Combining SP probabilities and RP discrete choice in departure time modelling: joint MNL and ML estimations

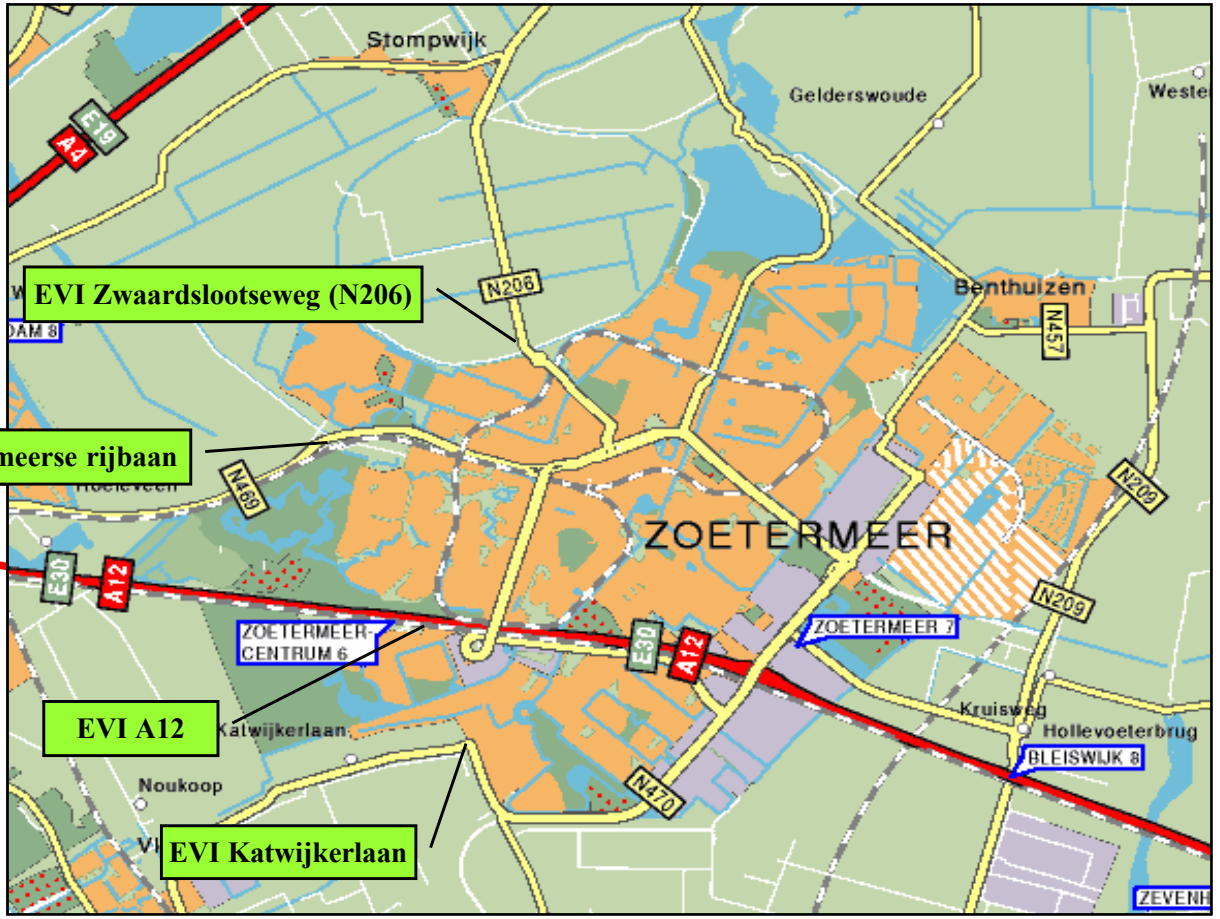
Jasper Knockaert
Yinyen Tseng
Erik Verhoef
Jan Rouwendal

Introduction

- Central question: How do RP and SP compare?
- Context of departure time and mode choice model:
  - Experiment with rewards
  - Micro data (~panel)
  - Discrete choice theory
  - Joint RP-SP estimation

Spitsmijden: Where & When?

