ANALYSIS OF THE IMPACT OF ACCESSIBILITY ON MODE CHOICE USING REVEALED PREFERENCE AND STATED PREFERENCE INFORMATION

CE271 Choice Modelling Project

Gavaskar Muppavaram
SR No 21661

1. Introduction:

Revealed preference data includes the choices reported by the travellers and are related to travel alternative characteristics available to them. Stated preference data is based on the responses to hypothetical situations; they have been used in cases where observed revealed preference data is inadequate to model the context of interest.

This study tries to combine both data and exploit the advantage of each type of data in drawing conclusions about mode choice behaviour [1]. The accuracy level and errors vary between revealed preference and stated preference data, and the errors will be reflected in the parameters estimated in the models. Hence the objective of the model is to combine the stronger features of RP and SP data.

The key advantage of combining both types of data is bias correction, efficiency and identification of new parameters or new alternatives that are not in the revealed preference data. A tree logit estimation technique is applied for modelling the data.

Earlier, Bradly & Daly [2] developed an integrated approach to combine SP and RP data using the logit estimation technique. The same framework is adopted for this study.

Combined SP RP model code is obtained from Apollo package. The theory behind the framework of the model developed is studied, and estimation is done using the same.

Overview:

This study is structured as follows:

Section 2 includes a literature review of the model used in the study.

Section 3 provides the theoretical framework of the integrated estimation approach, the model's structure and the joint utility function specification.

Section 4 describes the RP and SP experiment design and statistics of the data used in the study.

Section 5 presents the results and inference of the evaluated models.

Section 6 provides the conclusion of this study.

2. Literature review

***Combining Revealed and Stated Preference Data:* Ben-Akiva et al.**

This study describes the various statistical methodologies to combine multiple data sources in the estimation of choice models and current state-of-the-art data combination methods used in market research. The authors describe different types of survey data, some of which include revealed preference, stated intention, stated choice and attribute ratings. The paper also provides insights into research contexts in which they can be combined.

Estimating the model on both SP and RP data includes common parameters along with RP-specific and SP-specific model parameters. Due to the difference in the variance of RP and SP utility functions, the authors necessitate estimating a scale parameter to equalise their scale.

When multiple SP experiments are used, each data source has a unique scale parameter, and normalisation is achieved by setting the scale of RP data to one. The presence of one or more scale parameters introduces non-linearity requiring a special estimation procedure. A sequential estimation procedure [3] or estimating SP and PR data simultaneously [2] can be adopted.

When both RP and SP data are obtained from the same individuals, a correlation may exist between the RP and SP observations. More specifically, the random components of the two utility functions may be correlated due to unknown taste variation and omitted variables specific to the individual. This approach may involve complicated estimation techniques [4] but yields better statistical efficiency.

Estimation of Logit Choice Models Using Mixed Stated-Preference and Revealed-Preference Information

Bradly and Daly (1991)

This paper proposes using logit estimation techniques to use RP and SP data in an integrated model estimation framework.

The model adopted includes the RP utility function of RP-specific variables and the SP utility function of SP-specific variables. And a few variables which are common to both utility functions. The objective was to use data from two contexts and make a joint estimate. Independence between SP observations and between RP and SP observations from the same individual are assumed, and their correlation was not explored.

The authors suggest using the logit model in the case of binary choice models. However, in case of more than two alternatives, a tree logit model is suggested, or a probit model is based on the assumptions among the alternatives.

A scale parameter is used to scale the SP observations to equal the variance of error components of SP and RP observations, making it possible to perform a joint estimation of SP and RP data.

Finally, a case study compared the logit estimation of a data set on individual choice modelling to probit estimation on the same data. The tree logit estimate gave consistent results compared to the probit model for two alternative contexts.

3. Model Structure:

The utility maximized by travellers in their revealed preference U^{RP}, for a given traveller, for a given alternative is

$$U^{RP} = \beta X^{RP} + \alpha W + \varepsilon$$

Where

X^{RP}, W are vectors of measured variables influencing the RP decision.

β, α are vectors of unknown parameters to be estimated.

ε represents unmeasured utility components influencing RP decision.

The SP utility function is given as

$$U^{SP} = \beta X^{SP} + \eta Z + \rho$$

Where

X^{SP}, Z are vectors of measured variables influencing the SP decision.

