# Heterogeneous substitution patterns based on current mode and time of departure choices in the context of uncertainty

Ying Huang<sup>1</sup>, Brett Smith<sup>1,2</sup>, Doina Olaru<sup>1,2</sup>

- 1. University of Western Australia, WA
- 2. Planning and Transport Research Centre PATREC, WA

# ABSTRACT

Recent modelling approaches to jointly estimating revealed preference (RP) and stated preference (SP) have incorporated random effects or error-components to account for the panel data characteristics when observing multiple SP choices. This paper extends the panel data approach to SP data enrichment models by estimating substitution patterns that depend on the respondent's current mode (RP). This is relevant to a choice experiment which incorporates the time-of-day choice along with the possibility of switching modes. The *substitution dependence* error-components structure of the model is used to uncover whether a respondent perceives a change in mode or a change in departure time as the closer substitute to their current travel choice.

## 1 Introduction

It is has been some time since the fusion of revealed preference (RP) and stated preference (SP) data has advanced from a single estimate of a scale parameter to jointly estimate two multinomial logits. Pooling revealed and stated preference data helped analysts to enrich the data set and minimise the biases brought from the hypothetical choice scenarios. The technique was sometimes referred to as the Nested-Logit Trick (Bradley & Daly 1994;

<sup>&</sup>lt;sup>1</sup> Contact: M251 The University of Western Australia, 35 Stirling Highway Crawley WA, Australia 6009 +61 8 64883979; <u>Brett.Smith@uwa.edu.au</u>

Hensher & Bradley 1993). The majority of nested logit models focus on uncovering the scale difference or correlations between RP and SP data, which helps to improve the accuracy of model's parameter estimation (Ben-Akiva & Morikawa 1990; Adamowicz, Louviere & Williams 1994; Swait, Louviere & Williams 1994).

A state dependence parameter was introduced by Bhat and Castelar (2002) to examine the propensity for the respondent to choose their current mode (RP) for tasks in the stated preference survey In addition error components may be introduced to account for correlations in the unobserved utilities (Bhat & Castelar 2002; Hensher, Rose & Greene 2008; Hensher 2008). The structure of the correlations relates to a flexible substitution pattern where some alternatives are perceived as being closer substitutes by respondents. To this date, the literature on joint RP-SP modelling has considered that the patterns of substitution are uniform across the population and the SP error component structure is independent of the individual's current mode (RP). This paper examines the possibility that the substitution patterns are *substitution dependence*; that is the correlations between PnR and private vehicle depend on whether the respondent is a current car user or rides on other public transports.

The empirical component of the paper is based on a RP/SP survey undertaken between September 2013 and April 2014 in Perth focussing on park-and-ride (PnR) as a travel alternative. At the time parking availability was tight with PnR stations becoming full by 7:00am. As such the focus of the SP instrument was on time of departure and the risk of not finding a parking bay. The following section discusses these issues. The econometric model is presented in Section 3. This is followed by a description of the data (Section 4) and the estimation results (Section 5). Policy implications are discussed in Section 6.

# 2 Parking Availability, Travel Time Variation and Time of Departure

PnR, is seen as an attractive public transport alternative in low-density cities (e.g. Perth). Parking availability, which may be defined from the commuter's point of view as being their expectation on whether they will secure a parking bay, emerged as an important attribute affecting travellers' preference for PnR (Debrezion, Pels & Rietveld 2007; Chakour & Eluru 2014). Park-and-ride users tend to depart earlier than commuters using other modes, due to limited number of parking bays, indicating parking availability is highly correlated with departure time. In Perth, Western Australia, prior to July 2014, almost half of the 200,000 daily rail passengers access the stations by car. This put excessive pressure on the PnR

facilities with the free parking zones at some stations being at capacity well before 7:00am. In addition, the paid parking bays (AUD \$2 per day) were mostly occupied before 8:00am (Orr, 2014). On the 1st of July 2014, all bays attracted the \$2 fee. There is reported evidence that the fee has meant that commuters are arriving later at the PnR Stations (Acott, 2014) with major PnR stations (1,000 or more bays) being fully occupied by the end of peak hour (around 9:00am) but with smaller stations not reaching capacity. The results given here are from a study undertaken nine months prior to the introduction of the fee.

