



# Accommodating the Heterogeneity in Traveller's Responsiveness in Safe Route Choice Experiment – Study from India

Sivakumar Balakrishnan <sup>1\*</sup>, Krishnamurthy Karuppanagounder <sup>2</sup>

<sup>1</sup> *Research Scholar, Traffic and Transportation Planning, Department of Civil Engineering, National Institute of Technology, Calicut-673601, Kerala, India*

<sup>2</sup> *Professor, Traffic and Transportation Planning, Department of Civil Engineering, National Institute of Technology Calicut-673601, Kerala, India*

---

## Abstract

Developing countries like India lacks research works to examine how the heterogeneity in responsiveness affect the outcome of a stated choice experiment. This study is an attempt to observe the effects of variation in the responsiveness of a person to different levels of parameters in the selection of a safer route in a country like India. This paper applies a Mixed Logit (ML) model to analyze the safer route choices done by the two-wheeler road users. This model is extended to account for the heterogeneity and explain its impact on the willingness to pay (WTP) for accident reduction. The empirical study shows that if the heterogeneity across individuals are not accounted adequately it leads to a statistically inferior data fit and also to inappropriate evaluations. The calculated WTP values suggest that these values are sensitive to preference heterogeneity and are underestimated if an assumption of preference homogeneity is made.

*Keywords:* willingness to pay; stated preference survey; mixed logit; heterogeneity.

---

## 1. Introduction

Road crashes being one of the most serious impact on the economic development of a country, costing of traffic crashes is extremely inevitable to identify the sums of money that should be invested on the road safety improvement works. Due to global awareness about the financial implications of the crashes, public in developed countries have shown greater concern on the remedial measures and contributed extensively in the decision-making process to improve the transportation infrastructure safety. Costing of the road traffic accidents is essential to provide a guideline on the quantity of resources to be spent on the prevention of road traffic accidents or injuries. It is highly advisable to take into account the public approach towards safety and their Willingness to Pay (WTP) to bring

---

\* Corresponding author: Sivakumar Balakrishnan ([bsivakumar0071@gmail.com](mailto:bsivakumar0071@gmail.com))

down their involvement as a victim for evaluating the acceptance of any road safety improvement scheme. Models to calculate the WTP values against traffic crashes provide a tool to evaluate public response towards alternative traffic safety strategies. WTP values to reduce the involvement in road crashes have traditionally been obtained from simple models of safer route choice.

Accurate estimations of the public response towards a safer route choice require development of models which capture differences in intrinsic preferences and differences in sensitivity to various levels of attributes across individuals in the population (Hensher, 1981, as cited in Bhat, 1998). These models are based on the utility maximization hypothesis which assumes that an individual's route choice will reflect his underlying preferences for different combinations of attribute levels and the individual select the option with highest utility. The utility that an individual associates with his selection is comprised of two components: his socio-economic factors and the route characteristics which are heterogeneous due to observed (systematic heterogeneity) and unobserved (random heterogeneity) individual factors. Thus, he may have preference heterogeneity for the subjective evaluations of the utility of different attribute levels.

However, the tastes or characteristics of the population are assumed to be homogeneous in these models and a single WTP value is arrived for an imaginary average individual. This assumption may be too narrow as the evaluation may vary from person to person not only based on observable socio-economic characteristics, but also on unobservable characteristics. Daziano and Bolduc (2013) has shown that variables such as attitudes and perceptions which are not directly observable affects individual preferences and behaviour. Also, in Stated Choice (SC) data, each respondent is facing several choice scenarios with multiple responses which may result in heterogeneity of preference from the same individual because of varied interaction between his socio-economic characteristics and the attribute levels. So, instead of considering these heterogeneities, the assumption of homogeneity will result in biased, inconsistent parameters and estimates (Chamberlain, 1980). As a result, capturing the differences in the individual preferences and the associated differences in the sensitivity to changes in the attribute levels is essential for accurate estimations of the parameter values.

The possibility of incorporating the randomness of the parameters into the model helps the researchers in examining the effects of observed and unobserved heterogeneity of the individuals. Many studies have incorporated this heterogeneity into the means of the random parameters (e.g., Amador et al., 2005; Train, 2001). Advances in simulation techniques have enabled the use of complex models and Mixed Logit (ML) models have been proved to be such a useful tool to define wide behavioural patterns (Amador et al., 2005; McFadden and Train, 2000; Sillano and Ortúzar, 2005; Train, 2001).

Multiple choice situations in Stated Preference (SP) experiments for each respondent results in correlated responses across observations and it is a violation of the assumption of independence of observations. In ML models, this IIA property is relaxed and the correlation between different choice situations are allowed. ML model can determine the presence of heterogeneity of the parameters around the population mean through standard deviation of parameter distribution. ML model is also able to determine the possible source of heterogeneity, which is achieved through the interaction of random parameters with other variables in the model.

