Mode Availability and Partial Profiles: Insights into Nesting Structures in Stated and Revealed Choice Data

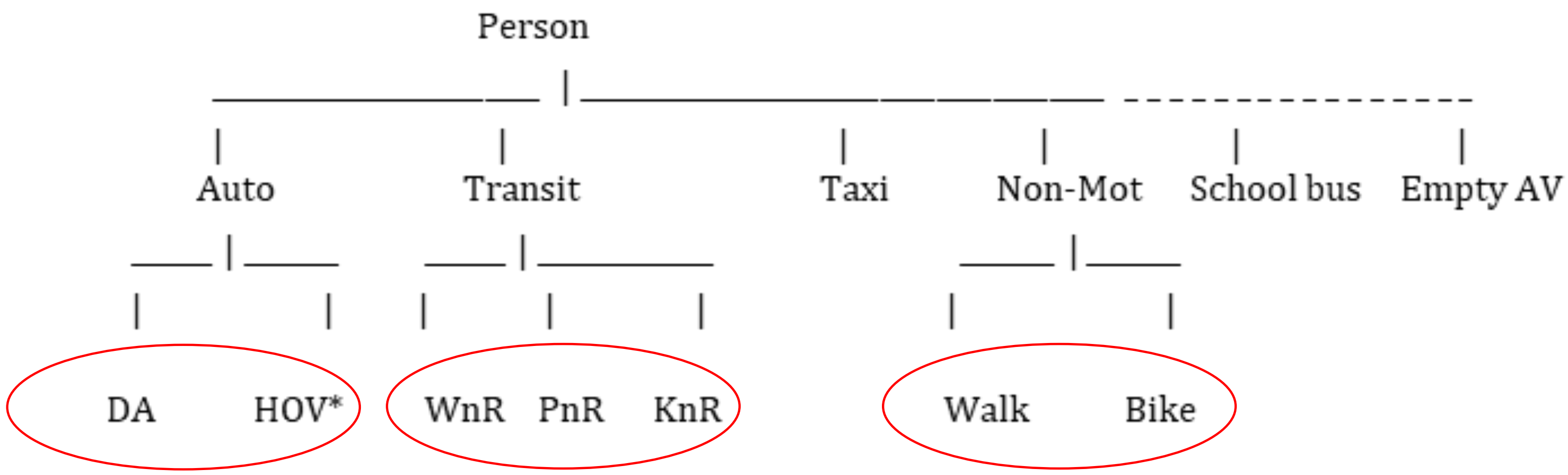
Research Team:
Brett Smith (UWA)
Thomas Hancock (Leeds)
John Buckell (Oxford)

PROJECT OVERVIEW AND PROBLEM

- Department of Transport of WA to develop and implement a tour-based strategic transport model.
- The project aim is to estimate a tour-based mode choice model on the revealed and stated preference data collected by PATHS.
- The alternative model from “donor” based on estimated mode choice models from the U.S.

(DONOR MODEL)