THE HAGUE



Reference: 2 weeks

Reward: 10 weeks

No Reward: 1 week

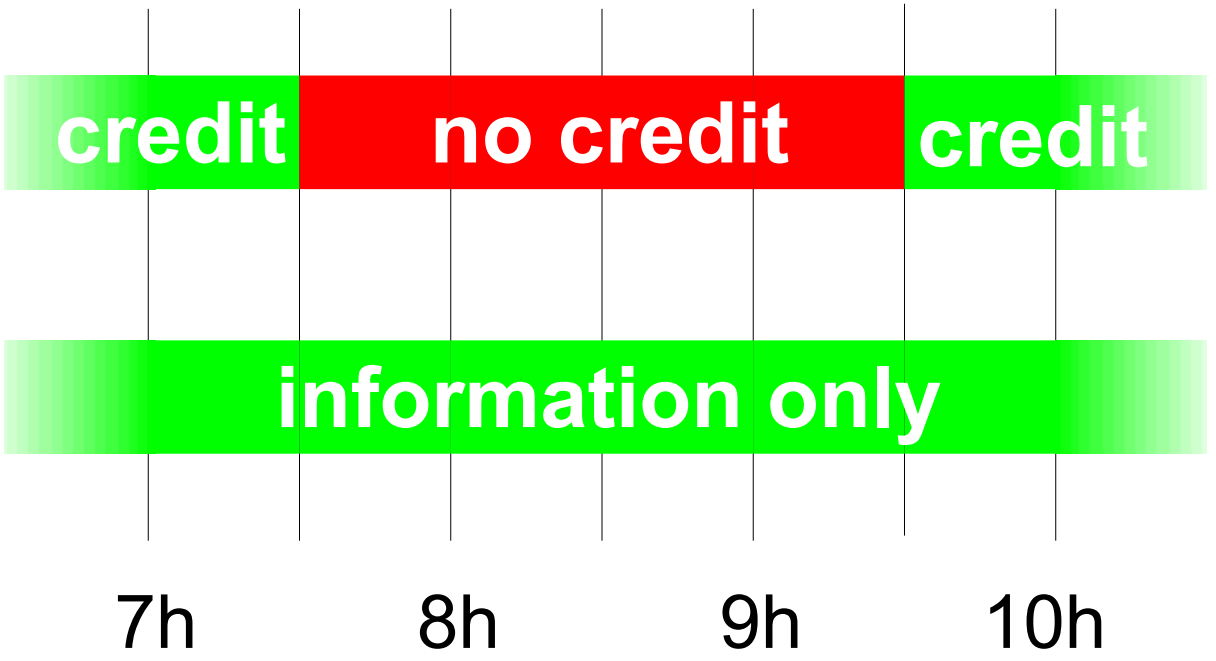
Sep 18 Okt 2 2006 Dec 4 Dec 11

Spitsmijden: What?

- Money: 3 types of reward, 232 participants

Spitsmijden: What?

- Yeti smartphone (108 participants):
  - Traffic information
  - Saving credits to keep phone at end of experiment



Spitsmijden: What?

- Reward adapted on basis of initial behaviour
  - Less frequent drivers are rewarded less frequent
- Rewards' regimes’ varied over participants over the duration of the experiment
  - Exclude the bias caused by learning impact
  - … or effects of change of season (Bikes,….)

Discrete choice theory

- Each choice alternative *j* has for each choice *m* by respondent *n* a random utility U_{jmn}:
  $U_{jmn} = \beta x_{jmn} + \varepsilon_{jmn} (+ \mu \eta_n)$
- We use linear utility specification:
  - generic explanatory variables: travel time, scheduling, reward
  - alternative specific coefficients
  - standard normal distributed variable η_n for inter personal choice heterogeneity (panel mixed logit)

Table 0.1: Definition of choice variables use in model estimates (choice situation m corresponds to a working day in RP or a choice set in SP)

attribute	unit	definition
t_{jm}	hour	car travel time corresponding to time interval j in choice situation m
m_{jmn}	euro	marginal loss of monetary reward for participant n of rush hour car travel in time interval j in choice situation m
y_{jmn}	credit	marginal loss of Yeti credit for participant n of rush hour car travel in time interval j in choice situation m
Y_{mn}	credit	number of credits needed by participant n to collect Yeti in choice situation m (varies across choice sets in SP but fixed across choice situations in RP)
p_n^*	hour	unrewarded passage time for participant n (based on RP observations in periods before/after experiment)
p_j	hour	actual passage time of rush hour car travel in time interval j
w_m	°C	maximum temperature observed on day m
C_{mode}		mode specific constant
η_n		panel specified error term: $\eta_n \sim \mathcal{N}(0,1)$

Table 0.5: Deterministic utility V_{jmn}^{RP} of alternative j faced on day m by participant n

alternative j	deterministic utility V_{jmn}^{RP}	mixed logit
private car at time p_j	$\beta_{euro} m_{jmn} + \beta_{yeti} y_{jmn} + \beta_{tt} t_{jm}$ $+\beta_{sde} \max(p_n^* - p_j, 0) + \beta_{sdl} \max(p_j - p_n^*, 0)$	$+\mu\eta_n$
bike	$\beta_{weather} w_m + C_{bike}$	
other mode	C_{mode}	