This paper, which is a part of a study to rank the road crash black spots for improvement in India by calculating crash costs through a SP experiment, attempts to examine the suitability of extending the scope and quality of ML model to account for the unobserved

heterogeneity in the preference parameters of an individual. Binary logic models were already tried to obtain the WTP of crash costs (Balakrishnan et al., 2015; Balakrishnan and Karuppanagounder, 2020; Karuppanagounder and Balakrishnan, 2016). Here, we formulate a ML based model of route selection to accommodate the preference heterogeneity in a comprehensive manner. Data for the study was collected through SP survey conducted in the northern part of the state of Kerala in India. The data includes socio-economic characteristics and the current trip characteristics (such as travel time, travel cost and crash cost rate) of the traveller. Various levels of the attributes travel time, travel cost and crash rate of the hypothetical route were generated based on the information on the current/recent trip. Further, to our knowledge, no previous study has conducted to account the heterogeneity of preference in route choice scenarios in India. This paper also evaluates what significant changes would the individual specific variations in the responsiveness will bring in the willingness of motorcyclists to pay to avoid involvement in road crashes.

The objectives of the study may be listed as follows:

- To identify existence of individual preference heterogeneity
- To evaluate the effect of the variation in the individual preference heterogeneity in WTP values

In the empirical part of the study, a basic Multinomial Logit (MNL) model, ML models with unconstrained and constrained random parameter distribution to identify the existence of preference heterogeneity and a model incorporating the heterogeneity are formed. We bring the triangular distribution promoted by Train (2001) and Green et al. (2006) also in addition to the normal distributions in the ML models to bring about a constrained nature for the distribution. Using the mean and standard deviation of the parameter values, further behavioural preferences are explained.

This paper is organized as follows. Second section presents a literature review on the previous studies done using ML model and the next section describes the theoretical explanations of the model and details on how the heterogeneity are identified followed by discussion on the estimated model results and finally a summary on our main conclusions.

## **2. Literature review**

Discrete models such as MNL based on Random Utility Theory were being popular for long time for the majority of practitioners (Domencich and McFadden, 1975). Basic assumption in these simple logit models is that the variances of the unobserved effects are the same for all alternatives. This concept of constant variances and zero covariances in these models results in a restrictive form. Recent interests (Brefle and Morey, 2000; Hess and Hensher, 2010) are to estimate the scale heterogeneity at the respondent level, even if it is a relatively old problem (Louviere et al., 2000). The main interest in using the ML model (they are also referred as Random Parameter Logit (RPL), Mixed Multinomial Logit (MMNL) or Kernel Logit (KL) model) instead of simpler logit model is to demonstrate the behavioural relevance of the individuals in the WTP values towards accident reduction. ML model gives a clear idea of the preference heterogeneity. In ML model the nature of heterogeneity of preferences for an attribute is explained through distribution of the marginal utilities/dis-utilities (Hensher et al., 2015). This model assumes that at least some of the parameters are random, following a certain probability distribution, which are assumed to be continuous over the sampled population. These distributions which, may take on different forms such as normal, lognormal, uniform or

triangular, are generally unrestricted over the domain but it can be constrained through additional assumptions (Hensher et al., 2005).

Many researchers have integrated the heterogeneity in responsiveness into their models (Bhat, 1998, 2000; Bhat and Zhao, 2002; Train, 2001). Bhat (1998, 2000) and Bhat and Zhao (2002) have treated the conditional variances of these distributions as constants, but Greene et al. (2006) explains the variance of the distribution to be a function of the individual-specific characteristics. Bhat mentions that if the variations in responsiveness across individuals were not accounted adequately, it will result in statistically inferior data fit and inappropriate evaluations of policy actions.

A traditional way of testing the heterogeneity of preferences has been to examine the interactions between attributes of the alternative and individual socio-economic characteristics such as age, sex, income level, number of dependants, distance travelled etc. (Revelt and Train, 1998). Here the interaction is observed through the variation in the parameter values with respect to the socio-econometric characteristics. But examination of this interaction may not be easy always as there may be individual random behaviour which cannot be observed or are difficult to measure. In such cases, the parameters can be specified as random parameters that vary randomly with personal tastes. If the parameter is considered as fixed and the error terms follow independent and identically distributed (IID) Gumbel distribution we will get the MNL model as described by Domencich and McFadden (1975). Instead on the other hand if we assume a randomness to the parameters, we will get a ML random parameter model (Train, 2003).

Mixed models can include the random effects because of the clustering factor associated with the repeated measurements as in SP experiments, to account for the influence of correlations. ML models, ML models with heterogeneity in means, ML models with heterogeneity in means and variances have all been successfully used in various studies to account for the significant effect of unobserved heterogeneity from various sources (Anastasopoulos and Mannering, 2011; Behnood and Fred, 2017). Behnood and Fred (2017) used ML model with heterogeneity in means and variances to account for the unobserved heterogeneity of the random parameters in their study to explore the possible factors that contribute the bicyclist injury severity.