Figure 27a. Model error structure from the Donor Model



TOUR COMPLEXITY AND TRAVEL DISTANCE

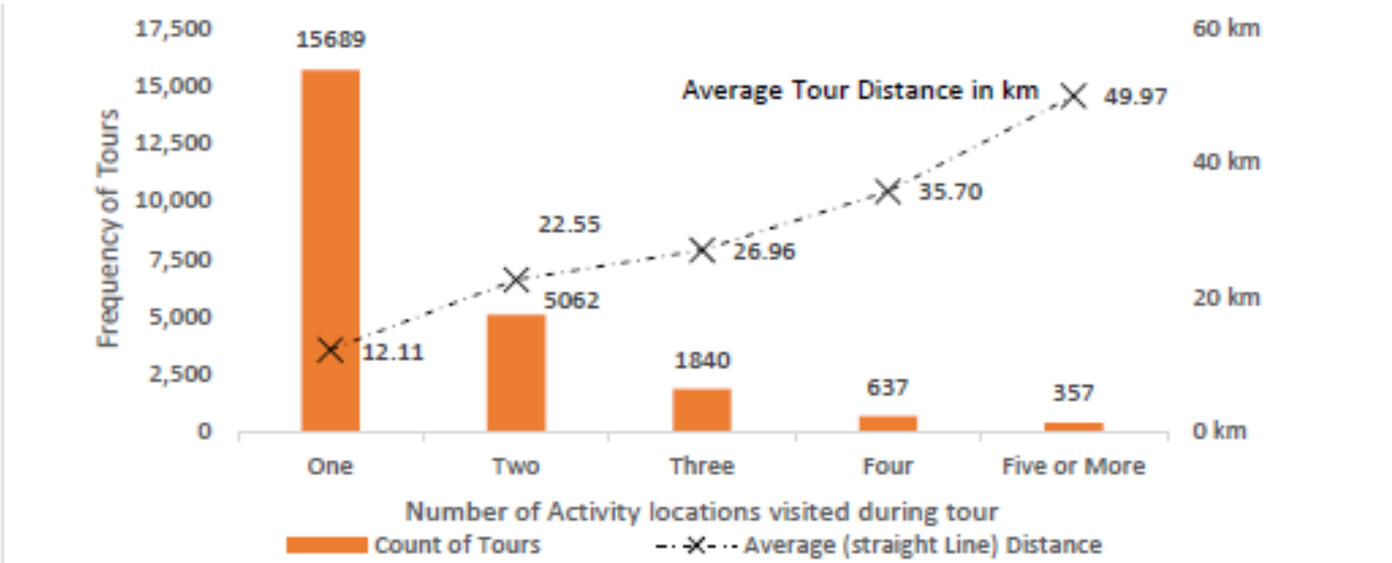


Figure 6. Frequency of Tours and Average Tour Distance

PRIVATE VEHICLE 90%

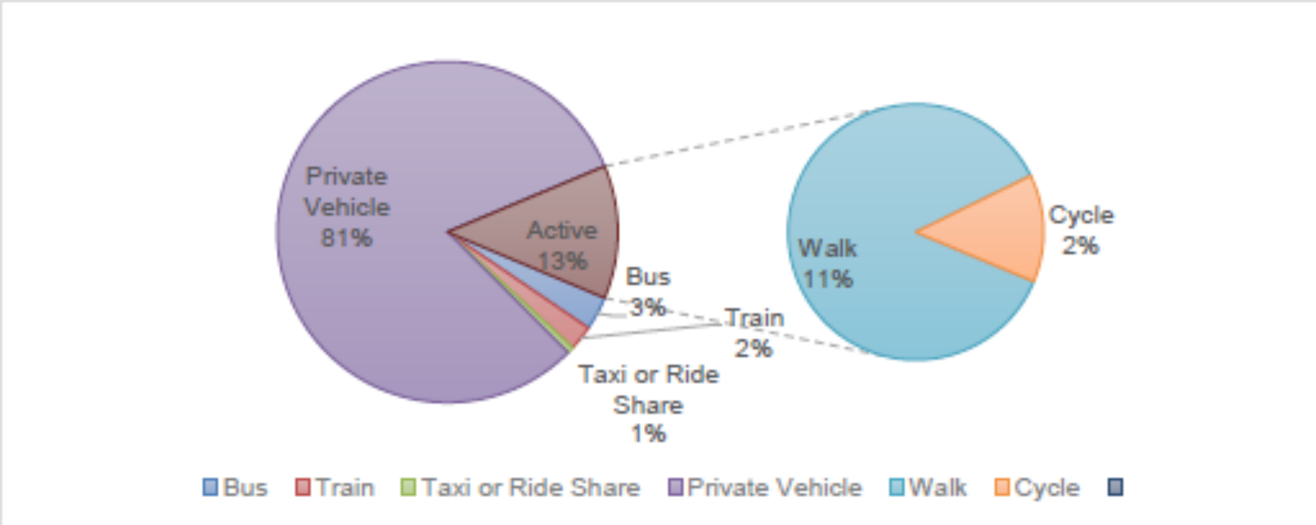
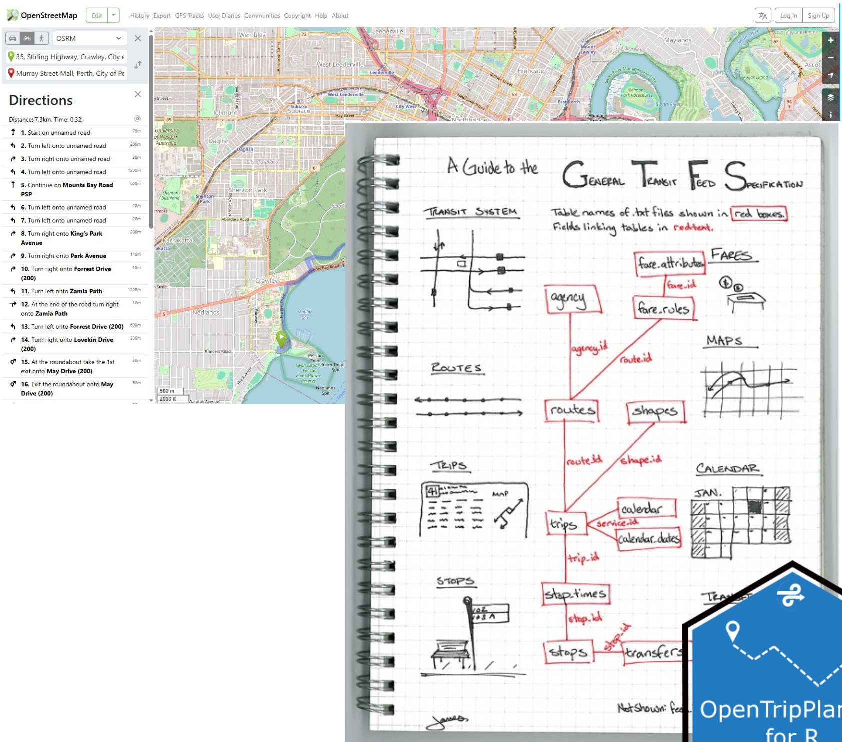


Figure 9. Distribution of Tours by Travel Modes

NETWORK SKIMS



	A	B	C	D	E	Ti
A	0	0	50	0	0	50
B	0	0	60	0	30	90
C	0	0	0	30	0	30
D	20	0	80	0	20	120
E	0	0	90	10	0	100
Tj	20	0	280	40	50	**390**

