Strategic Network Modelling for Passenger Transport Pricing

Smits, E.-S.

DOI
10.4233/uuid:4feb2454-7d0a-4481-b8c2-4bae411d2e4a

Publication date
2017

Document Version
Final published version

Citation (APA)
Smits, E.-S. (2017). *Strategic Network Modelling for Passenger Transport Pricing*. [Dissertation (TU Delft), Delft University of Technology]. TRAIL Research School. https://doi.org/10.4233/uuid:4feb2454-7d0a-4481-b8c2-4bae411d2e4a

STRATEGIC NETWORK MODELLING FOR PASSENGER TRANSPORT PRICING

Erik-Sander Smits

Delft University of Technology, 2017

This thesis is a result from a project funded by the Netherlands Organisation for Scientific Research (NWO) and Delft University of Technology.

Netherlands Organisation for Scientific Research

TU Delft Delft University of Technology

STRATEGIC NETWORK MODELLING FOR PASSENGER TRANSPORT PRICING

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof.ir. K.C.A.M. Luyben;
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op
dinsdag 16 mei 2017 om 12:30 uur

door

Erik-Sander SMITS

Master of Science in de Wiskunde en
Bachelor of Science in de Informatica,
Universiteit Utrecht, Nederland
geboren te Waalwijk, Nederland

This dissertation has been approved by the
promotors: Prof. dr. ir. B. van Arem and Prof. dr. M.C.J. Bliemer
copromotor: Dr. ir. A.J. Pel

Composition of the doctoral committee:

Rector Magnificus	chairman
Prof. dr. ir. B. van Arem	Delft University of Technology
Prof. dr. M.C.J. Bliemer	The University of Sydney, Australia
Dr. ir. A.J. Pel	Delft University of Technology

Independent members:

Prof. dr. K.W Axhausen	ETH Zürich, Switzerland
Prof. dr. E.T. Verhoef	VU University Amsterdam
Prof. dr. ir. C.G. Chorus	Delft University of Technology
Prof. dr. ir. C.M.J. Tampère	University of Leuven, Belgium
Prof. dr. ir. S.P. Hoogendoorn	Delft University of Technology, reserve member

This thesis is the result of a Ph.D. study carried out from 2010 to 2017 at Delft University of Technology, faculty of Civil Engineering & Geosciences, department of Transport & Planning.

TRAIL Thesis Series no. T2017/3, the Netherlands TRAIL Research School

TRAIL
P.O. Box 5017
2600 GA Delft
The Netherlands
Phone: +31 (0) 15 278 6046
Fax: +31 (0) 15 278 4333
E-mail: info@rsTRAIL.nl

ISBN 978-90-5584-222-3

Printed in the Netherlands

PREFACE

It's done. The time to reflect on how this dissertation was realized has – finally – arrived. After having finished my studies in Mathematics and Computational Sciences in 2010, there was this search of the next step. Some believe all mathematicians become math teachers, but in fact, it was not easy to find job opportunities that could satisfy my desire for applied mathematical challenges. When Michiel Bliemer asked me if I was interested in doing a PhD in Delft, I immediately knew that this opportunity provided the challenges that I was looking for. On the other hand, I knew that PhD research is sometimes individualistic, and that it is part of the deal to write a dissertation – a whole book –.

Now, seven years later, I can look back on a period with new friendships, in which I've explored many places in the world, and which provided me plenty of opportunities to develop myself. This final part of scientific education has been an excellent experience, and I'm grateful for the opportunities that I've received. I would like to address my gratitude to everyone who contributed to this.

Prof. dr. ir. B. van Arem, beste Bart. Bedankt voor je ondersteuning en het vertrouwen dat je vanaf het begin in mij hebt gehad. Door onze discussies werd de verhaallijn van dit proefschrift steeds helderder, en begon ik zelf ook te geloven dat het einde in zicht kwam. Tijdens mij sollicitatie vroeg je of ik liever veel aspecten onderzoek en de breedte in ga, of dat ik mij liever op één onderwerp focus en dat tot in detail uitwerk. Ik moet je het antwoord nog steeds schuldig blijven.

Prof dr. M.C.J. Bliemer, beste Michiel. Dankzij jou heb ik de mogelijkheid gekregen om aan dit onderzoek te beginnen. Vanaf het moment dat ik mijn afstudeerstage in Deventer begon, heb ik onze samenwerking als zeer prettig ervaren. Je beschikt over een onuitputtelijke bron met uitdagende vragen. Als geen ander wist je mij wegwijs te maken in de wetenschappelijke wereld van prijsbeleid en netwerkmodellen. Ik kijk met bijzonder veel plezier terug op het onderzoek dat we in Sydney hebben gedaan.

Dr. ir. A.J. Pel, beste Adam. Bedankt voor al je hulp en het grondig reviewen van al mijn werk. Ook waardeer ik het zeer dat je altijd klaar stond om mijn vragen binnen no-time te beantwoorden. Het wordt tijd om weer een keer de fiets te pakken.

Members of my doctoral committee. Thank you for taking place in the committee and for your valuable feedback on my dissertation.

Dr. ir. R. van Nes, beste Rob. Bedankt voor het overnemen van een gedeelte van de begeleiding. Je keek altijd met een andere blik naar mijn werk, waardoor ik weer anders ging nadenken over de relevantie van het onderzoek.

Colleagues in Delft. It has been great to work together with such highly skilled and motivated people. Besides that, we had fun lunches, ping-pong tournaments and drinx. My roommates, Gijs, Meng, Wouter, Kakpo, Jeroen Luuk, and I only needed one word to initiate our hourly walk to the coffee machine. TRAIL PhD Council, thank you for the nice cooperation. Priscilla, Dehlaila, Esther en Conchita, bedankt voor al jullie ondersteuning op het secretariaat en bij TRAIL. Het was altijd fijn om langs te lopen en even aan de wetenschap te ontsnappen.

Colleagues at ITLS Sydney. Thank you for the best cappuccino breaks and lunches in Hyde Park. It has been a true pleasure to visit the Institute. Michiel en Mark, wat hebben we in een korte periode veel kunnen doen!

Collega's bij Arane: Jaap, Serge, Gerard, Koen. Ik geniet er elke dag van om jullie samen te werken. Bedankt voor jullie interesse in mijn promotieonderzoek, en de ruimte om het af te ronden. De netwerkmodellen hebben bij mij nu plaatsgemaakt voor verkeersmanagement en voetgangersstromen. Uitdagingen genoeg, en het is fantastisch om onze oplossingen op straat te zien werken.

Al mijn vrienden. Bedankt voor jullie goede vriendschappen en zeer welkome afleiding van mijn proefschrift. Giselle, Gijs, and Jip, thank you for the Canadian road trip after the IATBR conference. I will never forget the karaoke and the horse racing. Meng, Gijs, Daniel, Pavle, Giselle, Olga, Tamara, Bernat, Mario, I am very happy that I found new friendships at the department. Let's have an international BBQ soon again. Oud huisgenoten, homies for life! Marlouke en Joost, jullie waren onze rots in de branding tijdens ons Haagse avontuur. Goof, laten we onze discussies, onder het genot van een perfecte cappuccino of goed speciaalbier, over de wetenschap, promoveren en het leven nog lang doorzetten. René, de heuvels staan bijna om de hoek: tijd om die fiets weer op te lappen!

Heren van de Societas. Bedankt voor het verlichten van de studie met jullie vriendschap. Deze stelling heeft het lijstje voorin mijn proefschrift net niet gehaald: "Mathematical rigor is the sparkling snowflake on the top of the iceberg called science". Tsjoch!

Tjibbe, Jip, Geert, Robert, Daan, Brinn, Martijn, Mark, en natuurlijk Jasper. Gasten, wat hebben we samen veel meegemaakt. Wat goed dat ik jullie heb! Als vierde, en ik denk als laatste, het proefschrift af: ik kijk er naar uit om dat met jullie te vieren.

Mijn paranimfen. Ik prijs me zeer gelukkig met de paranimfen die ik straks aan mijn zijde heb. Thomas, als een van de weinige beta's in het Utrechtse alfa-bolwerk heb ik altijd een goede partner in crime aan je gehad. Je hebt je doel helder op het vizier en ik twijfel er niet aan dat je deze haalt. Ik kijk uit naar de dag van jouw eerste ruimtevaart. Tjibbe, maatje, wat heeft die paar jaar op de Meck ons goed gedaan. We hebben de afgelopen tijd al veel reisjes gemaakt, en ik kijk uit naar de volgende, maar eerst heb ik je nodig bij mijn tocht langs de katheder.

Mijn schoonfamilie. Hanne, Egbert en Maud, bedankt voor alle warmte en gezelligheid. Maarten en Annemiek, het is heel fijn om bij jullie een tweede thuis te hebben. Bedankt dat ik de laatste stukken van dit proefschrift in alle rust in Rougeux heb kunnen schrijven.

Mijn lieve zusjes, Saskia en Karlijn. Bedankt dat jullie er altijd voor me zijn! Ik kan me geen betere zussen wensen.

Mijn ouders, Menno en Marja. Jullie hebben wat geduld moeten hebben, maar het einde is nu écht in zicht. En zonder jullie had ik dat nooit gehaald. Bedankt voor alle liefde en steun. Bedankt voor alles.

Emma, mijn allerliefste. Woorden schieten tekort om te beschrijven hoeveel je voor me betekent. Zonder jou had ik dit boek in elk geval nooit kunnen voltooien. En het wordt alleen maar mooier als we straks met z'n drieën zijn.

Erik-Sander Smits
May 2017

CONTENTS

NOTATION

- Finite sets (e.g., the route set, the set of modes) are denoted with calligraphic capitals $\mathcal{A} \ldots \mathcal{Z}$
- Bold letters are vectors
- Small capital sans serif NAMES are constants.
- Throughout this thesis, the product operator on sets is the standard Cartesian product.

GENERAL

$\mathbb{R}$	Set of real numbers	
$\mathbb{R}_+$	Set of non-negative real numbers	
$\lvert \cdot \rvert$	Operator: Number of elements in a set	
δ^i_j	Kronecker delta	$\left(\delta^i_j = \begin{cases} 1 \text{ if } i = j \\ 0 \text{ if } i \neq j \end{cases}\right)$
$\mathbb{E}(\cdot)$	Expected value	
$\mathrm{Var}(\cdot)$	Variance	
$\mathrm{Cov}(\cdot)$	Covariance	
$\mathrm{Corr}(\cdot,\cdot)$	Correlation	
$\sigma(\cdot)$	Standard deviation	
$\mathbb{P}(\cdot)$	Probability	
$F(\cdot)$	Cumulative Distribution Function	$(F(\cdot) : \mathbb{R} \to [01])$

STAKEHOLDERS

$\mathcal{S}$	Set of stakeholders	
s	Stakeholder index	$(s \in \mathcal{S})$
H_s	Objective function of stakeholder $s \in \mathcal{S}$	$(H_s : \Gamma \to \mathbb{R})$
C	Coalition of stakeholders	$(C \subseteq \mathcal{S})$
E	Traffic assignment function	$(E : \Pi \to \Gamma)$

PRICING MEASURES

$\mathcal{P}$	Set of pricing measures	
p	Pricing measure index	$(p \in \mathcal{P})$
Π_p	Feasible prices of measure $p \subseteq \mathcal{P}$	$(\Pi_p \in \mathbb{R})$
Π	Set of feasible prices	$(\Pi = \prod_{p \in \mathcal{P}} \Pi_p)$
π_p	Price of $p \in \mathcal{P}$	$(\pi_p \in \Pi_p)$
$\boldsymbol{\pi}$	Vector of prices, or pricing scheme	$(\boldsymbol{\pi} = \{\pi_p \vert p \in \mathcal{P}\} \in \Pi)$
π^*_p	Resulting price of $p \in \mathcal{P}$	$(\pi^*_p \in \Pi_p)$

$\boldsymbol{\pi}^*$	Resulting pricing scheme	$\left(\boldsymbol{\pi}^* = \left\{\pi^*_p \middle| p \in \mathcal{P}\right\} \in \Pi\right)$

Effects

$\mathcal{E}$	Set of (external) effects	
e	Effect index	$(e \in \mathcal{E})$
Γ_e	Feasible level set of effect $e \subseteq \mathcal{E}$	$(\Gamma_e \in \mathbb{R})$
Γ	Feasible effect levels set	$(\Gamma = \prod_{e \in \mathcal{E}} \Gamma_e)$
γ_e	Level of effect $e \in \mathcal{E}$	$(\gamma_e \in \Gamma_e)$
$\boldsymbol{\gamma}$	Vector of effect levels	$(\boldsymbol{\gamma} = \{\gamma_e \lvert e \in \mathcal{E}\} \in \Gamma)$

Network

$\mathcal{L}$	Set of links	
i	Link index	$(i \in \mathcal{L})$
$\mathcal{N}$	Set of nodes or intersections	
n	Node index	$(n \in \mathcal{N})$
$\mathcal{O}$	Set of origins	$(\mathcal{O} \subseteq \mathcal{N})$
o	Origin index	$(o \in \mathcal{O})$
$\mathcal{D}$	Set of destinations	$(\mathcal{D} \subseteq \mathcal{N})$
d	Destination index	$(d \in \mathcal{D})$
$\mathcal{OD}$	Set of Origin-Destination (O-D) pairs	$(\mathcal{OD} \subseteq \mathcal{O} \times \mathcal{D})$
od	O-D pair index	$(od \in \mathcal{OD})$

Choice alternatives

$\mathcal{M}$	Set of modes	
$\mathcal{R}^m_{od}$	Set of routes for mode $m \in \mathcal{M}$ and O-D pair $od \in \mathcal{OD}$	
$\mathcal{R}^m$	Set of all routes for mode $m \in \mathcal{M}$	$\left(\mathcal{R}^m = \cup_{od \in \mathcal{OD}} \mathcal{R}^m_{od}\right)$
$\mathcal{R}_{od}$	Set of all routes for O-D pair $od \in \mathcal{OD}$	$\left(\mathcal{R}_{od} = \cup_{m \in \mathcal{M}} \mathcal{R}^m_{od}\right)$
$\mathcal{R}$	Set of all routes	$(\mathcal{R} = \cup_{od \in \mathcal{OD}} \mathcal{R}_{od})$
r	Route index	$(r \in \mathcal{R})$
$\mathcal{T}$	Set of time-of-day periods	
T	Time-of-day period index	$(T \in \mathcal{T})$
$\mathcal{C}_{od}$	Set of choice alternatives for O-D pair $od \in \mathcal{OD}$	$(\mathcal{C}_{od} = \mathcal{T} \times \mathcal{R}_{od})$
$\mathcal{C}$	Set of all choice alternatives	$(\mathcal{C} = \cup_{od \in \mathcal{OD}} \mathcal{C}_{od})$
c	Choice alternative index	$(c \in \mathcal{C})$
V	Systematic utility	
U	Utility	
$\mathcal{U}$	Set of user-classes	
u	User-class index	$(u \in \mathcal{U})$
$D_{(u;od)}$	Total mobility demand for user-class $u \in \mathcal{U}$ for O-D pair $od \in \mathcal{OD}$	$(D_{(u;od)} \in \mathbb{R}_+)$
$\boldsymbol{D}_u$	Vector of mobility demand (O-D-matrix) for user-class $u \in \mathcal{U}$	$\left(\boldsymbol{D}_u = \{D_{(u;od)} \lvert od \in \mathcal{OD}\} \in \mathbb{R}_+^{\lvert\mathcal{OD}\rvert}\right)$

$\tau^{\text{IVT-FF}}_{(T;r)}$	Travel time: In-vehicle time during free flow conditions
$\tau^{\text{IVT-CONG}}_{(T;r)}$	Travel time: In-vehicle time during free flow condition
$\tau^{\text{IVT}}_{(T;r)}$	Travel time: General in-vehicle time (used for mode TRAIN)
$\tau^{\text{WAIT}}_{(T;r)}$	Travel time: Waiting time
$\tau^{\text{A-E}}_{(T;r)}$	Travel time: Access and egress time

Choice modelling

The notation for choice modelling has different contexts in Chapters 3, 4, and 7. This list provides general notation.

ε_c	Random variate of the analyst error of route $c \in \mathcal{C}$	
G	Generating function for choice set $\mathcal{C}$	$\left(G : \mathbb{R}_+^{\lvert \mathcal{C} \rvert} \to \mathbb{R}_+\right)$
y_c	Element of the generating vector for choice alternative $c \in \mathcal{C}$	
$\boldsymbol{y}$	Generating vector of choice set $\mathcal{C}$	$(\boldsymbol{y} = (y_1, \ldots, y_{\lvert \mathcal{C} \rvert}))$
G_{od}	Generating function for O-D pair od	$\left(G_{od} : \mathbb{R}_+^{\lvert \mathcal{C}_{od} \rvert} \to \mathbb{R}_+\right)$
μ	Overall scale parameter	$(\mu \in \mathbb{R}^+)$
P_c	Choice probability of choice alternative $c \in \mathcal{C}$	$(P_c \in [0,1])$

The following notation is specific for the overall framework:

$y_{(T;r;u)}$	Vector element for choice alternative $(T;r) \in \mathcal{C}$ of user-class $u \in \mathcal{U}$	$(y_{(T;r;u)}\mathbb{R}^-)$
$\boldsymbol{y}_{(od;u)}$	Generating vector of user-class $u \in \mathcal{U}$ for O-D pair $od \in \mathcal{OD}$	$\left(\boldsymbol{y}_{(od;u)} = \{y_{(T;r;u)} \mid T \in \mathcal{T}, r \in \mathcal{R}_{od}\}\right)$
μ^{MODE}	Mode-nest specific scale parameter	$(\mu^{\text{MODE}} \in \mathbb{R}^+)$
$\mu^{\text{T-O-D}}$	Time-of-day-nest specific scale parameter	$(\mu^{\text{T-O-D}} \in \mathbb{R}^+)$
$P_{(T;r;u)}$	Choice probability for alternative $(T;r) \in \mathcal{C}$ for user-class $u \in \mathcal{U}$	$(P_{(T;r;u)} \in [0,1])$

Network Loading

The given definitions below related to turns are specific for Chapter 3.

f_r	Number of trips for route $r \in \mathcal{R}$	$(f_r \in \mathbb{R}_+)$
f_i	Actual flow on link $i \in \mathcal{L}$	$(f_i \in \mathbb{R}_+)$
k_i	Density on link $i \in \mathcal{L}$	$(k_i \in \mathbb{R}_+)$
v_i	Speed on link $i \in \mathcal{L}$	$(v_i \in \mathbb{R}_+)$
F_i	Fundamental diagram (density $\to$ flow) of $i \in \mathcal{L}$	$(F_i : \mathbb{R}_+ \to \mathbb{R}_+)$
$\mathcal{W}$	Set of turns	$(\mathcal{W} \subseteq \mathcal{L} \times \mathcal{L})$
$\langle i,j \rangle$	Turn from link $i \in \mathcal{L}$ to link $j \in \mathcal{L}$	$(\langle i,j \rangle \in \mathcal{W})$
$\mathcal{W}^r$	Ordered set of turns for route $r \in \mathcal{R}$	$(\mathcal{W}^r \subseteq \mathcal{W})$
$\prec^r$	Turn order operator for route $r \in \mathcal{R}$	
$S_{\langle i,j \rangle}$	Demand for turn $\langle i,j \rangle \in \mathcal{W}$	$\left(S_{\langle i,j \rangle} \in \mathbb{R}_+\right)$
$\boldsymbol{S}$	Vector of turn demands	$\left(\boldsymbol{S} = \cup_{\langle i,j \rangle \in \mathcal{W}} S_{\langle i,j \rangle}\right)$
φ_i	Reduction (or squeezing) factor at the exit of $i \in \mathcal{L}$	$(\varphi_i \in (0,1])$
$\boldsymbol{\varphi}$	Vector of reduction vectors	$(\boldsymbol{\varphi} = \cup_{i \in \mathcal{L}} \varphi_i)$
Φ	Node model function (turn demands $\to$ reduction factors)	$\left(\Phi : \mathbb{R}_+^{\lvert \mathcal{W} \rvert} \to (0,1]^{\lvert \mathcal{L} \rvert}\right)$
Ψ	Queuing model function (reduction factors and route demands $\to$ densities per link)	$\left(\Psi : (0,1]^{\lvert \mathcal{L} \rvert} \times \mathbb{R}_+^{\lvert \mathcal{R} \rvert} \to \mathbb{R}_+^{\lvert \mathcal{L} \rvert}\right)$

Node Model

The given definitions below related to turns are specific for Chapter 5.

$\mathcal{I}$	Set of inlinks	$(\mathcal{I} \subseteq \mathcal{L})$
i	Inlink index	$(i \in \mathcal{I})$
$\mathcal{J}$	Set of outlinks	$(\mathcal{J} \subseteq \mathcal{L})$
j	Outlink index	$(j \in \mathcal{J}^n, n \in \mathcal{N})$
$\mathcal{W}$	Set of turns	$(\mathcal{W} \subseteq \mathcal{I} \times \mathcal{J})$
$\langle i,j \rangle$	Turn from inlink $i \in \mathcal{I}$ to outlink $j \in \mathcal{J}$	$(\langle i,j \rangle \in \mathcal{W})$
S_i	Demand for inlink $i \in \mathcal{I}$	$(S_i \in \mathbb{R}_+)$
R_j	Supply of outlink $j \in \mathcal{J}$	$\left(R_j \in \mathbb{R}_+\right)$
Q_i	Capacity of link $i \in \mathcal{I} \cup \mathcal{J}$	$(Q_i \in \mathbb{R}_+)$
q_i	Reduced capacity of inlink $i \in \mathcal{I}$	$(q_i \in \mathbb{R}_+)$
$\alpha_{\langle i,j \rangle}$	Turning fraction of turn $\langle i,j \rangle \in \mathcal{W}$	$\left(\alpha_{\langle i,j \rangle} \in [0,1]\right)$
f_i	Flow out of inlink $i \in \mathcal{I}$	$(f_i \in \mathbb{R}_+)$
f_j	Flow into outlink $j \in \mathcal{J}$	$\left(f_j \in \mathbb{R}_+\right)$
$f_{\langle i,j \rangle}$	Flow on turn $\langle i,j \rangle \in \mathcal{W}$	$\left(f_{\langle i,j \rangle} \in \mathbb{R}_+\right)$
h_i	Headways at the exit of inlink $i \in \mathcal{I}$	$(h_i \in \mathbb{R}_+)$
h_j	Headways at the entry of outlink $j \in \mathcal{J}$	$\left(h_j \in \mathbb{R}_+\right)$
$h_{\langle i,j \rangle}$	Headways on turn $\langle i,j \rangle \in \mathcal{W}$	$\left(h_{\langle i,j \rangle} \in \mathbb{R}_+\right)$
$d_{\langle i,j \rangle}$	Turn delay of turn $\langle i,j \rangle \in \mathcal{W}$	$\left(d_{\langle i,j \rangle} \in \mathbb{R}_+\right)$

Transferable Utility Game Theory

$\mathcal{S}$	Grand coalition (equals the set of stakeholders)	
v	Coalition value function	$\left(v : 2^{\mathcal{S}} \to \mathbb{R}\right)$
$(\mathcal{S}, v)$	Transferable Utility (TU)-game with grand coalition $\mathcal{S}$ and coalition values v	
χ_s	Allocation of stakeholder s	$(\chi_s \in \mathbb{R})$
χ	Vector of allocations	$\left(\chi = \{\chi_s \mid s \in \mathcal{S}\} \in \mathbb{R}^{\lvert\mathcal{S}\rvert}\right)$
Q	Partition of the grand coalition	(Q= partition of $\mathcal{S}$)
$K(\mathcal{S}, v)$	Core of TU-game $(\mathcal{S}, v)$	$\left(K(\mathcal{S},v) \subseteq \mathbb{R}^{(\lvert\mathcal{S}\rvert)}\right)$
$\zeta(\mathcal{S}, v)$	Shapley value of TU-game $(\mathcal{S}, v)$	$\left(\zeta(\mathcal{S},v) \in \mathbb{R}^{\lvert\mathcal{S}\rvert}\right)$
ω_s	Lower bound of stakeholder s	$(\omega_s \in \mathbb{R})$
ω	Vector of lower bounds	$\left(\omega = \{\omega_s \mid s \in \mathcal{S}\} \in \mathbb{R}^{\lvert\mathcal{S}\rvert}\right)$
Ω_s	Upper bound of stakeholder s	$(\Omega_s \in \mathbb{R})$
$\boldsymbol{\Omega}$	Vector of upper bounds	$\left(\boldsymbol{\Omega} = \{\Omega_s \mid s \in \mathcal{S}\} \in \mathbb{R}^{\lvert\mathcal{S}\rvert}\right)$
$\boldsymbol{\eta}(\mathcal{S}, v)$	Compromise value of TU-game $(\mathcal{S}, v)$	$\left(\boldsymbol{\eta}(\mathcal{S},v) \in \mathbb{R}^{\lvert\mathcal{S}\rvert}\right)$

Constants

GOV	Stakeholder: National government	(GOV $\in \mathcal{S}$)
TRAIN	Stakeholder: Train operator	(TRAIN $\in \mathcal{S}$)
AMS	Stakeholder: Amsterdam city council	(AMS $\in \mathcal{S}$)
LH	Effect: Loss hours	(LH $\in \mathcal{E}$)
TOLL	Effect: Total collected toll	(TOLL $\in \mathcal{E}$)

PROFIT Effect: Profit .... (PROFIT $\in \mathcal{E}$)
CO_2 Effect: CO_2 .... ($CO_2 \in \mathcal{E}$)
E-MAX-U Effect: Expected maximum utility .... (E-MAX-U $\in \mathcal{E}$)
KM-INC Effect: Income from kilometre charge .... (KM-INC $\in \mathcal{E}$)
CRDN-INC Effect: Income from cordon charge .... (CRDN-INC $\in \mathcal{E}$)
EMIS-STUDY Effect: Total value of emission in the study area .... (EMIS-STUDY $\in \mathcal{E}$)
EMIS-AMS Effect: Total value of emission in the Amsterdam area .... (EMIS-AMS $\in \mathcal{E}$)
LH-OTHER Effect: Loss hours of travellers exogenous to choice model .... (LH-OTHER $\in \mathcal{E}$)
ACCESS Effect: Accessibility of amsterdam .... (ACCESS $\in \mathcal{E}$)
LOSS Effect: Loss due to missed economic activity .... (LOSS $\in \mathcal{E}$)
TRAIN-INC Effect: Income from train ticket sales .... (TRAIN-INC $\in \mathcal{E}$)
TRAIN-COST Effect: Operating costs for the train operator .... (TRAIN-COST $\in \mathcal{E}$)
KM-ON Pricing measure: Kilometre charge — on peak .... (KM-ON $\in \mathcal{P}$)
KM-OFF Pricing measure: Kilometre charge — off peak .... (KM-OFF $\in \mathcal{P}$)
CRDN-ON Pricing measure: Cordon charge — on peak .... (CRDN-ON $\in \mathcal{P}$)
CRDN-OFF Pricing measure: Cordon charge — off peak .... (CRDN-OFF $\in \mathcal{P}$)
FARE-ON Pricing measure: train fare — on peak .... (FARE-ON $\in \mathcal{P}$)
FARE-OFF Pricing measure: Train fare — off peak .... (FARE-OFF $\in \mathcal{P}$)
CAR Mode: Car .... (CAR $\in \mathcal{M}$)
TRAIN Mode: Train .... (TRAIN $\in \mathcal{M}$)
HOME Mode (special): Stay-at-home or telework alternative .... (HOME $\in \mathcal{M}$)
R_0 Route: Dummy route for the Stay-at-home or telework alternative .... ($R_0 \in \mathcal{R}$)
$\overline{\text{CAR}}$ Route: Representative, or average, route for mode CAR .... ($\overline{\text{CAR}} \in \mathcal{R}$)
$\overline{\text{TRAIN}}$ Route: Representative, or average, route for mode TRAIN .... ($\overline{\text{TRAIN}} \in \mathcal{R}$)
PEAK Time-of-Day: Morning peak (7AM-9AM) .... (PEAK $\in \mathcal{T}$)
SHOULDER Time-of-Day: Morning shoulder (4AM-7AM and 9AM-12AM) .... (SHOULDER $\in \mathcal{T}$)
PEAK-PREF User-Class: Preference for peak travel .... (PEAK-PREF $\in \mathcal{U}$)
NO-PREF User-Class: No preference for peak travel .... (NO-PREF $\in \mathcal{U}$)

Chapter 1.

Introduction

This introduction describes important issues in transport such as congestion and emissions, and proposes pricing as a solution. The reader is presented with what they can expect from this dissertation, with the direction of the research, and with what the main societal and scientific contributions are in the dissertation. That is – in short – transport pricing measures (Section 1.1) affect the behaviours of travellers (Section 1.2). Strategic planning models (Section 1.3) capture these responses and are used to design and evaluate new pricing policies. Challenges regarding this process define the motivation of the research and its relevance for society (Section 1.4). Finally, the to be bridged scientific gaps (Section 1.5) are discussed based on the outline of the dissertation.

Travel occupies a substantial part of everyday life. A person usually has several activities each day, and if two successive activities are at different locations, one has to make a trip between them. Irrespective of the mode of transport (e.g, walk, cycle, public transport, car, or a combination of these), all trips together are straining the transport system. That system is vital for the economy and needs to be shared with commuters, tourists, and commercial transporters. Improvements and changes of mobility and transport infrastructure are frequently discussed by governments and stakeholders.

Congestion is an undesired effect of travel. There are stretches of road on which queues form almost every day. This over-saturation of the passenger transport system is also indicated by high emission levels and crowded public transport. The problems associated with mobility and passenger transport have engaged politicians, policy makers, economists, and engineers for decades already. However, despite all the effort, sustainable accessibility and mobility is still far from reality.

By making trips, travellers impose effects on others. One additional vehicle in a queue will increase the travel time of all other vehicles behind it. The emissions one causes by driving to work affect the residents and workers along that road. When one occupies the last seat in a bus, all following passengers have to stand. These negative effects, or external effects, travellers impose on others are not ‘paid for’ by the causer. Travel choices are generally made out of self-interest and do not take the burden experienced by others into account.

In addition, the transport system is not efficiently utilized. Congestion is at its worst during peak hours, but at other times of the day there is plenty of spare capacity. Because we impose similar working hours on ourselves, we overload the transport system all at the same time. Another inefficiency involves the spare capacity in public transport. Since taking the car is fast and convenient, many travellers do not take a public transport alternative if it takes longer and involves a transfer. For the system, however, the latter is the preferred option with less emissions and congestion.

If it were possible to have complete control and make everyone's travel choices centrally, then the total congestion, emissions, and other effects could be minimized jointly. That utopian situation, in the eyes of transport planners at governments, will never be feasible. However, policy instruments can be used to steer towards an optimal situation.

On the other hand, everyone has a different opinion about optimality. Environmentalists would like to put everyone in emission-free public transport alternatives, and insurance companies want safety on the road. But the owner of a trucking company wants high profits, and thus guaranteed high speeds on the freeway. Profit is also the objective of privately operated transit operators, while governments need to consider multiple issues at once. These conflicting preferences complicate decision making.

1.1. Passenger Transport Pricing

Changing the price for mobility can contribute to the reduction of external effects. Price incentives can change the behaviour of travellers and thus make the transport system more efficient. If the peak hours are avoided and more sustainable modes are used, then the queues will shrink and the air quality will improve. So by manipulating the travel choices by introducing a pricing measure, policy makers can work towards their goals.

Pricing strategies have been successfully deployed in numerous places worldwide. The cities of London, Stockholm, and Singapore charge users who want to enter the centres by car, and have successfully improved the transport system. Insurance companies are now implementing Pay-as-you-Drive policies, which primarily focus on safety and fairness – also external effects of the transport system. Special (fast) tolling lanes where the price depends on the number of vehicles on the (slower) main road (i.e., congestion) exist in Israel and the United States. In the Netherlands peak avoidance projects[1] are being deployed to reduce the number of vehicles on the road for limited periods, .e.g., during road works. Frequent users of a certain road receive a reward if they do not drive on this road during peak hours.

In addition to these innovative strategies, there are more traditional measures, such as fuel surcharges (also known as excise taxes), annual registration fees, parking fees, road taxes, and public transport fees. The innovative strategies are often more flexible than the traditional strategies, because the price incentive can be tailored for specific travellers, at specific locations, and at specific times of the day. To stimulate sustainability for example, road taxes can be differentiated towards fuel efficiency of the vehicles (i.e, economical cars have a reduced fee). The more aspects prices can be differentiated on in a pricing measure, the more innovative it is. Especially time-differentiated policies are regarded as innovative. Policy makers

[1] 'Spitsmijden' in Dutch

usually consider several measures simultaneously and are interested in their joint effect.

In the Netherlands only reward schemes (e.g. drivers are paid to avoid the peak) have been established as an innovative measure up-to now. Despite (or thanks to) decades of political discourse and many proposals, no innovative pricing measure has been implemented. From toll roads, via a peak hour sticker, to a kilometre charge, everything has failed due to lack of political support (see van der Sar and Baggen, 2005; Smaal, 2012; Ubbels, 2006).[2] Low public support also nourished the indecisiveness. Nor were the conflicts mentioned earlier between preferences of different stakeholders beneficial; for example, the travellers' association Royal Dutch Touring Club ('ANWB') has opposed against the proposed policies in the last decade of the previous century after they held a survey under its members.

Neglect of the conflicting preferences of stakeholders during the design and planning process of pricing policies finally blocks the implementation of it. To determine the details of the pricing policy only a single objective is considered, for example, the reduction of congestion or other external effects. Stakeholders like an automobile association may have different preferences, and since they represent a large part of the population their support is important. Since political and public support are key conditions for a successful implementation, multiple stakeholders, their preferences, and their influence have to be incorporated in planning pricing measures.

Innovative pricing measures do have a high potential. Advances in technology allow detailed and differentiated schemes that can influence the choices of specific groups of travellers. That means that congestion and emissions at specific times and at specific places can be mitigated. They also allow for improvements of economic factors such as equity and welfare. Specifically, fairness can be improved because travellers are charged for usage instead of having to pay a flat fee, as is now the fact with registration fees and road taxes. The potential in the Netherlands is demonstrated by peak-avoidance projects, where travellers are rewarded (see Knockaert et al., 2010).

The European Commission stimulates *user pays* and *polluter pays* principles and studies recommend highly differentiated kilometre charging (see van Essen et al., 2012). Furthermore, in none of the European Union members road transport is fully paid by its users. Verhoef et al. (2004a) conclude in their Dutch research report that pricing on the road is effective to mitigate congestion. This is especially true when (1) the measure is strongly differentiated, (2) perceptions of different users are taken into account, and (3) the revenue is recycled as a compensation for negative effects and in acquisition of public support.

The electricity market and electricity network shows interesting similarities with the passenger transport market and road network. The demand for electricity stems from people's activities, and electricity can not be stored, just like road capacity. Therefore, electricity rates are higher during peak periods in order suppress peak demand.

[2] Only a few tolled tunnels exist.

1.2. Travellers' Responses to Pricing Measures

Price incentives affect several choices that travellers make (Karlström and Franklin, 2009; Vrtic, 2009). The fact that travellers are sensitive to prices makes pricing an effective tool to improve the performance of the transport system. In order to get insights into the effects of pricing measures, it is important to understand these responses. One's demand for trips is directly derived from one's activities; therefore, price incentives work on two levels. They can affect activity patterns and they can affect the way trips are made. Also, not all pricing measures are able to change all types of behaviour.

Define the travel demand of a person as the desired movements the person would like to make. In economic terms, this demand for travel is a *derived demand* since travelling itself is not beneficial. The activities performed at different locations are the satisfiers, or benefits, of travelling. The activity patterns of a person determine the travel demand of that person. The combined activity patterns of all people in a region represent the total travel demand for that region.

Choices influencing activities include major life events: they are carefully considered and based on many factors. Examples are choosing a job and choosing a place to live. These choices are considered long-term responses to price incentives. They are closely related to location choice (i.e. choosing where to have your activities), and an important aspect of locations is their accessibility. This accessibility is in turn highly affected by the price of mobility, because pricing measures can increase perceived distances. Therefore, pricing measures cause a response in long-term choices. At the same time, many other aspects, like public transport availability, influence accessibility. Some residential areas are very well connected to the public transport system, but more remote residences can only be reached by private transport modes. So, transport pricing can not be solely considered to steer activity-related choices.

Choices related to consuming mobility (i.e. how to get from A to B) occur repetitively, are often habitual, but are also sensitive to prices. They are the choices that one makes on a daily basis, or that can at least be changed every day. These choices named travel choices and are considered as short-term responses to price incentives. For transport pricing measures, there is a high potential to the short-term choices because they are made so often. On the other hand, the volatility of these choices, i.e. the ease with which travellers can change them, makes it cumbersome to grasp them in models that try to predict these choices.

Figure 1.1 shows an example of a person's activity pattern and travel demand on a certain day. The person works, shops and goes to a movie on that day. All activities take place at different locations and they are connected by trips. This spatial travel demand is related one-to-one with the activity pattern and it is thus a derived demand. By choosing the departure times, transport modes, and routes for each leg of the travel demand, trips are the result. The work and shopping related trips are performed by car and follow three distinct routes. Public transport is taken on the trips to and from the movie and different service lines are used for each leg.

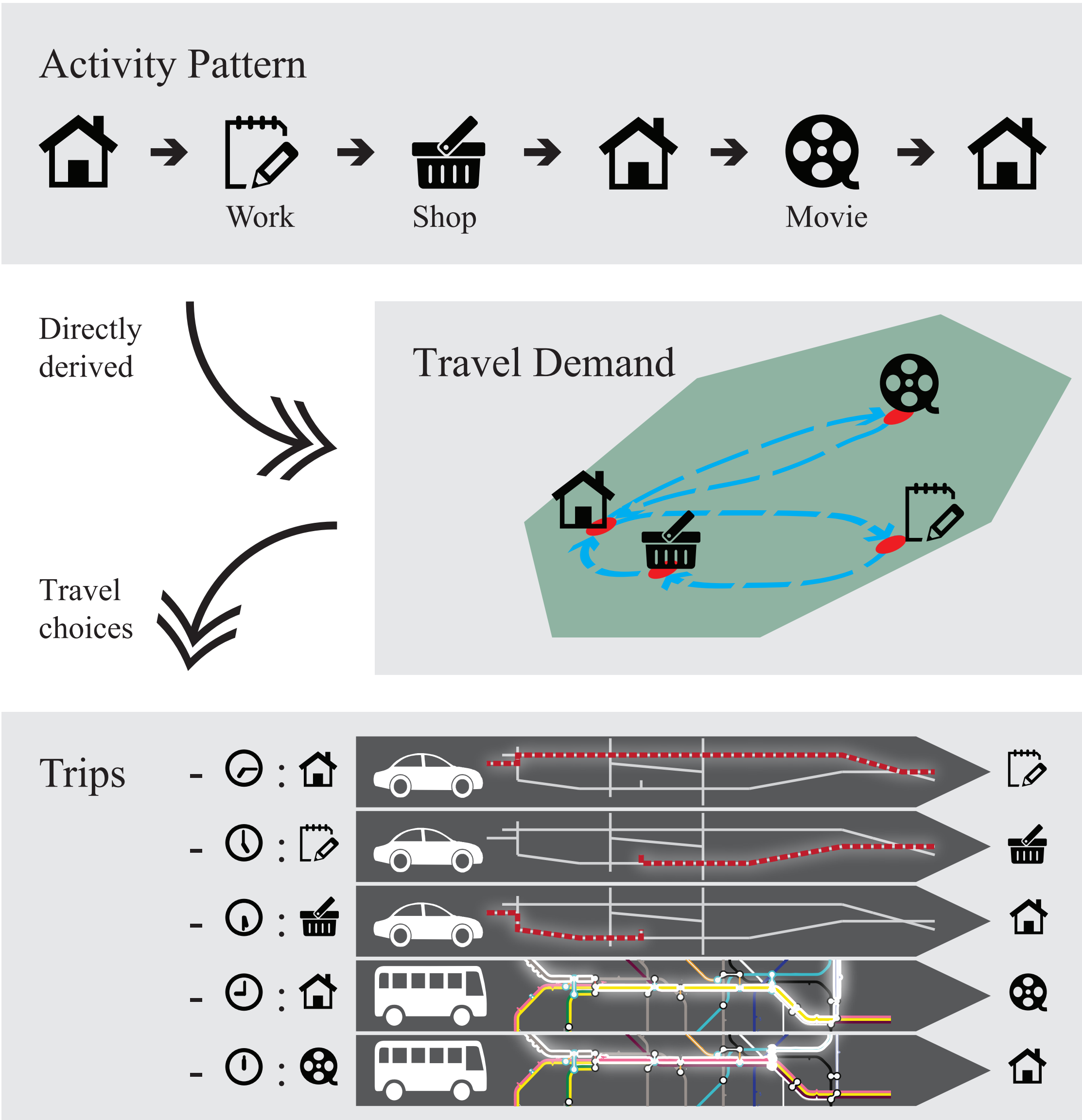


Figure 1.1.: Example of an activity pattern, its travel demand, and chosen trips.

We can distinguish the following travel choices:

- *Change departure time*
  Instead of leaving at the preferred departure time, one could prepone[3] or postpone the trip. Reasons for doing so can either be severe expected delays or a price incentive. Such a price incentive comes from a time differentiated pricing measure. In the Netherlands for example, train tariffs are time differentiated; season tickets for off-peak only are significantly cheaper than season tickets that are valid during the peak hours as well.
- *Change route*
  Deviating from the preferred route of a trip is very common. Frequently small detours are taken to avoid congestion or busy intersections. Although the taken routes will always alter, and hence there is no real standard route, introducing a toll at specific roads will make them less travelled. So, having location differentiated pricing measures will influence the route choice behaviour of travellers.
- *Change mode*
  For each trip a different set of modes is feasible and the chosen mode is habitually anchored. The modes cycling and public transport are frequently interchanged, but car users tend to stick to their familiar mode. However, substantial price incentives, such as high parking fees, can get drivers out of their cars and direct them towards public transport.
- *Stay at home*
  The do-not-travel alternative. If an activity is not mandatory, or if the activity can also be carried out at a persons' current location, then the choice can be to not travel at all. Teleworking (i.e., working from home) is a great example of this, and occurs more and more frequently. While the cost of mobility and travel is not the only reason for teleworking (others include the working environment), prices will influence the decision. The stay at home alternative is posted as a short term response; however, teleworking can be either sporadic or systematic. In the latter case it could be regarded more as a long term response.

1.3. Strategic Planning Models

The decision making process of policy makers is usually supported by strategic planning models that determine medium and long-term effects of policies. These models aim to forecast, and especially to assess the consequences of major alterations to the transport system. The latter can be infrastructure projects, but also new policies such as transport pricing. Planning models for pricing schemes determine expected changes in travel behaviour as well as their impact on effects. Multiple scenarios can be compared based on their performance for several indicators. Strategic models are often important and useful tools in the decision making process.

An important feature of strategic transport planning models in the passenger transport pricing context is the ability to reflect travellers' responses to changes in the transport system.

[3] i.e., bring forward to an earlier time

This implies that the behaviour has to be captured in mathematical equations, so that it can predict travel choices based on hypothetical costs. Different choices are described by different behavioural mechanisms and thus captured by different models.

Models are a simplified representation of reality, and - in transport - models can be found for many subjects and in a wide variety of levels of detail. The essence of models is to provide a (mathematical) abstraction of reality such that current situations can be reproduced and hypothetical situations can be evaluated through forecasting. It is beyond question that the transport system is complex. There is a large number of heterogeneous agents (e.g., travellers, public transport vehicles) interacting within the transport network, which is challenging to describe in models. The transport system itself consists of a huge number of roads sections and intersections. It is not feasible to model every detail due to limited computer power. Therefore, models exist at different scales, from nano-models that describe individual travellers to micro-models that analyse individual vehicles to macro-models that consider aggregate traffic flows.

The traditional basis of strategic planning models in transport is the 4-step model (see Figure 1.2). It consists of trip generation, trip distribution, mode choice and trip assignment. These four consecutive steps provide a very basic guide to transport modelling. Based on socio-economic data, the trip generation step determines the activities of people and how many trips they make. The trip distribution step then finds a destination for each activity, i.e., travel patterns are determined. A travel pattern consists of sequence of trips: movements from origins to destinations. Next, travel choices are modelled. Each trip is assigned to a mode of transport (e.g. car or public transport): the mode choice step. Finally, the best route for each trip using the specific mode is found, and flows are loaded onto the network, resulting in traffic and travel times.

The results from the 4-step model are the travel times for each trip, the number of cars on each road section, and the service level of public transport. From these quantities other external effects as emissions and noise levels can be estimated. The combined results indicate the performance of the transport system. Policy makers base their decisions on the performance indicators for different strategies.

Lack of realism induced countless adjustments and additions to the classical 4-step model. The choices existing in the 4-step model are trip choice (i.e. do I travel or not), destination choice, mode choice, and route choice. Each of them is tackled separately and has its own limitations. For example, the aggregated approach to destination choice groups people by location. It considers almost no individual characteristics and is not based on behavioural principles. A second example is the standard route choice technique which considers delays from congestion at the wrong location. To overcome these and other drawbacks, different approaches, such as disaggregate models, have been introduced and used during the last five decades or so.

There is a long history of transport model development[4] that has left its mark on the state of practice. Software packages are available that execute the traditional 4-step model, which makes it easy to apply them. Governments also standardized the tool-kit, leaving little room for innovations. However, developments due to scientific research are outstripping the state of

[4] For more background, see these well-known textbooks on transport modelling: (Cascetta, 2009; Ortúzar and Willumsen, 2011)

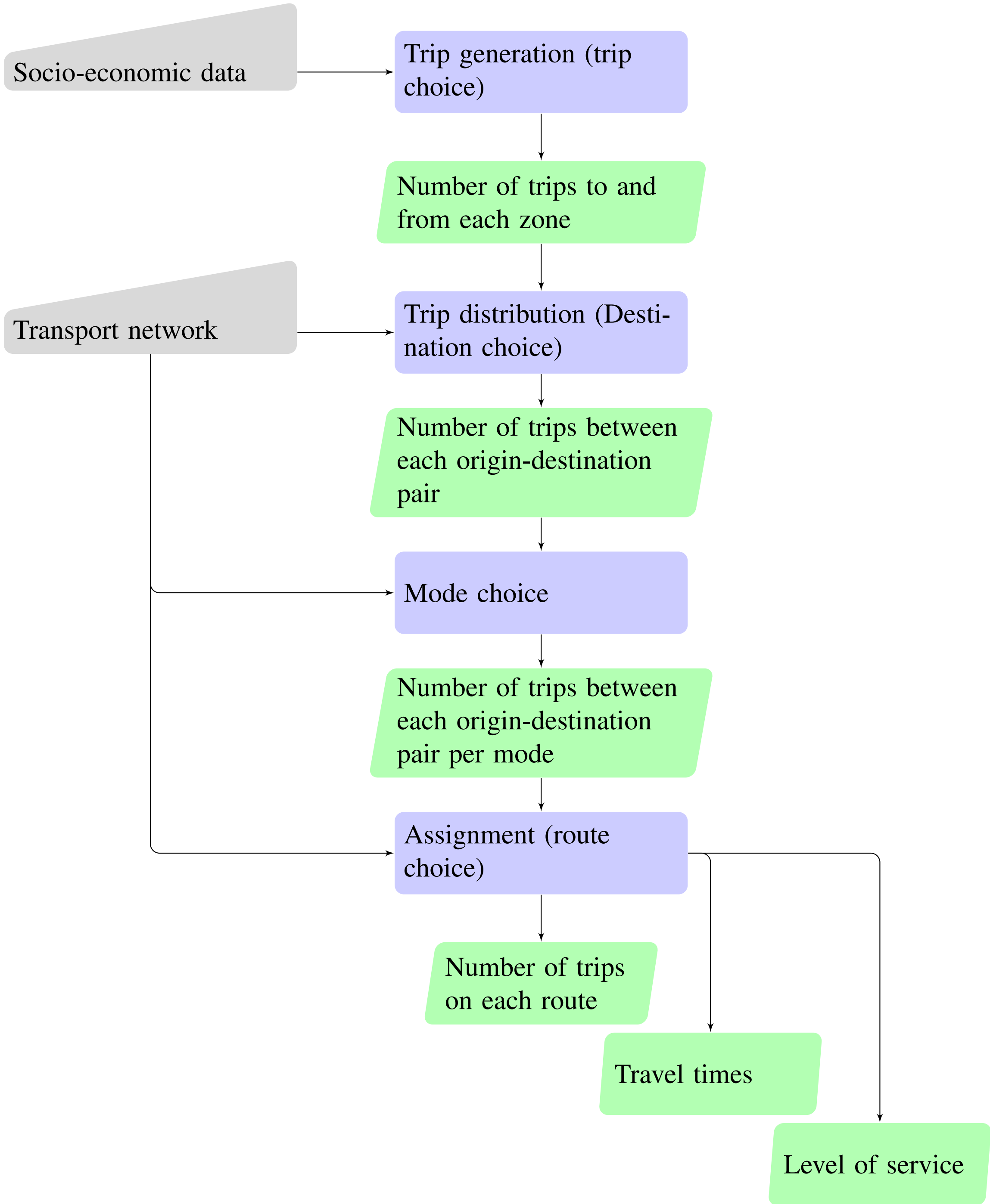


Figure 1.2.: 4-step model

practice. It has become feasible to achieve a much higher level of behavioural realism. In the next paragraphs three important drawbacks and potential solutions are presented.

One of the main disadvantages of traditional traffic models is their inability to identify bottlenecks or queue locations. A bottleneck is a location in the network where demand exceeds capacity. Simply stated, a bottleneck is a location in the network where more vehicles want to pass than physically possible. That leads to congestion and thus to delays. A very important observation is that the queue builds up before (i.e., upstream of) the bottleneck and not inside it. In the traditional static approach, the delay is predicted inside the bottleneck. These static models do not include a time dimension or traffic dynamics (e.g. queue formation). Modern dynamic and quasi-dynamic models overcome this problem and thus lead to more realistic queue formation. These models simulate the physical formation of queues.

Bottlenecks always arise at nodes in the road network (i.e., locations where homogeneous roads link); these can range from a simple lane drop on the highway to a complex signalized intersection. The main purpose of an node model is to identify bottlenecks by checking if there is sufficient supply (i.e., capacity) to accommodate the demand (i.e., incoming traffic). Furthermore, it needs to determine how severe the bottleneck is and what flow constraints to apply. Thus it determines how many vehicles can proceed. Finding behaviourally realistic models and solutions that reproduce intersection flows is not straightforward. For example, flows on priority intersections are inherently non-unique. Only recently Tampère et al. (2011) formulated the basic requirements for proper node models, and currently only two proper node models exist. Compared to other fields in transport modelling, node models received less attention from scientists, and one could argue that they are underdeveloped. The underlying behaviour of the existing models is not completely clear, nor do extensive empirical validations of the models exist.

A final disadvantage in the traditional model relates to the way choices are modelled. Choices (e.g., route choice) are often assumed to be made between distinct alternatives, meaning that modellers can enumerate the possibilities. In networks, all routes between an origin-destination pair form such a set of alternatives. The distance between each origin-destination pair is different. The logit model family is often used for route choice, because, compared to the traditional deterministic approach, it can handle different preferences between travellers and uncertainty about travel time and other route characteristics. The uncertainty about a route's travel time is proportional to the distance between the origin and destination. The frequently applied logit model family does not take this proportionality into account. This behavioural property makes them less suitable for applications to transport networks.

Strategic planning models that are acting as a decision support system should run quickly; especially when they are part of a design process. For example, let the location of a cordon toll and the price level of the toll be two variables for a pricing scheme. If ten possible choices exist for both, the model needs to calculate all one hundred combinations to find the optimal choices. In reality, often much more than two dimensions exist, and the number of variants that should be considered grows rapidly. Therefore, the running time of strategic planning models has to be low. So, while developing strategic transport models, the computational efficiency should always be kept in mind. This leads to a practical constraint: when one aims to increase the realism of strategic planning models, the computation time should not be impacted much.

1.4. Societal Motivation and Relevance

The previous sections describe societal issues that are the motivation for the conducted research. It can be decomposed in a problem formulation, a solution approach, and then some underlying challenges that hamper the implementation of this solution, in summary:

Problem formulation

The transport system functions inefficiently and suboptimally. By travelling, people put others at disadvantages, like congestion and emission exposure. If travellers would make different choices, e.g., if they would avoid peak periods or use more public transport, the total impact of the external effects would be smaller.

Solution approach

Provide incentives for travellers through innovative pricing strategies to change their behaviour. This improves the transport system and makes travellers responsible for the impacts their choices have on others. Innovative pricing strategies differentiate the price level for specific travellers, for specific locations, and/or specific times of the day. This particularly affects these travel choices: trip choice, mode choice, time-of-day choice, and route choice.

Challenges

- Public and political support for innovative pricing policies is often low.
- Strategic planning models that could support the decision making process lack realism at several aspects.
- The computation time of strategic planning models constrains the number of strategies that can be assessed.

The research in this thesis is motivated by these three challenges associated with transport pricing. The way to achieve this, is through the development of new and improved methods within strategic planning models. So, the research aims to increase public and political support by providing a strategic planning model that has increased realism at a similar computational efficiency. The focus of this disseratation is on the improvement and development of methodology, rather than specific applications, although feasibility is demonstrated by a case study. So, particular policy issues within the decision making process, and particular transport pricing projects are not within the scope of the research.

The following chapters present a toolbox that can be used by analysts and decision makers to improve public and political support. The basic principle applied to each tool is achieving a good balance between computational efficiency and realism, making them usable for analysts *and* credible for decision makers. In other words, the behavioural realism should be as high as possible such that results are trustworthy, and the computing costs should be as low as possible such that one is not limited in the amount of considered pricing schemes.

The way this dissertation aims to alleviate the challenges is twofold. On the one hand by providing a guide to model travellers' responses to pricing measures, and on the other hand by providing insight in the interactions at the negotiation table of decision makers. The first is in line with traditional and current applications of strategic planning models. while the latter is generally not part of the strategic planning models. Therefore, two main topics will be covered: (1) travellers' behaviour and especially their response in terms of travel choices to pricing measures and the state of transport system, and (2) the preferences and interactions of stakeholders at the negotiation table. The latter identifies possible conflicting interests, and tries to provide solutions for them.

Both topics have a different 'nature', meaning that the types of methods and histories in development do not overlap. Therefore, the organization of the dissertation follows the structure of the two topics, and it is therefore organized in two parts. Part I, Traffic Assignment, allows strategic planning models to be more realisic in terms of choice behaviour and queue formation without making it impossible to assess many different pricing schemes. Part II, Stakeholders & Pricing, provides a new approach to model how stakeholders negotiate, including solutions for conflicts.

Another boundary of the scope lies within the Traffic Assignment (TA) model of Part I. This model is limited to travellers' choices with respect to mobility consumption. Activity related long-term responses depend on many factors, and interact with other markets than the transport market. Therefore, the activity pattern is input for the lower level model, and thus assumed to be available. Mode, route, and departure time choice are modelled, and the stay at home alternative also exists. The latter still allows a thin connection with the activity-related responses, since the stay at home alternative can represent the economic concept of demand elasticity. Note that throughout this dissertation, *traffic assignment* is the simultaneous assignment of a mode, route and departure time to a movement of a traveller. This is contrary to the earlier mentioned traditional notion of assignment that only involves route choice.

As will become clear, the approach of the dissertation is primarily methodological. However, despite the mathematical character of these contributions, each topic has a clear societal relevance; they can all be retraced to the earlier mentioned challenges. One should bear in mind that the mathematical modelling tools in the toolbox provide abstractions of behaviours and systems, which makes them supportive in the real decision making process. Real societal added value arises when analysts apply the tools to assess meaningful transport pricing cases.

At the same time, acknowledged strategic planning studies on transport pricing remain a necessity to create political and societal support. The author believes that meaningful results can be achieved by applying the methods and tools in this dissertation. Chapter 7 illustrates this by analysing a transport pricing case study for the Randstad area in the Netherlands. It shows how conflicts between governments and a train operator can be resolved when they introduce a kilometre charge and change fares simultaneously.

1.5. Scientific Contributions

The scientific gaps that are addressed in this dissertation are discussed based on its outline (see Figure 1.3) and the structure of the proposed model framework, of which the latter is first briefly introduced. The tools are pinned down in an extensive model framework that exists of several sub-models. Chapter 2 sketches the outlines this framework based on a more extensive discussion on transport pricing than is provided in this introductory chapter. It consists of two levels; an upper level for the decision making process of stakeholders, and a lower level for the travellers' responses. The upper level determines a pricing scheme, and the lower level determines the impacts of that scheme in terms of effects.

The analysis of aspects related transport pricing Chapter 2 leads to a set of requirements for each of the two levels. It also deliberates on the several approaches to transport pricing (e.g. economics and policy/politics), and the relation to the engineering approach of this dissertation. Subsequently, the upper and lower level will be discussed in more detail in respectively Part II and Part I. Especially Chapters 3 and 6 introduce the framework in more detail.

The scientific contributions can be subdivided into three themes: the holistic approach, methodological advances in traditional transport modelling, and new methodology to analyse the decision making of multiple stakeholders.

1.5.1. Holistic approach

This dissertation combines methods in several disciplines (such as discrete choice analysis, traffic flow theory and game theory) into an extensive framework. All methods are available to assess innovative pricing measures, and the related short-term travel responses of travellers. As presented in the next chapter, many transport pricing studies' scopes are much narrower. Therefore, collecting and bringing together many theories within a hollistic approach is a contribution on its own.

The bi-level transport pricing framework uses state-of-the-art sub-models and captures many important aspects of innovative pricing measures. The upper level uses game theory and especially TU-games to address multiple stakeholders. The route, mode, and departure time choice of travellers is modelled with discrete choice, which captures individual preferences and perception biases. The node model satisfies the recently derived first-order requirements. Traffic flow and propagation is more realistically tackled with kinematic wave theory than in traditional static traffic assignment, while it is still has relatively low computational costs.

The dissertation finishes with an application of the extensive framework on the Randstad area in the Netherlands. In Chapter 7 the methodological improvements of Part I and the new game-theoretical multi-stakeholder approach come together in a case study in the Randstad area in the Netherlands. This case study illustrates how this dissertation collects and unifies leading models from different fields.

1.5.2. Methodological Advances in Transport Modelling

Second, several aspects of TA-(sub)models within existing strategic planning models are analysed and improved. The basis for this is so-called Quasi-Dynamic Traffic Assignment (QDTA), which allows to simulate more realistic queue formation, and is computationally efficient. It satisfies many of the formulated requirements. Chapter 3 introduces this QDTA-model, while the remainder of Part I provides in depth studies on node models and route choice models. Two highlights of corresponding contributions are:

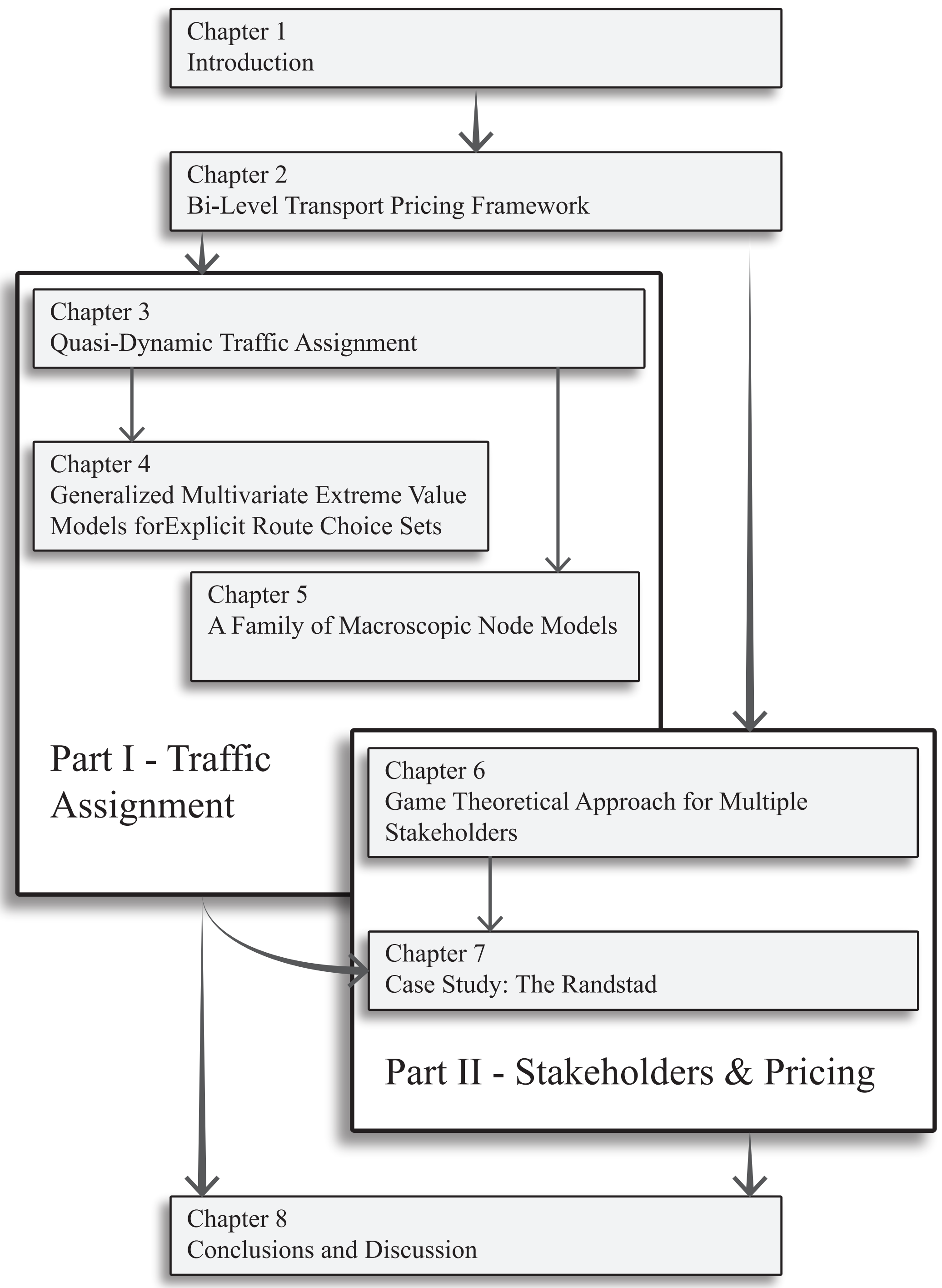


Figure 1.3.: Dissertation Outline with introductory Chapters 1 and 2, detailed lower level analysis in Part I, and detailed upper level analysis in Part II. All theory converges in the transport pricing case study of Chapter 7. Chapter 8 summarizes the conclusions and adds discussion.

.

1. The behavioural realism of route choice is improved because the perceived travel times and travel costs are better simulated. The framework of Random Utility Maximization (RUM), in which travellers choose the route with the lowest 'cost' (i.e. highest utility), is used. The utility is composed of a deterministic and a random part. Instead of taking the sum of these parts, this study takes the product. This especially allows that the uncertainty about a routes' travel time is not the same for short and long routes. That this is more realistic is underpinned with empirical data. (Chapter 4)
2. Node, or intersection, models are reconsidered. In such a model the flow on an intersection is determined based on boundary conditions. The new framework has implicit delays for every vehicle. That allows a much better behavioural interpretation of node model results. (Chapter 5)

1.5.3. Analysis of Multiple Stakeholders' Decision Making

Third, multiple stakeholders are included, and the potential benefit of cooperation between them can be determined. Multiple stakeholders, their preferences, and their interaction explicitly comprise the decision making upper level component. Stakeholders each have their own objectives in terms of effects, and also their own executive power. A game-theoretical approach captures the interaction between them; this is presented in Chapter 6. Cooperation and non-cooperation are two paradigms that are explicitly considered. The difference in effects between the two paradigms shows the potential improvement reached with cooperation. Political and public support, main factors of successful measures, can increase with this multi-stakeholder approach. It is the first to consider multiple objectives with a TU-game approach; furthermore, the price of non-cooperation can be determined.

CHAPTER 2.

BI-LEVEL TRANSPORT PRICING FRAMEWORK

This chapter describes the basic mechanisms involved in passenger transport pricing and provides a framework to simulate these mechanisms. First, several approaches (e.g. perspectives from economics and policy) to transport pricing are discussed and related to the engineering approach of this dissertation. Second, relevant aspects of transport pricing are introduced and briefly discussed. Finally, a bi-level formulation that can serve as a strategic planning model for innovative transport pricing measures will be presented. This formulation is accompanied by requirements for both levels. The upper level, which is the decision making level, has the distinguishing requirement of being able to handle multiple stakeholders. The lower level addresses provides a computational platform for travel choices and takes the transport infrastructure into account; its requirements concern primarily realism and computational efficiency.

Passenger transport pricing has been analysed extensively. Motivations for pricing have scientific underpinning from multiple theories and models. They describe the underlying mechanisms of mobility and its price. This work is multidisciplinary because social, political, technical, and operational aspects are involved. The implementations of pricing schemes – and the failure thereof – show the practical importance of these aspects wherein multiple research questions remain open. This chapter analyses transport pricing based on the literature, and uses these insights to develop a versatile modelling framework. Analysts can derive strategic planning models for transport pricing from this framework.

The abundance of literature about transport pricing marks its importance in science, and also depicts its versatility. Tsekeris and Voß (2009) present the state-of-the-art of the design and evaluation of road pricing. They cite well over four-hundred articles in their review, and point out the extensive choices for economic principles and underlying network performance models. De Palma and Lindsey (2011) restrict their overview to road congestion pricing (i.e., the objective is congestion relief by means of charging vehicles); they present several types of pricing schemes and discuss how to choose between them. Lawphongpanich et al. (2006); Yang and Huang (2005) present several economic and mathematical models for road pricing.

Other overview papers include (Hau, 2005a,b; Morrison, 1986; de Palma et al., 2006; Parry, 2009).

The first idea of pricing in transport was proposed by the economist Pigou (1920). In Section 2.1 his theory of marginal cost pricing is presented. All following transport pricing studies rely on this basic economic principle, while the literature disperses over the decades to other disciplines as engineering, mathematics, politics/policy, and psychology. This multidisciplinarity has increased over time.

In this chapter the most important aspects of transport pricing are topic-wise discussed. Each of these topics is important for a successful implementation of pricing schemes. Thus ideally, a strategic planning model takes them all into account. Therefore, this chapter also analyses the implications of the different aspects for the framework from which the strategic planning models are derived. This is done in terms of requirements for the framework and the models derived from the framework. Transport pricing studies in the literature only satisfy to a small subset of these requirements. They often present analytical results for which the assumptions are generally very strong, and for which the transport networks are extremely simplified. The achievable level-of-detail of the presented framework is much higher than the level-of-detail of the established (pricing) literature.

First, several transport pricing aspects will be discussed. Second, several pricing measures and (external) effects will be presented and discussed in order to provide more context on what is possible with transport pricing. Third, the mathematical framework will be introduced. To cover as many transport pricing aspects as possible, this framework will be set up holistically; it allows multiple stakeholders, multiple pricing measures, multiple effects, multiple modes, and multiple user-classes. This requires multiple modelling levels in the framework, with the most important distinction between an upper level decision making model for stakeholders and a lower level assignment model to assess the pricing scheme. The requirements for both levels of the framework are specifically stated.

2.1. Basic Principle of Transport Pricing

The basis of transport pricing is the Pigouvian toll (see Pigou, 1920), also referred to as marginal cost pricing, or first-best pricing. The underlying principle is that travellers are taxed on top of their private travel cost to compensate for the caused external effects. *External effects* are the effects that are caused by the traveller, but for which they do not take full responsibility. They are also called externalities; examples are congestion, emissions, noise, and unsafety. Consider Figure 2.1 with three main curves: inverse demand, private cost and social cost. The horizontal axis contains the number of trips and the vertical axis the marginal cost (i.e., the increase in total cost for one additional trip). The private costs are the ‘out-of-pocket’ costs for the traveller, and the social cost also include the delays of others and emissions, i.e., the external effects. The inverse demand curve becomes the demand curve by swapping the axes, and then represents the amount of travellers willing to make the trip given a certain price. The intersections of the cost curves with the inverse demand curve are the equilibria: the user equilibrium and the social optimum.

In the user equilibrium situation there is an overconsumption of mobility, i.e., the social cost

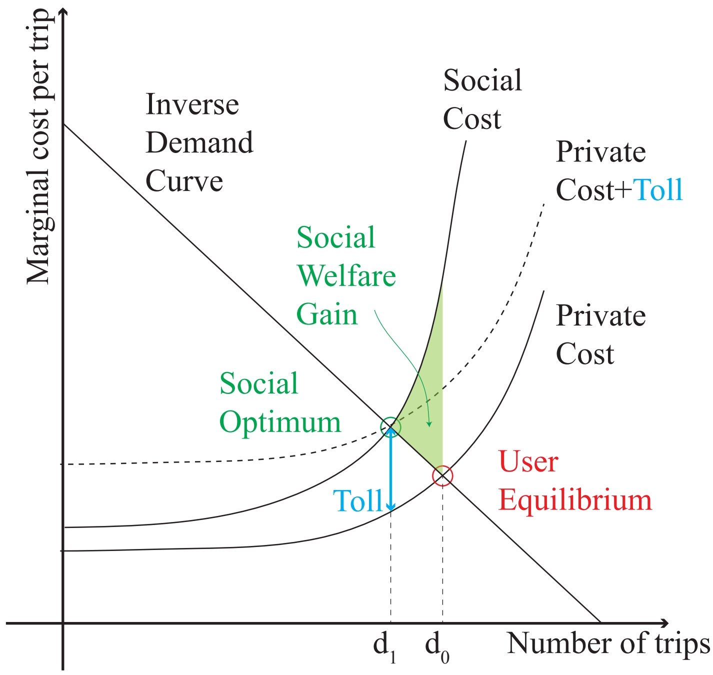


Figure 2.1.: Marginal cost pricing or first-best pricing. The toll equals the difference between the marginal private cost and marginal social cost in system optimum. The green area is the social welfare gain.

is not completely covered for trips made. By charging a toll that equals the difference between the marginal private cost and marginal social cost at the social optimum, the overconsumption will be corrected. When this toll is included in the private cost, the user equilibrium and the social optimum coincide. Furthermore, the green area in Figure 2.1 shows the gain in social welfare under the Pigouvian toll.

This economic theory cannot be translated directly to real transport networks. First, the toll would be different on every road segment because the social cost depends on the how and where congestion builds up, and on who are effected by emissions. Second, the value of time is assumed to be equal for all travellers. This is obviously not true in reality. Third, the inverse demand curve should be known. In reality this curve depends on many factors; think for example of O-D patterns, travel alternatives, and other markets (e.g., the labour and housing markets).

The economic literature has dealt with several of these issues and has relaxed some of the strict assumptions. This extended the marginal cost pricing principle in many directions. While more detail was added, the more generalizations were achieved. As a result, the theory has moved more and more from an economic to an engineering approach. Adding ‘real world’ constraints to the pricing problem is called second-best pricing in economic terms. The transport pricing aspects touched upon in Section 2.2 can all be considered as additional ‘real world’ constraints. Studies in the literature generally discuss one or a few of these aspects, while the framework in this chapter aims to capture most of them.

A seminal contribution after Pigou that has to be mentioned is the bottleneck model of Vickrey (1969), who considers a physical queue and different departure times of travellers. Congestion occurs because all travellers want to arrive at the same time, but this congestion can be mitigated with a toll. He ignited a stream of studies that all take temporal dynamics into accounts. The importance of these dynamics is beyond questioning, but the strategic network models have not been able to capture them for a long time (see also Section 3.1). Therefore, the time dimension, the different type of dynamics, and their relation with strategic network models is thoroughly discussed in Section 2.2.6.

2.2. Aspects of Transport Pricing

This section contains an enumeration of the most important aspects involved in transport pricing. The success of implementation of pricing is affected by its *political* and *public support*. The efficiency of pricing schemes is bounded by the *price of anarchy*. Price incentives cause *travellers to respond*, and in their consideration the *value of time* is very important. The way the *time dimension* is taken in, determines which temporal dynamics can be captured. The transport system’s representation also varies along other dimensions; inclusion of multiple *transport modes* provides alternatives, and the *network* and its traffic representation largely influence the level of realism. Finally, *charge collection technologies* are an important feasibility constraint.[1]

[1] The PhD thesis of Ubbels (2006) also discusses several aspects of road pricing, he takes an economic approach and he does not relate the aspects to strategic network models.

The transport pricing framework focusses on the decision making process itself and the behavioural responses of travellers. Some aspects are important for the decision making process, and others are related to the responses of the travellers. This section points out to which mechanisms the aspects apply. However, some aspects relate to neither, but since they are important for transport pricing, they are mentioned in this section. Examples are the revenue distribution that influences acceptability and the charge collection technology.

2.2.1. Decision Making & Political Support

In general, policy makers are responsible for the implementation of pricing schemes. Decisions are made after an interplay of planning studies and politics. Vonk Noordegraaf et al. (2014) identify and analyse the factors that play a role in the implementation and non-implementation of road pricing in several areas based on 106 scientific papers. Political and public support are the most important factors, but certainly not the only ones. In Norway the road authorities played a significant role, and in London the power of the mayor was important. They conclude that the decision makers have to manage a broad set of factors, which can be different for every case.

This indicates that in reality multiple stakeholders are involved, and they have different preferences and interaction. In fact, the implementation of pricing measures in the Netherlands has failed because of conflicts between stakeholders; in one case the negative standpoint of (the members of) the Royal Dutch Touring Association (ANWB) was one of the reasons for the government to abort the project.

Studies on highly simplified networks have shown that competition between stakeholders can reduce the efficiency of the transport system. Acemoglu and Ozdaglar (2007) show on a parallel network that monopolists can always achieve the social welfare optimum, but that competition in oligopolistic system leads to reduced efficiency. Van der Weijde et al. (2013) indicate that introducing dynamics in a multimodal duopolistic setting can lead to different conclusions (i.e., prices) for the players than in the static case. This shows the importance of realistic transport system representation.

This all shows that it is important to take decision making itself into account in strategic planning. Multiple stakeholders (e.g., governments, public transport operators, lobbies) interact, and they all have different objectives. Note that besides governments at different levels, also different departments of the same government have different objectives. For example, the transport department wants less congestion, and the planning department wants accessible cities. Implementation failures of several pricing measures and the identification that economics of the transport system changes when multiple stakeholders are active, are the main arguments to have a multiple stakeholder approach in the framework in this dissertation. This leads to a better understanding of conflicting interests, and might lead to higher political support.

On the contrary, the classical modelling approach sees decision making (i.e., price setting) as a single objective optimization problem. First, an objective function is defined (e.g., social welfare, profit), and constraints are formulated that capture the transport system. Second, an optimization methodology is applied to determine which pricing scheme will result in the optimal value of the objective function. This classical approach is straightforward. Although it can be a highly non-linear problem and difficult to solve, it remains restricted to a single objective function. The interactions and conflicts can not be captured by a single objective function; therefore, the currently used methods are not sufficient to model stakeholder behaviour. Only a few studies exist that take the different preferences and interactions of stakeholders into account, especially in combination with a network model (see e.g., Ohazulike, 2014).

2.2.2. User Equity & Public Support

In addition to policy makers, the public is a major other factor of influence. As mentioned in the previous subsection, Vonk Noordegraaf et al. (2014) find public support one of the main factors for successful implementation of pricing schemes. Hamilton (2012) points out the importance of acceptability by the public of pricing measures, and provides the most influential decisive factors: people's experience, how revenue is spent (revenue recycling), self-interest, and political attitudes. Therefore, it is important that the public support is taken into account during the design of pricing schemes.

Measures that relate to public support are social welfare and user equity. The first reports the overall gains of pricing, while the equity measure also take into account how these gains are distributed among the population. Levinson (2010) reviews the equity effects of road pricing, and his message clearly addresses the issue for strategic planning: "The perception of equity is highly subjective. A project that may appear equitable to an analyst across one set of dimensions may not to individuals affected by the project. Achieving consensus on decisions (thereby ensuring people believe the decision was equitable) may involve departure from objective 'engineering' rationality, moving into the realm of politics."

For strategic planning not all aspects of public support can be addressed. It is possible to adjust people's experiences with pricing by pilots, as done in Stockholm, and designing implementation paths. The scenarios can be designed with planning models, but it remains difficult to assess the change in attitude towards pricing. Further, equity depends on the costs of a trip and on the accessibility of areas. Those measures can be addressed with planning models. Ecola and Light (2009) advise that equity is considered in an early stage of decision making by using planning models. As an additional remark, revenue recycling is important for public support and part of the design of pricing schemes, but it does not directly influence the behaviour of travellers which leaves too little grip for planning models to assess revenue recycling.

Some studies in the literature dedicate the design to improve acceptance by the travellers. For example, it is possible to assess acceptable pricing schemes by means of Pareto-improving tolls (see e.g. Guo and Yang, 2010). Those consider pricing with revenue redistribution such that no traveller is worse off. Such schemes are equitable by design and are likely to be supported by the public. Wu et al. (2010) and Wu (2011) consider Pareto-improving pricing on multimodal networks, and thus design equitable and acceptable schemes. They also provide an overview of the literature on this topic.

2.2.3. Price of Anarchy

The Price of Anarchy is an interesting (game theoretical) concept that indicates the inefficiency of the transport system. There is a difference between the performance of the system optimum (i.e., the total cost is minimal) and the user equilibrium (i.e., each user chooses its best alternative). The simplest illustration hereof is the welfare gain described in Section 2.1. In general the price of anarchy can be defined as the ratio of a performance indicator, such as social welfare, between the system optimum and the user equilibrium of some system. Usually, pricing measures are used to steer towards the system optimum. Therefore, the price of anarchy is bounded from above by the gains of pricing measures.

The price of anarchy was initiated in 1999 in a conference paper version of Koutsoupias and Papadimitriou (2009). The paper 'How bad is selfish routing?' by Roughgarden and Tardos (2002) further introduces the topic (it provides a nice analogy with a system of strings and springs) and provides that for a very simplified case the total travel time in user equilibrium are at most $4/3$ times the total travel time in the system optimum. It has been analysed under less strict – but still strict – assumptions by Chau and Sim (2003); Correa et al. (2004); Han et al. (2008); Perakis (2007); Roughgarden (2003); Schulz and Stier-Moses (2006). Maillé and Stier-Moses (2009) investigate to what extent the price of anarchy can be resolved by rewarding mechanisms, and Schulz and Stier-Moses (2006) provide route guidance inspired by the price of anarchy.

Given certain pricing measures, planning models can derive the optimal strategy for a certain objective. The ratio between the original reference scenario without pricing and this optimal strategy then provides a specific price of anarchy for the analysed measure. The price of anarchy for transport systems seems to be more popular amongst game theoreticians and computer scientists than amongst transport economists and engineers; however, the measure can provide more insight in the maximum achievable gain of any transport policy, including transport pricing. Therefore, it is worth paying attention to the price of anarchy during the planning process of policies.

2.2.4. Responses of Travellers

Understanding the response of the traveller is an important aspect of transport pricing. The fact of the matter is that the main working mechanism of a pricing measure is to provide an incentive for travellers to change their behaviour. As discussed in Chapter 1 there are several possible responses, which can be categorised into short-term and long-term responses. The choices regarding the consumption of mobility (i.e., the short-term responses) are more important for the transport system (with respect to the interactions between them), and therefore only those will be considered in the remainder. See Section 1.2 for a brief discussion on the long-term responses.

As pointed out in Sections 1.2 and 1.3 the four most important choices that should be captured in strategic planning models for transport pricing are route choice, mode choice, departure time choice, and trip choice. In the past, several mechanisms have been used to capture the different choices. The traditional 4-step model (see Section 1.3) uses aggregate approaches to represent for example mode choice. The flexibility of the traditional approach is limited;

for example, in these aggregate approaches it is difficult include individual characteristics and preferences. The transport pricing studies with more economic approaches often distinguish between fixed and elastic demand. In such an approach the multiple choice aspects (e.g., budget constraints, mode substitutions) are captured with a single inverse demand function. The derivation or calibration of such functions is infeasible with the level-of-detail considered in this thesis. Recent studies on transport pricing tend to use discrete choice models. Vrtic (2009) uses for example stated preference data form a Swiss' survey to analyse mode, route and departure time choice for pricing measures. Nielsen (2004) shows empirically (with GPS devices) what the behavioural responses to a kilometre price were in an experiment in Copenhagen, Denmark; the main changes were new routes, new destinations for 'occasional trips', switching to off-peak, and travel less.

For strategic planning models, it is important that the interaction between the pricing mechanism and the corresponding travel choice is captured realistically. For example, when time-differentiated tolls are in place, travellers should be able to change their departure time in the planning model. In a recent cost-benefit analysis in the Netherlands, the authors conclude that their underlying transport model was unable to sufficiently capture departure time changes (Hilbers et al., 2015). Therefore, all expected responses of the travellers should be listed in the early design stage of strategic planning models. In addition, appropriate modelling paradigms should be considered to be able to capture these responses realistically.

2.2.5. Value of Time Distribution & User-Classes

As mentioned earlier in Section 2.1 regarding the transition from first-best pricing to second-best pricing, travellers are heterogeneous; they have different budgets and they value their time differently. Therefore, they will respond differently to pricing measures, which is a very important aspect for road pricing. For some travellers reducing travel time (and thus delay) is very important, and they are willing to pay a high price. Other travellers may simply minimize cost, and will easily adapt their behaviour when price incentives are provided. This is called taste heterogeneity in the literature.

Arnott et al. (1992) are one of the first to consider heterogeneity in a transport pricing context. They provide a first investigation of two user-classes on a network consisting of two parallel links. Verhoef et al. (2004b) further investigate heterogeneity and they conclude that the potential effectiveness of pricing measures can be underestimated when heterogeneity is ignored. That there is indeed variation in taste was confirmed by Small et al. (2005) with empirical data; they find substantial additional benefits in value pricing when taste variation is taken into account. Recent studies that discuss pricing for heterogeneous travellers include (van den Berg and Verhoef, 2011; Guo and Yang, 2010; Jiang et al., 2011). The review by Small (2012) provides an introduction to the value of travel in general, and he discusses what is known and what should be known in the future.

In addition to heterogeneity of travellers, vehicles also differ in performance (i.e., fuel consumption and emission). Distinguishing different vehicle types can also improve quantifying effects such as emissions.

It is important that strategic planning models capture that heterogeneity in preferences of travellers and heterogeneity in vehicles. With respect to taste heterogeneity two approaches are possible; either the differences are captured in continuous distributions of tastes (or willingnesses to pay), or the differences are captured in discrete classes where each class contains a part of the population that is assumed to have similar preferences.

2.2.6. Time Dimension: Time-of-Day & Static versus Dynamic

Time plays a major role in transport systems. Conditions change over time with recurrent patterns around rush hours. Travellers want to be on-time at their destinations with a minimal travel time. Public transport runs with changing frequencies throughout the day. The connection between travel choices and time is two-fold; (1) at the demand side, travellers choose their departure time, and (2) at the supply side, the state of the transport system is dynamic.

Firstly, travellers choose the time they travel, and each traveller has different preferences related to departure and arrival times. The bottleneck model of Vickrey (1969) is the first to account for a preferred arrival time. Van Amelsfort (2009) provides an in-depth analysis of the responses of drivers towards time-varying pricing schemes in a discrete choice framework. Distinctions are made between early departures, late departures, early arrivals, and late arrivals, and the choice model parameters have been derived from a stated choice experiment. Bottom line is that it is possible to capture the change in departure time of travellers under time differentiated pricing measures.

Secondly, conditions within the transport system change over time, and past conditions influence conditions in the near future. The foremost important example hereof is congestion that builds up and dissolves over time. It is this congestion that causes delays, and therefore it is important to capture queue dynamics on the road. In addition, other effects, such as emissions, also highly depend on time-varying road conditions.

The importance of the time dimension for planning models is therefore also twofold. The travellers are able to change their departure time in response to prices, and this should be reflected in the framework. This choice depends on their preferred arrival and/or departure time. And secondly, travel times are only realistic when they are based on a proper representation of how queues build up over time. Section 3.1 provides more discussion and references on the different models (i.e., static and dynamic) for congestion build up to determine travel times.

Note that in the literature the dynamics are categorized in within-day dynamics (that contains the two aspects in this section), and day-to-day dynamics. The latter relates to the learning of travellers and corresponding equilibrium concept. See Tsekeris and Voß (2009, Section 3.2) for a discussion on transport pricing with the within-day and day-to-day classification. Both dynamics can be incorporated in strategic planning models by using equilibrium concepts.

2.2.7. Transport Modes

Chapter 1 already described that transport pricing can cause travellers to switch to alternative modes. Some transport pricing studies incorporate no other modes than cars, and thus omit an important effect. In the Netherlands, it is standard practice to incorporate all modes in the strategic planning models. In addition, prices for the different modes interact with each other. Therefore, an integrated approach with a multi-modal transport system representation and pricing measures for all modes are desired.

A multi-modal network model with private and public transport pricing is presented by Hamdouch et al. (2007). Mode choice is modelled with a binomial logit model. The system optimum only considers the net user benefit of public transport users. Several methodologies for choosing valid tolls are discussed.

2.2.8. Network Representations

Economic approaches to road pricing often represent transport networks with parallel or serial links. Each link then represents a different route, of a different mode. This simplification allows the analyst to obtain analytical results, but substantially reduces the realism of the study. Only very basic relations between travel demand and 'infrastructure' supply can be captured (i.e., only direct mappings). However, the most important attribute for travellers to base their decisions on is travel time. Delay is indeed a imbalance between demand and supply, but these delays are a result of complex traffic movements with a highly dynamic nature and with important network effects. The latter includes for example that bottlenecks in one area can influence travel times in other areas due to spillback of congestion. Basic relationships between demand and supply do not capture network effects, and will not assign delay to the actually affected travellers. It is too simplistic to represent routes with serial or parallel links as done in some economic approaches, because they will not result in realistic travel times.