β, η are vectors of unknown parameters to be estimated.

ρ represents unmeasured utility components influencing SP decision.

The measured variables w may occur only in the RP context, and Z may occur only in the SP context.

Assumptions

The variable X is in both the utility functions and their coefficient can be estimated using both surveys. Hence, by using this formulation, the objective of exploiting data from both sources is accomplished. It becomes necessary to assume that the effect of variable X in SP and RP contexts is the same.

Interdependence between SP and RP unobserved factors is assumed to be independent, and the results from previous studies suggest that the problems caused by this assumption are limited in their practical effect.

The distributional assumption of unobserved components is assumed to be Gumbel, leading to the use of logit models for probability calculations.

The existence of interdependence between SP observations from the same individuals makes it impossible to assume that the variance of unobserved components ε and ρ is equal. Hence the utility function of SP data is scaled to make its variance equal to that of the RP utility function.

Estimation Procedure

The unobserved components ε and ρ are assumed to be independently distributed with Gumbel but with unequal variance.

$$\theta^2 = \frac{\mathrm{var}(\varepsilon)}{\mathrm{var}(\rho)}$$

The modified SP utility is

$$\theta U^{SP} = \theta \beta X^{SP} + \theta \eta Z + \theta \rho$$

Now, both the unobserved components have the same variance, and it is possible to use both RP and SP observations in a logit estimation procedure that requires equal variance across the observations.

The SP model is now non-linear, and a tree logit model, as shown in Figure 1, is used.

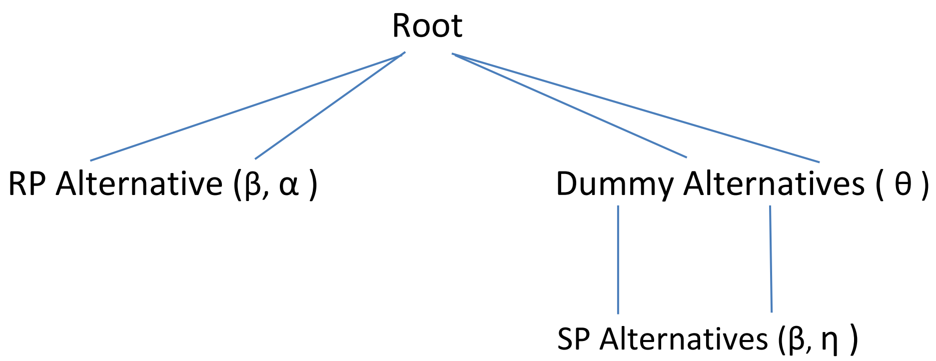

Figure 1. Tree structure for joint estimation

In this study, RP data is modelled first, and then the combined RP-SP model is estimated. RP data is modelled using Multinomial Logit Model (MNL). The description of Variables used in the data is given in Table 1

Variable	Description
TT_MODE	Travel time from source to target by that mode in minutes
COST_MODE	Travel cost from source to target by that mode in rupees
DIST	Distance from source to target in Kms
TRANSFERS	Number of transfers by bus from source to target
CROWD	Crowding levels in bus, categorial variable ranging from 1 to 4
WORK	If the trip is work-based, then it takes 1. Else, 0
INCOME	Categorial variable of income ranging from 1 to 9

Table 1. Description of variables in the model

Utility equations

V[Metro] = asc_metro + b_tt * TT_METRO*WORK +b_cost * COST_METRO/INCOME

V[Bus] = asc_bus + b_tt * TT_BUS *WORK + b_transfers_bus*TRANSFERS+b_cost *COST_BUS/INCOME +b_crowd *CROWD

V[Car] = asc_car +TT_DRIVE*b_tt*WORK + b_cost *COST_CAR/INCOME

V[TW]= asc_tw + b_tt *TT_DRIVE*WORK +b_cost* COST_TW/INCOME

4. Data Description:

The dataset includes CisTUP Travel Behaviour Survey in Bengaluru. The modal share in RP data is shown in Figure 2. The share of public transport is 38% only. The income levels of the individuals are shown in Figure 3. Most respondents have an income in the range of Rs. 10,000 to Rs. 80,000 per month. 538 RP survey respondents are considered, and 1856 SP responses from the same individuals are included in SP data.