### **3. Methodology**

The methodology used was a SP questionnaire-based survey, which was formulated as a binary route choice experiment. Attributes such as travel time (minutes), travel cost (rupees) and road crashes per year (numbers) were included to decide upon the choice sets. A fractional factorial design based on differences similar to the set up made by Rizzi and Ortúzar (2003) was used. The experimental design is based on attribute differences, which is what matters when estimating discrete choice models. Hypothetical route choices were designed, with travel times, costs and number of road crashes constructed as some percentage above or below those of the recent/regular trip. Each choice set included two alternatives and three levels of the three attributes leading to 27 possible combinations. In order to reduce the overburden of the respondents the 27 combinations were split into three blocks and each respondent was presented with a randomly selected block (Hensher et al., 2009; Rizzi and Ortúzar, 2003). Respondents were asked to describe their regular trip or a recent trip. After detailed explanation of different risk scenarios in comparison with that of their presently used routes, respondents were asked to choose among the hypothetical routes in terms of travel time, travel cost and number of accidents. The differences in attribute levels may not be directly meaningful to the

respondents and so the difference levels were attached to their recent/regular trip (i.e. the reference values were given directly by the respondents). The alternatives were formed by adding the difference levels to these reference values. Each sampled respondent completed 6 choice tasks under different attribute levels.

The attributes chosen with their levels used in the survey (Table 1) were identified through extensive literature reviews. For existing route, the attribute levels were pivoted off the attribute levels captured from the recent trip of the respondent. For the attributes ‘travel time’ and ‘travel cost’ the levels given are one lower level than the base and one higher level than the base level. The exception was for the attributed ‘crash rate’, for which both the levels were shown as reduction from the base level. Respondents were asked to compare present alternative that they used with the choices given (A sample choice set is given in Table 2).

Table 1: Attributes and Levels.

	<i>Level 1</i>	<i>Level 2</i>	<i>Level 3</i>
Travel time	Reduced by 25%	Base value	Increased by 25%
Travel cost	Reduced by 25%	Base value	Increased by 25%
Crash rate	Reduced by 25%	Base value	Reduced by 50%

Table 2: Choice Set 1.

	<i>Presently used mode</i>	<i>Alternatives</i>
Travel time (minutes)	TT	1.25TT
Travel cost (Rs)	TC	1.00TC
Crash rate	CR	0.75CR

### 3.1. Mixed Logit (ML) model and estimation of preference heterogeneity.

Based on the Random Utility Theory (Domencich and McFadden, 1975; Ortúzar and Luis, 2011) the utility  $U_{nsj}$  of the alternative  $J$  is obtained as

$$U_{nsj} = V_{nsj} + \varepsilon_{nsj} \quad (1)$$

where  $V_{nsj}$  is the indirect utility function on the alternative  $J$  for the individual  $n$ , which can be measured by the analyst, in each of  $s$  choice situations and  $\varepsilon_{nsj}$  is a stochastic component that reflects the non-measurable factors (Hensher et al., 2005). The deterministic component  $V_{nsj}$  is a function of explanatory variables  $x_{nsjk}$  (attributes of the alternatives, socio-economic characteristics of the individual, etc) observed by the analyst and  $\beta_{nk}$ , the marginal utility or parameter weight associated with the attribute  $K$  for the respondent  $n$ , which is usually considered constant for all individuals.

So, the associated utility can be expressed then as

$$U_{nsj} = \sum_{k=1}^K \beta_{nk} x_{nsjk} + \varepsilon_{nsj} \quad (2)$$

The unobserved components,  $\varepsilon_{nsj}$  is IID extreme value Type 1 across individuals, alternatives, and choice situations. This IID assumption is very restrictive and it does not allow for the error components of different alternatives to be correlated.

In an SP survey, each respondent is facing several choice scenarios, providing multiple responses which results in increased complexity of the modelling. So, it is required to account for the correlation among choices from the same individuals. A conventional form of reflecting heterogeneity of preferences is through introduction of interaction between attributes and characteristics of the individual such as age, personal income, etc.

(Ortúzar and Luis, 2011; Revelt and Train, 1998). This is achieved by making  $\beta_{nk}$  a function of individual's observed socio-economic characteristics, through introduction of additional stochastic elements that may be heteroskedastic and correlated across alternatives. Thus, the choice probabilities of the model can be summarized as (Hensher et al., 2015)

$$Prob(choice_{ns}=j|x_{nsj}, z_n, v_n) = \frac{\exp(V_{nsj})}{\sum_{j=1}^{J_{ns}} \exp(V_{nsj})} \quad (3)$$

where

$$V_{nsj} = \beta_n x_{nsj}$$

$$\beta_n = \beta + \Delta z_n + \Gamma v_n$$

$$\beta = \text{constant}$$

$\Delta z_n$  = term reflecting the observed heterogeneity

$\Gamma v_n$  = term representing the unobserved heterogeneity

$x_{nsj}$  = the  $K$  attributes of alternative  $j$  in choice situation  $s$  faced by individual  $n$ ,

$z_n$  = a set of  $M$  characteristics of individual  $n$  that influence the mean of the taste parameters

$v_n$  = a vector of  $K$  random variables with zero means and known (usually unit) variances and zero covariances and

$J_{ns}$  = total choices available to the individual  $n$  in choice situation  $s$ .