Table 0.6: Panel mixed logit model (RP dataset)

		robust statistics		
decision variable	coeff.	std err	t-stat	p-value
β_{euro} [euro] (monetary reward)	−0,350	0,0198	−17,70	0,00
β_{sde} [hour] (early arrival)	−1,95	0,126	−15,51	0,00
β_{sdl} [hour] (late arrival)	−2,24	0,0963	−23,23	0,00
β_{yeti} (smartphone reward)	−2,18	0,168	−12,96	0,00
β_t [hour] (motorway car travel time)	−1,54	0,372	−4,15	0,00
$\beta_{weather}$ [max temp in °C] (bike)	0,0999	0,0243	4,11	0,00
C_{py} (constant for public transport)	−2,24	0,173	−12,97	0,00
C_{bike} (constant for bike)	−4,98	0,522	−9,55	0,00
$C_{offpeak}$ (constant for car not 7h–10h)	−1,51	0,142	−10,66	0,00
$C_{rideshare}$ (constant for carpool)	−3,78	0,231	−16,39	0,00
C_{home} (constant for telework)	−3,20	0,173	−18,50	0,00
C_{other} (constant for other)	−4,01	0,358	−11,22	0,00
μ (inter personal heterogeneity)	1,69	0,0796	21,27	0,00
estimation statistics				
observations		12371		
individuals		322		
Halton draws		4000		
log-likelihood		−27043,676		

SP survey

- Completed by participants to reward experiment (before trial started)
- 6 choice sets randomly selected out of a 17 set design
- Choice alternatives:
  - 3 or 5 departure time intervals for private car
  - 4 other non-car alternatives (including no-travel)

Table 1 Example of the choice set presentation for monetary reward participants

Choice variant					
The reward you would receive if you travel by car in the following time slot:					
before 7.00 am	you receive	**7,00 €**	(**5** minute delay in the traffic jam)		
between 7.00 and 7.30 am	you receive	**3,50 €**	(**20** minute delay in the traffic jam)		
between 7.30 and 8.30 am	you receive	**no reward**	(**35** minute delay in the traffic jam)		
between 8.30 and 9.00 am	you receive	**3,50 €**	(**20** minute delay in the traffic jam)		
after 9.00 am	you receive	**7,00 €**	(**5** minute delay in the traffic jam)		
The travel time of **public transport** is **10** minutes longer than the travel time in the central peak.					
If you travel by public transport, bike, or work from home, you would receive the reward of **7,00 €**					
Suppose the experiment lasts for **50** days. How often would you travel in the following possibilities:					
By car		By public transport	By bike	Work from home	Other
before 7.00 am	: .... day(s)				
between 7.00 and 7.30 am	: .... day(s)				
between 7.30 and 8.30 am	: .... day(s)	.... day(s)	.... day(s)	.... day(s)	.... day(s)
between 8.30 and 9.00 am	: .... day(s)				
after 9.00 am	: .... day(s)				

Table 2 Example of the choice set presentation for yeti credit participants

Choice variant					
The reward you would receive if you travel by car in the following time slot:					
before 7.00 am	you receive	**1 credit**	(**5** minute delay in the traffic jam)		
between 7.00 and 7.30 am	you receive	**0.5 credit**	(**20** minute delay in the traffic jam)		
between 7.30 and 8.30 am	you receive	**no credit**	(**35** minute delay in the traffic jam)		
between 8.30 and 9.00 am	you receive	**0.5 credit**	(**20** minute delay in the traffic jam)		
after 9.00 am	you receive	**1 credit**	(**5** minute delay in the traffic jam)		
The travel time of **public transport** is **10** minutes longer than the travel time in the central peak.					
If you travel by public transport, bike, or work from home, you would receive **1 credit**.					
You would earn the Yeti if you could save up to **15 credits.** Suppose the experiment lasts for **25** days. How often would you travel in the following possibilities:					
By car		By public transport	By bike	Work from home	Other
before 7.00 am	: .... day(s)				
between 7.00 and 7.30 am	: .... day(s)				
between 7.30 and 8.30 am	: .... day(s)	.... day(s)	.... day(s)	.... day(s)	.... day(s)
between 8.30 and 9.00 am	: .... day(s)				
after 9.00 am	: .... day(s)				