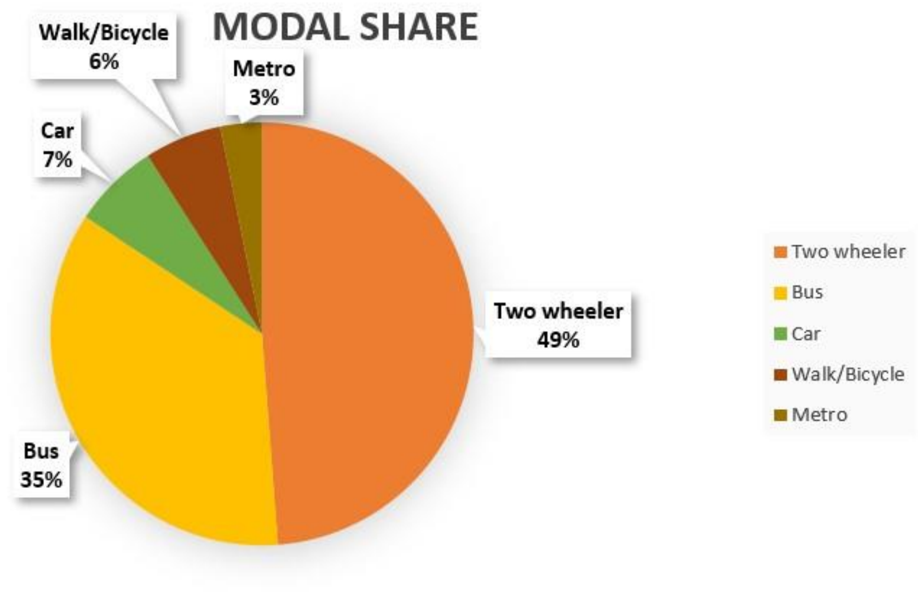

Figure 2: Modal share in collected Data

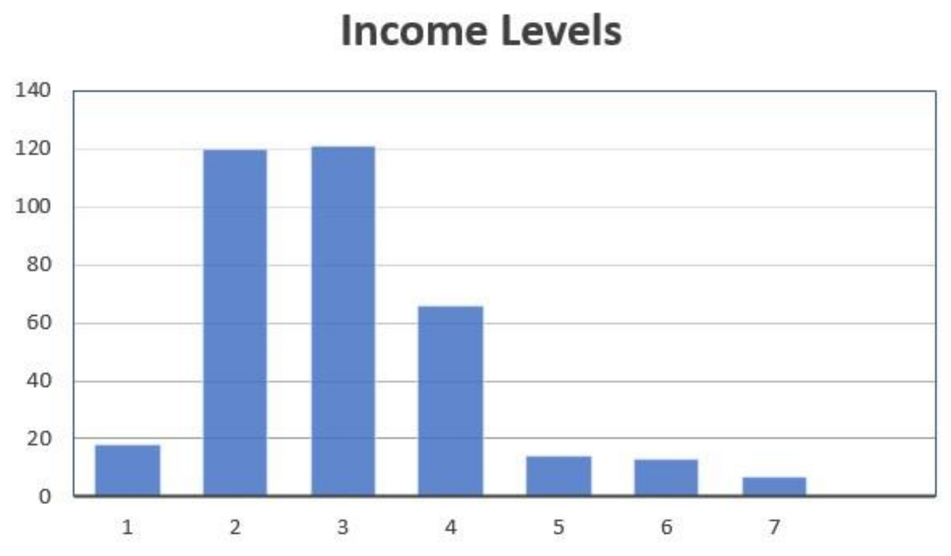

Figure 3. Income level distribution

RP questionnaire and SP experiment design:

Revealed Preference Questionnaire included two types of questions. 1) Socio-Demographic details age, income, gender, and vehicle ownership. 2)Travel information details travel time, primary mode, and travel cost. The attributes of the Stated Preference experiment, Socio-Demographic and their levels are shown in Table 2.