The ML class of models assumes a general distribution for  $\beta_{nk}$  and an IID extreme value Type 1 distribution for  $\varepsilon_{nsj}$ .  $\beta_{nk}$  can take on different distributional forms such as normal, lognormal, uniform, triangular, etc.

The popularity of ML model is on its ability to incorporate randomness of the parameters and its possibility to account for the heterogeneity around the mean of the distribution. Additionally, further behavioural information connected with the variance of the random parameter distribution can also be easily evaluated using these models.

Many literatures indicate that simple models like MNL fails to give correct estimates in the presence of unobserved heterogeneity. For examples, Bhut (1998) mention that ignoring the preference heterogeneity or response heterogeneity will lead to biased parameter values, elasticities and WTP values. He observed larger WTP values and lower elasticities in the MNL model than in the ML model. At the same time in another study Bhat (2000) found that MNL severely underestimated the WTP for travel time and over estimated the cost elasticities compared with his ML model. Train (1998) experienced that a general conclusion cannot be made whether the MNL model gives a good estimate for the WTP or whether it is stable for each data set. Contradict to these findings, Horowitz (1980) states that ignoring heterogeneity will not lead to a biased WTP, but at the same time the choice probabilities may get biased.

In a study, Amador et al (2005) examined the presence of preference heterogeneity using ML models and found that the benefit measures (subjective value of travel time), both at the individual and at the social level are sensitive to preference heterogeneity assumptions. They found that the subjective values and the welfare changes are underestimated if taste homogeneity is assumed and variations up to 30% is detected in their results. Hensher (2001) also concludes that MNL models underestimate the value of time and Lapparent et al. (2013) states that there is an improved representation of the observed choices when the unobserved taste heterogeneity is taken into account. In a SP experiment of choice of rail season cards conducted in Dutch, van den Berg et al. (2008) found that MNL model underestimates the price sensitivities of the respondents and their WTP for reduction in the restrictions. They concluded that the mixed logit model gives

the most satisfactory results as it controls for the unobserved response heterogeneity. In their study to investigate the preferences of inhabitants for the future of an old port in Trieste, Massiani and Rosato (2008) clearly indicates that ML estimates provided more convincing results while exploring the heterogeneity among interviewees' preferences. Eboli and Mazzulla (2008) also reports advantages of ML model over MNL model in their study to find out WTP for improvement in the service quality of bus services.

Another important concern is about the number of repetitions or draws to be done in the model to get a stable simulation. Traditional pseudo-random draws used in ML models requires a large number of repetitions to get an accurate result (Hensher et al. 2005). In this scenario Halton draws which are based on number theory and possessing a more uniform spread are being proposed by researchers (Bhat et al., 2000). Halton draws gives a more stable estimation with lesser number of repetitions. We used LIMDEP/NLOGIT to estimate the models in this paper, and for our final mixed logit estimation, 500 Halton draws were made (Hensher et al., 2005, 2015).

#### **4. Experimental Methods**

The above research activity is carried out using data collected through a SP survey conducted among two-wheeler riders. Data for this study were grouped as socio-economic data, travel data and road accident data based on earlier research works (Andersson and Lundborg, 2007; Bhattacharya et al., 2007). Traffic crash-related data was taken from police records and person to person interview were executed for collecting other data. Data was collected across the state of Kerala in India. Traffic accident rates per year at various locations were arrived from the FIR reports and were used while describing the accident scenario to the respondents. Socio economic and travel related aspects such as number of dependents in the family, marital status, personal monthly income, age of the respondent, mode of travel, etc, which proved to be affecting the decision-making process for WTP, were included in the questionnaires (Andersson and Lundborg, 2007; Bhattacharya et al., 2007).

Choices made by each individual, together with the values of each attribute in each choice situation, were the SP data used in the estimation of the models. These attributes were decided upon after a pilot study to check the apprehension level of the users on these various attributes and their levels.

In the face to face interview the respondents were handed over a hard copy of the questionnaire to fill it according to the instructions. The interviewer helped the respondent in answering the questions and provided clarifications. Stratified sampling method was used for sampling. A total of 1000 two-wheeler drivers were approached for data collection, but only 867 were ready to give the data (86.70%). From this, 48 responses were removed for being incomplete. Finally, 819 useful samples were obtained. On an average 30 minutes were taken for the survey per person. Respondents were not given any remuneration for participation.