The attributes of commuter mode alternatives are highly dependent on the time of departure. To this end, studies into the time-of departure choice are often studied in a joint setting with mode choice (Arnott et al., 1990; Bhat, 1998a; Hess et al., 2007a, 2007b). London commuters are willing to change their transport mode, more so than to substantially alter their departure time (Hess et al., 2007a, 2007b). New Jersey commuters tended to adjust their departure schedule to secure the travel time reliability rather than to change their routes (Ozbay & Yanmaz-Tuzel 2008). However, based on data collected in San Luis Obispo, Yan, Small and Sullivan (2002) concluded that individuals are more likely to change routes than departure time.

Commuters face day-to-day travel time variation (Senna 1994; Noland & Polak 2002) and crowding in public transport vehicles (Li & Hensher 2013), when making departure time and mode choices decisions. Travel uncertainty, especially road congestion, is highly correlated with the departure time (De Jong et al. 2003; Hess et al. 2007). Departure time choices are often studied in combination with the choice of mode (Arnott et al., 1990; Bhat, 1998a; Hess, Daly, et al., 2007; Hess, Polak, et al., 2007) or route (Bajwa et al., 2008; Lam & Small, 2001; Yan et al., 2002).

This paper makes use of scale difference, state dependence and error components to model unreflective substitution between modes and departure times with RP and SP panel data, through segmenting the sample into those who commuted by car, by PnR and those who rode public transport. Respondents were presented with uncertain travel times and levels of parking availability at railway station. The heterogeneity and dependence of error component on current mode users observed in RP data is examined. The analysis then aimed to highlight whether there are differences among these segments in terms of their willingness to switch to an alternative mode or to change their departure time, in addition to scale difference and state dependence between RP and SP data.

#### **3** The Econometric Model

The substitution between time-of-departure and mode alternatives is traditionally modelled as a mixed logit or error component logit model (McFadden & Train 2000; Revelt & Train 1998). A general form of the utilities for the joint RP-SP time of day choice model is given by:

$$U_{j1n(RP)} = \sum_{k=1}^{K} (\bar{\beta}_k + \gamma_k z_{j1nk}) X_{j1nk} + \varepsilon_{jnk}$$

$$U_{jtn(SP)} = \lambda_{sp} \left[ \sum_{k=1}^{K} (\bar{\beta}_k + \gamma_k z_{jtnk}) X_{jtnk} + (1 - k_{rp}) \vartheta_n \right] + \theta_m \mu_{mn} d_m + \theta_t \mu_{tn} d_t + \varepsilon_{jn}$$
(1)

where  $U_{jtn}$  is the utility for respondent *n* associated with alternative *j* for choice task *t* The RP choice task is observed only once and is assigned t=1. The attributes  $X_{jink}$  taste parameters associated with the mode *j* at time *t*, have a distribution of mean,  $\bar{\beta}_k$  and a standard deviation  $\gamma_k$  where  $z_{jnk}$  is a random draw from a standardised normal (in this case) distribution.  $\vartheta_n$  is the state dependence parameter of alternative *n* in SP data,  $k_{rp}$  is a binary variable which equals 1 when the SP alternative is the same as the currently chosen mode (RP).  $\lambda_{SP}$  represents the scale difference between SP and RP data by normalising RP scale to the unity. The residual component of the unobserved utilities,  $\varepsilon_n$ , is modelled as an independent Extreme Value type I distribution. Social Demographic variables may enter as interactions with alternative specific constants, with mode attributes or by specifying the random parameters as conditional variables on the characteristics of the individual (i.e.,  $\bar{\beta}_k$ ,  $\theta_m$ , or  $\theta_t$  could be functions of a socio-demographic variable).

For the SP alternatives respondents are offered the possibility to choose their current mode but an earlier or later departure time. The random variables  $\mu_{mn}$  and  $\mu_{tn}$  have a standardised normal distribution and  $\theta_m$  and  $\theta_t$  are the standard deviations of the error components for mode and time of departure respectively. The indicators  $d_m$  and  $d_t$  specify which of the modes and departure times are included in the error component for each individual. In a standard error component model it is assumed that the same error components are estimated for the entire sample. This paper explores the possibility that the error components differ across the sample based on the respondent's RP mode. By allowing substitution patterns (significant error components standard deviations) to be dependent on the respondent's current mode choice, we are picking up a form of state dependence, but in the unobserved utility components. The model incorporates *substitution dependence*.