2.2.9. Charge Collection Technologies

A final important aspect is the way the charges are collected. While toll booths were inevitable in the past, current technological advances make almost any type of measure practically feasible. Vehicles can be equipped with tracking and identification devices, and road-side systems allow electronic tolling.[2] Public transport operators switch to electronic fare collection systems; for example, in the Netherlands travellers check-in and check-out of the system with a 'chip-card' from which a kilometre-based fare is automatically deducted.

Such advanced systems allow spatial, temporal and individual differentiation of prices, and they can even be dependent on prevailing (traffic) conditions of the transport system. So, the technology itself is not an impediment for innovative transport pricing. However, such technologies have their price tags, and they bring privacy issues since the systems may track and store individual trips and travel patterns. These factors are an important constraint in the decision making process of the stakeholders. The political and public support is strongly influenced by costs and privacy concerns.

Although collection technologies cost money and yield privacy concerns, this dissertation does not compare different collection technologies. It is assumed that for the considered pricing measures some collection technology exists with an acceptable solution for the privacy issues. Fixed implementation and running costs can be added to the objective of the corresponding stakeholder. A recent cost-benefit analyses for several congestion charging schemes

[2] An overview of road congestion pricing technologies can be found in (de Palma and Lindsey, 2011)

in the Netherlands (Hilbers et al., 2015) provides estimates of the costs of different types of charge collection technologies; also, advantages and disadvantages of different technologies are discussed by Hilbers et al. (2015).

2.3. Pricing Measures

There are plenty of transport pricing measure types. This section presents an overview with a brief explanation of the different measures. They are classified in traditional and innovative measures. The traditional measures are widely implemented and they have been introduced at least a century ago. The innovative pricing are generally more elaborate in design, and they have been introduced only recently, in the last few decades. In general, innovative pricing measures are better capable in influencing travellers' behaviour towards a certain direction. That is also the main reason why policy makers consider them more and more frequently. Hensher and Bliemer (2014) refer to this as choice-pricing and no-choice-pricing.

2.3.1. Traditional Pricing Measures

Periodical Registration Fee Road authorities charge vehicle (usually car and truck) owners for the use of the vehicle. This is normally done with a periodical fee, that allows the vehicle to be used on public roads. This measure is very traditional in the sense that very little differentiations can be made. In the Netherlands there is only a differentiation to vehicle mass and fuel type. The measure is not suitable for paying per usage. Also, no time and spatial differentiations are possible.

Fuel Excise Tax Another frequently deployed measure is the fuel excise tax. This levy paid at the petrol station should, as all excise taxes do, discourage fuel usage. They are generally based on two motives, namely (1) to reduce environmental harm, and (2) to generate revenues. The advantages of fuel excise taxes are that it is a pay-per-usage charge, and that there is a direct relation between fuel consumption and emissions. Fuel excise taxes differentiate between fuel type and indirectly to vehicle type. On the other hand the fuel excise tax can neither differentiate with respect to time-of-day nor location, and is therefore less suitable to reduce congestion.

Toll Roads There is a long history of toll roads that dates back 2,700 years. Toll has been levied on the Roman roads and in other ancient empires. Tolls can either be for financing the construction and maintenance of roads, or for generating revenue on private roads. They run frequently over bridges and through tunnels since those come with high investments. Tolls naturally differentiate by location, and occasionally the toll depends on the time-of-day, or even on the level-of-service. In the Netherlands there is toll in the Westerscheldetunnel (N62) and in the Kiltunnel (N217); besides those there are also two tunnels with a shadow toll[3].

[3] Shadow tolls appear in public private partnerships, where the government pays the private party per user/vehicle of a road for the construction and maintenance. Shadow tolls do not influence travellers' be-

Public Transport Fares Tickets are required to use the public transport system. Fares depend on the distance travelled (with a few exceptions: e.g., the New York subway) and sometimes on the time-of-day. Depending on the charge system, the fare structure can be constructed such that crowding is minimized. Public transport services can generally be exploited more cost efficiently when its usage is high throughout the day, instead of confined to peak hours.

Paid Parking Parking in and around dense commercial areas is often not free. There is either a flat charge or a price per time unit spent. On-street parking is usually paid upfront, while parking in garages is paid afterwards. The price is usually different depending on the day of the week and the time-of-day (e.g., parking is free during the night). Charging for parking contributes to keep commercial areas accessible.

2.3.2. Innovative Pricing Measures

Kilometre Charge In the Netherlands, the kilometre charge is the most well-known innovative measure with a long history. The kilometre charges is, as the name suggests, a pricing measure that is charged per driven kilometre. Since it generally uses on-board technology, it can differentiate in a lot of manners (e.g., for road type, vehicle type, time-of-day, user). This provides plenty of possibilities in designing optimal pricing schemes. The largest disadvantages are the required advanced technology and the privacy related issues. This nourishes its lack of public and political support.

Peak Avoidance Rewarding Several peak avoidance rewarding schemes have been implemented successfully in the Netherlands. This pricing measure is often temporary (e.g., to reduce demand during extensive roadway maintenance) and applies to regular road users. A priori regular users of a road are identified using license plate recognition, and those are offered a reward when they avoid the peak hours. The traveller can choose to stay-at-home, switch modes, change their departure time, or possibly change route. The results of the peak avoidance schemes are promising (Knockaert et al., 2010). The main disadvantage of such schemes is that they are not a structural solution since it does not generate money, but rather only costs money.

Pay-as-you-Drive Insurance Instead of regular periodical vehicle insurance fees, more and more usage-based insurance contracts become available. This is advantageous for vehicle owners with a low travel demand. This type of pricing aims at directly internalising the unsafety effects of driving. Prices can be differentiated between road types, where relatively safe highways have a lower kilometre fee than urban roads. When the fee is, besides milage and road type, also based on driving behaviour, such as speeding or night-time driving, it is sometimes called Pay-how-you-Drive insurance. Such contracts can also differentiate between different driver types/'classes'. In-car technology is required for advanced pay-as-you-drive

haviour and are therefore out of the scope of this thesis.

insurance systems. In several countries, including the Netherlands, such insurance contracts are available.

Cordon Charge Since economic activity is concentrated around commercial areas, destinations of travellers are similar and therefore congestion is concentrated around these areas. A cordon charge (i.e., a fee to enter such an area by car) provides an incentive to use different types of transport. This is especially effective when good public transport alternatives are present. When the charge is time differentiated the departure choice can be influenced. Stockholm and London are well-known European examples of cities with a cordon charge.

2.4. Effects

Parry et al. (2007) describe the magnitude of external effects of road traffic. The most important ones in the United States and their marginal external costs (in US$ cent/km) are: congestion (6.21), accidents (3.11), local pollution (1.24), oil dependency (0.37), and greenhouse warming (0.19). They argue that "Electronic road pricing offers the only real hope of addressing relentlessly increasing gridlock, while encouraging a transition to mileage-based insurance would improve highway safety more effectively [than higher fuel taxes]". Note that the fuel related externalities only account for a small amount of the total marginal external cost. This is an argument to use innovative differentiated pricing measures instead of a fuel excise.

In this thesis the notion of effects of the transport system and transport pricing is considered in a broad sense. Everything that can be valued externally from the transport system is considered as an effect. Besides the general collective (social) costs as congestion and emissions – the external effects–, also the costs and benefits of individuals and stakeholders are called effects. The reason for this is that the stakeholders' objectives can be a set of effects. In the next paragraphs several private and collective effects are mentioned.

Emissions are important external effects. The emission of Carbon Monoxide (CO) and Carbon Dioxide (CO_2) influence greenhouse warming. Local air pollution is caused by Nitrogen Oxides (NO_X) and Particle Matter < 10 micrometre (PM_{10}). Wismans (2012) uses the ARTEMIS model to quantify the emissions based on network conditions. Such models are included in strategic planning models to connect the predicted traffic conditions with the predicted emissions.

Some forms of transport are ran by private companies (e.g., private (toll) roads, public transport operators), and those companies have a profit motive. The revenues from the pricing mechanisms are therefore of great importance for these companies, and considered as effects in this thesis. Private stakeholders usually have profit maximization as their objective. The revenue is influenced by the price and the demand for the product, which is captured within the transport pricing framework. So, the profit depends on the market effects.

Delay is another important external effect. It is probably, together with cost, the most important dis-utility for travellers and furthermore influences the accessibility of cities and areas. The dis-utility stemming from delay is also higher than that of normal travel time (Abrantes and Wardman, 2011). Related to delay, and probably equally important, is the unreliability of travel time (i.e., the variation of travel time over days). See Small (2012) for a review on the valuation of travel time, and see Wardman et al. (2012) for a European-wide meta-analysis of several types of values of time.

Social welfare is a measure that combines the costs and benefits of the whole population. The individual costs consist of the dis-utility from making the trip with travel time and travel costs as the most important components. The individual benefits in the transport market are not directly related to transport, but come from the activities at the different locations. So, only when travellers decide not to travel, the benefits are reduced. This relates to the fact that mobility is a derived demand. Besides indivudual costs and benefits there are collective costs and benefits. Dependending on how the revenue from transport pricing is recycled, population will benefit from these investments. The transaction associated with paying for the pricing measure often does not influence the social welfare, since the cost for the traveller is a revenue that cancels out. An example of collective (i.e., social) costs are emissions and noise. The social welfare measure should be the sum of all these costs and benefits.

2.5. Bi-Level Framework Formulation

This dissertation adopts the for transport pricing widely used bi-level modelling framework. The two levels correspond with the decision making processes of respectively the stakeholders and the travellers. The structure allows flexibility in the interpretation of both layers. The literature sometimes refers to bi-level (optimization) models as Stackelberg games or mathematical programs with equilibrium constraints. This section defines the required properties of both levels such that the important transport pricing aspects of Section 2.2 can be captured. Implementations of the lower and upper level are respectively presented in Chapters 3 and 6.

The upper level represents the stakeholders (i.e., the decision makers or price setters). These stakeholders all pursue their goals, and they are aware of each others' existence. The goal of the upper level is therefore to grasp the tactics and strategies of the stakeholder, and to discover which pricing scheme is the result of this 'game'. It is not obvious how the interaction between stakeholders will unfold, but game theoretical principles provide a direction for this. Also, it is not easy to quantify the exact objectives of the stakeholders. However, it is often clear what the purposes of stakeholders are, and these can generally be expressed in terms of effects. In this chapter some basic notions and requirements of a game theoretical (quantitative) approach to this problem will be introduced.

The stakeholders are assumed to have an objective based on the several effects (as introduced in the previous paragraph). Next, they try to adjust their strategy – in terms of a price for their 'travel product' – such that their objective will be optimized. To do this they have to take the responses of the travellers into account, as well as the actions of the other stakeholders. This problem formulation is much more complex than the traditional transport pricing problems with only one stakeholder since multiple objectives exist that can conflict.

The lower level describes the behaviour of travellers and their interaction with the transport infrastructure. This level is used to quantify the effect of the responses of the travellers towards the pricing scheme. Those responses will be revealed by adjustments in their travel choices. These changed choices will lead to a different demand in the transport system, and that will be expressed in different traffic flows. Finally, the traffic flows pass through the infrastructure

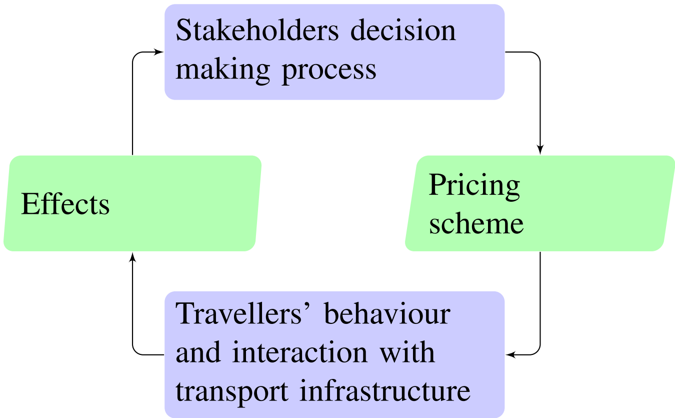


Figure 2.2.: Bi-level model: schematic overview

which determines the level of the effects in which the stakeholders are interested. The lower level model quantifies all these steps, which is not straightforward since the travellers themselves base their choices also on the effects. Therefore, feedback is introduced and the result of the lower level is an equilibrium.

The two levels interact by means of prices and effects, so they are mutually dependent. This is shown in the schematic overview of the bi-level model, see Figure 2.2. This dependency implies that is not possible to determine the optimal pricing scheme for each stakeholder with a 'one-shot' calculation. The resulting effects for multiple pricing schemes need to be determined with the lower level model to be able to have a final result of the upper level. Besides, the lower level is a simulation model which means that few analytical properties are available. Before these issues are discussed in more detail, the bi-level model is presented mathematically. Although this means moving from the comfort of a text-only manuscript, this allows a rigorous discussion on the properties of the problem formulations, and is a reference for the requirements of both levels.

First, the notations for the overall mathematical formulation are defined, followed by the problem definition; finally, a simple example illustrates the notation and problem formulation. The set of pricing measures[4] is denoted with $\mathcal{P}$ and denote the set of effects with $\mathcal{E}$. Each pricing measure $p \in \mathcal{P}$ has a set of feasible prices $\Pi_p \subseteq \mathbb{R}$ and a particular price level is written as $\pi_p \in \Pi_p$.[5] A combination of price levels for each price is denoted with price vector $\boldsymbol{\pi} = \left\{\pi_p \middle| p \in \mathcal{P}\right\} \in \Pi$, where $\Pi = \prod_{p \in \mathcal{P}} \Pi_p$.[6] Each effect $e \in \mathcal{E}$ has a feasible effect level set $\Gamma_e \subseteq \mathbb{R}$ and a particular effect level is written as $\gamma_e \in \Gamma_e$. A combination of levels for each effect is denoted with effect vector $\boldsymbol{\gamma} = \{\gamma_e | e \in \mathcal{E}\} \in \Gamma$, where $\Gamma = \prod_{e \in \mathcal{E}} \Gamma_e$. Futhermore, let $\mathcal{S}$ be the set of stakeholders, and let $H_s : \Gamma \to \mathbb{R}$ be the objective function (assume higher is better) of stakeholder $s \in \mathcal{S}$. Also assume that the unit of each stakeholder's objective function is equal, for example, a monetary unit or utility. Finally, let $E : \Pi \to \Gamma$ be a function that represents the

[4] The 'complete' pricing measure of a stakeholder can consist of multiple of these elementary pricing measures, e.g., on-peak and off-peak prices. This will be become formal in Chapter 6

[5] π_p is a tariff in some monetary unit.

[6] Here $\boldsymbol{\pi}$ can be interpreted as a pricing scheme and Π as the set of all feasible pricing schemes.

TA and thus the travellers' responses to the price and their interaction with the infrastructure.

The most elementary optimization problem associated with stakeholders setting a price by optimizing their objective is given by this Multiple Stakeholders Problem (MSP):

$$\max_{\pi \in \Pi} \{H_s(\boldsymbol{\gamma}) | s \in \mathcal{S}\} \quad \text{subject to } \boldsymbol{\gamma} = E(\boldsymbol{\pi}). \tag{MSP}$$

This mathematical program is a non-straightforward bi-level problem. Since multiple objective functions are considered simultaneously the solution is ambiguous, since there is no single (or general) solution to multi objective optimization problems. In addition, the constraint (i.e., lower level) is a TA problem which is a non-straightforward mathematical problem itself. The following example with two stakeholders illustrates the use of variables, as well as the MSP.

EXAMPLE: National government and train operator

The set of stakeholders consists of the national government GOV and the train operator TR-OP:

$$\mathcal{S} = \{\text{GOV}, \text{TR-OP}\}$$

They consider following effects: loss hours LH (non-negative, in hours), total collected toll, denoted TOLL (of which the level γ_{LH} is non-negative and in €), and the profit of the train operator PROFIT (of which the level γ_{PROFIT} is in €):

$$\begin{aligned}
\mathcal{E} &= \{\text{LH}, \text{TOLL}, \text{PROFIT}\} \\
\Gamma_{\text{LH}} &= \mathbb{R}_+ \\
\Gamma_{\text{TOLL}} &= \mathbb{R}_+ \\
\Gamma_{\text{PROFIT}} &= \mathbb{R} \\
\Gamma &= \Gamma_{\text{LH}} \times \Gamma_{\text{TOLL}} \times \Gamma_{\text{PROFIT}} \\
&= \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}
\end{aligned}$$

The government has the objective to maximise welfare and assumes a value of time of 15 €/hour, and the train operator maximizes its profit:

$$\begin{aligned}
H_{\text{GOV}}(\gamma_{\text{LH}}, \gamma_{\text{TOLL}}) &= -15\gamma_{\text{LH}} - \gamma_{\text{TOLL}} \\
H_{\text{TR-OP}}(\gamma_{\text{PROFIT}}) &= \gamma_{\text{PROFIT}}
\end{aligned}$$

The pricing measure of the government is a time differentiated kilometre charge (denoted KM-ON for on-peak and KM-OFF for off-peak, both their levels in €/km). The pricing measure of the train operator is changing the ticket price FARE with a percentage change. The kilometre charge can only be in whole cents with a maximum of twelve cents, while the fares can be in- and decreased by any percentage, as long as they not exceed ±10%, thus:

$$\begin{aligned}
\mathcal{P} &= \{\text{KM-ON}, \text{KM-OFF}, \text{FARE}\} \\
\Pi_{\text{KM-ON}} &= \{0; 0.01, \ldots, 0.11; 0.12\} \\
\Pi_{\text{KM-OFF}} &= \{0; 0.01, \ldots, 0.11; 0.12\} \\
\Pi_{\text{FARE}} &= [-10\%, 10\%] \\
\Pi &= \Pi_{\text{KM-ON}} \times \Pi_{\text{KM-OFF}} \times \Pi_{\text{FARE}} \\
&= \{0; 0.01; \ldots; 0.11; 0.12\} \times \\
&\quad \{0; 0.01; \ldots; 0.11; 0.12\} \times [-10\%, 10\%]
\end{aligned}$$

The corresponding MSP for the government and train operator becomes:

$$\max_{(\pi_{\text{KM-ON}}, \pi_{\text{KM-OFF}}, \pi_{\text{FARE}}) \in \Pi} \left\{ \begin{array}{c} -15\gamma_{\text{LH}} - \gamma_{\text{TOLL}} \\ \gamma_{\text{PROFIT}} \end{array} \right\}$$

$$\text{such that } (\gamma_{\text{LH}}, \gamma_{\text{TOLL}}, \gamma_{\text{PROFIT}}) = E(\pi_{\text{KM-ON}}, \pi_{\text{KM-OFF}}, \pi_{\text{FARE}})$$

Here E is not yet specified, but keep in mind that any price level can influence any effect level. Then it is already clear that there is no straightforward solution to this problem because the interaction between the government and the train operator is not yet specified. When the train operator adjusts the price such that their profit is maximal, the government can respond by changing the kilometre charge. Since that changes the profit of the train operator, its fare is not necessarily optimal any more.

Methodology is required that can solve the multi objective character of (MSP) on Page 30 by including the interaction between stakeholders, where in addition the TA model represented by function E contains plenty of travel behaviour and interaction between demand and supply. The next sections therefore introduce both modelling levels further, and formulates requirements for them.

2.5.1. Upper Level Specification & Requirements

It is far from straightforward to predict the outcome of the decision making process of one stakeholder. Although strategic planning models can indicate optimal policies, irrational behaviour and politics are by definition difficult to capture in a mathematical model. By introducing more stakeholders, and by observing that stakeholders respond to each other's action makes the modelling even less tractable. In addition, there is an ongoing discussion on the practical implementation of planning support systems. That started exclusively positive when the models and the computing power for large scale models became available; Clune et al. (1999) wrote "for transport planners themselves to devise prices [...] anywhere near the best would take (i) great effort, (ii) great insight, (iii) a long period of time, and (iv) luck", and they state that the 'tools' would be able to take over this task easily. At present, the view on this is more balanced since it turns out that tools do not take over all the planners' tasks, and the whole planning process should be analysed (i.e., by moving some emphasis from the tooling to the whole process) (Pelzer et al., 2014, 2015).

The focus of this dissertation is also on the (mathematical) modelling of transport pricing, and rationality of the stakeholders is still assumed. However, the main purpose of the upper level is to support the decision making process in a more elaborate manner, by providing arguments for the stakeholders, and by proposing solution concepts. The traditional notion of an unambiguous single solution to the transport pricing issue will be abandoned. Instead, paradigms will be allowed that provides different ranges of solutions that can be based on different assumption on the interaction between the stakeholders.

The field where mathematical models and rational decision makers intersect is game theory. This expansive field has multiple frameworks and theories about competing and cooperating decision makers that fit the MSP. Therefore, a game theoretical approach is useful for the upper level. The main challenge here is to numerically resolve the multiple objective functions while it reflects real stakeholders' interactions. The theses of Joksimovic (2007) and Ohazulike (2014) are the first to explore the field of game theory in a multi-stakeholder setting. Based on the discussion in Sections 2.2.1 and 2.2.2 two requirements for the upper level are formulated:

- *Rational stakeholder behaviour*
  Stakeholders' preferences should be reflected by their objective function. They will act only in favour of their objective. In case of negotiations and/or cooperation, the used strategy of each stakeholder is rational, meaning that they each optimize their own objective.
- *Reflection of different cooperation formations*
  The upper level should be able to analyse different mutual attitudes of stakeholders. Co-operative and competitive behaviour is considered as endogenous. This allows analysis of the price of competition.

2.5.2. Lower Level Specification & Requirements

The lower level TA function E in the MSP represents a complicated process with travellers' choices and traffic propagating over the network. The analyses of transport pricing aspects in Section 2.2 lead to the observation of important properties of the travellers and the transport system. These are translated to requirements for the TA model. In addition, general requirements, that also hold without the pricing application, for strategic planning models exist (see Bliemer et al., 2013). This leads to the following list of requirements:

Travel Choices

- *Incorporation of differences in travellers' responses*
  There is a large difference in choice behaviour between different (types of) travellers. The model should address this by either explicitly grouping travellers according to similar characteristics, or by taking the taste heterogeneity implicitly into account by random variates in the model.
- *Incorporation of different travel time types*
  Travellers evaluate their time differently per mode, but also, one hour of travelling in

congested conditions is experiences as a larger burden than one hour of travelling in free flow conditions. Furthermore, waiting, access and egress times for pubic transport should have different valuations.

- *Capturing overlap*
  When two alternatives share the same characteristics (e.g., the modes or time-of-days are equal) or when there is even physical overlap (i.e., road segments), the choice preference of a traveller will be similar for these alternatives. The red-bus blue-bus is the textbook example of this correlation. It is required that the choice model accounts for this overlap.
- *Choice opportunities*
  In planning models the possible responses of the travellers are reflected by their choice set. The larger this set, the more diverse the responses can be. In most transport systems public transport or slow modes are an attractive alternative, and so is deviating from the preferred departure time. The framework should be able to reflect all relevant alternatives present in the transport system in the choice set.
- *Individual choices based on alternatives' properties*
  Choices are made based on different observable properties (or attributes) of alternatives. Each traveller values these alternatives' properties in his/her own manner. Therefore, it is required that the model determines the choice probabilities based directly on these properties. Furthermore, travel cost and travel time of a trip are the most important properties related to pricing, since they are likely to change under pricing regimes. In addition, it should also be possible to capture travel time reliability of a trip. The model has to address these aspects based on physical infrastructure and on its prevailing performance given a certain pricing scheme.

Traffic Phenomena and Network Representation

- *Proper identification of bottlenecks*
  Bottlenecks are locations in the road network where the travel demand is higher than the capacity of the infrastructure. These occur usually at discontinuities in the transport network, which are located at nodes in the abstraction of the network. Node models capture the traffic phenomena and conditions at nodes and are therefore a strict requirement. Without a node model the location and severity of congestion cannot be determined.
- *Queue propagation by shock waves and with spillback*
  Bottleneck locations and conditions alone are not sufficient to represent traffic, because queues occupy physical space. Therefore, the spatial dimension of congestion has to be determined. Kinematic Wave Theory (KWT) is an elegant and simple theory that allows propagation of traffic conditions over links (details follow in Section 3.3.1). When congested conditions reach the beginning of a link, the node model – with new input – can determine the direction and severity of the spillback. A node model combined with KWT therefore captures important traffic flow phenomena, such as queue growth and spillback. Computational efficient methods exist that implement KWT. A traffic state that represents traffic flow, speed, and density at every location in the network is the provided output.
- *Representative travel time calculation*

Travellers make decisions based on their foreseen travel time. One would say it is rather simple to determine the travel times when speeds are known, and this is true when the traffic propagation is performed with KWT as stated in the previous requirement. However, this is listed as a separate requirement since plenty of standard models in Static Traffic Assignment (STA) (definition follows in Section 3.1) and other heuristic methods cannot determine these speeds appropriately under congested conditions, and thus cannot report representative travel times.

- *Representative (external) effects quantifications*
  For the stakeholders holds that they want to make decision about pricing measures based on reliable estimates of effects. Quantities like air and noise pollution can – just like travel time – be determined with traffic conditions as flow, density and speed. However, just like travel time, these have to be realistic to get a good result. Due to the importance of effects, also this is stated as a separate requirement.
- *Varying network conditions over the day*
  The (average) network conditions change over the periods within the day. It is required to capture this variability over the day by having representative time periods (e.g., hourly intervals or different peak hours). Since the second-to-second or minute-to-minute variations are not relevant for transport pricing, a continuous or strongly discretized approach is not required.

PART I.

TRAFFIC ASSIGNMENT

CHAPTER 3.

QUASI-DYNAMIC TRAFFIC ASSIGNMENT

The lower level model of the transport pricing framework (Chapter 2) is an assignment model in which mode, route and departure time choice of travellers have to be captured. This chapter implements the lower level by means of a Quasi-Dynamic Traffic Assignment (QDTA) framework that combines important traffic dynamics from Dynamic Traffic Assignment (DTA) and the computational efficiency of Static Traffic Assignment (STA). The mode, route, and departure time choices are combined to a single choice decision. This choice model can correct for similarities between alternatives (e.g., route overlap) and is feasible for application on real transport networks.

The Quasi-Dynamic Network Loading (QDNL) part of this chapter uses the theory presented in the following papers:

- Bliemer, M. C., Brederode, L., Wismans, L., and Smits, E.-S. (2012).
  Quasi-dynamic traffic assignment: static traffic assignment with queueing and spillback.
  In *The Transportation Research Board (TRB) 91st Annual Meeting, Washington DC, January 22-26, 2012 (paper no 12-0358).*, pages 1 – 24
- Bliemer, M. C., Raadsen, M. P., Smits, E.-S., Zhou, B., and Bell, M. G. (2014b).
  Quasi-dynamic traffic assignment with residual point queues incorporating a first order node model.
  Transportation Research Part B: Methodological, 68(0):363–384

This chapter starts with a concise introduction to different Traffic Assignment (TA) models in the light of the specifications given in Section 2.5.2. The different types and their advantages and disadvantages are given. Afterwards, QDTA is introduced which is a TA model that has a good balance between realism and efficiency for transport pricing applications. Elements hereof are a choice model which uses theory from Chapter 4 and QDNL which relies largely on node models. The behaviour of the latter is analysed in Chapter 5.

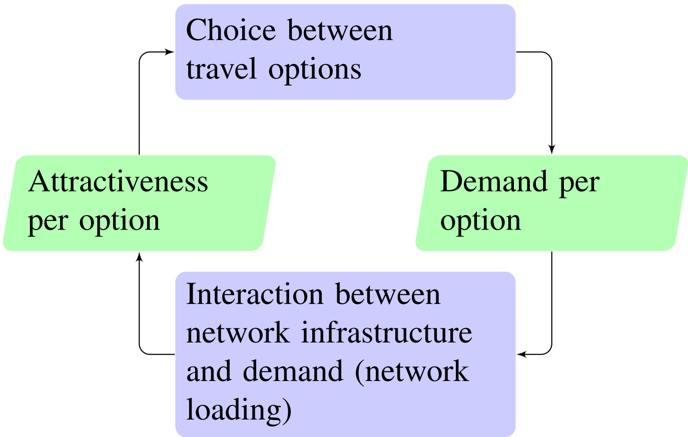


Figure 3.1.: Interaction between choices and network loading

3.1. Traffic Assignment Introduction

Traffic Assignment models are an important modelling tool to estimate current and predict future traffic flows and conditions of transport systems. Traditional TA models are used to describe travellers' route choice behaviour and the implications these choices have for traffic conditions. To determine the intensity, location and duration of congestion, traffic flows are propagated over the network. Applications of TA models in general include strategic transport planning (e.g., pricing of course), the design of networks, but also assessment of traffic management techniques (e.g. ramp metering or dynamic speed limits) with more advanced approaches. This section discusses which TA models are suitable for strategic transport planning. Typical TA models consist of two components, (1) a choice model for modelling travellers' decisions and (2) a network loading model to propagate traffic over networks. There is feedback since the network loading determines the attractiveness of the different choice options. Figure 3.1 depicts these interactions. Note that all choices related to the consumption of mobility can be included here, so they are not restricted to route choice, and that the specification of the network loading is key to the type of TA that is obtained.

Wardrop (1952) is the originator of basic principles in TA. He outlines in his seminal paper that "[...] speed is a function of flow, so that redistribution of traffic upsets the pattern of speeds. The problem is to discover how traffic may be expected to distribute itself over alternative routes, and whether the distribution adopted is the most efficient one.". This problem has inspired generations of transport scientists, and there is a huge repertory of solution methods. Wardrop (1952) himself already provided the two criteria on which the distribution of routes can be based, these a currently known as the user equilibrium (also known as the Wardrop equilibrium) and the system optimum. "The journey times on all the routes actually used are equal, and less than those which would be experienced by a single vehicle on any unused route." is the equilibrium in which no user can be better off by unilaterally changing routes. He describes system optimum as when "The average journey time is a minimum.".

The neat connection between pricing and Wardrop's criteria is that by pricing roads in the network, the system can move from the user equilibrium to the system optimum. This reasoning is analogous to marginal cost pricing (see Section 2.1), and indeed (Beckmann, 1965) showed that when on every link on the network the marginal cost is levied, the system moves from user equilibrium to system optimum. This is no reason for jubilation yet, since it is very difficult to determine the (social) costs precisely per link. In addition, charging the marginal costs per link is practically infeasible. This is due to the fact that no simple analytical model exists that satisfies the requirements stated in Section 2.5.2. These and other 'real-life' constraints bring forward second-best pricing, and it becomes key to analyse the user equilibria under such second-best prices (see also Section 2.1).

To start with the analysis, consider the network loading procedure. In real traffic on road networks several phenomena can be identified, network loading models should be able to reproduce these phenomena. Important phenomena include (1) congestion build up (location, intensity and speed), (2) shockwave propagation, and (3) spillback over nodes. Presence of these phenomena in a TA allows for a good assessment of the transport systems. Other phenomena include traffic instability, hysteresis, traffic heterogeneity, and lane change behaviour. These are considered less important in network loading for TA models, because these phenomena act on a high level of detail (i.e., individual vehicles), and do not affect the representative traffic conditions too much. Rule of thumb is that the more phenomena are modelled, the more demanding the solution methods are – hence the longer calculation times become. Practitioners should therefore consider the level of detail required for their application.

The foremost classification in network loading distinguishes between *static* and *dynamic* models. The static models typically consider a travel time function[1] for every link; an increase in flow leads to an increase in travel time. The static equilibrium model with such functions was defined by Beckmann et al. (1956). This has been the standard for equilibrium models for the following half-century. Nonetheless, Beckmann et al. (1956) state "The notion of static equilibrium of flow in a network may be thought somewhat limited because of the noted periodicity of traffic during the day, week, year, and perhaps business cycle.". They furthermore had the foreseeing insight that "While it is not difficult, by attaching time subscripts to the flow variables, to write down formally the equilibrium conditions [...] for a dynamic model, this merely makes the analysis more complicated without explaining much that is new.". This is foreseeing since it has been attempted to add the subscript, but these results never described physical queues adequately (Bliemer et al., 2014a, 2017). The user equilibria based on static models have the advantage that they have nice properties; existence and uniqueness can be proven, and efficient sophisticated solution methods are available[2]. On the contrary, they lack realism by allowing flow to exceed capacity, and by completely omitting queue build-up.

Dynamic Network Loading (DNL) models do provide the required realism; they can roughly be classified in microscopic and macroscopic models. The microscopic models simulate each vehicle separately, while the macroscopic models consider traffic as a continuous flow. Since microscopic modelling requires much computational power and memory storage, applications on large networks are infeasible. The advantage of microscopic simulators is that be-

[1] Which is usually the well-known Bureau of Public Roads (BPR) link delay function.

[2] A good model to use is the TAPAS method introduced by Bar-Gera (2010); Bar-Gera et al. (2012), as it is efficient and has unique resulting route flows. Also, the algorithm by Dial (2006) provides good performance. Recent overviews on static models and their extensions can be found in (Bliemer et al., 2014a,b, 2017).

haviour can be specified for each vehicle. Macroscopic models have a lower level of detail, but have more efficient solution methods. This makes them more suitable for transport models with large scale networks. Within macroscopic models, multiple approaches exist, the by far largest school of them relies on the Kinematic Wave Theory (KWT) developed by Lighthill and Whitham (1955); Richards (1956). They describe how traffic propagates as kinematic waves through the network. Traffic characteristics are flow and density which are related with the law of conservation of vehicles. Since KWT is of significant importance, an introduction to the main theory is provided in Section 3.3.1. For a full genealogy of traffic flow models see van Wageningen-Kessels et al. (2014)

In this thesis, TA is not restricted to how travellers are distributed over routes as Wardrop (1952) describes. In addition, the distribution over time and modes is included. This implies different choices for the travellers; they can depart earlier or later, or might switch modes. Since it is utopian to find and impose the system optimum with the required level of detail (from Section 2.5.2) under three choice dimensions, the focus of the TA in this thesis is to determine the user equilibrium based on the *simulation* of travel times and other conditions, given a particular pricing scheme.[3] Allowing travellers to distribute over time and modes has implications for network loading. These will be touched upon briefly, before the actual traveller's choice behaviour will be discussed.

In the developed model, travellers can choose between different modes which increases complexity of transport models. The 'dynamics' of road traffic and public transport are completely different. On the road, delays are due to congestion, while in public transport delays are not necessarily a direct consequence of an increase in demand. However, crowding in public transport makes this option less attractive, and crowding is a direct consequence of a high demand. This requires different approaches per mode. Contrary to road networks, there is no queue build-up in public transport networks with dedicated infrastructure. The public transport system is based on services with either a frequency or fixed time schedule, and crowding reveals itself by discomfort from overcrowded vehicles, and/or additional delay due to waiting for the next vehicles if the current vehicle is full, or in the case of disruptions. Buses can suffer from both crowding and congestion effects. de Cea and Fernández (1993); Spiess and Florian (1989) provide user equilibrium-like models with feedback from flow; for an overview of public transport related assignment see Ceder (2007).

Vickrey (1969) was the first to identify that travellers will distribute over time because they have different preferences for their departure and arrival times, but also because they avoid delays. Traffic conditions differ before, during and after peak hours. Fully dynamic models provide these fluctuations, while static models cannot capture them. This section is titled *quasi-dynamic* traffic assignment, because it does allow travellers to distribute over time, and the traffic conditions come from KWT, but the conditions represent the averages per time period (which is similar to static approaches). This allows the traveller to choose the time period in which they prefer to travel, but there is no distribution within each time period. Arnott et al. (1990a,b); Vickrey (1969) consider simple models with continuous departure

[3]With milder assumptions, analytical solutions for achieving the system optimum with pricing exist. Yang and Huang (2005) provide an overview of user equilibrium problems in their Chapter 2, and they provide a series of analytical marginal cost pricing solutions using these formulations

times. Huang and Lam (2002) utilize the discrete time approach, but they combine this with a simplified underlying network loading model.

The approaches of Wardrop (1952) and Beckmann et al. (1956) assume that travellers have perfect information about the attractiveness of their options, and that they all confine to the same choice process. However, different travellers perceive different routes in a different fashion. In other words, they might not know the exact travel time, and they can have an individual preference towards time evaluation. Stochastic models provide a method to capture this heterogeneity and uncertainty. The attractiveness of each option is sampled from a distribution, and the traveller uses some choice mechanism to choose one of the options. With only few exceptions, the Random Utility Maximization (RUM) framework is used to capture these choices, which assigns a choice probability to each travel option. Section 3.2.1 introduces this theory. User equilibria with non-deterministic choices are called stochastic user equilibria.

Stochastic user equilibria that use static network loading models were introduced by Daganzo and Sheffi (1977); Dial (1971); Fisk (1980). With logit-based choice models, the stochastic user equilibrium has nice properties (e.g., existence and uniqueness), and such an equilibrium can be found efficiently.[4] At the other end of the spectrum, user equilibria with a DNL model can only be approximated by means of simulations. These simulations try to find stable (i.e., fixed) points in the TA system as described in Figure 3.1. Since they are non-unique, the meaning of such a point is still an open question. TA models based on static network loading are called Static Traffic Assignment (STA) models, and those based on DNL are called Dynamic Traffic Assignment (DTA). Bliemer et al. (2014a, Section 2) describe STA and DTA models, and show what is available in the grey area in between them. Bliemer et al. (2017) provide a comprehensive overview of TA models for strategic planning. They classify the models with an analogy to genetics. Each model is build up from a spatial gene, a temporal gene, and a behavioural gene.

The QDTA model presented in this chapter, lies in between the static and dynamic approaches. The necessary traffic dynamics provided by KWT are included, but it is used to describe the average traffic state with a time-of-day period. Since no complete travel time profile over departure times is presented, QDTA is not a dynamic model. This simplification allows tremendous efficiency gains compared to complete dynamic models. Compared to static models, the level of realism is significantly larger. The user equilibrium based on QDTA has a neat mathematical problem definition. However, existence and uniqueness properties of this user equilibrium are not proven yet.

3.1.1. Quasi-Dynamic Traffic Assignment Model Framework

This section constructs the delimiters of the QDTA model by defining several model components and their interactions. So, the TA function $E : \Pi \rightarrow \Gamma$, that maps pricing schemes to effects, used in the Multiple Stakeholders Problem (MSP) is specified. Figure 3.2 shows the flowchart of QDTA. The pricing scheme, travel demand, and the transport network are the

[4] Note that this only holds if the choice alternatives are provided a priori, as is the case in this dissertation. If generation of choice alternatives is endogenous in the equilibrium model, an equilibrium does not necessarily exist (Watling et al., 2015).

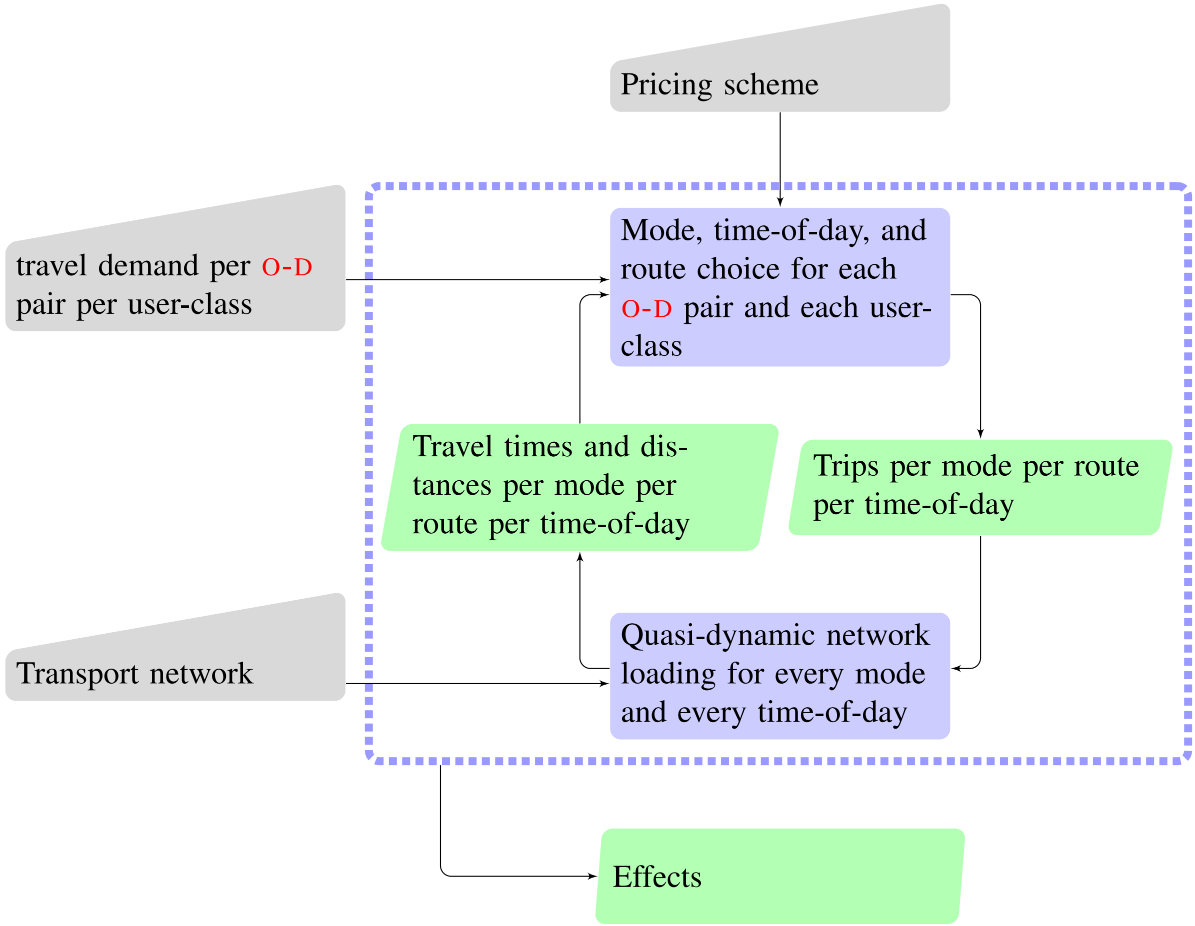


Figure 3.2.: Flowchart of the QDTA model

input, and the quantified effects are the output. Two main processes are identified: (1) the simultaneous mode, time-of-day and route choice, and (2) the QDNL model. The latter is independently executed for each mode and time-of-day combination, which assumes that traffic cannot transfer between modes or time-of-days. The model allows for taste heterogeneity by allowing different preference parameters per user-class. Let $\mathcal{U}$ be the set of user-classes and assume that each traveller belongs to one of the user-classes $u \in \mathcal{U}$.

Next, the main processes are individually discussed in detail. This requires some preliminaries on the notation of the input. Consider a transport network $(\mathcal{N}, \mathcal{L})$ consisting of nodes $\mathcal{N}$, and links $\mathcal{L}$. Trips start at origins $\mathcal{O} \subseteq \mathcal{N}$ and terminate at destinations $\mathcal{D} \subseteq \mathcal{N}$. The set of Origin-Destination (O-D) pairs with positive travel demand is $\mathcal{OD}$, the demand of user-class $u \in \mathcal{U}$ for O-D pair $od \in \mathcal{OD}$ is denoted with $D_{(u;od)} \in \mathbb{R}_+$, and write $\boldsymbol{D}_u = \{D_{(u;od)} | od \in \mathcal{OD}\} \in \mathbb{R}_+^{|\mathcal{OD}|}$ for the O-D matrix for user-class u.

3.2. Mode, Route and Departure Time Choice

This section specifies the four components (i.e., choice set, systematic utility, error distribution, and utility formula) that specify a RUM model, such that it can be used as the choice sub-model of the transport pricing framework. Extensive analyses on the error distributions and utility formulas applied to route choice can be found in Chapter 4. In this section mode and departure time choice are also considered, making it a generalization of Chapter 4, but fortunately the mathematical properties remain equal. Before the application of RUM in the QDTA-model is discussed, a basic introduction to RUM is provided in the following section.

3.2.1. Random Utility Maximization

Discrete choice analysis based on RUM is widespread in transport science. It provides a method to determine choice probabilities for a set of alternatives. RUM is an important ingredient of the QDTA model, since it captures the behaviour of travellers. Train (2009) is an excellent introductory book and reference work on the topic. A discrete choice model requires the analyst to specify four components: (1) a set of choice alternatives, i.e. C, (2) a systematic utility for each alternative based on observed attributes, (3) a randomly distributed set of error terms, and (4) a utility formula that determines how the systematic utility and error term are combined. The latter can for example be additive or multiplicative. Application of the RUM principle on these four components determine the choice probabilities.

Traditional RUM models write the utility U_c for every choice alternative $c \in C$ as

$$U_c = V_c + \varepsilon_c, \tag{3.1}$$

where V_c is the systematic utility and ε_c is the error term. The systematic utility usually is a linear combination of observed attributes of the alternative, for example travel time. The error term contains, amongst other randomness, the unobserved attributes of the alternative. The RUM principle is that the traveller chooses the alternative with the highest probability. The probability that a traveller chooses alternative $c \in C$ equals

$$P(c) = \mathbb{P}\left(U_c \geq U_{c'} | \forall c' \in C, c' \neq c\right). \tag{3.2}$$

The interesting part for the choice modeller starts when a distribution for $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_{|C|})$ is specified. When it is multivariate Gumbel[5] distributed, the logit model arises (McFadden, 1974), and with multivariate normal distributed error terms the probit model arises (Daganzo and Sheffi, 1977; Daganzo, 1979)[6]. The largest difference between them is that only the logit model has a closed-form probability expression, namely

$$P(c) = \frac{e^{V_c}}{\sum_{c' \in C} e^{V_{c'}}}, \quad \forall c \in C. \tag{3.3}$$

While for the probit model the choice probabilities have to be simulated with a Monte Carlo method. Therefore, applications of probit on large-scale networks is not feasible.

[5] To be exact, multivariate extreme value type I, see Section 4.3.1

[6] See the Nobel lecture of McFadden (2001) for the history of RUM models.

Furthermore, logit and probit models have different properties. The probit model offers flexibility since the means and (co)variances can be completely specified. For the logit models on the other hand, only the means can be specified easily. The variance of each marginal distribution has to be the same; however, covariances can be included since the work of McFadden (1978). The achievable covariance structure is limited, and significantly less flexible than what is achievable with probit models.[7]

In Chapter 4 a different type of models that is strongly related to the logit models are analysed. Instead of the additive utility formula of equation (3.1), a multiplicative utility formula is considered:

$$U_c = V_c \times \varepsilon_c, \quad \forall c \in C. \tag{3.4}$$

Under the assumption that the systematic utility is negative, and that $\boldsymbol{\varepsilon}$ is a multivariate reversed Weibull distribution, the probability formulation is again closed-form:

$$P(c) = \frac{\left(\frac{-1}{V_c}\right)^{\mu}}{\sum_{c' \in C} \left(\frac{-1}{V_{c'}}\right)^{\mu}}, \quad \forall c \in C, \tag{3.5}$$

where μ is a parameter of the model. Chapter 4 completely analyses these so called weibit models. The flexibility is different from the logit model, since the variance of utility is completely determined by the systematic utility. Translating the theory of (McFadden, 1978) to the weibit family provides a similar method to introduce covariances.

Chapter 4 also shows that the four ingredients of a choice model can be grasped by a choice set C, a generating vector $\boldsymbol{y} = (y_1, \ldots, y_{|C|})$ (which implies the systematic utility and the utility formula) and a generating function $G : \mathbb{R}^{|C|} \to \mathbb{R}$ (defines the covariances of the error terms). The next sections provides a specific choice for each of them, such that an appropriate choice model for QDTA unfolds.

3.2.2. Choice Alternatives

Consider the set of modes $\mathcal{M}$ containing car CAR, train TRAIN, and the stay-at-home (or telework) alternative HOME. The latter is considered as a special mode, since no trip is actually made. However, this is the most natural way to include it in the choice set, and it allows for some demand elasticity. A fixed set of modes – with one private mode and one public transport mode – is used in this dissertation, since this conveniently shows the mode specific behavioural responses and network effects. Extensions with other modes (e.g., bicycle, bus, metro) are straightforward. The modes introduced in this chapter also return in the case study presented in Chapter 7. For the generalization to multi-modal trips, see the discussion in Section 8.4. The model of Zhou et al. (2009) is another integrated approach where mode and route choice is included; however, they exclude departure time choice, but they include trip generation and distribution.

[7] Mixed logit and logit kernel models (for overviews of these models, see in Frejinger and Bierlaire, 2006; Prashker and Bekhor, 2004) allow more flexible covariance structures; however, these models require simulation of the choice probabilities and are therefore not considered (see also Section 4.4).

For every mode $m \in \mathcal{M}$ and every O-D pair $od \in \mathcal{OD}$ there is a set of routes $\mathcal{R}^m_{od}$. The set of all routes for mode $m \in \mathcal{M}$ is $\mathcal{R}^m = \cup_{od \in \mathcal{OD}} \mathcal{R}^m_{od}$, and the set of all routes for O-D pair $od \in \mathcal{OD}$ is $\mathcal{R}_{od} = \cup_{m \in \mathcal{M}} \mathcal{R}^m_{od}$. For mode HOME no routes exist of course; therefore we introduce a dummy route R_0, and set $\mathcal{R}^{\mathrm{HOME}}_{od} = \{\mathrm{R}_0\}, \forall od \in \mathcal{OD}$. By specifying as well $\mathcal{R}^m_{od}$, $\mathcal{R}^m$, and $\mathcal{R}_{od}$, the number of od and m subscripts/indices can be reduced in the remainder of this chapter.

The departure time choice has a discrete approach. That means that it is assumed that travellers make their trip in a certain time interval. The set of those so called time-of-days is denoted with $\mathcal{T}$. This implies that the demand pattern over time is piecewise constant. Furthermore it is assumed that the attributes related to each time-of-day interval $T \in \mathcal{T}$ represent averages over that interval. Thus for example, travel time attributes should represent the average travel time on routes, and should consider average queue lengths. Models with (discrete) time-of-day choice models appear in (Ettema et al., 2007; Hess et al., 2007; de Jong et al., 2003). More advanced approaches to model time-of-travel preferences and corresponding methodological issues can be found in Ben-Akiva and Abou-Zeid (2013).

Choice situations appear for every O-D pair, and each choice is the selection of a combination of a time-of-day and a route. Since routes are mode specific, choosing a route implies mode choice as well, so there is a simultaneous mode, time-of-day and route choice. Denote the set of choice alternatives for O-D pair $od \in \mathcal{OD}$ with $C_{od} = \mathcal{T} \times \mathcal{R}_{od}$,[8] and denote $(T; r) \in C_{od}$ as the choice alternative consisting of time-of-day T and route r. Denote all choice alternatives as $C = \cup_{od \in \mathcal{OD}} C_{od}$. The analyst can choose to add mode HOME to every time-of-day – leading to $|\mathcal{T}|$ HOME alternatives –, but it is more natural to retain a single HOME alternative (since travellers don't choose a time-of-day when they do not travel), as done in this thesis where it is denoted with $(\cdot, \mathrm{R}_0) \in C$. The correct definition of the choice set is therefore

$$C_{od} = (\mathcal{T} \times (\mathcal{R}_{od} \setminus \{\mathrm{R}_0\})) \cup \{(\cdot, \mathrm{R}_0)\}. \tag{3.6}$$

This is an important choice, since generally the *composition* of the set of choice alternatives itself influences the choice probabilities (see Bliemer and Bovy, 2008). The composition of a choice set is critical for the model results. The following aspects should be borne in mind during this process.

- When a mode is added that partly substitutes another mode, while leaving everything else unchanged, the predicted market share of the modes combined will be overestimated. This can be partly corrected by using a nested choice structure, or by changing the mode specific constants in the systematic utility.
- When a time-of-day is added (or split), additional alternatives arise for every mode, so there is not a direct over-/underestimation error for regular modes. However, the mode HOME is not time-of-day specific, and its market share will decrease.
- Route set generation remains an open question in science. The composition of the route set largely influences the choice probabilities of genuine route choice models (Bliemer and Bovy, 2008). Ideally routes are sampled such that they represent a specified distribution based on the choice model (Frejinger et al., 2009). Unfortunately, no route set generator exists that complies to this requirement. They are either inefficient (such as

[8] Without loss of generality it is possible to use $C_{od} \subset \mathcal{T} \times \mathcal{R}_{od}$, where it is possible to exclude certain mode, time-of-day combinations.

the sampler of Flötteröd and Bierlaire, 2013), or they are merely heuristics (see e.g., Bekhor et al., 2006; Bovy and Fiorenzo-Catalano, 2007; Ramming, 2002). For multimodal networks, van Eck et al. (2012) analyse these heuristics and show that methods merely based on shortest paths under-perform compared to approaches with labelling.

3.2.3. Systematic Utility Specification

The systematic utility for each choice alternative $(T;r) \in C$ should contain its observed attributes. In the QDTA model, the most important attributes of each alternative are travel costs and travel times. Travel costs largely depend on the considered pricing scheme and travel times are determined by the underlying QDNL model. Each user-class has a different value of time and preference towards time-of-days and modes. By normalizing the taste parameter for the cost attribute to -1, all other taste parameters read as willingness-to-pay (see Section 4.4.3). Therefore, the travel time related parameters are values of time. Denote for each time-of-day and route combination $(T;r) \in C$, the systematic utility $V_{(T;r;u)}$ for user-class u as,

$$V_{(T;r;u)} = \begin{cases} \mathrm{ASC}_{(T;\textsc{car};u)} - \kappa_{(T;r;u)} - \mathrm{VOT}^{\textsf{IVT-FF}}_{(\textsc{car};u)} \tau^{\textsf{IVT-FF}}_{(T;r)} - \mathrm{VOT}^{\textsf{IVT-CONG}}_{(\textsc{car};u)} \tau^{\textsf{IVT-CONG}}_{(T;r)} & \text{if } r \in \mathcal{R}^{\textsf{CAR}} \\ \mathrm{ASC}_{(T;\textsc{train};u)} - \kappa_{(T;r;u)} - \mathrm{VOT}^{\textsf{IVT}}_{(\textsc{train};u)} \tau^{\textsf{IVT}}_{(T;r)} & \\ \quad -\mathrm{VOT}^{\textsf{WAIT}}_{(\textsc{train};u)} \tau^{\textsf{WAIT}}_{(T;r)} - \mathrm{VOT}^{\textsf{A-E}}_{(\textsc{train};u)} \tau^{\textsf{A-E}}_{(T;r)} & \text{if } r \in \mathcal{R}^{\textsf{TRAIN}} \\ \mathrm{ASC}_{(T;\textsc{home};u)} - \kappa_{(T;r;u)} & \text{if } r \in \mathcal{R}^{\textsf{HOME}}, \end{cases} \tag{3.7}$$

where for each $m \in \mathcal{M}$,

$$\mathrm{ASC}_{(T;m;u)} = \text{the alternative specific constant of user-class } u \text{ for alternative } (T;r) \tag{3.8}$$

$$\kappa_{(T;r;u)} = \text{the travel cost of alternative } (T;r) \text{ for user-class } u \tag{3.9}$$

$$\tau^{\cdot}_{(T;r)} = \text{the travel time of alternative } (T;r) \tag{3.10}$$

$$\mathrm{VOT}^{\cdot}_{(m;u)} = \text{the value of travel time of user-class } u \text{ for mode } m. \tag{3.11}$$

Here, the $\cdot$ is a placeholder for different types of travel time; Equation (3.7) distinguishes in-vehicle travel time during free flow conditions IVT-FF, in-vehicle travel time during congested conditions IVT-CONG, general in-vehicle travel time IVT (here only used for mode TRAIN), waiting time WAIT, travel time for access and egress legs A-E. These particular choices for types of travel time are in line with the case study presented in Chapter 7; however, it is straightforward to remove and/or add different types. Section 2.2.5 discusses the different preferences towards these travel times.

For the used choice model it is required that the systematic utility is strictly negative. Therefore, it is assumed that travel costs are always positive, and that the alternative specific constants are always negative.

3.2.4. Error Term Distribution and Utility Formula

Section 4.3 will analyse how choice probabilities can be derived from the distribution of the error terms. All models with closed-form probability expressions can be captured with a single

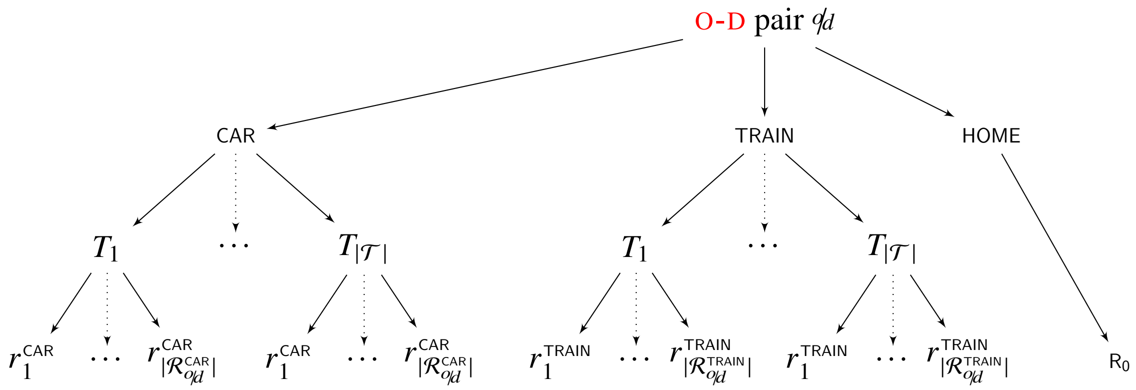


Figure 3.3.: Choice decision tree for O-D pair $\mathcal{od}$ at the top level. Consecutive levels are mode, time-of-day, and route.

formula for the choice probabilities, which is Equation (4.26). To specify the distribution of random utility, merely a generating function that satisfies some requirements (see Section 4.3.1.1) and a generating vector have to be specified. The function determines the covariance structure, while the vector determines how the systematic utility enters the general utility.

Choice situations occur for every O-D pair; the simultaneous choice is captured with a nested structure. The choice tree for the QDTA model is shown in Figure 3.3. First the mode, then the time-of-day, and finally, the route is chosen. The nested structure of the mode and time-of-day choice is considered fixed. Hess et al. (2007) shows how the best nesting structures can be derived from survey data. For the lowest level, i.e., route choice, any generating function from Section 4.3.2 can be implemented. The nested structure implies that the routes within a time-of-day are correlated, and that in-turn the time-of-days within the modes are correlated. Nested models can be considered as sequential multinomial models; where the systematic utility of a group of alternatives in the lower level is a combination of each alternative's systematic utility. Therefore, a scale parameter has to be defined for every level in the choice tree.

Define the generating function $G_{\mathcal{od}} : \mathbb{R}_+^{|C_{\mathcal{od}}|} \to \mathbb{R}_+$ for each O-D pair $\mathcal{od} \in \mathcal{OD}$ with the choice set $C_{\mathcal{od}}$ from Equation (3.6) as

$$G_{\mathcal{od}}(z) = \sum_{m \in \{\text{CAR},\text{TRAIN}\}} \left(\sum_{T \in \mathcal{T}} \left(G^{\text{ROUTE}}_{(\mathcal{od};T;m)}(z_{(T;m)}) \right)^{\frac{\mu^{\text{MODE}}}{\mu^{\text{T-O-D}}}} \right)^{\frac{\mu}{\mu^{\text{MODE}}}} + z_{\text{R}_0}^{\mu}, \tag{3.12}$$

where $0 < \mu < \mu^{\text{MODE}} < \mu^{\text{T-O-D}}$ holds for respectively overall, mode-nest specific, and time-of-day-nest specific scales μ, μ^{MODE} and $\mu^{\text{T-O-D}}$; and for each $T \in \mathcal{T}$ and $m \in \mathcal{M}$, $G^{\text{ROUTE}}_{(\mathcal{od};T;m)} : \mathbb{R}_+^{|\mathcal{R}^m_{\mathcal{od}}|} \to \mathbb{R}_+$ is a $\mu^{\text{T-O-D}}$-homogeneous generating function – which can be any of the Multinomial (MN), Path-Size (PS), Paired Combinatorial (PC), or Link-Nested (LN) generating functions –, and note that $z = \left(\cup_{m \in \mathcal{M}} \cup_{T \in \mathcal{T}} z_{(T;m)}\right) \cup z_{\text{R}_0}$.

Chapter 4 describes how different utility formulas lead to different generating vectors. Since the model should be feasible on networks, heteroscedastic utility distributions are required (see Sections 2.5.2 and 4.2). Therefore, the multiplicative utility formula of Equation (3.4) is chosen. This leads the following definition for the elements of the generating vectors:

$$y_{(T;r;u)} = \frac{-1}{V_{(T;r;u)}}, \quad \forall (T;r) \in \mathcal{C}, u \in \mathcal{U}. \tag{3.13}$$

When these are aggregated per O-D pair per user-class, then they can be combined with Equation (3.12) to produce choice probabilities. Define $\boldsymbol{y}_{(\mathcal{od};u)} = \{y_{(T;r;u)} | T \in \mathcal{T}, r \in \mathcal{R}_{\mathcal{od}}\}$ as the generating vector of user-class $u \in \mathcal{U}$ for O-D pair $\mathcal{od} \in \mathcal{OD}$, to be able to give the choice probabilities

$$P_{(T;r;u)} = \frac{y_{(T;r;u)} \dfrac{\partial G_{\mathcal{od}}\left(\boldsymbol{y}_{(\mathcal{od};u)}\right)}{\partial y_{(T;r;u)}}}{\mu G_{\mathcal{od}}\left(\boldsymbol{y}_{(\mathcal{od};u)}\right)}, \quad \forall (T;r) \in \mathcal{C}, u \in \mathcal{U}. \tag{3.14}$$

With the assumption that each G^{ROUTE} is the multinomial generating function (see Equation (4.40), the choice probability for alternative $(T;r) \in \mathcal{C}$ for user-class $u \in \mathcal{U}$ is

$$\begin{aligned}
P_{(T;r;u)} =& \left(\sum_{m' \in \{\text{CAR},\text{TRAIN}\}} \left(\sum_{T' \in \mathcal{T}} \left(\sum_{r' \in \mathcal{R}^{m'}_{\mathcal{od}}} \left(\frac{-1}{V_{(T';r';u)}} \right)^{\mu^{\text{T-O-D}}} \right)^{\frac{\mu^{\text{MODE}}}{\mu^{\text{T-O-D}}}} \right)^{\frac{\mu}{\mu^{\text{MODE}}}} + \left(\frac{-1}{V_{(\cdot;\text{R}_0;u)}} \right)^{\mu} \right)^{-1} \\
&\times \left(\sum_{T' \in \mathcal{T}} \left(\sum_{r' \in \mathcal{R}^{m}_{\mathcal{od}}} \left(\frac{-1}{V_{(T';r';u)}} \right)^{\mu^{\text{T-O-D}}} \right)^{\frac{\mu^{\text{MODE}}}{\mu^{\text{T-O-D}}}} \right)^{\frac{\mu}{\mu^{\text{MODE}}}-1} \\
&\times \left(\sum_{r' \in \mathcal{R}^{m}} \left(\frac{-1}{V_{(T;r';u)}} \right)^{\mu^{\text{T-O-D}}} \right)^{\frac{\mu^{\text{MODE}}}{\mu^{\text{T-O-D}}}-1} \\
&\times \left(\frac{-1}{V_{(T;r;u)}} \right)^{\mu^{\text{T-O-D}}}, \quad \text{if } r \neq \text{R}_0, \text{ and,} \\
P_{(T;r;u)} =& \left(\sum_{m' \in \{\text{CAR},\text{TRAIN}\}} \left(\sum_{T' \in \mathcal{T}} \left(\sum_{r' \in \mathcal{R}^{m'}_{\mathcal{od}}} \left(\frac{-1}{V_{(T';r';u)}} \right)^{\mu^{\text{T-O-D}}} \right)^{\frac{\mu^{\text{MODE}}}{\mu^{\text{T-O-D}}}} \right)^{\frac{\mu}{\mu^{\text{MODE}}}} + \left(\frac{-1}{V_{(\cdot;\text{R}_0;u)}} \right)^{\mu} \right)^{-1} \\
&\times \left(\frac{-1}{V_{(T;r;u)}} \right)^{\mu}, \quad \text{if } r = \text{R}_0,
\end{aligned} \tag{3.15}$$

where m is the mode of route r and $\mathcal{od}$ is the O-D pair of route r. Finally, the demand and the choice probabilities can be applied to obtain the number trips for each alternative:

$$f_{(T;r)} = \sum_{u \in \mathcal{U}} P_{(T;r;u)} D_{(u;\mathcal{od})}, \quad \forall (T;r) \in \mathcal{C}, u \in \mathcal{U}, \tag{3.16}$$

where $\mathcal{od}$ is the O-D pair of route r.

3.3. Quasi-Dynamic Network Loading of Vehicular Traffic

The interaction between travel demand and infrastructure supply for road networks is captured in this section. The network loading procedure for a single time-of-day will be presented; therefore, the time-of-day indices are omitted. The resulting traffic states (e.g., speeds, travel times and densities) represent the average conditions for that time-of-day. The section summarizes the QDNL method introduced in (Brederode et al., 2010; Bliemer et al., 2012, 2014b; Raadsen et al., 2016). This model is also known as Static Traffic Assignment with Queuing (STAQ). Before the details of the model are discussed, a basic introduction to KWT is presented in the next section.

3.3.1. First-order Kinematic Wave Theory

DNL models based on first order KWT have realistic bottleneck locations, congestion build up by shock wave propagation and allow for spillback. Therefore these types of models are often used in practice. The numerical solution methods used to solve the first order kinematic wave equations can be classified in space-discretized models and models based on variational theory or viability theory. With space-discretized models the link is divided into cells and traffic propagates through these cells in each timestep. The other models consider the link as a whole, leading to more efficient solution schemes.

The well known kinematic wave theory originates from the seminal papers of Lighthill and Whitham (1955); Richards (1956) The theory is especially useful since it captures important traffic flow phenomena like shockwaves and spillback. In Daganzo (1995a) a space-discretized solution method for the kinematic wave equations is presented, the well known Cell Transmission Model. The trilogy (Newell, 1993a,b,c) presents a *simplified* kinematic wave theory. Newell achieves a major improvement in tractability and simulation speed of first order kinematic wave theory by applying variational theory. To propagate traffic over a link according to Newell's theory, the Link Transmission Model (LTM) is developed. The LTM is an elegant and fast traffic simulation method that is not restricted to links, but also propagates traffic through networks by making use of a node model. Descriptions are found in (Yperman, 2007; Gentile, 2008). One advantage of the LTM is that it only requires a time discretization and thus not a space discretization.

Newell (1993a) describes how traffic on a road segment can be represented by a Moskowitz surface. This is a surface in the three-dimensional space spanned by time, space and vehicle number. The latter is equivalent to cumulative flow. Figure 3.4 depicts an example Moskowitz surface of a road with a temporary bottleneck. At the start no vehicle is on the road (constant line of the north-west boundary), and vehicles start to enter (increasing line at the south-west boundary). Vehicles propagate with free-flow speed (iso-colour lines are trajectories). Halfway there is a bottleneck (downstream of this bottleneck there is a flat surface), and vehicles are in congestion in the queue in front of it. After the bottleneck is resolved vehicles flow out with capacity (increasing slope of the line at the north-east boundary after the bottleneck's constant surface).

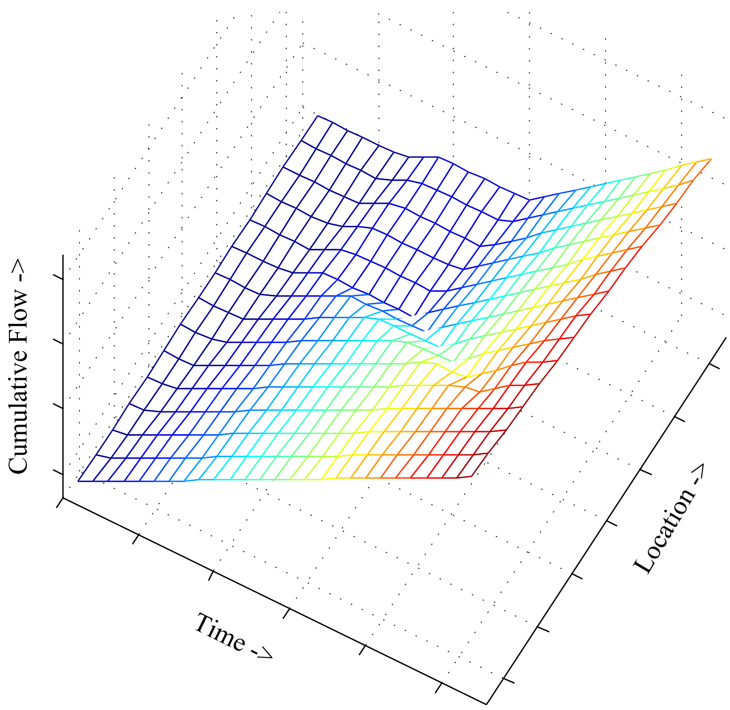


Figure 3.4.: The Moskowitz surface representing an initially empty road segment with constant inflow and a temporary bottleneck in the middle.

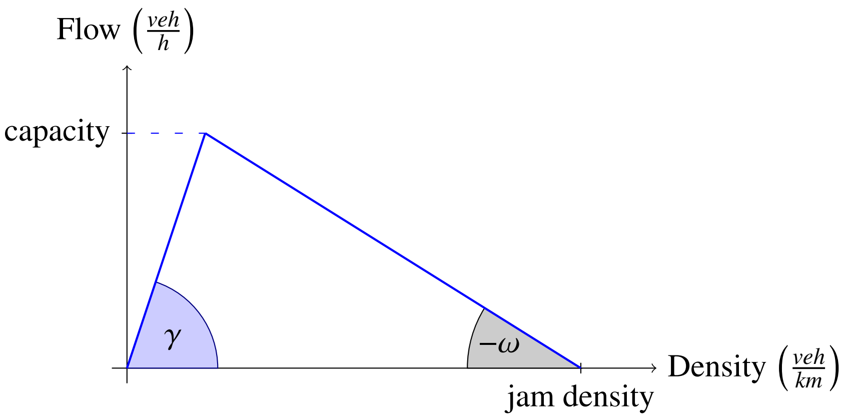


Figure 3.5.: Triangular density-flow fundamental diagram

The fundamental diagram and the conservation of vehicles are the two main ingredients of KWT; they are explained in the next paragraphs. The Moskowitz surface can be represented in the functional form $N(x,t)$, where N is the cumulative flow at space (i.e., position) x and time t. The equality of the two second-order derivatives then define the conservation of vehicles equation,

$$-\frac{\partial k(x,t)}{\partial t} = \frac{\partial^2 N(x,t)}{\partial x \partial t} = \frac{\partial^2 N(x,t)}{\partial t \partial x} = \frac{\partial f(x,t)}{\partial x}, \tag{3.17}$$

where $k(x,t)$ is density and $f(x,t)$ is flow at position x and time t. As the name suspects, there is a simple interpretation of this equation. Namely, imagine one takes a helicopter for a view on a road segment from above. Count the number of visible vehicles and keep that number in mind; wait for a while and add one for every vehicle that enters the segment and subtract one for every vehicle that exits the segment. Stop after some time, the number that is in mind now equals the number of vehicles that are on the road segment at that time. This simple equality illustrates the conservation of vehicles equation. The difference in number of vehicles on the road before and after illustrates the left-hand side of Equation (3.17), and right-hand side illustrates the outcome of the counting of vehicles entering and exiting the road.

Now define the fundamental diagram, that defines the relation between flow and density, as concave function $F : \mathbb{R}_+ \rightarrow \mathbb{R}_+$. This function can be substituted in differential equation (3.17), which leads to

$$\frac{\partial k(x,t)}{\partial t} + \frac{\partial F(k(x,t))}{\partial x} = 0. \tag{3.18}$$

Figure 3.5 shows an example fundamental diagram with a triangular shape. The branch left of the capacity describes free flow states (where the vehicle speed is γ), while the branch right of the capacity describes congested states (where the vehicle speed gradually decreases to 0 at the jam density). Points on the fundamental diagram represent feasible traffic states; this means that the first derivatives at any point of the Moskowitz surface (Figure 3.4), which are flow and density, coincide with a point on the fundamental diagram. This is of course not true with an empirical Moskowitz surface; therefore, the points on the fundamental diagram are also referred to as the stable or equilibrium states of traffic.

Also the fundamental diagram can be easily interpreted with a helicopter view. Assume that the conditions are constant over time, observe the road segment again, and count the number of vehicles on the segment. The fundamental diagram now prescribes how many vehicles will enter and exit the segment during the observation. So, in the free flow branch, the more vehicles are visible, the more will enter and exit. Up to the point when traffic breaks down and the congested branch starts; from then, the more vehicles are visible, the slower they drive, and the less they will enter and exit the road. When just up- or downstream of the observed segment, the number of vehicles is different (i.e., the conditions are not constant), then the flows over the boundaries will be influenced by both densities. KWT describes how these states interact.

In first-order KWT only states on the fundamental diagram are feasible. Furthermore, a direct transition between two states is possible, which implies an instantaneous speed change. (Note that the speed of traffic υ is easily determined with $\upsilon = \frac{F(k)}{k}$.) Second-order models allow for bounded acceleration, but this comes with other undesired properties such as negative speeds (Daganzo, 1995b). For strategic network models first-order KWT is sufficient to describe traffic dynamics since it satisfies the requirements stated in Section 2.5.2.

Differential equation (3.18) can describe (i.e., estimate and predict) traffic dynamics at other locations based on boundary conditions. Such boundary conditions are values of $N(x,t)$ for specific space-time locations (x,t); they can for example be the values of the cumulative flow (only as function of time) at the start, half-way (where the bottleneck is), and the end of the segment depicted in Figure 3.4. This means that if the traffic conditions at the start and the end of a road segment are known, the prevailing conditions for any location of the road segment can be determined. Mathematically, such a problem is a Hamilton-Jacobi partial differential equation (Lax, 1957; Hopf, 1970), for which efficient solution methods exist based on variational theory. Applying variational theory on KWT has been explored by Newell (1993a,b,c); Daganzo (2005a,b, 2006); Laval and Leclercq (2013). Recently, Claudel and Bayen (2010a,b) proposed methods to solve Hamilton-Jacobi partial differential equations based on viability theory (see also Aubin et al., 2011, Chapters 13&14). The network loading model in the QDTA model is based on viability theory.

Exploration of variational and viability theory started with Newell (1993a,b,c)'s trilogy named simplified KWT. The main contribution of Newell is that, given the past conditions at the boundaries, the conditions for the future can be analytically determined. This approach is particularly interesting when the fundamental diagram is piece-wise linear, or when the flows are piece-wise constant. The first case allows efficient variational theory methods, while the latter case allows application of viability theory. In any of these situations the future condition is the minimum of a finite number of possibilities. For each of these possibilities a kinematic wave is traced through time-space with its (constant) kinematic wave speed; the key property in both theories is that the cumulative flow over this trace changes with a constant rate.

The LTM (Yperman, 2007; Gentile, 2008) uses variational theory in combination with a node model to predict future traffic state in a road network gives the complete history of traffic states. Daganzo (2005a,b) uses fundamental diagrams with carefully chosen kinematic wave speeds, such that traffic conditions for a grid of points in the time-space plane can be determined analytically.

3.3.2. Quasi-Dynamic Network Loading

Brederode et al. (2010) proposed a *hybrid* model consisting of two phases, which was further developed by (Bliemer et al., 2012). In the first phase a static traffic assignment model with vertical queues and capacity constraints is applied, while in the second phase a dynamic model is applied that propagates traffic flow states according to KWT in order to describe physical queues. This dissertation adopts such a hybrid approach which applies a capacity constrained traffic assignment model with residual queues proposed by Bliemer et al. (2014b)[9] in the first phase, called *squeezing*, and we apply the dynamic event-based algorithm proposed by Raadsen et al. (2016) and based on the continuous-time LTM in the second phase, called *queueing*. This two phase hybrid approach will be referred to as QDNL. The squeezing phase leads to vertical residual queues in front of bottlenecks. The queueing phase uses an event-based algorithm based on variational theory propagates shock waves and allows for spillback to other upstream links.

Figure 3.6 shows the flowchart of the QDNL model. The inputs are the travel demand per route, and the road network. The latter specifies the links ($\mathcal{L}$) and nodes ($\mathcal{N}$), and as well fundamental diagrams per link that describe traffic characteristics (e.g., capacity, jam density) and determine the speeds of shock waves. The output is the traffic state which can be expressed as speeds, densities, and flows per link. Squeezing and queuing can be identified as two sequential iterative processes.

Squeezing determines the resulting residual queues, which are revealed in the form of demands for every inlink of every node; squeezing also fixes the turn fractions at every node. This is achieved by solving a fixed point problem, where the largest challenge is to determine input (i.e., the demands per turn) for a node model which depends on the output of all other node models. This node model input is obtained by walking through every route, where the 'flow' of the route is reduced after every bottleneck. After such a reduction the demand from that route for the following nodes is lower.

The second iterative process, queueing, propagates shockwaves for a predetermined length of time. Given a set of past boundary conditions for the links, events predict potential changes in these boundary conditions. A node model then determines if the potential change triggers any actual change on any of its adjoining links, which in turn leads to new events which are added to the event list. Figure 3.7 shows an example application of QDNL on a simple corridor with multiple bottlenecks. The example serves as an illustration of theory presented in the next sections. Bliemer et al. (2014b) present more examples of the squeezing phase that are more interesting since they contain nodes that go beyond lane drops.

3.3.3. Bottleneck Identification (squeezing)

As preliminary, write $\mathcal{W}^r$ as the ordered set of turns that belong to route $r \in \mathcal{R}$. To be able to test the ordering of two turns within a route define operator $\prec^r$ as

$$\langle i,j\rangle \prec^r \langle i',j'\rangle \Leftrightarrow \langle i,j\rangle \text{ preceeds } \langle i',j'\rangle \text{ in route } r \in \mathcal{R}, \quad \forall \langle i,j\rangle, \langle i',j'\rangle \in \mathcal{W}^r \tag{3.19}$$

[9] Note that Bliemer et al. (2012) also contains the squeezing phase; however, the squeezing phase methodology presented in that paper violates the conservation of turning fractions in nodes. This has been resolved in a revised squeezing method by Bliemer et al. (2014b).

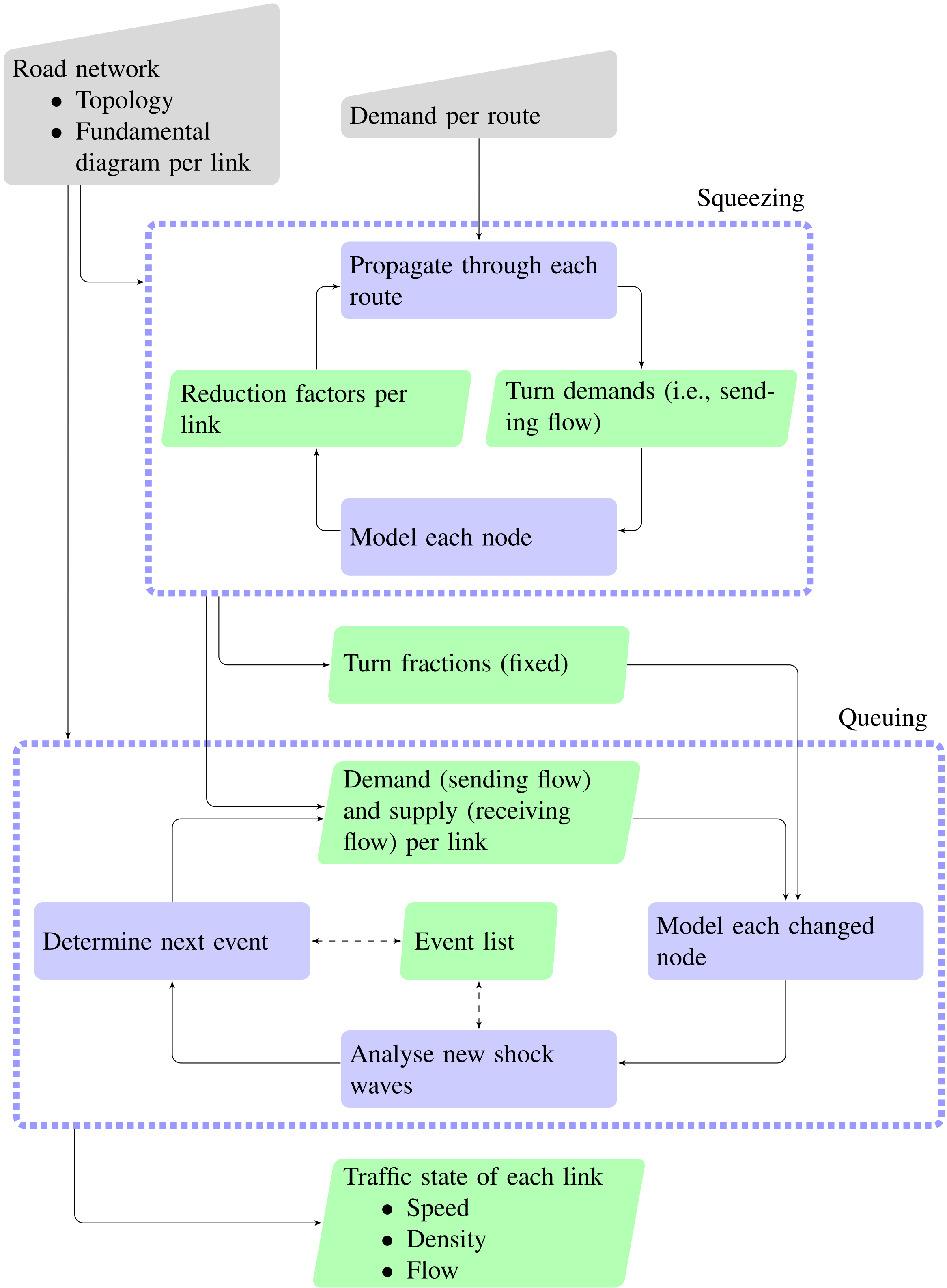


Figure 3.6.: Flowchart of the QDNL model

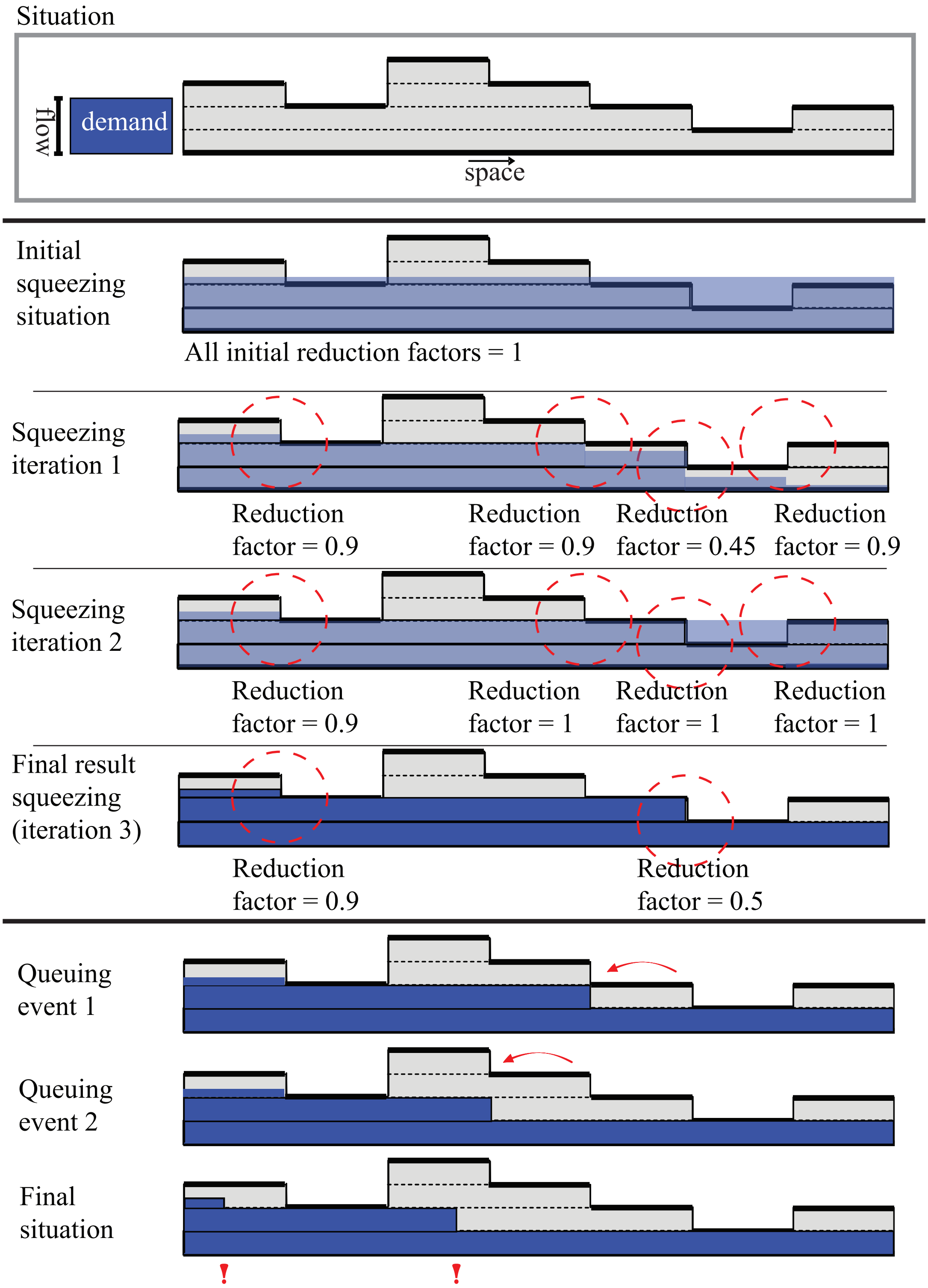


Figure 3.7.: An example of QDNL on a sequence of road segments. The squeezing phase ensures that the capacity constraints are not violated. In each iteration reduction factors are determined based on the demand and upstream reduction factors from the previous iteration. The queuing phase uses KWT to propagate queues based on events; the shock waves are only backwards in this example.