The concept of hypothetical alternatives was conveyed through illustrating real-life cases such as flyovers, modernised roads, improvised intersections etc. The collected data was analysed using spreadsheets and labelling was done to all the variables.

##### *4.1 Data description*

Socio-economic details, trip characteristics, and road crash rates of the location were obtained through the preliminary analysis of the collected data. The annual average

accident rates obtained from Police reports for the selected roads are varied from 13 to 240 for the different road stretches under study.

The two-wheeler respondents mainly were aged between 20 and 40 years (54%) representing the younger age group. About 76% of two-wheeler users declared that they are married. Nearly 56% of the two-wheeler riders claimed that they live in urban area. A great majority, 78%, said that they are highly worried about the road accidents and feel insecure about the road accidents.

Almost 53% of the two-wheeler riders had past experiences of road accident either directly or indirectly (for family or friends). The minimum, maximum and average household sizes for the sample are 1.0, 8.00 and 5.2. About 36% of the riders have monthly personal income between Rs. 10,000 and 20,000. Only 0.25% of the riders claimed a monthly income above Rs. 2 lakhs. The monthly household income category for 37% of two-wheeler users was Rs. 20,000 – Rs. 40,000. It can also be noted that 24% of the two-wheeler users are graduates or postgraduates. A significant percentage (about 31%) two-wheeler users are government employed. Another important observation was that a large portion of about 66% of two-wheeler users felt that the present roads are unsafe. The minimum, maximum and average numbers of dependents for two-wheeler users are 0.0, 5.0 and 1.5, respectively.

## **5. Results and discussions**

Detailed discussions on the various model analysis is given in the following sections.

### *5.1 Model analysis*

Table 3 presents the various model estimation results based on random parameters approach and heterogeneity in means and variances. Four models are formulated for the data. The first one is a an MNL model and the remaining three are ML models. These models examine the observed and unobserved effects of various Socio-Demographic Characteristics (SDCs) on the route choice. The route choice selected by the respondent was taken as the dependent variable. Final specification of each model is arrived through the systematic elimination of the insignificant variables.

The MNL model is a basic model with no consideration on the randomness and heterogeneity of the parameter. In the ML models, randomness of the parameters is taken into account; models A & B do not allow for the heterogeneity while model C does. Model A is a mixed logit with normal distribution (unconstrained distribution) for random parameters estimates and for Model B the parameters were drawn from constrained triangular distributions (Greene et al., 2006). In all these three ML models, the three route choice attributes (the travel cost, travel time and number of traffic crashes) were found to produce random parameters with heterogeneity through 500 Halton draws.

All the parameter estimates shown in the table are of expected sign and statistically significant at the 95 percent level of confidence. To keep the length of paper reasonable, we present and discuss only those variables that were found to be statically significant. The derived standard deviations of the distribution of the random parameters obtained in the Model B and C allow for testing statistical significance between these models.

All the negative coefficients indicate that an increase in these variables reduce the utility. It shows that the female and unmarried participants prefer the existing route. As age increase may be due to more family responsibilities, they prefer the safer routes.



Education level is significant in the model A only and is showing a disutility to the existing route as education level increase.

Table 3: Descriptive Statistics of the Significant Variables in the Models.

Variable	MNL		ML (Model A) unconstrained		ML (Model B) constrained t		ML (Model C) heterogeneity	
	Coeff.	t-ratio	Coeff.	t-ratio	Coeff.	t-ratio	Coeff.	t-ratio
<i>Random parameters in utility functions (trip attributes)</i>								
Travel time	-0.018	-3.259	-0.043	-3.237	-0.020	-3.493	-0.223	-5.284
Travel cost	-0.014	-4.080	-0.027	-3.431	-0.012	-3.818	-0.024	-1.950
Crash rate	-0.002	-2.129	-0.005	-5.103	-0.002	-2.093	-0.017	-2.261
<i>Non-random parameters in the utility function</i>								
Gender	0.724	7.233	1.962	6.313	0.724	7.225	2.573	9.168
Marital status	0.326	2.045	1.603	3.475	0.338	2.109	1.981	4.402
Age	-0.012	-3.780	-0.021	-2.582	-0.012	-3.769	-0.008	-0.946
Education			-0.132	-2.591				
Occupation	-0.264	-11.53	-0.422	-7.118	-0.264	-11.51	-0.252	-5.045
Vehicle ownership	-0.042	-2.232	-0.125	-2.725	-0.041	-2.193	-0.169	-3.739
Crash history	-0.290	-7.171	-0.739	-5.727	-0.290	-7.152	-0.848	-7.406
Total dependants	-0.235	-6.531	-0.580	-5.818	-0.236	-6.543	-0.603	-6.461
Household income	-0.307	-9.585	-0.781	-7.129	-0.307	-9.535	-0.505	-6.307
Distance travelled	-0.037	-7.288	-0.093	-5.157	-0.037	-7.251	-0.032	-2.709
<i>Derived standard deviations of parameter distributions</i>								
Travel time			0.152	4.574	0.020	3.493	0.223	5.284
Travel cost			0.000	0.041	0.012	3.818	0.024	1.960
Crash rate			0.054	5.882	0.002	2.093	0.017	2.261