PROBLEM

- A mode choice model for a strategic transport model
- Eight (+) elemental alternatives
- Unknown error structure
- Not all alternatives feasible for all trips

FLIGHT ITINERARY CHOICE SET

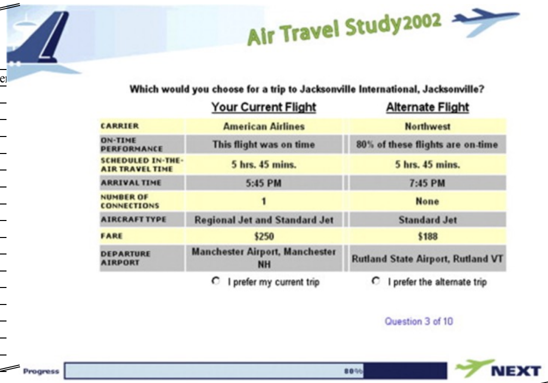
No.1

Airlines	Flight	Departure	Full Flex	Semi Flex	Non Flex
Major	1	8:00	￥600	Sold Out	Sold Out
Low Cost	2	8:00	-	-	￥350
Major	3	12:00	￥550	Sold Out	Sold Out
Major	4	18:00	￥550	Sold Out	Sold Out
Major	5	20:00	￥600	Sold Out	Sold Out
Low Cost	6	20:00	-	-	￥350