Bottlenecks are located at the start of links with insufficient capacity to accommodate the demand. After each bottleneck, only part of the route demand continues downstream. Such a reduction (or squeeze) occurs only at the end of each link, and due to the first-in first-out principle this factor of this reduction is equal for all routes on that link[10]. Therefore, we can define a single reduction factor $\varphi_i \in (0,1]$ per link $i \in \mathcal{L}$; this factor is applied at the end of a link, so nodes are positioned between the bottlenecks, and the locations where the reduction factor are applied. The vector of all reduction factors is denoted with $\boldsymbol{\varphi} = \{\varphi_i, i \in \mathcal{L}\}$. The turn demand is the sum of all reduced route demands after all their respective upstream reductions have been applied.

$$S_{\langle i,j\rangle} = \sum_{\{r\in\mathcal{R}|\langle i,j\rangle\in\mathcal{W}^r\}} \left(\prod_{\{\langle i',j'\rangle\in\mathcal{W}^r|\langle i',j'\rangle \prec^r \langle i,j\rangle\}} \varphi_{\langle i,j\rangle} \right) f_r, \quad \forall \langle i,j\rangle \in \mathcal{W}, \tag{3.20}$$

where for each route, the product of all reduction factors on all preceding turns is taken. Write $\boldsymbol{S} = \{S_{\langle i,j\rangle}, \langle i,j\rangle \in \mathcal{W}\}$ for the vector of turn demands.

To determine the reduction factors based on the turn demands, any of the node models discussed in Chapter 5 can be used since they all satisfy the requirements for proper node models. In this section a single function represents the application of the node model for all nodes[11] in the network. Define the node function for the squeezing phase $\Phi : \mathbb{R}_+^{|\mathcal{W}|} \to (0,1]^{|\mathcal{L}|}$ as

$$\boldsymbol{\varphi} = \Phi(\boldsymbol{S}). \tag{3.21}$$

The turn demands alone do not suffice as input, as discussed in Chapter 5 also the supply (or receiving flow) for each outlink of each node is required. In the squeezing phase no backward shock waves are considered; therefore, the supply always equals capacity and is thus constant. To be completely compatible with Chapter 5, which uses slightly different variables, the turn demands have to be converted to the demand per inlink and turn fractions. This can be achieved with a trivial rewriting per inlink.

By substituting equation (3.20) in (3.21), the squeezing phase is reduced to a fixed point problem:

$$\boldsymbol{\varphi} = \Phi\left(\left\{ \sum_{\{r\in\mathcal{R}|\langle i,j\rangle\in\mathcal{W}^r\}} \left(\prod_{\{\langle i',j'\rangle\in\mathcal{W}^r|\langle i',j'\rangle \prec^r \langle i,j\rangle\}} \varphi_{\langle i,j\rangle} \right) f_r, \forall \langle i,j\rangle \in \mathcal{W} \right\}\right). \tag{3.22}$$

This fixed problem is solved by iterating. In general, this is very efficient – also on large-scale networks. Bliemer et al. (2014b) prove existence of a fixed point, but it is not necessarily unique. Stylised networks exist for which multiple fixed points exist, but these situations are unlikely to occur in reality since a large amount of tail-biting behaviour is required, where the input of a node is almost completely dependent on its own output.

[10] The first-in first-out rule is equivalent with non-overtaking behaviour and also with the conservation of turning fraction rule which is often used in node modelling.

[11] Since for some nodes the supply is always sufficient (no matter which upstream situations occur), they can be completely omitted in the squeezing phase, which leads to significant efficiency improvements (see Bliemer et al., 2014b)

3.3.4. Queue Propagation (queuing)

This section discusses the queuing phase by explaining the underlying ideas, but without going into the mathematical details (for that, see Bliemer et al., 2012; Raadsen et al., 2016). The model is plainly represented by a function. The input is the fixed point of the squeezing phase, and the output is the traffic state which is represented by density. Before the queuing phase with event iterations starts, there is a conversion; the turn demand and squeezing factors are converted to turn fractions which remain fixed in the remainder. Also, the squeezing factors and turn demands are converted to the initial boundary conditions for the links, which consist of link inflows and outflows.

For each link $i \in \mathcal{L}$ write k_i as its density and denote $\boldsymbol{k} = \{k_i, i \in \mathcal{L}\}$ as the vector of densities. Let queuing function $\Psi : (0,1]^{|\mathcal{L}|} \times \mathbb{R}_+^{|\mathcal{R}|} \to \mathbb{R}_+^{|\mathcal{L}|}$ be

$$\Psi(\boldsymbol{\varphi}, \boldsymbol{S}) = \boldsymbol{k}. \tag{3.23}$$

The queuing function represents an event-based algorithm, as depicted in Figure 3.6. An event is the combination of a location (i.e., node) and time instance at which the traffic conditions of a link adjacent to this node potentially changes. In other words, an event is the time instance a shock wave reaches a node.

These conditions remain constant until the next event occurs. During an event a shock wave can arrive at some node, and this can lead to new in- and outflow on the links adjacent to this node. These are evaluated, and possibly events are added and removed from the event list. Then again the next event is determined. This iterative process continues until some pre-defined time limit is reached. This time limit is endogenous and can serve as a calibration parameter; the higher the time limit is, the longer the queues become.

The traffic conditions at that time are derived from the cumulative in- and outflows. For a link, the difference between them is the density, these can be translated to flows and speeds by using the fundamental diagram of the link (F_i).

$$f_i = F_i(k_i), \quad \forall i \in \mathcal{L}, \tag{3.24}$$

$$v_i = \frac{F_i(k_i)}{k_i}, \quad \forall i \in \mathcal{L}. \tag{3.25}$$

From the speeds, the travel time per route is calculated as

$$\tau_r = \sum_{i \in \mathcal{L}_r} \frac{l_i}{v_i}. \tag{3.26}$$

3.4. Variational Inequality Formulation

The QDTA can be seen as a simulation model, summarized by Figure 3.2. However, it can also be written in terms of a variational inequality problem, which is preferred for analytical analysis. In Section 4.4.2 the variational inequality formulations for stochastic user equilibria with different route choice models are derived. This can be extended in a straightforward manner to include mode and time-of-day choice into the variational inequality formulation. Contrary to mathematical programming formulations for user equilibria, their variational inequality formulations are sporadic in the literature. Chen (1999) provides several models, amongst other the stochastic user equilibrium under multinomial logit. Bliemer and Bovy (2003) provide a time-discrete multi-class dynamic traffic assignment model. Huang and Lam (2002) provide a variational inequality formulation for the time-of-day choice. Zhou et al. (2009) provide an approach to trip generation, trip distribution, modal split, and trip assignment based on a hierarchical model.

The variational inequality of Equation (4.87) is applied to the generating function (Equation (3.12)) and generating vector (Equation (3.13)) of the QDTA model. For notational convenience first define the stochastic generalized cost[12] for alternative $(T;r) \in C$ for user-class $u \in \mathcal{U}$ as

$$\begin{aligned} c_{(T;r;u)}(\boldsymbol{f}) = & -\ln\left(\left(\sum_{T'\in\mathcal{T}}\left(\sum_{r'\in\mathcal{R}^m_{od}}\left(\frac{-1}{V_{(T';r';u)}(\boldsymbol{f})}\right)^{\mu^{\text{T-O-D}}}\right)^{\frac{\mu^{\text{MODE}}}{\mu^{\text{T-O-D}}}}\right)^{\frac{\mu}{\mu^{\text{MODE}}}-1}\right) \\ & -\ln\left(\left(\sum_{r'\in\mathcal{R}^m}\left(\frac{-1}{V_{(T;r';u)}(\boldsymbol{f})}\right)^{\mu^{\text{T-O-D}}}\right)^{\frac{\mu^{\text{MODE}}}{\mu^{\text{T-O-D}}}-1}\right) \\ & -\ln\left(\mu\left(\frac{-1}{V_{(T;r;u)}(\boldsymbol{f})}\right)^{\mu^{\text{T-O-D}}}\right)+\ln\left(f_r\right), \quad \text{if } r \neq \text{R}_0, \text{ and,} \\ c_{(T;r;u)}(\boldsymbol{f}) = & -\ln\left(\mu\left(\frac{-1}{V_{(T;r;u)}(\boldsymbol{f})}\right)^{\mu}\right)+\ln\left(f_r\right), \quad \text{if } r = \text{R}_0, \end{aligned} \tag{3.27}$$

where m is the mode of route r and od is the O-D pair of route r. Here the systematic utility is made dependent of the flow. The travel times herein for mode CAR can be retrieved by the QDNL model. By using the fixed point solution of the squeezing phase and by substituting Equations (3.25) and (3.23) successively in Equation (3.26) the travel time for a route is

$$\tau_r(\boldsymbol{f}) = \sum_{i\in\mathcal{L}_r}\frac{l_i\Psi_i(\boldsymbol{\varphi}^*(\boldsymbol{f}),\boldsymbol{S})}{F_i(\Psi_i(\boldsymbol{\varphi}^*(\boldsymbol{f}),\boldsymbol{S}))}, \tag{3.28}$$

where $\boldsymbol{\varphi}^*(\boldsymbol{f})$ is the fixed point of Equation (3.22) based of flows $\boldsymbol{f}$.

Consider the following variational inequality formulation of the stochastic user equilibrium in the QDTA model; find equilibrium flow $\boldsymbol{f}^* \in \mathbb{R}^{|C|\times|\mathcal{U}|}$ such that

$$\begin{aligned} & \sum_{(T;r)\in C}\sum_{u\in\mathcal{U}} c_{(T;r;u)}(\boldsymbol{f}^*)(f_{(T;r;u)} - f^*_{(T;r;u)}) \geq 0, \quad \forall \boldsymbol{f} \in \Omega, \\ \text{where } \Omega = & \left\{\boldsymbol{f}\in\mathbb{R}^{|C|\times|\mathcal{U}|}\,\middle|\, f_{(T;r;u)} > 0;\ \sum_{(T;r)\in C_{od}} f_{(T;r;u)} = D_{(u;od)}, \forall u\in\mathcal{U}, od \in \mathcal{OD}\right\}. \end{aligned} \tag{3.29}$$

[12]This cost is unit-less. In some other formulations the cost is divided with the scale parameter, such that one of the terms equals the systematic utility (only for the MNL model).

There are several advantages of having a variational inequality formulation of the QDTA model. First, the theory of variational inequalities allows analysis of the existence and uniqueness of equilibria. Bliemer et al. (2014b) show that for route choice and when the queuing phase of the QDNL model is omitted, an equilibrium exists. They also show that it is likely that only under very specific circumstances this equilibrium is non-unique. Due to the simulation character of queuing that creates a lot of dependencies between links in the network, it will be cumbersome to retrieve analytical results, The second advantage is that the variational inequality allows the formulation of a gap function.

3.4.1. Gap Function

Facchinei and Pang (2003); Fukushima (1992); Solodov and Tseng (2000) provide methods to retrieve a gap functions based on variational inequalities. Gap functions provide a measure of the distance of some solution (i.e., flows) to an equilibrium, and they can be used in iterative methods as a stopping criterion. They are based on the fact that a generalized cost is included in the variational inequality formulation. Using the definitions of the previous section and the results of Section 4.4.2 that derives a gap function similar to that in (Bliemer et al., 2014b), it follows that when $\hat{\boldsymbol{f}} \to \boldsymbol{f}^*$, then

$$\frac{\sum\limits_{(T;r)\in C}\sum\limits_{u\in\mathcal{U}} \hat{f}_{(T;r;u)}\left(c_{(T;r;u)}\left(\hat{\boldsymbol{f}}\right) - \min\limits_{(T;r;u)\in C\times\mathcal{U}} c_{(T;r;u)}\left(\hat{\boldsymbol{f}}\right)\right)}{\sum\limits_{(T;r)\in C}\sum\limits_{u\in\mathcal{U}} \hat{f}_{(T;r;u)} \min\limits_{(T;r;u)\in C\times\mathcal{U}} c_{(T;r;u)}\left(\hat{\boldsymbol{f}}\right)} \to 0. \tag{3.30}$$

This gap can be used after each loop in Figure 3.2. The travel times of the QDNL model and the trips per route can be inserted in Equation (3.30), and when the value smaller than a threshold, the simulation can be aborted, and the final effects can be computed.

3.5. Synthesis

In this chapter an introduction to TA is presented, as well as the specific QDTA model that looks ahead to the methodological Chapters 4 and 5. This model consists of two main components being a simultaneous mode, time-of-day, and route choice model and a QDNL model. The choice model uses a multiplicative utility formula which is advantageous on large networks (see Chapter 4). Dependencies between choice alternatives are captured with a nest structure. A special dummy mode captures the stay-at-home alternative. The QDNL model is presented concisely and is based on the theory in (Bliemer et al., 2012, 2014b; Raadsen et al., 2014a); in Chapter 5 the node model part hereof is further analysed. The components merge in the variational inequality formulation of the stochastic user equilibrium. Chapter 7 will use the QDTA model to investigate the responses of travellers to pricing measures from Chapter 2's upper level problem within a case study with a large real network.

CHAPTER 4.

GENERALIZED MULTIVARIATE EXTREME VALUE MODELS FOR EXPLICIT ROUTE CHOICE SETS

This chapter analyses a class of route choice models with closed-form probability expressions, namely, Generalized Multivariate Extreme Value (GMEV) models. A large group of these models emerge from different utility formulas that combine systematic utility and random error terms. Twelve models are captured in a single discrete choice framework. The additive utility formula leads to the known logit family, being multinomial, path-size, paired combinatorial and link-nested. For the multiplicative formulation only the multinomial and path-size weibit models have been identified; this study also identifies the paired combinatorial and link-nested variations, and generalizes the path-size variant. Furthermore, a new traveller's decision rule based on the multiplicative utility formula with a reference route is presented. Here the traveller chooses exclusively based on the differences between routes. This leads to four new GMEV models. We assess the models qualitatively based on a generic structure of route utility with random foreseen travel times, for which we empirically identify that the variance of utility should be different from thus far assumed for multinomial probit and logit-kernel models. The expected travellers' behaviour and model-behaviour under simple network changes are analysed. Furthermore, all models are estimated and validated on an illustrative network example with long distance and short distance origin-destination pairs. The new multiplicative models based on differences outperform the additive models in both tests.

This chapter is a revised and extended version of:

- Smits, E.-S., Bliemer, M., Pel, A., and van Arem, B. (2014).
  On route choice models with closed-form probability expressions.
  In *The Transportation Research Board (TRB) 93^rd Annual Meeting, Washington DC, January 12-16, 2014 (paper no 14-3733).*, pages 1 – 24

4.1. Introduction

Route choice is important in transport applications such as network equilibrium modelling, day-to-day route choice decisions, and route guidance. The literature describes various methods to model the route choice behaviour of travellers. They range from simple deterministic shortest route choice to sophisticated stochastic models in which random error terms capture travel time uncertainty and taste heterogeneity amongst travellers. Commonly, the route choice model is applied in an iterative process, for example, to reach equilibrium in a congested network or for en-route decisions. For large networks the number of times route choice has to be simulated is very high. Therefore, a good balance between realism and computational efficiency is required.

4.1.1. Random utility maximization

A common choice mechanism is that travellers consider multiple routes, assign a subjective utility to each route, and choose the route with the highest utility. The discrete choice framework based on RUM allows random components in the utility formulations. In this case the utility of each route follows some random distribution. Observe that a route consists of a set of links, and we assume that a route's utility is (among other things) determined by the characteristics of the corresponding links. Then the joint probability distribution of route utilities contains dependencies when routes have overlap, and thus share (dis)utility from the same link(s). Furthermore the distribution is heteroscedastic (i.e., the variability of utility differs amongst routes) since routes have different lengths. Section 4.2 provides an in-depth analysis of desired random route utility properties, that especially explores the distribution of foreseen travel times besides the usual analyst error. This gives new insights into the variance and covariance structure of route utility.

The Multinomial Probit (MNP) model for route choice (see Daganzo and Sheffi, 1977; Yai et al., 1997) can easily address correlation and heteroscedasticity, but no closed-form formulation of the route probabilities exists. The latter leads to computationally expensive simulations. On the other hand the Multinomial Logit (MNL) model for route choice (see Daganzo and Sheffi, 1977)[1] has an elegant closed-form formulation for the route probabilities; however, route utilities have to be independent and homoscedastic. Extensions to mixed logit models (Bekhor et al., 2002) are possible to overcome some of the limitations, however, these models again have to rely on numerical solution methods requiring simulations which are often

[1] Daganzo and Sheffi (1977, Eqn. 10) are the first to present MNL route choice probabilities, based on the method of Dial (1971)

infeasible on large scale networks. In addition, the thus far proposed covariance structures in MNP and mixed logit are not in line with revealed foreseen travel times; the *variance* has been assumed proportional to the mean, whereas instead, data shows a linear relation between *standard deviation* and mean (see Section 4.2.1).

Several adaptations to the MNL model exist to address correlation due to route overlap (Prato, 2009). They either insert a correction term into the utility, or exploit the more general Multivariate Extreme Value (MEV) distribution for the error term. In Section 4.3.2 an overview of these methods is presented. The heteroscedasticity is less addressed in the literature. In practice, the scale parameter of MNL is sometimes considered to be proportional to the distance between origin and destination, but this does not address heteroscedasticity within the O-D route set (see Chen et al., 2012). Less pragmatic is the solution by Castillo et al. (2008) who assume Weibull distributed utilities leading to the Multinomial Weibit (MNW) model. Only slightly different is the approach of Gálvez (2002); Fosgerau and Bierlaire (2009) who use a multiplicative error term instead of an additive error term. The probabilities of the earlier-known Kirchhoff distribution of routes as presented by Fellendorf and Vortisch (2010) also coincide with MNW. Recently, Kitthamkesorn and Chen (2013) included the path-size factor into the MNW model to correct for correlation between routes.

4.1.2. Generalized Multivariate Extreme Value models

In this paper the state-of-the-art of models with closed-form utility formulations is reviewed, and they are gathered in a framework with a single, general form for the choice probabilities. Two classes of MEV can be identified – each based on a different utility formula –, which are unified in what we call Generalized Multivariate Extreme Value (GMEV). The generalization gives rise to three undiscovered route choice models based on the multiplicative utility formula. Furthermore, we recall that equivalent MEV formulations exist for models in which a correction term is inserted in the utility formulation (e.g. Path-Size Logit). All these models can therefore be captured in the GMEV framework based on either additive or multiplicative utility formulas. Section 4.4 assesses these models qualitatively by analysing their distributions and comparing them with the desired properties formulated in Section 4.2.3. Additive models prove not to be able to capture random foreseen travel time, but they do capture the analyst error. On the other hand, multiplicative models can handle both errors, but based on the same distribution (i.e., the foreseen travel time and analyst error are completely dependent).

One of the desired properties cannot be fulfilled by any existing model. Namely, in a transport network model the route choice model is applied to many route sets, so it is important that the model gives realistic results for all route sets existing in a network. In particular, the results should remain realistic when the network and/or routes change. With respect to this, both MNL and MNW have an undesired property. In MNL, the route probabilities will remain equal if a constant is added to each route's utility. In MNW, the route probabilities will remain equal if each route's utility is multiplied with a constant. Both these properties are unrealistic. Therefore, Xu et al. (2015) provide a hybrid method where the choice probabilities have a MNL en MNW component; however, this hybrid method stems from a combined impedance function (i.e., it defines the choice probabilities based on the choice probability functions of logit and weibit). Although the model can be useful in practice, the model does not stem from a utility formulation with an additive and multiplicative error term (Section 4.3.1.3 discusses how the hybrid models would fit in the GMEV framework).

To partially overcome the shortcomings of MNW, we introduce new GMEV route choice models based on the multiplicative utility formula and a reference route. The probability of choosing the reference route only depends on the non-overlapping differences with other routes, since the overlap is explicitly removed. The models fit in the same framework as the existing models. Furthermore, they are designed to be suitable under changing networks and routes. A qualitative assessment and a numerical benchmark on a small network example of the existing and new models show that the new models can resemble expected behaviour under more types of network changes than existing closed-form models can. For example, the model performs well if all route costs are multiplied with a factor, but also if a constant cost is added to all routes.

4.1.3. Route set generation

This paper assumes that a relevant route set is explicitly given. For an overview of route set generation techniques, see Frejinger et al. (2009); Prato (2012). However, it should be noted that the generation and composition of the route set has a large influence on model outcomes (Bliemer and Bovy, 2008; Cascetta et al., 2002; Prato, 2012). Another disadvantage with explicit route sets is that a correct sample of routes is required to obtain unbiased parameter estimates (Frejinger et al., 2009). For discussions on sampling routes based on distributions derived from route choice models see (Frejinger et al., 2009; Flötteröd and Bierlaire, 2013; Guevara and Ben-Akiva, 2013). Bierlaire et al. (2008) points out difficulties with selection bias in the estimation of additive MEV models from choice based samples. A recently revisited different approach – that overcomes route set generation/sampling related problems – is to implicitly generate routes as done by Dial (1971); Papola and Marzano (2013); Fosgerau et al. (2013a); Mai et al. (2015). However, these models either have a restricted route set (Dial, 1971; Papola and Marzano, 2013) (see Section 4.3.2.5), or contain all routes, even those with loops (Fosgerau et al., 2013a; Mai et al., 2015). In addition, no methods with implicit choice set generation that can handle non-additive travel costs are known to exist by the authors of this paper.

Since explicit finite route sets have practical advantages, they are commonly used in traffic assignment models to either avoid too computationally expensive (dynamic) shortest-path calculations, to allow dynamic network loading of path flows, or to allow for traffic assignment models with non-additive link costs. For example, Zhou et al. (2015) generates an a-priori route set in a static assignment context using the deterministic method of Bekhor et al. (2006), Bliemer et al. (2014b) generates an a-priori route set for a quasi-dynamic model using the stochastic method of Fiorenzo-Catalano et al. (2004).

4.1.4. Contribution

The contribution of this paper is summarized as follows. First, we describe the desired properties of random route utility which includes random foreseen travel times, and empirically verify the linear relation between mean and standard deviation of foreseen travel time. Second, existing closed-form route choice models based on random utility maximization are presented in a novel GMEV framework and it is shown that all existing closed-form models with explicit choice sets fit in the GMEV framework. Third, the unexplored area of multiplicative MEV models gives rise to three new closed-form route choice models, and the proposed multiplicative MEV models based on a reference route give rise to another four new closed-form route choice models. The working of this model is illustrated with a small network example, and its behaviour with respect to network changes is compared to the behaviour of existing models. Fourth, we show that the additive models, contrary to multiplicative models, can not capture random foreseen travel time. Fifth, we show the strength of a unified GMEV model by providing its stochastic user equilibrium formulation. Sixth, all twelve models (of which seven are new) are estimated on a carefully constructed small network example with multiple choice sets for origin-destination pairs with both short and long distances. The estimation is done twice with two different synthetic datasets generated by MNP simulations based on the desired properties of random route utility, this allows benchmarking the models on the other dataset. The benchmark provides insight in the performance of the model if its applied on different networks, without re-estimation. Overall, when the models are estimated on one synthetic dataset and validated on the other synthetic dataset, the new models based on reference routes have a better validation than existing models. They can better approximate – without simulations – the MNP probabilities that take, amongst others, heteroscedasticity and correlation into account.

4.2. Random Route Utility Formulation

The random utility maximisation (RUM) framework is commonly used for describing route choice. Each route alternative is associated with some utility and the traveller is assumed to choose the route that provides the maximum expected utility. Since utility is not directly observable, analysts typically assume a structure of utility that includes measurable components and random terms that describe unmeasurable components. This section provides a route utility formulation with random foreseen travel times, that is used in Section 4.4 to assess the GMEV models.

Let $\mathcal{R}$ denote the set of relevant routes between a certain O-D pair. The systematic utility V_r for a route $r \in \mathcal{R}$ is decomposed in a part based on travel time τ_r, and 'other' systematic utility V_r^0, thus $V_r = V_r^0 + \beta\tau_r$, where τ_r is a random variate representing foreseen travel time, β is the travel time parameter. Then we write the random utility for route $r \in \mathcal{R}$ as[2],

$$U_r = V_r + \varepsilon_r = V_r^0 + \beta\tau_r + \varepsilon_r, \tag{4.1}$$

where ε_r is the analyst error. V_r^0 is assumed to be deterministic and can contain all route attributes other than the foreseen travel time, such as travel distance, running cost, number of traffic lights encountered, number of left turns, travel time variability, etc. Typical random

[2]The O-D pair and individual are not indexed since they are not relevant for this study and omitting them makes the formulas more readable.

component ε_r includes attributes that are considered by travellers when choosing their routes, but that are not included in $V_r^0 + \beta\tau_r$. Since the analyst does not have perfect knowledge on how the traveller gathers information about its travel time, we explicitly consider foreseen travel time as a random distribution τ_r in this work. In addition, travellers can have different preferences with respect to travel time and they do not perfectly measure the actual values of the attribute levels, but rather have a subjective perception that deviates from the actual values. Denote the analyst's estimate of foreseen travel time with $\hat{\tau}_r$, which generally does not consider all conditions (e.g., exact departure time, weather, information sources) the traveller is aware of. The deviation between the actual and estimate of foreseen travel time is typically larger for longer trips (see next Section). For example, when asking a traveller about the travel time of a recent short trip, his or her expectation may be off by one or two minutes, while for a long trip the expectation may be off by 10 minutes or more. We would like to point out that the foreseen travel time distribution does not capture travellers' disproportionate resilience towards high travel time variability, which can be captured by an (un)reliability term in V_r^0.

The utility in Equation (4.1) is very general since the analyst can include any attribute in V_r^0, and the, in our opinion, two most important sources of randomness for route utility are captured. Extension for errors around other terms in the systematic utility can be made in a straightforward fashion. We focus in this paper on the foreseen travel time because it is the most important factor in route choice. In addition, data is available to support assumptions on its structure.

Both τ_r and ε_r are random variables and are described by probability distributions. A common assumption for analyst errors ε_r is that they are independently and identically distributed (i.i.d.). The next section discusses the foreseen travel time distribution in more detail. Since τ_r and ε_r capture different sources of randomness, they can be considered independent, i.e. $\text{Cov}(\tau_r, \varepsilon_r) = 0$ for all $r \in R$.

4.2.1. Structure of foreseen travel time

Consider two similar road segments with lengths respectively 1 and 5 kilometres, and suppose the distribution of foreseen travel times for the 1-km road segment is known. There are two natural ways to determine the foreseen travel time distributions of the 5-km road segments, which impose different relations between their means and variances. The first method – used in all previous Probit studies (e.g., Yai et al., 1997; Daganzo and Sheffi, 1977) – is based on link-additivity and postulates that the foreseen travel time of the 5-km road segment is equally distributed as the sum of foreseen travel times of five independent 1-km road segments. This implies for each set of road segments for which the *additivity postulate* holds, that the ratio between the mean and the *variance* is equal for each segment. Link-additivity is convenient in transport networks, since overlap and heteroscedasticity of routes can automatically be captured when the normal distributed travel costs are drawn for each link and then summed per route. The second method is based on scaling and postulates that the foreseen travel time of the 5-km road segment is equally distributed as five times the foreseen travel time of one 1-km road segment. This implies for each set of road segments for which the *scaling postulate* holds, that the ratio between the mean and the *standard deviation* is equal for each segment. Furthermore, the convenience of link-additivity does not apply under the scaling postulate.

When an analyst wants to specify means and variances of foreseen travel times of routes, these two postulates each imply a different structure. Under the natural assumption that the mean equals the analyst's estimate $\hat{\tau}_r$, the variance is linear in $\hat{\tau}_r$ under the linear-additivity postulate and quadratic in $\hat{\tau}_r$ under the scaling postulate. In order to examine which postulate is more plausible, we first analyse the relationship between mean travel times and standard deviation of travel times in two datasets, and then discuss findings in the literature related to the postulates.

We have analysed household travel survey data from the Netherlands for the years 2010, 2011, and 2012 (Onderzoek Verplaatsingen in Nederland, OViN). In this survey, travellers are asked to state the travel times from trips they have made during a single day. We pooled data from trips with motorised private vehicles (i.e., car-driver, car-passenger, and motor bicycle) per 4-digit postal code and municipality[3]. After excluding O-D pairs with less than 25 observations we obtained 686 municipality aggregated O-D pairs and 216 postal code aggregated O-D pairs containing 38,106 and 8,079 trips, respectively. Figure 4.1 shows the relationship between the average travel time and the standard deviation of the perceived travel times for each O-D pair. The data shows a strong correlation and a linear relation between the average travel time and the standard deviation of travel time. Under the assumption that the traveller had complete information, the perceived travel time and foreseen travel time coincide, and the data is then in accordance with the scaling postulate. Clearly, the spread in reported travel times here is also the result of variations in traffic conditions and travel demand, routing, and aggregation of different households in each O-D pair. Unfortunately, we do not have the exact data on the foreseen travel time per route in the household travel surveys; however, we do have data from another smaller study about route choice and information.

This second data set comes from a study where 32 commuters between The Hague and Delft in the Netherlands were asked about the used information sources (see Ramos et al., 2012; Ramos, 2015). Some of them were equipped with Global Positioning System (GPS) devices for real-time information. After every trip they made, they were asked to indicate what their expected route and foreseen travel time for their next trip are. We aggregated this data per route per respondent, and determined the mean and standard deviation for all routes with at least 10 entries. In total 407 entries for 21 routes from 19 respondents[4] are used for the results in Figure 4.2. This dataset also indicates a linear relationship between the average and standard deviation. Thus both datasets indicate that the scaling postulate is more plausible than the additivity postulate.

Mahmassani et al. (2012) investigate travel time variability based on GPS and simulated trajectories and they conclude that a linear relation exists between the standard deviation of observed travel times and the average travel times. They furthermore reject a square root and quadratic relation. Mahmassani et al. (2013) provide the same conclusion at a network level, rather than just at link level. Fosgerau (2010) and Yildirimoglu et al. (2013) use data on time-of-day dependent relations between mean and standard deviation to show hysteresis. Their plots also strongly suggest a linear relation. Furthermore, Hellinga et al. (2012) took a year

[3] The four large municipalities Amsterdam, Utrecht, The Hague and Rotterdam are subdivided in respectively 7, 3, 5 and 5 districts.

[4] Two respondents reported two routes at least 10 times; 17 other respondents reported one route at least 10 times.

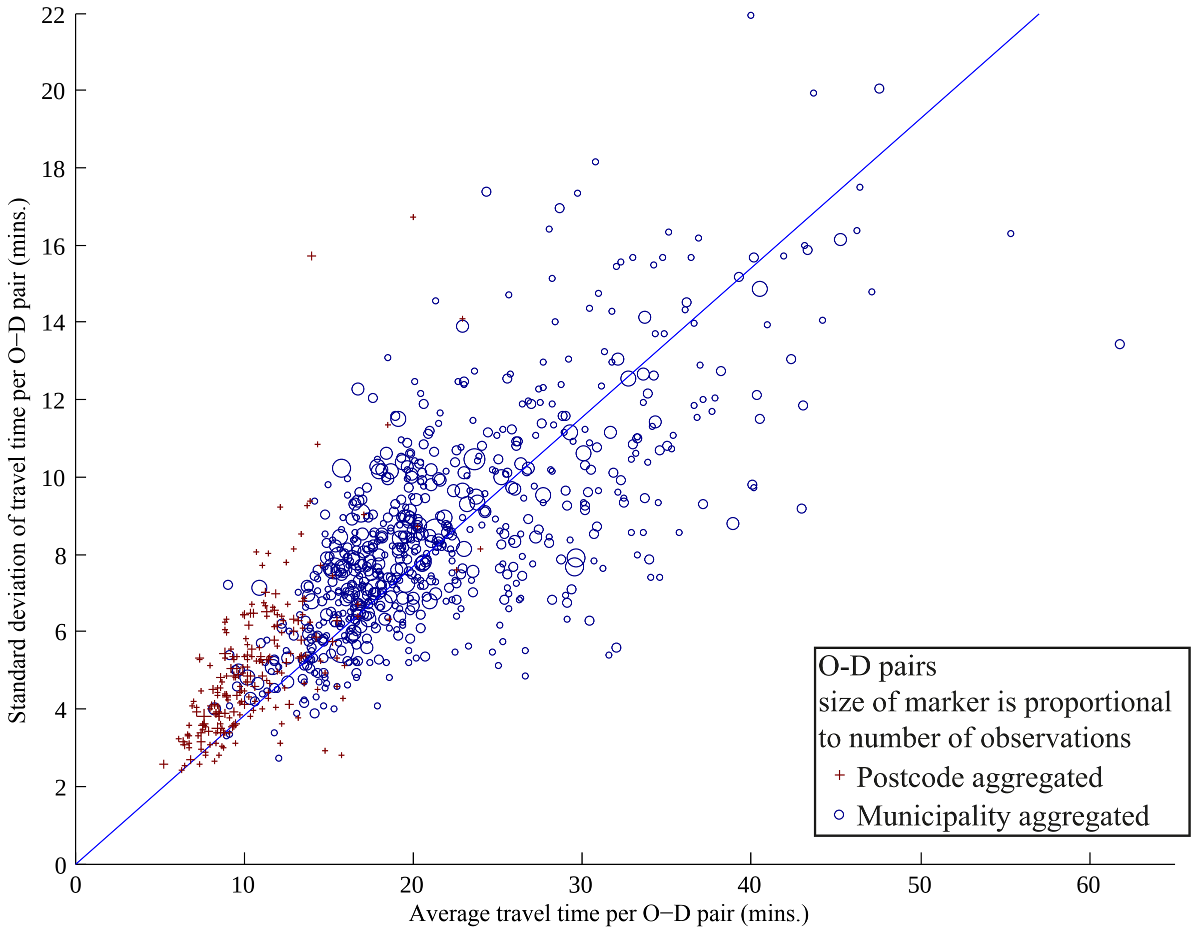


Figure 4.1.: Relation between the average and the standard deviation of revealed perceived travel times for an O-D pair, shown for 216 postcode O-D pairs and 686 municipality O-D pairs from Dutch national travel survey data. Linear regression between mean and standard deviation has a R-square of 0.9379. Linear regression between mean and variance has a R-square of 0.8016. The linear regression line has slope 0.3859.

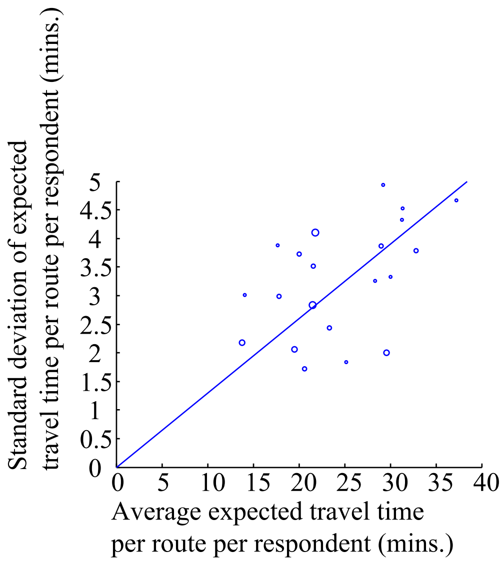


Figure 4.2.: Relation between the average and the standard deviation of foreseen travel times for a route from a respondent, shown for 21 routes from 19 respondents from trips between Delft and The Hague. The marker size is proportional to the number of entries per route. Linear regression between mean and standard deviation has an R-square of 0.9300. Linear regression between mean and variance has an R-square of 0.8401. The linear regression line has slope 0.1301.

of measurements from loop detector data on a motorway and aggregated each day into 15-minute averages. They also show a strong linear relationship between the standard deviation of the observed travel times and the average travel times. Despite that none of these studies analysed *foreseen* travel time distributions, they all indicate that the scaling postulate holds for their travel times. These results, the OViN data, and the actual foreseen travel time data, should convince that the scaling postulate is more plausible than the additivity postulate for distributions of foreseen route travel times.

Write $\sigma(\tau_r)$ as the standard deviation of foreseen travel time. Let this standard deviation be proportional to $\hat{\tau}_r$ with proportionality constant θ, then $\sigma(\tau_r) = \theta\hat{\tau}_r$

This finding contradicts the assumptions for MNP route choice models made by Daganzo and Sheffi (1977) and Yai et al. (1997). Their assumption of a linear relationship between the variance of travel times and $\hat{\tau}_r$, stems from the additivity postulate. In these models it is practical that route travel time can be readily obtained from the summation of normal distributed link travel times, but this seems to be unrealistic. It is possible to adjust the MNP-model to make it compatible with the scaling postulate by directly determining the (co)variance-matrix for all routes. However, this comes at the cost of link-additivity.

4.2.2. Covariances of foreseen travel times

Since routes may overlap, foreseen travel times are correlated. Route overlap induces a structure of covariances between travel times. This means that if two routes are largely overlapping, then also the foreseen travel time of both routes should be almost identical. Consider two routes, $r, s \in \mathcal{R}$. Let $\hat{\tau}_r = \hat{\tau}_{rs} + \hat{\tau}_{r\setminus s}$ be the decomposed travel time estimates of route r, where $\hat{\tau}_{rs}$ is the overlapping part with route s, and $\hat{\tau}_{r\setminus s}$ the non-overlapping part. The covariance between τ_r and τ_s should capture the overlap. We propose two candidate formulations for the covariance. One based on a correlation formulation derived from the literature, and one based on variance distribution over links.

The approaches from Daganzo and Sheffi (1977); Yai et al. (1997); Bekhor et al. (2002) rely on the additivity postulate, and the additivity over links leads to a definition of covariances. Under the scaling postulate these covariances are not directly usable due to the different relation between mean and standard deviation. However, the corresponding correlation, $\hat{\tau}_{rs}/\sqrt{\hat{\tau}_r\hat{\tau}_s}$, is independent of any relationship between mean and standard deviation, and can be used together with $\sigma(\tau_r) = \theta\hat{\tau}_r$. Then from the definition of correlation it follows that

$$\mathrm{Corr}(\tau_r,\tau_s) = \frac{\hat{\tau}_{rs}}{\sqrt{\hat{\tau}_r\hat{\tau}_s}} = \frac{\mathrm{Cov}(\tau_r,\tau_s)}{\sigma(\tau_r)\,\sigma(\tau_s)} \tag{4.2}$$

$$\Downarrow$$

$$\mathrm{Cov}(\tau_r,\tau_s) = \theta^2\hat{\tau}_r\hat{\tau}_s\frac{\hat{\tau}_{rs}}{\sqrt{\hat{\tau}_r\hat{\tau}_s}} = \theta^2\hat{\tau}_{rs}\sqrt{\hat{\tau}_r\hat{\tau}_s}. \tag{4.3}$$

An alternative formulation of the covariance can be obtained by using the definition $\mathrm{Var}(\tau_r - \tau_s) = \mathrm{Var}(\tau_r) + \mathrm{Var}(\tau_s) - 2\,\mathrm{Cov}(\tau_r,\tau_s)$ and an approximation of variance of the non-overlapping parts of travel time. Under the assumption that a route's variance is equally 'distributed' over the total travel time and that τ_r can be decomposed in independent parts, $\tau_r = \tau_{rs} + \tau_{r\setminus s}$, the variance associated with the non-overlapping part equals

$$\mathrm{Var}(\tau_{r\backslash s}) = \frac{\hat{\tau}_{r\backslash s}}{\hat{\tau}_r}\mathrm{Var}(\tau_r) = \frac{\hat{\tau}_{r\backslash s}}{\hat{\tau}_r}\theta^2\hat{\tau}_r^2 = \theta^2\hat{\tau}_r\hat{\tau}_{r\backslash s}. \tag{4.4}$$

Then the covariance equals

$$\begin{aligned}
\mathrm{Cov}(\tau_r,\tau_s) &= \frac{1}{2}\left(\mathrm{Var}(\tau_r)+\mathrm{Var}(\tau_s)-\mathrm{Var}(\tau_r-\tau_s)\right) && (4.5)\\
&= \frac{1}{2}\left(\mathrm{Var}(\tau_r)+\mathrm{Var}(\tau_s)-\mathrm{Var}(\tau_{r\backslash s}-\tau_{s\backslash r})\right) && (4.6)\\
&= \frac{1}{2}\theta^2\left(\hat{\tau}_r^2+\hat{\tau}_s^2-\hat{\tau}_r\hat{\tau}_{r\backslash s}-\hat{\tau}_s\hat{\tau}_{s\backslash r}\right) && (4.7)\\
&= \frac{1}{2}\theta^2\left(\hat{\tau}_r(\hat{\tau}_r-\hat{\tau}_{r\backslash s})+\hat{\tau}_s(\hat{\tau}_s-\hat{\tau}_{s\backslash r})\right) && (4.8)\\
&= \theta^2\hat{\tau}_{rs}\frac{\hat{\tau}_r+\hat{\tau}_s}{2}. && (4.9)
\end{aligned}$$

Notice that the covariance formulations of Equations (4.3) and (4.9) are respectively based on the geometric mean and arithmetic mean of travel times $\hat{\tau}_r$ and $\hat{\tau}_s$. Note that although we base the covariance on the travel time and its overlap, it is also possible to define it based on overlap of distance or total systematic utility.

4.2.3. Desired choice model properties

In summary, the assumptions on the error terms and the utility formulation in Equation (4.1) determine the expected value, variance and covariance of the utility.

- The expected value of utility of route $r \in \mathcal{R}$ is

$$\begin{aligned}
\mathbb{E}(U_r) &= \mathbb{E}\left(V_r^0+\beta\tau_r+\varepsilon_r\right), && (4.10)\\
&= \mathbb{E}\left(V_r^0\right)+\mathbb{E}\left(\beta\tau_r\right)+\mathbb{E}\left(\varepsilon_r\right), && (4.11)\\
&= V_r^0+\beta\hat{\tau}_r+\mathbb{E}\left(\varepsilon_r\right). && (4.12)
\end{aligned}$$

- The variance of the utility of route $r \in \mathcal{R}$ is

$$\begin{aligned}
\mathrm{Var}(U_r) &= \mathrm{Var}\left(V_r^0+\beta\tau_r+\varepsilon_r\right), && (4.13)\\
&= \mathrm{Var}\left(V_r^0\right)+\mathrm{Var}\left(\beta\tau_r\right)+\mathrm{Var}\left(\varepsilon_r\right), && (4.14)\\
&= \theta^2\hat{\tau}_r^2+\mathrm{Var}\left(\varepsilon_r\right). && (4.15)
\end{aligned}$$

- The covariance between routes $r,s \in \mathcal{R}$, $\mathrm{Cov}(U_r,U_s) = \mathrm{Cov}(\tau_r,\tau_s)$, equals either Equation (4.3) or (4.9).

Furthermore, when the route choice model is part of a transport network model, the following properties are desired:

- applicability to and compatibility between all types of (changing) networks, including those with O-D pairs with short and long distances, and including those with changing differences and ratios between routes;
- closed-form probabilities to avoid computationally expensive simulations;
- capture correlations due to overlap.

4.3. Random Route Utility Maximization models with Generalized Multivariate Extreme Value distributions

In this section an overview is given of choice models that are applicable for routes based on Random Utility Maximization (RUM) with closed-form expressions for the probabilities. This implies that MNP, Mixed Logit, and models with an error correction component (e.g., Logit kernel) are not considered; we refer to (Frejinger and Bierlaire, 2006; Prashker and Bekhor, 2004) for an overview of these simulation based models. However, since the MNP model is capable of completely capturing both the analyst error and the random foreseen travel time (and thus heteroscedasticity and correlation)[5], we will use MNP simulated data as ground truth in the illustrative example in Section 4.7. The existing models are all based on a single random variable, which should approximate both the analyst error and travel time distribution; Section 4.4 answers to which degree this is possible.

In this section the family of distributions will be deduced from RUM models with separable systematic and random utility components (i.e., U_r can be written as a function of V_r and ε_r). This leads to two types of Extreme Value (EV) models. First consider the additive utility formula $U_r = V_r + \varepsilon_r, \forall r \in \mathcal{R}$, where $(\varepsilon_1, \ldots, \varepsilon_{|\mathcal{R}|})$ is MEV distributed (McFadden, 1978); these models fall in the EV type I (Gumbel) category. Second, consider the multiplicative utility formula $U_r = V_r \varepsilon_r, \forall r \in \mathcal{R}$ and assume that $(-\ln \varepsilon_1, \ldots, -\ln \varepsilon_{|\mathcal{R}|})$ is MEV distributed (Fosgerau and Bierlaire, 2009; Gálvez, 2002); these models fall in the EV type III (reversed Weibull) category. The EV type II (Fréchet) is not suitable for route choice since Fréchet distributions have a lower bound on its support (i.e., there is a maximum on the route cost/travel time if such a model existed). The GMEV models consist of both the additive and multiplicative MEV models. Note that in older literature the MEV models were also named Generalized Extreme Value (GEV). GEV is a family of univariate distributions consisting of Gumbel, Fréchet and reversed Weibull. An MEV distribution is the joint distribution of multiple random variables with marginal distribution from one GEV type. MEV type I was discovered first by McFadden (1978) under the name 'The Generalized Extreme Value model'.

Other parametric approaches – not based directly on separable RUM – also exist. In (Castillo et al., 2008; Li, 2011; Nakayama, 2013; Nakayama and Chikaraishi, 2015; Chikaraishi and Nakayama, 2015) utilities $U_r, \forall r \in \mathcal{R}$ are independently Generalized Extreme Value (GEV) distributed. Li (2011) captures even a larger class of models, with independent route utilities. Fosgerau et al. (2013b) and Mattsson et al. (2014) provide the largest class of parametric approaches, where dependencies can be included as well. Note that all these models cannot be decomposed in systemtic and random components[6]. They directly parametrize the utility's probability distributions which is less explainable in behavioural terms. The exception is the derivation of the revised q-generalized logit model by Chikaraishi and Nakayama (2015) from the q-product function.

[5] Note that we use an adjusted MNP (compared to the literature) that is in line with the (co)variances under the scaling postulate.

[6] Except for their specific cases where they fit in the GMEV type.

4.3.1. The two types of Multivariate Extreme Value distributions

MEV distributions are very flexible since they allow for correlation between their variables. Generating functions are the core of MEV distributions and unfortunately these are not easily interpretable, making the theory rather complex. It is well-known that MEV type I models capture the MNL, Paired Combinatorial Logit and Link-Nested Logit models, but it is less known that C-Logit, Path-Size Logit and Path-Size Correction Logit belong to the MEV family. It is convenient that the latter models also fit in the generic framework since general results can be applied to them. All these mentioned 'logit based' models have an additive utility formula, so first the MEV derivation for this additive case is presented. Thereafter, the multiplicative MEV models are derived, and we show how they both fit in the GMEV framework.

4.3.1.1. Additive MEV models

The Additive (A) utility formula is $U_r^A = V_r + \varepsilon_r^A, \forall r \in \mathcal{R}$. Define the probability of choosing route $r \in \mathcal{R}$ as

$$P_r^A := \mathbb{P}\left(U_r^A \geq U_p^A, \quad \forall p \neq r\right) \tag{4.16}$$

$$= \mathbb{P}\left(V_r + \varepsilon_r^A \geq V_p + \varepsilon_p^A, \quad \forall p \neq r\right). \tag{4.17}$$

Following McFadden (1978), assume that the joint distribution of $\left(\varepsilon_1^A, \ldots, \varepsilon_{|\mathcal{R}|}^A\right)$ follows MEV type I distribution $\boldsymbol{\varepsilon}^{\text{MEV}}$ with cumulative distribution function

$$F_{\boldsymbol{\varepsilon}^{\text{MEV}}}(x_1, \ldots, x_{|\mathcal{R}|}) = e^{-G(e^{-x_1}, \ldots, e^{-x_{|\mathcal{R}|}})}, \tag{4.18}$$

where *generating function* $G : \mathbb{R}^{|\mathcal{R}|} \to \mathbb{R}$ satisfies

- μ-homogeneity:

$$G(az_1, \ldots, az_{|\mathcal{R}|}) = a^{\mu} G(z_1, \ldots, z_{|\mathcal{R}|}), \quad \forall a > 0 \tag{4.19}$$

- Limit property:

$$\lim_{z_r \to \infty} G(z_1, \ldots, z_{|\mathcal{R}|}) = \infty, \quad \forall r \in \mathcal{R} \tag{4.20}$$

- Alternating signs:

$$\frac{\partial^{|\hat{\mathcal{R}}|} G(z_1, \ldots, z_{|\mathcal{R}|})}{\prod_{r \in \hat{\mathcal{R}}} \partial z_r} (-1)^{|\hat{\mathcal{R}}|-1} \geq 0, \quad \forall \hat{\mathcal{R}} \subseteq \mathcal{R}, \hat{\mathcal{R}} \neq \emptyset. \tag{4.21}$$

The alternating signs property means that the sign of G switches every time an additional distinct partial derivative is taken. The generating function needs to satisfy the properties in order for $F_{\boldsymbol{\varepsilon}^{\text{MEV}}}$ to be a well-defined multivariate cumulative distribution function. With this utility formulation and distribution the following closed-form choice probabilities can be derived:

$$P_r^A = \mathbb{P}\left(U_r^A \geq U_p^A, \quad \forall p \neq r\right)$$
$$= \frac{y_r^A G_r(y_1^A, \ldots, y_{|\mathcal{R}|}^A)}{\mu G(y_1^A, \ldots, y_{|\mathcal{R}|}^A)}, \tag{4.22}$$

where $G_r(x_1,\dots,x_{|\mathcal{R}|}) = \frac{\partial G(x_1,\dots,x_{|\mathcal{R}|})}{\partial x_r}$ and $y_r^{\mathrm{A}} = e^{V_r}, \forall r \in \mathcal{R}$. More details on the generating function and the derivation of the choice probabilities can be found in (McFadden, 1978). By using Euler's homogeneous function theorem Equation (4.22) can be rewritten as

$$P_r^{\mathrm{A}} = \frac{y_r^{\mathrm{A}} G_r(y_1^{\mathrm{A}},\dots,y_{|\mathcal{R}|}^{\mathrm{A}})}{\sum_{p\in\mathcal{R}} y_p^{\mathrm{A}} G_p(y_1^{\mathrm{A}},\dots,y_{|\mathcal{R}|}^{\mathrm{A}})} \tag{4.23}$$

$$= \frac{e^{V_r} G_r(e^{V_1},\dots,e^{V_{|\mathcal{R}|}})}{\sum_{p\in\mathcal{R}} e^{V_p} G_p(e^{V_1},\dots,e^{V_{|\mathcal{R}|}})} \tag{4.24}$$

$$= \frac{e^{V_r + \ln G_r(e^{V_1},\dots,e^{V_{|\mathcal{R}|}})}}{\sum_{p\in\mathcal{R}} e^{V_p + \ln G_p(e^{V_1},\dots,e^{V_{|\mathcal{R}|}})}} \tag{4.25}$$

Since the denominator is independent of r and the formulation is similar to that of MNL, the probabilities provided in Equation (4.25) are the easiest to interpret. In the remainder of this paper the formulations of Equations (4.22) and (4.23) will be used because in the multiplicative case only the definition of the y_r^{A}'s will change, and these can then be substituted easily.

For a route set $\mathcal{R}$ the combination of a generating function G and a *generating vector* $\boldsymbol{y} = (y_1,\dots,y_{|\mathcal{R}|})$ completely determines choice probabilities with Equation (4.22). In other words, the route set, the generating function and the generating vector completely specify a choice model. Therefore, we can write the choice probabilities as a function of the generating function and generating vector for all types of GMEV models. For a choice model X applied on route set $\mathcal{R}$ with μ-homogeneous generating function G^X and generating vector $\boldsymbol{y}^X$, we can write the probability of choosing route r as

$$P_r^X(G^X;\boldsymbol{y}^X) := \frac{y_r^X G_r^X(\boldsymbol{y}^X)}{\mu G^X(\boldsymbol{y}^X)}, \quad \forall r \in \mathcal{R}. \tag{4.26}$$

Note that for the additive utility formula the generating vector is always equal to the exponentials of systematic utilities. Therefore define $\boldsymbol{y}^{\mathrm{A}} := (e^{V_1},\dots,e^{V_{|\mathcal{R}|}})$ as the *additive generating vector*. This additive generating vector should be used for all MEV models with an additive utility form, thus the generating function then defines choice probabilities.

4.3.1.2. Multiplicative MEV models

Before we present examples of the generating function, we first examine the multiplicative utility formula. Fosgerau and Bierlaire (2009) are the first to explore this field. In their analysis, the multiplicative formulation is transformed into an equivalent additive formulation. The analysis presented here is slightly different because we allow more flexibility on the scale parameter. The models in (Castillo et al., 2008; Kitthamkesorn and Chen, 2013; Nakayama, 2013) yield similar results for specific instances of generating functions, but the utility is not purely the product of systematic utility and an error term; instead, the utility is directly parametrized.

The Multiplicative (M) utility formula is

$$U_r^{\mathrm{M}} = V_r \varepsilon_r^{\mathrm{M}}, \quad \forall r \in \mathcal{R}, \tag{4.27}$$

where $V_r < 0$ and $\varepsilon_r^{\mathrm{M}} \geq 0$. This domain of the systematic utility is more restrictive than for additive MEV; however, in the route choice context it is natural that route costs are valued negatively, yielding negative systematic utility. By applying a log-transformation to Equation (4.16) and some further algebra, we can derive the probability of choosing route $r \in \mathcal{R}$ as:

$$P_r^{\mathrm{M}} = \mathbb{P}\left(V_r \varepsilon_r^{\mathrm{M}} \geq V_p \varepsilon_p^{\mathrm{M}}, \quad \forall p \neq r\right) \tag{4.28}$$

$$= \mathbb{P}\left(-V_r \varepsilon_r^{\mathrm{M}} \leq -V_p \varepsilon_p^{\mathrm{M}}, \quad \forall p \neq r\right) \tag{4.29}$$

$$= \mathbb{P}\left(\ln\left(-V_r \varepsilon_r^{\mathrm{M}}\right) \leq \ln\left(-V_p \varepsilon_p^{\mathrm{M}}\right), \quad \forall p \neq r\right) \tag{4.30}$$

$$= \mathbb{P}\left(\ln\left(-V_r\right) + \ln\left(\varepsilon_r^{\mathrm{M}}\right) \leq \ln\left(-V_p\right) + \ln\left(\varepsilon_p^{\mathrm{M}}\right), \quad \forall p \neq r\right) \tag{4.31}$$

$$= \mathbb{P}\left(-\ln\left(-V_r\right) - \ln\left(\varepsilon_r^{\mathrm{M}}\right) \geq -\ln\left(-V_p\right) - \ln\left(\varepsilon_p^{\mathrm{M}}\right), \quad \forall p \neq r\right) \tag{4.32}$$

Equation (4.32) has the same structure as the choice probabilities in Equation (4.17) of the additive choice model with systematic utilities $\tilde{V}_r = -\ln\left(-V_r\right), \forall r \in \mathcal{R}$ and error terms $\tilde{\varepsilon}_r^{\mathrm{A}} = -\ln\left(\varepsilon_r^{\mathrm{M}}\right)$. Therefore, we can apply the theory of the additive MEV with these variable substitutions. So in the multiplicative case, $\left(-\ln \varepsilon_1^{\mathrm{M}}, \ldots, -\ln \varepsilon_{|\mathcal{R}|}^{\mathrm{M}}\right)$ follows MEV type I distribution $\boldsymbol{\varepsilon}^{\mathrm{MEV}}$ (i.e., $(\varepsilon_1^{\mathrm{M}}, \ldots, \varepsilon_{|\mathcal{R}|}^{\mathrm{M}})$ is not similar to $\boldsymbol{\varepsilon}^{\mathrm{MEV}}$!), this distribution of $(\varepsilon_1^{\mathrm{M}}, \ldots, \varepsilon_{|\mathcal{R}|}^{\mathrm{M}})$ is called MEV type III. This is a multivariate distribution whose marginal distributions are of type reversed Weibull, and inherits the covariance structure from the generating function. So, it is the reversed Weibull equivalent of the model of McFadden (1978). Note that all generating functions can be applied on both types. The *multiplicative generating vector*, denoted with $\boldsymbol{y}^{\mathrm{M}}$, is derived from the additive generating vector and becomes

$$\boldsymbol{y}^{\mathrm{M}} = \left(e^{\tilde{V}_1}, \ldots, e^{\tilde{V}_{|\mathcal{R}|}}\right) = \left(e^{-\ln\left(-V_1\right)}, \ldots, e^{-\ln\left(-V_{|\mathcal{R}|}\right)}\right) = \left(\frac{-1}{V_1}, \ldots, \frac{-1}{V_{|\mathcal{R}|}}\right). \tag{4.33}$$

The probabilities of models with multiplicative utility formulas can be derived from Equation (4.26) by simply using the multiplicative generating vector $\boldsymbol{y}^{\mathrm{M}}$. By applying Euler's homogeneous function theorem again we can get

$$P_r = \frac{\frac{-G_r\left(-1/V_1, \ldots, -1/V_{|\mathcal{R}|}\right)}{V_r}}{\sum_{p \in \mathcal{R}} \frac{-G_p\left(-1/V_1, \ldots, -1/V_{|\mathcal{R}|}\right)}{V_p}}, \tag{4.34}$$

but the form of Equation (4.26) using $\boldsymbol{y}^{\mathrm{M}}$ is preferred. The main advantage of the multiplicative utility formula is that the utilities are automatically heteroscedastic. The standard deviation of the utility is proportional with the systematic utility and thus even the multiplicative equivalent of MNL is heteroscedastic.

A well-known property of additive models is that the probabilities are invariant under addition of a constant to all systematic utilities. Therefore, only differences in utilities matter in the MNL model. Analogously, the multiplicative models are invariant under multiplying all systematic utilities with the same factor (mini-proof: If $\lambda V_r \varepsilon_r^{\mathrm{M}} \geq \lambda V_p \varepsilon_p^{\mathrm{M}}$ then $V_r \varepsilon_r^{\mathrm{M}} \geq V_p \varepsilon_p^{\mathrm{M}}$), and only ratios between utilities matter in the multiplicative equivalent of MNL. In Section 4.4 its properties are further analysed.

4.3.1.3. Hybrid approach

Xu et al. (2015) introduce a hybrid model directly derived from the logit and weibit choice probability functions. This model inherits characteristics of as well logit as weibit. The model uses two scale parameters, one from logit (μ^{A}) and one from weibit (μ^{M}), and has choice probabilities equal to

$$P_r = \frac{e^{\mu^{\mathrm{A}} V_r}\left(-\frac{1}{V_r}\right)^{\mu^{\mathrm{M}}}}{\sum_{p\in\mathcal{R}} e^{\mu^{\mathrm{A}} V_p}\left(-\frac{1}{V_p}\right)^{\mu^{\mathrm{M}}}}, \quad \forall r \in \mathcal{R}. \tag{4.35}$$

Unfortunately, the model cannot be written as a RUM model with route r's utility in the form of $U_r = V_r \times \varepsilon^{\mathrm{M}} + \varepsilon^{\mathrm{A}}$ with a additive error term ε^{A} and multiplicative error term ε^{M}, which would have been a very neat solution. A straightforward hybrid approach in our framework would be to introduce generating vector $(-e^{V_1}/V_1,\ldots,-e^{V_{|\mathcal{R}|}}/V_{|\mathcal{R}|})$, i.e., the element-wise multiplication of $\boldsymbol{y}^{\mathrm{A}}$ and $\boldsymbol{y}^{\mathrm{M}}$. This will not lead to the approach of Xu et al. (2015), since only one scale parameter will be introduced when the vector is applied on a generating function.

However, the hybrid model of Xu et al. (2015) can be derived from either the A or M models from our framework by introducing an additional term in the systematic utility. In the additive form, the hybrid model of equation (4.35) is obtained with

$$U_r = V_r - \frac{\mu^{\mathrm{M}}}{\mu^{\mathrm{A}}}\ln(-x) + \varepsilon_r^{\mathrm{A}}, \quad \forall r \in \mathcal{R}, \tag{4.36}$$

where a logarithmic term is inserted to simulate the multiplicative model (Xu et al., 2015). In our framework, this can be represented with generating vector

$$\boldsymbol{y} = \left(e^{V_1-\rho\ln(-V_1)},\ldots,e^{V_{|\mathcal{R}|}-\rho\ln(-V_{|\mathcal{R}|})}\right) = \left(\frac{e^{V_1}}{(-V_1)^{\rho}},\ldots,\frac{e^{V_{|\mathcal{R}|}}}{(-V_{|\mathcal{R}|})^{\rho}}\right), \tag{4.37}$$

where $\rho > 0$ is a parameter representing the ratio between the scales. In the multiplicative form, the hybrid model of equation (4.35) is obtained with

$$U_r = V_r \times e^{-\frac{\mu^{\mathrm{A}}}{\mu^{\mathrm{M}}}V_r} \times \varepsilon_r^{\mathrm{M}}, \quad \forall r \in \mathcal{R}, \tag{4.38}$$

where a exponential factor is inserted to simulate the additive model. In the framework again, this becomes generating vector

$$\boldsymbol{y} = \left(\frac{-1}{V_1 e^{-\frac{V_1}{\rho}}},\ldots,\frac{-1}{V_{|\mathcal{R}|} e^{-\frac{V_{|\mathcal{R}|}}{\rho}}}\right) = \left(\frac{e^{\frac{V_1}{\rho}}}{-V_1},\ldots,\frac{e^{\frac{V_{|\mathcal{R}|}}{\rho}}}{-V_{|\mathcal{R}|}}\right), \tag{4.39}$$

with the same interpretation for ρ as above. The difference between the two generating vectors lies at the estimated scale parameter. With the latter formulation μ^{M} is estimated directly from the model and μ^{A} is derived from ρ. While with equation (4.37) μ^{A} is estimated directly and μ^{M} is derived from ρ. There is no difference in model outcomes and/or flexibility. Using any of the two approaches, it is possible to derive hybrid path-size, paired combinatorial, and link-nested models (see next section). The covariance structure provided in the latter two model instances are then only applied on the main error structure (i.e., dependent on which of the two generating functions is chosen).

4.3.2. Generating functions and model instances

To create instances of the family of GMEV route choice models, the generating function has to be specified. Together with one of the two derived generating vectors, Equation (4.26) will provide the closed-form choice probabilities of the choice model. Any function that satisfies Equations (4.19)-(4.21) is a generating function. The cross-nested logit uses one of the most general forms of the generating function and is analysed by Bierlaire (2006); Papola (2004). All presented generating functions here are also special cases of cross-nested logit models.

To address overlap on links, the presented model instances assume that each link is associated with one cost, namely systematic utility. This assumption is made for clarity reasons, differentiation into link length, travel and other cost can easily be done, also for the various factors and coefficients introduced below. Let $\mathcal{L}$ be a set of links and let the systematic utility of link $l \in \mathcal{L}$ be V_l. For each route $r \in \mathcal{R}$, $\mathcal{L}_r$ is the set of links of which r consists and the systematic route utility equals $V_r = \sum_{l \in \mathcal{L}_r} V_l$.

4.3.2.1. Multinomial

The well-known MNL model (see McFadden, 1974) is based on multinomial generating function

$$G^{\mathrm{MN}}(\boldsymbol{z}) := \sum_{r \in \mathcal{R}} z_r^{\mu}, \tag{4.40}$$

where $\mu > 0$ is the scale parameter. The traditional (additive) MN model, denoted Additive Multinomial (A-MN), has the following simple route choice probabilities,

$$P_r^{\text{A-MN}}(G^{\mathrm{MN}}; \boldsymbol{y}^{\mathrm{A}}) = \frac{y_r^{\mathrm{A}} G_r^{\mathrm{MN}}(\boldsymbol{y}^{\mathrm{A}})}{\mu G^{\mathrm{MN}}(\boldsymbol{y}^{\mathrm{A}})} = \frac{e^{\mu V_r}}{\sum_{p \in \mathcal{R}} e^{\mu V_p}}, \quad \forall r \in \mathcal{R}. \tag{4.41}$$

The multiplicative counterpart is the multinomial weibit model (or Kirchhoff distribution), denoted with Multiplicative Multinomial (M-MN) and has route choice probabilities

$$P_r^{\text{M-MN}}(G^{\mathrm{MN}}; \boldsymbol{y}^{\mathrm{M}}) = \frac{y_r^{\mathrm{M}} G_r^{\mathrm{MN}}(\boldsymbol{y}^{\mathrm{M}})}{\mu G^{\mathrm{MN}}(\boldsymbol{y}^{\mathrm{M}})} = \frac{\left(-\frac{1}{V_r}\right)^{\mu}}{\sum_{p \in \mathcal{R}} \left(-\frac{1}{V_p}\right)^{\mu}}, \quad \forall r \in \mathcal{R}. \tag{4.42}$$

4.3.2.2. Path-Size

We will first introduce a PS generating function to define a specific MEV type, and then we will show that models proposed in the literature are identical. For this PS generating function the path-size factor PS_r is required for each route $r \in \mathcal{R}$; this factor depicts the amount of overlap with other routes. Define PS generating function

$$G^{\mathrm{PS}}(\boldsymbol{z}) := \sum_{r \in \mathcal{R}} \mathrm{PS}_r^{\beta} z_r^{\mu}, \tag{4.43}$$

where $\mu > 0$ is the scale parameter and β is the path-size parameter. The Additive Path-Size (A-PS) model has choice probabilities

$$P_r^{\text{A-PS}}(G^{\text{PS}};\boldsymbol{y}^{\text{A}}) = \frac{y_r^{\text{A}} G_r^{\text{PS}}(\boldsymbol{y}^{\text{A}})}{\mu G^{\text{PS}}(\boldsymbol{y}^{\text{A}})} = \frac{\text{PS}_r^{\beta} e^{\mu V_r}}{\sum_{p\in\mathcal{R}} \text{PS}_p^{\beta} e^{\mu V_p}}, \quad \forall r \in \mathcal{R}. \tag{4.44}$$

The path-size-like models in the literature have a logarithmic term included in the utility. The MNL model with scale μ and systematic route utility $V_r + \gamma \ln x_r$ leads to the same choice probabilities as the additive MEV model with generating function $G^{\text{PS}}(\boldsymbol{z})$, generating vector $\boldsymbol{y} = (e^{V_1},\ldots,e^{V_{|\mathcal{R}|}})$, path-size factors $\text{PS}_r = x_r, \forall r \in \mathcal{R}$, and path-size parameter $\beta = \mu\gamma$. So, C-Logit, Path-Size Logit, and Path-Size Correction Logit can be translated into a A-PS model. C-Logit presented by Cascetta et al. (1996) adds the term $-\beta^{\text{CF}}\text{CF}_r$ to the utility, with parameter β^{CF} and commonality factor CF_r for each route r. This is equivalent to A-PS with $\text{PS}_r = \text{CF}_r$ and $\beta = -\mu\beta^{\text{CF}}$. Path-Size Logit presented by Ben-Akiva and Bierlaire (1999) adds the term $\beta^{\text{PS}}\tilde{\text{PS}}_r$ to the utility, with parameter β^{PS} and path-size factor $\tilde{\text{PS}}_r$ for each route r. This is equivalent to A-PS with $\text{PS}_r = \tilde{\text{PS}}_r$ and $\beta = \mu\beta^{\text{PS}}$. Path-Size Correction Logit (see Bovy et al., 2008) is equal in the same way. For those familiar with general cross-nested logit: note that G^{PS} is an instance of cross-nested logit with parametrized non-normalized inclusion factors with a single nest per alternative.

Different choices for PS_r are compared in (Frejinger and Bierlaire, 2006); they conclude that the best formulation for the path-size factor is

$$\text{PS}_r = \frac{\sum_{l\in\mathcal{L}_r} \frac{V_l}{\#_l}}{\sum_{l\in\mathcal{L}_r} V_l} = \frac{\sum_{l\in\mathcal{L}_r} \frac{V_l}{\#_l}}{V_r}, \tag{4.45}$$

where $\#_l = |\{r \in \mathcal{R} | l \in \mathcal{L}_r\}|$ is the number of routes using link l. If $\text{PS}_r = 1$ then r has no overlap with any other route. If $\text{PS}_r \to 0$ then r shares each link with many other routes.

The multiplicative counterpart of A-PS is denoted with Multiplicative Path-Size (M-PS) and has route choice probabilities

$$P_r^{\text{M-PS}}(G^{\text{PS}};\boldsymbol{y}^{\text{M}}) = \frac{y_r^{\text{M}} G_r^{\text{PS}}(\boldsymbol{y}^{\text{M}})}{\mu G^{\text{PS}}(\boldsymbol{y}^{\text{M}})} = \frac{\text{PS}_r^{\beta}\left(-\frac{1}{V_r}\right)^{\mu}}{\sum_{p\in\mathcal{R}} \text{PS}_p^{\beta}\left(-\frac{1}{V_p}\right)^{\mu}}, \quad \forall r \in \mathcal{R}. \tag{4.46}$$

Path-size is also added to MNW by Kitthamkesorn and Chen (2013, 2014), but they assume that $\beta = 1$, which thus leads to a specific instance of M-PS.

4.3.2.3. Paired Combinatorial

The (additive) PC Logit model as presented and analysed by Chen et al. (2003); Chu (1989); Gliebe et al. (1999); Koppelman and Wen (2000); Prashker and Bekhor (1998); Pravinvongvuth and Chen (2005) captures correlation between each pair of routes. For each route couple a nest is defined with a fixed nest specific scale. This scale is determined by the similarity index, denoted with φ_{rp}, for all $r \neq p \in \mathcal{R}$. Define paired combinatorial generating function

$$G^{\text{PC}}(\boldsymbol{z}) := \sum_{r\in\mathcal{R}} \sum_{p\in\mathcal{R}\setminus\{r\}} \left(z_r^{\frac{\mu}{1-\varphi_{rp}}} + z_p^{\frac{\mu}{1-\varphi_{rp}}} \right)^{1-\varphi_{rp}}, \tag{4.47}$$

where $\mu > 0$ is the scale parameter. Multiple definitions for the similarity index exist, but the most common one is

$$\varphi_{rp} := \frac{\sum_{l \in \mathcal{L}_r \cap \mathcal{L}_p} V_l}{\sqrt{V_r V_p}}, \quad \forall r,p \in \mathcal{R}. \tag{4.48}$$

The traditional Additive Paired Combinatorial (A-PC) model has route choice probabilities

$$P_r^{\text{A-PC}}(G^{\text{PC}};\mathbf{y}^{\text{A}}) = \frac{y_r^{\text{A}} G_r^{\text{PC}}(\mathbf{y}^{\text{A}})}{\mu G^{\text{PC}}(\mathbf{y}^{\text{A}})} = \frac{\sum_{p \in \mathcal{R}\setminus\{r\}} e^{\frac{\mu V_r}{1-\varphi_{rp}}} \left(e^{\frac{\mu V_r}{1-\varphi_{rp}}} + e^{\frac{\mu V_p}{1-\varphi_{rp}}} \right)^{-\varphi_{rp}}}{\sum_{r' \in \mathcal{R}} \sum_{p \in \mathcal{R}\setminus\{r'\}} \left(e^{\frac{\mu V_{r'}}{1-\varphi_{r'p}}} + e^{\frac{\mu V_p}{1-\varphi_{r'p}}} \right)^{1-\varphi_{r'p}}}, \quad \forall r \in \mathcal{R}. \tag{4.49}$$

The multiplicative counterpart Multiplicative Paired Combinatorial (M-PC) has route choice probabilities

$$\begin{aligned} P_r^{\text{M-PC}}(G^{\text{PC}};\mathbf{y}^{\text{M}}) &= \frac{y_r^{\text{M}} G_r^{\text{PC}}(\mathbf{y}^{\text{M}})}{\mu G^{\text{PC}}(\mathbf{y}^{\text{M}})} \\ &= \frac{\sum_{p \in \mathcal{R}\setminus\{r\}} \left(-\frac{1}{V_r}\right)^{\frac{\mu}{1-\varphi_{rp}}} \left(\left(-\frac{1}{V_r}\right)^{\frac{\mu}{1-\varphi_{rp}}} + \left(-\frac{1}{V_p}\right)^{\frac{\mu}{1-\varphi_{rp}}} \right)^{-\varphi_{rp}}}{\sum_{r' \in \mathcal{R}} \sum_{p \in \mathcal{R}\setminus\{r'\}} \left(\left(-\frac{1}{V_{r'}}\right)^{\frac{\mu}{1-\varphi_{r'p}}} + \left(-\frac{1}{V_p}\right)^{\frac{\mu}{1-\varphi_{r'p}}} \right)^{1-\varphi_{r'p}}}, \quad \forall r \in \mathcal{R}. \end{aligned} \tag{4.50}$$

4.3.2.4. Link-Nested

Vovsha and Bekhor (1998) are the first who applied the Cross-Nested Logit model to route choice such that links represent nests. In this (additive) LN Logit a nest is created for each link and all routes that use the link are included in the nest. The inclusion coefficient of each route in each link (or nest) is denoted with α_{lr} for all links $l \in \mathcal{L}$ and routes $r \in \mathcal{R}$. Define link-nested generating function

$$G^{\text{LN}}(\mathbf{z}) := \sum_{l \in \mathcal{L}} \left(\sum_{r \in \mathcal{R}} \alpha_{lr} z_r^{\mu_l} \right)^{\frac{\mu}{\mu_l}}, \tag{4.51}$$

where $\mu > 0$ is the scale parameter and $\mu_l > 0, \forall l \in \mathcal{L}$ are the link specific scale parameters. In general, it is difficult to estimate all link specific scale parameters in large networks. More on this topic can be found in (Bierlaire, 2006), where also the necessity of normalizing the inclusion coefficients is discussed. For this study the following normalized inclusion coefficients are applied

$$\alpha_{lr} = \begin{cases} \frac{V_l}{V_r} & \text{if } l \in \mathcal{L}_r \\ 0 & \text{otherwise} \end{cases}. \tag{4.52}$$

The traditional Additive Link-Nested (A-LN) has route choice probabilities

$$P_r^{\text{A-LN}}(G^{\text{LN}};\mathbf{y}^{\text{A}}) = \frac{y_r^{\text{A}} G_r^{\text{LN}}(\mathbf{y}^{\text{A}})}{\mu G^{\text{LN}}(\mathbf{y}^{\text{A}})} = \frac{\sum_{l \in \mathcal{L}} \alpha_{lr} e^{\mu_l V_r} \left(\sum_{p \in \mathcal{R}} \alpha_{lp} e^{\mu_l V_p} \right)^{\frac{\mu}{\mu_l}-1}}{\sum_{l \in \mathcal{L}} \left(\sum_{p \in \mathcal{R}} \alpha_{lp} e^{\mu_l V_p} \right)^{\frac{\mu}{\mu_l}}}, \quad \forall r \in \mathcal{R}. \tag{4.53}$$

The multiplicative counterpart Multiplicative Link-Nested (M-LN) with route choice probabilities

$$P_r^{\text{M-LN}}(G^{\text{LN}};\boldsymbol{y}^{\text{M}}) = \frac{y_r^{\text{M}} G_r^{\text{LN}}(\boldsymbol{y}^{\text{M}})}{\mu G^{\text{LN}}(\boldsymbol{y}^{\text{M}})} = \frac{\sum_{l\in\mathcal{L}} \alpha_{lr} \left(-\frac{1}{V_r}\right)^{\mu_l} \left(\sum_{p\in\mathcal{R}} \alpha_{lp} \left(-\frac{1}{V_p}\right)^{\mu_l}\right)^{\frac{\mu}{\mu_l}-1}}{\sum_{l\in\mathcal{L}} \left(\sum_{p\in\mathcal{R}} \alpha_{lp} \left(-\frac{1}{V_p}\right)^{\mu_l}\right)^{\frac{\mu}{\mu_l}}}, \quad \forall r \in \mathcal{R}. \tag{4.54}$$

4.3.2.5. Joint Network

Recently, Papola and Marzano (2013) introduced a new generating function based on an generalization of cross-nested generating functions to networks (see Daly and Bierlaire, 2006; Newman, 2008). Because their approach is based on links instead of routes, the generating function defines dependencies between links instead of routes. This is fundamentally different from the GMEV models presented in this work that is based on a route set and based on the error structure of random utility per route. The joint network formulation of Papola and Marzano (2013) is not a member of the GMEV models for route choice in this study since it cannot handle generic route sets as input. Their method has an implicit enumeration of routes. This has the advantage that the problem of choice set generation is dealt with internally. Unfortunately, this does not allow a utility formulation per route. Furthermore, the link based approach also delimits the routes that will be found. For their model specifically, scale parameters of 'nodes' depend on the shortest path to the destination. Their definition requires that a route advances through nodes for which the shortest path towards the destination decreases in every step. In real networks, the remaining shortest path will often increase when one deviates from the shortest path. This makes the model restrictive. Although a natural connection between a road network and a (recursive) network based generating function seems to be promising, Marzano (2014) suggests that the achievable covariance structure is not more advanced than that of cross-nested generating functions.

4.4. Qualitative Assessment of the Models

This section analyses the utility distribution of the GMEV models and assesses them qualitatively based on the desired properties presented in Section 4.2.3. The desired route utilities have two random variables (i.e., the random foreseen travel time and the analyst error) per route, but the models contain only one (GMEV -distributed) variable per route. We show that this variable can only resemble the analyst error in the additive models, and that it can resemble both errors in the multiplicative models, however, with completely correlated foreseen travel times and analyst errors.

4.4.1. Utility distribution

The results in this section for the additive models summarize the findings of McFadden (1978); Daly and Bierlaire (2006), while the derivations for the multiplicative models are slightly different from those in (Fosgerau and Bierlaire, 2009) that uses two scales. From the additive utility formulation and Equation (4.18) follows the multivariate utility distribution of the additive models:

$$F_{U^{\mathrm{A}}}(x_1,\ldots,x_{|\mathcal{R}|}) = \Pr\left(U_r^{\mathrm{A}} \leq x_r, \quad \forall r \in \mathcal{R}\right) \tag{4.55}$$

$$= \Pr\left(V_r + \varepsilon_r^{\mathrm{A}} \leq x_r, \quad \forall r \in \mathcal{R}\right) \tag{4.56}$$

$$= \Pr\left(\varepsilon_r^{\mathrm{A}} \leq x_r - V_r, \quad \forall r \in \mathcal{R}\right) \tag{4.57}$$

$$= F_{\boldsymbol{\varepsilon}^{\mathrm{MEV}}}\left(x_1 - V_1,\ldots,x_{|\mathcal{R}|} - V_{|\mathcal{R}|}\right) \tag{4.58}$$

$$= e^{-G\left(e^{-x_1+V_1},\ldots,e^{-x_{|\mathcal{R}|}+V_{|\mathcal{R}|}}\right)} \tag{4.59}$$

With some algebra, the multivariate utility distribution of the multiplicative models is written as

$$F_{U^{\mathrm{M}}}(x_1,\ldots,x_{|\mathcal{R}|}) = \Pr\left(U_r^{\mathrm{M}} \leq x_r, \quad \forall r \in \mathcal{R}\right) \tag{4.60}$$

$$= \Pr\left(V_r \varepsilon_r^{\mathrm{M}} \leq x_r, \quad \forall r \in \mathcal{R}\right) \tag{4.61}$$

$$= \Pr\left(-V_r \varepsilon_r^{\mathrm{M}} \geq -x_r, \quad \forall r \in \mathcal{R}\right) \tag{4.62}$$

$$= \Pr\left(\ln\left(-V_r \varepsilon_r^{\mathrm{M}}\right) \geq \ln(-x_r), \quad \forall r \in \mathcal{R}\right) \tag{4.63}$$

$$= \Pr\left(\ln(-V_r) + \ln\left(\varepsilon_r^{\mathrm{M}}\right) \geq \ln(-x_r), \quad \forall r \in \mathcal{R}\right) \tag{4.64}$$

$$= \Pr\left(-\ln\left(\varepsilon_r^{\mathrm{M}}\right) \leq \ln(-V_r) - \ln(-x_r), \quad \forall r \in \mathcal{R}\right) \tag{4.65}$$

$$= F_{\boldsymbol{\varepsilon}^{\mathrm{MEV}}}\left(\ln(-V_1) - \ln(-x_1),\ldots,\ln(-V_{|\mathcal{R}|}) - \ln(-x_{|\mathcal{R}|})\right) \tag{4.66}$$

$$= F_{\boldsymbol{\varepsilon}^{\mathrm{MEV}}}\left(\ln\frac{V_1}{x_1},\ldots,\ln\frac{V_{|\mathcal{R}|}}{x_{|\mathcal{R}|}}\right) \tag{4.67}$$

$$= e^{-G\left(e^{-\ln\frac{V_1}{x_1}},\ldots,e^{-\ln\frac{V_{|\mathcal{R}|}}{x_{|\mathcal{R}|}}}\right)} \tag{4.68}$$

$$= e^{-G\left(\frac{x_1}{V_1},\ldots,\frac{x_{|\mathcal{R}|}}{V_{|\mathcal{R}|}}\right)} \tag{4.69}$$

Then the marginal distribution for the utility of route $r \in \mathcal{R}$ is

$$F_{U_r^{\mathrm{A}}}(x_r) = \lim_{\{x_p \to \infty\}_{p \neq r}} F_{U^{\mathrm{A}}}\left(x_1,\ldots,x_{|\mathcal{R}|}\right) \tag{4.70}$$

$$= e^{-G\left(e^{-x_r+V_r}\mathbf{1}_r\right)} \tag{4.71}$$

$$= e^{-e^{\mu(-x_r+V_r)}G(\mathbf{1}_r)} \tag{4.72}$$

$$= e^{-e^{\mu(-x_r+V_r)+\ln G(\mathbf{1}_r)}} \tag{4.73}$$

$$= e^{-e^{\mu\left(-x_r+V_r+\frac{\ln G(\mathbf{1}_r)}{\mu}\right)}}, \tag{4.74}$$

for the additive case, which can be identified as a Gumbel distribution, and with notation $\mathbf{1}_r := (0,\ldots,0,1,0,\ldots,0)$ (1 on the r-th position).[7] For the multiplicative models the marginal

[7] Note that for $\mu = 1$ these are reversed exponential distributions.

distribution function of route r's utility is

$$F_{U_r^{\mathrm{M}}}(x_r) = \lim_{\{x_p \to 0\}_{p \neq r}} F_{U^{\mathrm{M}}}\left(x_1, \ldots, x_{|\mathcal{R}|}\right) \tag{4.75}$$

$$= e^{-G\left(\frac{x_r}{V_r}\mathbf{1}_r\right)} \tag{4.76}$$

$$= e^{-\left(\frac{x_r}{V_r}\right)^{\mu} G(\mathbf{1}_r)} \tag{4.77}$$

$$= e^{-\left(\frac{x_r}{V_r G(\mathbf{1}_r)^{\frac{-1}{\mu}}}\right)^{\mu}}, \tag{4.78}$$

which can be identified as a reversed Weibull distribution, where reversed means that $-U_r^{\mathrm{M}}$ is Weibull distributed .

The expected maximum utility is an important measure of choice models; it can be used to formulate the corresponding user equilibrium. Furthermore, it is input to define duality gaps. Denote $U^{*\mathrm{A}}$ as the maximum utility for the additive models, and $U^{*\mathrm{M}}$ as the maximum utility for the multiplicative models. Their distributions can be written as

$$F_{U^{*\mathrm{A}}}(x) = F_{U^{\mathrm{A}}}(x, \ldots, x) \tag{4.79}$$

$$= e^{-G\left(e^{-x+V_1}, \ldots, e^{-x+V_{|\mathcal{R}|}}\right)} \tag{4.80}$$

$$= e^{-e^{-\mu x} G\left(e^{V_1}, \ldots, e^{V_{|\mathcal{R}|}}\right)} \tag{4.81}$$

$$= e^{-e^{\mu\left(-x + \ln G\left(e^{V_1}, \ldots, e^{V_{|\mathcal{R}|}}\right)/\mu\right)}}, \text{ and} \tag{4.82}$$

$$F_{U^{*\mathrm{M}}}(x) = F_{U^{\mathrm{M}}}(x, \ldots, x) \tag{4.83}$$

$$= e^{-G\left(\frac{x}{V_1}, \ldots, \frac{x}{V_{|\mathcal{R}|}}\right)} \tag{4.84}$$

$$= e^{-(-x)^{\mu} G\left(\frac{-1}{V_1}, \ldots, \frac{-1}{V_{|\mathcal{R}|}}\right)} \tag{4.85}$$

$$= e^{-\left(\frac{x}{-G\left(\frac{-1}{V_1}, \ldots, \frac{-1}{V_{|\mathcal{R}|}}\right)^{\frac{-1}{\mu}}}\right)^{\mu}}, \tag{4.86}$$

which can be identified as Gumbel and reversed Weibull distributions again.

Table 4.1 shows the desired expected value, variance, and expected maximum utility together with the actual expected value, variance and expected maximum utility of the models. It is based on the identification of the marginal distributions as Gumbel and reversed Weibull, and on systematic utility specification $V_r = V_r^0 + \beta\tau_r$. Neither the additive nor the multiplicative formulation completely coincide with the desired result.