The first ML model, model A, is formed with 3 random parameters with unconstrained normal distribution to provides a good reference point to consider the alternative distributions (Hensher et al., 2015). The first and third part of the output in the ML models provides information on the random parameters estimated. Random parameters estimated within the most basic ML models are generally derived at the sampled population level only and are called unconditional parameter estimates as they are not conditioned on any particular individual's choice pattern but rather on the sample population as a whole.

The first series of output indicates whether the mean of the sample population random parameter is statistically different to zero. For example, for the travel cost attribute, the parameter estimates of -0.027 represents the mean of the 500 Halton draws. The p values of all the variables listed in the table are less than 0.05 indicating the mean parameter estimates given are statistically different to zero at the sample population level. While the first set of output relates to the means of the random parameters, the third set of the output indicates the amount of dispersion existing around the sample population. Insignificant parameter estimates in this output indicate that the dispersion around the mean is statistically equal to zero and that all information in the distribution is captured within the mean itself. On the other hand, statistically significant parameter estimates in the derived standard deviations indicates the existence of heterogeneity in the parameter estimates over the sampled population around the mean. It means that each individual possesses his own parameter estimates which will be different from the population mean values.

As per this explanation, the dispersion of the parameters travel time and crash rate are statistically significant with Wald statistics greater than the critical value. This means that all the respondents within the sample cannot be assigned by the parameters obtained in the first series of output. At the same time, dispersion of the parameter travel cost is statistically insignificant (Wald statistic less than the critical value of 1.96) and so this travel cost parameter should collapse to a single point estimate for the entire sampled population. This indicates that no heterogeneity exists over the population with respect to the parameter travel cost, at individual level.

At this point, to check the consistency of this behaviour, distribution of this parameter is re-estimated with constrained distributions imposed on all random parameters. We chose a constrained triangular distribution (with equal mean and standard deviation) which is increasingly being used in empirical studies to re-estimate the randomness of parameters (Hensher et al., 2015; Greene et al., 2006; Train, 2001). In the table we can see that the only difference is at the distribution of random parameters and the derived standard deviation for each parameter is exactly the same as in the mean estimate obtained in the first part of the output.

We can see from that all the three variance heterogeneities are statistically significant at 95% level indicating presence of heterogeneity in preferences of the individuals. Differences in the random parameter estimates among the models A & B are due to the different distributions employed in the estimation of the models.

In model C, which is an extension of the model B, the presence of preference heterogeneity around the mean is determined through the interaction of each random parameter with the variables which may have preference heterogeneity. The resulting interaction effects are shown in the Table 4. In this model every random parameter is conditioned on all the significant variables.

Differences in the model parameter values among the various models necessitates formal statistical tests to check the fit of the models. The significance of the models can be checked through the p values and likelihood values, and likelihood ratio test are used to find a better model or compare different models. Table 5 gives a comparison of the log likelihood values and the *pseudo R<sup>2</sup>* values and Table 6 shows values of the log likelihood ratio test for different models. The *p* values for the *Chi square* statistics for all the models are 0.0000 which is less than 0.05 indicate an overall model fit and so we conclude that these models are significant at 95% of confidence level.

Table 4: Interaction Effects Between the Random Parameter Estimates and Covariates.

<i>Variables</i>	<i>Travel Time</i>		<i>Travel Cost</i>		<i>Crash Rate</i>	
	<i>Coeff.</i>	<i>t-ratio</i>	<i>Coeff.</i>	<i>t-ratio</i>	<i>Coeff.</i>	<i>t-ratio</i>
Gender	-0.045	-2.861	-0.024	-2.515	0.022	6.610
Marital status	0.011	0.442	0.028	2.146	0.015	3.273
Age	-0.001	-1.508	-0.001	-1.334	-0.000	-1.840
Occupation	-0.014	-3.945	-0.004	-1.315	-0.001	-2.300
Vehicle ownership	-0.012	-3.161	-0.001	-0.266	0.001	2.734
Crash history	0.005	0.752	0.005	1.338	0.006	5.210
Total dependants	-0.016	-2.943	-0.012	-3.022	-0.004	-4.129
Household income	0.004	0.823	0.008	2.816	0.002	2.709
Distance travelled	-0.001	-2.067	0.001	1.989	-0.001	-8.877

Table 5: Measures of Model Fit.