Attribute	Mode	No of levels	Values	Units
Age	-	6	1-6	-
Income	-	7	1 to 7	Rupees
Distance Range	-	8	2,5,10,15,20,25,30	Km
Crowding	Bus	4	1 to 4	-

Table 2. SP and Socio-Demographic variables and their levels

Waiting time, in-vehicle travel time, and transfers Data for Bus trips are generated by **Google Maps API**. Google Application Programming Interface (API) is a service provided by Google and is used to generate accurate information on the travel attributes of a trip. Variables such as travel time and the number of transfers for bus mode are obtained accurately using API.

5. Results and discussion:

Variable	RP (MNL)	SP-RP
asc_car	- (base)	-(base)
asc_bus	6.71(6.19)	3.93 (10.97)
asc_metro	4.69 (4.61)	0.68 (2.59)
asc_tw	5.93 (5.91)	2.62 (10.17)
b_tt	-0.17 (-8.25)	-0.05 (-5.28)
b_cost	-	-0.10 (-4.72)
transfers	-0.34 (-2.91)	-0.87 (-11.10)
b_crowding	-0.15 (-1.1)	-0.18 (-2.10)
Scale (θ)	-	0.415
LL	-356.78	-1936.79
ρ^2	0.163	0.033
Adj ρ^2	0.149	0.029

Table 3. Model estimates

Variables vehicle ownership, frequency of routine travel, and categorial variable age were found to be insignificant with current specifications and hence were omitted. In the RP model, the cost variable was insignificant and was omitted from RP only model.

All the parameter estimates are statistically significant, with an 80% confidence level in the RP model and a 95% confidence level in the SP-RP joint model.

The signs of all the parameters obtained are logical. Consider, for instance, the travel time parameter, which is negative, implying that an increase in the cost of a mode for a work-based commute discourages them from taking that mode compared to trips for other purposes, as a work-based commute is a regular activity.

Cost is interacted with income, and the negative sign of the estimate Indicates that with the increase in income, cost does not play much significance in choosing a mode.

The scale parameter is less than 1, implying that the SP responses have significantly more unexplained variance than RP data.

Elasticities:

The effect of an increase in transfer by 1 lead to a shift of 47% of travellers from Bus to other modes in the case of RP-only mode. In the case of SP data, the shift is 17%

Increasing the cost of bus fare up to 20% doesn't impact change in ridership in the case of both SP and RP data.

To increase ridership in public transit, if we penalise the car and two-wheeler users with up to 20% of their travel cost using private vehicles as a toll, then the observed mode shift is negligible.

If by using dedicated bus lanes, reducing the number of stops by increasing fleet size, we reduce the travel time of bus alternative by 15%, then the increase in ridership of bus mode observed is 3.4 % in the case of RP data and about 1% increase in case of SP data is predicted by the model.

6. Conclusion:

In the present model specification, the unobserved variables are assumed to be Gumbel distributed, but normal distribution can be considered, which leads to a model with better sensitivity to outliers, as evident from previous studies [4]

Also, the assumption of no correlation between SP observations from the same individual can be relaxed. Also, interdependence between RP and SP observations from the same individual can be addressed by considering correlation.

When multiple SP experiment data are used, we can estimate a separate scale parameter for each data. In this model, only one scale parameter is estimated.

Another Model can be made comprising only the users captive by choice, not by force. That is, studying the mode choice of Individuals who own personal vehicles will help the agency to pull the private vehicle users and increase the public transit ridership by improving the variables suggested by the model.

Reference:

1. Ben-Akiva, M., Bradley, M., Morikawa, T. et al. Combining revealed and stated preferences data. Market Lett 5, 335–349 (1994).
2. Bradley, Mark & Daly, Andrew & Stopher, Peter & Lee-Gosselin, M. (1996). Estimation of Logit Choice Models Using Mixed Stated-Preference and Revealed-Preference Information.
3. Ben-Akiva M., and T. Morikawa, (1990b). "Estimation of Travel Demand Models from Multiple Data Sources." In M. Koshi (ed.), Transportation and Tr«~ffi« Theory. New York: Elsevier, pp. 461-476.
4. Morikawa, T., Ben Akiva, M. and McFadden, D. "Incorporating Psychometric Data in Econometric Demand Models". Prepared for Banff Invitational Symposium on Consumer Decision Making and Choice Behavior. Canada. May 1990