The expected value is not a problem for any model; any value can be achieved after normalization and identification, and the location of utility is not decisive for choice probabilities. On the other hand, the variance of the additive models is constant and the variance of the multiplicative models is affected by V_r^0, and for the latter $\mathrm{Var}(\varepsilon_r)$ is not directly represented. However, as the next section describes, the constant in the systematic utility of the multiplicative model does not have to be normalized. This constant will return as a constant term in

Table 4.1.: Expected value and variance of the desired, multiplicative and additive utility formulations

	Desired	Additive MEV (γ is Euler's constant)	Multiplicative MEV ($\Gamma(\cdot)$ is the Gamma function)
Utility U_r	$V_r^0 + \beta\tau_r + \varepsilon_r$	$V_r + \varepsilon_r^{\mathrm{A}} = V_r^0 + \beta\hat{\tau}_r + \varepsilon_r^{\mathrm{A}}$	$V_r \times \varepsilon_r^{\mathrm{M}} = \left(V_r^0 + \beta\hat{\tau}_r\right) \times \varepsilon_r^{\mathrm{M}}$
Expected value	$\underbrace{V_r^0 + \beta\hat{\tau}_r + \mathbb{E}(\varepsilon_r)}_{\text{affine in } \hat{\tau}_r}$	$\underbrace{V_r^0 + \beta\hat{\tau}_r + \frac{\ln G(\mathbf{1}_r) + \gamma}{\mu}}_{\text{affine in } \hat{\tau}_r}$	$\underbrace{\frac{V_r^0 + \beta\hat{\tau}_r}{G(\mathbf{1}_r)^{1/\mu}}\Gamma\left(1 + \frac{1}{\mu}\right)}_{\text{affine in } \hat{\tau}_r}$
Variance	$\underbrace{\theta^2\hat{\tau}_r^2 + \mathrm{Var}(\varepsilon_r)}_{\text{quadratic in } \hat{\tau}_r}$	$\underbrace{\frac{\pi^2}{6}\left(\frac{1}{\mu}\right)^2}_{\text{constant}}$	$\underbrace{\frac{\left(V_r^0 + \beta\hat{\tau}_r\right)^2}{G(\mathbf{1}_r)^{2/\mu}}\left(\Gamma\left(1 + \frac{2}{\mu}\right) - \Gamma\left(1 + \frac{1}{\mu}\right)^2\right)}_{\text{quadratic in } V_r^0 + \beta\hat{\tau}_r}$
Expected $\max_{r\in\mathcal{R}} U_r$	Not closed form	$\frac{\ln G\left(e^{V_1}, \ldots, e^{V_{\lvert\mathcal{R}\rvert}}\right) + \gamma}{\mu}$	$-G\left(\frac{-1}{V_1}, \ldots, \frac{-1}{V_{\lvert\mathcal{R}\rvert}}\right)^{\frac{-1}{\mu}} \Gamma\left(1 + \frac{1}{\mu}\right)$

the variance, and can thus resemble Var(ε_r). Despite that the multiplicative models' variances can capture the variances from the foreseen travel time distribution and analyst error simultaneously, they are both represented by ε_r and thus completely dependent. This is not in line with the desired independence.

Closed form formulations of the covariance of GMEV models are not known. Marzano et al. (2013); Marzano (2014) present a more tractable expression and a method to calculate them for additive MEV. The covariances of the additive models are based on the constant variance, and will therefore not change when the systemic utility changes. However, the systematic utility enters the variance quadratically in the multiplicative model, and therefore its covariance will also in- and decrease together with it. Thus, the multiplicative models capture the covariances better.

The expected maximum utility for the A-MN model indeed return in the well-known logsum that can by used for an hierarchical derivation of the other models. This was first identified by Ben-Akiva (1973).

4.4.2. Stochastic user equilibrium formulation

The derived GMEV route choice models all have an equivalent stochastic user equilibrium formulation. Mathematical programming formulation are known for all logit based models (Bekhor and Prashker, 1999; Fisk, 1980), for the M-MN and (simplified) M-PS models (Kitthamkesorn and Chen, 2013, 2014), for the q-generalized logit model (Chikaraishi and Nakayama, 2015), and for the hybrid logit-weibit model (Xu et al., 2015). It is straightforward to derive the variational inequality formulation – that we discuss – from a mathematical programming formulation. Chen (1999); Guo et al. (2010) describe a stochastic user equilibrium for the MNL model using a variational inequality. Zhou et al. (2012) provide two variational inequality formulations for the C-logit model. Their first formulation is specific for C-logit and a special case of the formulation we present below. Their second formulation is more generic and works with any choice probability formula, but it returns different generalised costs than the approach we present below.

Since the choice probabilities of each GMEV model can be written solely in terms of the generating function and generating vector, this also holds for the corresponding variational inequality formulation. Denote the demand with D, the flow for route $r \in \mathcal{R}$ as f_r, and $\boldsymbol{f}$ as the vector of flows. Since, amongst other attributes, the travel time in the systematic utility depends on the flow, write that the generating vector now depends on $\boldsymbol{y}(\boldsymbol{f})$[8]. For clarity, but without loss of generality, only one route set (i.e., one O-D pair) is considered.

Consider the following Variational Inequality (VI) formulation; find equilibrium flow $\boldsymbol{f}^* =$

[8]When the generating vector is substituted, just replace V_r with $V_r(\boldsymbol{f})$ to denote dependency of attributes on flow.

$(f_1^*, \ldots, f_{|\mathcal{R}|}^*)$, such that

$$\sum_{r \in \mathcal{R}} \left(-\ln\left(y_r(\boldsymbol{f}^*) G_r(\boldsymbol{y}(\boldsymbol{f}^*))\right) + \ln(f_r^*)\right)(f_r - f_r^*) \geq 0, \quad \forall \boldsymbol{f} \in \Omega,$$
$$\text{where } \Omega = \left\{ \boldsymbol{f} \in \mathbb{R}^{|\mathcal{R}|} \middle| f_r > 0, \sum_{r \in \mathcal{R}} f_r = D \right\}. \tag{4.87}$$

Define the generalized stochastic cost for route $r \in \mathcal{R}$ under flow $\boldsymbol{f}$ as[9]

$$c_r(\boldsymbol{f}) := -\ln\left(y_r(\boldsymbol{f}) G_r(\boldsymbol{y}(\boldsymbol{f}))\right) + \ln(f_r). \tag{4.88}$$

The corresponding Karush-Kuhn-Tucker system (see e.g., Facchinei and Pang, 2003, Proposition 1.2.1) is then to find multipliers κ and $\lambda_1, \ldots, \lambda_{|\mathcal{R}|}$ for which

$$0 = -\ln\left(y_r(\boldsymbol{f}^*) G_r(\boldsymbol{y}(\boldsymbol{f}^*))\right) + \ln(f_r^*) + \kappa - \lambda_r, \quad \forall r \in \mathcal{R}, \tag{4.89}$$
$$\sum_{r \in \mathcal{R}} f_r = D, \tag{4.90}$$
$$0 \leq \lambda_r, \quad \forall r \in \mathcal{R}, \tag{4.91}$$
$$0 < f_r, \quad \forall r \in \mathcal{R}, \text{ and} \tag{4.92}$$
$$0 = \sum_{r \in \mathcal{R}} \lambda_r f_r \tag{4.93}$$

hold. Equations (4.91), (4.92) and (4.93) imply that $\lambda_r = 0$ for all $r \in \mathbb{R}$. Rewriting equation (4.89) then gives

$$\ln\left(y_r(\boldsymbol{f}^*) G_r(\boldsymbol{y}(\boldsymbol{f}^*))\right) = \ln(f_r^*) + \kappa, \quad \forall r \in \mathcal{R} \tag{4.94}$$
$$y_r(\boldsymbol{f}^*) G_r(\boldsymbol{y}(\boldsymbol{f}^*)) = f_r^* e^{\kappa}, \quad \forall r \in \mathcal{R}. \tag{4.95}$$

Summing over routes gives

$$\sum_{r \in \mathcal{R}} y_r(\boldsymbol{f}^*) G_r(\boldsymbol{y}(\boldsymbol{f}^*)) = D e^{\kappa}, \quad \forall r \in \mathcal{R}. \tag{4.96}$$

Finally, dividing equation (4.95) by equation (4.96) gives

$$\frac{y_r(\boldsymbol{f}^*) G_r(\boldsymbol{y}(\boldsymbol{f}^*))}{\sum_{s \in \mathcal{R}} y_s(\boldsymbol{f}^*) G_s(\boldsymbol{y}(\boldsymbol{f}^*))} = \frac{f_r^*}{D}, \tag{4.97}$$

which actually coincides with the choice probability definition of Equation (4.23). Therefore, VI problem (4.87) describes the stochastic user equilibrium with the choice model that is defined by generating function G and generating vector $\boldsymbol{y}$.

Such a VI formulation is useful, since the corresponding theory can be used to analyse existence and uniqueness of stochastic user equilibria (as for example done by Bliemer et al.

[9] Note that this cost is unit-less. For MNL, it is possible to normalize it such that the systemic utility appears as a term.

(2014b); Nagurney (1998)). In addition, the VI formulation provided the generalized stochastic cost per route – which should all be equal in equilibrium –, that can be used in assignment algorithms to determine duality gaps. Namely, as $\hat{\boldsymbol{f}} \to \boldsymbol{f}^*$, then

$$\frac{\sum_{r\in\mathcal{R}} \hat{f}_r \left(c_r(\hat{\boldsymbol{f}}) - \min_{s\in\mathcal{R}} c_s(\hat{\boldsymbol{f}})\right)}{\sum_{r\in\mathcal{R}} \hat{f}_r \min_{s\in\mathcal{R}} c_s(\hat{\boldsymbol{f}})} \to 0. \tag{4.98}$$

4.4.3. Normalization, identification, and invariance

The constant in the systematic utility (i.e., the alternative specific constant) of the logit-based (i.e., additive) models has to be normalized because only differences in utility matter. Furthermore, one of the attribute parameters or the scale has to be normalized due to identification. If the cost parameter is normalized to one, all other attribute parameters can be interpreted as willingness to pay for its attribute.

The multiplicative models however, do not require the constant in the utility to be normalized. This constant is also multiplied with the error term and thus this term is not equal amongst alternatives. This allows the modeller to use an additional parameter. Similar as for the additive case, one of the attribute parameters or the scale has to be normalized due to identification.

For logit based additive models only differences between utilities matter for the choice probabilities. Similarly, only the ratios between utilities matter for the choice probabilities in the multiplicative models (Fosgerau and Bierlaire, 2009). This means that the additive models are invariant under addition with a constant and the multiplicative models are invariant under multiplication with a constant. As the next example shows, it is doubtful that these properties are realistic and they limit both models. For the additive case consider two route sets that only differ by a constant, one with route costs (e.g., travel time) $\{1,6\}$ and one with $\{100,105\}$; it is not expected that the choice probabilities are the same for these routes sets. Analogously, route costs $\{2,3\}$ and $\{40,60\}$ should neither reflect the same probabilities for the multiplicative case. Nevertheless, since the constant in the systematic utility does not have to be normalized, the property is less restrictive for the multiplicative case. Section 4.6 provides further analysis.

4.5. Multiplicative MEV Models with Explicit Removal of Overlap

We present an adjustment to the multiplicative model based on the decision rule that travellers only compare the non-overlapping part of routes compared to a reference route. This relaxes the invariance properties to a certain extent, and implies a different choice mechanism for travellers. Probabilities for switching to another route and staying at the reference route are determined. This different decision rule can be retrieved from an alternative utility formulation, which allows a qualitative analysis of the model.

Using the systematic utility per link, and the multiplicative utility formula, let

$$U_r^{\mathrm{M}} = \left(\sum_{l\in\mathcal{L}_r} V_l\right)\varepsilon_r^{\mathrm{M}}, \quad \forall r \in \mathcal{R}.$$

The choice probability for alternative $r \in \mathcal{R}$ is

$$\begin{aligned}
P_r^{\mathrm{M}} &= \mathbb{P}(U_r^{\mathrm{M}} \geq U_p^{\mathrm{M}}, \forall p \neq r) && (4.99)\\
&= \mathbb{P}\left(\left(\sum_{l\in\mathcal{L}_r} V_l\right)\varepsilon_r^{\mathrm{M}} \geq \left(\sum_{l\in\mathcal{L}_p} V_l\right)\varepsilon_p^{\mathrm{M}}, \quad \forall p \neq r\right) && (4.100)\\
&= \mathbb{P}\left(\sum_{l\in\mathcal{L}_r} V_l\varepsilon_r^{\mathrm{M}} \geq \sum_{i\in\mathcal{L}_p} V_l\varepsilon_p^{\mathrm{M}}, \quad \forall p \neq r\right) && (4.101)\\
&= \mathbb{P}\left(\sum_{l\in\mathcal{L}_r\setminus\mathcal{L}_p} V_l\varepsilon_r^{\mathrm{M}} + \sum_{l\in\mathcal{L}_r\cap\mathcal{L}_p} V_l\varepsilon_r^{\mathrm{M}} \geq \sum_{l\in\mathcal{L}_p\setminus\mathcal{L}_r} V_l\varepsilon_p^{\mathrm{M}} + \sum_{l\in\mathcal{L}_r\cap\mathcal{L}_p} V_l\varepsilon_p^{\mathrm{M}}, \quad \forall p \neq r\right). && (4.102)
\end{aligned}$$

In this formulation, each part of random utility is associated with a part of systematic utility according to the scaling postulate. The systematic utility that is shared amongst a pair of alternatives appears at both sides of the inequality. The shared links for route pair r,p are contained in $\mathcal{L}_r \cap \mathcal{L}_p$. An individual will evaluate the links independent of the alternatives they belong to. Therefore we assume that the part of the random utility belonging to these shared links is fully correlated. This means that for all $l \in \mathcal{L}_r \cap \mathcal{L}_p$ the terms $V_l\varepsilon_r^{\mathrm{M}}$ and $V_l\varepsilon_p^{\mathrm{M}}$ in Equation (4.102) are equal and can be subtracted from both sides of the inequality.

This can be derived in an econometrical sound fashion by reconsidering the utility formula based on a reference route. Let r be the reference route, and denote the utility for each route $p \in \mathcal{R}$ as

$$U_p^{\mathrm{M}\Delta,r} := \begin{cases} \sum_{l\in\mathcal{L}_r} V_l\varepsilon_r^{\mathrm{M}} & \text{if } p = r \\ \sum_{l\in\mathcal{L}_p\setminus\mathcal{L}_r} V_l\varepsilon_p^{\mathrm{M}} + \sum_{l\in\mathcal{L}_r\cap\mathcal{L}_p} V_l\varepsilon_r^{\mathrm{M}} & \text{otherwise.} \end{cases} \tag{4.103}$$

The utility that overlaps with the reference route, is multiplied with the error term of the reference route, while the remainder has its own error term. We refer to this type of probabilities as the *Multiplicative with Reference Route* (MΔ)*-case* with reference route r. The probability of choosing route $p \in \mathcal{R}$, conditional on reference route r is

$$P_p^{\mathrm{M}\Delta,r} = \begin{cases} \mathbb{P}\left(\sum_{l\in\mathcal{L}_{p\setminus s}} V_l\varepsilon_p^{\mathrm{M}} \geq \sum_{l\in\mathcal{L}_{s\setminus p}} V_l\varepsilon_s^{\mathrm{M}}, \quad \forall s \neq p\right) & \text{if } p = r \\ \mathbb{P}\left(\sum_{l\in\mathcal{L}_{p\setminus r}} V_l\varepsilon_p^{\mathrm{M}} + \sum_{l\in\mathcal{L}_r\cap\mathcal{L}_p} V_l\varepsilon_r^{\mathrm{M}} \geq U_s^{\mathrm{M}\Delta,r}, \quad \forall s \neq p\right) & \text{otherwise,} \end{cases} \tag{4.104}$$

with notation $\mathcal{L}_{r\setminus p} = \mathcal{L}_r \setminus \mathcal{L}_p$, and the overlapping parts are subtracted for case $p = r$ (see Equation (4.102)). First, we analyse the case $p = r$. That choice probability reads that the utility of non-overlapping links of routes s with p is smaller than the utility of the non-overlapping links of route p with each respective s. Thus, it is the probability that the traveller does not benefit from swapping some links and change route. Now isolate ε_r on the left hand side;

$$\text{if } p=r\text{, then } P_p^{\mathrm{M}\Delta,r} = \mathbb{P}\left(\left(\sum_{l\in\mathcal{L}_{p\setminus s}} V_l\right)\varepsilon_p^{\mathrm{M}} \geq \left(\sum_{l\in\mathcal{L}_{s\setminus p}} V_l\right)\varepsilon_s^{\mathrm{M}}, \quad \forall s\neq p\right) \tag{4.105}$$

$$= \mathbb{P}\left(\varepsilon_p^{\mathrm{M}} \leq \frac{\sum_{l\in\mathcal{L}_{s\setminus p}} V_l}{\sum_{l\in\mathcal{L}_{p\setminus s}} V_l}\varepsilon_s^{\mathrm{M}}, \quad \forall s\neq p\right). \tag{4.106}$$

With the same algebra that derives Equation (4.32) from Equation (4.28), this multiplicative formulation can be transformed to an additive formulation:

$$\text{if } p=r\text{, then } P_p^{\mathrm{M}\Delta,r} = \mathbb{P}\left(\varepsilon_p^{\mathrm{M}} \leq \frac{\sum_{l\in\mathcal{L}_{s\setminus p}} V_l}{\sum_{l\in\mathcal{L}_{p\setminus s}} V_l}\varepsilon_s^{\mathrm{M}}, \quad \forall s\neq p\right) \tag{4.107}$$

$$= \mathbb{P}\left(\ln(\varepsilon_p^{\mathrm{M}}) \leq \ln\left(\frac{\sum_{l\in\mathcal{L}_{s\setminus p}} V_l}{\sum_{l\in\mathcal{L}_{p\setminus s}} V_l}\varepsilon_s^{\mathrm{M}}\right), \quad \forall s\neq p\right) \tag{4.108}$$

$$= \mathbb{P}\left(\ln(\varepsilon_p^{\mathrm{M}}) \leq \ln\left(\frac{\sum_{l\in\mathcal{L}_{s\setminus p}} V_l}{\sum_{l\in\mathcal{L}_{p\setminus s}} V_l}\right) + \ln\left(\varepsilon_s^{\mathrm{M}}\right), \quad \forall s\neq p\right) \tag{4.109}$$

$$= \mathbb{P}\left(-\ln(\varepsilon_p^{\mathrm{M}}) \geq -\ln\left(\frac{\sum_{l\in\mathcal{L}_{s\setminus p}} V_l}{\sum_{l\in\mathcal{L}_{p\setminus s}} V_l}\right) - \ln\left(\varepsilon_s^{\mathrm{M}}\right), \quad \forall s\neq p\right) \tag{4.110}$$

For the case $p\neq r$ we can show that;

$$\text{if } p\neq r \text{ then } P_p^{\mathrm{M}\Delta,r} = \mathbb{P}\left(\sum_{l\in\mathcal{L}_{p\setminus r}} V_l\varepsilon_p^{\mathrm{M}} + \sum_{l\in\mathcal{L}_r\cap\mathcal{L}_p} V_l\varepsilon_r^{\mathrm{M}} \geq U_s^{\mathrm{M}\Delta,r}, \quad \forall s\neq p\right) \tag{4.111}$$

$$= \mathbb{P}\left(\frac{\sum_{l\in\mathcal{L}_{p\setminus r}} V_l}{\sum_{l\in\mathcal{L}_{r\setminus p}} V_l}\varepsilon_p^{\mathrm{M}} \leq \varepsilon_r^{\mathrm{M}} \wedge \left(\frac{\sum_{l\in\mathcal{L}_{p\setminus s}} V_l}{\sum_{l\in\mathcal{L}_{s\setminus p}} V_l}\varepsilon_p^{\mathrm{M}} \leq \frac{\sum_{l\in\mathcal{L}_{s\setminus p}} V_l}{\sum_{l\in\mathcal{L}_{p\setminus s}} V_l}\varepsilon_s^{\mathrm{M}}, \forall s\neq r,p\right)\right) \tag{4.112}$$

This equality is not trivial, one should check that Equation (4.112) holds or fails for all six orderings of U_r, U_p, and U_s by using and/or substituting the condition in Equation (4.107). So, if route p is chosen under reference route r, then it is beneficial to switch from route r to route p, and this improvement is larger than to switch to any other route s. Equation (4.112) can be rewritten in an additive form similar to (4.110) by taking the logarithm transformation. Assume – similar as in the M-models – that in the derived additive forms $\left(-\ln(\varepsilon_1^{\mathrm{M}}),\ldots,-\ln(\varepsilon_{|\mathcal{R}|}^{\mathrm{M}})\right)$ follows MEV distribution $\boldsymbol{\varepsilon}^{\mathrm{MEV}}$, then we have another type of GMEV route choice models. The generating vector is reference route specific, denoted with $\boldsymbol{y}^{\mathrm{M}\Delta,r}$, and given by

$$y_p^{\mathrm{M}\Delta,r} = \begin{cases} 1 & \text{if } r=p \\ \frac{\sum_{l\in\mathcal{L}_{r\setminus p}} V_l}{\sum_{l\in\mathcal{L}_{p\setminus r}} V_l} & \text{otherwise} \end{cases}, \quad \forall r,p\in\mathcal{R}, \tag{4.113}$$

where we used that $e^0 = 1$ and

$$e^{-\ln\left(\frac{\sum_{l\in\mathcal{L}_{p\setminus r}} V_l}{\sum_{l\in\mathcal{L}_{r\setminus p}} V_l}\right)} = \frac{\sum_{l\in\mathcal{L}_{r\setminus p}} V_l}{\sum_{l\in\mathcal{L}_{p\setminus r}} V_l}.$$

Given any generating function G, reference route r and MΔ *generating vector* $\mathbf{y}^{\text{M}\Delta,r}$ the choice probabilities can be derived with Equation (4.26). By then applying Euler's homogeneous function theorem, the probability of choosing route $p \in \mathcal{R}$ with reference route $r \in \mathcal{R}$ is:

$$P_p^{\text{M}\Delta,r} = \frac{y_p^{\text{M}\Delta,r} G_p(\mathbf{y}^{\text{M}\Delta,r})}{G_r(\mathbf{y}^{\text{M}\Delta,r}) + \sum_{\{p\in\mathcal{R}|p\neq r\}} \frac{\sum_{l\in\mathcal{L}_{p\setminus r}} V_l}{\sum_{l\in\mathcal{L}_{r\setminus p}} V_l} G_p(\mathbf{y}^{\text{M}\Delta,r})}. \tag{4.114}$$

Note that if no overlap exists the MΔ-models collapse to the M-models (i.e., they return the same probabilities).

The previous analysis was based on one reference route, but for applications there is not always a (single) reference route available since the (current) reference of the travellers is not known. The final step is to handle this uncertainty. Denote $P^{\text{ref}}(r)$ as the probability that route r is the reference route. The final choice probability for route p then becomes

$$P_p^{\text{M}\Delta} = \sum_{r\in\mathcal{R}} P_p^{\text{M}\Delta,r} P^{\text{ref}}(r). \tag{4.115}$$

We consider three ways to determine the $P^{\text{ref}}(r)$. First, it is possible to assign every route as reference route with the same probability (i.e., $P^{\text{ref}}(r) = 1/|\mathcal{R}|, \forall r \in \mathcal{R}$). This seems only realistic when no irrelevant routes exist in the choice set, since it is not plausible that travellers use an irrelevant route as reference. Second, it is very natural to set the probability that a route is chosen equal to the probability that a route is the reference route. Then Equation (4.115) becomes a system of equations: $P_p^{\text{M}\Delta} = \sum_{r\in\mathcal{R}} P_p^{\text{M}\Delta,r} P_p^{\text{M}\Delta}$ (where $P_p^{\text{M}\Delta}, p \in \mathcal{R}$ are the unknowns). Furthermore, $\sum_{p\in\mathcal{R}} P_p^{\text{M}\Delta} = 1$, thus this system of equations can be identified as a Markov chain. Since $P_p^{\text{M}\Delta,r} > 0, \forall p,r \in \mathcal{R}$, a steady state exists which can be found by solving the system of equations. A third possibility is to fix one route (e.g., the fastest in free-flow conditions) as the reference route r, and to only determine the choice probabilities of MΔ,r. This means that $P^{\text{ref}}(r) = 1$ for exactly one $r \in \mathcal{R}$. This might not be the realistic for standard equilibrium models since this creates asymmetry. However, such an approach would be very feasible for en-route decisions, where the current route serves as reference point. Also in day-to-day models, the previously chosen trip can be a natural reference route.

This new family of MΔ-models does not account automatically for all overlap with the multinomial generating function. Any network with overlap in which all routes have the same length, will give equal choice probabilities; the network structure does not have any influence on the choice probabilities. Also, the conditional choice situation with reference route r does account for overlap between r and all other alternatives, but cannot capture dependencies between all these other alternatives. So, a specific generating function that can handle these dependencies has to be specified per reference route.

4.5.1. Model instances

Using the four available generating functions we can derive several models based on multiplicative utility formulas based on reference routes. Here we only present the choice probabilities of route r conditional on r being the reference route, i.e., using $p = r$ in Equation (4.114)[10]. The derived choice probabilities seem to be rather complex; however, that is only due to the asymmetry of the generating vector. Basically, the choice probabilities of a MΔ,r-model are not more complex than those for a A or M model, and they can be derived easily for applications.

Multiplicative Multinomial with Reference Route (MΔ-MN): The disadvantage of independence in basic A-MN is inherited. Not all overlap is explicitly removed in the MΔ case, dependencies between non-reference routes maintain. The choice probabilities are

$$P_r^{\text{M}\Delta\text{-MN},r}(G^{\text{MN}}; \boldsymbol{y}^{\text{M}\Delta,r}) = \frac{1}{1 + \sum_{p \in \mathcal{R} \setminus \{r\}} \left(\frac{\sum_{l \in \mathcal{L}_{r \setminus p}} V_l}{\sum_{l \in \mathcal{L}_{p \setminus r}} V_l} \right)^{\mu}}, \quad \forall r \in \mathcal{R}. \tag{4.116}$$

Multiplicative Path-Size with Reference Route (MΔ-PS): The path-size choice probabilities are

$$P_r^{\text{M}\Delta\text{-PS},r}(G^{\text{PS}}; \boldsymbol{y}^{\text{M}\Delta,r}) = \frac{\text{PS}_r^{\beta}}{\text{PS}_r^{\beta} + \sum_{p \in \mathcal{R} \setminus \{r\}} \text{PS}_p^{\beta} \left(\frac{\sum_{l \in \mathcal{L}_{r \setminus p}} V_l}{\sum_{l \in \mathcal{L}_{p \setminus r}} V_l} \right)^{\mu}}, \quad \forall r \in \mathcal{R}. \tag{4.117}$$

As it is possible to convert the path-size factor back into the utility formulation for the A-PS and M-PS models (Kitthamkesorn and Chen, 2013), this is also possible for the MΔ-PS model. This will lead to:

$$U_p^{\text{M}\Delta\text{-PS},r} := \begin{cases} \frac{\sum_{l \in \mathcal{L}_r} V_l \varepsilon_r^{\text{M}}}{\text{PS}_r^{\beta}} & \text{if } p = r \\ \frac{\sum_{l \in \mathcal{L}_p \setminus \mathcal{L}_r} V_l \varepsilon_p^{\text{M}}}{\text{PS}_p^{\beta}} + \frac{\sum_{l \in \mathcal{L}_r \cap \mathcal{L}_p} V_l \varepsilon_r^{\text{M}}}{\text{PS}_r^{\beta}} & \text{otherwise.} \end{cases} \tag{4.118}$$

Having a reference route allows one to revisit the used path-size factors. These can be obtained by excluding all links of the reference route. This will lead to a formulation with reference route specific path-size parameters. Similar to equation (4.45), the formulation for the reference route r specific path-size factor $\text{PS}_{p,r}$ for route p is

$$\text{PS}_{p,r} = \begin{cases} \frac{\sum_{l \in \mathcal{L}_{p \setminus r}} \frac{V_l}{\#_l}}{\sum_{l \in \mathcal{L}_{p \setminus r}} V_l}, & \text{if } p \neq r \\ 1 & \text{if } p = r. \end{cases} \tag{4.119}$$

Multiplicative Paired Combinatorial with Reference Route (MΔ-PC): The paired combinatorial choice probabilities for in the MΔ case are

[10] Under the Markov chain assumption, these are the probabilities for staying in the same state.

$$P_r^{\text{M}\Delta\text{-PC},r}(G^{\text{PC}};\boldsymbol{y}^{\text{M}\Delta,r}) = \frac{\sum_{p\in\mathcal{R}\setminus\{r\}}\left(1+\left(\frac{\sum_{l\in\mathcal{L}_{r\setminus p}}V_l}{\sum_{l\in\mathcal{L}_{p\setminus r}}V_l}\right)^{\frac{\mu}{1-\varphi_{rp}}}\right)^{-\varphi_{rp}}}{\sum_{r'\in\mathcal{R}\setminus\{r\}}\left(1+\left(\frac{\sum_{l\in\mathcal{L}_{r\setminus r'}}V_l}{\sum_{l\in\mathcal{L}_{r'\setminus r}}V_l}\right)^{\frac{\mu}{1-\varphi_{rr'}}}\right)^{1-\varphi_{rr'}} + \sum_{p\in\mathcal{R}\setminus\{r,r'\}}\left(\left(\frac{\sum_{l\in\mathcal{L}_{r\setminus r'}}V_l}{\sum_{l\in\mathcal{L}_{r'\setminus r}}V_l}\right)^{\frac{\mu}{1-\varphi_{r'p}}}+\left(\frac{\sum_{l\in\mathcal{L}_{r\setminus p}}V_l}{\sum_{l\in\mathcal{L}_{p\setminus r}}V_l}\right)^{\frac{\mu}{1-\varphi_{r'p}}}\right)^{1-\varphi_{r'p}}}, \tag{4.120}$$

$\forall r\in\mathcal{R}$. This formula looks complicated, as the cumbersome indexing is required because the non-symmetric generating vector $\boldsymbol{y}^{\text{M}\Delta,r}$ is substituted, and then some terms collapse to 1. Without this substitution the probabilities are

$$P_r^{\text{M}\Delta\text{-PC},r}(G^{\text{PC}};\boldsymbol{y}^{\text{M}\Delta,r}) = \frac{\sum_{p\in\mathcal{R}\setminus\{r\}}\left({y_{rr}^{\text{M}\Delta}}^{\frac{\mu}{1-\varphi_{rp}}}+{y_{rp}^{\text{M}\Delta}}^{\frac{\mu}{1-\varphi_{rp}}}\right)^{-\varphi_{rp}}}{\sum_{r'\in\mathcal{R}}\sum_{p\in\mathcal{R}\setminus\{r'\}}\left({y_{rr'}^{\text{M}\Delta}}^{\frac{\mu}{1-\varphi_{r'p}}}+{y_{rp}^{\text{M}\Delta}}^{\frac{\mu}{1-\varphi_{r'p}}}\right)^{1-\varphi_{r'p}}},\quad \forall r\in\mathcal{R}. \tag{4.121}$$

Multiplicative Link-Nested with Reference Route (MΔ-LN): The link-nested choice probabilities in the MΔ case are

$$P_r^{\text{M}\Delta\text{-LN},r}(G^{\text{LN}};\boldsymbol{y}^{\text{M}\Delta,r}) = \frac{\sum_{m\in\mathcal{L}}\alpha_{mr}\left(\alpha_{mr}+\sum_{p\in\mathcal{R}\setminus\{r\}}\alpha_{mp}\left(\frac{\sum_{l\in\mathcal{L}_{r\setminus p}}V_l}{\sum_{l\in\mathcal{L}_{p\setminus r}}V_l}\right)^{\mu_m}\right)^{\frac{\mu}{\mu_m}-1}}{\sum_{m\in\mathcal{L}}\left(\alpha_{mr}+\sum_{p\in\mathcal{R}\setminus\{r\}}\alpha_{mp}\left(\frac{\sum_{l\in\mathcal{L}_{r\setminus p}}V_l}{\sum_{l\in\mathcal{L}_{p\setminus r}}V_l}\right)^{\mu_m}\right)^{\frac{\mu}{\mu_m}}},\quad \forall r\in\mathcal{R}. \tag{4.122}$$

4.5.2. Model properties

Since for the MΔ case utilities are sums of two variates of the MEV distributions, it is not possible to obtain a convenient closed-form for the joint density function of the utilities. Therefore, we cannot derive the expected maximum utility. However, from the marginal distributions of $\boldsymbol{\varepsilon}^{\text{M}}$ it is possible to obtain the expected value and variance of the MΔ utilities conditional on a reference route, which are:

$$\mathbb{E}(U_p^{\text{M}\Delta,r}) = \begin{cases}\frac{\sum_{l\in\mathcal{L}_r}V_l}{G(\mathbf{1}_r)^{1/\mu}}\Gamma\left(1+\frac{1}{\mu}\right) & \text{if } p=r\\ \left(\frac{\sum_{l\in\mathcal{L}_p\setminus\mathcal{L}_r}V_l}{G(\mathbf{1}_p)^{1/\mu}}+\frac{\sum_{l\in\mathcal{L}_r\cap\mathcal{L}_p}V_l}{G(\mathbf{1}_r)^{1/\mu}}\right)\Gamma\left(1+\frac{1}{\mu}\right) & \text{otherwise}\end{cases},\ \text{and,} \tag{4.123}$$

$$\text{Var}(U_p^{\text{M}\Delta,r}) = \begin{cases}\frac{\left(\sum_{l\in\mathcal{L}_r}V_l\right)^2}{G(\mathbf{1}_r)^{2/\mu}}\left(\Gamma\left(1+\frac{2}{\mu}\right)-\Gamma\left(1+\frac{1}{\mu}\right)^2\right) & \text{if } p=r\\ \left(\frac{\left(\sum_{l\in\mathcal{L}_p\setminus\mathcal{L}_r}V_l\right)^2}{G(\mathbf{1}_p)^{2/\mu}}+\frac{\left(\sum_{l\in\mathcal{L}_r\cap\mathcal{L}_p}V_l\right)^2}{G(\mathbf{1}_r)^{2/\mu}}\right)\left(\Gamma\left(1+\frac{2}{\mu}\right)-\Gamma\left(1+\frac{1}{\mu}\right)^2\right) & \text{otherwise.}\end{cases} \tag{4.124}$$

Table 4.2.: Choice probabilities for the MΔ-MN model on the simple network. The probabilities conditional on reference routes, as well as the two solution methods are provided.

	Upper	Middle	Lower
Ref. route = Upper	$\frac{2}{5}$	$\frac{1}{5}$	$\frac{2}{5}$
Ref. route = Middle	$\frac{8}{17}$	$\frac{4}{17}$	$\frac{5}{17}$
Ref. route = Lower	$\frac{5}{14}$	$\frac{4}{14}$	$\frac{5}{14}$
Equal $P^{\text{ref}}(\cdot)$ solution	≈ 0.409	≈ 0.240	≈ 0.350
Markov chain solution	≈ 0.401	≈ 0.239	≈ 0.359

Furthermore, when the reference route is fixed (i.e., $P^{\text{ref}}(r) = 1$ for exactly one $r \in \mathcal{R}$), then SUE formulation (4.87) also holds for the MΔ,r-case.

4.5.3. Simple network

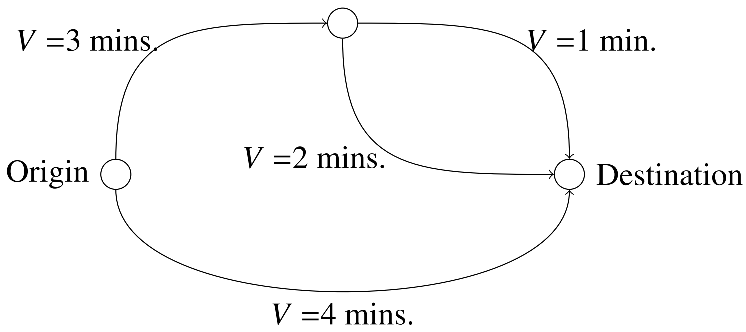


Figure 4.3.: Simple three route overlap network.

To provide more insight in the working of MΔ-models, the MΔ-MN probabilities are provided for a simple network. Figure 4.3 shows an origin and destination with three routes in between. Systematic utility is assumed to equal foreseen travel time, and no route specific constant is included. The upper and middle routes have overlap, and the upper route is faster. The lower route has no overlap with the other two, and is equally fast as the upper route. Table 4.2 shows the choice probabilities for each reference route based on $\mu = 1$, as well as, the solutions based on equal reference route probabilities and the Markov chain approach.

As expected, the slowest middle route has lowest choice probability. This choice probability is significantly smaller than it would be under A- and M-models, which is realistic since changing from the upper to the middle route means that the non-overlapping travel time doubles. Also, the two different solution methods for dealing with multiple reference routes do not differ that much.

On the other hand, there is a higher preference for the upper route than for the lower route while they have equal travel times. Based on the theory in Section 4.2, one would argue that the opposite should be true. Having the upper or lower route as the reference route, causes no difference in choice probabilities between the upper and lower route. Only because the middle route exists, and since it is more 'profitable' (i.e. the factor 2 between the non-overlapping parts between middle and lower is higher than the factor 4/3 between middle and lower route) to switch to the upper route, there is a slight preference for the upper route. This means that by having routes included as reference routes, the final probabilities change (similar to what Bliemer and Bovy (2008) show).

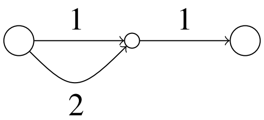
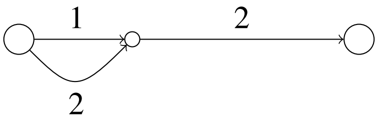
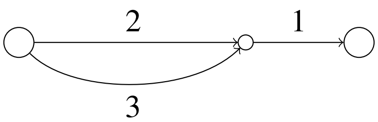
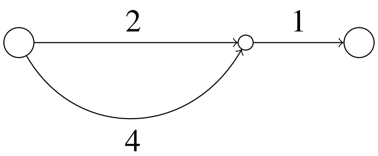
So, this simple example provides insight in the working of the MΔ-models, but does not show its full potential. Clearly, the MΔ-models have different behaviour than all other known route choice models. The next two sections provide more insight in the differences between the A-, M- and MΔ-models – and their suitability for traffic assignment applications –. First, we analyse their basic behaviour under simple changes in the network configuration, and second, we analyse all models under more complex network variations. The latter quantitative test shows which model can best approximate the generic utility formulation of Section 4.2. These Sections provide the full potential of the MΔ-models.

4.6. Basic model behaviour under simple network changes

In this section we discuss the change in choice behaviour under three simple network adjustments. Consider the four networks depicted in the first column of Table 4.3. Network A is the basis and has 2 routes with different lengths of which the final parts overlap. In network B the overlapping part is extended, while in networks C1 and C2 the non-overlapping parts are extended by respectively adding a constant length and by multiplying their lengths. The models are so simple that, regardless the *exact* choice behaviour, the trend in choice probabilities for the two routes is known. The trend can either be that the probabilities remain equal, converge, or diverge. It is desired that choice models can reproduce the expected trend for each network change. However, not all trends can be reproduced by all choice models. Therefore, we analyse the behaviour of the three basic (-MN) models. We take advantage of having only two routes for the MΔ-MN model here, which avoids dealing with reference routes. However, similar results are obtained when extending to more than two routes, as is analysed in the next section.

Consider the switch from network A to network B. Since the non-overlapping parts remain equal, the choice probabilities for both routes obviously also remain equal. In the A-MN model only the difference between the routes matters, and since this difference does not change, the probabilities also remain equal for that model. For the M-MN model however, only the ratio between the routes matters, and this ratio changes. Because the non-overlapping part increases, the (lower:upper)-ratio decreases, and the choice probabilities will converge. The expected behaviour can not be reproduced by any M-MN model instance. The MΔ-MN model in the end, merely considers the ratio of the difference between routes, i.e., the ratio of the non-

Table 4.3.: **Behaviour under changing networks.** Networks B, C1 and C2 are slight variations on network A. Each network has two routes; the table shows the trend of the route probabilities when one switches for network A to any of the other three networks. They can either converge, diverge, or remain equal (depicted with arrows). The expected trend and achievable trends for three models are provided (see main text).

		Expected choice behaviour	A-MN		M-MN		MΔ-MN	
A								
B		=	=	✓	↓↑	✗	=	✓
C1		↓↑	=	✗	↓↑	✓	↓↑	✓
C2		↑↓	↑↓	✓	↑↓	✓	↑↓	✓

overlapping parts. Since the non-overlapping part does not change, the choice probabilities do not change. This holds for all MΔ-MN model instances

Consider the switch from network A to network C1. When a constant length is added to both routes, they become more similar the choice probability of the shortest route increases. So, it is expected that the route probabilities will converge[11]. The difference between the two routes will not change, therefore no A-MN model instance can reproduce the expected behaviour. On the other hand, the M-MN and MΔ-MN models will reproduce the expected behaviour since the (lower:upper)-ratio between routes (in- and excluding the overlapping part) decreases. The probabilities converge for all M-MN and MΔ-MN model instances.

Consider the switch from network A to network C2. When a both routes are multiplied by the same factor, the absolute detour of the longest route will increase, and will therefore be chosen less. So, it is expected that route probabilities will diverge[12]. The difference between the two routes will increase, therefore the probabilities in the A-MN models will diverge, which coincides with the expected behaviour. The ratio of the non-overlapping parts of the routes remains equal, but due to the overlapping part, the ratio between the whole routes changes. This increase in the (lower:upper)-ratio will lead to the desired diverging probabilities in the M-MN model. Of course, the divergence 'speed' depends on the length of the overlapping part. However, this can be adjusted by using the constant in the systematic utility specification. This constant is also the reason that the diverging probabilities are obtainable in the MΔ-MN model; here it also completely determines the divergence 'speed', which is an advantage compared to M-MN. All M-MN model instances diverge for this network example, but the behaviour remains dependent on the overlapping length. Almost all MΔ-MN model instances will have the expected behaviour, except for those with a constant equal to zero.

This analysis, summarized in Table 4.3, on the most simple and basic network variation shows that only the MΔ-MN model can reproduce all expected behaviours. We believe no simple network change exists of which the expected behaviour can be captured by A-MN or M-MN, but not by MΔ-MN. Since real networks do not consist of merely simple network changes, the next section analyses the competitiveness of all models on more comprehensive networks based on the route utility formulation of section 4.2, including the PS, PC, and LN variants, and including more than two routes which requires the use of reference routes for the MΔ-models.

4.7. Network Example

The route choice models are applied on a network with three routes to show the advantage of the MΔ-models compared to the others. Bliemer and Bovy (2008) compare route choice models on different route sets. They demonstrate that existing closed-form route choice models (except for A-MN) are sensitive to irrelevant route alternatives. Hence, if the route sets in application are different from the route sets in estimation, the results may be poor. This network

[11] As an easy example, consider routes with length x and $x+1$; the choice probabilities will become 50 %-50% if $x \to \infty$ and 100%-0% if $x \to 0$

[12] As an easy example, consider routes with length x and $2 \times x$; the choice probabilities will become 50 %-50% if $x \to 0$ and 100%-0% if $x \to \infty$

example considers very different choice situations and the models are estimated and validated on different subsets of the data.

In Figure 4.4 the network with link travel times and routes is depicted. This network has been carefully constructed to put the choice models under stress, namely with and without overlap, and with short and long distance O-D pairs. Also, the relative influence of the analyst and random foreseen travel time changes. The travel times of links 1, 2, 4 and 5 depend on variable x and by varying x different choice situations are created. Each value of x can represent a different O-D pair in a transport network. The graph shows how route travel times increase with x; for $x = 0$ the difference (respectively ratio) between the slowest and fastest routes is four minutes (respectively 0.67), and this increases to eight minutes (respectively 0.90) for $x = 40$. Furthermore, overlap occurs on links 1 and 5.

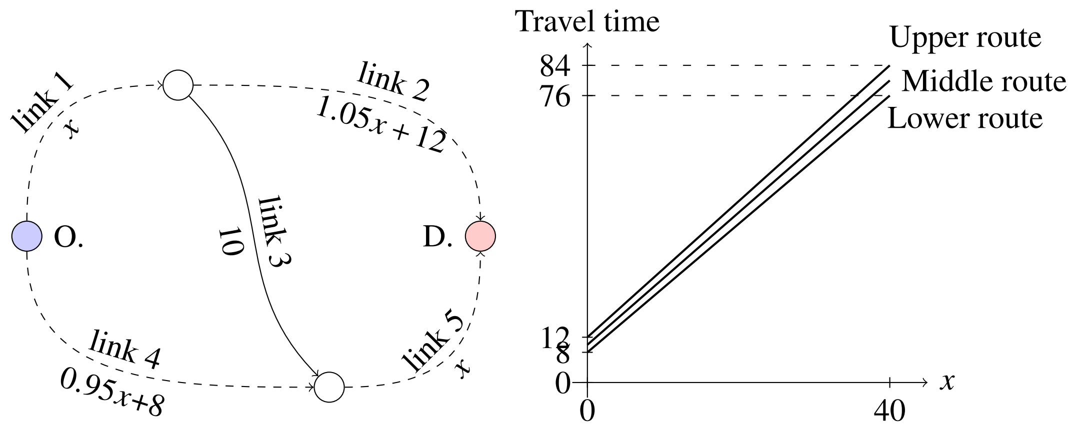


Figure 4.4.: Network and travel times. Dashed links have variable costs.

For the experiment, the probabilities from the MNP model – that can properly handle both the analyst error and random foreseen travel time, and the desired variance-covariance structure based on the scaling postulate – are the ground truth. Thus, the route utilities are jointly distributed following a multivariate normal distribution. This distribution is specified in line with the random route utility of Section 4.2. Assume that there is no systematic utility other than travel time, then identification leads to $V_r^0 = 0$ for the A-models, and $V_r^0 = c$ for the M- and MΔ-models (see Section 4.4.3). For the normalization set the expected analyst error $\mathbb{E}(\varepsilon_r)$ to 0 and the travel time parameter β to -1. For the proportionality parameter (regarding the standard deviation of the foreseen travel time), linear regression on the OViN data of Figure 4.1 leads to $\theta = 0.3859$, and linear regression on the route survey leads to $\theta = 0.1301$; however, the first value is too high (see Section 4.2.1) and the latter too low (since it is the response of only one traveller), thus we assume $\theta = 0.2$. The covariances between the random foreseen travel times are assumed to be based on the arithmetic mean, see Equation (4.9). Note that these definitions of the (co)variances are different from those in the literature that assume proportionality between variance and mean. The standard deviation of the analyst error is set to

$\sigma(\varepsilon_r) = 10$ (minutes). This leads to the following multivariate normal utility distribution:

$$\mathcal{N}\left(\underbrace{-\begin{pmatrix} 2.05x+12 \\ 2x+10 \\ 1.95x+8 \end{pmatrix}}_{\mathbb{E}(\boldsymbol{U})}, \underbrace{0.2^2\begin{pmatrix} (2.05x+12)^2 & 2.025x^2+11x & 0 \\ 2.025x^2+11x & (2x+10)^2 & 1.975x^2+9x \\ 0 & 1.975x^2+9x & (1.95x+8)^2 \end{pmatrix}}_{\text{covariance matrix of } \boldsymbol{\tau}} + \underbrace{\text{diag}\begin{pmatrix} 100 \\ 100 \\ 100 \end{pmatrix}}_{\text{variance matrix of } \boldsymbol{\varepsilon}}\right). \tag{4.125}$$

While x increases, not only the distance between origin and destination increases, but also the influence of the randomness from foreseen travel time compared to the randomness from the analyst error increases. For $x = 0$ only 3.8 percent of the variance of utility is due to the foreseen travel time; however, for $x = 40$, 76.4 percent of the variance is due to the foreseen travel time. Furthermore, the part of utility with overlap increases with x. In the development of this network example, the link travel times were chosen given the assumptions on the errors, and such that the route choice probabilities are more or less stable.

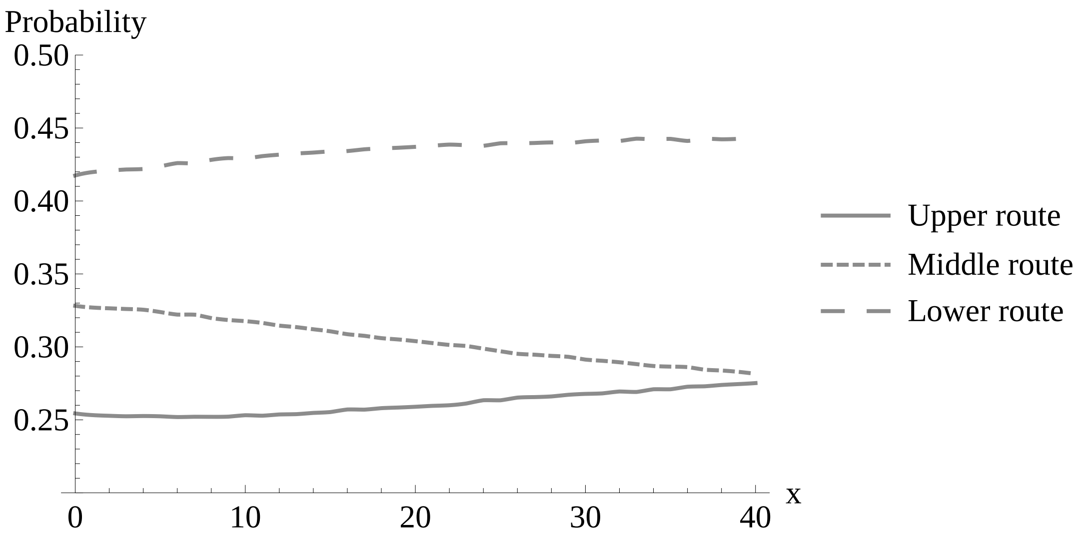


Figure 4.5.: Ground truth probabilities based on a million multivariate normal samples for each $x \in \{0,\dots,40\}$.

For all $x \in \{0,1,\dots,40\}$ the choice situation is simulated a million times by sampling from the multivariate normal distribution (Equation (4.125)), see Figure 4.5. The models are estimated twice; once on the dataset $x = \{5,\dots,15\}$, and once on dataset $x = \{25,\dots,35\}$ using log-likelihood maximization. So, eleven million 'observations' are used to estimate each model. For the MΔ-PS the path-size formulation of equation (4.45) is chosen[13]. For the LN models link specific scales are estimated for links 1 and 5; the other links – read nests – contain only one route and thus 'collapse': $(y^{\mu_l})^{\mu/\mu_l} = y^{\mu}$. The Markov chain approach for reference routes is chosen for MΔ-MN and MΔ-PS, and the 'equal probability reference route' approach

[13]This is chosen since equation (4.119) leads to much more path-size factors

Table 4.4.: Parameter estimates for every model for the two datasets

Model	Parameter estimates					
	Dataset $x = \{5,\ldots,15\}$			Dataset $x = \{25,\ldots,35\}$		
	Scale $\mu \approx$	Constant $\approx$	Other	Scale $\mu \approx$	Constant $\approx$	Other
A-MN	0.107			0.0699		
A-PS	0.107		$\beta \approx 0.182$	0.0681		$\beta \approx 0.501$
A-PC	0.0935			0.0576		
A-LN	0.0438		$\mu_1 \approx 0.095$ $\mu_5 \approx 0.818$	0.0225		$\mu_1 \approx 0.0225$ $\mu_5 \approx 1.091$
M-MN	11.518	-77.635		5.593	=0	
M-PS	12.932	-91.038	$\beta \approx 0.173$	8.690	-47.403	$\beta \approx 0.490$
M-PC	12.983	-108.293		7.809	-55.479	
M-LN	0.490	-142.935	$\mu_1 \approx 73.916$ $\mu_5 \approx 1175.47$	0.319	-0.258	$\mu_1 \approx 18.235$ $\mu_5 \approx 118.725$
MΔ-MN	7.475	-43.589		4.619	=0	
MΔ-PS	10.613	-72.889	$\beta \approx 0.145$	7.381	-41.176	$\beta \approx 0.432$
MΔ-PC	14.328	-126.223		7.396	-62.199	
MΔ-LN	0.475	-111.01	$\mu_1 \approx 58.367$ $\mu_5 \approx 7518.45$	0.569	=0	$\mu_1 \approx 15.265$ $\mu_5 \approx 61.098$

is chosen for MΔ-PC and MΔ-LN[14]. For the validation, the models' log-likelihood on the other dataset is determined, so the parameters from the estimation on $x = \{5,\ldots,15\}$ are applied to $x = \{25,\ldots,35\}$, and vice versa.

Table 4.4 shows the estimated parameters.[15] The standard errors are very low due to the large artificial dataset and therefore not reported. Figure 4.6 shows the log-likelihoods of the estimation and validation results. Probabilities of the models are found in Figures 4.7 and 4.8. In every graph two instances of one model and the ground truth are shown, the blue lines

[14] For these two models the software (Wolfram Mathematica) couldn't find analytical soltions of the choice probabilities using the Markov chain with unknown x. General applications do not have this parametrization on x, and thus shouldn't be problematic.

[15] The q-generalized logit model by Nakayama (2013) that captures both A-MN and M-MN has also been estimated; the results are not presented since the resulting model was always equivalent to the M-MN model (i.e., not to A-MN)

depict the route probabilities for the model estimated on $x = \{5,\dots,15\}$, the red lines depict the route probabilities for the model estimated on $x = \{25,\dots,35\}$, and the grey lines depict the MNP (i.e., ground truth) probabilities.

The estimation results show that the MN models have the poorest log-likelihoods of all on the $\{25,\dots,35\}$-data; these models cannot address the overlap of routes properly. The validation results of the MN models are also worse than the other types, so they cannot be transferred between short distance and long distance O-D pairs. Another remark is that the PC models, which are designed to capture overlap, perform poorest on the $\{5,\dots,15\}$-data; this is due the fact that the covariance between routes is low (because the analyst error is dominating), while the PC model always imposes dependencies.

In all cases, the A-models have the worst estimation result compared to the M and MΔ-models. Remarkably, the multiplicative models also outperform the additive models on the $\{5,\dots,15\}$-data, where the influence of the foreseen travel time is relatively small. The MΔ-models have better results on all four MN and PS estimations, while the M-models have better results on three out of four PC and LN estimations. This is because the additional parameter, the constant c, for the multiplicative models, which leads to a better fit.

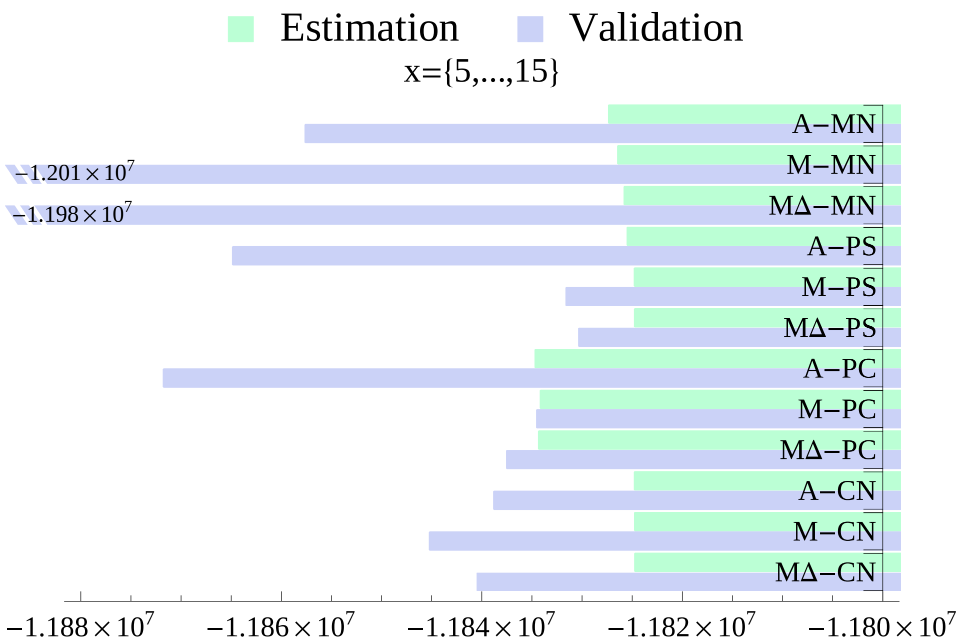
The multiplicative errors capture the analyst error with constant c in the systematic utility. This constant is indeed larger for the models estimated on the $\{5,\dots,15\}$-data, where the analyst error is dominant. Furthermore, note that all scale parameters μ are smaller for the $\{25,\dots,35\}$-data (i.e., the variance is higher); this reflects the heteroscedasticity of route utility.

We would like to mention that the PC and LN estimation were problematic on other network configurations that we have tried. For the multiplicative models extremely high scales occurred, which can lead to numerical problems. They also generated very unrealistic probabilities outside the area they were estimated on. As mentioned earlier, it is infeasible to estimate all link-specific nest-scales in large networks for LN models.

4.8. Conclusions and Discussion

This paper presented twelve route choice models – of which seven are new – in a single framework, and assessed them qualitatively and quantitatively. Choice probabilities for all models have the same closed form expression, namely Equation (4.26), based on a generating function and a generating vector. The generating function determines how route overlap is captured, which is either multinomial (MN), path-size (PS), paired combinatorial (PC), or link-nested (LN). The generating vector determines the utility formula, which is either additive RUM (A), multiplicative RUM (M), or the newly presented multiplicative RUM based on reference routes that only considers differences between routes (MΔ).

For the qualitative assessment a basic structure of utility with random foreseen travel time was presented (see Section 4.2). We base our analysis on two postulates on random travel times of road segments that each lead to a different structure of randomness. Empirical evidence provides the new insight that the foreseen travel time distribution's mean and standard deviation have a linear relationship, contrary to a linear relationship between its mean and variance. The homoscedastic additive models are not able to capture the random foreseen travel



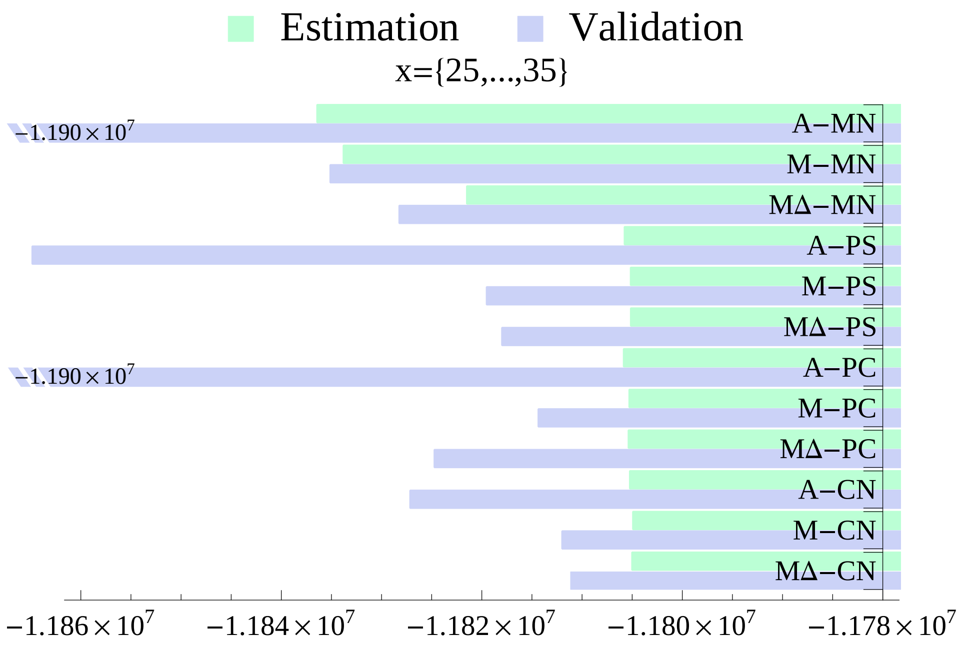


Figure 4.6.: Log-likelihoods of model instances for estimation and validation. Parameter estimates of the left estimation are used for the validation in the right, and vice versa.

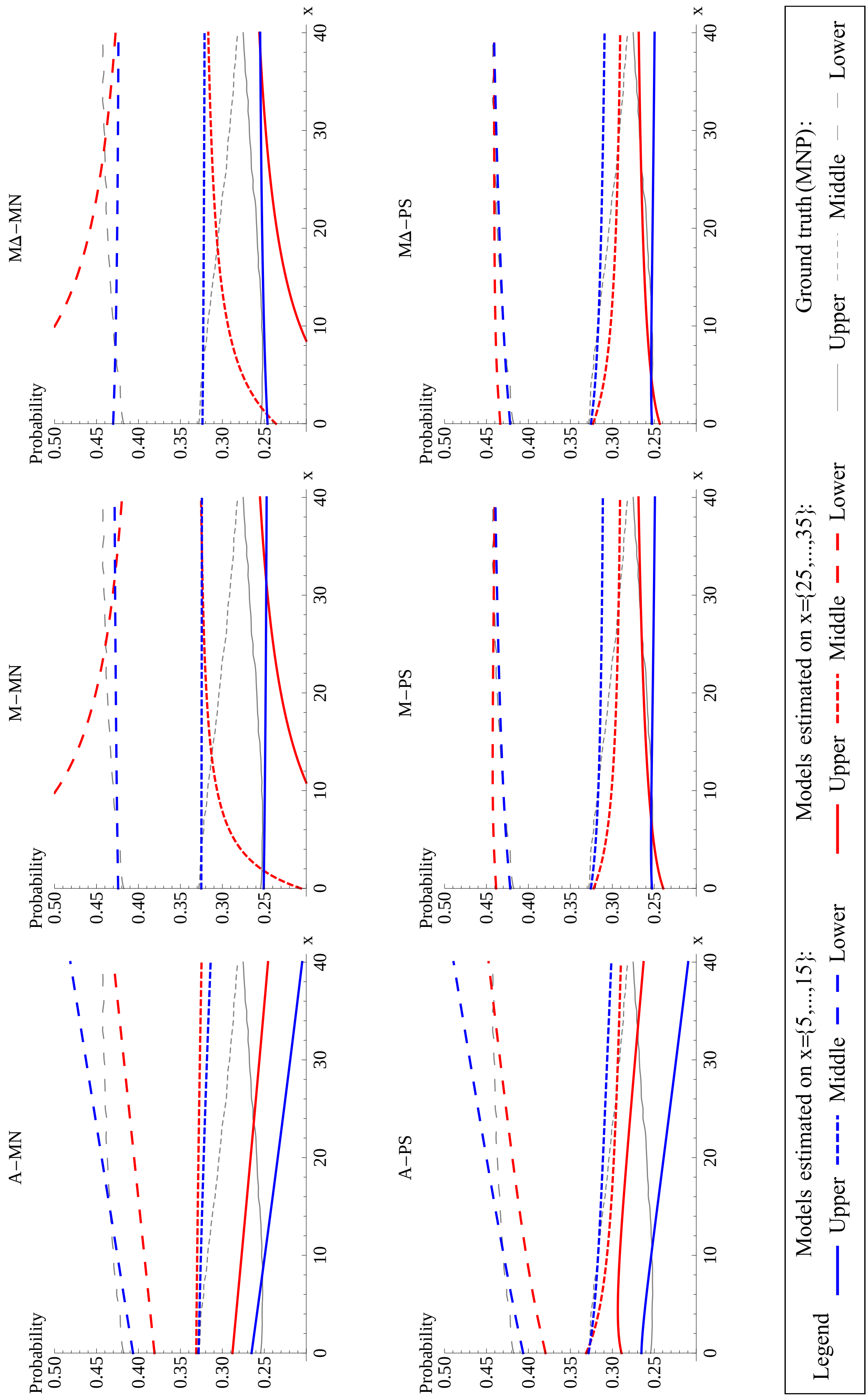


Figure 4.7.: Choice probabilities of the MN- and PS -models.

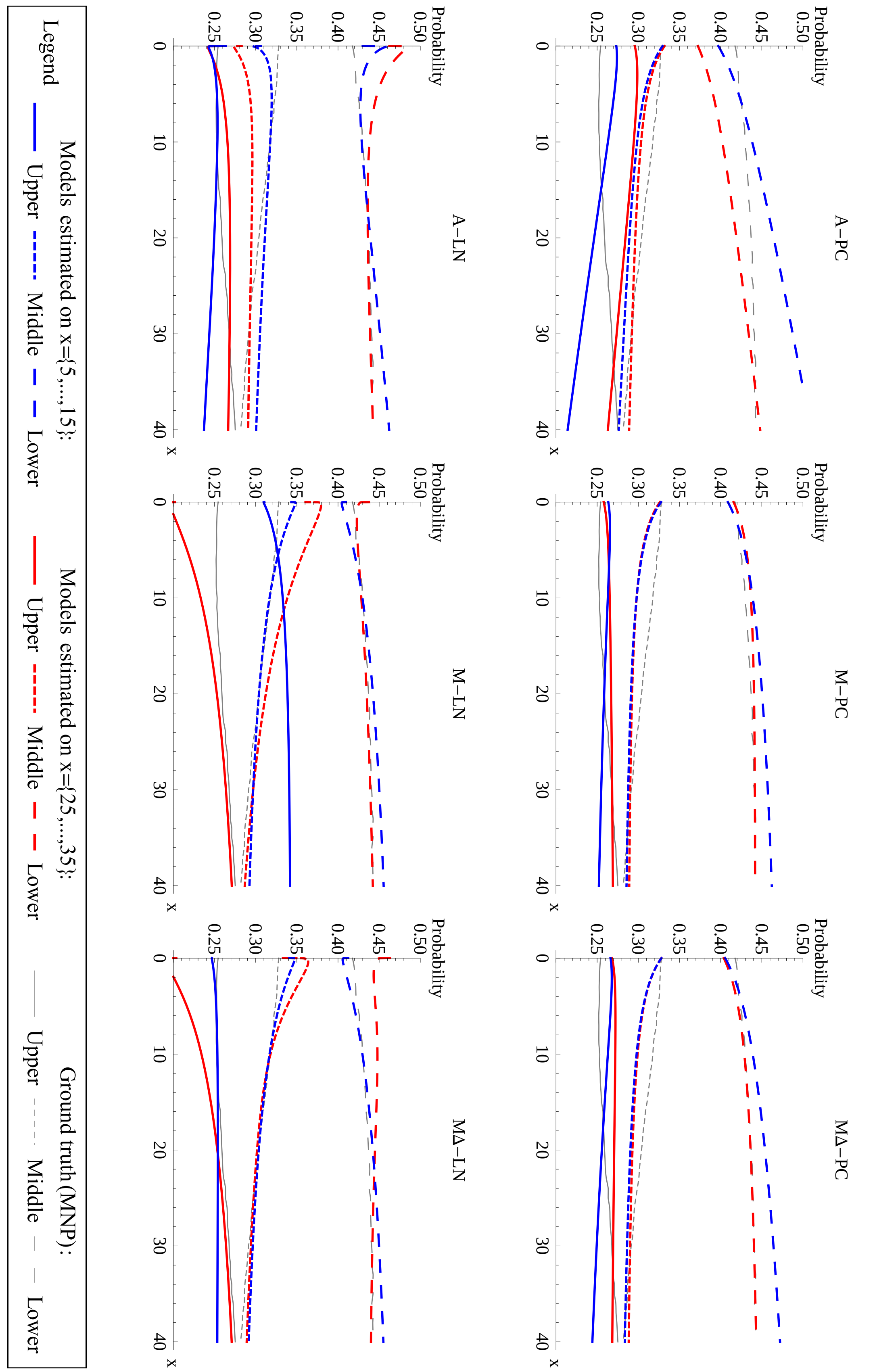


Figure 4.8.: Choice probabilities of the PC- and LN-models.

time, but multiplicative models do allow for this. Furthermore, differences in normalization, identification, and invariance are pointed out. The constant in systematic utility in multiplicative models does not have to be normalized. This allows more degrees of freedom and a better fit on the data, but makes it more difficult to compare models from the different paradigms directly. One main advantage of the generic GMEV framework is that it can be analysed as a whole, we show this by providing the equivalent stochastic user equilibrium formulation for all models.

To show the distinctiveness of the MΔ-models, each model's behaviour under basic network changes was analysed. Only the MΔ-models can reproduce realistic behaviour when the characteristics of parallel and serial links change.

To test the models' potential on real networks, and to test whether they can be applied on datasets on which they are not estimated, a carefully constructed network example was presented. Based on our analyses, we expect good performance of the M-PS and MΔ-PS models for route choice on real networks. They can capture overlap sufficiently, and they can handle random foreseen travel time. Also Fosgerau and Bierlaire (2009) and Chikaraishi and Nakayama (2015) have both compared additive with multiplicative formulations on multiple datasets, and they found that the multiplicative models have a better fit for all datasets.[16] However, additional empirical estimation and validation is required to conclusively assess all models. Finally, the PC and LN models are problematic to estimate on some other networks we tried, and they can cause numerical problems.

This chapter does not discuss the route generation or sampling related to the explicit route sets of the models. As pointed out earlier, a correct sample of routes is required to obtain unbiased parameter estimates (Frejinger et al., 2009). For econometrically sound applications, sampling techniques as presented by (Frejinger et al., 2009; Flötteröd and Bierlaire, 2013; Guevara and Ben-Akiva, 2013) have to be adapted for the new models. Models with implicit route sets (Dial, 1971; Fosgerau et al., 2013a; Papola and Marzano, 2013; Mai et al., 2015) do not have this issue, but they might lead to unrealistic routes. The connection between the link-based MEV model of (Papola and Marzano, 2013) with the route-based GMEV framework does not seem feasible due to the different base units, but if it exists, it might lead to new GMEV model instances. On the other hand, Prato (2012) points out conceptual and empirical reasons that plea for the explicit approach.

Acknowledgements

We would like to thank Michel Bierlaire for the discussion with him on an early version of this paper, Kees van Goeverden for preparing the OViN data, and Giselle de Moraes Ramos-Heydendael for providing the foreseen travel time data. The study is performed as part of the Innovative Pricing for Sustainable Mobility (iPriSM) project in the Sustainable Accessibility of the Randstad (SAR) program of the Dutch National Science Foundation (NWO). The main

[16] Note that Chikaraishi and Nakayama (2015) state their model equals logit when $q = 0$ and to weibit when $q = 1$; however, they do not fully use the flexibility of weibit. When they would have estimated a constant for the weibit models (similar to the approach in this paper), the q-generalized logit models with $0 < q < 1$ will also become weibit.

theory in this chapter is initiated during a visit of the corresponding author to the Institute of Transport and Logistics Studies at the University of Sydney.

CHAPTER 5.

A FAMILY OF MACROSCOPIC NODE MODELS

The family of macroscopic node models which comply to a set of basic requirements is presented and analysed. Such models are required in macro-, mesoscopic traffic flow models, including dynamic network loading models for dynamic traffic assignment. Based on the behaviour of drivers approaching and passing through intersections, the model family is presented. The headway and the turn delay of vehicles are key variables. Having demand and supply as input creates a natural connection to macroscopic link models. Properties like the invariance principle and the conservation of turning fractions are satisfied. The inherent non-uniqueness is analysed by providing the complete set of feasible solutions. The node models proposed by Tampère et al. (2011); Flötteröd and Rohde (2011); Gibb (2011) are members of the family. Furthermore, two new models are added to the family. Solution methods for all family members are presented, as well as a qualitative and quantitative comparison. Finally, an outlook for the future development of empirically verified models is given.

5.1. INTRODUCTION AND BACKGROUND TO MACROSCOPIC NODE MODELS

This paper is a slightly adapted version of:

- Smits, E.-S., Bliemer, M. C., Pel, A. J., and van Arem, B. (2015).
  A family of macroscopic node models.
  Transportation Research Part B: Methodological, 74:20–39

A core component of every dynamic transportation model is to compute the time-varying traffic conditions (described by, e.g., flows, densities, headways, speeds, travel times, etc.) on a network once the dynamic travel demand from origins to destinations is given. This traffic simulation procedure is often referred to as DNL. Hence, the main purpose of DNL models is to determine the emerging traffic conditions as a result of the interaction between infrastructure supply and travel demand. To this end, DNL models typically consist of a link model and a node model.

The link model computes the dynamic traffic flow propagation along homogeneous road stretches, while the node model computes the traffic conditions at discontinuities in the network, such as bottlenecks and intersections. For the link model, fundamental diagrams are used to describe the traffic dynamics and underlying driving behaviour. Depending on the level of aggregation in the representation of traffic, models can be categorised as microscopic, mesoscopic, or macroscopic, and similarly the fundamental diagrams describe relationships for pace-headway, spacing-speed, and density-flow (Laval and Leclercq, 2013).

Unfortunately, for nodes no general driving behaviour representative (as the fundamental diagram is for links) is known. Nevertheless, this study shows that cumulative flow curves on the incoming links, turns, and outgoing links of a node can be derived and provide a complete representation of the traffic dynamics. Furthermore, this paper shows how these cumulative flow curves can be used to describe driving behaviour at nodes in terms of time-headways. The benefit hereof is twofold. First, this newly introduced way to represent traffic at nodes allows to derive and analyse the full family of node models consistent with the requirements for the Generic Class of first-order Node Models (GCNM) as presented by Tampère et al. (2011). Second, this representation of traffic based on a time-headway relationship yields descriptive variables that can be interpreted at the level of individual driving behaviour. Hence, earlier developed models belonging to the GCNM can now be analysed according to their assumptions on the underlying driving behaviour at nodes. Furthermore, new models fulfilling the requirements for the GCNM can be derived based on explicit behavioural assumptions.

In the transition from static to dynamic transportation models, the field of traffic flow theory has studied the propagation of traffic dynamics along homogeneous road stretches (links) exhaustively. However, where the link model propagates these traffic conditions along the links, the node model determines most congestion seeds where queues originate (due to insufficient downstream capacity) as well as the direction (i.e. the upstream links) towards which these queues spill back. The validity of the node model is therefore particularly important for traffic assignment studies and traffic flow on dense (urban) networks. Although node models have been studied in several papers through the last decades, they received significantly less attention than link models. Early contributions on node models include (Daganzo, 1995a; Lebacque, 1996), and were in the next decade followed by (Jin and Zhang, 2003, 2004; Ni and Leonard II, 2005; Bliemer, 2007; Jin, 2010, 2012a).

Several studies have posed requirements for the validity of a node model. Lebacque and Khoshyaran (2005) identified two invariance principles to ensure consistency between traffic flows on links and nodes. The *first invariance principle* requires that when spillback occurs on an upstream link, the outflow of that link is invariant to an increase of the upstream demand. The *second invariance principle* requires that when the supply (or capacity) is not fully utilized at a downstream link, the inflow of that link is invariant to a decrease in the supply. Later, Tampère et al. (2011) constructed a complete set of requirements for node models including these invariance principles as well as demand and supply constraints, requirements for the conservation of turning fractions (due to first-in first-out), and individual flow maximization. Also general applicability, i.e. any number of in- and outlinks, is a requirement. If a node model satisfies these requirements it belongs to the earlier mentioned GCNM; call this set the *generic requirements.*

Tampère et al. (2011) argue that if any of the generic requirements is violated, the model is either not applicable on general networks, or does not comply with basic traffic characteristics. Therefore, the generic requirements are necessary to ensure the behavioural validity of a generally applicable node model with any number of in- and outlinks. At present only two node models exist that satisfy the generic requirements, namely the model presented by Tampère et al. (2011) and Flötteröd and Rohde (2011), and another model by Gibb (2011). The generic requirements are not sufficient to guarantee a unique solution. In order to create a specific node model, additional constraints need to be introduced. Those determine how the downstream supply is distributed. In the capacity proportional model of Tampère et al. (2011) and Flötteröd and Rohde (2011) the capacity of outgoing links is divided among competing flows proportional to the capacity[1] of the corresponding incoming links. This is achieved in (Tampère et al., 2011) by adding a capacity proportional supply constraint interaction rule (SCIR) to the total flow maximization problem. In (Flötteröd and Rohde, 2011) instead, the incremental transfer principle is used where the incremental node model solution is the stationary point of a dynamic system. Notwithstanding that these two models are derived from different principles, their solutions are identical. In the capacity consumption equivalence model by Gibb (2011) the underlying assumption is that traffic towards a saturated (downstream) link consumes more capacity of its (upstream) link. Although both node models satisfy all requirements, they lack behavioural foundation (the first model to a larger extent than the latter), complicating a assessment of their validity. Furthermore, non-iterative and non-repetitive[2] solution methods are lacking for both models which, especially in Gibb's case, lead to long calculation times. Finally, the formulations of these models are not compatible as they build on different (algorithm driven) variables, making it impossible to capture one model in the others' framework.

This paper (i) presents the representation of traffic flows at nodes according to the time-headway relationship and turn delays (a concept that will be introduced in more detail in this paper) as decision variables (ii) shows how all models belonging to the GCNM (i.e., satisfying the generic requirements) fit into this framework yielding a family of node models where the challenge for each model is to find a set of turn delays, (iii) presents two new node models, (iv) shows how the two existing node models by Tampère et al. (2011) and Flötteröd and Rohde (2011), and by Gibb (2011) are specific cases of the family, (v) presents the mathematical optimization problem for the node model family as multi-objective optimization problem and discusses Pareto optimality and the output relevant feasible solution set[3], (vi) provides solution methods for all models, and (vii) analyses the family of node models (including specific cases) on an illustrative three-leg node.

In the past other models are presented that do no satisfy all generic requirements. Corthout (2012, Chapter 3) provides an extensive literature overview of all models that do not satisfy all requirements; Table 3-1 shows which models satisfy which requirements. It should be noted that the problem was already solved for merges and diverges[4] by Daganzo (1995a).

[1] More precisely, the *directed* capacity, as explained in depth in Section 5.4.3
[2] I.e., does not repeat similar calculations.
[3] Section 5.5.1 clarifies what output relevant solutions are.
[4] These are the most important nodes at highways, and widely used in traffic flow theory.

These nodes have respectively one outgoing or one incoming link. For the merge a priority parameter is required. Node models that do not satisfy the invariance principles may lead to non-stationary turn flow rates in kinematic wave models. For example, calculating an exact solution to the first-order kinematic wave model in continuous time with a node model that does not satisfy the first invariance principle, and using an event-based algorithm such as presented in Raadsen et al. (2014b) leads to an infinite number of events being generated because of flip-flopping flow rates. Finding an approximate solution using the cell transmission model is problematic since it can take long before the flows stabilize. The converge speed furthermore depends on the discretization of space, which is undesirable. Clearly, dynamic network loading procedures that do not average traffic conditions over space, such as the link transmission model, can not handle unstable flow rates and require node models that satisfy the invariance principles.

The focus of this paper is on non-signalized intersections with no detailed geometrical specification. The presented theory serves as fundamental for more specific intersection types. Furthermore the intersection is considered as a point, contrary to spatial extensions. The latter could result in flip-flop effects (see Corthout et al., 2012, Fig. 7), and non-uniqueness remains an issue.

5.2. Descriptive Variables for Traffic Representation at Nodes

To present the traffic representation at nodes, it is helpful to start by introducing the recent analysis by Laval and Leclercq (2013) on the traffic representation on links. Here traffic is represented by a surface in the three-dimensional space spanned by time, location and cumulative flow[5]. Key is that all three dimensions are continuous, including the cumulative flow or ‘vehicle number’. The fact that cumulative flow is represented as continuous is not an assumption nor restrictive for the representation because individual vehicle characteristics are derived by the discretization with interval one. In fact, through this discretization the relations between microscopic and macroscopic models can be explored. Time t, location x and cumulative flow n are continuous, and any of the three variables can be expressed uniquely as a function of the other two. This yields three functional representations of the Moskowitz-surface relating to macroscopic, microscopic, and mesoscopic model formulations.

First, writing the cumulative flow $n = N(x,t)$ as a function of location x and time t is the most common form, for which the two partial derivatives represent negative *density:* $-k = \partial N(x,t)/\partial x$ and *flow:* $f = \partial N(x,t)/\partial t$. The density-flow fundamental diagram is thus a functional relation between the partial derivatives of $N(x,t)$. Also note that the identity of the two second-order derivatives yields the conservation law:

$$-\frac{\partial k}{\partial t} = \frac{\partial^2 N(x,t)}{\partial x \partial t} = \frac{\partial^2 N(x,t)}{\partial t \partial x} = \frac{\partial f}{\partial x}. \tag{5.1}$$

[5]This surface is often referred to as the Moskowitz-surface. See (Newell, 1993a) for historical notes on this representation.

Therefore, $N(x,t)$ and the density-flow fundamental diagram represent the well-known first-order kinematic wave theory model of Lighthill and Whitham (1955); Richards (1956) (LWR). The relation between cumulative flow and the LWR theory was first identified by Newell (1993a).

Second, writing location $x = X(n,t)$ as a function of cumulative flow and time has partial derivatives representing negative *spacing:* $-s = \partial X(n,t)/\partial n$ and *speed:* $v = \partial X(n,t)/\partial t$, which in turn are functionally related by the spacing-speed fundamental diagram. This resembles car-following models, and in particular the theory in (Newell, 2002).

Third, writing time $t = T(n,x)$ as a function of cumulative flow and location is less common. The partial derivatives represent *time-headway:* $h = \partial T(n,x)/\partial n$ and *pace:* $p = \partial T(n,x)/\partial x$, which can be functionally related with a pace-headway fundamental diagram. The model of Leclercq and Bécarie (2012) is based on this representation.

The strength of these three different representations is that the behaviour of the traffic flow is specified through the relationship of two variables with clear physical meaning (i.e., the fundamental diagrams). The first-order dynamics can be solved with the theory of Hamilton-Jacobi differential equations. Besides that the approach is very elegant, it also renews the qualification of macro-, meso-, and microscopic approaches[6]. Macroscopic models correspond with the $N(x,t)$ representation. Mesoscopic models correspond with the $T(n,x)$ representation. And microscopic models correspond with the $X(n,t)$ representation.

The majority of existing link models can be derived in this manner from the Moskowitz-surface and a fundamental diagram. Unfortunately, a straightforward extension of the approach of Laval and Leclercq (2013) towards node models is not possible due to the fact that location is not a continuous variable at nodes. This is a crucial difference, since it means that all earlier presented partial derivatives with a differential of x in the enumerator or denominator only exist in the limit on link extremes, but not on the node itself. Locations that can be identified at nodes are the exits of incoming (upstream) links, the entries of outgoing (downstream) links, and the turns (as the virtual point where a pair of incoming and outgoing links is connected). Location is therefore a discrete variable at nodes. Note that the turns are identified as points, so the node is not spatially expanded. Although location is not continuous, traffic can still be represented by a curve in the time-cumulative flow space at each of these discrete locations, such as shown in Figure 5.1. These curves can then be represented by either one of the functional forms $n = N(t)$ or $t = T(n)$. This means that flow $f = \mathrm{d}n/\mathrm{d}t$ and time-headway $h = \mathrm{d}t/\mathrm{d}n$ are the only descriptive variables that can be derived from the traffic representation at nodes .

For reasons of brevity, in the remainder of this paper when we write headway, we are referring to the time-headway (as opposed to the distance-headway). Disregarding the spatial dimension, flow and headway are each other's reciprocals. That is,

$$\frac{\mathrm{d}n}{\mathrm{d}t} \times \frac{\mathrm{d}t}{\mathrm{d}n} = 1. \tag{5.2}$$

If flow is positive, i.e., $\mathrm{d}N(t)/\mathrm{d}t > 0$, then $N(t)$ and $T(n)$ are each other's inverse functions.

[6]Note that with this qualification, some models that are currently labelled as mesoscopic become different discretizations of micro- or macroscopic models. We follow the mesoscopic definition of Leclercq and Bécarie (2012).

For zero flow the functional form $T(n)$ is not well defined, and the headway would be infinite. So, for all n and t with positive flow we have

$$\frac{\mathrm{d}T(n)}{\mathrm{d}n} = h(n) = \frac{1}{f(t)} = \left(\frac{\mathrm{d}N(t)}{\mathrm{d}t}\right)^{-1}, \tag{5.3}$$

where the headway is strictly given as a function of n, but can be easily rewritten as a function of t by $h(t) = \mathrm{d}T(N(t))/\mathrm{d}n$. As the flow is also given as a function of t, thus the descriptive variables (or functions) for the traffic representation at nodes are $h(t)$ and $f(t)$.

To further clarify the traffic representations at a node in terms of flows $f(t)$ or headways $h(t)$ we consider the topology of the node. Let $\mathcal{I}$ denote the set of incoming (upstream) links (*inlinks*) and $\mathcal{J}$ denote the set of outgoing (downstream) links (*outlinks*) of a node (where for sake of simplicity the index for the node is omitted). Furthermore, we define a turn $\langle i,j\rangle$ for each pair of inlink $i \in \mathcal{I}$ and outlink $j \in \mathcal{J}$, and let $\mathcal{W} = \{\langle i,j\rangle | i \in \mathcal{I}, j \in \mathcal{J}\}$ be the set of all turns. The flows and headways on turns are then defined as $f_{\langle i,j\rangle}(t)$ and $h_{\langle i,j\rangle}(t)$ respectively. Note that the traffic representation on the node is given on turn level, but can be easily aggregated for inlinks and outlinks, where by definition it holds that

$$f_i(t) = \sum_{j\in\mathcal{J}} f_{\langle i,j\rangle}(t), \quad ,\forall i \in \mathcal{I} \quad \text{and,} \quad f_j(t) = \sum_{i\in\mathcal{I}} f_{\langle i,j\rangle}(t) \quad ,\forall j \in \mathcal{J}, \tag{5.4}$$

and it follows that

$$h_i(t) = \frac{1}{\sum_{\langle i,j\rangle\in\mathcal{W}_i^+} \frac{1}{h_{\langle i,j\rangle}(t)}}, \quad \forall i \in \mathcal{I}, \quad \text{and,} \tag{5.5}$$

$$h_j(t) = \frac{1}{\sum_{\langle i,j\rangle\in\mathcal{W}_j^+} \frac{1}{h_{\langle i,j\rangle}(t)}}, \quad \forall j \in \mathcal{J}, \tag{5.6}$$

where $\mathcal{W}_i^+ \subseteq \mathcal{W}$ is the set of turns with positive flow exiting inlink i, and $\mathcal{W}_j^+ \subseteq \mathcal{W}$ is the set of turns with positive flow entering outlink j, and we can define $\mathcal{W}^+ \subseteq \mathcal{W}$ as the set of turns with positive flow. Note that these '+-sets' can change over time with varying travel demand.