Summary Statistics	MNL		ML (Model A) unconstrained		ML (Model B) constrained t		ML (Model C) heterogeneity	
Log-likelihood value at convergence	-2010.187	-1987.670	-2010.860	-1887.918	-2010.187	-1987.670	-2010.860	-1887.918
Number of parameters	12	17	13	40	12	17	13	40
Chi squared at p=0.0000	417.384	2836.911	2790.530	3036.414	417.384	2836.911	2790.530	3036.414
Pseudo R-squared	0.09405	0.41644	0.40963	0.4457	0.09405	0.41644	0.40963	0.4457

Table 6: Likelihood Ratio Test ( $\chi^2$  Values in the Bracket).

LLRT between	$\chi^2$ (Observed test statistic)	Degree of freedom	$\chi^2$ (Critical at 95% confidence level)
MLN – ML (A)	45.03	5	11.07
MLN – ML (C)	244.54	28	41.34
ML (A) – ML (C)	245.88	23	40.11
ML (B) – ML (C)	199.54	27	35.17

From the log likelihood function values, we may see that the constrained Model B with a value -2010.187 is not as good as the unconstrained Model A with a value -1987.67. But this is a comparison of different distributions and for an unconstrained triangular distribution the value obtained is -2009.243 which is almost equal to the value mentioned above (not reported here). With lowest log-likelihood value, Model C shows a better fit than all other models.

The overall model fit in terms of *pseudo-R*<sup>2</sup> index, analogue to the multiple correlation coefficients, provides a practical and theoretically sound index of the goodness of fit (Domencich and McFadden, 1975; Louviere et al., 2000). McFadden gives values of 0.2 to 0.4 for the *pseudo R*<sup>2</sup>, for an excellent fit. Here, all the ML models perform very well with very good value of *pseudo R*<sup>2</sup> index than the MNL model.

The likelihood ratio test between two models is done using the formula  $-2*(LL_{largest} - LL_{smallest})$ . This value is compared to critical  $\chi^2$  values with degree of freedom equals the difference between the number of parameters estimated by the two models. Based on these values we can state that the log likelihood of the model C incorporating the heterogeneity correlation, is statistically closer to zero than that of the other models and so we can accept that this model statistically improved the LL over the previous models and this model has a good fit (Hensher et al., 2005). ML models shows superior performance than MNL model and the ML model C outperform the ML models A & B.

Thus, we can see that the model C with the heterogeneity interaction out-performs the models A & B which is consistent with the earlier findings where, substantial improvements in model fit were experienced when unobserved heterogeneity had incorporated in the models (Amador et al., 2005; Behnood and Fred, 2017; Revelt and Train, 1998).

The interaction terms obtained between the random variables and the other variables in the model C explains as to why the heterogeneity may exist. A statistically significant interaction term indicates that the differences in marginal utilities held for that particular random attribute may be explained by the differences in the levels of the interacting variable.

From the interaction matrix (Table 4) we can see that the variables marital status, age, crash history and household income do not explain preference heterogeneity in the Travel Time random parameter. However, other variables listed in the table are statistically significant (t-ratio greater than 1.96 at 95% confidence level) in the case of Travel Time random parameter. The heterogeneity in the mean parameter estimate for the *travel-time x gender* of -0.045 indicates that sensitivity of male respondents increases with the travel time. Or in other words male respondents prefer routes of lesser travel time. Similarly, we can see that unmarried persons, respondents those who have more crash history and more household income are less sensible to travel time. But, on the other hand age, businessman, persons own individual vehicles, respondents with more dependants at home and those who travel lot shows more sensitivity to the travel time. These indications seem to be logic.

Similar behaviour is shown towards the travel cost and crash rate variations except in the case of distance travelled for travel cost and gender and vehicle ownership for crash rate. We can see that individuals with larger trip distances are more sensitive to travel cost. The positive gender response towards crash rate suggests that men are not much bothered about the crash rate. Similarly, individuals who own more vehicles are less sensitive to the crash rate.

The ratio of parameter of non-monitory attribute (number of crashes per year) and monitory attribute (travel cost), keeping all other parameters constant, gives the WTP. WTP values against reduction of the risk of being injured or killed in the road crashes are shown in Table 7. Comparing to an earlier finding (Balakrishnan et al., 2015; Balakrishnan and Karuppanagounder, 2020; Karuppanagounder and Balakrishnan, 2016) the WTP value in the model C shows an increment of about 35% of the value (Rs. 0.527 per crash) obtained in a binary logit model. These results show that the WTP values derived from a model with homogeneous preferences and a model with heterogeneity in preferences shows significant differences. This suggests that a restrictive specification in the models will underestimate the values of WTP (Hensher, 2001). One possible reason for this huge change in the values may be the re-scaling of the parameters when they are allowed to vary randomly (Sillano and Ortúzar, 2005). As the random parameter model explains the un-observed characteristics too, the error component may be smaller and so it would have contributed in re-scaling of the parameter values. This is analogous to the explanation given by Horowitz (1981) where the parameters are reconstructed to accommodate the extra explanation added by the random parameter. Thus, based on the variables included, variables taken as random and the distribution selected for the random parameters, the utility function may change and thus will lead to change in the estimate of values. These findings show the importance of consideration of heterogeneity in the variance of unobserved effects.