Three sources of uncertainty are considered relevant in choosing between mode and departure time: a) parking availability; b) crowding; c) travel time reliability. This measure of crowding on public transport is related to discomfort which is incorporated in the models by multiplying the probability of finding a seat by the duration of the trip. Parking availability is modelled as a risk-adjusted cost. It is assumed that commuters prefer to secure free parking bays at the station, but face the uncertainty of having to pay for parking or being unable to locate a bay, thus incurring additional costs or penalties. Expected car travel time is used to describe travel time based on the objective values (fast travel time, usual travel time and slow travel time associated with probabilities) presented in the stated preference survey.

#### 4 The Survey Instrument

Perth, capital of Western Australia, has two million residents, spreading along the coast of the Indian Ocean for more than 150km. There is a radial rail network with five corridors and 69 stations. The East-West lines were built in the late 19th century (heritage lines), the North line opened in 1992, and the South line in 2007. Public transport has less than a 15% (9.8% train and 4.4% bus) modal share. Survey results show that 34.8% of train commuters are PnR users. Numerous planning initiatives to curb urban sprawl and create transit oriented developments around (mainly new) numerous train stations are expected to change the modal split.

Two household revealed preference (RP) and stated preference (SP) survey were undertaken between September 2013 and April 2014. The sample was drawn from the catchments of seven train stations with PnR facilities in Perth, WA. The attribute levels for the stated preference were pivoted on the revealed preference responses for labelled alternatives. Each respondent received six stimuli. An efficient design (Rose & Bliemer 2008) was developed and optimised using genetic algorithm (Olaru et al., 2011). In total, 215 commuters responded both the RP and SP surveys. This paper is based on 122 who indicated some flexibility in their morning commute and were, therefore, presented with both mode and departure time options. An extract from the survey is presented in Figure 1.

The SP show cards always included the respondent's current (revealed preference) mode with an earlier or later departure time alternatives, pivoted on the current time of departure. The purpose of including a time-of-day alternative was to present attributes conditioned on the greater/lesser degree of uncertainty faced when travelling within or outside the peak.



Figure 1: Revealed Preference Survey and pivoted Stated Preference instrument

## 5 Results and analysis

Three mixed logit models were estimated: a random parameter logit (RPL) model, an error component mixed logit model, EC1 mixed logit, that has standard error components for mode choice and a second error component mixed logit model, EC2 mixed logit, incorporates *substitution dependence*. All models allow for taste heterogeneity in expected car travel time, expected PnR parking cost as well as crowding on public transport, which are associated with travel uncertainty during daily commuting.

In terms of substitution among modes, the difference between error components mixed logits, EC1 and EC2, is whether error component estimation is dependent or independent on respondents' current mode choices observed in RP data. Thus the former model captures the substitution between alternatives across all respondents. The latter formed under the assumption that the flexible substitution effects are based on the respondent's current mode (RP) choice.