Figure 5.1 shows the traffic representation of a three-legged node without U-turns. The time-cumulative flow curves for the individual turns are presented. Both the flows and headways on inlinks, outlinks, and turns can be read from the figure. Note that for each turn $\langle i,j\rangle \in \mathcal{W}^+$ both $N(t)$ and $T(n)$ yield a valid functional representation of the vehicles that pass the turn. The difference between these two representations is that the axes are inverted.

As mentioned earlier, in a node model the demand and supply constraints are given while the flows or headways on the turns are to be computed. Note that the demand and supply constraints for the link extremes are given from the conditions at the boundaries of the links. More precisely, the demand is the maximum outflow of the upstream inlinks (without downstream supply constraints), while the supply is the maximum inflow of the downstream outlinks (without upstream demand constraints) (see also Lebacque and Khoshyaran, 2005). In this paper we focus on the node model and thus assume that the demand, turning fractions (as determined by the route choice model), and supply[7] are exogenously given as input. Therefore the challenge is to find the resulting flows or headways on turns (and links).

[7]Note that spillback from downstream links is captured in this supply.

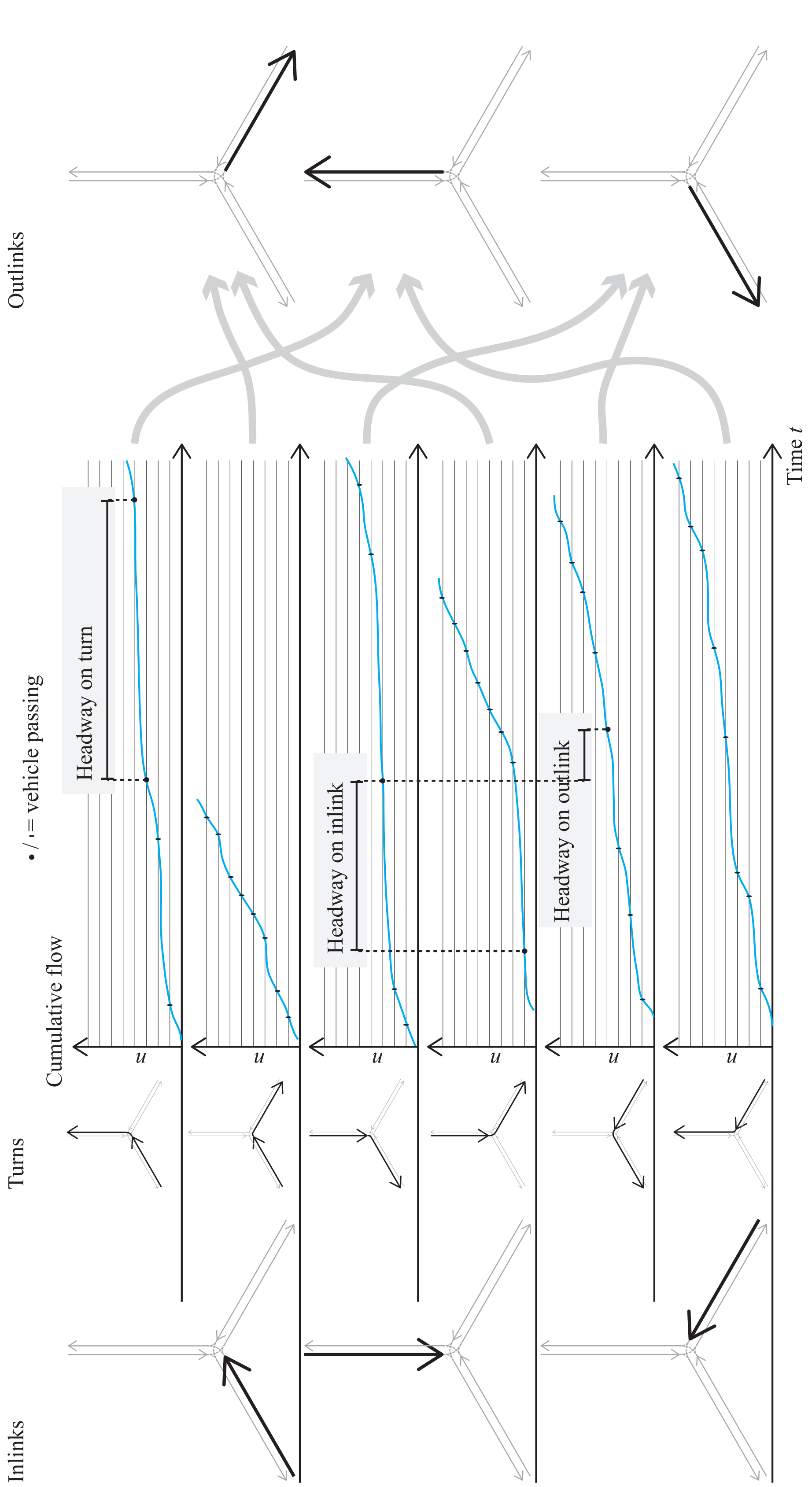


Figure 5.1.: Traffic representation of a 3 by 3 node. Curves in the time-cumulative flow $(t - n)$ space are given for each turn, and the dots indicate when a front bumper passes. Wider dots are used to show example headways.

In this paper we argue that for behavioural interpretation the headway lends itself better for clear analysis and explanation (which we will come back to in the next section). Hence, in the ensuing we discuss the traffic representation and traffic dynamics in terms of headways. Recall that speed and density (and their reciprocals pace and spacing) are rejected as potential variables to describe behaviour at nodes due to the discontinuity of vehicle speeds and distance headways at nodes.

5.3. Family of Node Models Satisfying the Generic Requirements

This section introduces turn delays as descriptive variables, and then shows how all generic requirements (i.e., the requirements for the GCNM in (Tampère et al., 2011)) can be captured in a concise optimization problem in terms of turn delays.

5.3.1. Concept of Turn Delays

In the ensuing, as in most node model formulations, the time index is dropped. This can be done because the problem either considers a time interval or specific time instance, where the solution only depends on the prevailing conditions (i.e., the problem is memory-less). Furthermore, let Q_i be the capacity of link $i \in \mathcal{I} \cup \mathcal{J}$. Let $S_i > 0$ be the demand (or sending flow) from inlink $i \in \mathcal{I}$, and let R_j be the available supply (or receiving flow) at outlink $j \in \mathcal{J}$. Turnfraction $\alpha_{\langle i,j \rangle}$ is the fraction of flow coming from inlink $i \in \mathcal{I}$ which is heading for outlink $j \in \mathcal{J}$, so $\sum_{j \in \mathcal{J}} \alpha_{\langle i,j \rangle} = 1, \forall i \in \mathcal{I}$. Note that $\alpha_{\langle i,j \rangle} = 0$ if $\langle i,j \rangle \notin \mathcal{W}^+$, which is thus a simple test to determine membership of $\mathcal{W}^+$.

With the traffic representation in terms of headways presented in the previous section we can now introduce the concepts of occupancy times and turn delays (that will later enable the behavioural interpretation of the specific cases within the family). The *occupancy time* of a vehicle on an inlink is defined as the time that the vehicle 'occupies' the exit of the inlink in the sense that no other vehicle can exit the inlink during that time. The occupancy time includes the time that the vehicle physically occupies the end of the inlink when passing to the next downstream outlink link, as well as the time that the vehicle virtually occupies the end of the inlink to ensure a safe headway. More formal: the occupancy time interval of a vehicle at a location x starts at the moment x is included in its safety headway, and ends when its rear bumper crossed x.[8] In case of no downstream restrictions, the occupancy time equals $1/Q_i$, that we will call the *capacity occupancy*. Note that the capacity occupancy is a lower bound on the occupancy time, as the occupancy time of vehicles on an inlink can increase in case of downstream restrictions. This increase in occupancy times (additional to the capacity occupancy) that vehicles experience due to downstream restrictions for a specific turn is defined as the *turn delay*.

[8]In strict free flow conditions there will be gaps between occupancy time intervals of successive vehicles, while in congested conditions, these intervals are adjoined.

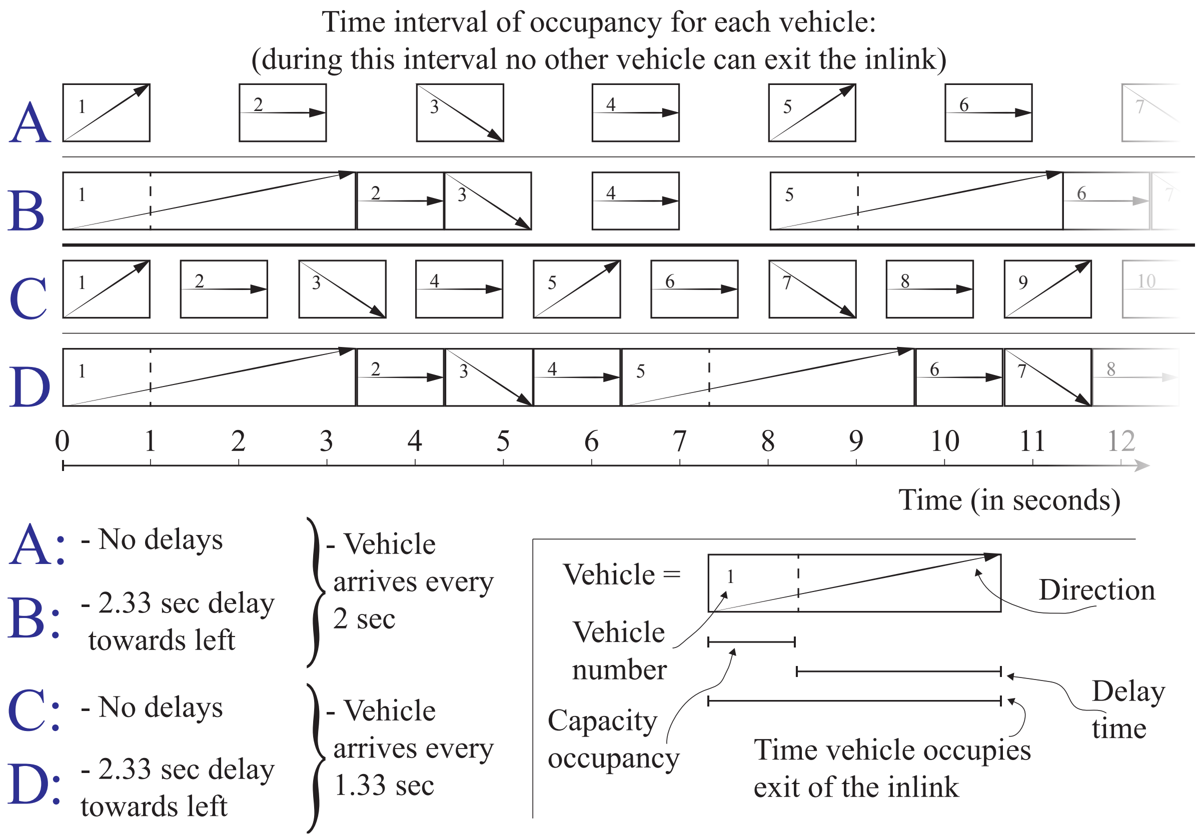


Figure 5.2.: Occupancy of an inlink exit for four possible traffic flows. For each vehicle the time interval it occupies the inlink is presented. Three turns are used: left (25%, vehicles 1,5,9,...), straight (50%, vehicles 2,4,6,...), and right (25%, vehicles 3,7,11,...). Situations A and B have a lower demand than C and D. Situations B and D have a delay for left-turning vehicles. Spillback occurs only in situation D.

To illustrate these concepts, consider the four situations depicted in Figure 5.2. The figure shows the occupancy times of a stream of vehicles in four different situations. The demand is equal in situations A and B, as well as in situations C and D, where the demand is higher for the latter two cases. Furthermore, the occupancy times are equal to the capacity occupancy in situations A and C, while in situations B and D the left-turning vehicles (directions are indicated by the arrows inside the blocks) experience a turn delay. Note that in situations A and C no spillback occurs and the (average) headways are determined by the demand on the upstream inlink (i.e., the flow exiting the inlink is restricted by the demand, but not by the downstream supply). In situation B the headways are affected by the delay times of the left-turning vehicles, however, the delay times are not sufficient to sustain a queue and generate spillback. Hence, the average headways are still determined by the demand on the upstream inlink. Only in situation D spillback occurs due to the combination of high demand (i.e., shorter inter-arrival times) and delay times (for left-turning vehicles). In this situation the (average) headways are no longer only dependent on the demand, but instead restricted by the (downstream) supply constraints. Note that in this representation spillback may also occur when the demand is relatively low, but the fraction of left-turning vehicles with delay time is sufficiently high.

Another way to describe this behaviour is that the slack-time – the white space between vehicles in situations A and C – is sufficient to accommodate a delay to the left in the first case (i.e., transition from situation A to B), but insufficient in the second case (i.e., transition from situation C to D).

Assume that the turn delays $d_{\langle i,j \rangle}$ are known for all turns $\mathcal{W}^+$. Then it is possible to determine whether the combination of demand and turn delays will result in spillback. In case of no spillback, the average headway is completely determined by demand and equals $h_i = 1/S_i$. In case of spillback, the average headway is determined by the capacity occupancy and turn delays per turn with

$$h_i = \frac{1}{Q_i} + \sum_{\langle i,j \rangle \in \mathcal{W}_i} \alpha_{\langle i,j \rangle} d_{\langle i,j \rangle}, \tag{5.7}$$

where the summation term is the average delay per vehicle. In case the average headway exceeds the vehicle inter-arrival time for the demand on that inlink, spillback occurs. It is important to note that here the average (or expected) headway is determined. In reality the arrival process of vehicles is not uniformly distributed over time. However, the computation of the average headway is consistent with the manner in which aggregated macroscopic link flow models also consider average (or expected) link flows, c.q. demand.

Since the candidate solutions for spillback and non-spillback situations both yield lower bounds for the headway, we can determine the resulting headway at inlinks by taking the maximum. Thus

$$h_i = \max \left\{ \frac{1}{Q_i} + \sum_{\langle i,j \rangle \in \mathcal{W}_i} \alpha_{\langle i,j \rangle} d_{\langle i,j \rangle}, \frac{1}{S_i} \right\}, \quad \forall i \in \mathcal{I}. \tag{5.8}$$

The concept of turn delays and equation (8) show that (occupancy times and) turn delays determine the headways on turns and inlinks, yielding a complete representation of the traffic dynamics on a node. Furthermore, the concept of turn delays allows for behavioural interpretation. The key point of a node model is therefore to find turn delays that adequately describe driving behaviour at nodes. These turn delays can be based on capacity restrictions on the downstream outlink (where vehicles have to wait until enough space is available) as well as based on, for instance, priority rules, traffic signals, or internal capacity restrictions. Before discussing the challenges related to determining proper turn delays, we first elaborate more on the behavioural interpretation, and secondly, we formulate the generic requirements (based on the GCNM) in terms of our newly proposed traffic representation and show how these requirements together with equation (5.8) fully specify the complete family.

5.3.2. Observing Delays and Occupancies

The introduction of occupancy times and turn delays allow a behavioural interpretation of node models. The previous section described the effect of delays at inlinks. This section shows how to observe delays and occupancy on complete nodes. The main difficulty in this is that, as pointed out before, the headways describe average behaviour. Given the layout of an intersection and a set of corresponding turn delays, it is not straightforward to retrieve collision-free vehicle trajectories. This is because the description does not take the 'synchronisation' between in- and outlinks into account. For a merge however, the trajectories can easily be determined as shown next.

Figure 5.3 shows the relation between the occupancy times and the actual traffic at a symmetrical merge. Recall that the occupancy time of a vehicle consists of the time its safety headway and the vehicle itself occupy a location. Vehicles at inlink-1 and -2 proceed to the outlink in turn. After a vehicle proceeds to the end of the inlink, it has to wait until the outlink becomes available. During that time it still occupies the inlink, and the follower has to wait in the queue on the link. The lower part of the figure depicts the traffic situation for time instances 4 and 5 (note that it shows *space*-headways, which are significantly different from *time*-headways). At time=4 vehicle #3 just arrived at the front of the inlink-1 and starts waiting, while vehicle #2 just cleared from inlink-2 which allows vehicle #4 to proceed to the front. At time=5 vehicle #3 is halfway its wait (i.e. delay), and vehicle #4 is halfway its occupancy time.

To be able to validate models, the occupancy times and especially the delays have to be observed. Real traffic does not depict the average situations, but is subject to arrival processes and rather fluctuates. Therefore measurements have to be averaged over a time period. For every vehicle in such an interval its turn delay and its turn direction have to be measured; of which the latter is straightforward. The capacity occupancy can be derived directly from the capacity of the inlink. So when the occupancy time can be measured, the turn delay is acquired by subtracting the capacity occupancy.

So the remaining question is how to measure the occupancy times. A distinction can be made between situation A+C, B, and D from Figure 5.2. Situation D is the most important one because it describes spillback, and it is also the easiest situation to determine the occupancy times since no slack time exists; the occupancy time equals the time between two successive rear-bumper passages at the end of the inlink. If all inlinks are in situations A and C no supply constraint is active, and therefore all turn delays are zero and nothing has to be measured. If it can not be observed that no supply constraint is active, or if situation B occurs, then the

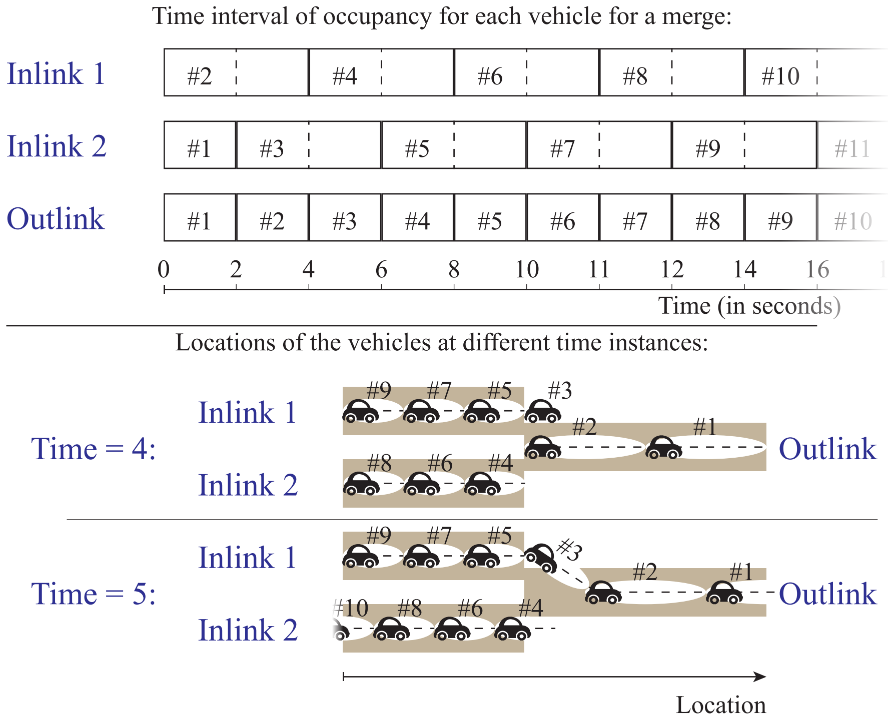


Figure 5.3.: Occupancy times at a merge where the capacity of each link is 1800 vehicles per hour, and traffic spills back to both inlinks. The occupancy times consist of a capacity occupancy and the turn delay (see also Figure 5.2), and are shown with white ellipses for each vehicle. The empty area in front of each vehicle represents its safety headway.

difficulty is to determine when a vehicles occupancy time begins (since this starts when the exit of the inlink is part of the safety headway). One approach to solve this is taking a fixed safety headway, and subtract this from the time the front bumper passes the exit of the inlink[9]. Section 5.7 suggests future research on this topic.

The existing models that go beyond merges and diverges lack this link with observed traffic, where only Flötteröd and Rohde (2011) verify their model for one conflict point at a signalized intersection. This shows feasibility for continuous models and models with small time steps. There is a need for more empirical verification or calibration of macroscopic point-wise node models for specific intersections. The connection to observed traffic is an important property of the formulation in terms of turn delays, and therefore an important contribution in this paper.

5.3.3. Generic Requirements

The requirements that form the first-order GCNM as described in (Tampère et al., 2011) can be stated in terms of flows and headways. The original formulation is based on flows. In this section the requirements are formulated based on headways, and are interpreted by observing what happens at inlinks (since inlinks are particularly interesting as this is where queues can start and spillback can occur). For a full derivation of the requirements, we refer to Tampère et al. (2011). For reasons of clarity, the requirements follow in the same order as in (Tampère et al., 2011).

General applicability, irrespective of the number of incoming and outgoing links. As stated by Tampère et al. (2011): “node models should be applicable to any combination of number of incoming and outgoing links”. This requirement relates to the mathematical problem formulation of the node model, and is independent of the traffic representation at the node.

Maximizing flows. The flow maximization requirement has this explanation: “each flow should be actively restricted by one of the constraints, otherwise it would increase until it hits some constraint”. The underlying behavioural assumption is that drivers accelerate whenever possible. In terms of turn delays the equivalent formulation is behaviourally more elegant and intuitive, namely that turn delays are minimized,

$$\min \left\{ d_{\langle i,j \rangle} \middle| \langle i,j \rangle \in \mathcal{W} \right\}. \tag{5.9}$$

This multi-objective problem minimizing individual delays is in accordance with the individual flow maximization formulated by Tampère et al. (2011). However, note that in their mathematical formulation the *sum* of flows is maximized, which appears less realistic from a behavioural perspective.

Non-negativity. As traffic does not flow upstream, flows are required to be non-negative. Equivalently, we can state that headways are non-negative,

$$h_{\langle j,j \rangle} \geq 0, \quad \forall \langle i,j \rangle \in \mathcal{W}^{+}. \tag{5.10}$$

[9]However, in this case the occupancy time becomes completely dependent on the speed and length of the vehicle which might not be realistic.

Conservation of vehicles. Since the node is not a source or sink for traffic, the total flows from the inlinks should equal the total flows into the outlinks. Considering that flow and headway are inversely related, the equivalent requirement in terms of headways can be simply stated as,

$$\sum_{i \in \mathcal{I}} \frac{1}{h_i} = \sum_{j \in \mathcal{J}} \frac{1}{h_j} \tag{5.11}$$

Note that the conservation of vehicles is guaranteed through the definitions in equations (5.5,5.6).

Satisfying demand and supply constraints. Once again considering the inverse relationship between flow and headway, we can state that the demand constraint requires the headway on an inlink not to be smaller than the inverse of the demand (or sending flow) for that inlink,

$$h_i \geq \frac{1}{S_i}, \quad \forall i \in \mathcal{I}, \tag{5.12}$$

and the supply constraint requires the headway on an outlink not to be smaller than the inverse of the available supply (or receiving flow) for that outlink,

$$h_j \geq \frac{1}{R_j}, \quad \forall j \in \mathcal{J}. \tag{5.13}$$

Obeying conservation of turning fractions. The conservation of turning fractions (due to the first-in-first-out assumption at links and that turning fractions are exogenously determined by the route choice model) and equation (5.5) give a very straightforward relation between headways on exits of inlinks and on turns, that satisfy the conservation of turning fractions constraint, namely

$$h_{\langle i,j \rangle} = \frac{h_i}{\alpha_{\langle i,j \rangle}}, \quad \forall \langle i,j \rangle \in \mathcal{W}^+. \tag{5.14}$$

Compatibility with link traffic flow dynamics: satisfaction of the invariance principles. The derivation of headways at inlinks in equation (5.8) resolves the first invariance principle (related to inlinks) in a natural way. Under the assumption that the demand is *only* included as the constraints in equation (5.12), the headways are determined by demand if and only if there is no spillback on that link. In other words, only when term $1/S_i$ is smaller than term $1/Q_i + \sum_{\langle i,j \rangle \in \mathcal{W}_i} \alpha_{\langle i,j \rangle} d_{\langle i,j \rangle}$ spillback will occur; the resulting headways are then independent of demand S_i. For the second invariance principle it is required that the supply is *only* included as the constraint in equation (5.13). In the next section we show that this supply constraint can be written in terms of turn delays and that a set of turn delays is the solution of a node model. When for such a solution the supply constraint holds with strict inequality, we can remove this supply constraint without changing the solution. Therefore the second invariance principle holds if the supply is only used as supply constraint. (Note that this argument also holds for the demand constraints related to the first invariance principle).

5.3.4. Problem Formulation based on Turn Delays

With the newly introduced traffic representation at nodes and the generic requirements, the family of node models can be formulated as a mathematical optimization problem where the turn delays are the only unknowns. As discussed above, most requirements are automatically satisfied with the definition of turn delays and by equation (5.8). The remaining requirements are the supply constraints and flow maximization. The former constraints can be added in the latter optimization problem. That leads to a mathematical problem which completely reflects the generic requirements and where the unknown variables are the turn delays.

Equation (5.8) can be rewritten by substituting equations (5.6), (5.14), and (5.8) in that order, which results in:

$$\frac{1}{\sum_{i\in\mathcal{I}}\left(\frac{\max\left\{\frac{1}{Q_i}+\sum_{\langle i,j'\rangle\in\mathcal{W}_i}\alpha_{\langle i,j'\rangle}d_{\langle i,j'\rangle},\frac{1}{S_i}\right\}}{\alpha_{\langle i,j'\rangle}}\right)^{-1}} \geq \frac{1}{R_j}, \quad \forall j \in \mathcal{J}. \tag{5.15}$$

This can be rewritten as

$$\sum_{i\in\mathcal{I}} \min\left\{\alpha_{\langle i,j\rangle}S_i, \frac{\alpha_{\langle i,j\rangle}Q_i}{1+Q_i\sum_{\langle i,j'\rangle\in\mathcal{W}_i}\alpha_{\langle i,j'\rangle}d_{\langle i,j'\rangle}}\right\} \leq R_j, \quad \forall j \in \mathcal{J}. \tag{5.16}$$

These constraints can be combined to form a single constraint by observing that the minimum difference between the right-hand-side and left-hand-side should be positive,

$$\min_{j\in\mathcal{J}}\left\{R_j - \sum_{i\in\mathcal{I}} \min\left\{\alpha_{\langle i,j\rangle}S_i, \frac{\alpha_{\langle i,j\rangle}Q_i}{1+Q_i\sum_{\langle i,j'\rangle\in\mathcal{W}_i}\alpha_{\langle i,j'\rangle}d_{\langle i,j'\rangle}}\right\}\right\} \geq 0. \tag{5.17}$$

Then the turn delay minimization requirement is captured in the multi-objective optimization Node Problem (NP)

$$\begin{aligned} &\min_{d_{\langle i,j\rangle}\geq 0, \langle i,j\rangle\in\mathcal{W}} \left\{d_{\langle i,j\rangle} \middle| \langle i,j\rangle \in \mathcal{W}\right\} \qquad \text{such that} \\ &\min_{j\in\mathcal{J}}\left\{R_j - \sum_{i\in\mathcal{I}} \min\left\{\alpha_{\langle i,j\rangle}S_i, \frac{\alpha_{\langle i,j\rangle}Q_i}{1+Q_i\sum_{\langle i,j'\rangle\in\mathcal{W}_i}\alpha_{\langle i,j'\rangle}d_{\langle i,j'\rangle}}\right\}\right\} \geq 0. \end{aligned} \tag{NP}$$

This problem formulation is completely consistent and equivalent with the GCNM. Also, when the turn delays $d_{\langle i,j\rangle}$ are computed, then the headways at inlinks and outlinks follow directly, and can be translated to flow with:

$$f_i = \min\left\{\frac{Q_i}{1+Q_i\sum_{j\in\mathcal{J}}\alpha_{\langle i,j\rangle}d_{\langle i,j\rangle}}, S_i\right\}, \quad \forall i \in \mathcal{I}, \tag{5.18}$$

$$f_j = \sum_{i\in\mathcal{I}} \alpha_{\langle i,j\rangle} f_i, \quad \forall j \in \mathcal{J}. \tag{5.19}$$

However, (NP) has multiple objectives, and thus multiple solutions, and it has a very non-linear constraint. All Pareto optimal solutions of (NP) are solutions consistent with the generic requirements. Because the problem is underspecified and does not have a unique solution, additional behavioural assumptions and accompanying constraints should be added to make the problem tractable and to provide a unique solution. The way that these additional constraints are chosen then determines the specific case of a node model within the node model family. Many assumptions can be made, yielding different additional constraints, and hence are different specific members of the family. Note however that such constraints can not depend directly on the demand or the supply due to the invariance principles.[10] Existing and new node models can be analysed by specifying their additional constraint and the corresponding underlying behavioural assumption in terms of turn delays. In the next section the characteristics of four members of the node model family are analysed.

5.4. New and Existing Members of the Family

At present, the only node models satisfying the generic requirements are the models presented in the papers by Gibb (2011) and Tampère et al. (2011) (equivalent to the node model presented by Flötteröd and Rohde (2011)). In this section we show how these models are captured as specific cases within the node model family, and which additional constraints are added to (NP). Furthermore, two new members of the family are introduced based on two different behavioural assumptions.

5.4.1. Single Server

The first new member is the first-come-first-serve single server node. In this simple model vehicles have to take turns to pass the node, and they all get a delay – regardless of the possibility that their outlink is freely available. Such a situation could occur on highly saturated nodes, where the middle of the intersection is blocked by vehicles. In this case, the additional information (i.e., constraints) for (NP) is that the delays are inversely proportional to the capacity with the same proportionality constant for every turn, i.e.

$$d_{\langle i,j\rangle} Q_i = d_{\langle i',j'\rangle} Q_{i'}, \quad \forall \langle i,j\rangle, \langle i',j'\rangle \in \mathcal{W}. \tag{5.20}$$

The reason that delays are inversely proportional with the capacity in such a system is best understood when thinking of two competing inlinks. Inlink-1 has one lane, and inlink-2 has two lanes. Traffic arriving at the end of inlink-2 can take its turn when the single lane of inlink-1 is not occupied anymore. On the other hand, an arriving vehicle at the end of inlink-1 can take its turn when *both* lanes of inlink-2 become free. Therefore, the delay at inlink-1 is twice as long as the delay at inlink-2. This behaviour is equivalent to the *capacity-based weighted fair queuing* principle for merges of Ni and Leonard II (2005) (see Section 5.5.4).

In order to combine equation (5.20) and problem (NP), introduce delay constant c such that $d_{\langle i,j\rangle} = c/Q_i, \forall \langle i,j\rangle \in \mathcal{W}$. This constant can be interpreted in terms of occupancy time and capacity occupancy introduced in Section 5.3.1. For every turn, the occupancy time is a multiple of the capacity occupancy, and the multiplication constant is equal, namely $1+c$. So, if the delay constant $c = 1$, then the occupancy time is twice its capacity occupancy.

[10] This is why the model by Jin and Zhang (2003) does not satisfy the first invariance principle. Their priority parameters depend directly on the demand.

Equation (5.20) actually connects (i.e. weighs) all turn delays with each other such that (NP) becomes single objective. Substituting the turn delays with the delay constants leads to the following Single Server node Problem (SSP):

$$\min_{c \geq 0} c \qquad \text{such that}$$
$$\min_{j \in \mathcal{J}} \left\{ R_j - \sum_{i \in \mathcal{I}} \min \left\{ \alpha_{\langle i,j \rangle} S_i, \frac{\alpha_{\langle i,j \rangle} Q_i}{1+c} \right\} \right\} \geq 0. \tag{SSP}$$

It has become equivalent to minimize delay constant c now, because the turn delays are strictly increasing in c.

5.4.2. Equal Delay at Outlink

A more natural member of the family emerges when turn delays are equal for every outlink. In other words, that for each turn towards an outlink the turn delays are equal; this leads to the following additional constraints for (NP):

$$d_{\langle i,j \rangle} = d_{\langle i',j \rangle}, \quad \forall \langle i,j \rangle, \langle i',j \rangle \in \mathcal{W}_j, \forall j \in \mathcal{J}. \tag{5.21}$$

These constraints have a clear and straightforward behavioural interpretation, namely that each vehicle has to wait the same amount of time to enter a particular outlink. Denote the turn delay towards outlink $j \in \mathcal{J}$ as $\tilde{d}_j$. Then the Equal Delay node Problem (EDP) becomes

$$\min_{\tilde{d}_j \geq 0, j \in \mathcal{J}} \left\{ \tilde{d}_j \middle| j \in \mathcal{J} \right\} \qquad \text{such that}$$
$$\min_{j \in \mathcal{J}} \left\{ R_j - \sum_{i \in \mathcal{I}} \min \left\{ \alpha_{\langle i,j \rangle} S_i, \frac{\alpha_{\langle i,j \rangle} Q_i}{1 + Q_i \sum_{\langle i,j' \rangle \in \mathcal{W}_i} \alpha_{\langle i,j' \rangle} \tilde{d}_{j'}} \right\} \right\} \geq 0. \tag{EDP}$$

The number of objectives is reduced significantly since turn delays are fixed to each other; however, the solution is not straightforward as will be shown later.

5.4.3. Directed Capacity Proportional

The node model by Tampère et al. (2011) and Flötteröd and Rohde (2011) assumes directed capacity proportionality when supply is distributed over demand at outlinks. Directed capacity is defined for each turn, and equals the proportion of its inlink's capacity that is assigned to that turn (i.e. $\alpha_{\langle i,j \rangle} Q_i, \forall \langle i,j \rangle \in \mathcal{W}$). The proportionality at outlinks means that the supply is divided among competing flows proportional to the directed capacity of the corresponding turns. The second assumption is that the flow at an inlink is completely determined by either its demand or *one* supply constraint. This implies that only one turn of each inlink can have a positive delay.

As already mentioned in Section 5.1 of this paper, both papers have a different approach to reach the same result. The equivalent solutions are repeated procedures that fix at least one inlink at every step/repetition. So, the models are computationally efficient since nodes generally do not have more than four inlinks (and thus at most four repetitions are required). In every step two scenarios can occur: (1) For some inlinks it is guaranteed that no supply constraint can restrict them (i.e. no turn delay will ever be big enough to cause spillback), and thus the flow is set to demand. (2) The final turn delays for an outlink can be determined; this fixes all inlinks connected to that outlink. In the latter scenario the turn delays have to be determined according to the directed capacity proportionality principle. To do that the inlinks that are already marked 'demand constrained' by the first scenario and the ones already blocked by another supply constraint are omitted (i.e. not competing). Therefore, the proportionality only holds for inlinks that are actively blocked during that step.

The relation with turn delays, and thus with traffic behaviour, can be analysed in a similar fashion as for the single server model. However, the translation from flow to headways needs to be made, and the fact that inlinks can only be blocked by one outlink has to be taken into account. Whenever a supply constraint becomes active at an outlink it will determine all (positive) delays for turns going into that link. As the previous paragraph describes, some of the competing inlinks are already fixed, and the remaining ones are resolved with the directed capacity proportionality rule. This rule can be written as

$$\frac{f_{\langle i,j\rangle}}{\alpha_{\langle i,j\rangle}Q_i} = \frac{f_{\langle i',j\rangle}}{\alpha_{\langle i',j\rangle}Q_{i'}}, \quad \forall\langle i,j\rangle,\langle i',j\rangle \in \hat{\mathcal{W}}_j, \forall j \in \mathcal{J}, \tag{5.22}$$

where $\hat{\mathcal{W}}_j$ is the set of *competing* turns; note that $\hat{\mathcal{W}}_j \subseteq \mathcal{W}_j^+$ holds because turns with no flow cannot compete. These equalities can be rewritten in terms of headways as

$$h_{\langle i,j\rangle}\alpha_{\langle i,j\rangle}Q_i = h_{\langle i',j\rangle}\alpha_{\langle i',j\rangle}Q_{i'}, \quad \forall\langle i,j\rangle,\langle i',j\rangle \in \hat{\mathcal{W}}_j, \forall j \in \mathcal{J}. \tag{5.23}$$

Since each of these inlinks is constrained by this supply *only* the headway can – in this case – be written as

$$h_{\langle i,j\rangle} = \frac{1}{(\alpha_{\langle i,j\rangle}Q_i)} + d_{\langle i,j\rangle}, \quad ,\forall\langle i,j\rangle \in \hat{\mathcal{W}}_j. \tag{5.24}$$

Note that this is only possible because all other turn delays are equal to zero, otherwise $1/\alpha_{\langle i,j\rangle}Q_i$ (the rate which vehicles can arrive at turn $\langle i,j\rangle$ when the inlink is congested) would be affected by these delays. Substituting Equation (5.24) in Equation (5.23) leads to

$$1 + d_{\langle i,j\rangle}\alpha_{\langle i,j\rangle}Q_i = 1 + d_{\langle i',j\rangle}\alpha_{\langle i',j\rangle}Q_{i'}, \quad \forall\langle i,j\rangle,\langle i',j\rangle \in \hat{\mathcal{W}}_j, \forall j \in \mathcal{J}. \tag{5.25}$$

When the ones at both sides are removed, this shows that the turn delays and the directed capacity are inversely proportional with the same proportionality constant. Introduce, similarly as for the single server node, c_j as the delay constant at outlink $j \in \mathcal{J}$ such that $d_{\langle i,j\rangle} = c_j/(\alpha_{\langle i,j\rangle}Q_i)$. When this is finally substituted in (NP), then, by using that only the most restrictive delay constant is 'active', the Directed Capacity Proportional node Problem (DCPP) becomes

$$\min_{c_j \geq 0, j\in\mathcal{J}} \left\{ c_j \middle|, j \in \mathcal{J} \right\} \qquad \text{such that}$$

$$\min_{j\in\mathcal{J}} \left\{ R_j - \sum_{i\in\mathcal{I}} \min\left\{ \alpha_{\langle i,j\rangle}S_i, \frac{\alpha_{\langle i,j\rangle}Q_i}{1 + \max_{\{j' | \langle i,j'\rangle \in \mathcal{W}_i^+\}} c_{j'}} \right\} \right\} \geq 0. \tag{DCPP}$$

Again, it is equivalent to minimize the delay constants because each turn delay is strictly increasing in its c_j.

Now it is interesting to analyse the behavioural meaning of constants c_j. Where the inverse proportionality at the single server node was induced by turn-taking behaviour, the underlying behaviour here is more unusual, especially when the turn flow is a small part of the total demand of the inlink. Gibb (2011) also points this out with a more detailed example. Assume some $c_j > 0$ for some outlink, and assume that two inlinks with equal capacity are competing. The inlinks are saturated and thus have a demand equal to capacity. Ten percent of the vehicles at inlink-1, and all vehicles at inlink-2 bound towards the outlink. The ratio between the directed capacities is then 1 to 10. This means that the experienced turn delay of vehicles from inlink-1 is ten times larger than the experienced turn delay of vehicles from inlink-2. If $c_j = 1$, then each vehicle at inlink-1 has to wait for ten vehicles to pass from inlink-2. This is unrealistic when one remembers that the capacities of both inlinks are identical. Furthermore, the vehicles towards other directions at inlink-1 are affected by the long delay of this particular turn.

The challenge of the model is to find the correct c_j's, which is solved by Tampère et al. (2011) as well as Flötteröd and Rohde (2011). Tampère et al. (2011) look for capacity to flow restriction factors β_j. The relations between β_j and c_j are

$$c_j = \frac{1-\beta_j}{\beta_j}, \quad \beta_j = \frac{1}{1+c_j}, \quad \forall j \in \mathcal{J}.$$

Instead of directed capacity proportionality other priority coefficients can be included (Flötteröd and Rohde, 2011; Tampère et al., 2011; Corthout et al., 2012). The coefficients are inlink specific and should be calibrated to model specific intersections[11]. However, the flexibility of this approach is limited since such coefficients are often turn specific instead of inlink specific. Flötteröd and Rohde (2011) also provide a fixed point method to include additional node supply constraints that can for example represent conflicting turns. Similar additional internal node constraints are also presented in (Tampère et al., 2011; Corthout et al., 2012), but no concrete solution method is provided. Uniqueness in terms of flow is not guaranteed for these generalizations (see Section 5.5.2).

The generalizations in the latter paragraph still have the very restrictive property that flows at inlinks can only be constrained by a single constraint (either demand, supply, or internal). Interaction between multiple supply constraints at an inlink can be achieved with the notion of turn delays.

5.4.4. Capacity Consumption Equivalence

The node model by Gibb (2011) is based on capacity consumption equivalence, assuming that traffic towards an outlink that operates at supply consumes more capacity of its inlink. The capacity consumption of a vehicle depends on the outlink j it is heading towards. Denote per outlink $j \in \mathcal{J}$ capacity consumption factor $e_j \geq 1$. The factor equals 1 if the supply is not fully

[11] Note that this calibration can be cumbersome because only observed flows and predicted flows can be compared.

utilized on the outlink. The factor determines for each vehicle the time it occupies the inlink (i.e. the length of the vehicle 'blocks' in Figure 5.2). A vehicle on turn $\langle i,j\rangle \in \mathcal{W}$ occupies the end of the inlink e_j/Q_i time units, that is e_j times the capacity occupancy. Gibb uses this time to determine a new 'demand' with increased capacity consumption for each inlink $i \in \mathcal{I}$ via $\sum_{\langle i,j\rangle\in\mathcal{W}_i} e_j\alpha_{\langle i,j\rangle}S_i$. If this 'demand' is larger than capacity, then a reduction factor is applied to the demand, this factor equals

$$\theta_i = \min\left\{\frac{Q_i}{\sum_{\langle i,j\rangle\in\mathcal{W}_i} e_j\alpha_{\langle i,j\rangle}S_i}, 1\right\}, \quad \forall i \in \mathcal{I}. \tag{5.26}$$

The resulting flows of Gibb's model are $f_i = \theta_i S_i, \forall i \in \mathcal{I}$. By using that flow and headway are reciprocals, and by using equation (5.26) the headways can be written as

$$h_i = \max\left\{\sum_{\langle i,j\rangle\in\mathcal{W}_i} \alpha_{\langle i,j\rangle}\frac{e_j}{Q_i}, \frac{1}{S_i}\right\}, \quad \forall i \in \mathcal{I}, \tag{5.27}$$

which is very similar to equation (5.8). By observing that the first elements of the max operators are equal, the relation between turn delays and capacity consumption factors can be derived, which is

$$d_{\langle i,j\rangle} = \frac{e_j - 1}{Q_i}, \quad \forall \langle i,j\rangle \in \mathcal{W}. \tag{5.28}$$

The Capacity Consumption Equivalence node Problem (CCEP) is derived by substituting equation (5.28) in NP:

$$\min_{e_j\geq 1, j\in\mathcal{J}} \left\{e_j \middle| j \in \mathcal{J}\right\} \qquad \text{such that}$$
$$\min_{j\in\mathcal{J}}\left\{R_j - \sum_{i\in I}\min\left\{\alpha_{\langle i,j\rangle}S_i, \frac{\alpha_{\langle i,j\rangle}Q_i}{\sum_{\langle i,j'\rangle\in\mathcal{W}_i}\alpha_{\langle i,j'\rangle}e_{j'}}\right\}\right\} \geq 0. \tag{CCEP}$$

It has become equivalent to minimize to capacity consumption factors because each turn delay is strictly increasing in its determining e_j.

5.5. Analysis of the Node Model Family

As mentioned earlier, each set of turn delays that is a Pareto optimal solution of (NP) satisfies all generic requirements. Indeed, each of these sets of turn delays resembles different behaviour at the node. The discussed members of the family add mechanisms and constraints to select a single solution. However, (EDP), (DCPP), and (CCEP) still have multiple objectives, and it is not straightforward to show that they lead to a unique solution. It is actually true that multiple Pareto optimal solutions of these problems exist. Despite this, their result in terms of resulting flows is always the same. That multiple Pareto optimal solutions of the turn delays can yield the same resulting flows (at inlinks and outlinks) is clarified in the following section.

5.5.1. Reduced Capacity and Model Equivalence

Because multiple sets of turn delays can lead to the same flows on a node, the result (or output) relevant set of solutions is analysed. Due to the conservation of turning fractions the ratio between the flows on two turns exiting the same inlink is equal for each solution; so when the flow at the inlinks is known, then the flows at the outlinks are also known. In addition, the resulting flows at inlinks are solely determined with equation (5.18). This equation consists of two terms, the demand (which is input), and the maximum flow in case of spillback under the turn delays. The latter will be called the *reduced capacity* $q_i, i \in \mathcal{I}$, and is defined with

$$q_i = \frac{Q_i}{1 + Q_i \sum_{\langle i,j\rangle \in \mathcal{W}_i} \alpha_{\langle i,j\rangle} d_{\langle i,j\rangle}}, \quad \forall i \in \mathcal{I}. \tag{5.29}$$

This is exactly the reciprocal of the average headway at inlinks during spillback as defined in equation (5.7). The reduced capacity is determined by the capacity of the inlink and the delays at all turns connected to the inlink.

Two Pareto optimal solutions of (NP) yield the same flows if and only if the reduced capacity of the inlinks with spillback are equal. In other words, the results of two sets of turndelays, $\{d_{\langle i,j\rangle} | \langle i,j\rangle \in \mathcal{W}\}$ and $\{d'_{\langle i,j\rangle} | \langle i,j\rangle \in \mathcal{W}\}$ are *equivalent* when

$$\frac{Q_i}{1 + Q_i \sum_{\langle i,j\rangle \in \mathcal{W}_i} \alpha_{\langle i,j\rangle} d_{\langle i,j\rangle}} = \frac{Q_i}{1 + Q_i \sum_{\langle i,j\rangle \in \mathcal{W}_i} \alpha_{\langle i,j\rangle} d'_{\langle i,j\rangle}}, \quad \forall i \in \hat{\mathcal{I}}, \tag{5.30}$$

where $\hat{\mathcal{I}}$ is the set of inlinks in spillback conditions under either set of turn delays.

By using the reduced capacities, the result relevant solutions set can be analysed. The number of 'unknowns' is much lower than when turn delays are considered. By substituting equation (5.29) in the constraint of (NP), the feasible set of reduced capacities is determined by

$$\min_{j \in \mathcal{J}} \left\{ R_j - \sum_{i \in \mathcal{I}} \min \left\{ \alpha_{\langle i,j\rangle} S_i, \alpha_{\langle i,j\rangle} q_i \right\} \right\} \geq 0, \tag{5.31}$$

furthermore, $0 < q_i \leq Q_i, \forall i \in \mathcal{I}$ due to non-negativity and finiteness of turn delays. In Section 5.5.5 the feasible region for an example node problem is presented.

5.5.2. Non-Uniqueness

Models can be unique at three different levels. First, they can be unique in terms of resulting flows; this is a desirable property for applications. Second, they can be unique in terms of turn delays. This latter is a less important model property for application, because two sets of turn delays can lead to the same flows. And third, uniqueness in any of the other describing variables, such as reduced capacity, can be determined. Since this in general has no important behavioural interpretation, we analyse uniqueness only in terms of flow and turn delay.

Corthout et al. (2012) analyse non-uniqueness based on flows. They show that for a generalisation of the directed capacity proportional node model with arbitrary priority parameters (i.e., not directed capacity), multiple solutions exist. This type of non-uniqueness is subsistent in traffic, and they reproduce this with microscopic simulations. The four models in the node model family in this paper are all unique in terms of flows, which is discussed per model in the next section.

The single server problem is unique in terms of turn delays as well. For the models, a symmetrical diverge with spillback does not have unique solutions in terms of turn delays (see also discussions in next section). For the directed capacity proportional model for example, a delay exists only towards the outlink that is considered first in the algorithm, which is arbitrary (see Sections 5.4.3 and 5.5.3.3). Ideally, the additional behavioural constraints of a model are such that they determine the turn delays uniquely; in that case the connection to underlying behaviour is unambiguous.

5.5.3. Solution Methods

Before the methods are presented for finding the solution for each member, an overview of the additional assumptions is presented. Table 5.1 contains the unknowns, reduced capacity definitions, turn delay definitions, and some remarks of every known member of the node model family.

Key in finding the solution of the node problem is the analysis of the constraint of (NP). Obviously, when $d_{\langle i,j\rangle} \equiv 0$ is feasible, then it is the solution. In this case no spillback occurs and the supply is sufficient to accommodate the demand. Otherwise, a solution has to be found that lies on the boundary of the constraint (i.e., for the solution the constraint holds with equality). The problem then becomes finding a point on this boundary that satisfies the additional constraints as defined for each member of the family. The solution methods below focus on problems with non-trivial solutions (i.e., $\exists d_{\langle i,j\rangle} > 0$).

5.5.3.1. Single Server

The single server problem (SSP) is the only model with one unknown. Finding c is equivalent to finding $1/(1+c)$ which will be denoted with y. The solution can be found by solving

$$\min_{j\in\mathcal{J}}\left\{R_j - \sum_{i\in\mathcal{I}} \min\left\{\alpha_{\langle i,j\rangle}S_i, \alpha_{\langle i,j\rangle}Q_i y\right\}\right\} = 0, \tag{5.32}$$

which is firstly rewritten to

$$\max_{j\in\mathcal{J}}\left\{-R_j + \sum_{i\in\mathcal{I}} -\alpha_{\langle i,j\rangle}Q_i \max\left\{-y, -\frac{S_i}{Q_i}\right\}\right\} = 0. \tag{5.33}$$

This is a polynomial equation over the algebraic structure $(\mathbb{R},\max,+)$ (i.e. the max-plus algebra). Linear time algorithms to rewrite and solve polynomial equations in the max-plus algebra are presented in Cuninghame-Green and Meijer (1980) and Cuninghame-Green (1995).

First, rewrite the inner sum of maximizations in equation (5.33) to a minimization of sums. Assume that $\mathcal{I} = \{1,\ldots,|\mathcal{I}|\}$ is sorted such that $-S_i/Q_i \le -S_{i+1}/Q_{i+1}$ for $i = 1,\ldots,|\mathcal{I}|-1$. Define for all $j \in \mathcal{J}$ the constants $g_{ij} = \sum_{i'=1}^{i} -\alpha_{\langle i',j\rangle}Q_{i'}$ for $i = 1,\ldots,|\mathcal{I}|$, and $g_{0j} = 0$; and

Table 5.1.: Summary of GCNM family member characteristics

	Unknowns (#)	Reduced capacity q_i	Turn delay $d_{\langle i,j\rangle}$	Remarks
(NP)	$d_{\langle i,j\rangle}$ $(\lvert\mathcal{I}\rvert\times\lvert\mathcal{J}\rvert)$	$= \dfrac{Q_i}{1+Q_i\sum_{j\in\mathcal{J}}\alpha_{\langle i,j\rangle}d_{\langle i,j\rangle}}$	$= d_{\langle i,j\rangle}$	The most general formulation
(SSP)	c (1)	$= \dfrac{Q_i}{1+c}$	$= \dfrac{c}{Q_i}$	Every vehicle is delayed, also vehicles towards unsaturated links
(EDP)	$\tilde{d}_j$ $(\lvert\mathcal{J}\rvert)$	$= \dfrac{Q_i}{1+Q_i\sum_{j\in\mathcal{J}}\alpha_{\langle i,j\rangle}\tilde{d}_j}$	$= \tilde{d}_j$	At every outlink, the delay is equally distributed amongst vehicles
(DCPP)	c_j $(\lvert\mathcal{J}\rvert)$	$= \dfrac{Q_i}{1+\max_{\{j\mid\langle i,j\rangle\in\mathcal{W}_i^+\}}c_j}$	$= \dfrac{c_j}{\alpha_{\langle i,j\rangle}Q_i}$	Inlinks can only be blocked by one outlink
(CCEP)	e_j $(\lvert\mathcal{J}\rvert)$	$= \dfrac{Q_i}{\sum_{j\in\mathcal{J}}\alpha_{\langle i,j\rangle}e_j}$	$= \dfrac{e_j-1}{Q_i}$	Delay depends on inlink and outlink characteristics

$\upsilon_{ij} = -R_j + \sum_{i'=i+1}^{|\mathcal{I}|} \alpha_{\langle i',j\rangle} S_{i'}$ for $i = 0,\ldots,|\mathcal{I}|-1$, and $\upsilon_{|\mathcal{I}|j} = -R_j$. Then in the same line[12] as Lemma 6.1 and its preceding theory in Cuninghame-Green (1995) equation (5.33) can be written as

$$\max_{j\in\mathcal{J}} \min\left\{\upsilon_{0j}, \min_{i\in\mathcal{I}} \upsilon_{ij} - g_{ij} y\right\} = 0. \tag{5.34}$$

The left-hand-side is a piece-wise linear function in y. Because the g_{ij}s are non-positive, the $\min\{\cdots\}$ part is a non-decreasing function which is equal to constant υ_{0j} for large enough y. Solving this for equality to 0 leads to no solution for y if $\upsilon_{0j} < 0$, to $y = \max_{i\in\mathcal{I}} \upsilon_{ij}/g_{ij}$ if $\upsilon_{0j} > 0$, and all $y \in [\max_{i\in\mathcal{I}} \upsilon_{ij}/g_{ij}, \infty)$ are solutions if $\upsilon_{0j} = 0$. Unfortunately, this would lead to division by 0 if $g_{ij} = 0$, however, all these terms can be omitted since its corresponding υ_{ij} equals υ_{0j}, and those are already included. By using this intermediate result equation (5.34) is easily solved. Additionally note that $y = 1$ for the trivial solution, thus the solution of problem (SSP) is

$$y = \min\left\{\min_{j\in\mathcal{J}} \max_{\{i\in\mathcal{I}|g_{ij}>0\}} \frac{\upsilon_{ij}}{g_{ij}}, 1\right\}, \tag{5.35}$$

from which $c = (1-y)/y$ and $d_{\langle i,j\rangle} = c/Q_i, \forall\langle i,j\rangle \in \mathcal{W}$ can be retrieved.

The complexity of this method is $O(|\mathcal{I}|\ln|\mathcal{I}| + |\mathcal{I}||\mathcal{J}|)$; first the inlinks have to be sorted, and then the additional constants (i.e., the gs and υs)[13], and the solution can be determined in linear time. This complexity means that the method is very efficient. Especially because the number of turns is generally low, no computational efficiency issues are expected for large scale road networks. Obviously, equation (5.35) is single valued and based on input constants, thus the solution is unique.

5.5.3.2. Equal Delay at Outlink

The equal delay at outlinks problem (EDP) is more complicated than the single server problem. Delays at every outlink influence the occupancy time at every inlink. Furthermore, multiple unknowns are involved and thus uniqueness is not trivial. A fixed point method that iteratively determines the headways at every turn and successively the delays at every outlink is presented. Given a set of delays $\left\{\tilde{d}_j, j \in \mathcal{J}\right\}$, equations (5.8) and (5.14) provide that the headways as a function of the delays are

$$h_{\langle i,j\rangle}\left(\left\{\tilde{d}_j, j \in \mathcal{J}\right\}\right) = \max\left\{\frac{1}{\alpha_{\langle i,j\rangle} Q_i} + \frac{\sum_{\langle i,j'\rangle\in\mathcal{W}_i} \alpha_{\langle i,j'\rangle}\tilde{d}_{j'}}{\alpha_{\langle i,j\rangle}}, \frac{1}{\alpha_{\langle i,j\rangle} S_i}\right\}, \quad \forall\langle i,j\rangle \in \mathcal{W}^+. \tag{5.36}$$

However, these headways might violate the supply constraints. Therefore, the delay at each outlink is updated based on current headways. These current headways at outlinks consist of the inter-arrival time of a vehicle from any inlink to that outlink plus the current delay for that

[12] The difference is that $-S_i/Q_i$ is always negative in this case, so the resulting canonical form is based on a minimization instead of a maximization.

[13] Note that these can be constructed incrementally, without evaluating the sum completely every time (see Cuninghame-Green, 1995).

outlink. Hence, for a given set of headways $\left\{h_{\langle i,j\rangle}, \langle i,j\rangle \in \mathcal{W}^+\right\}$ and 'old' delays $\left\{\tilde{d}'_j, j \in \mathcal{J}\right\}$, the inter-arrival time at outlink $j \in \mathcal{J}$ equals (using equation (5.6))

$$\frac{1}{\sum_{\langle i,j'\rangle \in \mathcal{W}_j^+} \frac{1}{h_{\langle i,j'\rangle}} - \tilde{d}'_j}, \quad \forall j \in \mathcal{J}. \tag{5.37}$$

The renewed delay $\tilde{d}_j$ has to be added to this term, and set such that the supply constraint (equation (5.13)) is satisfied. Thus, the delays as a function of headways and previous delays become

$$\tilde{d}_j\left(\left\{h_{\langle i,j\rangle}, \langle i,j\rangle \in \mathcal{W}^+\right\}, \left\{\tilde{d}'_j, j \in \mathcal{J}\right\}\right) = \max\left\{0, \tilde{d}'_j + \frac{1}{R_j} - \frac{1}{\sum_{\langle i,j'\rangle \in \mathcal{W}_j^+} \frac{1}{h_{\langle i,j'\rangle}}}\right\}. \tag{5.38}$$

Since these new delays do not take the changed delays at other outlinks into account, the inter-arrival time is not consistent anymore, and new headways have to be determined. Substituting equation (5.36) into equation (5.38) provides a fixed point problem in terms of delays at outlinks.

This fixed point problem can be solved by iterating over the delays, which may converge quickly (only a few steps) if there is only one supply constraint active, or slowly when multiple supply constraints are active. The delays at outlinks are not necessarily unique for symmetrical problems; however, it seems that the resulting reduced capacities in the solution are always unique. Proofs for convergence and uniqueness of reduced capacity are left for future research.[14]

5.5.3.3. Directed Capacity Proportional

Solution methods for the directed capacity proportional problem (DCPP) can be found in (Flötteröd and Rohde, 2011; Tampère et al., 2011). For the convenience of the reader, Algorithm 1 provides the method with the variables names used in this paper to retrieve the capacity to flow restriction factors (i.e. β_js). The equations in Section 5.4.3 can be used to convert the reduction factors to turn delays, reduced capacities, and resulting flows.

The complexity of the method is $O(|\mathcal{I}|^2|\mathcal{J}|)$; the outer-loop iterates at most $|\mathcal{I}|$ times and within this loop several loops over all turns are made. Tampère et al. (2011) and Flötteröd and Rohde (2011) prove that this algorithm converges to a unique point in terms of resulting flow.

5.5.3.4. Capacity Consumption Equivalence

A solution method for the capacity consumption equivalence problem (CCEP) can be found in Gibb (2011). Similar to the solution of the equal delay model, a fixed point approach is used. Given a set of capacity consumption factors $\left\{e'_j, j \in \mathcal{J}\right\}$, new capacity consumption

[14] Proof methods similar to those in Gibb (2011) are very likely to work as well for this model. Also, the low-dimensionality of Pareto optimal solutions (see Section 5.5.5) can be used to prove uniqueness.

Algorithm 1 Directed Capacity Proportional Model: Solution Method

while $\mathcal{I} \neq \emptyset$ **do**
    for all $j \in \mathcal{J}$ **do**
        $\beta_j \leftarrow \frac{R_j}{\sum_{i \in \mathcal{I}} \alpha_{\langle i,j \rangle} Q_i}$
        # Determine potential outlink restriction factors based on competing inlinks
    end for
    $j^* \leftarrow \arg\min_{j \in \mathcal{J}} \beta_j$
    $\hat{\mathcal{I}} \leftarrow \left\{ i \in \mathcal{I} \middle| \beta_{j^*} \alpha_{\langle i,j^* \rangle} Q_i > \alpha_{\langle i,j^* \rangle} S_i \right\}$
    # $\hat{\mathcal{I}}$ is the set of all inlinks that exceed demand if restricted by j^*
    if $\hat{\mathcal{I}} \neq \emptyset$ **then**
        for all $i \in \hat{\mathcal{I}}$ **do**
            for all $j \in \mathcal{J}$ **do**
                $R_j \leftarrow R_j - \alpha_{\langle i,j \rangle} S_i$ # Update supply j by subtracting the demand from i
            end for
            $\mathcal{I} \leftarrow \mathcal{I} \setminus \{i\}$
            # i will not compete anymore, it is guaranteed demand constrained
        end for
    else
        $\mathcal{I}^* \leftarrow \left\{ i \in \mathcal{I} \middle| \alpha_{\langle i,j^* \rangle} > 0 \right\}$
        # $\mathcal{I}^*$ is the set of all inlinks that compete for j^*, and will be completely determined by j^*
        for all $i \in \mathcal{I}^*$ **do**
            for all $j \in \mathcal{J}$ **do**
                $R_j \leftarrow R_j - \beta_{j^*} \alpha_{\langle i,j^* \rangle} Q_i$
                # Update supply at other outlinks j by subtracting the final flow on turn $\langle i,j \rangle$.
            end for
            $\mathcal{I} \leftarrow \mathcal{I} \setminus \{i\}$
            # i will not be considered anymore, its flow is guaranteed determined by j^*
        end for
        $\mathcal{J} \leftarrow \mathcal{J} \setminus \{j^*\}$
        # j^* will not be considered anymore, and β_{j^*} is definitive
    end if
end while
return $\left\{ \beta_j \middle| j \in \mathcal{J} \right\}$

equivalence factors can be determined with

$$e_j\left(\left\{e'_j, j \in \mathcal{J}\right\}\right) = \max\left\{1, \frac{e'_j \sum_{i\in\mathcal{I}} \alpha_{\langle i,j\rangle} S_i \min\left\{1, \frac{Q_i}{\sum_{j'\in\mathcal{J}} e'_{j'} \alpha_{\langle i,j'\rangle} S_i}\right\}}{R_j}\right\}, \quad \forall j \in \mathcal{J}. \tag{5.39}$$

Solving this fixed point problem iteratively converges to a unique solution in terms of reduced capacities (for proofs, see Gibb, 2011). The capacity consumption equivalence factors can be easily converted to turn delays, reduced capacities, and resulting flows with the equations in Section 5.4.4.

5.5.4. Diverges and Merges

Node models for merges and diverges are easier to solve because they lack the interaction between multiple in- and outlinks. Daganzo (1995a) has solved the problem for merges and diverges. Recently, Jin (2010) has analysed merges based on the demand-supply framework, and Jin (2012b) has analysed merge-diverge networks further based on kinematic wave theory. The diverge is straightforward since the result is completely determined by a single supply constraint. Due to the conservation of turning fractions and the fact that only one inlink exists – thus there is no competition – there is only one reduced capacity that solves (NP) for diverges. All the models in this paper provide indeed the same result for diverges. On the other hand, competition is involved at merges, which requires additional priority information.

The priority parameters of Daganzo (1995a) describe how competing inlinks share the supply. Either none, all, or a strict subset of inlinks will be in spillback conditions[15], and those are the competing inlinks. The connection between the priority parameters and the framework of this paper is simple: the ratio between the priority parameters of inlinks equals the ratio of the reduced capacities of those inlinks.

Ni and Leonard II (2005) introduce *capacity-based weighted fair queuing* where the priority parameters equal the capacities of the inlinks; they verify this with empirical data. It can easily be seen that the ratio between two reduced capacities – in case of a merge – for problems (SSP), (DCPP), and (CCEP) equal the ratio of the capacities of the two inlinks. So these three models coincide with the model of Ni and Leonard II (2005) in case of a merge.

Problem (EDP) with the equal delays is different because the priority parameters depend on the severity of the bottleneck. The ratio between the reduced capacities of two inlinks at a merge, say q_1/q_2, equals

$$\frac{Q_1 + Q_1 Q_2 d}{Q_2 + Q_1 Q_2 d}, \tag{5.40}$$

with only one delay d for all turns. So if the turn delay is very low (i.e., $d \to 0$), then the solution coincides with that of Ni and Leonard II (2005), but if the delay increases the ratio tends towards 1. In the latter case all inlinks have equal priority.

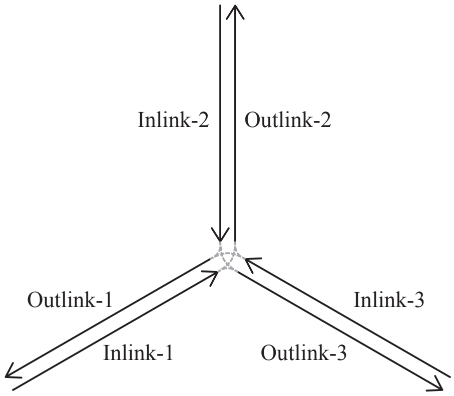


Figure 5.4.: Topology of the three-legged node

Table 5.2.: Example node input variables: demand, capacity and supply (all in vehicles per hour)

from \ to		Demand 1	2	3	all	Capacity
Demand	1	0	1200	600	1800	2000
	2	300	0	600	900	1000
	3	1200	600	0	1800	2000
	all	1500	1800	1200		
Capacity		1000	1000	1000		

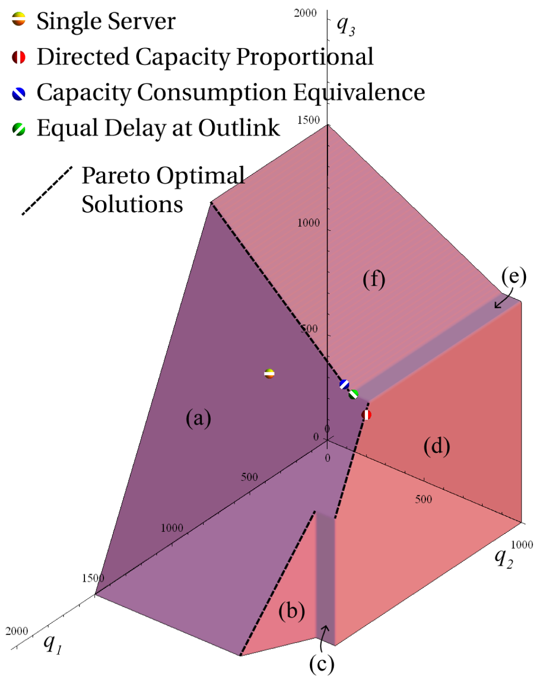


Figure 5.5.: Feasible reduced capacities, Pareto optimal solutions, and solutions of the known GCNM family members for the three-legged node example

5.5.5. Three-legged Node Example

To point out the differences between the models, a simple three-legged example is presented. Consider the node in Figure 5.4; there are three inlinks and three outlinks, and there are no U-turns, so in total six turns are present. Table 5.2 contains the capacity, demand and supply (equal to capacity in this case) for every link. The turn fractions can also be derived from the directional demand. For example, $\alpha_{\langle 2,3\rangle} = 2/3$, $S_3 = 1800$, $R_1 = 1000$. It is clear that there is a conflict at every outlink since the supply is insufficient to accommodate all demand, so spillback will occur at at least one inlink.

Figure 5.5 shows the boundary of the feasible solutions in terms of reduced capacities (i.e., $q_i, i \in \mathcal{I}$). Because merely the feasible reduced capacities are plotted it is possible to show it in a three dimensional space. All reduced capacities on the shown polyhedron are feasible, also all strictly positive reduced capacities within the area between the origin and the surface are feasible. However, the latter are obviously not Pareto optimal. More than that, not all points on the surface are Pareto optimal either. Only the points on the dashed lines are in fact Pareto optimal solutions in terms of reduced capacities; this will be explained in the next paragraph.

Each of the faces in Figure 5.5 represents a constraint, and are labelled with a letter. Face (a) represents the supply constraint at outlink-2; inlink-1 and -3 are actively constrained; however, any *interior point* of this face is not Pareto optimal since reduced capacity q_2 can be increased. Face (b) and (c) represents the supply constraint corresponding with outlink-3; inlink-1 is

[15]This relates one-to-one to Daganzo's causality regimes.

actively constrained at faces (b) and (c), while inlink-2 is only actively constrained at face (b). Here as well, any interior point of these faces is not Pareto optimal since reduced capacity q_1 and/or q_2 can be increased. Face (d) reflects $q_2 \leq 1000$ due to positiveness of turn delays; on its interior, q_1 and q_2 can be increased and those points are thus neither Pareto optimal. Finally, faces (e) and (f) represent the supply constraint at outlink-1, and analogously to faces (b) and (c) reduced capacities q_2 and/or q_3 can be increased. The points on the edge between faces (a) and (c) & (e) are neither Pareto optimal because q_2 can be increased.

While looking at the solutions of the GCNM family members in Figure 5.5, the single server solution is remarkable because it does not reflect a Pareto optimal solution. The reason for this is the added constraint that all turn delays are proportional to each other, and thus that each vehicle encounters a delay (also those heading to outlink-3). Therefore, the solution of the single server problem can be found so efficiently; it is the intersection between the surface and the vector with direction $(1/Q_i, i \in \mathcal{I})$.

The solutions for all models on the example are listed in Table 5.3. The resulting flows, as well as the turn delays and reduced capacities are given. The characteristics of every model as presented in Section 5.4 can be identified in the results:

- The *Single Server* result is completely determined by the supply at outlink-2. It has equal delays for turns originating at inlink-1 and -2, which have the same capacity. This is also the only model with positive turn delays towards outlinks that are not at capacity. Therefore, the total throughput is lowest, and the result is not a Pareto optimal solution of the original problem (NP).
- The *Equal Delay at Outlink* model has positive delays towards outlink-1 and -2 that are indeed equal for each of them. The behaviour at inlink-2 is also interesting because its capacity is reduced from 1000 to 913, but the demand is still decisive. Thus this is the only inlink in situation type B of Figure 5.2. This model also has the highest total throughput in this example.
- The *Directed Capacity Proportional* model is again completely determined by the supply outlink-2. Note that only one positive turn delay exists for every inlink. Turn $\langle 3,2 \rangle$ has the longest delay of all models for this example, which eventually also hampers vehicles at turn $\langle 3,1 \rangle$ due to spillback.
- The *Capacity Consumption Equivalence* model has positive delays towards outlink-1 and -2. It shows many similarities with the Equal Delay at Outlink model, but inlink-2 is also in spillback condition in this model. This is due to different capacity of inlink-2 and -3, which results in a larger delay at turn $\langle 2,1 \rangle$

Note as well that the results of the different models are close to each other; this is result of selecting an example that clearly depicts the characteristics of each model. In other examples, especially those in which small turn fractions exist, the results of the different models are significantly different.

5.6. Concluding Remarks

This chapter has presented the Generic Class of first-order Node Models (Tampère et al., 2011) as a family of models based on turn delays. A turn delay is the additional time a vehicle

Table 5.3.: Solutions of the three-legged node for the GCNM family members: flow, reduced capacities (both in vehicles per hour), and turn delays (in seconds per vehicle)

Single Server

from \ to		Flow 1	Flow 2	Flow 3	Flow all	Reduced capacity	from \ to		Turn delay 1	Turn delay 2	Turn delay 3
Flow	1	0	667	333	1000	1000	Turn delay	1		1.8	1.8
	2	167	0	333	500	500		2	3.6		3.6
	3	667	333	0	1000	1000		3	1.8	1.8	
	all	833	1000	667							

Equal Delay at Outlink

from \ to		Flow 1	Flow 2	Flow 3	Flow all	Reduced capacity	from \ to		Turn delay 1	Turn delay 2	Turn delay 3
Flow	1	0	650	325	925	925	Turn delay	1		2.84	0
	2	300	0	600	900	913		2	1.02		0
	3	700	350	0	1050	1050		3	1.02	2.84	
	all	1000	1000	925							

Directed Capacity Proportional

from \ to		Flow 1	Flow 2	Flow 3	Flow all	Reduced capacity	from \ to		Turn delay 1	Turn delay 2	Turn delay 3
Flow	1	0	667	333	1000	1000	Turn delay	1		2.7	0
	2	300	0	600	900	1000		2	0		0
	3	667	333	0	1000	1000		3	0	5.4	
	all	967	1000	933							

Capacity Consumption Equivalence

from \ to		Flow 1	Flow 2	Flow 3	Flow all	Reduced capacity	from \ to		Turn delay 1	Turn delay 2	Turn delay 3
Flow	1	0	643	322	965	965	Turn delay	1		2.90	0
	2	286	0	572	858	858		2	1.79		0
	3	714	357	0	1071	1071		3	0.90	2.90	
	all	1000	1000	894							

occupies an inlink when it heads for a certain direction, and is thus easily interpretable. The complete family is represented by multi-objective optimization problem (NP). Any model that finds a Pareto optimal solution of (NP) is a member of the family.[16] The new *single server* and *equal delay at outlink* models are presented as well as the existing *directed capacity proportional* (Flötteröd and Rohde, 2011; Tampère et al., 2011) and *capacity consumption equivalence* (Gibb, 2011) models. Their four problem formulations are respectively given in (SSP), (EDP), (DCPP), and (CCEP).

It is shown that solving these problems is not straightforward due to multiple objectives and unknowns. For (SSP) a very efficient solution method based on the theory of polynomials in the max-plus algebra is presented. The known method for (DCPP) is slightly less efficient. These are the only two models for which the exact solution can be found in finite time, and can thus be incorporated in large scale DNL models. On the other hand, they have reduced behavioural realism. The single server model assumes that all vehicles have to wait, also those heading to an empty outlink; this assumption is only realistic under very saturated conditions where the intersection is blocked. The directed capacity proportional model can lead to very high delays on turns with little demand.

The other two models have fixed point methods that can only determine the solution up to convergence. This is a major drawback when the models have to be applied repetitively in (large) DNL problems. Some convergence criterion (e.g., the flows of consecutive iterations differ less than one vehicle) should be set to terminate the iterations. However, their underlying behavioural assumptions are plausible.

Several sets of turn delays leading to the same resulting flows can be found. The notion of reduced capacity helps to identify whether two results are equivalent. It is shown that for diverges all family members are equivalent and that three out of four members yield identical results at merges. The relation between priority parameters at merges and the family is also presented. The additional relations between turn delays, required to select a Pareto optimal solution in terms of reduced capacities, imply a part of the behaviour at intersections. Especially at a supply constraint, the relation between competing turns should be determined in terms of delays.

5.7. Future Research and Empirical Validation

The turn delays are key in finding new models, which should be based on data analysis. By observing traffic at intersections under different spillback configurations, the capacity occupancy, turn delays, and – most important – the relations between turn delays can be retrieved. Because problem (NP) is underspecified, additional information about the behaviour of traffic at intersections is required, and it can only be included through data analysis. This not only applies to uncontrolled equal-priority intersections, but particularly to intersections with priority rules, traffic signals, roundabouts, and other infrastructures. These are actually the members of the family that are really required for good applications.

[16] Note that if additional constraints are added the solution is not necessarily Pareto optimal for (NP); however, it is Pareto optimal for the specific problem. (SSP) is an example of this.

Section 5.3.2 introduces how occupancy times and turn delays can be observed. The discrepancy between average headways used in the models and vehicle trajectories is not completely solved. What are the exact locations (i.e. where is the exit of an inlink exactly), where should headways should be measured, and how should multiple lanes be captured? The first should be defined such that when a vehicle clears the inlink, it does not hamper its follower that wants to take another turn. The latter can be solved by either modelling the different lanes as different links, or by an integrated approach.

Additional infrastructure, priorities, and/or conflict points are ideally captured by additional constraints on the turn delays. For additional infrastructure and conflict points these constraints are either likely to have the same form as the supply constraints (see equation (5.13)), or as direct functional relation between couples or groups of turn delays. Priorities are of a different nature and we expect that the delays at yielding turns can be expressed in the headways on the priority road. Such a relation could be found by using gap acceptance theories and conflict theory. Finally, the geometric design of a turn could cause the driver to increase its safety headway; this would induce a lower bound on the considered turn. Important for all of these additional constraints is that they cannot directly include the demand or the supply because otherwise an invariance principle can be violated. However, it is possible to incorporate the headways on inlinks, turns, and outlinks by using equation (5.8).

As we already pointed out in Section 5.5.2, Corthout et al. (2012) have shown that when turns of the same inlink have different priorities at different constraints non-uniqueness in flows can occur. This is a subsistent characteristic of traffic. Thus adding realism with constraints can lead to non-uniqueness. Therefore, in the search of new node models, a balance needs to found between two desired properties: realism and uniqueness.

Where data of course contributes in identifying realistic members of the family, also methodological advances in the field of max-plus polynomials will help. (SSP) is now solved by finding the root of a univariate max-plus polynomial. If the root of a multivariate min-max-plus polynomial could be found efficiently, the Pareto optimal solutions of (NP) in terms of reduced capacities can be found (i.e. solve equation (5.31) for equality)(see de Schutter and de Moor, 1996).

Acknowledgements

We would like to thank the reviewers for their sharp and useful comments. The study is performed as part of the Innovative Pricing for Sustainable Mobility (iPriSM) project in the Sustainable Accessibility of the Randstad (SAR) programme of The Netherlands Organisation for Scientific Research (NWO). The main theory in this paper is developed during a visit of the corresponding author to the Institute of Transport and Logistics Studies at The University of Sydney.

PART II.

STAKEHOLDERS & PRICING

CHAPTER 6.

GAME THEORETICAL APPROACH FOR MULTIPLE STAKEHOLDERS

This chapter introduces the game theoretical approach to solve the upper level problem of the transport pricing framework presented in Chapter 2. This approach with multiple stakeholders assumes that each stakeholder has control over some pricing mechanism. The analysis starts with non-cooperative behaviour, which then provides input for cooperative solution concepts. The theory of cooperative Transferable Utility (TU)-games provides a framework that allows to capture the preferences and interactions between stakeholders, but requires details about each coalition that can be formed between stakeholders. Therefore, for different coalition settings (non-cooperative) Nash-equilibria are defined and analysed. This analysis of competition provides all resources to define a cooperative TU-game. Multiple solution concepts of TU-games exist, and the three most important of them will be discussed, namely the core, the Shapley value and the compromise value. Besides the solutions based on cooperation between stakeholders from TU-games, also the non-cooperative solution rolls out, which is the equilibrium without any cooperation and without the formation of any coalition. The theory is illustrated with several examples, and characteristics of solution concepts are provided.

As discussed in Chapter 2.5, solving the Multiple Stakeholders Problem (MSP) is not straightforward due to the presence of (possibly opposing) multiple objectives. No approach exists that can solve it in a satisfying manner. Especially, little insight in negotiation processes is available. Therefore, this chapter starts with an overview of the available literature and subsequently introduces a novel approach to multiple stakeholders in transport pricing. Notice that the literature on transport pricing with multiple stakeholders varies in detail and methodology. It would be unfeasable to capture them in a unified framework.

The PhD theses of Joksimovic (2007) and Ohazulike (2014) present a game theoretical multi-level approach to road pricing. Joksimovic (2007) considers a single stakeholder and identifies the monopoly game, Stackelberg game and Cournot game. The first assumes the leader (i.e., the road authority) determines the responses of the travellers (e.g., their route choice). As such, this could reflect the best possible outcome fot the system – conform a system optimum –. This is unrealistic since travellers will act based on their own preferences, and they will respond to implemented pricing measures. Therefore, the monopoly game does not reflect a realistic situation. The Stackelberg game allows travellers to respond as followers to the prices imposed by the leader. In the Stackelberg setting both the road authority and travellers act based on their preferences. The road authority determines the prices, and the travellers respond by making their travel choices. Ohazulike (2014) includes multiple stakeholders by analysing Nash-equilibria between stakeholders. In a Nash-equilibrium no player can be better off by changing its own strategy (Nash et al., 1950; Nash, 1951); the user equilibrium in Traffic Assignment (TA) is also a Nash-equilibrium. He also adds another level on top of the stakeholders, the grand leader, that can induce a mechanism to the bi-level game to steer towards the optimal Nash-equilibrium. Such a situation can arise when local governments interact; the national government can than resemble the grand leader. However, it is not feasible that such a grand leader always exists in reality. Also, Levinson (1998) discusses multiple stakeholders in his dissertation. He analyses how jurisdictions choose to finance their roads. He considers decision making on two levels (i.e., to tax or to toll?, and what is the price?), and he makes notice of different outcomes under cooperation and non-cooperation. Of these three theses, only Ohazulike (2014) provides a solution to the MSP.

Levinson (2005) has a different game theoretical approach to pricing, where the travellers are the players. Two players can choose to depart early, on-time, or late, and based on these (pure[1]) strategies congestion can occur. The approach shows similarities with the well-known prisoners' dilemma. Also, Levinson (1999) provides a prisoner's dilemma analogy with tolling at frontiers. In a non-game-theoretical setting, Levinson and Chang (2003) provide a model for optimizing electronic toll collection systems.

Several economic studies analyse competition between multiple stakeholders. In general they provide analytical 'solutions' to the pricing question; however, their models have too many assumptions to serve as practical strategic planning models. The simple transport connections are not explicitly represented by the transport system and network. See de Borger and Proost (2012b); Ubbels and Verhoef (2008) for literature overviews of these economic approaches. An exception to these is the simulation study of Proost and Sen (2006) that considers a model of Brussels under different types of behaviour of regional and local governments. Nevertheless, the transport network is represented by a single link. The review papers do not include Zhang et al. (2011) who compare competitive, cooperative and Stackelberg congestion pricing for multiple regions on simple networks. Different from these approaches is the political economy model of de Borger and Proost (2012a) that considers the decisions of voters based on pricing schemes in a highly simplified manner.

All previous models either assume competition between stakeholders or a clear hierarchy of authority. Other forms of interaction (such as collaboration and coalition formation) have not been investigated. None of the models mentioned above use a rigorous mathematical framework based on game theory that includes coalition formation and interactions between stakeholders. Such an approach is provided in this chapter; it is based on the notion of Nash-

[1] Pure strategies involve deterministic choices, contrary to mixed strategies where stakeholders can make decisions with a certain probability.

equilibria for (groups of) stakeholders and the corresponding TU-games. It allows an (elaborate) underlying transport model with a user equilibrium that models the 'game' travellers.[2] Sets of stakeholders can form coalitions, which means that they optimize their combined objectives. This provides more insights in the stakes at the negotiation table of the stakeholders.

From an economics (first-best pricing) perspective, an optimal transport pricing strategy would only show up if all individuals join the negotiation table, and if unlimited monetary transfers between them would be possible. With that assumption it is sufficient to consider one stakeholder with one objective function (i.e., social welfare). However, reality is more complex. Not every individual participates at the negotiations and governments generally represent the interests of large groups of individuals. In turn, governments make their decisions also based on politics, rather than only social welfare arguments (see Section 2.2.1). In this, it is not true that if all stakeholders cooperate, one ends up in a system optimum – in which social welfare is optimal –. In addition, social welfare is the objective sum of all individual interests, that does not imply that at the system optimum, individuals (or groups of individuals, or stakeholders) experience their own optimum. Welfare can be very unevenly distributed over individuals in the optimal situation.[3] Since no 'grand leader' exists that can rule and divide welfare, but rather multiple governments and organisations with power exist – that represent individuals and each optimize their own objectives –, the eventual 'system' is the result of negotiations between them. Hereby, not all individuals are equally represented at the negotiation table, which means that the optimum from an economics perspective will not be realized.

Therefore, the approach presented in this chapter aims to provide insight in what may happen at the negotiation table. What is the effective power of each stakeholder? Which arguments can be put down? What can be achieved by cooperating? How to form coalitions, and what is the value of coalitions? This multi-stakeholder framework delivers knowledge about the effectiveness and feasibility of pricing measures, compared to when only a single welfare optimizing stakeholder is considered. In other words: the approach researches a 'second-best' solution in which the 'first-best' assumption that one actor with one objective – from the perspective that every individual is equally represented at the negotiation table – is released, and in which the negotiation table is interpreted with all its participating stakeholders.

6.1. Problem Formulation

Chapter 2 has introduced the MSP (see page 30); this problem formulation will be analysed and addressed in this chapter.[4] It is the starting point of the game theoretical analysis of the interactions between stakeholders. That general problem formulation does not define which stakeholder can set which pricing measure. In the following it is assumed that each stakeholder has full control over its own, and only its own, pricing measure(s). This assumption implies

[2] Travellers' behaviour is not discussed in this chapter, but it is taken implicitly into account in the TA. The resulting equilibrium is considered as a constraint here.

[3] Welfare may also include equity concerns where the interest of some target groups is weighed higher than that of others, or where measures such as the Gini-index are included. However, that objective is then still pursued by, e.g., a government that represents individuals.

[4] The notation of Chapter 2 is adopted in this chapter.

that each measure has one rationale (i.e., underlying goal), and it is captured with this slightly more specific problem formulation,

$$\left\{ \max_{\pi_{p_s} \in \Pi_{p_s}} H_s(\boldsymbol{\gamma}) \middle| s \in \mathcal{S} \right\} \\ \text{subject to } \boldsymbol{\gamma} = E(\boldsymbol{\pi}), \tag{6.1}$$

where p_s represents the pricing measures controlled by stakeholder s, with corresponding feasible price levels Π_{p_s}.[5] Pricing measures per stakeholders are now easily distinguished. With this formulation, the market power of each stakeholder is completely determined by the pricing measures that it controls. For example, if a government introduces a kilometre charge, other stakeholders can only influence the government by setting its own price.

In the problem formulation (Equation 6.1), several optimizations exist in parallel with one shared constraint. The primary solution of problem (6.1) is the resulting pricing scheme $\boldsymbol{\pi}^* \in \Pi$ which is the set of resulting prices per stakeholder $\left\{\pi_p^* \middle| p \in \mathcal{P}\right\}$. The secondary results are the achievements of each stakeholder that primarily consists of its objective function, but which may also contain other components (e.g., monetary transaction) as the result of negotiations with other stakeholders.

