The case of attraction and compromise effects. *Journal of Consumer Research*, 19, 158-174
- Simonson, I., 1992. The influence of anticipating regret and responsibility on purchasing decisions. *Journal of Consumer Research*, 19(1), 105-119

- Starver, T., 2008. Anticipating regret: Why fewer options may be better. *Econometrica*, 76(2), 263-305
- Steenburgh, T.J., 2008. The invariant Proportion of Substitution Property (IPS) of discrete-choice models. *Marketing Science*, 27(2), 300-307
- Train, K.E., 2003. *Discrete choice methods with simulations*. Cambridge University Press, Cambridge UK
- Wernerfelt, B., 1995. A rational reconstruction of the compromise effect: Using market data to infer utilities. *Journal of Consumer Research*, 21(4), 627-633
- Zeelenberg, M., 1999. The use of crying over spilled milk: A note on the rationality and functionality of regret. *Philosophical Psychology*, 12(3), 325-340
- Zeelenberg, M., Pieters, R., 2007. A theory of regret regulation 1.0. *Journal of Consumer Psychology*, 17(1), 3-18