	RPL		EC1 Mixed logit		EC2 Mixed logit				
Variable									
	Est.	t	Est.	t	Est.	t			
Alternate Specific Constants									
Constant for Car* in RP	-0.226	0.18	2.609	1.60	1.980	1.64			
Constant for PnR in RP	-3.398	3.66	0.093	0.14	0.151	0.25			
Constant for KnR in RP	-2.889	3.27	-3.259	3.55	-3.234	3.65			
Constant for Bus in RP	1.667	1.96	-0.010	0.01	-0.246	0.28			
Constant for FnR in RP	-0.309	0.58	-0.445	0.61	-0.411	0.79			
Constant for Car* in SP	-3.310	3.73	-1.680	1.14	-2.201	1.99			
Constant for PnR in SP	-4.016	5.89	-1.787	2.05	-1.730	2.63			
Constant for KnR in SP	-0.965	1.94	-1.728	2.19	-1.601	2.29			
Constant for Bus in SP	-0.359	0.62	-2.321	2.59	-2.430	3.30			
Constant for FnR in SP	-0.843	1.96	-1.449	1.96	-1.195	2.11			
Means and Standard Deviations for Random Parameters									
Expected car travel time	-0.153	5.79	-0.100	2.43	-0.069	2.13			
Expected cost for PnR	-0.532	3.79	-0.407	3.37	-0.548	3.19			
Crowding on PT**	-0.051	2.18	-0.061	2.17	-0.063	2.73			
St.Dev for Average car travel time	0.105	6.66	0.047	2.27	0.039	1.65			
St.Dev for Expected cost for PnR	0.468	3.91	0.017	0.04	0.332	1.74			
St.Dev for Crowding on PT	0.118	4.60	0.026	0.29	0.030	0.57			
Non-random parameters									
Travel cost (\$)	-0.333	9.34	-0.366	8.06	-0.366	8.07			
Access time for PT	-0.070	3.11	-0.082	2.67	-0.081	3.32			
In-vehicle travel time for PT	-0.110	5.47	-0.061	2.26	-0.052	2.36			
Early departure time for modes (excluding	0 905	2 40	0 820	1 27	1 006	1 60			
PnR)	-0.000	3.40	-0.820	1.37	-1.000	1.09			
Late departure time for modes (excluding	-2 270	7 07	-3 100	1 10	_2 1/0	3 10			
PnR)	-2.219	1.21	-3.400	4.40	-3.140	J.40			
Early departure time for PnR	-0.357	0.41	-0.906	0.42	-2.205	1.64			
Late departure time for PnR	-3.252	1.73	-5.323	1.47	-7.015	2.04			

#### Table 1 Estimated Random Parameter Logit (RPL), Error Component ML models

\*Alternative specific constant for Train with walk as access mode is zero in both RP and SP data.

\*\* PT include bus and train with dropped off (KnR), feeder bus (FnR), walk (WnR) access modes as well as PnR (sometimes a reference is made to all other public modes as PT excluding PnR)

Table 1 Continued, Estimated Random Parameter Logit (RPL), Error Component ML models:
Error Component Structure

Variable	RPL		EC1 Mixed logit		EC2 Mixed	
			ECT MIXed logit		logit	
	Est.	t	Est.	t	Est.	t
Error Component Value						
Car drivers departure time			2.894	2.60	3.208	2.61
PnR users departure times			2.234	1.46	2.100	2.76
PT (excluding Pnr) commuters departure			0.740	4 77	0.500	0.04
time			3.710	1.77	3.530	3.04
Substitutions between car and bus			2.105	4.54		
Current mode car or bus					2.528	5.71
Current PT (including PnR)					0.025	0.04
Substitutions between car and PnR			0.790	1.26		
Current mode Car and PnR					1.086	1.52
Current PT (excluding PnR)					0.329	0.34
Substitutions between PT and PnR			2.369	2.57		
Current mode PnR and all other PT					2.846	4.03
Current mode Car					1.495	1.67
Substitution between PT modes			0.000	0.07	4 000	0.54
(excluding PnR)			2.203	3.67	1.929	3.54
Scale parameter***	0.715	2.82	0.514	8.23	0.715	2.82
State dependence						
State dependence for car	1.687	2.46	2.87	1.65	3.094	2.50
State dependence for PnR	1.823	2.71	3.40	2.13	3.889	2.78
No. of observations		8	s)			
Information Criterion: AIC/N	1.862		1.704		1.701	
Log-likelihood	-735.246		-667.266		-662.805	
pseudo r <sup>2</sup>	0.311		0.375		0.379	

\*\*\* Scale parameter for best choice data is normalised to 1.0 and the values 0.514 and 0.715 are estimated from non-linear RPL model, which is not presented here.

Results indicate the negative relationship between expected car trave time and car drivers' utility function. The random parameter models also capture significant taste heterogeneity among car drivers. The expected parking cost at railway station and crowding on public transport are key determinants affecting commuters' preference for PnR and public transport. The estimation does not show significant taste variation for crowding in error component mixed models.