Note that the word 'optimal' is deliberately avoided because optimality in multi-objective optimization is usually associated with Pareto-optimal solutions. A solution is Pareto-optimal if no objective can be improved without degrading any other objective. Many Pareto-optimal solutions can exist, and without any preference information, it is difficult to choose between them. Although it is interesting to identify the Pareto-optimal solutions, also called the Pareto-frontier, it might not provide sufficient insight in the underlying mechanisms.

6.1.1. Assumptions

The earlier mentioned assumption that each stakeholder has full control over its own, and only its own, pricing measure(s), is an important difference with the MSP of Chapter 2. Although, it does not explicitly make a hierarchy or ranking between stakeholders' market power, the 'power' of the measures implicitly reflects the power of each stakeholder. Each stakeholder's power is defined by the difference it can make with the pricing measures he controls. In addition to this assumption, the approach that will be presented in this chapter makes the following assumptions:

- Each stakeholder has perfect information over the other stakeholders' objectives.
- Each stakeholder has perfect information over the outcome of the mapping from price levels to effect levels (i.e., the outcome of the underlying transport model).
- Stakeholders have pure strategies (rather than mixed stategies).
- Each stakeholder has a 'fixed' objective function. I.e., the stakeholder arrive at the negotiation table with predefined objective functions. They do not adjust their objective functions for strategic purposes. The latter would add another strategic layer to the game.

[5] To be complete, assure that every pricing measure is controlled by exactly one stakeholder, i.e., $\cup_{s \in \mathcal{S}} p_s = \mathcal{P}$, and $p_{s_1} \cap p_{s_2} = \emptyset, \forall s_1 \neq s_2 \in \mathcal{S}$

- There are no costs or gains involved when stakeholders enter or leave coalitions.

6.2. Approach Towards TU-games

This section analyses the interaction between stakeholders, and introduces a novel connection between cooperative and non-cooperative game theory. Nash equilibria and cooperative Transferable Utility (TU) games are the main ingredients. Branzei et al. (2008); Peters (2008) provide extensive introductions to cooperative game theory. As its name suggest the theory assumes that stakeholders cooperate when they come to a solution; however, they still pursue their own, and only their own, objective. To get insight in the interaction between stakeholders and in the negotiation positions, TU-game theory analyses coalitions of stakeholders. Define a coalition as a set of stakeholders that cooperate. Cooperating stakeholders decide together the price levels of all their available pricing measures, and they are willing to compensate each other (by monetary transfers). The *grand coalition* is the coalition in which all stakeholders reside, so the grand coalition equals the set of stakeholders $\mathcal{S}$.[6] In total $2^{|\mathcal{S}|}$ coalitions can be formed.[7] For a TU-game a certain value is assigned to each coalition $C \subseteq \mathcal{S}$, this *coalition value* is defined with function $v(C) : 2^{\mathcal{S}} \rightarrow \mathbb{R}$.[8] The analysis considers the residency of stakeholders in multiple coalitions. The coalition value represents the combined surplus value of the stakeholders in the coalition. The values reflect the negotiation positions of stakeholders in the coalition. The pair $(\mathcal{S}, v)$ formally defines a TU-game.

A standard example of coalition values are two merchants that each have a batch of respectively left-handed and right-handed gloves. A 'coalition' with one merchant will have zero value since he cannot sell anything. However, the coalition of both merchants will be equal to market value of all pairs of gloves that can now be sold. Determining the value of coalitions in the context of transport pricing and the MSP is much less trivial. To achieve this Nash-equilibria between stakeholders have to be analysed. Also, the behaviour of stakeholders outside the coalition (and the behaviour of the travellers) has to be considered. This section first works towards a coalition value for each stakeholder and then discusses how these can be used with TU-game theory to find solutions.

6.2.1. The Value of a Coalition

To the author's knowledge, no method exists that can translate the MSP into coalition values. In this section such a method is provided. It is natural to base the value of a coalition on the best achievable objective function values of the coalition's stakeholders. This raises two questions that should be answered. Firstly, "what is best achievable for a coalition?", and secondly, "which pricing scheme is in place?", because the latter completely determines the objective function values. In addition, each stakeholder – in- and outside the coalition – will employ its power (to set the price of its own pricing measure(s)) in favour of its own objective.

[6] In game theory the grand coalition is usually denoted with N; in this thesis the names *grand coalition* and *set of stakeholders* can be used interchangeably and are both denoted with $\mathcal{S}$

[7] This includes the empty coalition $\emptyset$ that has a formal role in TU-game theory.

[8] $2^{\mathcal{S}}$ is the power set of $\mathcal{S}$ and contains all possible coalitions.

This can lead to the situation where some seemingly achievable objective values for a coalition are infeasible in reality, because one or more stakeholders outside the coalition influence the outcome by changing their price. The following definition of a coalition value considers such scenarios.

Remember that all stakeholder's objective functions have the same unit (e.g., a monetary unit), and that objective values can thus be compared and added. The result of a coalition given some pricing scheme is the sum of the corresponding objective function values, i.e. in shorthand $\sum_{s \in C} H_s$. This sum differs with each pricing scheme and, more specifically, with the price level of *each* pricing measure. To assign a value to the coalition one needs to find a result that corresponds to a pricing scheme that satisfies a stakeholders equilibrium condition. This is defined in such a manner that it answers the question which pricing schemes *could* be in place. The corresponding equilibrium condition for this question states that no stakeholder, in- or outside the coalition, wants to change its price and thereby the coalition's result. Above all, note that the stakeholders outside the considered coalition can also form coalitions; therefore, also all possible coalition formations have to be considered. To deal with this complexity, two definitions will be introduced that together provide the formal equilibrium. The first definition considers the stability of a single coalition, and the second definition incorporates the coalition forming process. The former is:

Definition 6.1 (Pricing scheme stability for coalition)**.** *A pricing scheme $\boldsymbol{\pi}$ is stable for coalition C if the combined objective function of stakeholders in C, i.e., $\sum_{s \in C} H_s(\boldsymbol{\gamma})$, where $\boldsymbol{\gamma} = E(\boldsymbol{\pi})$, cannot be increased by changing the price level π_p of pricing measures $p \in \cup_{s \in C} p_s$ controlled by a stakeholder in coalition C.*

A stable pricing scheme for a coalition could be seen as a conditioned optimum, with the condition that stakeholders outside the coalition do not act. However, to reach an equilibrium the stakeholders outside the coalition should also be considered, since they can influence the coalitions outcome as is described before. And since those stakeholders can also form one or more coalitions which can change their preferences, it is important to consider all possibilities of coalition formation. Therefore denote *partition* Q as a mathematical partition (i.e., a set of coalitions) of the grand coalition $\mathcal{S}$; a partition divides the stakeholders into coalitions, and satisfies

$$\cup_{C \in Q} = \mathcal{S} \tag{6.2}$$

$$C \neq \emptyset, \forall C \in Q \tag{6.3}$$

$$C_1 \cap C_2 = \emptyset, \forall C_1, C_2 \in Q \text{ s.t. } C_1 \neq C_2. \tag{6.4}$$

For a grand coalition with three stakeholders, say {A,B,C}, the following five partitions are possible: {{A},{B},{C}}, {{A,B},{C}}, {{A,C},{B}}, {{A},{B,C}}, and {{A,B,C}}. The following definition of the *Nash-equilibrium* ensures that, given a partition, no coalition wants to adjust the price level of any of its involved stakeholders.

Definition 6.2 (Nash-equilibrium for partitions)**.** *A pricing scheme $\boldsymbol{\pi}$ is a Nash-equilibrium for the partition Q when $\boldsymbol{\pi}$ is stable for all coalitions $C \in Q$.*

Note that multiple Nash-equilibria can exist for the MSP, and that this multiplicity has two bases. First, multiple pricing schemes can be in equilibrium for a given partition. Second, multiple partitions are possible for which equilibria can exist. To define the value of a coalition, all Nash-equilibria – within and between[9] partitions – have to be considered. It is logical to choose the worst-case equilibrium (i.e., the one with the lowest value for the coalition) within the partition, since the other competing coalitions can steer towards it by adjusting their pricing levels. Also between partitions the worst-case equilibrium is chosen, because coalition forming process of the other stakeholders is not controlled. This leads to the following definition of the value of a coalition:

Definition 6.3 (Coalition value)**.** *The value of non-empty coalition $C \subseteq \mathcal{S}$ is*

$$v(C) = \min_{\{Q|C\in Q\wedge Q \text{ is a partition of } \mathcal{S}\}} \quad \min_{\{\boldsymbol{\pi}\in\Pi|\boldsymbol{\pi} \text{ is a Nash-equilibrium for } Q\}} \sum_{s\in C} H_s(\boldsymbol{\gamma}) \qquad (6.5)$$

such that $\boldsymbol{\gamma} = E(\boldsymbol{\pi})$

Note that the value of the grand coalition, $v(\mathcal{S})$, equals the *maximum* over all pricing schemes of the sum of all stakeholders' objective functions.[10] Two notes have to be made. The first regards the well-definedness of equation (6.5) since at least one Nash-equilibrium has to exist for a partition that includes C.[11] If no Nash-equilibria exist for any of the partitions where the coalition is part of, assume that the value of this coalition equals zero.[12] Secondly, Definition 6.3 does not address the value of the 'empty coalition' since that never appears in a partition; therefore, it is defined separately:

Definition 6.4 (Empty coalition value)**.** *The value of the empty coalition ($C = \emptyset$) is*

$$v(\emptyset) = 0. \qquad (6.6)$$

The Nash-equilibria and the coalition value are a first step to solve the MSP, and problem (6.1) in particular. The concept of the Nash-equilibrium provides the preconditions for a

[9]Note that coalitions can appear in multiple partitions, in the previous example coalition {A} appears in two partitions.

[10]This is because the grand coalition appears in only one partition, without any stakeholder outside the coalition.

[11]This existence will be assumed in the remainder of this chapter. Its detailed analysis is out of the scope of this thesis, but here some leads are provided. The existence cannot easily be proved. The pricing measures and underlying transport model can be so complex that it is likely that carefully constructed counter examples are possible if no further assumptions are made (e.g., see example 'Three conflicting players' later in this chapter). Game theory provides some insight in existence in more controlled situations. If the price levels are discrete, existence cannot be proved; however, if mixed strategies are allowed (i.e., stakeholders can set prices with a certain probability) then Nash's theorem can be applied to show existence. In the case of continuous price levels, existence can be shown under a quasi-concavity condition. This condition can be released again when mixed (continuous) strategies are allowed (by using Glicksberg's theorem)(Sion and Wolfe, 1957).

[12]One could allow for mixed strategies to avoid this situation; however, the computation of the equilibria will become complex and the interpretation of stable pricing schemes becomes more abstract. Moreover, for the transport pricing analysis performed in Chapter 7 at least one Nash-equilibrium exists per coalition.

pricing scheme to be a feasible outcome. While in addition, the coalition value starts to include competition between coalitions, and provides the negotiation position of each coalition. The last step is to resolve these negotiations, and to provide the (or a) cooperative solution to problem (6.1) by using TU-game theory solution concepts. Note that the Nash-equilibria can also be used to analyse the non-cooperative solutions.

6.3. Solution Concepts

The coalition values from Definitions 6.3 and 6.4 determine a complete TU-game $(\mathcal{S}, v)$; they are the building blocks of general TU-games for which a wide variety of solution concepts exist, and which are extensively discussed in the literature. Although that the Nash-equilibria that lead to the coalition values are an indication of what realistic pricing schemes are, they do not directly allow for an unique ultimate solution of the TU-game. The different solution concepts lead to multiple possibilities. Such a solution consists of the allocations χ_s for all stakeholders $s \in \mathcal{S}$, and is called the allocation vector $\boldsymbol{\chi} \in \mathbb{R}_+^{|\mathcal{S}|}$ which equals $\{\chi_s | s \in \mathcal{S}\}$. Another frequently used name for allocation vector is payoff vector. Since multiple solution concepts are available, no unique and nor an optimal allocation vector exists. It is not even guaranteed that an allocation vector exists that all stakeholders will accept. On the other hand, the 'played' or resulting pricing scheme is the pricing scheme $\boldsymbol{\pi}^*$ that corresponds to the value of the grand coalition $v(\mathcal{S})$, because the stakeholders cooperate and π_s^* leads to largest aggregate objective. The difference between each stakeholder's allocation χ_s and each stakeholder's individual objective H_s under $\boldsymbol{\pi}^*$ will be settled with monetary transactions.

So the negotiation table in TU-games is all about monetary transfers between stakeholders. After all, the played pricing scheme is the best scheme for the grand coalition. The discussion is therefore on the allocation for each stakeholder, or in other words, on how to share the cake. Solution concepts for TU-games seek for allocation vectors that are accepted by all stakeholders and all coalitions of stakeholders, and they often include a fairness argument. For some TU-games no allocation vector exists that is not objected to by one or more stakeholders or coalitions. Those situations bring the negotiations to a deadlock.

This chapter discusses three TU-game solution concepts and thereby provides insight in solutions for the MSP. The *core*, the *Shapley value*, and the *compromise value* will be discussed. The core stands out from the other two since it describes a set of the allocation vectors that satisfy several basic properties (and it does not choose a single allocation from this set), as will be described in the next paragraph. The other two determine specific allocation vectors. Another solution concept for TU-games is the *nucleolus*. However, its calculation is cumbersome (Guajardo and Jornsten, 2014) and its interpretation is abstract. Therefore, this concept is considered not applicable to the MSP.

6.3.1. Core

The core, introduced by Gillies (1959) in its modern form, defines allocations that are feasible outcomes of the negotiation. This means that the stakeholders are efficient (the payoff of the grand coalition, $v(\mathcal{S})$, is allocated) and no coalition wants to deviate from the grand coalitions.

Coalitions will deviate from the grand coalition if their total allocation is less than its coalition value. So, the core $K(\mathcal{S},v)$ of TU-game $(\mathcal{S},v)$ is defined by

$$K(\mathcal{S},v) = \left\{ \chi \in \mathbb{R}_+^{|\mathcal{S}|} \middle| \sum_{s \in \mathcal{S}} \chi_s = v(\mathcal{S}) \wedge \sum_{s \in C} \chi_s \geq v(C), \forall C \subset \mathcal{S} \right\}. \tag{6.7}$$

The core is stable regarding coalition dissidence; no coalition has a rightful argument to object towards any of the allocations in the core because their allocated value is at least as high as its coalition value. This also means that none of the stakeholders wants to deviate from the grand coalition (since this also holds for single stakeholder coalitions).

6.3.2. Shapley value

The Shapley value derives the allocated value for each stakeholder on what it can contribute to existing coalitions. It is a concept first presented by Shapley (1952) and it leads to a single solution. Each stakeholder receives the average of its marginal contributions over all different orderings in which the stakeholders can enter the grand coalition. The marginal contribution is the additional value a stakeholder brings to the table when it is added to the coalition, i.e., $v(C) - v(C \setminus \{s\})$ is the marginal contribution of stakeholder s to coalition C. To calculate the Shapley value marginal values for each stakeholder for each permutation[13] of the grand coalition are determined. The Shapley value is the average over all $|\mathcal{S}|!$ possible permutations. The Shapley value is unique.

For the transport pricing problem MSP, the marginal contribution of a stakeholder to a coalition is not directly intuitive. The coalition values arise from Nash-equilibria in which per definition all stakeholders are involved. Comparing the definition of the marginal contribution with Definition 6.5, reveals that the marginal value of a stakeholder will depend on the magnitude of its objective function (since this enters into the total). In addition, one can see that its marginal contribution also lies in the external influence the stakeholder can exert on the coalition it is about to enter. When that happens the potential partition that forced the coalition to minimal objective value(s) is cancelled.

6.3.3. Compromise value

The compromise value, introduce by Tijs (1981), is an allocation in which all stakeholders perform equally 'well' according to their own frames of reference – i.e., all stakeholder hold the same position compared to their best and their worst achievable result. To achieve this, for every stakeholder an upper and lower bound for the allocation is determined that can reasonably be negotiated.

The upper bound Ω_s for stakeholder s is the marginal contribution of the stakeholder when it is added as the final stakeholder to the grand coalition, i.e.,

$$\Omega_s = v(\mathcal{S}) - v(\mathcal{S} \setminus \{s\}). \tag{6.8}$$

[13] Permutations are unique orderings in which stakeholders enter one by one to a coalition.

This is the best possible result because when the stakeholder claims more, coalition $\mathcal{S} \setminus \{s\}$ will separate from the stakeholder and will leave him behind with less than he claims. Denote the vector of upper bounds as $\boldsymbol{\Omega} = \{\Omega_s | s \in \mathcal{S}\}$.

For the lower bound ω_s of stakeholder s consider the remaining value for a stakeholder in all possible coalitions and when all competitors took as much as possible – their upper bound. The maximum over the remaining value per coalition is the lower bound. That is

$$\omega_s = \max_{\{C|s\in C\}} \left\{ v(C) - \sum_{\{s' \in C, s' \neq s\}} \Omega_{s'} \right\}. \tag{6.9}$$

For intuition, imagine that a stakeholder chooses with which other stakeholders he starts negotiating. They consider how much they can divide, and all other stakeholders get their maximum (based on their upper bound) and the stakeholder himself claims the remainder. This remainder is always achievable since the other parties are 100% satisfied. Denote the vector of lower bounds as $\boldsymbol{\omega} = \{\omega_s | s \in \mathcal{S}\}$.

Finally, the compromise value $\boldsymbol{\eta}(\mathcal{S}, v)$ of TU-game $(\mathcal{S}, v)$ is the unique and equal convex combination of upper and lower bounds for all stakeholders such that the allocation is efficient, i.e.,

$$\begin{aligned} \boldsymbol{\eta}(\mathcal{S}, v) &= \alpha \boldsymbol{\Omega} + (1-\alpha)\boldsymbol{\omega} \\ \text{with } \alpha &= \frac{v(\mathcal{S}) - \sum_{s\in\mathcal{S}} \omega_s}{\sum_{s\in\mathcal{S}} \Omega_s - \sum_{s\in\mathcal{S}} \omega_s}. \end{aligned} \tag{6.10}$$

For example, when stakeholder A and stakeholder B have bounds of respectively [0,10] and [10,20], and the total budget is 14, then both stakeholders will end up at 20% of their reach: A receives 2 and B receives 12. The compromise value is unique. Note that the compromise value only exists for *compromise admissible* TU-games.[14] The from the MSP derived games in this thesis are all compromise admissible. In the literature, the compromise value is also known as the τ-value (Tijs, 1981; Tijs and Otten, 1993).

6.4. Properties of Solution Concepts

Each solution concept satisfies several properties that are relevant for analysts, policy makers, and other users of this method. The purpose of this section is to provide a reference for the properties. For details on the analysis of the several solution concepts, and for additional solution concepts, we refer to Peters (2008) for the core and Shapley value and Tijs and Otten (1993) for the compromise value. In the next section multiple illustrative examples are provided that should provide more insight in the working and characteristics of the solution concepts.

[14] A game is compromise admissible if for each stakeholder the upper bound is larger than or equal to the lower bound ($\omega_s, \forall s \in \mathcal{S}$), and if the value of the grand coalition lies between the sum of lower bounds and sum of upper bounds ($\alpha \in [0,1]$ in Equation (6.10)).

6.4.1. Core

The core's important properties can be derived directly from its definition. Let $(\mathcal{S}, v)$ be a TU-game. Its core $K(\mathcal{S}, v)$ has the following properties:

- **Efficiency:** The payoff of the grand coalition is allocated, i.e., $\sum_{s\in\mathcal{S}} \chi_s = v(\mathcal{S}), \forall \chi \in K(\mathcal{S}, v)$.
- **Coalition rationality:** The total allocation of each coalition is at least as much the coalition value:
$$\sum_{s\in C} \chi_s \geq v(C), \quad \forall C \subseteq \mathcal{S}, \quad \forall \chi \in K(\mathcal{S}, v). \tag{6.11}$$
- The core can be empty.

6.4.2. Shapley value

The Shapley value and its correponding allocation $\zeta(\mathcal{S}, v) = \chi$ have the following properties:

- **Efficiency:** The payoff of the grand coalition is allocated, i.e., $\sum_{s\in\mathcal{S}} \chi_s = v(\mathcal{S})$.
- **Symmetry:** The stakeholders of whose marginal contributions to each coalition is equal – these stakeholders are equivalent – have an equal allocation.
- **Dummy player:** For all stakeholders $s \in \mathcal{S}$ that have marginal contributions equal to $v(\{s\})$ to all coalitions (i.e., $v(C \cup \{s\}) = v(C) + v(\{s\}), \forall C \subseteq \mathcal{S} \setminus \{s\}$), it holds that $\zeta_s(\mathcal{S}, v) = v(\{s\})$.
- **Linearity:** This mathematical property says that if a linear combination of games is studied, the Shapley value can be retrieved directly from the linear combination. Formally, let $(\mathcal{S}, v)$, $(\mathcal{S}, v')$ be TU-games and let $a, b \in \mathbb{R}$, then
$$\zeta(\mathcal{S}, av + bv') = a\zeta(\mathcal{S}, v) + b\zeta(\mathcal{S}, v'). \tag{6.12}$$
- **Zero player:** A player s that adds no value to any coalition (i.e., all its marginal contributions are zero), then the allocated value of the Shapley value is also zero ($\chi_s = 0$).
- **Existence:** The Shapley value can always be calculated and always exists.
- The Shapley value is not necessarily part of the core. Possibly, coalition rationality is not satisfied

6.4.3. Compromise value

The compromise value and its correponding allocation $\eta(\mathcal{S}, v) = \chi$ has the following properties (Tijs and Otten, 1993):

- **Efficiency:** The payoff of the grand coalition is allocated, i.e., $\sum_{s\in\mathcal{S}} \chi_s = v(\mathcal{S})$.
- **Individual rationality:** The allocation of each stakeholder is at least as much as its single-stakeholder coalition value:
$$\chi_s \geq v(s), \quad \forall s \in \mathcal{S}. \tag{6.13}$$
- **Symmetry:** The stakeholders of whose marginal contributions to each coalition is equal – these stakeholders are equivalent – have an equal allocation.

- **Dummy player:** For all 'dummy' stakeholders $s \in \mathcal{S}$ that have marginal contributions equal to $v(\{s\})$ to all coalitions (i.e., $v(C \cup \{s\}) = v(C) + v(\{s\}), \forall C \subseteq \mathcal{S} \setminus \{s\}$), it holds that $\boldsymbol{\eta}_s(\mathcal{S}, v) = v(\{s\})$.
- **Covariance:** This mathematical property says that if a game is multiplied with a positive constant, and a vector is added, then the compromise value can be retrieved directly from the combination. Formally, let $a \in \mathbb{R}$ and $b \in \mathbb{R}^{|\mathcal{S}|}$, then

$$\boldsymbol{\eta}(\mathcal{S}, av + b) = a\boldsymbol{\eta}(\mathcal{S}, v) + b. \tag{6.14}$$

- The compromise value only exists for compromise admissible TU-games (see Footnote 14 on page 150).

More properties (dummy out, complementary monotonicity, restricted proportionality, and minimal right) of the compromise value are presented by Tijs and Otten (1993).

6.5. Illustrative Examples

In this section the solution concepts are illustrated with some example MSP instances. Continuous as well as discrete price levels are covered. The general approach is to find all stable pricing measures for each coalition, and then to analyse the possible partitions with Nash-equilibria. From these the coalition values can be derived. Finally, the solution concepts can be computed. For explanation purposes the objective values are directly written in terms of price levels, i.e., the underlying transport model is very simple and the intermediate effect levels are not named.

The purpose of the first example is to show how coalition values can be determined on a setting with continuous price levels, and simple objective functions. Only two stakeholders exist, so one can easily handle coalition formation. Using mathematical analysis the exact value of each coalition can derived.

Example: Two players with continuous price levels

Consider grand coalition $\mathcal{S} = \{\text{A}, \text{B}\}$, with $\Pi_\text{A} = \Pi_\text{B} = [0,1]$,

$$H_\text{A}(\pi_\text{A}, \pi_\text{B}) = -(\pi_\text{A} - 0.25)^2 - \frac{(\pi_\text{B} - 0.25)^2}{2}, \text{ and} \tag{6.15}$$

$$H_\text{B}(\pi_\text{A}, \pi_\text{B}) = -\frac{(\pi_\text{A} - 0.75)^2}{4} - (\pi_\text{B} - 0.75)^2. \tag{6.16}$$

First, analyse the stable pricing schemes for each coalition.

For coalition $\{\text{A}\}$ the stable pricing schemes can be found with $\partial H_\text{A} / \partial \pi_\text{A} = 0$ (since H_A is continuous and concave). This implies that stakeholder A always plays $\pi_\text{A} = 0.25$, regardless the action of stakeholder B. Note that the absolute optimum for A lies at $(\pi_\text{A}, \pi_\text{B}) = (0.25, 0.25)$. This is illustrated in Figure 6.1 below.

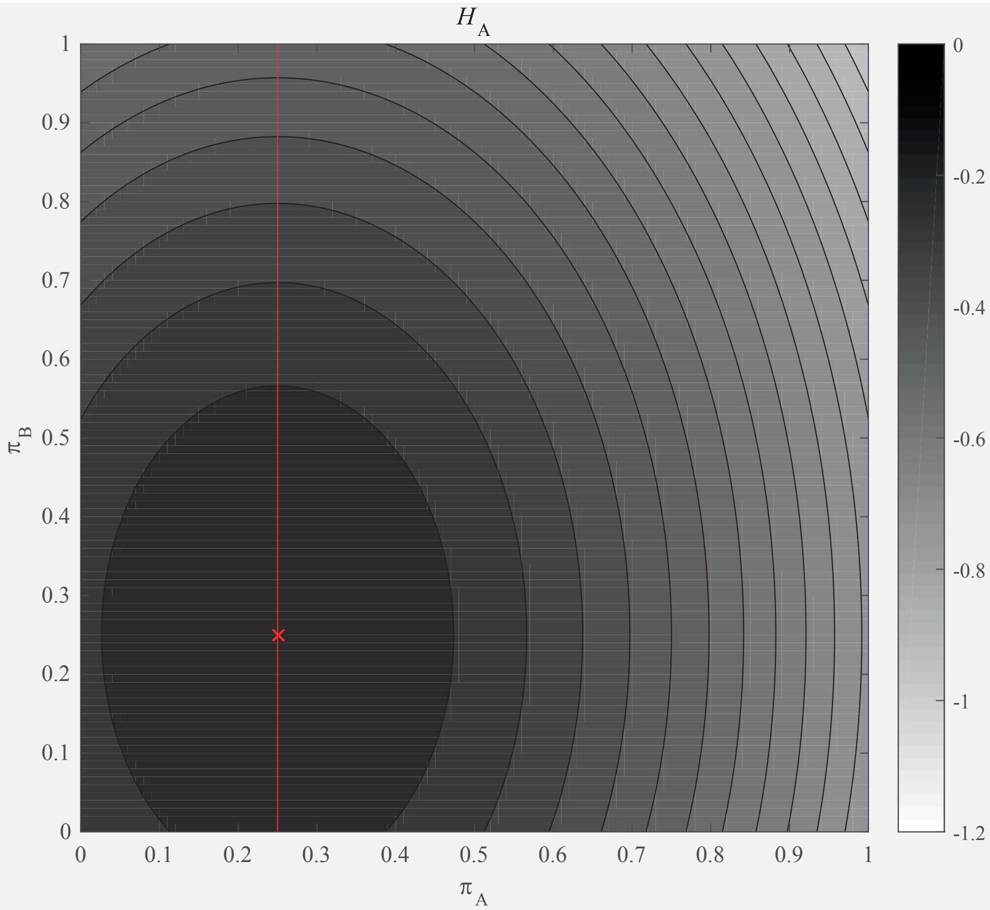


Figure 6.1.: Objective function contour plot for the coalition with stable pricing schemes (red line) and optimal point (red cross).

For coalition {B} a similar analysis can be made leading to its stable price level $\pi_B = 0.75$ and optimal point $(\pi_A, \pi_B) = (0.75, 0.75)$. See Figure 6.2 below.

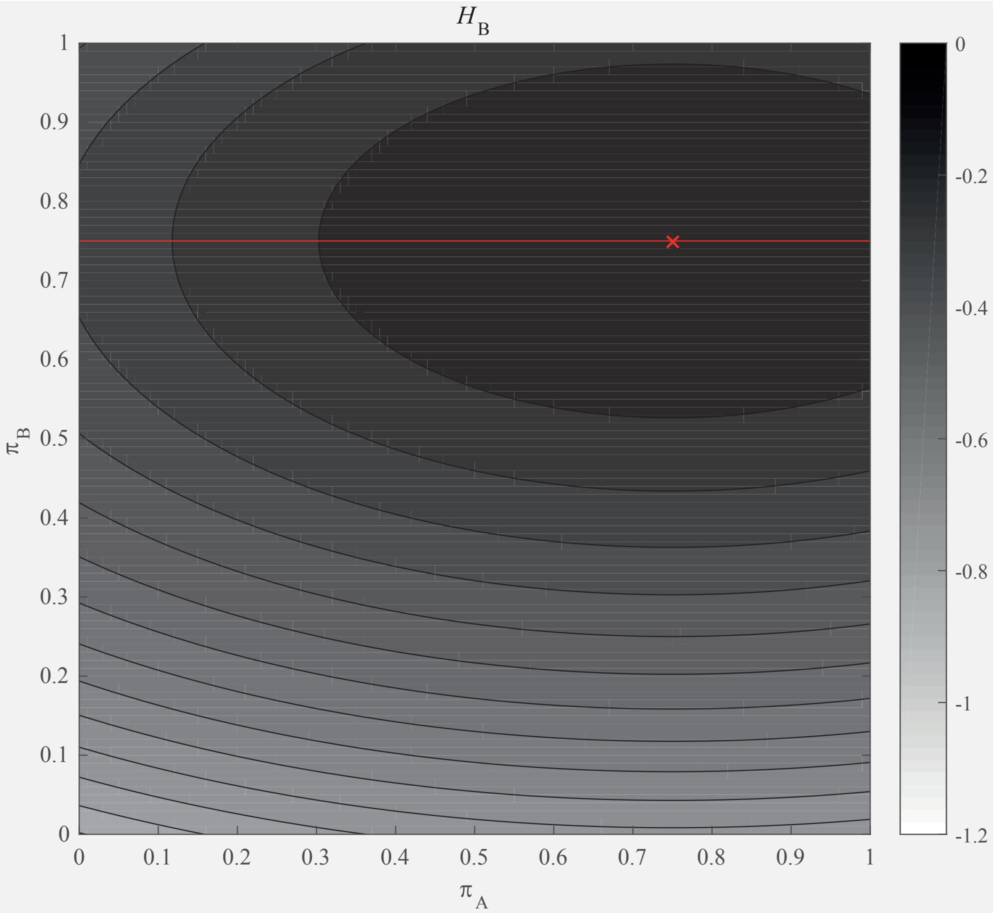


Figure 6.2.: Objective function contour plot for the coalition with stable pricing schemes (red line) and optimal point (red cross).

Obviously, none of these single-stakeholder coalitions can reach their optimal point because the other stakeholder will move him away from the optimum. The only partition in which any (and both) of the single-stakeholders coalitions exist is $\{\{A\},\{B\}\}$. The only pricing scheme for which both these coalitions are stable is $(\pi_A,\pi_B) = (0.25,0.75)$, i.e., the point where the two red lines in Figures 6.1 and 6.2 intersect. This is the only Nash-equilibrium for this partition, and the objectives correspond with $H_A(0.25,0.75) = -1/8$ and $H_B(0.25,0.75) = -1/16$.

For partition with the grand coalition $\{\{A,B\}\}$ the combined objective function $H_A + H_B$ has to be analysed to find stable pricing schemes. The grand coalition controls all pricing measures, and can thus set the pricing scheme equal to its optimal value. By solving these equations,

$$\frac{\partial(H_{\mathrm{A}}(\pi_{\mathrm{A}},\pi_{\mathrm{B}}) + H_{\mathrm{B}}(\pi_{\mathrm{A}},\pi_{\mathrm{B}}))}{\partial\pi_{\mathrm{A}}} = 0, \tag{6.17}$$

$$\frac{\partial(H_{\mathrm{A}}(\pi_{\mathrm{A}},\pi_{\mathrm{B}}) + H_{\mathrm{B}}(\pi_{\mathrm{A}},\pi_{\mathrm{B}}))}{\partial\pi_{\mathrm{B}}} = 0, \tag{6.18}$$

one can find resulting scheme $\boldsymbol{\pi}^* = (\pi_{\mathrm{A}},\pi_{\mathrm{B}}) = (7/20, 7/12)$, with corresponding value $-2/15$ of the combined objective function. For all TU-game solution concepts, this pricing scheme is eventually played by the stakeholders under cooperation, see Figure 6.3 for an illustration.

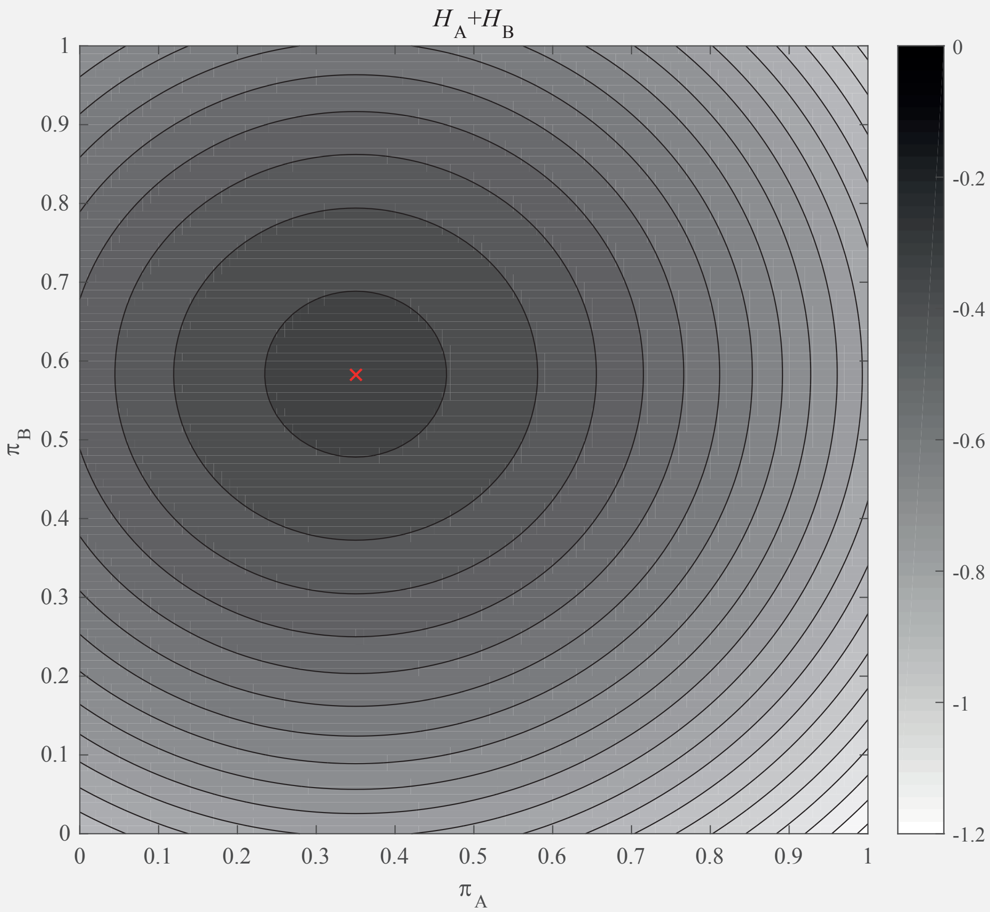


Figure 6.3.: Objective function contour plot for the coalition with optimal point (red cross).

Applying the definition of coalition values leads the coalition values in Table 6.1.

Table 6.1.: Coalition values

C	$v(C)$
$\emptyset$	0
$\{\text{A}\}$	$-\frac{1}{8}$
$\{\text{B}\}$	$-\frac{1}{16}$
$\{\text{A},\text{B}\}$	$-\frac{2}{15}$

The TU-game theoretical analysis of the game presented in Table 6.1 is rather straightforward since only two stakeholders are involved. Therefore, just the results are presented; the next examples aim to provide more insight in the computation and intuition of these solution concepts. Although, note that due to the negative coalition values this game is rather to split losses than to share profit. The core equals

$$K(\mathcal{S},v) = \left\{ \chi_{\text{A}}, \chi_{\text{B}} \in \mathbb{R} \middle| \chi_{\text{A}} + \chi_{\text{B}} = -\frac{2}{15} \wedge \chi_{\text{A}} \geq -\frac{1}{8} \wedge \chi_{\text{B}} \geq -\frac{1}{16} \right\} \tag{6.19}$$

The Shapley- and compromise value coincide and equal allocation

$$(\chi_{\text{A}}, \chi_{\text{B}}) = \left(-\frac{47}{480}, -\frac{17}{480} \right) \approx (-0.0979, -0.0354). \tag{6.20}$$

Remember that the differences per stakeholder between the stakeholder's objective values in the optimal pricing scheme of the grand coalition and the chosen allocation resemble the resulting (monetary) transfers. In this game the stakeholder's objective values for the optimal pricing scheme are $H_{\text{A}}(7/20, 7/12) = -59/900 \approx -0.656$ and $H_{\text{B}}(7/20, 7/12) = -61/900 \approx -0.678$. This means stakeholder A has to compensate stakeholder B substantially if the single-point solution concepts' allocation is effectuated.

In the following example, three stakeholders are considers. Instead of continuous price levels, the example deals with discrete price levels. By enumerating all different pricing schemes, the computation of coalition values based on all Nash-equilibria is illustrated.

Example: Three players with binary price levels

Consider grand coalition $\mathcal{S} = \{\text{A},\text{B},\text{C}\}$ with $\Pi_{\text{A}} = \Pi_{\text{B}} = \Pi_{\text{C}} = \{0,1\}$, so each stakeholder can either implement their pricing scheme or not, but no specific level can be set. They consider the following objective functions:

$$H_{\mathsf{A}}(\boldsymbol{\pi}) = \sum_{s \in \mathcal{S}} \pi_s \tag{6.21}$$

$$H_{\mathsf{B}}(\boldsymbol{\pi}) = \pi_{\mathsf{B}} \tag{6.22}$$

$$H_{\mathsf{C}}(\boldsymbol{\pi}) = -\pi_{\mathsf{B}} - \pi_{\mathsf{C}} \tag{6.23}$$

Table 6.2.: Coalition objective function values for all pricing schemes

Pricing schemes			Coalition objectives $\sum_{s \in C} H_s(\boldsymbol{\pi})$ for $C =$						
π_{A}	π_{B}	π_{C}	{A}	{B}	{C}	{A,B}	{A,C}	{B,C}	$\mathcal{S}$
0	0	0	0	0	0	0	0	0	0
0	0	1	1	0	-1	1	0	-1	0
0	1	0	1	1	-1	2	0	0	1
0	1	1	2	1	-2	3	0	-1	1
1	0	0	1	0	0	0	1	0	1
1	0	1	2	0	-1	1	1	-1	1
1	1	0	2	1	-1	2	1	0	2
1	1	1	3	1	-2	3	1	-1	2

Table 6.2 presents all eight possible pricing schemes with resulting objective functions. Note that especially C has conflicting interests with the other stakeholders.

By applying Definition 6.1 on this table, the stable pricing schemes for each coalition can be retrieved. Fix the price levels of the stakeholders outside the coalition and check if the objective can not be improved by switching the own price levels. Table 6.3 shows the results. Note that pricing scheme $(\pi_{\mathsf{A}}, \pi_{\mathsf{B}}, \pi_{\mathsf{C}}) = (1,1,0)$ is stable for all coalitions.

Table 6.3.: Stable coalitions per pricing scheme

Pricing schemes			Pricing scheme stability for coalition C =						
π_A	π_B	π_C	{A}	{B}	{C}	{A,B}	{A,C}	{B,C}	$\mathcal{S}$
0	0	0			✓			✓	
0	0	1							
0	1	0		✓	✓			✓	
0	1	1		✓					
1	0	0	✓		✓		✓	✓	
1	0	1	✓				✓		
1	1	0	✓	✓	✓	✓	✓	✓	✓
1	1	1	✓	✓		✓	✓		✓

Now all Nash-equilibria (see Definition 6.2) can be retrieved by looking up per pricing scheme and per partition if all involved coalitions are stable. Table 6.4 shows all Nash-equilibria for this game. Only three pricing schemes correspond to Nash-equilibria, while for all partitions a Nash-equilibrium exists.

Table 6.4.: All Nash-equilibria

Pricing schemes			Nash-equilibrium for partition Q =				
π_A	π_B	π_C	{A} {B} {C}	{A,B} {C}	{A,C} {B}	{B,C} {A}	$\mathcal{S}$
0	0	0					
0	0	1					
0	1	0					
0	1	1					
1	0	0				✓	
1	0	1					
1	1	0	✓	✓	✓	✓	✓
1	1	1			✓		✓

Finally, the TU-game can be defined by considering the worst Nash-equilibrium (see Definition 6.3) for each coalition. In this case each coalition has to consider its combined objective under scheme $(\pi_A,\pi_B,\pi_C) = (1,1,0)$, while coalitions {A,C}, {B} and $\mathcal{S}$ also have to consider scheme $(\pi_A,\pi_B,\pi_C) = (1,1,1)$, and coalitions {B,C} and {A} also have to consider scheme $(\pi_A,\pi_B,\pi_C) = (1,0,0)$. Table 6.5 shows the complete TU-game.

Table 6.5.: Coalition values

C	$v(C)$
$\emptyset$	0
{A}	1
{B}	1
{C}	-1
{A,B}	3
{A,C}	1
{B,C}	0
$\mathcal{S}$	2

Note that for the grand coalition the two equilibria correspond to the same value. So, two indifferent resulting pricing schemes $\boldsymbol{\pi}^*$ are possible; choosing one of them will only affect the eventual monetary transfers. Next, consider the several solution concepts starting with the core. For this TU-game the core equals

$$\begin{aligned} K(\mathcal{S},v) = \{\boldsymbol{\chi} | \\ \chi_A + \chi_B + \chi_C &= 2, \\ \chi_A &\geq 1, \\ \chi_B &\geq 1, \\ \chi_C &\geq -1, \\ \chi_A + \chi_B &\geq 3, \\ \chi_A + \chi_C &\geq 1, \\ \chi_B + \chi_C &\geq 0\}, \end{aligned} \tag{6.24}$$

which boils down to this single point core: $K(\mathcal{S},v) = \{(\chi_A,\chi_B,\chi_C) = (2,1,-1)\}$. Next, the marginal contributions $(v(C) - v(C \setminus \{s\}))$ of stakeholders entering the coalition one-by-one is calculated to obtain the Shapley value. This is done for each permutation (i.e., ordering in which stakeholders can enter) of the grand coalition. Table 6.6 shows the results, including the average over the orderings, which equals the Shapley value.

Table 6.6.: Shapley value calculation

Ordering	Marginal contributions A	B	C
A → B → C	1	2	-1
A → C → B	1	1	0
B → A → C	2	1	-1
B → C → A	2	1	-1
C → A → B	2	1	-1
C → B → A	2	1	-1
$\zeta(\mathcal{S},v)$	$1\frac{2}{3}$	$1\frac{1}{6}$	$-\frac{5}{6}$

Note that only for the orderings in which A enters first erratic marginal contributions appear. This is due to the coalition value of {B,C}, which is the only two-stakeholder coalition for which no surplus is generated when the stakeholders cooperate (i.e., $v(\{A,B\}) = v(\{A\}) + v(\{B\}) + 1$ and similar for {A,C}, but *not* for {B,C}).

It is remarkable that the Shapley value $\zeta(\mathcal{S},v) = \left(1\frac{2}{3}, 1\frac{1}{6}, -\frac{5}{6}\right)$ is not in the core of the game. In general the Shapley value does not have to be stable against coalition deviations, which is here also the case. Coalitions {A,B} and {A,C} can make rightful objections against the allocation of the Shapley value, since their coalition value is not met.

The final analysed solution concept is the compromise value. Computing equations (6.8) and (6.9) lead to rather boring bounds as shown in Table 6.7. There is no need to use Equation (6.10), nor to calculate the α.

Table 6.7.: Compromise value of TU-game $(\mathcal{S},\bar{v})$ for each stakeholder $s \in \mathcal{S}$

s	$\boldsymbol{\Omega}$	$\boldsymbol{\omega}$	$\boldsymbol{\eta}(\mathcal{S},v)$
A	2	2	2
B	1	1	1
C	-1	-1	-1

The compromise value lies in – actually equals – the core. The Shapley value and compromise value are not equal for this TU-game.

In the final example, a similar discrete price level setting is provided, but the players have more conflicting objectives. Not for all coalitions a Nash-equilibrium exists, meaning that these coalitions will have zero value. In addition, the core of game in this example is larger than a single point (as it was in the previous example). This allows – literately – an illustration of the differences between the solution concepts.

EXAMPLE: Three conflicting players

Again, consider grand coalition $\mathcal{S} = \{\text{A},\text{B},\text{C}\}$ with $\Pi_\text{A} = \Pi_\text{B} = \Pi_\text{C} = \{0,1\}$, but now the stakeholders consider the following objective functions:

$$H_\text{A}(\boldsymbol{\pi}) = -2 \times (-1)^{\sum_{s\in\mathcal{S}} \pi_s} \tag{6.25}$$

$$H_\text{B}(\boldsymbol{\pi}) = (-1)^{\sum_{s\in\mathcal{S}} \pi_s} \tag{6.26}$$

$$H_\text{C}(\boldsymbol{\pi}) = \sum_{s\in\mathcal{S}} \pi_s \tag{6.27}$$

Table 6.8.: Coalition objective function values for all pricing schemes

Pricing schemes			Coalition objectives $\sum_{s\in C} H_s(\boldsymbol{\pi})$ for $C =$						
π_A	π_B	π_C	{A}	{B}	{C}	{A,B}	{A,C}	{B,C}	$\mathcal{S}$
0	0	0	-2	1	0	-1	-2	1	-1
0	0	1	2	-1	1	1	3	0	2
0	1	0	2	-1	1	1	3	0	2
0	1	1	-2	1	2	-1	0	3	1
1	0	0	2	-1	1	1	3	0	2
1	0	1	-2	1	2	-1	0	3	1
1	1	0	-2	1	2	-1	0	3	1
1	1	1	2	-1	3	1	5	2	4

Table 6.8 presents all eight possible pricing schemes with resulting objective functions. Note that all objective values only depend on the number of prices 'implemented' and that stakeholders A and B have completely conflicting objectives. The stable pricing schemes for each coalition (see Table 6.9) are therefore also 'spread out' over the pricing schemes.

Table 6.9.: Stable coalitions per pricing scheme

Pricing schemes			Pricing scheme stability for coalition C =						
π_A	π_B	π_C	{A}	{B}	{C}	{A,B}	{A,C}	{B,C}	$\mathcal{S}$
0	0	0		✓					
0	0	1	✓		✓	✓	✓		
0	1	0	✓			✓			
0	1	1		✓	✓			✓	
1	0	0	✓			✓	✓		
1	0	1		✓	✓			✓	
1	1	0		✓				✓	
1	1	1	✓		✓	✓	✓		✓

Table 6.10 shows all Nash-equilibria for this game. Only three Nash-equilibria exist, which only involve partitions {{A,B},{C}} and $\mathcal{S}$, and pricing schemes (0,0,1) and (1,1,1). This means that not all partitions have corresponding Nash-equilibria, and the majority of coalitions can not even be associated with stable pricing schemes. This has its influence on the TU-game coalition values (presented in Table 6.11) where these coaltions will have a value equal to zero. The resulting pricing scheme equals $\boldsymbol{\pi}^* = (1,1,1)$.

Table 6.10.: All Nash-equilibria

Pricing schemes			Nash-equilibrium for partition Q =				
π_A	π_B	π_C	{A} {B} {C}	{A,B} {C}	{A,C} {B}	{B,C} {A}	$\mathcal{S}$
0	0	0					
0	0	1		✓			
0	1	0					
0	1	1					
1	0	0					
1	0	1					
1	1	0					
1	1	1		✓			✓

Table 6.11.: Coalition values

C	$v(C)$
$\emptyset$	0
{A}	0
{B}	0
{C}	1
{A,B}	1
{A,C}	0
{B,C}	0
$\mathcal{S}$	4

Next, the solution concepts are analysed and visualised. The core equals

$$\begin{aligned} K(\mathcal{S},v) = \{\chi \in \mathbb{R}^3 | \\ \chi_{\mathsf{A}} + \chi_{\mathsf{B}} + \chi_{\mathsf{C}} = 4, \\ \chi_{\mathsf{A}} \geq 0, \\ \chi_{\mathsf{B}} \geq 0, \\ \chi_{\mathsf{C}} \geq 1, \\ \chi_{\mathsf{A}} + \chi_{\mathsf{B}} \geq 1, \\ \chi_{\mathsf{A}} + \chi_{\mathsf{C}} \geq 0, \\ \chi_{\mathsf{B}} + \chi_{\mathsf{C}} \geq 0\}, \end{aligned} \tag{6.28}$$

which is a (non-empty) polyhedron in $\mathbb{R}^3$. This non-empty core means that although the objective functions are conflicting, still allocations exist for which no coalition has a valid argument to object. Table 6.12 shows the Shapley value computation. The Shapley value assigns equal allocations to all stakeholders because each of them appears in exactly one coalition (other than the grand coalition) with coalition value equal to one, while the other values equal zero. No distinction is made between the stakeholders.

Table 6.12.: Shapley value calculation

Ordering	Marginal contributions A	B	C
A → B → C	0	1	3
A → C → B	0	4	0
B → A → C	1	0	3
B → C → A	4	0	0
C → A → B	-1	4	1
C → B → A	4	-1	1
$\zeta(\mathcal{S},v)$	$\frac{4}{3}$	$\frac{4}{3}$	$\frac{4}{3}$

The final solution concept is the compromise value, see Table 6.13. In this case α equals 0.3. Stakeholders C has a slight advantage over the other two stakeholders here. $v(\{A,B\}) = 1$ limits C's upper bound to 3, but $v(\{C\}) = 1$ provides a good lower bound. Since the budget (4) is closer to the lower bound than to the upper bound, C's differences ultimately lead to an advantage.

Table 6.13.: Compromise value of TU-game $(\mathcal{S},\bar{v})$ for each stakeholder $s \in \mathcal{S}$

s	$\boldsymbol{\Omega}$	$\boldsymbol{\omega}$	$\boldsymbol{\eta}(\mathcal{S},v)$
A	4	0	$\frac{6}{5}$
B	4	0	$\frac{6}{5}$
C	3	1	$\frac{8}{5}$

Games with $|\mathcal{S}| = 3$ have the advantage that the core (and the solution concepts that lie in the core) can be nicely visualised. Figure 6.4 provides this illustration for the game. The core is symmetrical for A and B, which is also reflected in the one-point solution concepts. At the negatiation table, stakeholder C has the advantage that it is associated with more Nash-equilibria as a single-stakeholder coalition, which provides him arguments (i.e., the compromise value) to receive a higher amount than the other stakholders. On the other hand, the pricing scheme of the grand coalition's equilibrium (i.e., $\boldsymbol{\pi}^*$) leads to direct objective values of 2, −1 and 3 for respectively A, B and C. If the final allocation ends up close to the Shapley or compromise value, it is inevitable that stakeholder B receives a monetary compensation from both other stakeholders.

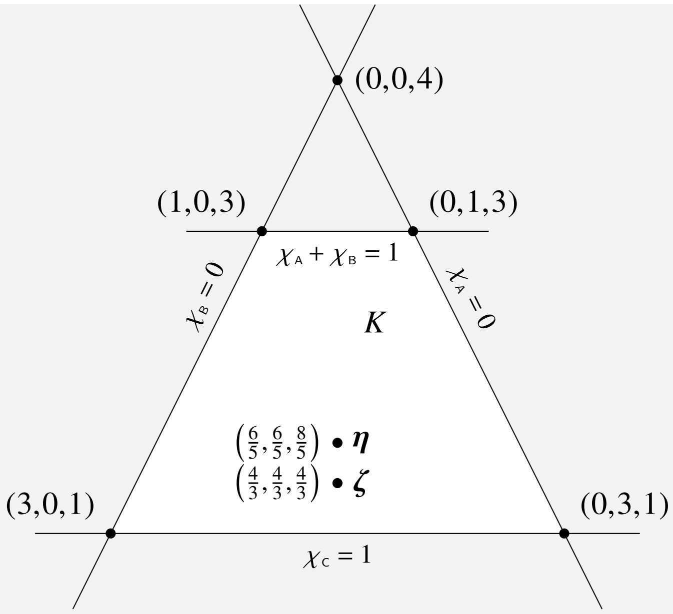


Figure 6.4.: A graphical representation of the solution concepts for the TU-game $(\mathcal{S}, v)$ with the core (white area, K), Shapley value (ζ), and compromise value (η). The drawn plane is $\sum_{s \in \mathcal{S}} \chi_s = v(\mathcal{S}) = 4$, also the equalities that define the boundaries of the core are depicted.

6.6. Extensions

In this section three possible extensions of the approach are provided and discussed.

6.6.1. Cost of coalition formation

One of the assumptions of the approach is that no costs are involved when stakeholders enter or leave coalitions. Perhaps surprisingly, introducing such costs can provide more 'stable' solutions. By allowing the analyst (i.e., 'game leader') to impose a cost when a stakeholders enters or leaves the grand coalition, the analysis of the TU-game would become simpler since then epsilon-Nash-equilibria can be enforced (which always exist).

6.6.2. Multiplicity of effects

Notice that if effect levels appear in multiple objective function, then they are counted multiple times in the combined objective function of, e.g., the grand coalition. This is not problematic for the approach, since this price level adds to the objective of each individual stakeholders of which the total sum will be optimized, but it can be counter-intuitive. Especially 'societal wide' (external) effects (i.e., those effects that represent utility of populations) demonstrate this. For example, if social welfare of a region is included in as well its covering national government's as its own (regional) government's objective function, it will appear twice in the objective function of all coalitions in which they both appear. This diminishes the influence of other effect levels by these and other stakeholders in the coalition. On the other hand, effects that read as 'a celebration will take place', which is added to multiple stakeholders' objectives as 'participate in the celebration', it is completely logical to repeat the effect in the coalition's objective.

Since in general stakeholders have their own rationale of existence and their own executive power; there is no bias in the TU-game derived from an MSP. However, if the additive behaviour of effect levels is undesirable (i.e., weighting factors in combined objective functions become erratic), then the framework can easily be adjusted by defining coalition specific objective functions. [15]

6.6.3. Generalization to shared pricing measures

In reality, not all pricing measures ae controlled by a single stakeholders. For example, travellers' association like the Dutch ANWB do have influence on policy making, but do not have control. If such 'lobbies' are key to analyse, then additional structures have to be introduced. The easiest way to approach this is to add a term in the objective function of the empowered stakeholder of the pricing measure. For example, by introducing a penalty (or veto) if a lobby group does not agree. Another approach is to introduce an intermediate layer in which multiple stakeholders have to decide on the measure. The latter will require a different form of modelling decision making that is out of the scope of this thesis.

6.7. Synthesis

This chapter has introduced a solution method to approach the MSP, and aimed at bridging the gap between multi-objective optimization and cooperative game theory. From objective functions and pricing measures per stakeholder, a TU-game has been defined by using Nash-equilibria as an intermediate result. The solution concepts core, Shapley value, and compromise value to TU-games are presented. This section discusses practical aspects and policy implications of this framework.

Compared to the traditional *social welfare optimization*, a more detailed view on possible gains and losses of each stakeholder is provided, under cooperation as well as under competition. The assumptions of *first-best pricing* and social welfare optimization that every individual of a population participates in the negotiations and that unlimited mutual transactions between them can exist are relaxed. The provided framework provides a concrete interpretation of the negotiation table. The resulting game theoretical solutions provide more feasible and more realistic 'maximum achievements', than can be obtained under merely social welfare optimization.

The negotiation table is made concrete in the following manner. Each stakeholder arrives with an objective function and pricing measure(s) (with feasible price levels). The market power of a stakeholder is completely determined by the pricing measures it controls. Based

[15] Instead of defining stakeholder objective functions $H_s, \forall s \in \mathcal{S}$ and deriving *coalition objective functions* implicitly by $H'_C := \sum_{s \in C} H_s, \forall C \subseteq \mathcal{S}$, these coalition objectives H'_C can be defined directly.

on all possible pricing schemes that can be constructed out of the pricing measures, the corresponding objective value of each stakeholder becomes clear. Subsequently, combined objectives of coalitions of stakeholders are computed. The provided notions of stability and equilibrium provide each stakeholder (and the coalitions it belongs to) with arguments. Finally, the TU-game solution concepts provide suggestions for final allocations. The stakeholders have to discuss and negotiate on those allocations to reach the final result. It is strongly suggested to implement resulting pricing scheme $\boldsymbol{\pi}^*$, since then the grand coalition's objective is optimal. Unfortunately, if the core is empty, it will be difficult to reach an agreement, since then there exists a stakeholder or coalition that will object to it.

Chapter 7.

Case Study: The Randstad

The case study in this chapter considers three stakeholders and three pricing measures in the Randstad area in the Netherlands. It uses a hypothetical setting to show how the tools provided in this dissertation can be brought to practice. The methodological advances in traffic assignment described in Part I are applied on a large real network. Based on the new stakeholders model described in Chapter 6 the interaction between stakeholders is analysed. In addition the impact of the pricing schemes on the transport system is determined in terms of congestion and emissions. The case study reveals that transport pricing is an effective measure to mitigate external effects, and it reveals that the governments and train operator have conflicting objectives. Finally, the computed price of competition (i.e., the difference between non-cooperative and cooperative outcomes) exposes a remarkable impact on social welfare when stakeholders start to cooperate.

This chapter is a translated and extended version of:

- Smits, E.-S., Pel, A., van Nes, R., and van Arem, B. (2016). Modelleren mobiliteitsbeprijzing met meerdere actoren. *Tijdschrift Vervoerwetenschap*. ISSN 1571-9227

The Randstad is a metropolitan area in the Netherlands that covers multiple cities (among them are Amsterdam, Rotterdam, The Hague, and Utrecht) with more than 7 million inhabitants. The densely populated urban areas are connected with road and rail networks. Congestion, crowding and emissions cause hinder on a daily basis. Innovative pricing measures, such as a kilometre charge, have never been deployed due to lack of political and public support (see e.g., Smaal, 2012; Vonk Noordegraaf et al., 2014). A kilometre charge, cordon charge, and/or different train tariffs could change the mobility consumption pattern of travellers.

The purpose of the case study is to show how the theories and methods presented in this dissertation can be brought to practice. The aim is to use a large network and as much 'real' data as possible; this shows that the methods are practically applicable given their computation time. The pricing measures and objectives of the stakeholders are hypothetical, and are chosen to illustrate how possibly conflicting interests could be handled. Therefore, the policy implications of the game theoretical approach are conditional to these hypotheses as well.

The case study presented in this chapter studies different pricing measures and models the preferences of the train operator, and national and local governments. The Amsterdam city council aims to reduce congestion and to improve its economic position by introducing a cordon charge. The national government wants to reduce travel times and emissions and to improve the social welfare by introducing a kilometre charge. The train operator wants to spread the demand to reduce the number of costly additional services in the peak, and thereby maximizes its profit. All pricing measures are time differentiated for the peak and shoulder periods.

7.1. Set-Up

Following the line of Chapter 2, the model has two levels. The three stakeholders and their pricing measures are captured in the upper level, while the lower captures the responses of the travellers regarding their used mode, time-of-day, and route. The upper level stakeholders model determines pricing schemes which are evaluated in the lower level assignment model. The returned effects are travel times, emission levels, expected maximum utility of each traveller, modal splits, time-of-day splits, route usage, profit, and revenues from pricing measures. The scope of the case study is the morning commute in the year 2020. All reported values represent a single morning commute.

The upper level model adopts the game theoretical approach as presented in Chapter 6. Based on the objective per stakeholder, the formation of coalitions and notions of stability and equilibrium a TU-game will be formulated. This game is analysed and the core, Shapley value and compromise value are computed. In addition to the TU-game, the Pareto-frontier is presented and the price of competition is discussed.

The lower level adopts the Quasi-Dynamic Traffic Assignment (QDTA) model of Chapter 3. The main data source for the transport model is the Nederlands Regionaal Model (Dutch Regional Model) (NRM). The morning peak Origin-Destination (O-D) demand and the road network from the NRM-west version form the basis. The NRM-west is one of the four NRM versions, and it covers the western part of the Netherlands which contains the Randstad area as well. Travellers consider mode, time-of-day, and route choices. Traffic is propagated using hybrid quasi-dynamic network loading model described in Section 3.3, and is available in the OmniTRANS software package under the name Static Traffic Assignment with Queuing (STAQ). Mode and time-of-day choice are modelled simultaneously, but separately from route choice. Because a route choice equilibrium model is integrated within STAQ, this is a more practical approach. O-D matrices and travel times are used to connect both choice models. The structure of the models is derived from the framework developed in this thesis. Model parameters and calibration data come from multiple sources, such as the original NRM-west, socio-economic statistics, and willingness to pay studies.

The network, see Figure 7.1, reaches till outside the Netherlands and contains 87 504 links. There are 3 608 zones of which 2 122 are within the study area. The train network is just considered indirectly by copying travel times and costs fron the NRM-west. The modelled

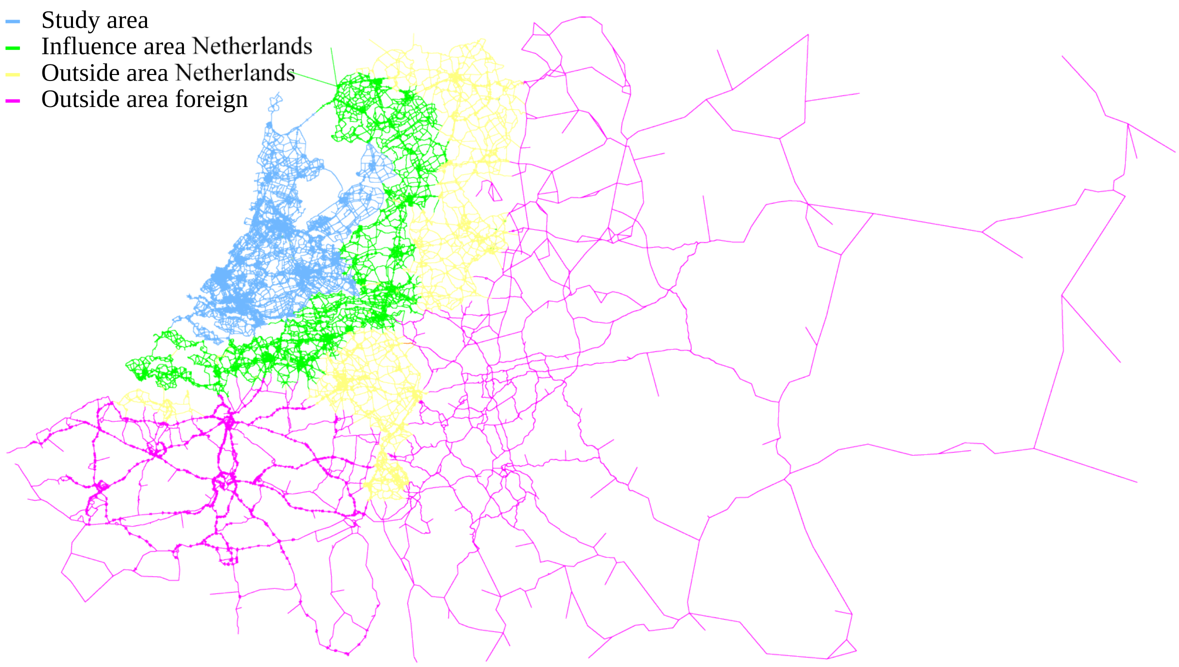


Figure 7.1.: Road network, the blue area indicates the study area

area and network are not restricted to the study area to be able to model the influence from trips starting or ending outside the study area. However, the resulting effects are determined for the study area only.

7.1.1. Stakeholders and Pricing Measures

The three stakeholders each have their own pricing measure to obtain their objective. Each objective function is a weighted sum of effect levels. Since objective functions are compared in the TU-game approach, they are monetized. This allows monetary transfers when the solution concepts are considered. The objectives of the stakeholders are introduced briefly. After the underlying TA-model is introduced, the effect levels and objective functions are further formalized.

National Government The national government, denoted with GOV, considers a kilometre charge which is assumed to affect all roads. During peak and shoulder periods, different price levels can be determined. It is assumed that the price per kilometre in the shoulder is never higher than the price in the peak to reduce the amount of feasible pricing schemes. The (*on*-)peak kilometre charge is denoted as KM-ON, and the shoulder kilometre charge is denoted as KM-OFF. Each of the associated levels can be either €0.00, €0.05 or €0.10, i.e., $\Pi_{\text{KM-ON}} = \Pi_{\text{KM-OFF}} = \{0, 0.05, 0.1\}$. It is assumed that on-board technology is available to collect the charges. Its objective function contains expected maximum utility of each commuter, the income from the kilometre and cordon charge, the emissions, and the loss hours of other

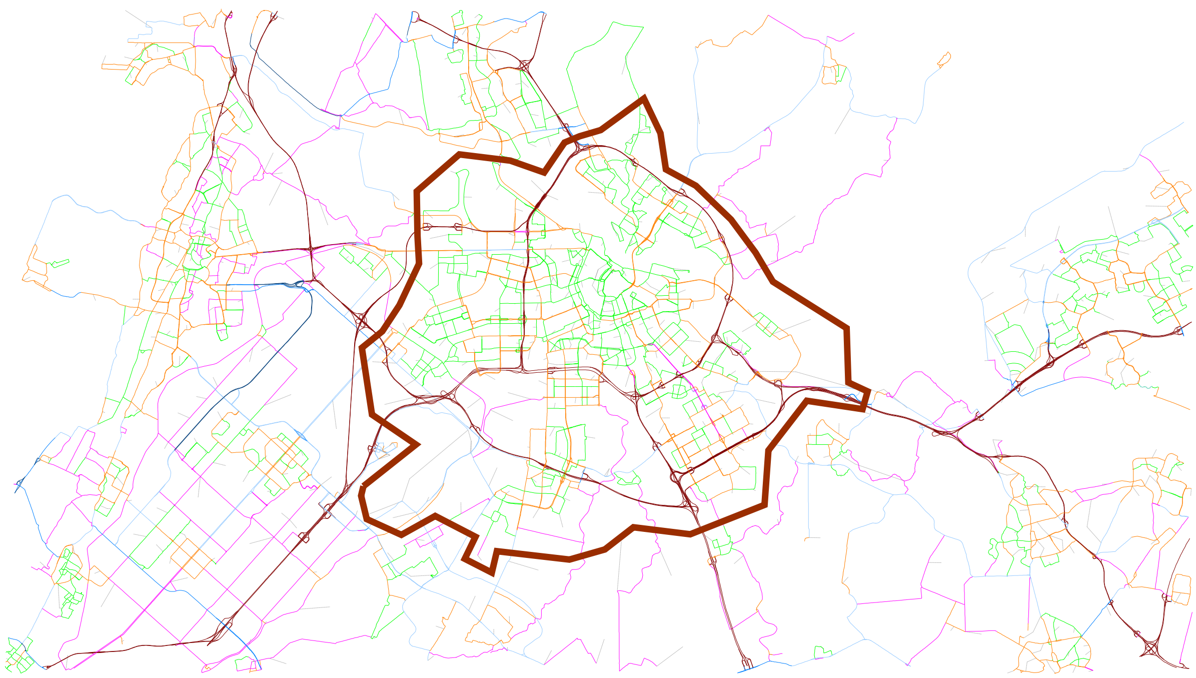


Figure 7.2.: Cordon around Amsterdam

travellers.

Municipality of Amsterdam The local government in Amsterdam, denoted with AMS, considers a cordon charge to improve the economic position of the city. Figure 7.2 shows the location of the cordon. Travellers have to pay when they enter the cordon, e.g., by use of electronic toll gates. The peak hour toll is denoted as CRDN-ON and the shoulder toll is denoted as CRDN-OFF. Each of the associated levels can be either €0.00, €4.00 or €8.00, i.e., $\Pi_{\text{CRDN-ON}} = \Pi_{\text{CRDN-OFF}} = \{0,4,8\}$. Also for the cordon charge it is assumed that the shoulder price can not exceed the peak price to reduce the numver of feasible pricing schemes. The objective function of AMS is deliberately different from social welfare to create contrast in interests between GOV and AMS; it contains the following effects: Accessibility of Amsterdam, loss due to missed economic activity, and emissions within Amsterdam. Since the cordon charge revenues will be recycled in the transport system, it is not part of Amsterdam's objective.

Train Operator The train operator, denoted by TR-OP, considers the train fares. Both the peak and shoulder fare, denoted respectively with FARE-ON and FARE-OFF, can be in- and decreased with 20%. The feasible price levels are written as a factor compared to the original prices from the reference model: $\Pi_{\text{FARE-ON}} = \Pi_{\text{FARE-OFF}} = \{0.8,1,1.2\}$. The shoulder fare can not increase more than the peak fare. The objective equals profit, so the objective function is the difference between income from ticket sales and costs. The costs consist of operating costs, where it is relatively expensive to run additional services during the peak to accommodate higher peak demand. Subsidies from the government are assumed constant in this case study; therefore, they do not affect the train operator's objective.

7.2. Underlying Traffic Assignment Model

The focus of the case study is on a typical 2020 morning commute. Therefore, the considered set of time-of-days is $\mathcal{T} = \{\text{PEAK}, \text{SHOULDER}\}$, in which time-of-day PEAK lasts from 07:00 AM to 09:00 AM, and time-of-day SHOULDER consists of two time intervals: 04:00 AM to 07:00 AM, and 09:00 AM to 12:00 AM (see Figure 7.3). The choice model determines the total demand for each time-of-day, the hourly matrices are respectively obtained by division with 2 (PEAK) and 6 (SHOULDER). The hourly demands are considered representative or each time-of-day. Furthermore, the simplifying assumption is made that average network conditions are equal for both shoulders.

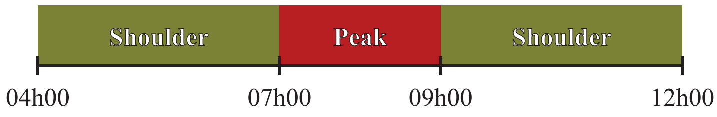


Figure 7.3.: Peak and shoulder time-of-days

Since the lengths of the time-of-days differ, the choice model should correct for this by adding a correction term (Ben-Akiva and Abou-Zeid, 2013). It is not straightforward to calculate this constant for nested Generalized Multivariate Extreme Value (GMEV)-models. As explained later in this section, a time-of-day alternative specific constant is used in the case study. As a pragmatic solution, it is assumed that this constant contains the correction term.

Set of modes $\mathcal{M} = \{\text{CAR}, \text{TRAIN}, \text{HOME}\}$ contains car, train, and stay-at-home alternatives. Modes CAR and TRAIN can be combined with both time-of-days (PEAK and SHOULDER), and thus four alternatives are available. However, special mode HOME is not related to a time-of-day choice, therefore, it leads to a single alternative. Route choice is only modelled for mode CAR, and is determined separately from the main choice model. Route choice for mode TRAIN is omitted because route choice effects due to pricing are expected to be minimal; the train tariffs are not differentiated per train service, and crowding is not incorporated as an attribute in the choice model.

Consider the set of user-classes $\mathcal{U} = \{\text{PEAK-PREF}, \text{NO-PREF}\}$. The only distinction in user-classes for this case study is the preferred time-of-day. Having two user-classes simplifies the case study, but allows stating and discussing all preferences and prices.[1] User-class PEAK-PREF has the preference to travel during time-of-day PEAK, for example due to fixed working hours, or habit. The choice model penalizes alternatives during the SHOULDER time-of-day for this user-class. On the other hand, user-class NO-PREF has no preference for either travelling during the PEAK or the SHOULDER.

Figure 7.4 depicts the flowchart of the underlying assignment model. The doubly iterative process maps a pricing scheme to a set of effects. The outer iterative process searches for an

[1] More realistic case studies should contain more user-classes with different time valuations.

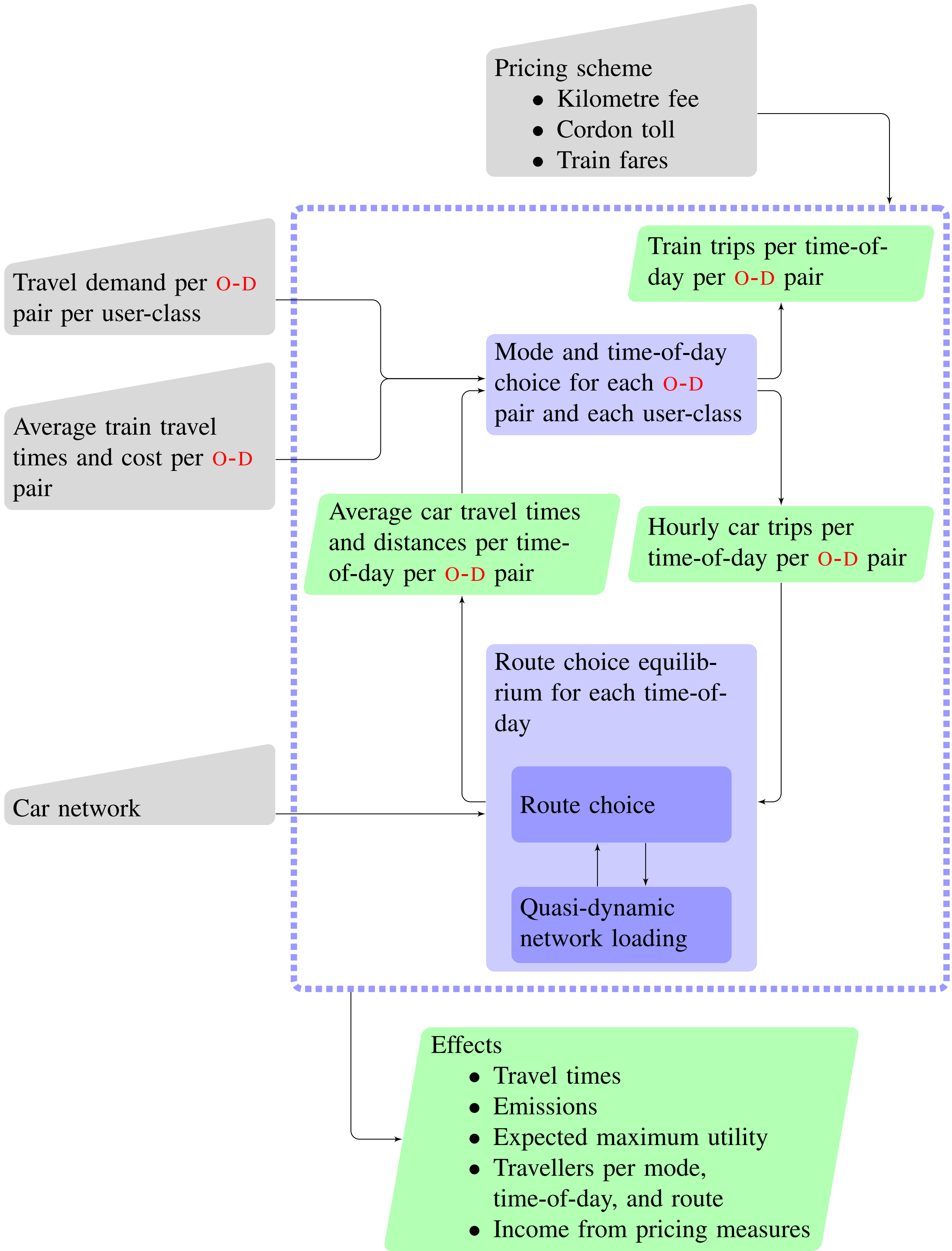


Figure 7.4.: Flowchart of the underlying traffic assignment model

equilibrium in mode and time-of-day choice. In every iteration the simultaneous mode and time-of-day choice model[2] computes hourly car trip matrices per time-of-day that is input for the route choice equilibrium model[3], which translate into travel impacts in the form of travel times and distances. The latter is input for the choice model again. The route choice equilibrium model is the inner iterative process; it iterates route choice and network loading.

The outer process is repeated 12 times with the method of successive averages (with diminishing weighting factors). The inner process is initiated with a 'warm start': the fully converged route choice probabilities of the zero-pricing scheme are used for the first network loading procedure. For the first eight outer iterations, 25 inner iterations are used, for the final four outer iterations, 50 inner iterations are used. After each step the duality gap for the simultaneous mode and time-of-choice model is computed, see Equation (3.30).

7.2.1. Mode and Time-of-Day Choice

This section develops the main choice model in the case study that captures mode and time-of-day choice. The route choice model within STAQ in the OmniTRANS software package is described in the next section, which we simply adopt in this case study. The choice probability for each alternative is determined using the Random Utility Maximization (RUM) framework and its methods from Chapter 4.[4] Recall that the combination of a set of alternatives, a generating vector, and a generating function completely determine the choice probabilities. A nested-logit structure captures correlations between alternatives with the same mode. Furthermore, a multiplicative error term formulation ensures the heteroskedasticity of random utilities of alternatives.

According to the framework in Chapter 3[5] the set of choice alternatives $\mathcal{C}$ is a subset of time-of-day and route combinations (i.e., $\mathcal{T} \times \mathcal{R}$). By grouping the alternatives per O-D-pair, the relevant choice sets $\mathcal{C}_{od}$ are formed. In that framework, the mode is a characteristic of a route. Since for this main choice model route choice is exogenous, it considers a single route per mode per O-D pair od. Average travel times and average travel costs reflect all routes per mode per O-D-pair, and are captured with so called representative routes $\overline{\text{CAR}}_{od}$ and $\overline{\text{TRAIN}}_{od}$ ($\in \mathcal{R}$) for respectively modes CAR and TRAIN, and for each $od \in \mathcal{OD}$. This leads to five choice alternatives per O-D-pair, i.e.

$$\mathcal{C}_{od} = \left\{(\text{PEAK},\overline{\text{CAR}}_{od});(\text{PEAK},\overline{\text{TRAIN}}_{od});(\text{SHOULDER},\overline{\text{CAR}}_{od});(\text{SHOULDER},\overline{\text{TRAIN}}_{od});(\cdot;\text{R}_0)\right\}, \tag{7.1}$$

where the dot $\cdot$ means that no time-of-day is specified for route R_0, which is the dummy route for mode HOME (see Section 3.2). Denote all CAR representative routes as $\overline{\text{CAR}} = \cup_{od\in\mathcal{OD}}\overline{\text{CAR}}_{od}$, and all TRAIN representative routes as $\overline{\text{TRAIN}} = \cup_{od\in\mathcal{OD}}\overline{\text{TRAIN}}_{od}$.

Using representative routes is a pragmatic approach that disconnects the route choice from the other choices. Ideally, one uses the expected maximum utility of the route choice model

[2] Implemented in Matlab
[3] Available in OmniTRANS
[4] Although the scope of that chapter is route choice, the methodology with multiplicative utility formulations is used here. Chapter 3 explains how the application of its theory can be applied to simultaneous mode, route and time-of-day choice. To keep route choice exogenous, representative routes are used for each mode.
[5] The notation from that chapter is adopted here.

under equilibrium (see Table 4.1 for its closed form formulas for GMEV-models). This expected maximum utility can then be substituted for all route-sums in Equation (3.15). Since STAQ does not use the exact generating vector and generating function as described in Section 3.2.4, this pragmatic approach is taken. Note, that all averages for representative routes calculated from the route choice equilibrium model are weighted with the internal route choice probabilities. These are dependent on the time-of-day of course.

Recall that the used systematic utility, adopted from Equation (3.7), is:

$$V_{(T;r;u)} = \begin{cases} \mathrm{ASC}_{(T;\mathrm{CAR};u)} - \kappa_{(T;r;u)} - \mathrm{VOT}^{\mathrm{IVT\text{-}FF}}_{(\mathrm{CAR};u)} \tau^{\mathrm{IVT\text{-}FF}}_{(T;r)} - \mathrm{VOT}^{\mathrm{IVT\text{-}CONG}}_{(\mathrm{CAR};u)} \tau^{\mathrm{IVT\text{-}CONG}}_{(T;r)} & \text{if } r \in \overline{\mathrm{CAR}} \\ \mathrm{ASC}_{(T;\mathrm{TRAIN};u)} - \kappa_{(T;r;u)} - \mathrm{VOT}^{\mathrm{IVT}}_{(\mathrm{TRAIN};u)} \tau^{\mathrm{IVT}}_{(T;r)} & \\ \qquad + \mathrm{VOT}^{\mathrm{WAIT}}_{(\mathrm{TRAIN};u)} \tau^{\mathrm{WAIT}}_{(T;r)} + \mathrm{VOT}^{\mathrm{A\text{-}E}}_{(\mathrm{TRAIN};u)} \tau^{\mathrm{A\text{-}E}}_{(T;r)} & \text{if } r \in \overline{\mathrm{TRAIN}} \\ \mathrm{ASC}_{(T;\mathrm{HOME};u)} - \kappa_{(T;r;u)} & \text{if } r = \mathrm{R}_0 \end{cases} \tag{7.2}$$

The alternative specific constants include two elements: a penalty for travelling outside the preferred time-of-day and the mode specific constant. The latter completely represents the systematic utility of mode HOME. The constants are determined with calibration as described in the next section. To avoid an underspecified system in the calibration proces, the alternative specific constants for mode TRAIN[6] and time-of-day PEAK are normalized to zero. This is the general formulation for the alternative specific constant for time-of-day $T \in \mathcal{T}$, mode $m \in \mathcal{M}$, and user-class $u \in \mathcal{U}$:

$$\mathrm{ASC}_{(T;m^r;u)} = \delta_T^{\mathrm{SHOULDER}} \delta_u^{\mathrm{PEAK\text{-}PREF}} \mathrm{ASC}^{\mathrm{SHOULDER}} + \delta_m^{\mathrm{CAR}} \mathrm{ASC}^{\mathrm{CAR}} + \delta_m^{\mathrm{HOME}} \mathrm{ASC}^{\mathrm{HOME}} \tag{7.3}$$

The route choice and the mode and time-of-day choice models are separated. Existing routines of the OmniTRANS software package are used to perform the route choice in an iterative manner with the Quasi-Dynamic Network Loading (QDNL). The mode and time-of-day choice uses the framework of Chapter 4, and is simulated in Matlab. As concluded in that chapter, the multiplicative utility formulation is more suitable for large real networks. Therefore, we use the generating vector $\boldsymbol{y}^{\mathrm{M}}$ as defined in Equation (4.33). Furthermore, a nested structure is assumed to capture correlation between the two time-of-day alternatives for each mode. The generating function (see Section 3.2.4; Equation (3.12)) for each choice set C_{od} is defined as

$$G(z) = \sum_{m \in \{\mathrm{CAR},\mathrm{TRAIN}\}} \left(\sum_{T \in \{\mathrm{PEAK},\mathrm{SHOULDER}\}} z_{(T,m)}^{\mu^{\mathrm{MODE}}} \right)^{\frac{\mu}{\mu^{\mathrm{MODE}}}} + z_{\mathrm{R}_0}^{\mu}. \tag{7.4}$$

The choice probabilities are similar to Equation (3.15), but do not include the route choice

[6] Testing both options showed that $\mathrm{ASC}^{\mathrm{CAR}} < \mathrm{ASC}^{\mathrm{TRAIN}}$, and since a negative systematic utility is required, $\mathrm{ASC}^{\mathrm{TRAIN}}$ – and not $\mathrm{ASC}^{\mathrm{CAR}}$ – is normalized.

nesting. This leads to the following choice probabilities:

$$P_{(T;r;u)} = \left(\sum_{m' \in \{\text{CAR},\text{TRAIN}\}} \left(\sum_{T' \in \mathcal{T}} \left(\frac{-1}{V_{(T';r^{m'};u)}} \right)^{\mu^{\text{MODE}}} \right)^{\frac{\mu}{\mu^{\text{MODE}}}} + \left(\frac{-1}{V_{(\cdot;\text{R}_0;u)}} \right)^{\mu} \right)^{-1}$$
$$\times \left(\sum_{T' \in \mathcal{T}} \left(\frac{-1}{V_{(T';r;u)}} \right)^{\mu^{\text{MODE}}} \right)^{\frac{\mu}{\mu^{\text{MODE}}}-1}$$
$$\times \left(\frac{-1}{V_{(T;r;u)}} \right)^{\mu^{\text{MODE}}}, \quad \text{if } r \neq \text{R}_0, \text{ and,} \tag{7.5}$$
$$P_{(T;r;u)} = \left(\sum_{m' \in \{\text{CAR},\text{TRAIN}\}} \left(\sum_{T' \in \mathcal{T}} \left(\frac{-1}{V_{(T';r^{m'};u)}} \right)^{\mu^{\text{MODE}}} \right)^{\frac{\mu}{\mu^{\text{MODE}}}} + \left(\frac{-1}{V_{(\cdot;\text{R}_0;u)}} \right)^{\mu} \right)^{-1}$$
$$\times \left(\frac{-1}{V_{(T;r;u)}} \right)^{\mu}, \quad \text{if } r = \text{R}_0,$$

with $r \in \{\overline{\text{CAR}}_{\mathcal{od}}, \overline{\text{TRAIN}}_{\mathcal{od}}, \text{R}_0\}$, $r^{\text{CAR}} = \overline{\text{CAR}}_{\mathcal{od}}$, and $r^{\text{TRAIN}} = \overline{\text{TRAIN}}_{\mathcal{od}}$, where $\mathcal{od}$ is the O-D-pair that corresponds with route r. By applying the choice probabilities to the demand, one can obtain the number of trips for each mode and time-of-day combination:

$$f_{(T;r)} = \sum_{u \in \mathcal{U}} P_{(T;r;u)} D_{(u;\mathcal{od})}, \quad \forall (T;r) \in \mathcal{C}, u \in \mathcal{U}, \tag{7.6}$$

where $\mathcal{od}$ is the O-D pair corresponding to (representative) route r. Note that the

7.2.2. Route Choice Equilibrium Model

The route choice equilibrium model is available in the OmniTRANS software. The route choice model is the Multinomial Logit (MNL) model, i.e., the Additive Multinomial (A-MN) model of Chapter 4, see Equation (4.39). The utility formulas are determined internally based on the pricing measures and travel times. The network propagation model is STAQ, as described in Section 3.3. The used node model is the directed capacity proportional node model that solves the (Directed Capacity Proportional node Problem (DCPP)) on page 122. More details can be found in Section 5.4.3.