Ratio scale

- Classic setup in estimation:
  $P_{mn} = \sum_j y_{jmn} P_{jmn} = \prod_j P_{jmn}^{y_{jmn}}$
  $LL_{mn} = \ln P_{mn}$
- Ratio scale format departs from classic setup
- Obviously, not 25 (or 50) independent choices
- Formulate as one choice to get correct results
- Python Biogeme log likelihood specification:
  $P_{mn} = \prod_j P_{jmn}^{p_{jmn}}$
  $LL_{mn} = \ln P_{mn}$ $(= \sum_j p_{jmn} \ln P_{jmn})$

Table 0.2: Deterministic utility V_{jmn}^{SP} of alternative j faced in choice set m by participant n

alternative j	deterministic utility V_{jmn}^{SP}	mixed logit
private car at time p_j	$\beta_{euro} m_{jmn} + \beta_{yeti} y_{jmn} / Y_{mn} + \beta_{tt} t_{jm}$ $+\beta_{sde} \max(p_n^* - p_j, 0) + \beta_{sdl} \max(p_j - p_n^*, 0)$	$+\mu\eta_n$
public transport	$C_{pt} + \beta_{pt} t_{jm}$	
other mode	C_{mode}	

Table 0.3: Multinomial logit model (SP dataset)

		robust statistics		
model coefficient	value	std err	t-stat	p-value
β_{euro} [euro] (monetary reward)	−0,104	0,0219	−4,77	0,00
β_{sde} [hour] (early arrival)	−1,25	0,0535	−23,30	0,00
β_{sdl} [hour] (late arrival)	−1,37	0,0636	−21,61	0,00
β_{yeti} (smartphone reward)	−19,7	2,07	−9,54	0,00
β_{pt} [hour] (excess pt travel time)	−0,0507	0,346	−0,15	0,88
β_t [hour] (car congestion delay)	−0,932	0,313	−2,98	0,00
C_{pt} (constant for public transport)	−1,56	0,258	−6,05	0,00
C_{bike} (constant for bike)	−2,57	0,105	−24,58	0,00
C_{home} (constant for telework)	−2,82	0,0977	−28,87	0,00
C_{other} (constant for other modes)	−2,98	0,114	−26,21	0,00
estimation statistics				
observations		1530		
log-likelihood		−2681,507		

Table 0.4: Panel mixed logit model (SP dataset)

		robust statistics		
model coefficient	value	std err	t-stat	p-value
β_{euro} [euro] (monetary reward)	−0,115	0,0246	−4,67	0,00
β_{sde} [hour] (early arrival)	−1,25	0,106	−11,73	0,00
β_{sdl} [hour] (late arrival)	−1,37	0,121	−11,37	0,00
β_{yeti} (smartphone reward)	−20,6	2,83	−7,27	0,00
β_{pt} [hour] (excess pt travel time)	−0,0653	0,339	−0,19	0,85
β_t [hour] (car congestion delay)	−0,745	0,286	−2,61	0,01
C_{pt} (constant for public transport)	−1,79	0,301	−5,96	0,00
C_{bike} (constant for bike)	−2,81	0,213	−13,20	0,00
C_{home} (constant for telework)	−3,06	0,190	−16,13	0,00
C_{other} (constant for other modes)	−3,22	0,233	−13,84	0,00
μ (inter personal heterogeneity)	−1,48	0,104	−14,21	0,00
estimation statistics				
observations		1530		
individuals		286		
Halton draws		1000		
log-likelihood		−2591,148		

Joint RP-SP estimation

- Common approach: scale the utility U_{jmn} of one of both:
  $U_{joint} = s_{RP}\ U_{RP} + s_{SP}\ U_{SP}$
- Panel: consistency across all observations for same respondent *n*

Table 0.7: Multinomial logit model (joint RP-SP dataset) 1/2

		robust statistics		
model coefficient	value	std err	t-stat	p-value
joint coefficients				
β_{euro} [euro] (monetary reward)	−0,324	0,00680	−47,65	0,00
revealed preference coefficients				
β_{sde} [hour] (early arrival)	−1,85	0,0339	−54,50	0,00
β_{sdl} [hour] (late arrival)	−2,20	0,0374	−58,82	0,00
β_{yeti} (smartphone reward)	−2,29	0,0743	−30,78	0,00
β_{t} [hour] (motorway car travel time)	−1,67	0,237	−7,05	0,00
$\beta_{weather}$ [max temp in °C] (bike)	0,115	0,0170	6,78	0,00
C_{pt} (constant for public transport)	−1,88	0,0433	−43,32	0,00
C_{bike} (constant for bike)	−4,85	0,288	−16,84	0,00
$C_{offpeak}$ (constant for car not 7h–10h)	−1,15	0,0389	−29,47	0,00
$C_{rideshare}$ (constant for carpool)	−3,42	0,0758	−45,07	0,00
C_{home} (constant for telework)	−2,84	0,0606	−46,79	0,00
C_{other} (constant for other modes)	−3,65	0,0835	−43,68	0,00