In non-random parameters estimation, results also indicate that travel time and travel costs are important factors commuters' mode choices. Commuters other than PnR users are less prone to shift current departure time, if it results in the loss of punctuality. Moreover, commuters are more reluctant to depart late than early. In terms of PnR users, the tendency of shifting to earlier departure time is unclear, while their strong objection to late departure is undoubtedly due to competition to secure parking bays at railway station.

The SP scale parameter of less than one (0.715 and 0.514) indicates more unobserved disturbances in SP data, which is consistent with previous empirical exercises (Brownstone, Bunch & Train 2000; Hensher, Rose & Greene 2008). A Significant and positive state dependence parameter for car<sup>2</sup> and for PnR suggests a degree of attachment to the current mode (Bhat & Castelar, 2002). The inclusion of error components in the utility functions to account for flexible substitution between modes and departure times provides a better explanation of the choice data (EC1 compared to RPL). The error component estimates suggest that car drivers perceive the time-of-departure alternatives as being closer substitutes than other modes – all else being equal – which is in line with conclusions from (Bhat 1998; Hess et al. 2007). PnR users perceive less similarity (in the unobserved utilities) in alternate departure times than with competing public transport (PT) modes. However, this may be an artefact of the design which systematically offered the later departure time with lower probabilities for securing a parking bay. Like the car drivers, commuters on all other forms of public modes (excluding PnR) consider an alternate departure time as being a closer substitute than finding a second public mode.

The EC2 mixed logit model further uncovers substitution patterns that relate to the respondent's current mode (RP) choice. The higher the error component estimate in Table 1 indicates a greater cross elasticity for the identified group of commuters. The relatively larger standard deviations for car to bus and car to public transport indicate car drivers have a closer substitution between car and bus. However non-car or bus users do not appear to have a significant correlation between these two modes. The same inference can be made

 $<sup>^{2}</sup>$  The state dependence parameter for car is significant at the 5% level for two of the three models and at the 10% level of significance for estimation model EC1.

on the substitution pattern for PnR and all other PT modes. The significance of the substitution parameter for car and PnR is unclear when using a single error component or when separating the sample into those who use these modes currently and those who do not. However, by splitting the unobserved effects into these groups reveals the general pattern of the larger t-value being associated with the current users of the relevant modes. Interestingly, the substitution pattern for within PT modes did not appear to vary across the sample: i.e., a similar parameter value is estimated for car or PnR commuters and riders on PT modes (excluding PnR) riders.

### 6 Conclusions

This paper investigates commuters' choices of time of day and transport mode in the context of travel uncertainty, using revealed preference and stated preference survey data collected in Perth, Western Australia. Error components are used to capture substitution effects between adjacent departure times and alternative modes. The respondents' current mode choices influence their responses to the stated preference instrument. As in previous studies (Bhat & Castelar 2002; Hensher, Rose & Greene 2008; Hensher 2008) the choice data revealed an inertia or habit by way of a significant state dependence parameter. However, an examination of the correlations in the unobserved utilities uncovered flexible substitutions that were also dependent on the respondent's current mode.

Compared with the traditional standard error components structure, the *substitution dependence* error component mixed logit model provides a significant improvement in understanding substitution patterns for each subsample. The segment of commuters' substitution patterns could potentially influence policy because it uncovers the strength of substitution for targeted commuters. By incorporating substitution dependence assists forecasting responses (more or less likely to switch to one mode or another) to road travel conditions, changes in working schedules, improvement of public transport service or the construction of more parking bays at railway station.

Whilst this paper makes a contribution towards better modelling approaches to enriching RP data with SP experiments, the magnitude of the policy implications has not been examined. However, the authors feel that such an examination requires more than one data set. Future work will investigate whether the finding here are robust across a number of RP/SP mode choice surveys as well as explore the policy implications for urban transport policy.