OmniTRANS reads the O-D trip demands per time-of-day and produces travel times, costs, and distances per time-of-day. These are averaged (weighted by route choice probabilities) per O-D-pair to retrieve the attributes needed in Equation (7.2). To determine the cordon costs per route, the number of passages over the cordon is computed. For the kilometre charge, the distance of each route is computed.

The NRM network was not directly feasible for STAQ, since priorities on roundabouts are not automatically realised in the used node model. That caused gridlocks on roundabouts. Therefore, the capacity of roundabouts (i.e., the four small link arches) have been doubled. This implies that traffic on the roundabout has priority over other traffic.

Input for the model is the flow/demand $f_{(T;r)}$ for both time-of-days $T \in \mathcal{T}$, and all CAR representative routes $r \in \overline{\text{CAR}}$ from Equation (7.6); these are converted to hourly matrices. In addition, freight traffic and non-commuter traffic are included as exogenous fixed O-D-matrices for both time-of-days (PEAK and SHOULDER). These matrices come directly from the NRM-west model. Only route choice is modelled for trucks and non-commuters, which is equal to the commuters' route choice. The passenger car equivalent of a truck is 1.75; this means that every truck is counted as 1.75 cars in the network loading model. Trucks and non-commuters are exogenous in the pricing model, which is a simplification of the case study.

Output of the model are the related variables in the systematic utility for all CAR representative routes $r \in \overline{\text{CAR}}$, i.e., the route costs $\kappa_{(T;r;u)}$ per user-class u[7] and time-of-day T, and the free-flow and congested travel times $\tau^{\text{IVT-FF}}_{(T;r)}$ and $\tau^{\text{IVT-CONG}}_{(T;r)}$ per time-of-day T. These travel impacts are generated for the reference route per O-D-pair. To aggregate the individual route data per O-D-pair, the travel times and travel costs are averaged with weights per route equal to the choice probabilities. In addition, OmniTRANS determines and exports the emission levels per link and area.

7.2.3. Calibration

The total travel demand per user-class, the scale parameters, and the alternative specific constant are calibrated using market shares. Data from the Dutch Central Bureau of Statistics (CBS) is used to determine the market share of each mode - time-of-day combination. The hourly matrix for mode CAR during time-of-day PEAK is obtained from the NRM-west. This matrix acts as the main O-D data source, and is up-scaled to the two total travel demands for each user-class (based on the choice model probabilities).

The used market shares are assumed to represent user-class PEAK-PREF. No figures are available on the number of travellers that have a preference to travel in the peak period. Therefore, it is assumed that 10% of the commuters has no preference for the time-of-day he or she travels in, i.e. they belong to user-group NO-PREF. The other 90% belongs to user-group PEAK-PREF. Almost one third (32%) of the Dutch commuters have the possibility to work from home (Beerepoot and Dijkers, 2013). The commuters that decide to stay-at-home, work at average 6 hours per week from home. We estimate that this leads to the average avoidance of one leg per week (e.g., by working from home one day per two weeks). This leads to a market share of 3.2% for mode HOME. The estimated amount of travellers that avoid the peak varies over studies (Schaap et al., 2014). Based on matrix totals from the NRM, data from the Dutch CBS, and expert judgement, the market shares in Table 7.2 are derived.

Based on these market shares and the $f_{(\text{PEAK};\overline{\text{CAR}})}$ (from NRM-west) several parameters of the choice model are calibrated. By means of the method of least squares, three alternative specific constants (ASC^{CAR}, $\text{ASC}^{\text{SHOULDER}}$, and ASC^{HOME}) and the scale parameter μ are computed. The non-linear optimization toolbox of Matlab has been used to perform the optimization. See Table 7.1 for the results. An additional product of this calibration is the total demand $\boldsymbol{D}_u$ per user-class u. The demands $\boldsymbol{D}_u$ have been post-processed to assure that for each O-D-pair the total demand is either zero or at least one. This is done by geographical grouping and avoids

[7] Note that costs are actually not dependent on user-classes in this case study.

Table 7.1.: Parameter values for the choice model. Each value (except for the scales) can be interpreted as negative willingness-to-pay.

Parameter	Value	Description
$\text{ASC}^{\text{SHOULDER}}$	−3.205	Penalty (in €) for travelling in the shoulder by user-class PEAK-PREF. Obtained by calibration.
ASC^{CAR}	−1.505	Penalty (in €) for travelling by mode CAR. Obtained by calibration.
ASC^{HOME}	−29.91	Penalty (in €) for the stay-at-home mode HOME. Obtained by calibration.
$\text{VOT}^{\text{IVT-FF}}_{(\text{CAR};u)}$	−9.25	Value of time (in €/hour) for in-vehicle time in the car during free-flow conditions. Retrieved from Warffemius (2013).
$\text{VOT}^{\text{IVT-CONG}}_{(\text{CAR};u)}$	−14.245	Value of time (in €/hour) for in-vehicle time in the car during congested conditions. Based on a factor 1.54 on top of the free-flow travel time (Abrantes and Wardman, 2011, Table 13).
$\text{VOT}^{\text{IVT}}_{(\text{TRAIN};u)}$	−11.50	Value of time (in €/hour) for in-vehicle time in the train. Retrieved from Warffemius (2013).
$\text{VOT}^{\text{WAIT}}_{(\text{TRAIN};u)}$	−26.74	Value of time (in €/hour) for waiting for the train. Based on a factor 2.32 on top of the in-vehicle time of the train (Abrantes and Wardman, 2011, Table 14).
$\text{VOT}^{\text{A-E}}_{(\text{TRAIN};u)}$	−16.445	Value of time (in €/hour) for access and egress times to the train. Based on a factor 1.43 on top of the in-vehicle time of the train (Abrantes and Wardman, 2011, Table 13).
μ	3.856	Overall scale. Obtained by calibration.
μ^{MODE}	9.498	Mode-specific scale. Fixed at $\mu/0.406$, where 0.406 is the ratio between general and mode-specific scales in the nested logit structure of the NRM for commuters.

Table 7.2.: Market shares for user-class with a preference to travel in the peak period.

Mode	Time-of-day	Market share
CAR	PEAK	75.97%
CAR	SHOULDER	7.48%
TRAIN	PEAK	12.15%
TRAIN	SHOULDER	1.20%
HOME	·	3.20%

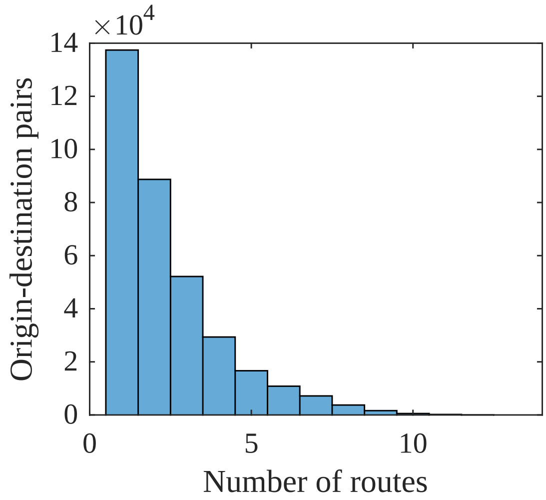


Figure 7.5.: Histogram of the number of routes per O-D-pair.

many computations with small amounts of traffic in the simulation.

The final step of the set-up is to generate routes of each O-D-pair with a positive demand. This is done by means of a Static Traffic Assignment (STA) equilibrium using the *volume averaging* method in OmniTRANS, allowing up to twelve routes per O-D-pair. For the almost 350 000 O-D-pairs with positive demand, a total of almost 840 000 routes has been generated, see Figure 7.5 for the distribution of number of routes. In the calibrated QDNL-model the travellers of the two user-classes experience more than 110 000 loss hours each morning, of which almost 90 000 are experienced in the PEAK period.

7.3. Case Study Problem Formulation

In the previous section almost all ingredients have been presented to describe the overall problem formulation of the transport pricing case study for the Randstad. The lower level and its notation is summarized in Figure 7.6; this is same flow chart as presented earlier, but it now shows the corresponding resources. The last ingredient is the formalization of the effects with their levels and the formulation of the stakeholders' objective functions.

National Government

- **Expected maximum utility of each commuter.** (E-MAX-U) The utility function for each traveller is normalized such that its unit is euro. This utility contains regular travel time, delay, and travel costs. Since travelling is associated with disutility, this component will be negative. By the way, the benefits of travelling are taken into implicitly since a stay-at-home alternative exists with a strong disutilty. Based on the distribution of the maximum utility of multiplicative choice models (see Equation (4.86)), the expected

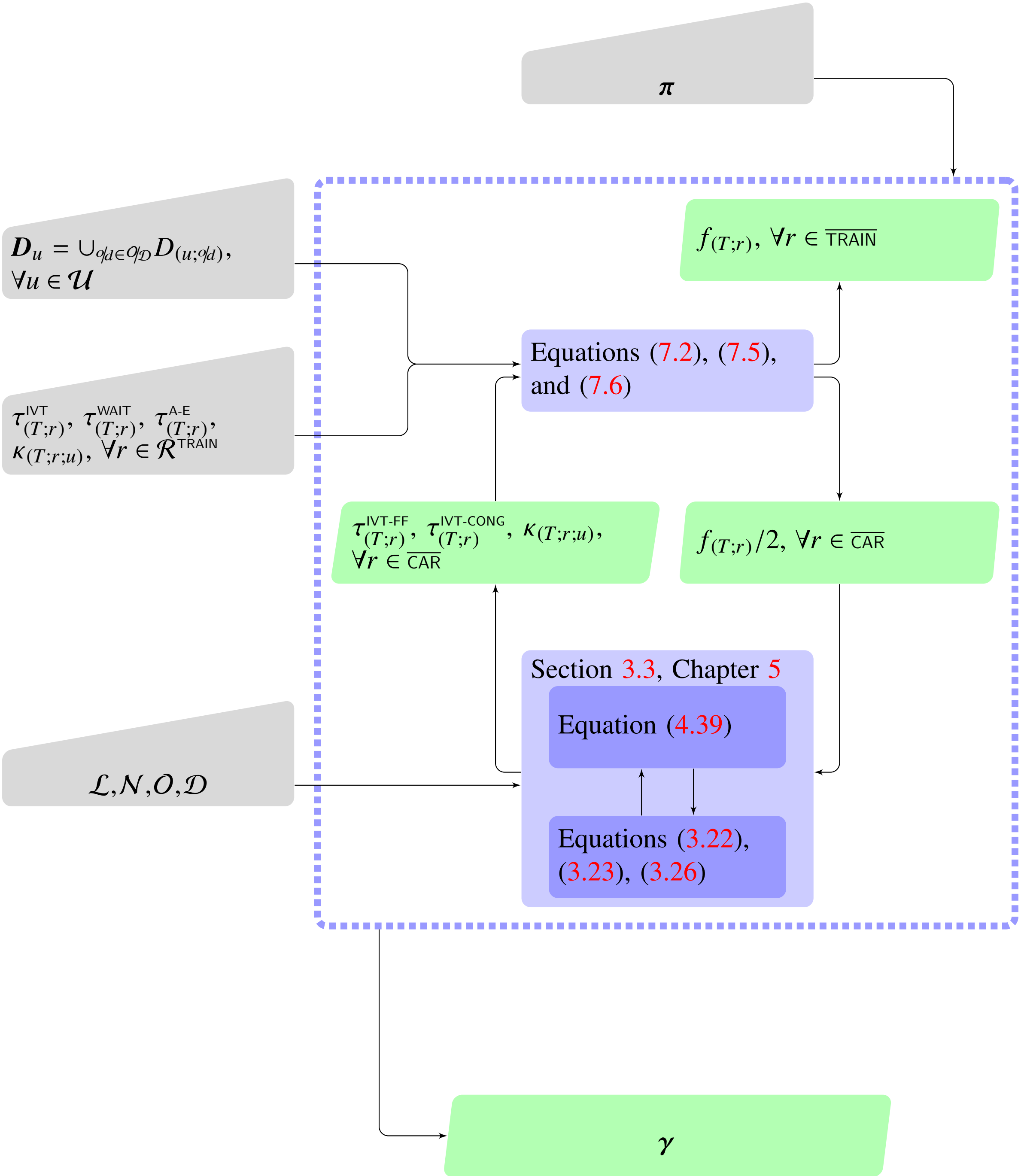


Figure 7.6.: Flowchart identical to the flowchart of Figure 7.4, but now with corresponding variables and equations.

maximum utility can be derived. Denote

$$U^{\max}_{(od;u)} = \mathbb{E} \max_{c \in C_{od}} U_{(c,u)} \tag{7.7}$$

as the expected maximum utility of user-class $u \in \mathcal{U}$ for O-D-pair $od \in \mathcal{OD}$. Its closed formula is provided in the last row and last column of Table 4.1. Then define the effect level as:

$$\gamma_{\text{E-MAX-U}} = \sum_{u \in \mathcal{U}} \sum_{od \in \mathcal{OD}} U^{\max}_{(od;u)} D_{(u,od)}. \tag{7.8}$$

- **Income from the kilometre charge.** (KM-INC) Assume that the revenue of the kilometre charge is recycled in the transport system. For example, by lowering other taxes, such as the annual registration fee.[8] The charge accounts negatively in the expected maximum utility of the travellers. But since it is recycled, it has a neutral influence on social welfare. Therefore, this component is introduced to compensate for the disutility. Let $l_{(T;r)}$ be the length of route $r \in \overline{\text{CAR}}$ (the weighted average from the route choice equilibrium model) for time-of-day T (it depends on the time-of-day due to the weighting), then the effect level is denoted as

$$\gamma_{\text{KM-INC}} = \sum_{r \in \overline{\text{CAR}}} \sum_{T \in \mathcal{T}} l_r f_{(T;r)}. \tag{7.9}$$

- **Income from the cordon charge.** (CRDN-INC) Similar as above. In spite of the fact that the municipality controls the cordon charge, also these revenues are recycled. Therefore also the cordon charge is neutral with respect to social welfare. Let $n_{(T;r)}$ be the number of times route $r \in \mathcal{R}$ enters the cordon for time-of-day T (again the weighted average over all O-D-pair routes), then the effect level $\gamma_{\text{CRDN-INC}}$ is computed by:

$$\gamma_{\text{CRDN-INC}} = \sum_{r \in \overline{\text{CAR}}} \sum_{T \in \mathcal{T}} n_r f_{(T;r)}. \tag{7.10}$$

- **Total emissions in the study area.** (EMIS-STUDY) In the study area, emissions of vehicular traffic are modelled according to the ARTEMIS model (see Wismans (2012) for more details on this emission model). The detriment per emission type is €26.60 per tonne Carbon Monoxide (CO), €25.00 per tonne Carbon Dioxide (CO_2), €10.60 per kilogram Nitrogen Oxides (NO_X), and €64.80 per kilogram Particle Matter < 10 micrometre (PM_{10}) (according to de Bruyn et al., 2010, Table 50). Denote its effect level with $\gamma_{\text{EMIS-STUDY}}$.
- **Total value of loss hours for the other travellers.** (LH-OTHER) The choice behaviour of non-commuters is not addressed in this case study. This is a small group of travelers that is exogenous to the choice model, but that does experience travel time. This group is therefore neither included in the maximum expected utility. To compensate for this the value of travel time of this group is also included separately. Let $\rho_{(T;r)}$ be the number of other travellers for the O-D-pair that belongs to $r \in \mathcal{R}$ during time-of-day T, then the effect level $\gamma_{\text{LH-OTHER}}$ is computed by:

$$\gamma_{\text{LH-OTHER}} = \sum_{r \in \overline{\text{CAR}}} \sum_{T \in \mathcal{T}} \rho_{(T;r)} \times \text{VOT}^{\text{IVT-CONG}}_{(\text{CAR};u)} \times \tau^{\text{IVT-CONG}}_{(T;r)}. \tag{7.11}$$

[8] Assume that the recycling does induce a change in travel behaviour.

Finally, the objective function of GOV is rather simple because all effect levels are already monetized. It equals

$$H_{\text{GOV}}(\boldsymbol{\gamma}) = \gamma_{\text{E-MAX-U}} + \gamma_{\text{KM-INC}} + \gamma_{\text{CRDN-INC}} + \gamma_{\text{EMIS-STUDY}} + \gamma_{\text{LH-OTHER}}. \tag{7.12}$$

Municipality of Amsterdam

- **Accessibility: value of loss hours of traffic towards Amsterdam.** (ACCESS) For all traffic that has its destination within Amsterdam, the delay is monetized. Since the morning commute is considered, this is a good indication of the accessibility of the city. Let ρ_r be as above and let $\mathcal{OD}^{\text{AMS}} \subset \mathcal{OD}$ be the O-D-pairs for which it holds that destination d is within Amsterdam, and define the effect level as

$$\gamma_{\text{ACCESS}} = \sum_{u \in \mathcal{U}} \sum_{od \in \mathcal{OD}^{\text{AMS}}} \sum_{\{(T;r) \in C^{od} | r \in \overline{\text{CAR}}\}} (\rho_{(T;r)} + P_{(T;r;u)} D_{(u;od)}) \times \text{VOT}^{\text{IVT-CONG}}_{(\text{CAR};u)} \times \tau^{\text{IVT-CONG}}_{(T;r)}. \tag{7.13}$$

- **Loss of €15.- for missed economic activity.** (LOSS) When commuters decide to work from home, less economic activity occurs in the city. Less office space is occupied and less expenses are made within the city. The value of this loss (€15.-) is a coarse estimation by the author. It is only assigned to travellers with (intentional) destination Amsterdam. Let $\mathcal{OD}^{\text{AMS}}$ as above, and define the effect level as

$$\gamma_{\text{LOSS}} = \sum_{u \in \mathcal{U}} \sum_{od \in \mathcal{OD}^{\text{AMS}}} \sum_{\{(T;r) \in C^{od} | r = \text{R}_0\}} 15 \times P_{(T;r;u)} D_{(u;od)}. \tag{7.14}$$

- **Total emissions within Amsterdam.** (EMIS-AMS) Similar as for social welfare, but now only for traffic within the city's boundaries. Also here the detriment per emission type is €26.60 per tonne CO, €25.00 per tonne CO_2, €10.60 per kilogram NO_X, and €64.80 per kilogram PM_{10} (according to de Bruyn et al., 2010, Table 50). Denote its effect level with $\gamma_{\text{EMIS-AMS}}$.

Finally, the objective function of AMS equals

$$H_{\text{AMS}}(\boldsymbol{\gamma}) = \gamma_{\text{ACCESS}} + \gamma_{\text{LOSS}} + \gamma_{\text{EMIS-AMS}}. \tag{7.15}$$

Train Operator

- **Income from ticket sales.** (TRAIN-INC) The income is based on the number of travellers, the original price in NRM-west, and the price factor. Let $\tilde{\pi}_r$ be the current fare for route $r \in \overline{\text{TRAIN}}$, then the total income is denoted as

$$\gamma_{\text{TRAIN-INC}} = \sum_{r \in \overline{\text{TRAIN}}} \pi_{\text{FARE-ON}} \tilde{\pi}_r f_{(\text{PEAK};r)} + \pi_{\text{FARE-OFF}} \tilde{\pi}_r f_{(\text{SHOULDER};r)}. \tag{7.16}$$

- **Costs.** (TRAIN-COST) Using the status quo, marginal costs for passenger kilometres during the peak and shoulder periods are derived. From the assumption that spare capacity exists in the shoulder, and that creating additional capacity in the peak period is costly, the marginal costs per passenger kilometre are €0.14 in the peak and €0.04 in the shoulder. It is assumed that currently no loss or profit is made by setting $\gamma_{\text{TRAIN-COST}} = \gamma_{\text{TRAIN-INC}} := \gamma^0_{\text{TRAIN-COST}}$, i.e., the income under the zero-scenario. And let $f^0_{(T;r)}$ be the flows under the scenario for route $r \in \overline{\text{TRAIN}}$ during time-of-day $T \in \mathcal{T}$. This effect level is then defined as

$$\gamma_{\text{TRAIN-COST}} = \left(\sum_{r \in \overline{\text{TRAIN}}} \sum_{T \in \mathcal{T}} \left(f_{(T;r)} - f^0_{(T;r)} \right) \times (0.14 \delta_T^{\text{PEAK}} + 0.04 \delta_T^{\text{SHOULDER}}) \right) - \gamma^0_{\text{TRAIN-COST}}. \quad (7.17)$$

The difference between income and cost, i.e., profit, is the objective function of TR-OP:

$$H_{\text{TR-OP}}(\boldsymbol{\gamma}) = \gamma_{\text{TRAIN-INC}} + \gamma_{\text{TRAIN-COST}}. \quad (7.18)$$

The corresponding MSP of the case study can now be formulated, and equals:

$$\left\{ \max_{\pi_{p_{\text{GOV}}} \in \Pi_{p_{\text{GOV}}}} H_{\text{GOV}}(\boldsymbol{\gamma}); \max_{\pi_{p_{\text{AMS}}} \in \Pi_{p_{\text{AMS}}}} H_{\text{AMS}}(\boldsymbol{\gamma}); \max_{\pi_{p_{\text{TR-OP}}} \in \Pi_{p_{\text{TR-OP}}}} H_{\text{TR-OP}}(\boldsymbol{\gamma}) \right\} \quad (7.19)$$
$$\text{subject to } \boldsymbol{\gamma} = E(\boldsymbol{\pi}),$$

Note that by applying Equations (3.27) and (3.29), function E can be captured in a variational inequality problem formulation. The next section will analyse this problem formulation.

7.4. Results

For all 216 possible pricing schemes the underlying traffic assignment model has been simulated. The calculation time to reach an equilibrium was about 23 hours for each pricing, and it was possible to analyse three schemes in parallel on an Intel i5 3.2 GHz, 16GB RAM machine. First, consider the results when only a single stakeholder is active. Tables 7.3, 7.4 and 7.5 show the effect of the respectively stakeholder GOV's, AMS's and TR-OP's pricing measure on its objective function. Note that the shown results are relative to the zero scenario in the remainder of this chapter, i.e. $H_s(\boldsymbol{\pi}) := H_s(\boldsymbol{\pi}) - H_s(\boldsymbol{\pi}^0), \forall s \in \mathcal{S}$, with $\boldsymbol{\pi}^0 \equiv 0$. Also note that these first results assume that the other stakeholders do not act. All results represent one morning commute.

For each of the stakeholders, time differentiated price levels are more beneficial. The national government is best off with a peak charge of €0.10 and a shoulder charge of €0.05. For the municipality of Amsterdam it holds that a peak hour cordon charge of €8 combined with a €4 cordon charge during the shoulder period leads to highest improvement in its economic position. Finally, the profit of the train operator is maximized when the fare is increased with 20 % in the peak period only. In general, one could state time differentiated prices are better than constant prices. This indicates that there is indeed spare capacity in the transport system during the shoulder period, and that it is beneficial to give an incentive to travellers to switch to the shoulder in order to reduce external effects in the peak period. Another remark is that for stakeholders GOV and TR-OP it is beneficial to employ their maximum pricing level during the peak period; it could be interesting to investigate even higher prices, although those might not be acceptable. These basic results do not show what happens when they are introduced simultaneously.

Table 7.3.: Social welfare improvement by introducing a kilometre charge (in k€)

	KM-ON		
KM-OFF	€0.00	€0.05	€0.10
€0.00	0	901	1 161
€0.05	—	572	1 202
€0.10	—	—	795

Table 7.4.: Improvement of the economic position of Amsterdam by introducing a cordon charge (in k€)

	CRDN-ON		
CRDN-OFF	€0.00	€4.00	€8.00
€0.00	0	68	47
€4.00	—	47	69
€8.00	—	—	51

Table 7.5.: Shift in profit by adjusting a train fares (in k€)

	FARE-ON		
FARE-OFF	−20%	0%	20%
−20%	−174	2	130
0%	—	0	139
20%	—	—	135

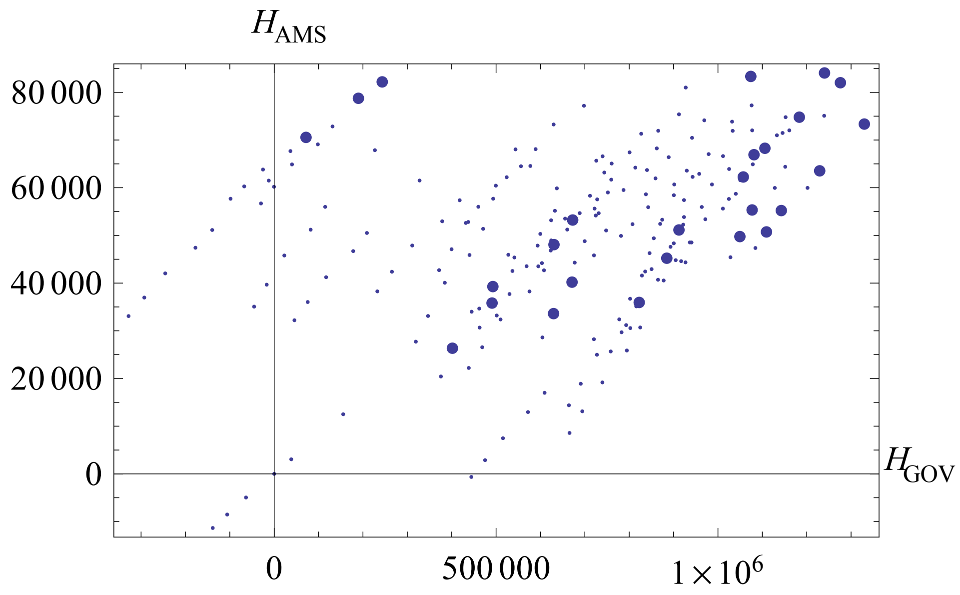


Figure 7.7.: Scatter plot to indicate relation between objective functions (in €). Large dots lie in the Pareto frontier based on all three stakeholders.

Figures 7.7, 7.8 and 7.9 each show scatter plots of two out of three objective functions based on the pricing schemes' simulation results. Also the Pareto-optimal solutions are shown; these are the point for which no objective can be improved without worsening any of the other objectives. This points form the so-called Pareto-frontier, which – contrary to the Nash-equilibrium – does not account for which stakeholder controls which pricing scheme. However, it does give a good overview of which pricing schemes are 'good' and 'stable'. With three variables, the Pareto-frontier is a (curved) plane in $\mathbb{R}^3$. Figures 7.7, 7.8 and 7.9 provide frontal, side and top views of this space. Notice that for $H_{\text{TR-OP}}$ constantly three groups (or clouds) of points can be identified; these are constituted by the train fare during peak hours. Additionally, H_{GOV} and H_{AMS} are positively correlated, while $H_{\text{TR-OP}}$ is negatively correlated with both other stakeholders' objective functions. This negative correlation points at conflicting interests of the train operator with the other stakeholders.

7.4.1. The TU-game

Next, the theory of Chapter 6 is applied. Table 7.6 provides an overview of the results of all coalitions for all relevant pricing schemes; the first columns define the pricing schemes, and the latter columns provide the combined objective function values for coalitions. A pricing scheme is called relevant (and is thus included) if it provides the overall maximum objective of a coalition, or if it is used to derive the TU-game coalition values (i.e., when a Nash-equilibrium exists for the pricing scheme). The rows are sorted on the objective function of the grand coalition. So, the first row shows the resulting pricing scheme $\boldsymbol{\pi}^*$. A total sum of 1 397 k€ can be allocated by the cooperative solution concepts. The results table shows again

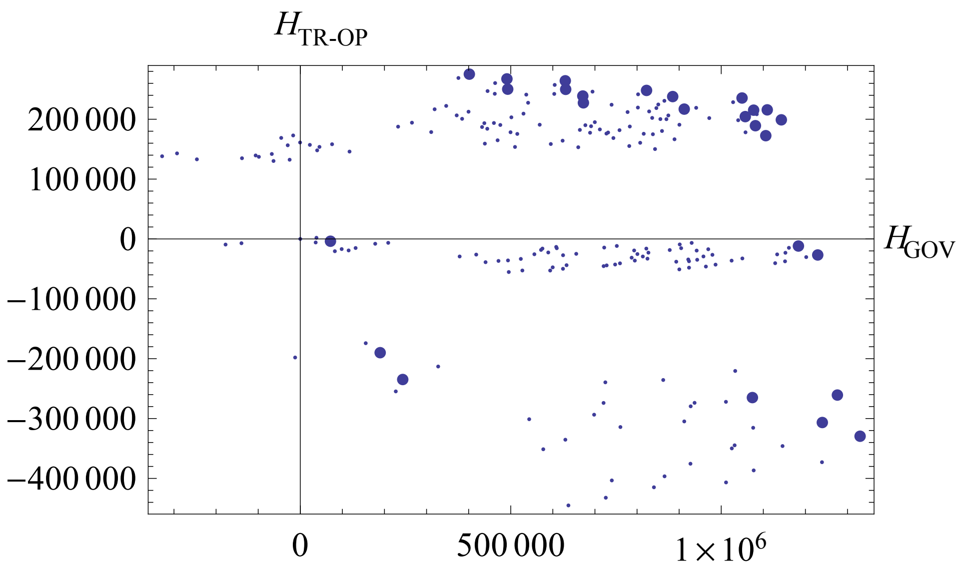


Figure 7.8.: Scatter plot to indicate relation between objective functions (in €). Large dots lie in the Pareto frontier based on all three stakeholders.

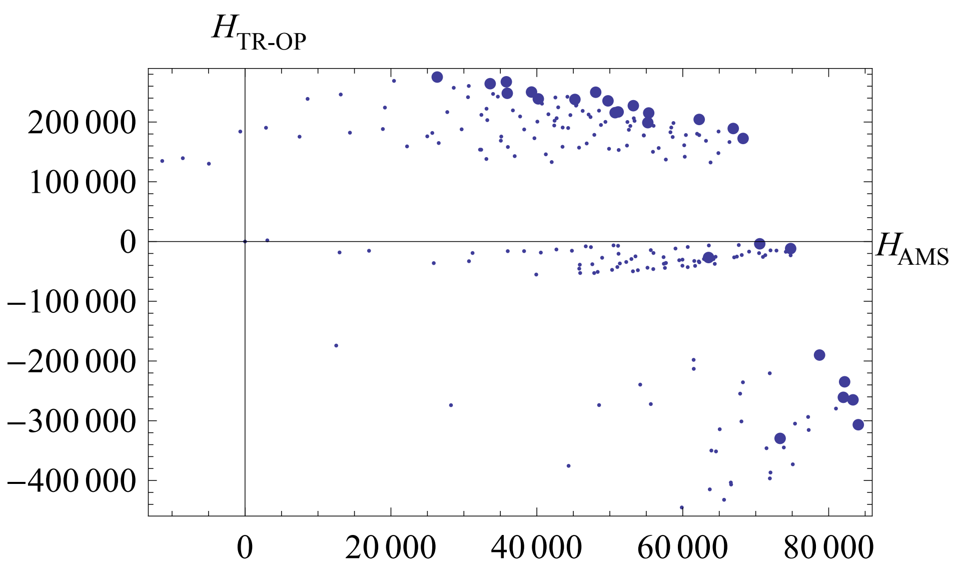


Figure 7.9.: Scatter plot to indicate relation between objective functions (in €). Large dots lie in the Pareto frontier based on all three stakeholders.

Table 7.6.: Overview of the most important pricing schemes

Pricing schemes						Coalition objectives $\sum_{s \in C} H_s(\boldsymbol{\pi})$ (in k€) for C =						
$\pi_{\text{KM-ON}}$ (in €)	$\pi_{\text{KM-OFF}}$ (in €)	$\pi_{\text{CRDN-ON}}$ (in €)	$\pi_{\text{CRDN-OFF}}$ (in €)	$\pi_{\text{FARE-ON}}$ (factor)	$\pi_{\text{FARE-OFF}}$ (factor)	{GOV}	{AMS}	{TR-OP}	{GOV, AMS}	{GOV, TR-OP}	{AMS, TR-OP}	S
0.10	0.05	0	0	1.2	0.8	1 143	55	199	1 198	**1 342**	254	**1 397**
0.10	0.05	0	0	1.2	1	1 109	51	216	1 160	1 325	266	1 376
0.10	0.05	4	4	1.2	0.8	1 077	55	215	1 132	1 292	270	1 347
0.10	0.00	4	4	1.2	1	1 057	62	204	1 119	1 261	267	1 324
0.10	0.05	0	0	0.8	0.8	**1 330**	73	−330	**1 403**	1 000	−256	1 074
0.10	0.00	4	4	0.8	0.8	1 240	**84**	−307	1 324	934	−223	1 018
0.10	0.10	8	4	1.2	1	491	36	267	527	758	**303**	794
0.10	0.10	8	8	1.2	1	402	26	**275**	428	677	302	703

that stakeholder TR-OP's interests are conflicting with the other two.

Using Definitions 6.1 and 6.2 the stability and equilibrium conditions of the objective values are determined. Table 7.7 shows all stable coalitions per pricing scheme and Table 7.8 shows all Nash-equilibria.

The second row shows, amongst others, the Nash-equilibrium with competition between all stakeholders. When it is assumed no cooperation exists between stakeholders and TU-game theory is omitted, then this pricing scheme is the resulting scheme. It is remarkable that (1) the only strategic difference between the cooperative and non-cooperative solutions is the train fare in the shoulder, and (2) that no cordon charge will be implemented around Amsterdam. The latter is due to the large improvement of the city's accessibility and economic position by solely implementing the kilometre charge. The additional delay reduction by a cordon charge is not profitable against the additional economic loss by the increasing amount of travellers that decide not to travel towards Amsterdam.

Table 7.9 shows the TU-game with the coalition values based on Definition 6.3. Only the first four rows will be used to determine the coalition values. For each partition, and thus for each coalition, at least one Nash-equilibrium exists.[9] Therefore all coalition values can be retrieved from the simulation results, and no coaltion will be assigned with zero value. For coalition {GOV} and {AMS} the minimum over partition' Nash-equilibria actually compared multiple pricing schemes (respectively over rows 2&4 and rows 2&3). The second row corresponds with each for coalition {TR-OP} relevant partition's Nash-equilibrium. All other coalitions appear in only one partition.

For the calculation of the solution concepts it is more convenient to work with zero-normalized

[9]This implies that no issues with existence of equilibria are apparent for the case study.

Table 7.7.: Stable coalitions per pricing scheme

Pricing schemes						Pricing scheme stability for coalition C =						
$\pi_{\text{KM-ON}}$ (in €)	$\pi_{\text{KM-OFF}}$ (in €)	$\pi_{\text{CRDN-ON}}$ (in €)	$\pi_{\text{CRDN-OFF}}$ (in €)	$\pi_{\text{FARE-ON}}$ (factor)	$\pi_{\text{FARE-OFF}}$ (factor)	{GOV}	{AMS}	{TR-OP}	{GOV, AMS}	{GOV, TR-OP}	{AMS, TR-OP}	S
0.10	0.05	0	0	1.2	0.8	✓			✓	✓		✓
0.10	0.05	0	0	1.2	1	✓	✓	✓	✓			
0.10	0.05	4	4	1.2	0.8		✓			✓		
0.10	0.00	4	4	1.2	1	✓		✓			✓	
0.10	0.05	0	0	0.8	0.8	✓			✓			
0.10	0.00	4	4	0.8	0.8	✓	✓					
0.10	0.10	8	4	1.2	1			✓			✓	
0.10	0.10	8	8	1.2	1			✓				

Table 7.8.: Nash-equilibria per pricing scheme

Pricing schemes						Nash-equilibrium for partition Q =				
$\pi_{\text{KM-ON}}$ (in €)	$\pi_{\text{KM-OFF}}$ (in €)	$\pi_{\text{CRDN-ON}}$ (in €)	$\pi_{\text{CRDN-OFF}}$ (in €)	$\pi_{\text{FARE-ON}}$ (factor)	$\pi_{\text{FARE-OFF}}$ (factor)	{{GOV}, {AMS}, {TR-OP}}	{{GOV, AMS}, {TR-OP}}	{{GOV, TR-OP}, {AMS}}	{{AMS, TR-OP}, {GOV}}	{S}
0.10	0.05	0	0	1.2	0.8					✓
0.10	0.05	0	0	1.2	1	✓	✓			
0.10	0.05	4	4	1.2	0.8			✓		
0.10	0.00	4	4	1.2	1				✓	
0.10	0.05	0	0	0.8	0.8					
0.10	0.00	4	4	0.8	0.8					
0.10	0.10	8	4	1.2	1					
0.10	0.10	8	8	1.2	1					

Table 7.9.: The TU-game coalition values and the zero-normalized game $\bar{v}$

C	$v(C)$	$\bar{v}(C)$
$\emptyset$	0	0
{GOV}	1 057	0
{AMS}	51	0
{TR-OP}	216	0
{GOV, AMS}	1 160	0
{GOV, TR-OP}	1 292	19
{AMS, TR-OP}	267	52
$\mathcal{S}$	1 397	73

games in which every single-stakeholder coalition has zero value. Therefore, Table 7.9 also provides the equivalent zero-normalized game $(\mathcal{S},\bar{v})$, with $\bar{v}(C) := v(C) - \sum_{s\in C} v(\{s\}), \forall C \subseteq \mathcal{S}$.

7.4.2. Cooperative solution concepts

The cooperative solution concepts for TU-game $(\mathcal{S},v)$ as defined in Table 7.9 lead to possible allocations $(\chi_{\text{GOV}},\chi_{\text{AMS}},\chi_{\text{TR-OP}}) \in \mathbb{R}^3$. The Shapley and compromise value are computed based on zero-normalized game $(\mathcal{S},\bar{v})$; solution concept on original and zero-normalized games are equivalent. If $\bar{\chi}$ is an allocation of $(\mathcal{S},\bar{v})$, then the corresponding $(\mathcal{S},v)$-game allocation equals $\chi_s = \bar{\chi}_s + v(\{s\}), \forall s \in \mathcal{S}$.

The core of the original game equals

$$\begin{aligned} K(\mathcal{S},v) = \{\chi | \\ \chi_{\text{GOV}} + \chi_{\text{AMS}} + \chi_{\text{TR-OP}} &= 1\,397, \\ \chi_{\text{GOV}} &\geq 1\,057, \\ \chi_{\text{AMS}} &\geq 51, \\ \chi_{\text{TR-OP}} &\geq 216, \\ \chi_{\text{GOV}} + \chi_{\text{AMS}} &\geq 1\,160, \\ \chi_{\text{GOV}} + \chi_{\text{TR-OP}} &\geq 1\,292, \\ \chi_{\text{AMS}} + \chi_{\text{TR-OP}} &\geq 267\}. \end{aligned} \tag{7.20}$$

The Shapley value computation for the zero-normalized game is presented in Table 7.10, which also presents the Shapley value of the original game. Determining upper and lower bound before applying Equation (6.10) to compute the compromise value leads to the results in Table 7.11 (with $\alpha = 36/73$).

With exactly three stakeholders elegant visualisations of the solution concepts can be made by drawing the intersection of $\mathbb{R}^3$ with efficient allocations. This is done in Figure 7.10, where

Table 7.10.: Shapley value computation of TU-game $(\mathcal{S},\bar{v})$

Ordering	Marginal contributions GOV	AMS	TR-OP
GOV → AMS → TR-OP	0	52	21
GOV → TR-OP → AMS	0	54	19
AMS → GOV → TR-OP	52	0	21
AMS → TR-OP → GOV	73	0	0
TR-OP → GOV → AMS	19	54	0
TR-OP → AMS → GOV	73	0	0
$\zeta(\mathcal{S},\bar{v})$	$36\frac{1}{6}$	$26\frac{2}{3}$	$10\frac{1}{6}$
$\zeta(\mathcal{S},v)$	$1\,093\frac{1}{6}$	$77\frac{2}{3}$	$226\frac{1}{6}$

Table 7.11.: Compromise value computation of TU-game $(\mathcal{S},\bar{v})$ for each stakeholder $s \in \mathcal{S}$

s	Ω	ω	$\eta(\mathcal{S},\bar{v})$	$\eta(\mathcal{S},v)$
GOV	73	0	$36\frac{1}{148}$	$1\,093\frac{1}{148}$
AMS	54	0	$26\frac{47}{74}$	$77\frac{47}{74}$
TR-OP	21	0	$10\frac{53}{148}$	$226\frac{53}{148}$

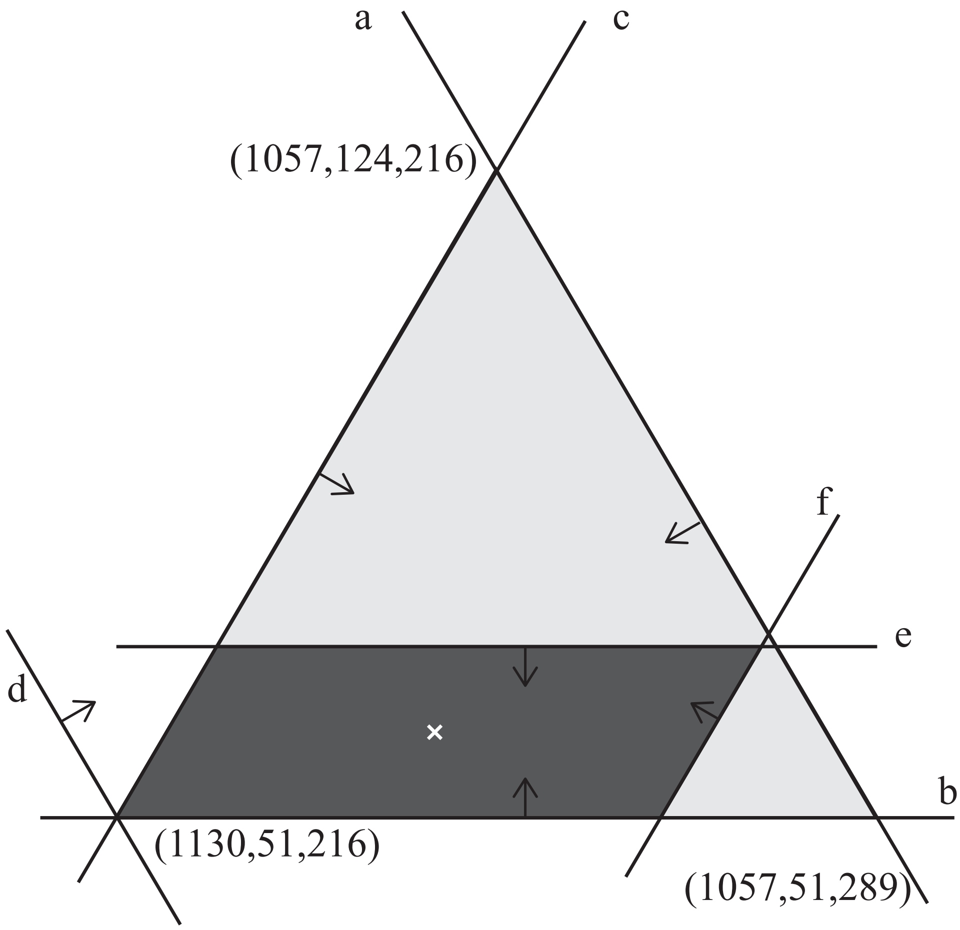


Figure 7.10.: Cooperative game theory solution concepts presented in the plane in $\mathbb{R}^3$ defined by $\sum_{s\in\mathcal{S}} \chi_s = v(\mathcal{S})$. The light grey area represents all efficient and stakeholder rational allocations (i.e., the so-called imputations); the dark grey area represents the core; the white crosses lie very close to each other and represent the Shapley and compromise values. The lines represent the inequalities that define the core.

the labeled lines correspond with coalition dissidence constraints from the core (Equation 7.20)): a→ {GOV}, b→ {AMS}, c→ {TR-OP}, d→ {AMS,TR-OP}, e→ {GOV,TR-OP}, and f→ {GOV,AMS}. The light grey triangle, determined by lines a, b and c, represents the so-called imputations. In every corner one of the stakeholder receives it maximal achievable allocation when the others claim their minimum. These allocation are labelled $(\chi_{\text{GOV}}, \chi_{\text{AMS}}, \chi_{\text{TR-OP}})$. The set of imputations does not consider coalitions. As shown, the core in this game is not empty, which means that there are stable allocations. The Shapley value and the compromise value are almost equal and also depicted in the plot.

As discussed in Section 6.4, both the Shapley and compromise value can be derived from axioms and/or fairness criteria. For this case study the Shapley value lies within the core, which is desirable, but not true in general. The fact that $\zeta(\mathcal{S},v)$ and $\boldsymbol{\eta}(\mathcal{S},v)$ are almost equal, makes this allocation very favourable. The fairness criteria of both values hold. So, the average marginal contributions and the compromise between the upper and lower bound are almost similar. When this allocation is compared to the actual objectives retrieved by resulting pricing scheme $\boldsymbol{\pi}^*$ (Table 7.6, first row), then it becomes clear that GOV has to pay around 50 k€ to the other stakeholders (23 k€ to AMS, 27 k€ to TR-OP).

Some other characteristics of the resulting pricing scheme $\boldsymbol{\pi}^*$ – compared to the zero-scenario – is that the loss hours decline with 60% in the peak period and 45% overall; the emissions inside the study area declines with about 6%; stay-at-home alternative mode HOME increases with 60%; and there is an increase of 58% of train passengers. These results do of course come with the disclaimer that the purpose of the case study is to show the working of the applied methods, and to get insight in how stakeholders interact. Some assumptions need to be released and more details need to be modelled, before the case study is a direct tool to assist Dutch policy makers.

7.4.3. The Price of Competition

Finally, consider the price of competition for this case study. This concept is similar to the price of of anarchy as discussed in Section 2.2.3. The price of competition can be interpreted as the price of anarchy with respect to cooperation and competion between stakeholder (i.e., decision makers). This is different from the original price of anarchy that analyses the ratio between the user equilibrium and system optimum assignments, which is not considered here. One could say that the price of competition resembles the *price of stakeholder anarchy*, versus the original *price of travellers anarchy*.

Assume that in the cooperative game the stakeholders accept the single point solution concepts and that they agree on allocation $\boldsymbol{\chi}^* = (1\,093, 78, 226)$, and remember that resulting pricing scheme $\boldsymbol{\pi}^*$ equals the first row of Table 7.6. If there was no cooperation whatsoever, then competition would lead to the Nash-equilibrium on the second row of 7.6; denote the corresponding pricing scheme for competition as $\tilde{\boldsymbol{\pi}}$. Although that the difference between $\boldsymbol{\pi}^*$ and $\tilde{\boldsymbol{\pi}}$ is small, the difference in outcomes is substantial. Table 7.12 summarizes these results.

The overall price of competition is small in the relative sense (101.45%); however, still 18 k€ is saved every morning and similar results might apply to the afternoon/evening. Remarkable are the prices of competition per stakeholder. TR-OP can improve with about 5% under cooperation. AMS on the other, has rather low resulting objective functions compared to the

Table 7.12.: The cooperative solution, the non-cooperative solution, and the price of competition for each stakeholder $s \in \mathcal{S}$

s	$H_s(\boldsymbol{\pi}^*)$	χ^*	$H_s(\tilde{\boldsymbol{\pi}})$	Price of Competition
GOV	1 143	1 093	1 109	98.56%
AMS	55	78	51	150.98%
TR-OP	199	226	216	104.63%
Total	1 397	1 397	1 379	101.45%

resulting allocation, and has a large price of competition of 150.98%. Finally, the most remarkable result is that GOV does not benefit from cooperation and has a price of competition lower than 100%. The questions that rise are: "How is it possible that χ^*_{GOV} is smaller than $H_{\text{GOV}}(\tilde{\pi}^*)$? " and "Why does GOV accept this allocation?". The answers lie in the coalition that can be formed by AMS and TR-OP. In the non-cooperative equilibrium, this coalition is not allowed to be formed because all stakeholders compete. This causes GOV to end up in a relatively good situation. In the cooperative setting coalition {AMS, TR-OP} is not stable for $\tilde{\boldsymbol{\pi}}$; therefore, GOV can not guarantee its objective value that corresponds to $\tilde{\boldsymbol{\pi}}$ when the others cooperate.

7.5. Synthesis and discussion

The case study presented in this chapter has formulated and analysed a bi-level MSP for the Randstad area in the Netherlands. Three stakeholders were involved that want to implement three different pricing measures (a kilometre charge, a cordon charge, a train fare adjustment). The upper level has used the cooperative game theory framework presented in Chapter 6. To map pricing schemes onto effect levels in the lower level a comprehensive traffic assignment model has been used. The basis for this is QDTA as presented in Chapter 3. The travellers' response to the pricing measures was captured in terms of mode choice, time-of-day choice, and route choice with GMEV models (see Chapter 4). Amongst other, the external effects emissions and delay were computed in detail; this has allowed a fairly detailed calculation of the welfare effects.

The main results of the case study are the core, Shapley value and compromise value. Both the values lie inside the core – meaning that no stakeholder or coalition has a direct objection – and they are almost equal. So this allocation satisfies multiple fairness criteria. In addition, the non-cooperative Nash-equilibrium (i.e., the result under competition) and an analysis of the price of competition, which is not significant, is provided. A remarkable result is that one of the stakeholders, the national government, is worse off under cooperation. Paradoxically, cooperation leads to a social welfare reduction due to the power of the other stakeholders.

It has become clear that the national government and the train operator have conflicting interests. However, under cooperation its combined objective is maximized. The insights based on the game theoretical solution concepts would not have been retrieved from traditional social welfare optimization. In addition, more realism is obtained with the chosen assignment model, and the calculation of delay and emissions. As much as possible, data from the NRM-west-model is used; this model is extensively estimated and calibrated.

However, the case study uses simplifying assumptions and has hypothetical objectives of stakeholders. For this reason, the policy implications have to be regarded in the correct context. Some important simplyfying assumptions include: (1) The population is differentiated into only two user-classes; therefore, taste heterogeneity between commuters is limited. In future studies, (simultaneous) choice models should be estimated with more user-classes to provide more realism. (2) Future studies should also have feedback from public transport crowding in the choice models. (3) The objective functions, especially that of the stakeholder TR-OP should be more detailed, and they should be set up in dialogues with each stakeholder. (4) The costs of implementing and operating new transport pricing systems has not been considered. These costs could be a substantial, and could therefore suppress welfare gains.

CHAPTER 8.

CONCLUSIONS AND DISCUSSION

This final chapter summarizes the findings and conclusions of the presented research by assessing to which extent the requirements for strategic planning models for passenger transport pricing are satisfied. The tools provided in this dissertation can be used to satisfy these requirements. The added value of the research for practice and policy makers is stated separately, as well as the methodological contributions. Finally, future directions for strategic network modelling for passenger transport pricing are provided that show which challenges still exist.

The introductory chapter described how the transport system functions inefficiently and suboptimally. By travelling, people put others at disadvantages, like congestion and exposure to emissions. If travellers would make different choices, e.g., if they would avoid peak periods or use more public transport instead of the car, the total external effects and their impacts would be smaller. By providing incentives for travellers – through innovative pricing strategies – to change their behaviour, the performance of the transport system could improve. By innovative pricing, travellers can be made responsible for their caused effects. However, practically applicable strategic planning models that could support the decision making process lack realism at several aspects. In addition, public and political support for innovative pricing policies is often low.

The research presented in this dissertation addresses aspects of strategic planning models for transport pricing. The aim has been to increase realism of existing strategic transport planning tools, while at the same time not significantly increasing the computational complexity of these models in order to keep them practically applicable on large scale networks. This could improve the credibility of strategic planning models, and could increase public and political support. The presented holistic framework allows to incorporate most relevant travel responses to pricing measures, and a stakeholder model is included that can identify potential conflicts of interest between multiple stakeholder. The latter game theoretical model also provides solution concepts to settle identified conflicts.

This dissertation has adopted the for transport pricing widely used bi-level modelling framework presented in Chapter 2, see Figure 8.1. The two levels correspond with the decision making processes of respectively the stakeholders and the travellers. The upper level (presented in

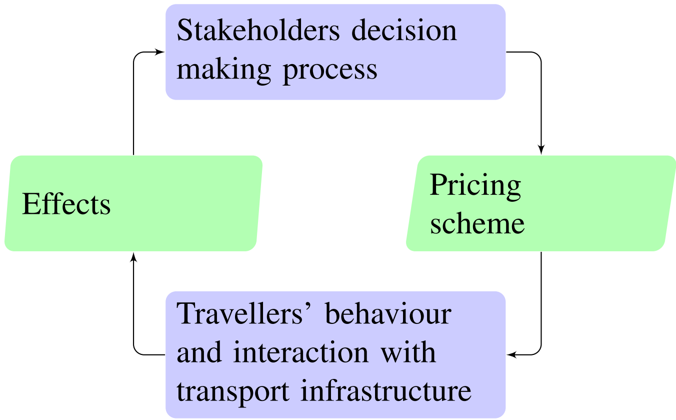


Figure 8.1.: Bi-level model: schematic overview

Part II) represents the stakeholders (i.e., the decision makers or price setters). The stakeholders all have objectives based on several effects (e.g., congestion, emissions). They set their strategy such that their objective will be optimized. To do this they have to take the responses of the travellers into account, as well as the actions of the other stakeholders. This problem formulation extends the traditional transport pricing problems with only one stakeholder, and is resolved with Transferable Utility (TU)-game theory. This theory provides multiple solution concepts.

The lower level describes (presented in Part I) the behaviour of travellers and their interaction with the transport infrastructure. A Quasi-Dynamic Traffic Assignment (QDTA) approach is adopted here that is used to quantify the effect of the responses of the travellers towards a pricing scheme. It consists of a travel behaviour component, and a network loading component.

The first component describes and simulates how travellers make their choices. This will ultimately be expressed in different traffic flows. The component makes use of Generalized Multivariate Extreme Value (GMEV) models that consider mode, route and time-of-day choice simultaneously. For the network loading component we use Static Traffic Assignment with Queuing (STAQ) as a Quasi-Dynamic Network Loading (QDNL) procedure in OmniTRANS that propagates traffic flows by applying capacity constraints following from a proper node model. Both components are applied in an iterative fashion to establish stable solution called a user equilibrium in which no traveller can improve their utility by unilaterally changing route, mode, or time-of-day it travels.

The upper and lower level meet in the case study for the Randstad area presented in Chapter 7. The study considers three stakeholders, contains time-differentiated pricing measures, contains multiple modes, contains a large road network, considers two user-classes, and compares multiple solution concepts. The case study illustrates the application of developed methodologies, and feasibility of 'large-scale' modelling for transport pricing.

8.1. Assessment of the Requirements

In Sections 2.5.1 and 2.5.2 several requirements for the bi-level strategic planning model have been identified. This section repeats those requirements and assesses how they can be satisfied in the provided modelling framework. The requirements were categorized into three types: the upper level as a whole, and the travel choices and traffic representation of the lower level. For each of the requirements the accompanying sections in the dissertation are provided.

Upper level

- *Rational stakeholder behaviour*
  "Stakeholders' preferences should be reflected by their objective function. They will act only in favour of their objective. In case of negotiations and/or cooperation, the used strategy of each stakeholder is rational, meaning that they each optimize their own objective."
  SECTION 2.5 The Multiple Stakeholders Problem (MSP) is defined such that it uses an objective function for each stakeholder, and each of these objectives is optimized. This implies rational stakeholder behaviour.
  SECTION 6.3 The presented solution concepts of the TU-game seek for solutions of the MSP that is accepted by all stakeholders. If a stakeholder can 'claim' more than the amount that is assigned to it in some solution, the stakeholder will reject the solution. However, differences in solution concepts still exist. Section 6.4 describes relevant (mathematical) properties of the concepts.
- *Reflection of different cooperation formations*
  "The upper level should be able to analyse different mutual attitudes of stakeholders. Cooperative and competitive behaviour is considered as endogenous. This allows analysis of the price of competition."
  CHAPTER 6 The theory of TU-games is based on cooperation between players. Therefore, this framework is very well suited to reflect cooperative stakeholder behviour.
  SECTION 7.4.3 Solutions based on non-cooperative behaviour are a side product of the TU-game framework. Within the case study a specific comparison between cooperation and competition has been carried out. This has led to the so-called price of competition.
  SECTION 6.2.1 A new method is presented to translate the MSP into a TU-game formulation. This translation considers stability of pricing schemes for coalitions and Nash-equilibria for partitions. This ensures that the value (i.e., 'profit') of a coalition resembles worst-case behaviour of all stakeholders outside the coalition.

Lower level: Travel Choices

- *Incorporation of differences in travellers' responses*
  "There is a large difference in choice behaviour between different (types of) travellers. The model should address this by either explicitly grouping travellers according to similar characteristics, or by taking the taste heterogeneity implicitly into account by random variates in the model."

SECTION 3.1.1 The QDTA distinguishes different user-classes. Each of them can have different valuations of time.

SECTION 4.2 The desired utility formulation of each choice alternative contains a distribution of foreseen travel times. Section 4.4 shows that the multiplicative Multivariate Extreme Value (MEV) models resemble all systematic components in the utility formulation as random variates.

- *Incorporation of different travel time types*
  "Travellers evaluate their time differently per mode, but also, one hour of travelling in congested conditions is experienced as a larger burden than one hour of travelling in free flow conditions. Furthermore, waiting, access and egress times for pubic transport should have different valuations."

  SECTION 3.2.3 The systemic utility specification used throughout the dissertation uses two types of travel time for cars (viz., free-flow and congested), and three types of train related travel time (viz., in-vehicle, waiting, and access-egress). Analysts can easily increase or reduce the number of travel time types.
- *Capturing overlap*
  "When two alternatives share the same characteristics (e.g., the modes or time-of-days are equal) or when there is even physical overlap (i.e., road segments), the choice preference of a traveller will be similar for these alternatives. The red-bus blue-bus is the textbook example of this correlation. It is required that the choice model accounts for this overlap."

  SECTION 4.3.2 Within the family of GMEV models, generating functions determine dependence between choice alternative's utility distributions. Therefore, the generating function determines how overlap is handled. This dissertation discusses multinomial (i.e., no depedencies between alternatives), path-size (i.e., a correction of the systematic utility), and two nested (i.e., impose actual covariance structures) generating functions. The latter are bounded in covariance structure by restrictions on the generating function that are required to obtain a sound Random Utility Maximization (RUM) model.

  SECTION 3.2.4 The transport pricing application presented in this dissertation uses a nested structure to correct for overlap (see Fig. 3.3). The travel modes reside in the first level, the time-of-days reside in the second level, and the routes reside in the last level. The framework allows an (additional) generating function (e.g., path-size) to correct for overlap on the route choice level.
- *Choice opportunities*
  "In planning models the possible responses of the travellers are reflected by their choice set. The larger this set, the more diverse the responses can be. In most transport systems public transport or slow modes are an attractive alternative, and so is deviating from the preferred departure time. The framework should be able to reflect all relevant alternatives present in the transport system in the choice set."

  SECTION 3.2 Any number of modes, times-of-day and routes can be modelled with the QDTA model. The associated choices are identified as the most important ones for transport pricing assessment. The transport pricing applications in this dissertation contains modes car, train, and a stay-at-home alternative; two time-of-day alternatives (viz., on-peak and off-peak); and, a variable number of routes per Origin-Destination (O-D)-pair.

- *Individual choices based on alternatives' properties*
  "Choices are made based on different observable properties (or attributes) of alternatives. Each traveller values these alternatives' properties in his/her own manner. Therefore, it is required that the model determines the choice probabilities based directly on these properties. Furthermore, travel cost and travel time of a trip are the most important properties related to pricing, since they are likely to change under pricing regimes. In addition, it should also be possible to capture travel time reliability of a trip. The model has to address these aspects based on physical infrastructure and on its prevailing performance given a certain pricing scheme."
  SECTION 3.2 By choosing the RUM framework for choice model, this requirement has been satisfied. In specific, Equation (3.7) shows which attributes of alternatives are included in the applications of this dissertation. Reliability has not been included as an attribute.

Lower level: Traffic Phenomena and Network Representation

- *Proper identification of bottlenecks*
  "Bottlenecks are locations in the road network where the travel demand is higher than the capacity of the infrastructure. These occur usually at discontinuities in the transport network, which are located at nodes in the abstraction of the network. Node models capture the traffic phenomena and conditions at nodes and are therefore a strict requirement. Without a node model the location and severity of congestion cannot be determined."
  SECTION 3.3 The QDTA-model uses the a QDNL-model for vehicular traffic propagation that is named STAQ. The squeezing phase of this model adds capacity constraints stemming from bottlenecks identified by node models to Static Traffic Assignment (STA). The traffic flow downstream of such a bottleneck is reduced, meaning that downstream nodes are less likely to be considered wrongfully as bottlenecks.
  CHAPTER 5 The chapter on node models shows how the confrontation of demand and supply occurs at discontinuities or intersections. The underlying behaviour (i.e., the delay drivers experience when they cross a node) of four node model instances have been compared. Although no empirical validations of the node models exist yet, the research provides directions on how to execute such validations.
- *Queue propagation by shock waves and with spillback*
  "Bottleneck locations and conditions alone are not sufficient to represent traffic, because queues occupy physical space. Therefore, the spatial dimension of congestion has to be determined. Kinematic Wave Theory (KWT) is an elegant and simple theory that allows propagation of traffic conditions over links. When congested conditions reach the beginning of a link, the node model – with new input – can determine the direction and severity of the spillback. A node model combined with KWT therefore captures important traffic flow phenomena, such as queue growth and spillback. Computational efficient methods exist that implement KWT. A traffic state that represents traffic flow,

speed, and density at every location in the network is the provided output."

SECTION 3.3 The QDTA-model uses the a QDNL-model for vehicular traffic propagation that is named STAQ. The queueing phase of this model propagates shockwaves over the network, and resembles spillback. However, full dynamic models remain more realistic in terms of dynamic growth of queue and temporal differences. Also, queueing phase does not consider second order effects (e.g., the capacity drop). Still, QDNL has a realistic average queue representation over time-of-days – especially when compared to traditional static models.

- *Representative travel time calculation*
  "Travellers make decisions based on their foreseen travel time. One would say it is rather simple to determine the travel times when speeds are known, and this is true when the traffic propagation is performed with KWT as stated in the previous requirement. However, this is listed as a separate requirement since plenty of standard models in STA and other heuristic methods cannot determine these speeds appropriately under congested conditions, and thus cannot report representative travel times."

  SECTION 3.3 The simulated traffic state of the network is directly derived from the fundamental diagrams for each link. Travel times are easily deduced from the traffic state of each link and are (potentially) observable; travel times are thus representative.

- *Representative (external) effects quantifications*
  "For the stakeholders holds that they want to make decisions about pricing measures based on reliable estimates of effects. Quantities like air and noise pollution can – just like travel time – be determined with traffic conditions as flow, density and speed. However, just like travel time, these have to be realistic to get a good result. Due to the importance of effects, also this is stated as a separate requirement."

  SECTION 7.3 The case study uses emissions as an important external effect. These are quantified with an off-the-shelf model that bases the emission levels on density and flow.

- *Varying network conditions over the day*
  "The (average) network conditions change over the periods within the day. It is required to capture this variability over the day by having representative time periods (e.g., hourly intervals or different peak hours). Since the second-to-second or minute-to-minute variations are not relevant for transport pricing, a continuous or strongly discretized approach is not required."

  SECTION 3.2 Any number of time-of-days can be included in the QDTA-model. The applications in this thesis distinguish on-peak and off-peak traffic. However, continuous time representations are not possible with the QDTA-model, and successive time periods are independent. It is not possible to consider residual traffic to tranfer from period to period. Still, travel times and e.g. emission levels represent averages over specific times-of-day.

8.2. Contributions for Practice and Policy Makers

The framework and methods provided in this thesis are a toolbox to design strategic planning models for transport pricing. Such a model can provide (necessary) insights in the impacts of innovative pricing measures. Besides more detailed forecasts of traditional indicators as accessibility and emissions, also potential conflicts between stakeholders can be identified. The game theoretical upper level approach provides multiple solution concepts on which stakeholders (i.e., decision makers) can base and/or alter their decisions. It suggests allocations that define monetary transfers between them.

When innovative pricing measures are considered often many alternatives have to be studied, or price levels need to be optimized. This leads to many different model runs. The work presented in this thesis seeks a good balance between realism and computation time. With the presented methodological advances, more detailed and more accurate analyses of innovative pricing measures become feasible. Strategic planning models for decision support can cover multiple modes, multiple stakeholders and large networks. This is illustrated by the case study presented in Chapter 7.

Lack of public and political support are two of the main hurdles of innovative pricing measures. This research could alleviate these obstacles by providing models that provide insight in possible conflicts between stakeholders on one hand, and provide more realism – and therefore more credible results – on the other hand. The main improvement in realism is realized by using the QDTA-model (see Chapter 3). This model calculates averages per time-of-day, which is similar to traditional static models, but with the main advantages that (1) capacity constraints are calculated much more accurately due to the inclusion of a proper node model, and (2) queues are placed upstream of the bottleneck (instead of inside the bottleneck). In addition, the QDTA-model works with state-of-the-art choice models (e.g., all logit and weibit based models). This also includes some newly presented choice models that have some favorable properties (e.g., it resembles expected behaviour under simple network changes). Furthermore, insight is provided in which models can be applied to other O-D-pairs than they were estimated on, without having re-estimate the parameters. See Chapter 4 for more analyses and details on choice models.

The case study for the Randstad area in the Netherlands presented in Chapter 7 illustrates the earlier mentioned improvements. Three stakeholders were involved that want to implement three different pricing measures (a kilometre charge, a cordon charge, a train fare adjustment). The underlying network is large (3 608 zones, almost 90 000 road links), and travellers could choose between modes, whether to travel during the peak, and between multiple routes. For 216 feasible pricing schemes a network equilibrium was computed with the QDTA-model. Three solution concepts from TU-game theory have been analysed: the core, the Shapley value and the compromise value. Each of them satisfies multiple fairness criteria.

The case study uses hypothetical stakeholder objectives and more simplifying assumptions have been made. So note that policy implications presented here can only be made conditional to these hypotheses and assumptions. One of the results is that the national government and the train operator have conflicting interests. These insights – based on the game theoretical solution concepts – would not have been retrieved from traditional social welfare optimization. Another remarkable result is that one of the stakeholders, the national government, is worse off under cooperation, than under competition. Paradoxically, cooperation leads to a social welfare reduction due to the power of the other stakeholders.

8.2.1. Tools for Practitioners

By using the theories and methods in this dissertation, practitioners (i.e., model developers and consultants) can improve their advice to policy makers. The theory allows for the following possibilities:

- Innovative pricing measures can be assessed. That means that pricing schemes can be differentiated over times of the day, locations, and/or user-classes. The response of the traveller to such a pricing scheme is determines with respect to mode, time-of-day, and route choice. Innovative pricing measures have more potential to improve the efficiency of the transport system than traditional pricing measures. For example, charges that differentiate between on-peak and off-peak could motivate travellers to travel off-peak, and thus alleviate on-peak congestion, while this is not possible with flat charges.
- The QDTA has extended traditional STA-models by adding capacity constraints from proper bottleneck identification. In addition, physical queues are simulated upstream of each bottleneck, under the assumption that they all start to grow at the same moment. At the same time, computational costs remain reasonable. This provides practitioners realistic travel times and emission levels to respond on. The QDNL model, or parts thereof, has been implemented in commercial packages (e.g., as STAQ in OmniTRANS), and is therefore available to practitioners.
- The QDTA model identifies bottlenecks with node models. The presented family of node models provides behavioural insight in the origin and cause of the bottleneck. The framework provides the delay per direction on intersections. This can for example be used to validate the used node model, and to develop new node models.
- Practitioners can use the GMEV framework to easily implement and analyse different types of choice models. All currently known logit- and weibit-based models with explicit route sets fit in the framework. A generating function and a generating vector are the only ingredients to retrieve close form choice probability formulas. Also, a new model type (MΔ) is presented, that is the only model that resembles expected behaviour under all simple network changes.
- The decision making process can be supported with the game theoretical multiple stakeholder approach. The possible arguments put down at the negotiation table are used to derive several solution concepts. The theory also takes the formation of coalitions into account. Conflicting interests between stakeholders are tackled by looking for Nash-equilibria. This approach is more comprehensive than single objective transport pricing approaches, such as social welfare optimization. As a side products, the practitioner can get insight in the price of competition.

8.3. Methodological Contributions

This section provides a list of methodological contributions of the research. The contributions are grouped per topic.

Traffic Assignment (TA)-models for transport pricing (Chapters 2 and 3)

- For the transport pricing application as presented in Chapter 2, the choice model uses a multiplicative utility formula which is advantageous on large networks (see Chapter 4). Dependencies between choice alternatives are captured with a nest structure. A special dummy mode captures the stay-at-home alternative. A simultaneous mode, time-of-day, and route choice model with a multiplicative utility formula has not been applied before. Any of the multiplicative Multinomial (MN), Path-Size (PS), Paired Combinatorial (PC), or Link-Nested (LN) can be substituted for the lowest route nests.
- The used QDNL-model based on the hybrid approach presented by Brederode et al. (2010); Bliemer et al. (2012) consisting of a squeezing phase and a queueing phase provides a balanced trade-off between efficiency and realism. The ultimately used squeezing phase is presented by Bliemer et al. (2014b) and extends STA with capacity constraint stemming from a node model. The ultimately used queueing phase is presented by Raadsen et al. (2014a) and provides a Dynamic Traffic Assignment (DTA)-model. The author has contributed to the development of the problem formulation and solution algorithms of the queueing phase (Bliemer et al., 2014b). This QDNL-model has not been applied to transport pricing problems before, and is integrated within the equilibrium formulation in the modelling framework.
- The variational inequality formulation of the QDTA-model has been presented. This allows the calculation of a gap function that shows to what extent an equilibrium has been achieved in each iteration of the computations.

Discrete Choice Modelling: the GMEV route choice models (Chapter 4)

- Twelve route choice models – of which seven are new – have been presented in a single framework. All models have the same closed form expression for the choice probabilities. They have been assessed on multiple criteria. No framework existed that includes all these models.
- Empirical evidence has provided this new insight: The foreseen travel time (i.e., the travel time on which travellers base their decision) distribution's mean and standard deviation have a linear relationship, contrary to a linear relationship between its mean and variance. The additive models (i.e., the logit family) are not able to capture the random foreseen travel time in this manner, but multiplicative models (i.e., the weibit family) do allow for this.
- The constant in systematic utility of multiplicative models does not have to be normalized. This allows more degrees of freedom and a better fit on the data.
- One main advantage of the generic framework is that it can be analysed as a whole, we show this by providing the equivalent stochastic user equilibrium formulation for all models. This is done in the form of a variational inequality.

- Four out of seven new models are based on a different decision rule: the Multiplicative with Reference Route (MΔ)-models. Those are the only ones that can reproduce realistic behaviour under three different basic network changes. The (logit based) Additive (A)-models cannot handle a change of parallel links' characteristics, and the (weibit based) Multiplicative (M)-models cannot handle a change of serial links' characteristics.
- The models Multiplicative Path-Size (M-PS) and Multiplicative Path-Size with Reference Route (MΔ-PS) are expected to have the best behaviour on real networks. They can capture overlap sufficiently, and they can handle random foreseen travel time. Furthermore, they perform good on tests with carefully constructed toy networks. The models have been estimated on one dataset and validated on another dataset to assess transferability. However, additional empirical estimation and validation is required to conclusively assess all models.

Node models (Chapter 5)

- The Generic Class of first-order Node Models (Tampère et al., 2011) can be formulated as a family of models based on turn delays in the form of a multi-objective optimization problem. A turn delay is the additional time a vehicle occupies an inlink when it heads for a certain direction, and is thus easily interpretable. Two existing (viz., *directed capacity proportional* and *capacity consumption equivalence*) and two new node models are member of the family.
- Two new models have been presented: the *single server* model where all vehicles hinder each other at an intersection, and the *equal delay at outlink* model where each vehicle with same destination undergoes an equal amount of delay.
- Solving these problems is not straightforward due to multiple objectives and unknowns. For the single server model a very efficient solution method based on the theory of polynomials in the max-plus algebra is presented.
- It holds for each currently known model that a balance between plausibility of the underlying behaviour and computational efficiency has to be made.
- Several sets of turn delays leading to the same resulting flows can be found. The notion of reduced capacity has been introduced to help identifying whether two results are equivalent.
- It has been shown that for diverges all family members are equivalent and that three out of four members yield identical results at merges. The relation between priority parameters at merges and the family is also presented.
- The main advantage of the node model family is the behavioural interpretability of the models due to the use of turn delays. The additional relations between turn delays, required to select a Pareto optimal solution in terms of reduced capacities, imply a part of the behaviour at intersections. Especially at a supply constraint, the relation between competing turns should be determined in terms of delays.

Game Theory with Multiple Stakeholders (Chapter 6)

- The gap between multi-objective optimization (i.e., the MSP) and cooperative TU-game theory has been bridged. By taking notion of coalitions and by using Nash-equilibria –

based on the optimization of stakeholders' objective functions – a TU-game with coalition values has been derived.

- The solution concepts core, Shapley value, and compromise value to TU-games have been presented in general (i.e., without the transport pricing application). The practical aspects and policy implications of these solution concepts have been presented.
- Compared to the traditional *social welfare optimization*, the TU-game approach provides a more detailed view on possible wins and losses of each stakeholder, as well under cooperation as under competition. The assumptions of *first-best pricing* and social welfare optimization that every individual of a population participates in the negotiations and that unlimited mutual transactions between them can exist are released. The provided framework provides a concrete interpretation of the negotiation table. The resulting game theoretical solutions provide more feasible and more realistic 'maximum achievements', than can be obtained under merely social welfare optimization.

8.4. Future Directions

This final section discusses open problems in transport pricing and strategic network modelling. It covers topics that have not been in the scope of this thesis, but that deserve some words. Suggestions for research and a look forward is provided.

Origins and Destinations There are plenty of sources of error and uncertainty in strategic network models. For example, parameters of choice models and network loading models have to be calibrated with care, and the uncertain forecasts of socio-economic indicators and data are the basis for future demand. The latter is included in – and given as – the travel demand per O-D-pair in this thesis. The amount of unknowns in such a matrix equals the square of the number of considered zones in the network. This amount of unknowns is much higher than the amount of available data. Even when traffic is counted on every road, there is – by far – not enough data.

Therefore, it is valuable to investigate other approaches than those with O-D matrices. A possible direction is to consider sampling methods that sample trips and their travel choices simultaneously. Then a 'true' O-D matrix is not required anymore. Davidson (2011) proposes such an approach. The major disadvantage would be that the modelling results are merely one draw/sample from a distribution of outcomes. Since that implies that the outcome can be an outlier (or black swan), some repeated simulation by means of a Monte Carlo method is required.

Another direction would be to take a data-driven approach. Floating car data become more and more available. Such data contains trajectories (i.e., origin, destination, route, mode, time) of trips which are very rich. The penetration grade of floating car data providers might be limited, and the sample might be biased (e.g., towards highway users), but the data might be rich enough to deduce travel demand. Unfortunately, such data does not provide insight in the underlying behaviour, e.g. the choice process, so it is more difficult to assess future scenarios.

Networks and Intersections The case study in this thesis has shown that the network configuration and specification is extremely important for the QDNL model. Since congestion builds up upstream of bottlenecks, and spills back over intersections, it is important that the direction (over an intersection) to which queues build up is correct. For example, at roundabouts the queues quickly form a gridlock (in a model) because the vehicles entering the roundabout do not give way. The QDNL model was only usable when this was resolved. Also, several errors in the road capacity and intersection configuration were present in the Nederlands Regionaal Model (Dutch Regional Model) (NRM)-network. They also had to be resolved to avoid wrongly placed and overly long queues. Compared to traditional static models, capacity constrained models are more sensitive to changes in capacity. Models with spillback also propagate the erroneous traffic states and therefore affect larger parts of the network.

Since the node model is the mechanism that determines the severity and direction of queues, node models need to be realistic. It should be possible to calibrate them using observations at intersections. For highways, lane drops, merges, diverges and weaving sections can cause bottlenecks. Empirical evidence should be provided which of the node models presented in this thesis has the best ‘fit’ to observed traffic states. For urban road networks, a similar approach can be taken for unsignalized and/or prioritized intersections and roundabouts. For signalized intersections, information of the layout and priorities within the controller could be exploited.

The node model family with turn delays presented in this thesis, can be used to develop new models that are based on data analysis. By observing traffic at intersections under different conditions, relations between turn delays on different turns can be retrieved. This additional information is required, and it can only be included through empirical analysis. Section 5.3.2 introduces how occupancy times and turn delays can be observed.

Section 5.5.2 and Corthout et al. (2012) have shown that when turns of the same inlink have different priorities at different conflicts non-uniqueness within the node model can occur. This is a subsistent characteristic of traffic. Thus adding realism by adding a conflict, can lead to non-uniqueness. Therefore, in the search of new node models, a balance needs to found between two desired properties: realism and uniqueness.

Big Data The large amounts of data that are becoming more and more available will help to make strategic network models more realistic. Road networks can be generated automatically, which avoids manual coding errors. GPS, GSM and other data from personal devices will allow a better estimation and calibration of the models.

On the other hand, big data will not be the holy grail. This has several reasons:

- In transport it is very important to know what the non-chosen alternatives are. The latent demand for a road when its number of lanes is doubled can not be retrieved by big data. And – more related to this thesis – the non-chosen alternatives of a choice set can not be observed. Big data can only reveal chosen alternatives
- Big data is associated with privacy concerns. The main hurdle is the fact that movements could be mapped one-to-one to individuals. That is one reason why data providers hesitate to provide raw data (another is based on commercial reasons).

Prediction uncertainty Big data, travel surveys, and other sources can measure, validate, and calibrate the base year scenario in order to approximate reality better and better. However, future scenarios always rely on the predictive power of the model, and even with all the knowldge of the world, there is always a limit to the accuracy of the prediction. The magnitude of the congestion problems in the Netherlands are in line with the economic recession and its resolution. No model has predicted the economic development over the last recession correctly. Therefore, one should be careful with predictions; one could for example use bandwidths or ensembles of model outcomes.

Multimodal trips TA-models become way more complex when multiple modes are included. Especially, when multiple modes can be combined within a single trip. In this dissertation the train mode has separate access and egress modes, but there is no full flexibility on combining modes. For example, a cordon charge or increased parking fees might stimulate park and ride facilities. An extension of the model to allow multimodal trips would therefore be of added value for the evaluation of innovative pricing measures. Recent research has provided more insight on this topic. Van Eck et al. (2014) describe these model complexities and formulate requirements for multimodal transport networks. Solehmainen (2011) uses a unified discrete choice framework for multimodal trips.

Choice sets This dissertation has an explicit choice set, which assumes that all choice alternatives are given exogenously. Especially, for route sets it is difficult to determine these a priori. When the transport system changes, excluded routes might become attractive alternatives. Many route generation algorithms and some route sampling approaches exist. It is important to work with a correct sample of routes is required, for example, to obtain unbiased parameter estimates of choice models (Frejinger et al., 2009). So, for econometrically sound applications, sampling techniques as presented by (Frejinger et al., 2009; Flötteröd and Bierlaire, 2013; Guevara and Ben-Akiva, 2013) have to be adapted to generate choice sets. Since practical route generation algorithms usually do not produce correct samples, and sampling approach are practically less feasible, it often remains a practical challenge to produce choice sets. Section 3.2.2 also points out the importance of the composition of the choice set.

Another approach is to avoid the use of a priori generated choice sets. Route choice models with implicit route sets (Dial, 1971; Fosgerau et al., 2013a; Papola and Marzano, 2013; Mai et al., 2015) have endogenously generated routes that can more easily satisfy desired econometrical properties. Unfortunately, these routes are not always realistic, and can sometimes contain loops. So, both approaches have their advantages and disadvantages. Future developments in both approaches will tell which one is the best for practical applications.

Mutual influence of stakeholders The provided game theoretical approach to solve the MSP assumes that each stakeholder has complete power over its own pricing measure. In reality, there is not a single decision maker nor a single objective function. An example hereof are lobby groups that influence politics. If such 'lobbies' are key to analyse, then additional structures have to be introduced to the framework. Tripsigala (2014) allows these influences within the TU-game framework by providing veto power to stakeholders without their own pricing measure. Another approach is to introduce an intermediate (decision) layer in which multiple stakeholders have to decide on the measure. The latter will require a completely different form of modelling decision making than provided this thesis.

Travel time reliability Travel time realibility (i.e., the certainty by which one can plan a trip using some travel time) becomes more and more important for policy evaluations. Innovative pricing measures could also be used to pursue travel time reliability. For example, Tirachini et al. (2014) accounts for travel time variability in the setting of optimal car and bus pricing; furthermore, they provide a literature review on travel time reliability with a focus on pricing.

Bibliography

Abrantes, P. A. and Wardman, M. R. (2011). Meta-analysis of uk values of travel time: An update. *Transportation Research Part A: Policy and Practice*, 45(1):1–17.

Acemoglu, D. and Ozdaglar, A. (2007). Competition and efficiency in congested markets. *Mathematics of Operations Research*, 32(1):1–31.

Arnott, R., de Palma, A., and Lindsey, R. (1990a). Departure time and route choice for the morning commute. *Transportation Research Part B: Methodological*, 24(3):209–228.

Arnott, R., de Palma, A., and Lindsey, R. (1990b). Economics of a bottleneck. *Journal of Urban Economics*, 27(1):111–130.

Arnott, R., de Palma, A., and Lindsey, R. (1992). Route choice with heterogeneous drivers and group-specific congestion costs. *Regional Science and Urban Economics*, 22(1):71–102.

Aubin, J., Bayen, A., and Saint-Pierre, P. (2011). *Viability Theory: New Directions*. Modern Birkhèauser classics. Springer.

Bar-Gera, H. (2010). Traffic assignment by paired alternative segments. *Transportation Research Part B: Methodological*, 44(8Ű9):1022–1046.

Bar-Gera, H., Boyce, D., and Nie, Y. M. (2012). User-equilibrium route flows and the condition of proportionality. *Transportation Research Part B: Methodological*, 46(3):440–462.

Beckmann, M. (1965). On optimal tolls for highways, tunnels and bridges. In *Vehicular Traffic Science*. Elsevier.

Beckmann, M., McGuire, C., and Winsten, C. B. (1956). Studies in the economics of transportation. Technical report, Yale University Press.

Beerepoot, R. and Dijkers, J. (2013). Gedragsmeting beter benutten; 2e meting. Technical report, I&O Research.

Bekhor, S., Ben-Akiva, M., and Ramming, M. (2006). Evaluation of choice set generation algorithms for route choice models. *Annals of Operations Research*, 144(1):235–247.

Bekhor, S., Ben-Akiva, M., and Scott Ramming, M. (2002). Adaptation of logit kernel to route choice situation. *Transportation Research Record: Journal of the Transportation Research Board*, 1805:78–85.

Bekhor, S. and Prashker, J. N. (1999). Formulations of extended logit stochastic user equilibrium assignments. In *14th International Symposium on Transportation and Traffic Theory*.

Ben-Akiva, M. and Abou-Zeid, M. (2013). Methodological issues in modelling time-of-travel preferences. *Transportmetrica A: Transport Science*, 9(9):846–859.

Ben-Akiva, M. and Bierlaire, M. (1999). Discrete choice methods and their applications to short term travel decisions. In Hall, R., editor, *Handbook of Transportation Science*, volume 23 of *International Series in Operations Research & Management Science*, pages 5–33. Springer US.

Ben-Akiva, M. E. (1973). *Structure of passenger travel demand models.* PhD thesis, Massachusetts Institute of Technology.

Bierlaire, M. (2006). A theoretical analysis of the cross-nested logit model. *Annals of Operations Research*, 144(1):287–300.

Bierlaire, M., Bolduc, D., and McFadden, D. (2008). The estimation of generalized extreme value models from choice-based samples. *Transportation Research Part B: Methodological*, 42(4):381 – 394.

Bliemer, M. (2007). Dynamic queuing and spillback in analytical multiclass dynamic network loading model. *Transportation Research Record: Journal of the Transportation Research Board*, 2029:14–21.

Bliemer, M. and Bovy, P. (2008). Impact of route choice set on route choice probabilities. *Transportation Research Record: Journal of the Transportation Research Board*, 2076:10–19.

Bliemer, M., Raadsen, M., Brederode, L., Bell, M., and Wismans, L. (2014a). A unified framework for traffic assignment: deriving static and quasi-dynamic models consistent with general first order dynamic traffic assignment models. Technical Report ITLS-WP-14-08. ISSN 1832-570X, Institute of Transport and Logistics Studies, The University of Sydney.

Bliemer, M., Raadsen, M., de Romph, E., and Smits, E.-S. (2013). Requirements for traffic assignment models for strategic transport planning: A critical assessment. In *Proceedings of the 36th Autralasian Transport Research Forum.*

Bliemer, M. C. and Bovy, P. H. (2003). Quasi-variational inequality formulation of the multiclass dynamic traffic assignment problem. *Transportation Research Part B: Methodological*, 37(6):501–519.