Not Travel ❑

PARTIAL PROFILES

	Alter															
Choice Set	1	2	3	4	5	6	7	8								
1	P	P	P	P	A	A	A	A								
2	A	A	A	A	P	P	P	P								
3	A	A	A	A	A	A	A	A								
4	A	A	A	A	A	A	A	A								
5	P	A	A	A	P	A	A	A								
6	A	P	A	A	A	P	A	A								
7	A	A	P	A	A	A	P	A								
8	A	A	A	P	A	A	A	P								
9	P	A	A	A	A	P	A	A								
10	A	P	A	A	P	A	A	A								
11	A	A	P	A	A	A	A	P								
12	A	A	A	P	A	A	P	A								
13	P	A	A	A	A	A	A	P								
14	A	A	A	P	P	A	A	A								
15	A	P	A	A	A	A	P	A								
16	A	A	P	A	A	P	A	A								
17	P	A	A	A	A	A	P	A								
18	A	A	P	A	P	A	A	A	A	P	A	A	A	A	A	P
19	A	A	A	P	A	P	A	A	P	A	A	A	A	A	P	A
20	A	P	A	A	A	A	A	P	A	A	P	A	P	A	A	A



Qantas

Fare

Departure time

…

Qatar

Fare

Departure time

- Lazari and Anderson (1994) Orthogonal X2 and cross effects
- Wang et al. (2013) Orthogonal / Efficient and cross effects
- Rose at al (2013) Efficient / Efficient

ESTIMATION ON A PARTIAL PROFILE (SIMULATED DATA)

(a) Non-Business Segment

64-Scenario Efficient Master and Sub Design						
	Coefficient			T-Test		
MFF	-0.05			-0.16		
MSF	1.14***			5.38		
NFF	-0.25			-0.77		
NSF	1.24***			5.37		
MNF	1.86***			11.76		
NNF	1.80***			11.60		
Price	-0.003***			-10.86		
$\chi^2(30)$	247.2 (43.7)					
Cross-Effects	With the Presence of					
Choice of	MFF	MSF	NFF	NSF	MNF	NNF
MFF	--	-0.34***			-0.58***	
MSF		--			-0.39***	0.11**
NFF			--	-0.33***		-0.73***
NSF				--	0.13**	-0.30***
MNF		-0.07*			--	0.12***
NNF		0.07*		-0.07*	0.13***	--

64-Scenario Orthogonal Master and Efficient Sub Design						
	Coefficient			T-Test		
MFF	0.20			0.83		
MSF	1.28***			7.34		
NFF	0.25			1.05		
NSF	1.26***			7.26		
MNF	1.79***			14.75		
NNF	1.77***			14.53		
Price	-0.003***			-12.94		
$\chi^2(30)$	243.7 (43.7)					
Cross-Effects	With the Presence of					
Choice of	MFF	MSF	NFF	NSF	MNF	NNF
MFF	--	-0.29***			-0.40***	0.12*
MSF		--			-0.30***	
NFF			--	-0.17**		-0.38***
NSF		0.08**		--		-0.30***
MNF					--	0.07**
NNF					0.05*	--

Note: ***, **, * significant at 1%, 5% and 10% level.

The Squiggles

- Regularity implies availability parameters are less-than or equal to zero
- To preserve IIA (MNL) availability parameters are equal to zero
- Correlation implies availability parameters are less-than zero
- Under limited settings the scale parameter (correlation) provides insight into the relative magnitude
- Estimation is murky

2 Alternatives with and without j (MNL/IIA implies $\gamma = 0$)

Consider three alternatives i, j, k, with k always present. We compare two choice sets:

$$C_0 = \{i, k\} \quad \text{(alternative } j \text{ absent)}, \qquad C_1 = \{i, j, k\} \quad \text{(alternative } j \text{ present)}.$$

True MNL utilities (no presence term). Let the true utilities be

$$U_r = V_r + \varepsilon_r, \qquad r \in \{i, j, k\},$$

where ε_r are i.i.d. Type I Extreme Value (Gumbel). Under MNL, the IIA property implies that for any pair of alternatives appearing in both choice sets (here i and k),

$$\frac{P_i(C_0)}{P_k(C_0)} = \frac{P_i(C_1)}{P_k(C_1)}. \qquad (1)$$

That is, adding j cannot change the odds of i versus k.