#### REFERENCES

- Adamowicz, W, Louviere, J & Williams, M 1994, 'Combining revealed and stated preference methods for valuing environmental amenities', *Journal of environmental economics and management*, vol. 26, no. 3, pp. 271-292.
- Ben-Akiva, M & Morikawa, T 1990, 'Estimation of switching models from revealed preferences and stated intentions', *Transportation Research Part A: General*, vol. 24, no. 6, pp. 485-495.
- Bhat, CR 1998, 'Analysis of travel mode and departure time choice for urban shopping trips', *Transportation Research Part B: Methodological*, vol. 32, no. 6, pp. 361-371.
- Bhat, CR & Castelar, S 2002, 'A unified mixed logit framework for modeling revealed and stated preferences: formulation and application to congestion pricing analysis in the San Francisco Bay area', *Transportation Research Part B: Methodological*, vol. 36, pp. 593-616.
- Bradley, M & Daly, A 1994, 'Use of the logit scaling approach to test for rank-order and fatigue effects in stated preference data', *Transportation*, vol. 21, no. 2, pp. 167-184.
- Brownstone, D, Bunch, DS & Train, K 2000, 'Joint mixed logit models of stated and revealed preferences for alternative-fuel vehicles', *Transportation Research Part B: Methodological*, vol. 34, no. 5, pp. 315-338.
- Chakour, V & Eluru, N 2014, 'Analyzing commuter train user behavior: a decision framework for access mode and station choice', *Transportation*, vol. 41, no. 1, pp. 211-228.
- De Jong, G, Daly, A, Pieters, M, Vellay, C, Bradley, M & Hofman, F 2003, 'A model for time of day and mode choice using error components logit', *Transportation Research Part E: Logistics and Transportation Review*, vol. 39, no. 3, pp. 245-268.
- Debrezion, G, Pels, E & Rietveld, P 2007, 'Choice of departure station by railway users', *European Transport*, vol. 37, pp. 78-92.
- Hensher, DA 2008, 'Empirical approaches to combining revealed and stated preference data: Some recent developments with reference to urban mode choice', *Research in Transportation Economics*, vol. 23, no. 1, pp. 23-29.
- Hensher, DA & Bradley, M 1993, 'Using stated response choice data to enrich revealed preference discrete choice models', *Marketing Letters*, vol. 4, no. 2, pp. 139-151.
- Hensher, DA, Rose, JM & Greene, WH 2008, 'Combining RP and SP data: biases in using the nested logit 'trick'–contrasts with flexible mixed logit incorporating panel and scale effects', *Journal of Transport Geography*, vol. 16, no. 2, pp. 126-133.
- Hess, S, Polak, JW, Daly, A & Hyman, G 2007, 'Flexible substitution patterns in models of mode and time of day choice: new evidence from the UK and the Netherlands', *Transportation*, vol. 34, no. 2, pp. 213-238.
- Li, Z & Hensher, D 2013, 'Crowding in Public Transport: A Review of Objective and Subjective Measures', *Journal of Public Transportation*, vol. 16, no. 2, pp. 107-134.
- McFadden, D & Train, K 2000, 'Mixed MNL models for discrete response', *Journal of applied Econometrics*, vol. 15, no. 5, pp. 447-470.
- Noland, RB & Polak, JW 2002, 'Travel time variability: a review of theoretical and empirical issues', *Transport Reviews*, vol. 22, no. 1, pp. 39-54.
- Ozbay, K & Yanmaz-Tuzel, O 2008, 'Valuation of travel time and departure time choice in the presence of time-of-day pricing', *Transportation Research Part A: Policy and Practice*, vol. 42, no. 4, pp. 577-590.
- Revelt, D & Train, K 1998, 'Mixed logit with repeated choices: households' choices of appliance efficiency level', *Review of economics and statistics*, vol. 80, no. 4, pp. 647-657.
- Rose, JM & Bliemer, MC 2008, 'Stated preference experimental design strategies', Handbook of Transport Modelling, Elsevier, Oxford, pp. 151-180.
- Senna, LA 1994, 'The influence of travel time variability on the value of time', *Transportation*, vol. 21, no. 2, pp. 203-228.

- Swait, J, Louviere, JJ & Williams, M 1994, 'A sequential approach to exploiting the combined strengths of SP and RP data: application to freight shipper choice', *Transportation*, vol. 21, no. 2, pp. 135-152.
- Yan, J, Small, KA & Sullivan, EC 2002, 'Choice models of route, occupancy, and time of day with value-priced tolls', *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1812, no. -1, pp. 69-77.