Bliemer, M. C., Brederode, L., Wismans, L., and Smits, E.-S. (2012). Quasi-dynamic traffic assignment: static traffic assignment with queueing and spillback. In *The Transportation Research Board (TRB) 91st Annual Meeting, Washington DC, January 22-26, 2012 (paper no 12-0358).*, pages 1 – 24.

Bliemer, M. C., Raadsen, M. P., Brederode, L. J., Bell, M. G., Wismans, L. J., and Smith, M. J. (2017). Genetics of traffic assignment models for strategic transport planning. *Transport reviews*, 37(1):56–78.

Bliemer, M. C., Raadsen, M. P., Smits, E.-S., Zhou, B., and Bell, M. G. (2014b). Quasi-dynamic traffic assignment with residual point queues incorporating a first order node model. *Transportation Research Part B: Methodological*, 68(0):363–384.

Bovy, P., Bekhor, S., and Prato, C. (2008). The factor of revisited path size: Alternative derivation. *Transportation Research Record: Journal of the Transportation Research Board*, 2076:132–140.

Bovy, P. H. and Fiorenzo-Catalano, S. (2007). Stochastic route choice set generation: behavioral and probabilistic foundations. *Transportmetrica*, 3(3):173–189.

Branzei, R., Dimitrov, D., Tijs, S., Dimitrov, D., and Tijs, S. (2008). *Models in cooperative game theory*, volume 556. Springer.

Brederode, L., Bliemer, M., and Wismans, L. (2010). *STAQ: Static Traffic Assignment with Queing*. Citeseer.

Cascetta, E. (2009). *Transportation systems analysis: models and applications*, volume 29. Springer Science & Business Media.

Cascetta, E., Nuzzolo, A., Russo, F., and Vitetta, A. (1996). A modified logit route choice model overcoming path overlapping problems. sepcification and some calibration results for interurban networks. In *Proceedings of the 13 th International Symposium on Transportation and Traffic Theory*.

Cascetta, E., Russo, F., Viola, F. A., and Vitetta, A. (2002). A model of route perception in urban road networks. *Transportation Research Part B: Methodological*, 36(7):577–592.

Castillo, E., Menéndez, J. M., Jiméez, P., and Rivas, A. (2008). Closed form expressions for choice probabilities in the weibull case. *Transportation Research Part B: Methodological*, 42(4):373–380.

Ceder, A. (2007). *Public transit planning and operation : theory, modelling and practice*. Elsevier, London; Burlington, MA.

Chau, C. K. and Sim, K. M. (2003). The price of anarchy for non-atomic congestion games with symmetric cost maps and elastic demands. *Operations Research Letters*, 31(5):327–334.

Chen, A., Kasikitwiwat, P., and Ji, Z. (2003). Solving the overlapping problem in route choice with paired combinatorial logit model. *Transportation Research Record: Journal of the Transportation Research Board*, 1857(-1):65–73.

Chen, A., Pravinvongvuth, S., Xu, X., Ryu, S., and Chootinan, P. (2012). Examining the scaling effect and overlapping problem in logit-based stochastic user equilibrium models. *Transportation Research Part A: Policy and Practice*, 46(8):1343–1358.

Chen, H.-K. (1999). *Dynamic travel choice models: A variational inequality approach*. Springer.

Chikaraishi, M. and Nakayama, S. (2015). How well drivers can estimate travel time? estimating travel-time-dependent variance based on a q-generalized logit model. In *Proceedings of the 14th International Conference on Travel Behaviour Research, Windsor, UK.*

Chu, C. (1989). A paired combinatorial logit model for travel demand analysis. In *Proceedings of the Fifth World Conference on Transportation Research*, volume 4, pages 295–309.

Claudel, C. and Bayen, A. (2010a). Lax-hopf based incorporation of internal boundary conditions into hamilton-jacobi equation. part i: Theory. *Automatic Control, IEEE Transactions on*, 55(5):1142–1157.

Claudel, C. and Bayen, A. (2010b). Lax-hopf based incorporation of internal boundary conditions into hamilton-jacobi equation. part ii: Computational methods. *Automatic Control, IEEE Transactions on*, 55(5):1158–1174.

Clune, A., Smith, M., and Xiang, Y. (1999). A theoretical basis for implementation of a quantitative decision support system – using bilevel optimisation. In Ceder, A., editor, *14th International Symposium on Transportation and Traffic Theory*, pages 489–513.

Correa, J. R., Schulz, A. S., and Stier-Moses, N. E. (2004). Selfish routing in capacitated networks. *Mathematics of Operations Research*, 29(4):961–976.

Corthout, R. (2012). *Intersection Modelling and Marginal Simulation in Macroscopic Dynamic Network Loading*. PhD thesis, Katholieke Universiteit Leuven, Belgium.

Corthout, R., Flötteröd, G., Viti, F., and Tampère, C. M. (2012). Non-unique flows in macroscopic first-order intersection models. *Transportation Research Part B: Methodological*, 46(3):343–359.

Cuninghame-Green, R. (1995). Maxpolynomial equations. *Fuzzy Sets and Systems*, 75(2):179–187.

Cuninghame-Green, R. and Meijer, P. (1980). An algebra for piecewise-linear minimax problems. *Discrete Applied Mathematics*, 2(4):267–294.

Daganzo, C. (1979). *Multinomial probit: the theory and its application to demand forecasting*. Academic Press New York.

Daganzo, C. F. (1995a). The cell transmission model, part ii: Network traffic. *Transportation Research Part B: Methodological*, 29(2):79–93.

Daganzo, C. F. (1995b). Requiem for second-order fluid approximations of traffic flow. *Transportation Research Part B: Methodological*, 29(4):277–286.

Daganzo, C. F. (2005a). A variational formulation of kinematic waves: basic theory and complex boundary conditions. *Transportation Research Part B: Methodological*, 39(2):187–196.

Daganzo, C. F. (2005b). A variational formulation of kinematic waves: Solution methods. *Transportation Research Part B: Methodological*, 39(10):934–950.

Daganzo, C. F. (2006). On the variational theory of traffic flow: Well-posedness, duality and applications. Technical report, Institute of Transportation Studies, UC Berkeley.

Daganzo, C. F. and Sheffi, Y. (1977). On stochastic models of traffic assignment. *Transportation Science*, 11(3):253–274.

Daly, A. and Bierlaire, M. (2006). A general and operational representation of generalised extreme value models. *Transportation Research Part B: Methodological*, 40(4):285–305.

Davidson, P. (2011). A new approach to transport modelling-the stochastic segmented slice simulation (4S) model and its recent applications. In *Australian Transport Research Forum*. Citeseer.

de Borger, B. and Proost, S. (2012a). A political economy model of road pricing. *Journal of Urban Economics*, 71(1):79–92.

de Borger, B. and Proost, S. (2012b). Transport policy competition between governments: A selective survey of the literature. *Economics of Transportation*, 1(1Ű2):35–48.

de Bruyn, S., Korteland, M., and Markowska, A. (2010). Bijlagen - handboek schaduwprijzen. Technical report, CE Delft.

de Cea, J. and Fernández, E. (1993). Transit assignment for congested public transport systems: An equilibrium model. *Transportation Science*, 27(2):133–147.

de Jong, G., Daly, A., Pieters, M., Vellay, C., Bradley, M., and Hofman, F. (2003). A model for time of day and mode choice using error components logit. *Transportation Research Part E: Logistics and Transportation Review*, 39(3):245–268.

de Palma, A. and Lindsey, R. (2011). Traffic congestion pricing methodologies and technologies. *Transportation Research Part C: Emerging Technologies*, 19(6):1377–1399.

de Palma, A., Lindsey, R., and Proost, S. (2006). Research challenges in modelling urban road pricing: An overview. *Transport Policy*, 13(2):97–105. Modelling of Urban Road Pricing and its Implementation.

de Schutter, B. and de Moor, B. (1996). A method to find all solutions of a system of multivariate polynomial equalities and inequalities in the max algebra. *Discrete Event Dynamic Systems: Theory and Applications*, 6:115–138.

Dial, R. B. (1971). A probabilistic multipath traffic assignment model which obviates path enumeration. *Transportation Research*, 5(2):83–111.

Dial, R. B. (2006). A path-based user-equilibrium traffic assignment algorithm that obviates path storage and enumeration. *Transportation Research Part B: Methodological*, 40(10):917–936.

Ecola, L. and Light, T. (2009). Equity and congestion pricing: A review of the evidence. Technical report, RAND Corporation, Santa Monica, CA.

Ettema, D., Bastin, F., Polak, J., and Ashiru, O. (2007). Modelling the joint choice of activity timing and duration. *Transportation Research Part A: Policy and Practice*, 41(9):827–841. Selected Papers on Applications of Discrete Choice Models Presented at the European Regional Science Conference, Amsterdam, August 2005.

Facchinei, F. and Pang, J.-S. (2003). *Finite-dimensional variational inequalities and complementarity problems*, volume 1. Springer.

Fellendorf, M. and Vortisch, P. (2010). Microscopic traffic flow simulator vissim. In Barceló, J., editor, *Fundamentals of Traffic Simulation*, volume 145 of *International Series in Operations Research & Management Science*, pages 63–93. Springer New York.

Fiorenzo-Catalano, S., Van Nes, R., and Bovy, P. H. (2004). Choice set generation for multimodal travel analysis. *European journal of transport and infrastructure research EJTIR, 4 (2)*.

Fisk, C. (1980). Some developments in equilibrium traffic assignment. *Transportation Research Part B: Methodological*, 14(3):243–255.

Flötteröd, G. and Bierlaire, M. (2013). Metropolis–hastings sampling of paths. *Transportation Research Part B: Methodological*, 48(0):53–66.

Flötteröd, G. and Rohde, J. (2011). Operational macroscopic modeling of complex urban road intersections. *Transportation Research Part B: Methodological*, 45(6):903–922.

Fosgerau, M. (2010). On the relation between the mean and variance of delay in dynamic queues with random capacity and demand. *Journal of Economic Dynamics and Control*, 34(4):598–603.

Fosgerau, M. and Bierlaire, M. (2009). Discrete choice models with multiplicative error terms. *Transportation Research Part B: Methodological*, 43(5):494–505.

Fosgerau, M., Frejinger, E., and Karlstrom, A. (2013a). A link based network route choice model with unrestricted choice set. *Transportation Research Part B: Methodological*, 56:70–80.

Fosgerau, M., McFadden, D., and Bierlaire, M. (2013b). Choice probability generating functions. *Journal of Choice Modelling*, 8:1–18.

Frejinger, E. and Bierlaire, M. (2006). Capturing correlation in large-scale route choice models. Technical report, École polytechnique fédérale de Lausanne. RO-060106.

Frejinger, E., Bierlaire, M., and Ben-Akiva, M. (2009). Sampling of alternatives for route choice modeling. *Transportation Research Part B: Methodological*, 43(10):984–994.

Fukushima, M. (1992). Equivalent differentiable optimization problems and descent methods for asymmetric variational inequality problems. *Mathematical Programming*, 53(1-3):99–110.

Gálvez, T. E. (2002). Elmodelo powit. *Urbana*, 7(30).

Gentile, G. (2008). The general link transmission model for dynamic network loading and a comparison with the due algorithm. In *Proceedings of the Second International Symposium on Dynamic Traffic Assignment DTA 2008*.

Gibb, J. (2011). Model of traffic flow capacity constraint through nodes for dynamic network loading with queue spillback. *Transportation Research Record: Journal of the Transportation Research Board*, 2263(-1):113–122.

Gillies, D. B. (1959). Solutions to general non-zero-sum games. *Contributions to the Theory of Games*, 4(40):47–85.

Gliebe, J., Koppelman, F., and Ziliaskopoulos, A. (1999). Route choice using a paired combinatorial logit model. In *78th TRB Meeting, Washington DC*.

Guajardo, M. and Jornsten, K. (2014). Common mistakes in computing the nucleolus. *NHH Dept. of Business and Management Science Discussion Paper*.

Guevara, C. A. and Ben-Akiva, M. E. (2013). Sampling of alternatives in multivariate extreme value (mev) models. *Transportation Research Part B: Methodological*, 48(0):31–52.

Guo, X. and Yang, H. (2010). Pareto-improving congestion pricing and revenue refunding with multiple user classes. *Transportation Research Part B: Methodological*, 44(8-9):972–982.

Guo, X., Yang, H., and Liu, T.-L. (2010). Bounding the inefficiency of logit-based stochastic user equilibrium. *European Journal of Operational Research*, 201(2):463–469.

Hamdouch, Y., Florian, M., Hearn, D. W., and Lawphongpanich, S. (2007). Congestion pricing for multi-modal transportation systems. *Transportation Research Part B: Methodological*, 41(3):275–291.

Hamilton, C. (2012). Decisive factors for the acceptability of congestion pricing. Technical report, CTS Working Paper, Stockholm.

Han, D., Lo, H. K., and Yang, H. (2008). On the price of anarchy for non-atomic congestion games under asymmetric cost maps and elastic demands. *Computers & Mathematics with Applications*, 56(10):2737–2743.

Hau, T. D. (2005a). Economic fundamentals of road pricing: a diagrammatic analysis, part i - fundamentals. *Transportmetrica*, 1(2):81–117.

Hau, T. D. (2005b). Economic fundamentals of road pricing: a diagrammatic analysis, part ii - relaxation of assumptions. *Transportmetrica*, 1(2):119–149.

Hellinga, B., van Lint, H., and Hofman, F. (2012). The use of exogenously defined standard deviation versus mean travel time relationships for estimating the impact of policy measures on reliability. In *91st Annual Meeting of the Transportation Research Board, Washington DC, USA*.

Hensher, D. A. and Bliemer, M. C. (2014). What type of road pricing scheme might appeal to politicians? viewpoints on the challenge in gaining the citizen and public servant vote by staging reform. *Transportation Research Part A: Policy and Practice*, 61(0):227–237.

Hess, S., Daly, A., Rohr, C., and Hyman, G. (2007). On the development of time period and mode choice models for use in large scale modelling forecasting systems. *Transportation Research Part A: Policy and Practice*, 41(9):802–826. Selected Papers on Applications of Discrete Choice Models Presented at the European Regional Science Conference, Amsterdam, August 2005.

Hilbers, H., van Meerkerk, J., Verrips, A., van der Straaten, W. W., and Zwaneveld, P. (2015). Maatschappelijke kosten en baten prijsbeleid personenauto's. Technical report, Centraal Planbureau & Planbureau voor de Leefomgeving.

Hopf, E. (1970). On the right weak solution of the cauchy problem for a quasilinear equation of first order. *Indiana Univ. Math. J.*, 19:483–487.

Huang, H.-J. and Lam, W. H. (2002). Modeling and solving the dynamic user equilibrium route and departure time choice problem in network with queues. *Transportation Research Part B: Methodological*, 36(3):253–273.

Jiang, L., Mahmassani, H. S., and Zhang, K. (2011). Congestion pricing, heterogeneous users, and travel time reliability. *Transportation Research Record: Journal of the Transportation Research Board*, 2254(1):58–67.

Jin, W. and Zhang, H. (2003). On the distribution schemes for determining flows through a merge. *Transportation Research Part B: Methodological*, 37(6):521–540.

Jin, W. and Zhang, H. (2004). Multicommodity kinematic wave simulation model for network traffic flow. *Transportation Research Record: Journal of the Transportation Research Board*, 1883(-1):59–67.

Jin, W.-L. (2010). Continuous kinematic wave models of merging traffic flow. *Transportation Research Part B: Methodological*, 44(8-9):1084–1103.

Jin, W.-L. (2012a). A kinematic wave theory of multi-commodity network traffic flow. *Transportation Research Part B: Methodological*, 46(8):1000–1022.

Jin, W.-L. (2012b). The traffic statics problem in a road network. *Transportation Research Part B: Methodological*, 46(10):1360–1373.

Joksimovic, D. (2007). *Dynamic bi-level optimal toll design approach for dynamic traffic networks*. PhD thesis, Delft University of Technology.

Karlström, A. and Franklin, J. P. (2009). Behavioral adjustments and equity effects of congestion pricing: Analysis of morning commutes during the stockholm trial. *Transportation Research Part A: Policy and Practice*, 43(3):283–296. Stockholm Congestion Charging Trial.

Kitthamkesorn, S. and Chen, A. (2013). A path-size weibit stochastic user equilibrium model. *Transportation Research Part B: Methodological*, 57:378–397.

Kitthamkesorn, S. and Chen, A. (2014). Unconstrained weibit stochastic user equilibrium model with extensions. *Transportation Research Part B: Methodological*, 59:1–21.

Knockaert, J., Bakens, J., Ettema, D., and Verhoef, E. (2010). Spitsmijden: een beloning als middel tegen congestie. *Tijdschrift Vervoerswetenschap*, 46:166–175.

Koppelman, F. S. and Wen, C.-H. (2000). The paired combinatorial logit model: properties, estimation and application. *Transportation Research Part B: Methodological*, 34(2):75–89.

Koutsoupias, E. and Papadimitriou, C. H. (2009). Worst-case equilibria. *Computer Science Review*, 3(2):65–69.

Laval, J. A. and Leclercq, L. (2013). The Hamilton-Jacobi partial differential equation and the three representations of traffic flow. *Transportation Research Part B: Methodological*, 52(0):17–30.

Lawphongpanich, S., Hearn, D., and Smith, M. (2006). *Mathematical and Computational Models for Congestion Charging*. Applied Optimization. Springer.

Lax, P. D. (1957). Hyperbolic systems of conservation laws ii. *Communications on Pure and Applied Mathematics*, 10(4):537–566.

Lebacque, J. and Khoshyaran, M. (2005). First order macroscopic traffic flow models: intersections modeling, network modeling. In *Proceedings of the 16th ISTTT*.

Lebacque, J.-P. (1996). The godunov scheme and what it means for first order traffic flow models. In *Proceedings of the 13th international symposium on transportation and traffic theory (ISTTT)*, pages 647–677.

Leclercq, L. and Bécarie, C. (2012). Meso Lighthill-Whitham and Richards model designed for network applications. In *Proceedings of the 91st Transportation Research Board Annual Meeting (TRB)*.

Levinson, D. (2005). Micro-foundations of congestion and pricing: A game theory perspective. *Transportation Research Part A: Policy and Practice*, 39(7Ű9):691–704.

Levinson, D. (2010). Equity effects of road pricing: A review. *Transport Reviews*, 30(1):33–57.

Levinson, D. and Chang, E. (2003). A model for optimizing electronic toll collection systems. *Transportation Research Part A: Policy and Practice*, 37(4):293–314.

Levinson, D. M. (1998). *On whom the toll falls: A model of network financing*. PhD thesis, Institute of Transportation Studies, University of California at Berkeley.

Levinson, D. M. (1999). Tolling at a frontier: A game theoretic analysis. In Ceder, A., editor, *Proceedings of the 14th International Symposium on Transportation and Traffic Theory*, pages 173–187.

Li, B. (2011). The multinomial logit model revisited: A semi-parametric approach in discrete choice analysis. *Transportation Research Part B: Methodological*, 45(3):461–473.

Lighthill, M. J. and Whitham, G. B. (1955). On kinematic waves. ii. a theory of traffic flow on long crowded roads. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 229(1178):pp.317–345.

Mahmassani, H., Hou, T., and Dong, J. (2012). Characterizing travel time variability in vehicular traffic networks. *Transportation Research Record: Journal of the Transportation Research Board*, 2315:141–152.

Mahmassani, H., Hou, T., and Saberi, M. (2013). Connecting network-wide travel time reliability and the network fundamental diagram of traffic flow. *Transportation Research Record: Journal of the Transportation Research Board*, In Press:1–22.

Mai, T., Fosgerau, M., and Frejinger, E. (2015). A nested recursive logit model for route choice analysis. *Transportation Research Part B: Methodological*, 75:100 – 112.

Maillé, P. and Stier-Moses, N. E. (2009). Eliciting coordination with rebates. *Transportation Science*, 43(4):473–492.

Marzano, V. (2014). A simple procedure for the calculation of the covariances of any generalized extreme value model. *Transportation Research Part B: Methodological*, 70(0):151–162.

Marzano, V., Papola, A., Simonelli, F., and Vitillo, R. (2013). A practically tractable expression of the covariances of the cross-nested logit model. *Transportation Research Part B: Methodological*, 57:1–11.

Mattsson, L.-G., Weibull, J. W., and Lindberg, P. O. (2014). Extreme values, invariance and choice probabilities. *Transportation Research Part B: Methodological*, 59:81–95.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In Zarembka, P., editor, *Frontiers in Econometrics*, pages 105–142. Academic Press, New York.

McFadden, D. (1978). Modelling the choice of residential location. In Karlqvist, A., Lundqvist, L., Snickars, F., and Weibull, J., editors, *Spatial Interaction Theory and Planning Models*. North Holland.

McFadden, D. (2001). Economic choices. *The American Economic Review*, 91(3):pp.351–378.

Morrison, S. A. (1986). A survey of road pricing. *Transportation Research Part A: General*, 20(2):87–97. Special Issue Road Pricing.

Nagurney, A. (1998). *Network economics: A variational inequality approach*, volume 10. Springer.

Nakayama, S. (2013). q-generalized logit route choice an network equilibrium model. *Procedia - Social and Behavioral Sciences*, 80:753–763. 20th International Symposium on Transportation and Traffic Theory (ISTTT 2013).

Nakayama, S. and Chikaraishi, M. (2015). A unified closed-form expression of logit and weibit and its application to a transportation network equilibrium assignment. *Transportation Research Procedia (of ISTTT21)*, 7:59–74.

Nash, J. (1951). Non-cooperative games. *Annals of mathematics*, pages 286–295.

Nash, J. F. et al. (1950). Equilibrium points in n-person games. *Proc. Nat. Acad. Sci. USA*, 36(1):48–49.

Newell, G. (1993a). A simplified theory of kinematic waves in highway traffic, part i: General theory. *Transportation Research Part B: Methodological*, 27(4):281–287.

Newell, G. (1993b). A simplified theory of kinematic waves in highway traffic, part ii: Queueing at freeway bottlenecks. *Transportation Research Part B: Methodological*, 27(4):289–303.

Newell, G. (1993c). A simplified theory of kinematic waves in highway traffic, part iii: Multi-destination flows. *Transportation Research Part B: Methodological*, 27(4):305–313.

Newell, G. (2002). A simplified car-following theory: a lower order model. *Transportation Research Part B: Methodological*, 36(3):195–205.

Newman, J. P. (2008). Normalization of network generalized extreme value models. *Transportation Research Part B: Methodological*, 42(10):958–969.

Ni, D. and Leonard II, J. D. (2005). A simplified kinematic wave model at a merge bottleneck. *Applied Mathematical Modelling*, 29(11):1054–1072.

Nielsen, O. A. (2004). Behavioral responses to road pricing schemes: Description of the danish akta experiment. *Journal of Intelligent Transportation Systems*, 8(4):233–251.

Ohazulike, A. (2014). *Road Pricing Mechanisms: A Game Theoretic and Multi-level Approach*. PhD thesis, University of Twente.

Ortúzar, J. d. D. and Willumsen, L. G. (2011). *Modelling Transport*. John Wiley & Sons, Ltd.

Papola, A. (2004). Some developments on the cross-nested logit model. *Transportation Research Part B: Methodological*, 38(9):833–851.

Papola, A. and Marzano, V. (2013). A network generalized extreme value model for route choice allowing implicit route enumeration. *Computer-Aided Civil and Infrastructure Engineering*, 28(8):560–580.

Parry, I. W. (2009). Pricing urban congestion. *Annual Review of Resource Economics*, 1(1):461–484.

Parry, I. W. H., Walls, M., and Harrington, W. (2007). Automobile externalities and policies. *Journal of Economic Literature*, 45(2):pp.373–399.

Pelzer, P., Geertman, S., van der Heijden, R., Hu, H., Hooimeijer, P., Huang, L., and Kang, J. (2015). Knowledge in communicative planning practice: a different perspective for planning support systems. *Environment and Planning B*, pages 0–0.

Pelzer, P., Geertman, S., van der Heijden, R., and Rouwette, E. (2014). The added value of planning support systems: A practitioner's perspective. *Computers, Environment and Urban Systems*, 48(0):16–27.

Perakis, G. (2007). The "price of anarchy" under nonlinear and asymmetric costs. *Mathematics of Operations Research*, 32(3):614–628.

Peters, H. (2008). *Game Theory: A Multi-Leveled Approach*. Springer-Verlag.

Pigou, A. C. (1920). *The economics of welfare*. McMillan&Co., London.

Prashker, J. and Bekhor, S. (1998). Investigation of stochastic network loading procedures. *Transportation Research Record: Journal of the Transportation Research Board*, 1645(-1):94–102.

Prashker, J. N. and Bekhor, S., S. (2004). Route choice models used in the stochastic user equilibrium problem: A review. *Transport Reviews*, 24(4):437–463.

Prato, C. G. (2009). Route choice modeling: past, present and future research directions. *Journal of Choice Modelling*, 2:65–100.

Prato, C. G. (2012). Meta-analysis of choice set generation effects on route choice model estimates and predictions. *Transport*, 27(3):286–298.

Pravinvongvuth, S. and Chen, A. (2005). Adaptation of the paired combinatorial logit model to the route choice problem. *Transportmetrica*, 1(3):223–240.

Proost, S. and Sen, A. (2006). Urban transport pricing reform with two levels of government: a case study of brussels. *Transport Policy*, 13(2):127–139.

Raadsen, M. P., Bliemer, M. C., and Bell, M. G. (2014a). An efficient event-based algorithm for solving first order dynamic network loading problems. Technical Report ITLS-WP-14-10. ISSN 1832-570X, Institute of Transport and Logistics Studies, The University of Sydney.

Raadsen, M. P., Bliemer, M. C., and Bell, M. G. (2014b). An efficient event-based algorithm for solving first order dynamic network loading problems. In *Proceedings of the 5th International Symposium on Dynamic Traffic Assignment, Salerno, Italy*.

Raadsen, M. P., Bliemer, M. C., and Bell, M. G. (2016). An efficient and exact event-based algorithm for solving simplified first order dynamic network loading problems in continuous time. *Transportation Research Part B: Methodological*, 92, Part B:191 – 210. Within-day Dynamics in Transportation Networks.

Ramming, M. S. (2002). *Network knowledge and route choice*. PhD thesis, Massachusetts Institute of Technology. Dept. of Civil and Environmental Engineering.

Ramos, G. M. (2015). *Dynamic Route Choice Modelling of the Effects of Travel Information using RP Data*. PhD thesis, Delft University of Technology.

Ramos, G. M., Frejinger, E., Daamen, W., and Hoogendoorn, S. (2012). A revealed preference study on route choices in a congested network with real-time information. In *Proceedings of the 13th International Conference on Travel Behaviour Research-Toronto*, pages 15–20.

Richards, P. I. (1956). Shock waves on the highway. *Operations Research*, 4(1):42–51.

Roughgarden, T. (2003). The price of anarchy is independent of the network topology. *Journal of Computer and System Sciences*, 67(2):341–364. Special Issue on STOC 2002.

Roughgarden, T. and Tardos, E. (2002). How bad is selfish routing? *J. ACM*, 49(2):236–259.

Schaap, N., Jorritsma, P., Derriks, H., and Berveling, J. (2014). Meer tijd- en plaatsonafhankelijk werken: kansen en barrières. Technical report, Kennisinstituut voor Mobiliteitsbeleid.

Schulz, A. S. and Stier-Moses, N. E. (2006). Efficiency and fairness of system-optimal routing with user constraints. *Networks*, 48(4):223–234.

Shapley, L. S. (1952). A value for n-person games. Technical report, DTIC Document.

Sion, M. and Wolfe, P. (1957). On a game without a value. *Contributions to the theory of games*, 3:299–306.

Smaal, M. (2012). *Politieke strijd om de prijs van automobiliteit*. PhD thesis, Tilburg University.

Small, K. A. (2012). Valuation of travel time. *Economics of Transportation*, 1(1Ű2):2–14.

Small, K. A., Winston, C., and Yan, J. (2005). Uncovering the distribution of motorists' preferences for travel time and reliability. *Econometrica*, 73(4):1367–1382.

Smits, E.-S., Bliemer, M., Pel, A., and van Arem, B. (2014). On route choice models with closed-form probability expressions. In *The Transportation Research Board (TRB) 93rd Annual Meeting, Washington DC, January 12-16, 2014 (paper no 14-3733).*, pages 1 – 24.

Smits, E.-S., Bliemer, M. C., Pel, A. J., and van Arem, B. (2015). A family of macroscopic node models. *Transportation Research Part B: Methodological*, 74:20–39.

Smits, E.-S., Pel, A., van Nes, R., and van Arem, B. (2016). Modelleren mobiliteitsbeprijzing met meerdere actoren. *Tijdschrift Vervoerwetenschap*. ISSN 1571-9227.

Solehmainen, J. (2011). Choice models in relation with road pricing schemes and designs. Master's thesis, Aalto University.

Solodov, M. and Tseng, P. (2000). Some methods based on the d-gap function for solving monotone variational inequalities. *Computational Optimization and Applications*, 17(2-3):255–277.

Spiess, H. and Florian, M. (1989). Optimal strategies: A new assignment model for transit networks. *Transportation Research Part B: Methodological*, 23(2):83–102.

Tampère, C. M., Corthout, R., Cattrysse, D., and Immers, L. H. (2011). A generic class of first order node models for dynamic macroscopic simulation of traffic flows. *Transportation Research Part B: Methodological*, 45(1):289–309.

Tijs, S. and Otten, G.-J. (1993). Compromise values in cooperative game theory. *Top*, 1(1):1–36.

Tijs, S. H. (1981). Bounds for the core and the τ-value. *Game Theory and Mathematical Economics, North-Holland, Amsterdam*, pages 123–132.

Tirachini, A., Hensher, D. A., and Bliemer, M. C. (2014). Accounting for travel time variability in the optimal pricing of cars and buses. *Transportation*, pages 1–25.

Train, K. (2009). *Discrete Choice Methods with Simulation*. Cambridge University Press.

Tripsigala, E. (2014). A game theoretical approach to road pricing with multiple stakeholders. Master's thesis, TU Delft, Delft University of Technology.

Tsekeris, T. and Voß, S. (2009). Design and evaluation of road pricing: state-of-the-art and methodological advances. *Netnomics*, 10:5–52.

Ubbels, B. (2006). *Road Pricing: Effectiveness, Acceptance and Institutional Aspects*. PhD thesis, Vrij Universiteit Amsterdam.

Ubbels, B. and Verhoef, E. T. (2008). Governmental competition in road charging and capacity choice. *Regional Science and Urban Economics*, 38(2):174–190.

van Amelsfort, D. H. (2009). *Behavioural responses and network effects of time-varying road pricing*. PhD thesis, Delft University of Technology.

van den Berg, V. and Verhoef, E. T. (2011). Congestion tolling in the bottleneck model with heterogeneous values of time. *Transportation Research Part B: Methodological*, 45(1):60–78.

van der Sar, J. and Baggen, J. (2005). Prijsbeleid op de weg in nederland. een historisch overzicht. In *Colloquium Vervoersplanologisch Speurwerk*.

van der Weijde, A. H., Verhoef, E. T., and van den Berg, V. A. (2013). Competition in multi-modal transport networks: A dynamic approach. *Transportation Research Part B: Methodological*, 53(0):31–44.

van Eck, G., Brands, T., Wismans, L., Pel, A., and van Nes, R. (2014). Model complexities and requirements for multimodal transport network design. *Transportation Research Record: Journal of the Transportation Research Board*, 2429(-1):178–187.

van Eck, G., van Nes, R., and van Arem, B. (2012). Efficient choice set generation in large scale dynamic multimodal transport networks. In *Proceedings of the 13th International Conference on Travel Behaviour Research-Toronto*.

van Essen, H., Nelissen, D., Smit, M., van Grinsven, A., Aarnink, S., Breemersch, T., Martino, A., Rosa, C., Parolin, R., and Harmsen, J. (2012). An inventory of measures for internalising external costs in transport. Technical report, European Commission.

van Wageningen-Kessels, F., van Lint, H., Vuik, K., and Hoogendoorn, S. (2014). Genealogy of traffic flow models. *EURO Journal on Transportation and Logistics*, pages 1–29.

Verhoef, E., Koopmans, C., Bliemer, M., Bovy, P., Steg, L., and van Wee, B. (2004a). Vormgeving en effecten van prijsbeleid op de weg. Technical Report ISBN 90-6733-275-5, Stichting voor Economisch Onderzoek (SEO).

Verhoef, E. T., Small, K. A., and Small, K. A. (2004b). Product differentiation on roads: constrained congestion pricing with heterogeneous users. *Journal of transport economics and policy*, pages 127–156.

Vickrey, W. S. (1969). Congestion theory and transport investment. *The American Economic Review*, 59(2):pp.251–260.

Vonk Noordegraaf, D., Annema, J. A., and van Wee, B. (2014). Policy implementation lessons from six road pricing cases. *Transportation Research Part A: Policy and Practice*, 59(0):172–191.

Vovsha, P. and Bekhor, S. (1998). Link-nested logit model of route choice: Overcoming route overlapping problem. *Transportation Research Record: Journal of the Transportation Research Board*, 1645:133–142.

Vrtic, M. (2009). The impacts of road pricing on route and mode choice behaviour. *Journal of Choice Modelling*, 3(1):109–126.

Wardman, M., Chintakayala, P., de Jong, G., and Ferrer, D. (2012). European wide meta-analysis of values of travel time. Technical report, University of Leeds. Prepared for the European Investment Bank.

Wardrop, J. (1952). Some theoretical aspects of road traffic research. *Proceedings of the Institution of Civil Engineers, Part II*, 1(36):352–362.

Warffemius, P. M. J. (2013). *De maatschappelijke waarde van kortere en betrouwbaardere reistijden*. Kennisinstituut voor Mobiliteitsbeleid.

Watling, D. P., Rasmussen, T. K., Prato, C. G., and Nielsen, O. A. (2015). Stochastic user equilibrium with equilibrated choice sets: Part i – model formulations under alternative distributions and restrictions. *Transportation Research Part B: Methodological*, 77(0):166–181.

Wismans, L. J. J. (2012). *Towards sustainable dynamic traffic management*. University of Twente.

Wu, D. (2011). *Designing more acceptable and equitable congestion pricing schemes for multimodal transportation networks*. PhD thesis, University of Florida.

Wu, D., Yin, Y., and Lawphongpanich, S. (2010). Pareto-improving congestion pricing on multimodal transportation networks. *European Journal of Operational Research*, In Press, Corrected Proof:–.

Xu, X., Chen, A., Kitthamkesorn, S., Yang, H., and Lo, H. K. (2015). Modeling absolute and relative cost differences in stochastic user equilibrium problem. *Transportation Research Part B: Methodological*, pages –.

Yai, T., Iwakura, S., and Morichi, S. (1997). Multinomial probit with structured covariance for route choice behavior. *Transportation Research Part B: Methodological*, 31(3):195–207.

Yang, H. and Huang, H.-J. (2005). *Mathematical and Economic Theory of Road Pricing*. Elsevier.

Yildirimoglu, M., Koymans, A., and Geroliminis, N. (2013). Exploring the properties of mean-variance relations in freeway travel times. In *Transportation Research Board 92nd Annual Meeting*.

Yperman, I. (2007). *The Link Transmission Model for dynamic network loading*. Open access publications from katholieke universiteit leuven, Katholieke Universiteit Leuven.

Zhang, X., Zhang, H., Huang, H.-J., Sun, L., and Tang, T.-Q. (2011). Competitive, cooperative and stackelberg congestion pricing for multiple regions in transportation networks. *Transportmetrica*, 7(4):297–320.

Zhou, B., Bliemer, M. C. J., Bell, M. G. H., and He, J. (2015). Two new methods for solving the path-based stochastic user equilibrium problem. *Computer-Aided Civil and Infrastructure Engineering*, pages n/a–n/a.

Zhou, Z., Chen, A., and Bekhor, S. (2012). C-logit stochastic user equilibrium model: formulations and solution algorithm. *Transportmetrica*, 8(1):17–41.

Zhou, Z., Chen, A., and Wong, S. (2009). Alternative formulations of a combined trip generation, trip distribution, modal split, and trip assignment model. *European Journal of Operational Research*, 198(1):129–138.

Acronyms

CO_2 Carbon Dioxide. 27, 182, 183
NO_X Nitrogen Oxides. 27, 182, 183
PM_{10} Particle Matter $<$ 10 micrometre. 27, 182, 183

A Additive. 73, 76, 90, 92, 93, 96, 99, 206
A-LN Additive Link-Nested. 79, 98
A-MN Additive Multinomial. 77, 84, 90, 93–95, 98, 177
A-PC Additive Paired Combinatorial. 79, 98
A-PS Additive Path-Size. 78, 90, 98

CCEP Capacity Consumption Equivalence node Problem. 124, 127, 129, 131, 136
CO Carbon Monoxide. 27, 182, 183

DCPP Directed Capacity Proportional node Problem. 122, 124, 127, 129, 131, 136, 177
DNL Dynamic Network Loading. 39, 41, 49, 105, 106, 136
DTA Dynamic Traffic Assignment. 37, 41, 205

EDP Equal Delay node Problem. 121, 124, 127, 128, 131, 136
EV Extreme Value. 72

GCNM Generic Class of first-order Node Models. 106, 107, 112, 115, 117, 119, 127, 133–135
GEV Generalized Extreme Value. 72
GMEV Generalized Multivariate Extreme Value. 61, 63–65, 72–74, 77, 80, 84, 88, 103, 173, 176, 194, 198, 200, 204, 205, 233
GPS Global Positioning System. 67

KWT Kinematic Wave Theory. 33, 34, 40, 41, 49, 51–53, 55, 201, 202

LN Link-Nested. 47, 79, 95, 97, 99, 102, 103, 205
LTM Link Transmission Model. 49, 52, 53

M Multiplicative. 74, 76, 88–90, 92, 93, 99, 206
M△ Multiplicative with Reference Route. 87, 89–93, 95, 96, 99, 103, 206
M△-LN Multiplicative Link-Nested with Reference Route. 91, 98
M△-MN Multiplicative Multinomial with Reference Route. 90, 92–95, 97, 98
M△-PC Multiplicative Paired Combinatorial with Reference Route. 90, 98
M△-PS Multiplicative Path-Size with Reference Route. 90, 97, 98, 103, 206
M-LN Multiplicative Link-Nested. 80, 98

SUMMARY

STRATEGIC NETWORK MODELLING FOR PASSENGER TRANSPORT PRICING

In the last decade the Netherlands has experienced an economic recession. Now, in 2017, the economy is picking up again. This growth does not only come with advantages because economic growth demands more from the transport system. Congestion is increasing again, the capacity of the train system is now insufficient during peak hours, and the world faces environmental challenges that are partly due to emissions caused by travellers. These negative effects worsen as travellers make rational choices, which could be undesirable from a system, or social welfare, perspective. For example, car drivers do often not choose public transport options, because it costs them more effort; however, if they choose public transport options, then the system improves since congestion and emissions will reduce. Or another example, if travellers choose to avoid peak hours, they might not arrive at their desired time, but then they do not contribute to peak hour congestion or crowding. In addition, the capacity of the transport system is more effectively used if travellers spread out over the day.

Passenger transport pricing can be an incentive for travellers to change their choices, and can therefore be used to mitigate congestion, emissions, and other undesirable effects. Passenger transport pricing is the umbrella term for measures that make passengers pay for their travels. Traditional pricing measures are for example: fuel excise taxes, public transport fares, and periodical registration fees for vehicles. More innovative measures are cordon charges (e.g., in London, Stockholm, and Singapore), special tolling lanes, and peak avoidance projects. When such an innovative measure has different prices for times of the day, and for different locations (i.e., it is time- and space-differentiated), travellers' choices related to route, mode and departure time can be influenced. By changing these choices, the overall performance of the transport system can improve. Travellers have differences regarding time valuation, preferred departure or arrival times, and car ownership. Therefore, a measure can become even more effective if it also allows to differentiate amongst characteristics of travellers.

However, innovative pricing measures have not been implemented widely across the globe, despite their potential to reduce congestion and emissions. This is primarily due to lack of public and political support. The Netherlands has experienced decades of political discourse and many failed proposals. Low public support did not contribute to (political) agreement either, because it has always fuelled the discussion with dissenting opinions. In the process of designing policies and making decisions, *strategic planning models* usually estimate (or forecast) the effects of the policy. The preferences of travellers and the transport system are captured by mathematical equations. Such models are always a simplified representation of reality. To apply them to asses pricing measures, they should capture the underlying mecha-

nisms that are important for transport pricing as realistic as possible.

This dissertation identifies disadvantages of current strategic network models for passenger transport pricing and provides methodological advances to resolve them. This is done with a holistic approach that combines game theory, discrete choice analysis, traffic flow theory, and tranport economics into one modelling framework. This framework has many sub-models and provides a toolbox for analysts to determine the effects of innovative pricing schemes. The basic principle for each tool is to make them realistic (so that the results are credible for decision makers), and computationally efficient. The latter means that many different pricing schemes can be computed within reasonable time. By providing the methodological advances, that are briefly discussed in the next sections, this dissertation aims to improve public and political support. For example, the preferences of multiple stakeholders can be considered, the possible conflicts between them can be identified, and solutions based on concepts that aim to resolve these conflicts can be computed.

Quasi-Dynamic Traffic Assignment

The holistic model framework consists of two parts. The upper level discusses the price setting and decision making; this leads to a certain pricing scheme. The lower level, Quasi-Dynamic Traffic Assignment (QDTA), is presented in Part I of this dissertation. It calculates the effects of the pricing scheme in terms of effect levels. Examples of these effects are congestion, emissions, and revenues. Usually, the effect levels can be retrieved from the conditions within the transport system. To determine these, the model computes the choices that travellers make. The QDTA model includes mode, route and departure time choice. The consequences of these choices are subsequently calculated by the Quasi-Dynamic Network Loading (QDNL) model that determines where vehicles have to queue and how much congestion occurs.

The embedded choice model is a Random Utility Maximization (RUM) model that models route, mode, and time-of-day choice simultaneously. Contrary to traditional logit models, the error term enters the utility formulation as a factor. Dependencies between choices are captured with a nest structure. A special mode represents the stay-at-home alternative. This provides trip choice (i.e., do I travel or not?) for travellers implicitly. By having these choices available, the models is able to capture the most important short-term responses of travellers to innovative pricing measures.

The used QDNL model takes a hybrid approach that extends static assignment models (e.g., models based on travel time functions) with capacity constraints. This means that there is a limit on the number of vehicles that enter a road. If for some road more vehicles want to enter than it has as capacity, this road is called a bottleneck. In reality queues will form upstream of bottlenecks. Traditional static models calculate delays – inside – the bottleneck. Therefore, it is difficult to calculate plausible delay times, and vehicles are potentially assumed to be unaffected. The first phase of the QDNL model identifies bottlenecks in the network. The second phase determines how long the queues upstream of each of the bottlenecks are. This approach makes the representation of congestion more realistic, with a limited increase in computation costs.

Finally, the QDTA model, that combines the simultaneous choice model and the network loading model, is presented as a variational inequality. This is a mathematical problem formulation that allows derivation of analytical properties. For example, the duality gap can be computed that allows the model developers to determine the accuracy of the simulated solutions, which in turn can be used to improve the solution algorithms.

Generalized Multivariate Extreme Value Choice Models

To select a choice model for the QDTA model, an in-depth analysis of various RUM-based route choice models has been performed. RUM is a frequently used framework to analyse discrete choices of travellers. It assumes that travellers assign a utility to each choice alternative, and that they subsequently choose the alternative with the highest utility. Since the utilities are random variates, the choices become probabilities. The group of analysed route choice models consists of those models for which the routes are known in advance, and for which the choice probability formula is closed-form (i.e., no simulation is required to calculate them). These types of route choice models are frequently used in practice.

Several ongoing challenges exist for these models. It is not straightforward to handle route overlap with such models. This is important because similar alternatives should not be accounted as independent. To see this, consider the classical red bus, blue bus example. Assume a commuter chooses between a red bus and the train with equal probability. If a blue bus is introduced which is the same as the red bus, except for the colour, then it is expected that the commuter chooses between train and *any* of the buses with equal probability. So, with a probability of 25% for each bus. Choice models that do not account for overlap assume that all three alternatives are independent, meaning that each alternative has a choice probability of 1/3, which is not expected. Another challenges is that the uncertainty of a long route's utility is higher than the uncertainty of a short route's utility. So, the variance of route utility depends on the length of the route. If one would like apply the same choice model on all possible origin-destination pairs in a network, the variance of route utilities should depend on the distance between the origin and destination.

In this dissertation, twelve models are analysed on, amongst others, these characteristics. In addition, seven of the twelve presented models are new. De described models are all Generalized Multivariate Extreme Value (GMEV)-models. The models differ in how the error terms enter the utility formulation, and in how the dependencies are introduces between error terms for different routes. The error terms can either enter as a term (additive models), or as a factor (multiplicative models); in addition, the dissertation presents multiplicative models that are relative to a reference route, for which the error only enter as a factor for the non-overlapping parts with the reference route. Four different structures of dependency can be added to the error terms, leading to four model types (viz., multinomial, path-size, paired-combinatorial, and link-nested). All models are assessed qualitatively and quantitatively.

The qualitative assessment is based on desired properties of the random utility formulation. An important – empirically validated – property is the linear relation between the average and the standard deviation of travel time. This contradicts the assumption of previously often used route choice models that a linear relation between the average and the variance of travel time exists. Analysis of the behaviour of the models under simple network changes shows that only the multiplicative models based on reference routes can reproduce the expected behaviour.

The quantitative assessment constitutes the estimation and validation of the models on two route sets in a carefully constructed network. This network has overlapping routes, and short and long distances. All models are estimated on the first route set and then validated on the second set and vice versa. The assessment shows that good performance on real networks can be expected from both multiplicative path-size models (with and without reference routes). Additional empirical validation is required to conclusively assess all models.

Macroscopic Node Models

An important part of the QDNL model is the node model. Such a model computes the flow at for example merges, diverges, lane drops, and intersections in the road network. Node models are also used to identify bottlenecks in the network. Only in 2011 a set of requirements for node models has been formulated. Compared to link models (i.e., models that propagate traffic over homogeneous road stretches), node models have not received a lot of attention in the literature. In addition, no complete insight on the underlying behaviour of vehicles at nodes is known for existing models.

This dissertation analyses and discusses four node models – of which two are new – that satisfy the recently formulated requirements. Discussion on the representation of traffic at nodes leads to the concept of turn delays, which are the additional time vehicles have to stay at the node. Turn delays are variables in which node models can be formulated, and that simultaneously provide a behavioural interpretation of the underlying mechanisms of the model.

The set of requirements is reformulated in terms of turn delays. This leads to a family of node models represented by a multi-objective optimization problem. Any method that finds a Pareto optimal solution of this problem is member of the node model family. Two different sets of turn delays can lead to the same traffic flows over the node. Therefore, also the notion of reduced capacity is introduced to show if two results are equivalent. It turns out that for diverges, all four models are equivalent, and that for merges three out of four models are equivalent.

One existing and one new model have efficient solution methods. The new model calculates the turn delays based on the theory of polynomials in the max-plus algebra. On the other hand, these two models have less behavioural realism than the other two. Unfortunately, the latter use fixed point methods that require much more computation time. So, while choosing a node model, one has to make a trade-off between computation time and realism.

Game Theory for Multiple Stakeholders

The upper level of the holistic model framework is the decision making level that represents the interactions between multiple stakeholders, and is presented in Part II of this dissertation. An important characteristic is that each stakeholder pursues its goal by applying some strategy. The latter consists of a price for their ‘travel product’. The stakeholders need to consider the responses of the travellers, and the resulting effects, when they determine their strategy. This response is captured in the lower level QDTA-model. Transport economists sometimes use social welfare optimization and first-best pricing to resolve the upper level. The multi-stakeholder approach in this dissertation relaxes their assumption that every individual participated at the negotiation table, and unlimited transactions between them can exist. This is done by exploring game theories that have not been applied to similar problems before.

The corresponding mathematical optimization problem of the upper level is a multi-objective optimization problem with equilibrium constraints. This means that solutions are not straightforward to determine. This dissertation uses game theory to retrieve a solution for this problem. Game theory allows one to capture the behaviour of interacting stakeholders at the negotiation table. Stakeholders can form coalitions, and stakeholders in coalitions can cooperate to thwart other stakeholders and coalitions. This dissertation introduces a method using Nash-equilibria and the notion of coalition formation to convert the multi-objective problem formulation to a Transferable Utility (TU)-game. In such a game every coalition has a certain value that it can bring to the negotiation table.

Game theory provides multiple solution concepts for TU-games. The core, Shapley value, and compromise value are presented as potential solutions for the upper level. The pricing scheme that optimizes the sum of all stakeholders' objectives is played in all solutions of the TU-game. Each of the solution concepts suggests monetary transfers that should be fair for all of the stakeholders. Some stakeholders could be compensated for missed gains in the overall best pricing scheme. Each concept uses a different interpretation of what is 'fair'.

Case Study: The Randstad

To show the feasibility of the holistic model framework approach for real applications, a case study of the Randstad area is presented. All methodological advances come together in the case study. Compared to traditional static models, realism has been added in the lower level QDTA-model, without requiring much more computation time. Multiple stakeholders, their objectives, and their possible conflicts at the negotiation table, can be analysed with the game theory-based upper level. Three stakeholders are involved, each with a different objective and pricing measure. The national government aims to improve social welfare by implementing a kilometre charge. The municipality of Amsterdam wants to improve its economic position with a cordon charge. Finally, the train operator optimizes revenues by changing the train fares. All stakeholders can differentiate their prices for peak and shoulder periods. Although the objectives of the stakeholders are hypothetical, most of the used sources are from existing strategic planning models and empirical research. Almost 90 000 links exist in the road network. Travellers can choose between car and public transport (train) options, between peak and shoulder periods to travel, or they can choose to stay at home. In addition, car drivers have many route alternatives.

The QDTA-equilibrium, with a multiplicative nested choice model and the QDNL-model, has been computed for 216 feasible pricing schemes. This forecasts, amongst others, the congestion levels, emission levels, and revenues for each stakeholder under each pricing scheme. Based on these evaluations, a TU-game is formulated. Successively, three solution concepts from TU-game theory are analysed (viz., the core, Shapley value and compromise value). They provide possible arguments that can be used by the stakeholders at the negotiation table.

Despite the case study uses some simplifying assumptions and hypothetical stakeholder objectives, it provides some interesting results. The national government and the train operator have conflicting interests. After negotiations, the national government has to compensate the train operator. Also, if the stakeholders are not allowed to cooperate, the national government will be better off. This might sound paradoxically, but the explanation lies in the coalition that can be formed between the municipality of Amsterdam and the train operator. This coalition can use its combined pricing measures to steer away from the good situation of the national government.

Comparing the zero-scenario with the solutions under pricing, shows the potential of the innovative pricing measures. In the morning peak, pricing achieves a reduction of 60% of loss hours. Overall, a 45% reduction is possible. The emissions in the Randstad area can be reduced by about 6%. This is primarily due to a large modal shift towards the train, and a steep increase of passengers that choose not to travel. The sum of the three stakeholders' objectives (i.e., social welfare, economic position of Amsterdam, and the profit of the train operator) increases with 1.4 M€ during each morning commute.

The conclusion of the case study is primarily that it is feasible to assess innovative pricing measures in a multi-stakeholder setting and by means of model components for which the realism has been improved. To retrieve conclusive results on the effects of pricing, and the interactions between stakeholders, a more detailed study has to be set up, with more modes, user-types, and stakeholder objectives that stem from interviews.

DUTCH SUMMARY (SAMENVATTING)

STRATEGISCHE NETWERKMODELLERING VOOR BEPRIJZING VAN PERSONENVERVOER

Gedurende het afgelopen decennium kende Nederland periodes van economische recessie. Tegenwoordig, in 2017, groeit de economie weer. Dit brengt niet alleen voordelen met zich mee, want de economische groei leidt ook tot een zwaardere belasting van het vervoerssysteem. De congestie neemt weer toe, de capaciteit van het treinsysteem is nu onvoldoende tijdens de spits, en de wereld staat voor milieuproblematiek waar de emissies veroorzaakt door reizigers gedeeltelijk debet aan zijn. Deze negatieve effecten verslechteren omdat reizigers rationeel hun keuzes maken, welke vanuit het oogpunt van het systeem of sociale welvaart ongewenst is. Bijvoorbeeld, automobilisten kiezen niet voor alternatieven met het openbaar vervoer omdat deze meer inspanning vergen; daarentegen zouden files en de uitstoot van schadelijke gassen afnemen wanneer vaker voor het openbaar vervoer gekozen wordt. Een ander voorbeeld is dat als reizigers ervoor kiezen buiten de spits te reizen, dat ze dan wellicht niet op hun gewenste aankomsttijd arriveren, maar dat ze dan geen bijdrage leveren aan de congestie of volle treinen. Bovendien wordt de capaciteit van het vervoerssysteem beter benut wanneer reizigers zich spreiden over de dag.

Beprijzing van personenvervoer kan als stimulans dienen voor reizigers om hun keuzes aan te passen, en prijsbeleid kan daarom gebruikt worden om congestie, emissies en andere ongewenste effecten te verminderen. Beprijzing van personenvervoer is een overkoepelende term voor maatregelen die reizigers laten betalen voor het gebruik van het vervoerssysteem. Traditionele beprijzingsmaatregelen zijn bijvoorbeeld: accijnzen, tarieven van het OV en periodieke motorrijtuigbelasting. Meer innovatieve maatregelen zijn cordon heffingen (e.g., in London, Stockholm en Singapore moet betaald worden om de stad in te rijden), speciale rijstroken met tol, en spitsmijden projecten (waarbij reizigers beloond worden om niet in de stpits te rijden). Indien de tarieven voor zulke innovatieve maatregelen onderscheiden zijn tussen spits- en dalperiodes, en tussen verschillende locaties (i.e., de prijzen zijn gedifferentieerd in tijd en ruimte), dan kunnen de keuzes van reizigers met betrekking tot route, vervoerswijze en vertrektijdstip beïnvloed worden. Door deze keuzes te veranderen kan de prestatie van het vervoerssysteem verbeterd worden. Reizigers zijn daarnaast verschillend met betrekking tot hoe ze tijd waarderen, hun gewenste vertrek- en aankomsttijd, en of ze een auto bezitten. Daarom kunnen maatregelen nog effectiever worden gemaakt wanneer ze onderscheid maken tussen reizigers met verschillende eigenschappen.

Echter, innovatieve beprijzingsmaatregelen zijn wereldwijd niet standaard geïmplementeerd – ondanks hun potentie om congestie en emissies te verminderen–. Dat komt hoofdzakelijk door het gebrek aan politieke en publieke steun. Nederland heeft de afgelopen decennia ruime ervaring opgedaan met politiek discours en vele mislukte voorstellen. Dat er weinig publieke steun was heeft niet begedragen aan het bereiken van een (politiek) akkoord omtrent innovatieve maatregelen. Dat heeft slechts de discussie gevoed met afwijzende meningen. Gedurende het proces van beleidsontwikkeling en besluitvorming worden doorgaans *strategische planningsmodellen* gebruikt om de effecten van het beleid te schatten (of te voorspellen). Het vervoerssysteem en de reizigersvoorkeuren worden daarin beschreven met wiskundige vergelijkingen. Zulke modellen zijn altijd een versimpelde weergave van de werkelijkheid. Om ze toe te kunnen passen om beprijzingsmaatregelen te evalueren, dienen de onderliggende mechanismen die belangrijk zijn voor het beprijzen zo realistisch mogelijk worden meegenomen, want alleen dan kunnen ze door beleidsmakers als geloofwaardig worden gezien.

Dit proefschrift identificeert de nadelen van huidige strategische netwerkmodellen die gebruikt worden voor het beoordelen van beprijzing van personenvervoer en levert methodologische ontwikkelingen om deze op te lossen. Dit wordt gedaan met een holistische aanpak waarin speltheorie, analyse van discrete keuzes, verkeersstroomtheorie en vervoerseconomie gecombineerd worden binnen één raamwerk. Dit raamwerk bevat verscheidende deelmodellen en levert een gereedschapskist voor analisten op, die daarmee de effecten van innovatieve beprijzingsmaatregelen kunnen bepalen. Het uitgangspunt bij ieder instrument is dat deze zowel realistisch (zodat ze geloofwaardig zijn voor beleidsmakers) als rekenefficiënt zijn. Het laatste betekent dat veel verschillende beprijzingsschema's doorgerekend kunnen worden binnen afzienbare tijd. Door het leveren van methodologische ontwikkelingen, welke in de volgende paragrafen beknopt worden toegelicht, richt dit proefschrift zich op het verbreden van politieke en publieke steun voor innovatieve beprijzingsmaatregelen. Zo kunnen bijvoorbeeld de voorkeuren van meerdere actoren meegenomen worden, hierdoor is het mogelijk eventuele conflicten tussen actoren te identificeren en oplossingen op basis van verschillende concepten voor deze conflicten te berekenen.

Quasi-Dynamische Verkeerstoedeling

Het holistische raamwerk bestaat uit twee onderdelen (of niveaus). Het bovenste niveau bespreekt het bepalen van de prijs en de besluitvorming. Het onderste niveau, Quasi-Dynamische Verkeerstoedeling (Quasi-Dynamic Traffic Assignment (QDTA) in het Engels), wordt besproken in Deel (Part) I van dit proefschrift. Het berekent de effecten van een beprijzingsschema en levert deze op in termen van effect niveaus. Voorbeelden hiervan zijn congestie, emissies en (tol-)opbrengsten. Doorgaans kunnen deze effecten worden afgeleid uit de condities die gelden binnen het vervoerssysteem. Om deze te kunnen bepalen berekent het model de keuzes die reizigers maken. Het QDTA model bevat vervoerswijze, route en verstrektijdstip keuze. De consequenties van deze keuzes worden vervolgens berekend middels een Quasi-Dynamisch Netwerkbelasting (Quasi-Dynamic Network Loading (QDNL) in het Engels) model dat bepaalt welke voertuig in een wachtrij komen te staan en hoeveel vertraging daarbij opgelopen wordt.

Het ingebouwde keuzemodel is een stochastisch nutsmaximalisatie (Random Utility Maximization (RUM) in het Engels) model dat route, vervoerswijze en periode (i.e., moment van de dag) simultaan modelleert. In tegenstelling tot traditionele logit modellen komt de fout-term als een factor terecht in de nutsformulering. Afhankelijkheden tussen de keuzeopties zijn opgevangen met een nest-structuur. Een speciale vervoerswijze representeert het thuisblijf-alternatief, welke impliciet een reiskeuze (i.e., ga ik wel of niet op pad?) levert. Door deze verschillende keuzes voorhanden te hebben, is het mogelijk alle belangrijke korte termijn reacties van reizigers op innovatieve beprijzingsmaatregelen mee te nemen binnen het model.

Het gebruikte QDNL model bestaat uit een hybride aanpak die statische toedelingsmodellen (e.g., modellen die op gebaseerd zijn op reistijdfuncties) uitbreid met capaciteitsrestricties. Dat betekent dat er een limiet zit aan het aantal voertuigen dat een weg op kan rijden – de capaciteit. Indien voor een bepaalde weg de verkeersvraag in voertuigen hoger is dan de capaciteit, dan is die weg het zogeheten kiemlocatie. In werkelijkheid zullen de wachtrijen stroomopwaarts van deze kiemlocaties ontstaan. De traditionele statische modellen berekenen de vertraging – binnenin – de kiemlocatie. Daarom is het lastig daarmee plausibele vertragingstijden te berekenen, en kan het zo zijn dat wordt aangenomen dat voertuigen onterecht niet beïnvloed zijn. De eerste fase van het QDNL model identificeert alle kiemlocaties in het netwerk. De tweede fase bepaalt hoe lang de wachtrijen stroomopwaarts van de kiemlocaties zijn. Deze aanpak verbetert de manier waarop congestie gerepresenteerd wordt in modellen aanzienlijk, terwijl de afname van rekenefficiëntie slechts beperkt is.

Tenslotte is het QDTA-model, dat het simultane keuzemodel en de het netwerkbelasting model combineert, geformuleerd als een variationele ongelijkheid. Dat is een wiskundige formulering die het mogelijk maakt eigenschappen van het model analytisch te herleiden. Zo is het bijvoorbeeld mogelijk het dualiteitsgat te berekenen dat de ontwikkelaar in staat stelt om de nauwkeurigheid van de oplossing te bepalen, wat vervolgens weer gebruikt kan worden om oplossingsalgoritmes te verbeteren.

Gegeneraliseerde Multivariate Extreme Waarde Keuzemodellen

Om een keuzemodel te selecteren voor het QDTA-model is een uitgebreide analyse uitgevoerd van verschillende RUM-gebaseerde routekeuzemodellen. RUM is een veelgebruikt raamwerk om de discrete keuzes van reizigers te analyseren. Het gaat er vanuit dat een reiziger een bepaalde hoeveelheid nut toekent aan ieder keuzealternatief, en dat deze vervolgens het alternatief kiest met het hoogste nut kiest. Aangezien het nut stochastisch is, wordt ieder alternatief met een bepaalde kans gekozen. De groep van routekeuzemodellen die meegenomen zijn binnen de analyse, zijn die modellen waarvan wordt aangenomen dat de routes vooraf bekend zijn, en waarvoor de keuzekans formule een gesloten vorm heeft (i.e., er zijn geen simulaties nodig om deze te berekenen). Dit type routekeuzemodellen zijn vaak gebruikt in de praktijk.

Er zijn meerdere uitdagingen voor deze modellen. Het is niet eenvoudig om overlappende routes goed mee te nemen in deze modellen. Dat is belangrijk omdat op elkaar lijkende alternatieven niet als onafhankelijk mogen worden gezien. Beschouw het volgende klassieke voorbeeld met een rode bus en een blauwe bus. Neem aan dat een forens met gelijke kans kiest tussen een rode bus en een trein. Wanneer er nu een blauwe bus wordt geïntroduceert die volledig, op de kleur na dan, gelijk is aan de blauwe bus, dan is verwachting dat de forens met gelijke kans kiest tussen de trein en *één van de* bussen. Dus, de keuzekans voor iedere bus is 25%. Keuzemodellen die deze overlap, of gelijkenis, niet corrigeren gaan er vanuit gaan dat alle drie de alternatieven onafhankelijk zijn, wat leidt tot een keuzekans van $1/3$ voor ieder alternatief, wat niet in de lijn der verwachting ligt. Een andere uitdaging is dat de onzekerheid over het nut van een korte route kleiner is dan de onzekerheid over het nut van een lange route. Dus, de mate van variantie van route-nut hangt af van de lengte van de route. Als men hetzelfde routekeuzemodel wil toepassen op alle mogelijke herkomst-bestemmings-paren binnen een netwerk, dan dient de variantie van het route-nut af te hangen van de afstand tussen de herkomst en de bestemming.

In dit proefschrift worden twaalf modellen geanalyseerd op deze en andere eigenschappen. Bovendien zijn zeven van de twaalf modellen nieuw. De beschreven modellen zijn allen gegeneraliseerde multivariate extreme waarde keuzemodellen. Het verschil tussen de modellen wordt gekenmerkt door hoe de foutterm terugkomt in de nutsformulering, en hoe afhankelijkheid tussen fouttermen voor verschillende routes wordt toegevoegd. De foutterm kan als term (additieve modellen), of als factor (multiplicatieve modellen) worden toegevoegd aan de nutsformule. Daarnaast introduceert dit proefschrift multiplicatieve modellen die relatief zijn aan een referentieroute, daarbij komt de foutterm allen terug als factor op niet-overlappende gedeeltes met de referentieroute. Vier verschillende structuren van afhankelijkheden tussen de fouttermen kunnen worden toegevoegd, wat leidt tot vier model types (viz., multinomiaal, pad-grootte, paarsgewijs-gecombineerd, en schakel-genest). Alle modellen zijn kwalitatief en kwantitatief beoordeeld.

De kwalitatieve beoordeling is gebaseerd op gewenste eigenschappen van de stochastische nutsformulering. Een belangrijke – empirisch gevalideerde – eigenschap is het lineaire verband tussen het gemiddelde en de standaard deviatie van reistijd. Dit is tegenstrijdig met de aanname van eerder veelgebruikte routekeuzemodellen die zegt dat er een lineair verband is tussen het gemiddelde en de variantie van de reistijd. Onderzoek naar het gedrag van de modellen onder eenvoudige netwerkaanpassingen laat zien dat alleen de multiplicatieve modellen gebaseerd op referentieroutes het verwachte gedrag kan reproduceren.

De kwantitatieve beoordeling beslaat het schatten en valideren van de modellen op routeverzamelingen binnen een zorgvuldig geconstrueerd netwerk. Dit netwerk heeft overlap tussen routes en er bestaan zowel kleine als grote afstanden. Alle modellen zijn geschat op de eerste routeverzameling en zijn vervolgens gevalideerd op de tweede routeverzameling, en vice versa. De beoordeling laat zien dat van beide multiplicatieve pad-grootte modellen (met en zonder referentieroute) een goede prestatie verwacht kan worden op echte netwerken. Om echter een definitief oordeel te vellen over alle modellen is verdere empirische validatie van alle modellen vereist.

Macroscopische Knoopmodellen

Een belangrijk onderdeel van het QDNL model is het knoopmodel. Een dergelijk model berekent de stromen over bijvoorbeeld samenvoegingen, splitsingen, wegversmallingen, en kruispunten in het wegennetwerk. Knoopmodellen worden ook gebruikt om kiemlocaties in het netwerk te identificeren. Pas in 2011 is een verzameling eisen opgesteld waar knoopmodellen aan dienen te voldoen. In vergelijking met schakelmodellen (i.e., modellen die verkeer propageren over homogene stukken weg) hebben knoopmodellen weinig aandacht gekregen in de wetenschappelijke literatuur. Bovendien is er geen compleet beeld van het onderliggende voertuiggedrag bij knopen voor de bestaande modellen.

Dit proefschrift analyseert en bespreekt vier knoopmodellen – waarvan er twee nieuw zijn – die voldoen aan de recent geformuleerde eisen. Het behandelen van de representatie van verkeer op knopen leidt tot het concept van richtingsvertragingen. Richtingsvertragingen zijn variabelen waarin het knoopmodel geformuleerd kan worden, en die tegelijkertijd een gedragsmatige interpretatie geven van de onderliggende mechanismen van het model.

De verzameling eisen wordt herformuleerd in termen van richtingsvertragingen. Dat leidt tot en familie van knoopmodellen die gerepresenteerd worden door een optimalisatieprobleem meer meerdere doelfuncties. Elke methode die een Pareto-optimale oplossing vindt van dit probleem is lid van de knoopmodel-familie. Twee verschillende verzamelingen van richtingsvertragingen kunnen leiden tot dezelfde verkeersstromen over de knoop. Daarom wordt ook de notie van gereduceerde capaciteit geïntroduceerd, welke laat zien wanneer resultaten gelijkwaardig zijn. Het blijkt dat voor splitsingen van wegen alle vier de modellen gelijkwaardig zijn, en dat voor samenvoegingen van wegen drie van de vier modellen gelijkwaardig zijn.

Één bestaand en één nieuw model kennen efficiënte oplossingsmethodes. Het nieuwe model berekent de richtingsvertragingen op basis van de theorie van polynomen in de max-plus algebra. Anderzijds zijn deze modellen in gedragsmatig opzichten minder realistisch dan de andere twee modellen. Helaas gebruiken de andere twee modellen dekpuntmethodes die veel meer rekentijd vergen. Dit komt er op neer dat wanneer met een knoopmodel kiest, men een afweging dient te maken tussen rekentijd en realisme.

SPELTHEORIE VOOR MEERDERE ACTOREN

In het bovenste niveau van het holistische modelraamwerk bevindt zich de besluitvorming die wordt gerepresenteerd door de interacties tussen meerdere actoren; dit wordt beschreven in Deel (Part) II van dit proefschrift. Een belangrijke eigenschap is dat iedere actor zijn eigen doel nastreeft door een bepaalde strategie toe te passen. Deze strategie bestaat uit het bepalen van de prijs van hun 'reisproduct'. De actoren dienen rekening te houden met de reacties van de reizigers en de resulterende effecten wanneer de strategie bepaald wordt. De reactie wordt behandeld binnen het onderste niveau, het QDTA-model. Vervoerseconomen gebruiken soms sociale welvaartsoptimalisatie en zogeheten 'first-best' beprijzing om het bovenste niveau in te vullen. Twee aannames daarbij zijn dat iedereen deelneemt aan de onderhandelingstafel en dat er onbeperkt transacties tussen individuen plaats kunnen vinden. De multi-actor aanpak in dit proefschrift haalt deze aannames weg door speltheoriën te verkennen die nog niet eerder zijn toegepast op dergelijke vraagstukken.

Het bijbehorende wiskundige optimalisatieprobleem van het bovenste niveau heeft meerdere doelfuncties en kent onderliggende evenwichtsvoorwaarden. Dat betekent dat het niet eenvoudig is oplossingen te bepalen. Dit proefschrift gebruikt speltheorie om oplossingen van het probleem te achterhalen. Met speltheorie heeft men de mogelijkheid het gedrag en de interacties van actoren aan de onderhandelingstafel mee te nemen. Actoren kunnen coalities vormen en actoren binnen coalities kunnen samenwerken om andere actoren en coalities tegen te werken. Dit proefschrift introduceert een methode om middels Nash-evenwichten en de notie van coalitieformering het wiskundige optimalisatieprobleem om te zetten in een spel van overdraagbaar nut (Transferable Utility (TU) in het Engels). In zo'n TU-spel heeft iedere coalitie een bepaalde waarde die meegebracht kan worden naar de onderhandelingstafel.

Speltheorie biedt meerdere oplossingsconcepten voor TU-spellen. De kern, de Shapley waarde en de compromiswaarde worden gepresenteerd als potentiële oplossingen voor het bovenste niveau. Het beprijzingsschema dat de som van de doelen van alle actoren optimaliseert wordt uiteindelijk uitgevoerd bij ieder oplossingsconcept voor TU-spellen. Ieder oplossingsconcept stelt vervolgens geldtransacties tussen actoren voor welke eerlijk zijn voor alle actoren en coalities. Sommige actoren kunnen daarbij gecompenseerd worden voor gemiste inkomsten binnen het algeheel beste en gekozen beprijzingsschema. Ieder concept levert een oplossing die gebaseerd is op andere eerlijkheidsprincipes (i.e., andere interpretaties van wat 'eerlijk' is).

CASESTUDIE: DE RANDSTAD

Om de haalbaarheid van het holistische modelraamwerk voor echte toepassingen aan te tonen wordt een casestudie van de Randstad gepresenteerd. Alle methodologische ontwikkelingen komen samen binnen deze casestudie. In vergelijking met traditionele statische modellen is er realisme aan het QDTA-model in het onderste niveau van het raamwerk toegevoegd zonder dat er veel meer rekentijd nodig is. Meerdere actoren, met hun doelen en hun mogelijke conflicten aan de onderhandelingstafel, kunnen geanalyseerd worden met behulp van speltheorie in het bovenste niveau van het raamwerk. Er zijn drie actoren binnen de casestudie, ieder met hun eigen doel en beprijzingsmaatregel. De nationale overheid streeft ernaar om de sociale welvaart te verbeteren door een kilometerheffing te implmenteren. De gemeente Amsterdam wil haar economische positie verbeteren door een cordonheffing in te voeren. De spoorwegen willen tenslotte hun bedrijfsresultaat maximaliseren door de traintarieven aan te passen. Alle actoren kunnen onderscheid maken met hun prijzen tussen de spits- en de dalperiode. Ondanks dat de gekozen doelfuncties van de actoren hypothetisch zijn, zijn zoveel mogelijk (data-) bronnen uit bestaande strategische planningsmodellen gehaald. Het wegennetwerk bestaat uit bijna 90 000 schakels. Reizigers kunnen kiezen uit auto en openbaar vervoer (trein) opties, tussen de spits- en dalperiode, of ze kunnen besluiten thuis te blijven. Bovendien zijn er met de auto veel verschillende routes mogelijk.

Het QDTA-evenwicht, met een multiplicatief genest keuzemodel en het QDNL-model, is berekend voor 216 mogelijke beprijzingsschema's. Dit model voorspelt onder andere het congestieniveau, emissieniveau's en baten voor iedere actor. Op basis van deze evaluaties is een TU-spel geformuleerd. Vervolgens zijn drie oplossingsconcepten voor TU-spellen geanalyseerd (viz., de kern, de Shapley waarde en de compromiswaarde). Deze leveren mogelijke argumenten die de actoren kunnen gebruiken aan de onderhandelingstafel.

Ondanks dat de casestudie een aantal versimpelende aannames kent en hypothetische doelfuncties van de actoren gebruikt, komen er interessante resultaten uit. De nationale overheid en de spoorwegen hebben conflicterende belangen. Na onderhandelingen dient de nationale overheid de spoorwegen te compenseren. Daarnaast is het zo dat wanneer het niet is toegestaan om samen te werken, de nationale overheid beter af is. Dit klinkt misschien paradoxaal, maar de verklaring moet gezocht worden in de coalitie die gevormd kan worden tussen de gemeente

Amsterdam en de spoorwegen. Deze coalitie zal door het combineren van hun beprijzingsmaatregelen de oplossing wegsturen van de voor de nationale overheid goede situatie.

Het vergelijken van het nul-scenario met de oplossingen met beprijzingsmaatregelen laat de potentie van innovatief prijsbeleid zien. In de ochtendspits kan beprijzing de verliesuren met 60% verminderen. Algeheel gezien is een vermindering van 45% mogelijk. De uitstoot binnen de Randstad kan met 6% verminderd worden. Dit wordt met name bereikt doordat er een vervoerwijzeverschuiving is richting de trein, en er een sterke groei is van het aantal personen dat besluit helemaal niet meer te reizen. De som van de doelfuncties van de actoren (i.e., socale welvaart, de economische positie van Amsterdam en het bedrijfsresultaat van de spoorwegen) stijgt met 1,4 M€ tijdens iedere ochtendspits.

De hoofdconclusie van de casestudie is dat het mogelijk is innovatief prijsbeleid waarbij meerdere actoren betrokken zijn te beoordelen. Dat gaat middels modelonderdelen waarvan het realisme verbeterd is. Om een eindoordeel te verkrijgen over de effecten van prijsbeleid en de interacties tussen actoren zal een meer gedetailleerde studie opgezet moeten worden, met meer vervoerswijzen en gebruikerstypes, en waarbij doelfuncties van de actoren worden gebruikt die verkregen zijn middels interviews.

About the Author

Erik-Sander Smits was born in Waalwijk, the Netherlands on July 26, 1983. He finished the Dutch pre-university education program 'Voorbereidend Wetenschappelijk Onderwijs' from Regionale Scholengemeenschap Broklede, Breukelen in 2002. He decided to start two Bachelor programs at Utrecht University: Mathematics and Computational Sciences. After obtaining both Bachelor of Science degrees, he decided to continue only with Mathematics. For his master thesis about the estimation of origin-destimation matrices, he performed an internship at Omnitrans International/Goudappel Coffeng in Deventer, the Netherlands. In 2010, he obtained his Master of Science degree in Mathematical Sciences from Utrecht University.

Since 2010, he has been conducting doctoral research at the Transport & Planning department of the faculty of Civil Engineering & Geosciences of Delft University of Technology. His research topic is strategic network modelling applied to innovative passenger transport pricing. This research is part of the Innovative Pricing for Sustainable Mobility (iPriSM) project in the Sustainable Accessibility of the Randstad (SAR) programme of the Netherlands Organisation for Scientific Research (NWO). In 2013, he visited the Institute of Transport and Logistics Studies at The University of Sydney to develop some of the methods presented in this dissertation. In 2010, he won the 'Poster of the Year Award' at the TRAIL congress.

At the Transport & Planning department, Erik-Sander has assisted in teaching the courses 'Transport & Spatial Modelling' and 'Advanced Transport Modelling', and he has supervised two master students. He was chairman of the PhD meeting, and he joined the general board of the department as the representative of the PhD students. He also represented the PhD students of Delft University of Technology in the PhD council of research school TRAIL.

Since 2015, he is an advisor at Arane Adviseurs in Gouda, the Netherlands. He works primarily on projects about integrated network management, empirical traffic flow analysis, and pedestrian flows at public transport terminals.

List of Publications

Adams, K., van Kooten, J., and Smits, E.-S. (2016). Gecoördineerd netwerkbreed verkeersmanagement: van tekentafel naar toepassing in de praktijk. In *Proceedings Nationaal Verkeerskunde Congres*.

Bliemer, M., Brederode, L., Wismans, L., and Smits, E.-S. (2012a). Quasi-dynamic network loading: Adding queuing and spillback to static traffic assignment. Technical Report ITLS-WP-12-03. ISSN 1832-570X, Institute of Transport and Logistics Studies, The University of Sydney.

Bliemer, M., Raadsen, M., de Romph, E., and Smits, E.-S. (2013). Requirements for traffic assignment models for strategic transport planning: A critical assessment. In *Proceedings of the 36th Autralasian Transport Research Forum*.

Bliemer, M. C. J., Brederode, L., Wismans, L., and Smits, E.-S. (2012b). Quasi-dynamic traffic assignment: static traffic assignment with queueing and spillback. In sn, editor, *The Transportation Research Board (TRB) 91st Annual Meeting, Washington DC, January 22-26, 2012 (paper no 12-0358).*, pages 1 – 24. TRB.

Bliemer, M. C. J., Raadsen, M. P., Smits, E.-S., Zhou, B., and Bell, M. G. (2014). Quasi-dynamic traffic assignment with residual point queues incorporating a first order node model. *Transportation Research Part B: Methodological*, 68(0):363 – 384.

Schreiter, T., Smits, E.-S., Lint, J. v., and Hoogendoorn, S. (2010). The cell transmission model with capacity drop. In *Proceedings 11th TRAIL Congress*, pages 1–4, Rotterdam.

Smits, E.-S. (2010). Innovative road pricing. Poster for Sustainable Accessibility Randstad conference.

Smits, E.-S. (2012a). Modeling transport pricing with multiple stakeholders - working paper: Methodology and a case study. In *Proceedings 12th TRAIL Congress*, pages 1–8, Rotterdam.

Smits, E.-S. (2012b). Prijsbeleid modellering met meerdere actoren. Presented at Platos Colloquium.

Smits, E.-S., Bliemer, M., Pel, A., and van Arem, B. (2014). On route choice models with closed-form probability expressions. In sn, editor, *The Transportation Research Board (TRB) 93rd Annual Meeting, Washington DC, January 12-16, 2014 (paper no 14-3733).*, pages 1 – 24. TRB.

Smits, E.-S. and Bliemer, M. C. J. (2010). Innovative transport pricing: deriving the optimal pricing scheme: A multi-objective multi-modal dynamic network perspective. In *Proceedings 11th TRAIL Congress*, pages 1–4, Rotterdam.

Smits, E.-S., Bliemer, M. C. J., Pel, A., and van Arem, B. (2013). Family of macroscopic node models. In *Proceedings of hEART2013 - 2nd Symposium of the European Association for Research in Transportation*, Stockholm.

Smits, E.-S., Bliemer, M. C. J., Pel, A. J., and van Arem, B. (2015). A family of macroscopic node models. *Transportation Research Part B: Methodological*, 74:20–39.

Smits, E.-S., Bliemer, M. C. J., and van Arem, B. (2011). Dynamic network loading of multiple user-classes with the link transmission model. In *Proceedings MT-ITS 2011*, pages 1–4, Leuven.

Smits, E.-S., Bliemer, M. C. J., Verhoef, E., and van Arem, B. (2012a). Modelling stakeholders behaviour in transport pricing. In *Proceedings 13th International Conference of the International Association for Travel Behavior Research*, pages 1–4, Toronto.

Smits, E.-S., Pel, A., Bliemer, M., and van Arem, B. (2017). Generalized multivariate extreme value models for explicit route choice sets. Under review at Transportation Research Part B: Methodological.

Smits, E.-S., Pel, A., and van Arem, B. (2012b). A multi-class macroscopic intersection model. In *Proceedings of the 1st European Symposium on Quantitative Methods in Transportation Systems*, pages 1–7, Lausanne.

Smits, E.-S., Pel, A., van Nes, R., and van Arem, B. (2016a). Modelleren mobiliteitsbeprijzing met meerdere actoren. *Tijdschrift Vervoerwetenschap*. ISSN 1571-9227.

Smits, E.-S., van Kooten, J., and Hoogendoorn, S. (2016b). Towards efficient mitigation of freeway congestion by integrated network management. In *Proceedings of hEART 2016 - 5th Symposium of the European Association for Research in Transportation*, Delft.

TRAIL Thesis Series

The following list contains the most recent dissertations in the TRAIL Thesis Series. For a complete overview of more than 150 titles see the TRAIL website: www.rsTRAIL.nl.

The TRAIL Thesis Series is a series of the Netherlands TRAIL Research School on transport, infrastructure and logistics.

Smits, E.-S., *Strategic Network Modelling for Passenger Transport Pricing*, T2017/3, May 2017, TRAIL Thesis Series, the Netherlands

Tasseron, G., *Bottom-Up Information Provision in Urban Parking: An in-depth analysis of impacts on parking dynamics*, T2017/2, March 2017, TRAIL Thesis Series, the Netherlands

Halim, R.A., *Strategic Modeling of Global Container Transport Networks: Exploring the future of port-hinterland and maritime container transport networks*, T2017/1, March 2017, TRAIL Thesis Series, the Netherlands

Olde Keizer, M.C.A., *Condition-Based Maintenance for Complex Systems: Coordinating maintenance and logistics planning for the process industries*, T2016/26, December 2016, TRAIL Thesis Series, the Netherlands

Zheng, H., *Coordination of Waterborn AGVs*, T2016/25, December 2016, TRAIL Thesis Series, the Netherlands

Yuan, K., *Capacity Drop on Freeways: Traffic dynamics, theory and Modeling*, T2016/24, December 2016, TRAIL Thesis Series, the Netherlands

Li, S., *Coordinated Planning of Inland Vessels for Large Seaports*, T2016/23, December 2016, TRAIL Thesis Series, the Netherlands

Berg, M. van den, *The Influence of Herding on Departure Choice in Case of Evacuation: Design and analysis of a serious gaming experimental set-up*, T2016/22, December 2016, TRAIL Thesis Series, the Netherlands

Luo, R., *Multi-Agent Control of urban Transportation Networks and of Hybrid Systems with Limited Information Sharing*, T2016/21, November 2016, TRAIL Thesis Series, the Netherlands

Campanella, M., *Microscopic Modelling of Walking Behavior*, T2016/20, November 2016, TRAIL Thesis Series, the Netherlands

Horst, M. van der, *Coordination in Hinterland Chains: An institutional analysis of port-related transport*, T2016/19, November 2016, TRAIL Thesis Series, the Netherlands

Beukenkamp, W., *Securing Safety: Resilience time as a hidden critical factor*, T2016/18, Oc-

tober 2016, TRAIL Thesis Series, the Netherlands

Mingardo, G., *Articles on Parking Policy*, T2016/17, October 2016, TRAIL Thesis Series, the Netherlands

Duives, D.C., *Analysis and Modelling of Pedestrian Movement Dynamics at Large-scale Events*, T2016/16, October 2016, TRAIL Thesis Series, the Netherlands

Wan Ahmad, W.N.K., *Contextual Factors of Sustainable Supply Chain Management Practices in the Oil and Gas Industry*, T2016/15, September 2016, TRAIL Thesis Series, the Netherlands

Liu, X., *Prediction of Belt Conveyor Idler Performance*, T2016/14, September 2016, TRAIL Thesis Series, the Netherlands

Gaast, J.P. van der, *Stochastic Models for Order Picking* Systems, T2016/13, September 2016, TRAIL Thesis Series, the Netherlands

Wagenaar, J.C., *Practice Oriented Algorithmic Disruption Management in Passenger Railways*, T2016/12, September 2016, TRAIL Thesis Series, the Netherlands

Psarra, I., *A Bounded Rationality Model of Short and Long-Term Dynamics of Activity-Travel Behavior*, T2016/11, June 2016, TRAIL Thesis Series, the Netherlands

Ma, Y., *The Use of Advanced Transportation Monitoring Data for Official Statistics*, T2016/10, June 2016, TRAIL Thesis Series, the Netherlands

Li, L., *Coordinated Model Predictive Control of Synchromodal Freight Transport Systems*, T2016/9, June 2016, TRAIL Thesis Series, the Netherlands

Vonk Noordegraaf, D.M., *Road Pricing Policy Implementation*, T2016/8, June 2016, TRAIL Thesis Series, the Netherlands

Liu, S., Modeling, *Robust and Distributed Model Predictive Control for Freeway Networks*, T2016/7, May 2016, TRAIL Thesis Series, the Netherlands

Calvert, S.C., *Stochastic Macroscopic Analysis and Modelling for Traffic Management*, T2016/6, May 2016, TRAIL Thesis Series, the Netherlands

Sparing, D., *Reliable Timetable Design for Railways and Connecting Public Transport Services*, T2016/5, May 2016, TRAIL Thesis Series, the Netherlands

Rasouli, S, *Uncertainty in Modeling Activity-Travel Demand in Complex Urban Systems*, T2016/4, March 2016, TRAIL Thesis Series, the Netherlands

Vries, J. de, *Behavioral Operations in Logistics*, T2016/3, February 2016, TRAIL Thesis Series, the Netherlands

Goñi-Ros, B., *Traffic Flow at Sags: Theory, Modeling and Control*, T2016/2, March 2016, TRAIL Thesis Series, the Netherlands