Add a "presence" effect for i. Now suppose we estimate a model in which the utility of i includes a presence dummy that switches on when j is in the set:

$$\tilde{U}_i = V_i + \gamma d_i + \varepsilon_i, \qquad \tilde{U}_k = V_k + \varepsilon_k,$$

where

$$d_i = \begin{cases} 0, & \text{in } C_0 = \{i, k\} \text{ (since } j \text{ is absent)}, \\ 1, & \text{in } C_1 = \{i, j, k\} \text{ (since } j \text{ is present)}. \end{cases}$$

We do not need to specify $\tilde{U}_j$ for the IIA argument since we only use the pair (i, k).

Odds ratios under the presence specification. Under this specification, the logit odds of i versus k in each set are:

$$\frac{\tilde{P}_i(C_0)}{\tilde{P}_k(C_0)} = \frac{\exp(V_i + \gamma \cdot 0)}{\exp(V_k)} = \exp(V_i - V_k),$$

and

$$\frac{\tilde{P}_i(C_1)}{\tilde{P}_k(C_1)} = \frac{\exp(V_i + \gamma \cdot 1)}{\exp(V_k)} = \exp(V_i - V_k + \gamma) = \exp(\gamma)\exp(V_i - V_k).$$

IIA implies $\gamma = 0$. If IIA holds, the two odds ratios must be equal:

$$\exp(V_i - V_k) = \exp(\gamma)\exp(V_i - V_k).$$

CASE STUDY 2: LEEDS RP DATA

- ‘DECISIONS’ data, following 500+ individuals travelling around Leeds, UK, for 2 weeks.
- total of 11,176 trips
- 6 different modes: Car, Bus, Rail, Taxi, Walk, Cycle
- Thousands of possibilities for NL and CNL!
- Hancock et al. (2024) model results include BIC values for MNL (9154) and a CNL with 2 nests (9144).
- Can availability MNL models allow us to identify a better CNL?

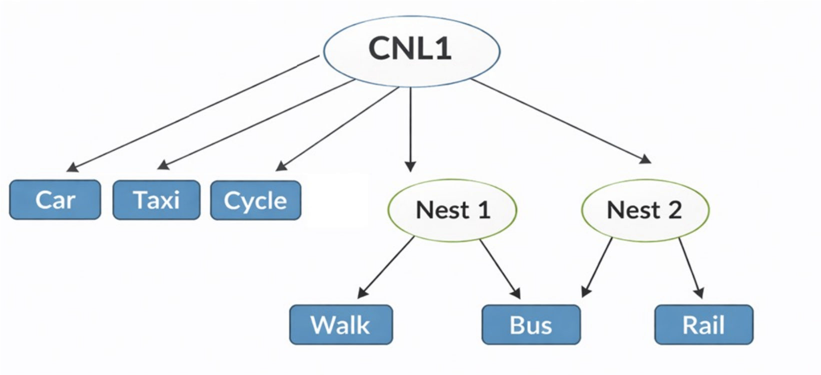
LEEDS RP DATA (MODEL PERFORMANCE)

Model	pars.	LL	BIC
MNL	20	-4,484	9,154
CNL	22	-4,469	9,144
MNL with availability	44	-4,433	9,275
MNL with availability (trimmed)	28	-4,468	9,196
New CNL	39	-4,433	9,230
New CNL (trimmed)	27	-4,442	9,136