Table 0.8: Multinomial logit model (joint RP-SP dataset) 2/2

		robust statistics		
model coefficient	value	std err	t-stat	p-value
stated preference coefficients				
β_{sde} [hour] (early arrival)	−3,95	0,794	−4,98	0,00
β_{sdl} [hour] (late arrival)	−4,37	0,875	−4,99	0,00
β_{yeti} (smartphone reward)	−58,3	8,70	−6,70	0,00
β_{pt} [hour] (excess pt travel time)	−0,207	1,11	−0,19	0,85
β_t [hour] (car congestion delay)	−3,07	1,50	−2,05	0,04
C_{pt} (constant for public transport)	−5,00	1,30	−3,85	0,00
C_{bike} (constant for bike)	−8,27	1,76	−4,71	0,00
C_{home} (constant for telework)	−9,06	1,91	−4,76	1,00
C_{other} (constant for other modes)	−9,56	2,01	−4,76	0,00
scale parameters				
s^{RP}	1	fixed		
s^{SP}	0,312	0,0640	4,88	0,00
estimation statistics				
SP observations		1530		
RP observations		12371		
log-likelihood		−31335,698		

Table 0.9: Panel mixed logit model (joint RP-SP dataset) 1/2

		robust statistics		
model coefficient	value	std err	t-stat	p-value
joint coefficients				
β_{euro} [euro] (monetary reward)	−0,350	0,0198	−17,68	0,00
revealed preference coefficients				
β_{sde} [hour] (early arrival)	−1,95	0,125	−15,55	0,00
β_{sdl} [hour] (late arrival)	−2,23	0,0962	−23,22	0,00
β_{yeti} (smartphone reward)	−2,18	0,168	−13,00	0,00
β_{t} [hour] (motorway car travel time)	−1,54	0,371	−4,14	0,00
$\beta_{weather}$ [max temp in °C] (bike)	0,0999	0,0243	4,12	0,00
C_{pt} (constant for public transport)	−2,23	0,173	−12,87	0,00
C_{bike} (constant for bike)	−4,97	0,522	−9,52	0,00
$C_{offpeak}$ (constant for car not 7h–10h)	−1,49	0,141	−10,61	0,00
$C_{rideshare}$ (constant for carpool)	−3,77	0,232	−16,25	0,00
C_{home} (constant for telework)	−3,19	0,174	−18,36	0,00
C_{other} (constant for other modes)	−4,00	0,359	−11,14	0,00
μ (inter personal heterogeneity)	1,70	0,0784	21,69	0,00

Table 0.10: Panel mixed logit model (joint RP-SP dataset) 2/2

		robust statistics		
model coefficient	value	std err	t-stat	p-value
stated preference coefficients				
β_{sde} [hour] (early arrival)	−3,84	0,678	−5,66	0,00
β_{sdl} [hour] (late arrival)	−4,23	0,746	−5,67	0,00
β_{yeti} (smartphone reward)	−49,4	8,18	−6,04	0,00
β_{pt} [hour] (excess pt travel time)	−0,540	1,09	−0,50	0,62
β_t [hour] (car congestion delay)	−3,36	1,23	−2,73	0,01
C_{pt} (constant for public transport)	−4,90	1,16	−4,21	0,00
C_{bike} (constant for bike)	−8,18	1,54	−5,31	0,00
C_{home} (constant for telework)	−8,92	1,61	−5,54	0,00
C_{other} (constant for other modes)	−9,38	1,74	−5,38	0,00
μ (inter personal heterogeneity)	1,75	0,447	3,92	0,00
scale coefficients				
s^{RP}	1	fixed		
s^{SP}	0,336	0,0633	5,31	0,00
estimation statistics				
SP observations		1530		
RP observations		12371		
individuals		322		
Halton draws		250		
log-likelihood		−29681,480		

Conclusions

- Ratio scale allows for consistent discrete choice model estimation
- SP-RP on microdata scheduling model:
  - Time and scheduling valuation about factor 2 higher in SP than RP
  - Consistent panel specification makes sense
- Paper will appear at http://www.tinbergen.nl