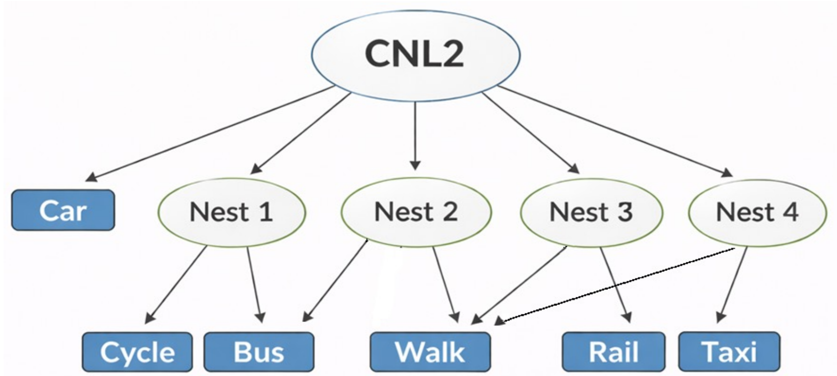
- 8 significant availability interactions pars (improved LL, but worse BIC)
- Suggests including 8/15 pairs (nests) in the CNL
- Some nesting parameter alphas -> 1, so the structure can be simplified
- Final model has a different (better!) structure to the original CNL

LEEDS RP DATA (NEW CNL MODEL)

Original CNL structure:



New CNL structure:



KEY DIFFERENCES IN MODEL STRUCTURE (DONOR VS. ESTIMATED MODEL)

Figure 27a. Model error structure from the Donor Model

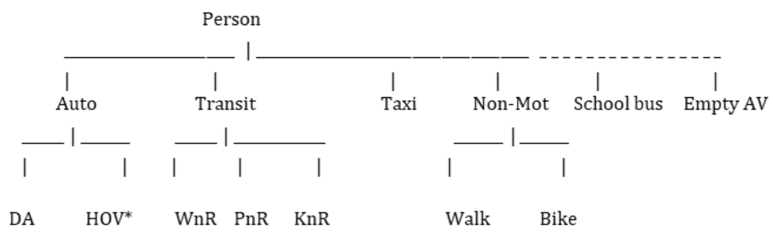
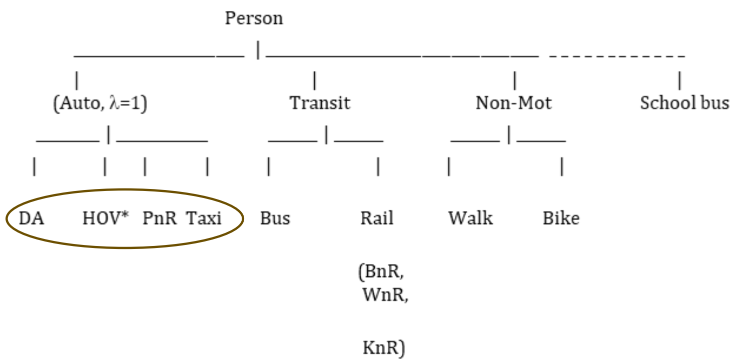


Figure 27b. Model error structure for PATHS estimates



TESTING

Model	Log Likelihood	McFadden's R^2	λ (PT)	λ (Active)	λ (Car)
Model 1: Simplified estimation model with new PATHS structure	–2,823.88	0.302	0.35	0.87	0.97
Model 2: Donor structure, free estimation	–2,836.35	0.299	1.06	0.88	1.31
Model 3: Donor structure with parameter constraints	–3,129.82	0.226	0.75	0.36	20.49
Model 4: New PATHS structure with Donor parameter constraints	–3,132.14	0.226	0.57	0.32	1.2

SUMMARY

- Stated Preference
  - Labelled alternatives
  - Number of alternatives is large (say > 4)
  - Partial Design used to assist respondent process or to mimic the market
  - Availability parameters help identify possible structure
- Revealed Preference
  - Strategic Transport Models
  - National Transport Models
  - Feasible / infeasible alternatives
  - Availability parameters help identify possible structure

Thank you

Planning and Transport Research Centre
The University of Western Australia (M087)
35 Stirling Highway, Crawley, WA 6009
https://patrec.org/

brett.smith@uwa.edu.au

doina.olaru@uwa.edu.au

tristan.reed@uwa.edu.au