Table 7: Willingness to Pay Values Against Crash Reduction.

<i>LLRT between</i>	<i>MNL</i>	<i>ML Model A</i>	<i>ML Model B</i>	<i>ML Model C</i>
WTP (Rs/no. of crash)	0.143	0.185	0.167	0.708

## 6. Conclusion

Various new analyses are being brought forward to identify sources of heterogeneity associated with the mean of population parameters and the variances of random parameter distributions. Heterogeneity in responsiveness to different levels of parameters is the

focus of this paper. We compared different discrete choice models to identify the presence of heterogeneity of preferences in the sample. The ML models were applied to estimate this heterogeneity in responsiveness to the different route choices in simple and comprehensive manner. The parameter distributions chosen by the analyst helps in providing an additional dimension to closely track the unobserved heterogeneity compared to the standard mixed logit models and the addition of this heterogeneity into the model provides additional flexibility for capturing the unobserved variations in the decision-making process. This study reveals that accommodating the heterogeneity of preference in the individual responsiveness improves the overall model fit and provides some new information.

It was found in this study that travel cost and number of road crashes of the route selected, age, distance travelled and household income have significant effect on the decision-making process. Different models were formed to check the existence of heterogeneity in preference based on attributes of the route. From the basic random parameter model, the t-test values for the spread show that the parameters travel time and crash rate exhibit significant heterogeneity over the population and no heterogeneity in the parameter travel cost. Thus, the basic MNL model is rejected in favour of this random parameter (ML) model based on the likelihood ratio test. The consistency of this finding is checked through a constrained model and it is found that travel time and crash rate continue to be heterogenous and the travel cost again shown heterogenous behaviour. As the presence of heterogeneity is identified, the significance of existence of any interaction effects between these random parameters and significant socio-economic variables were examined through another model incorporating the interaction of random parameter heterogeneity. The model results suggest that the individual decision making is sensitive to the assumptions about preference heterogeneity and the traditional assumption of taste homogeneity is found to underestimate the subjective values. Our findings regarding the effect of unobserved heterogeneity on the performance of various models are comparable to the findings of Bhat (1998, 2000).

This paper is a part of a work to investigate the WTP values against road traffic accident in the context of a developing country, India, which is a first attempt of this kind. Findings in this particular paper emphasize the importance of incorporating the heterogeneity in the means and variance of parameters in the choice-based experiments. The results reveal that inclusion of heterogeneity of the random parameters improves the model fit and provides new insights into the effect of socio-economic variables on the decision-making process. The methodology used in this study provides a framework for the understanding of how the heterogeneity in responsiveness in stated choice experiments affect the decision-making process and can offer better answers to the prediction of people behaviour in the specific field of choice among road crash reduction alternatives. It also indicates the flexibility achieved in observing the heterogeneity in responsiveness compared to the traditional methods. The distributional assumptions and the approach of capturing the heterogeneity in means and variances improve the model a lot. Identification of the parameters that influence the choice decision and how they affect the WTP values, attempted in this research, could be useful for assisting in planning/implementation of appropriate traffic safety programs in future. The model proposed in this study was from a specific region of the country and so the existence of people behaviour referred in this study should be carefully considered before applying these findings in local, regional, national level projects of other geographic locations and time periods. In future research, the evidence herein can be disaggregated by road type,

distance, etc., to provide inputs for various types of analysis like benefit-cost analysis, in contrast to the aggregated findings presented herein.

### *References*

- Amador, J. F., González, R. M., Ortúzar, J. de. D. (2005) “Preference Heterogeneity and Willingness to Pay for Travel Time Savings”, *Transportation*, 32 (6), pp. 627–647.
- Anastasopoulos, P. Ch., Mannering, F. L. (2011) “An Empirical Assessment of Fixed and Random Parameter Logit Models Using Crash-And Non-Crash-Specific Injury Data”, *Accident Analysis and Prevention*, 43 (3), pp. 1140-1147.
- Andersson, H., Lundborg, L. (2007). “Perception of Own Death Risk, an Analysis of Road-Traffic and Overall Mortality Risks”, *Journal of Risk and Uncertainty*, 34, pp. 67–84.
- Balakrishnan, S., Ashok, G., and Karuppanagounder, K. (2015) “Estimation of Road Accident Cost in India Through Willingness to Pay Approach”, *Proceedings from the 3<sup>rd</sup> Conference of Transportation Research Group of India (3rd CTRG)*, Kolkata, India.
- Balakrishnan, S., Karuppanagounder, K. (2020) “Estimating the Cost of Two-Wheeler Road Accident Injuries in India Using the Willingness to Pay Method”, *Australian Journal of Civil Engineering*, 18 (1), pp. 65-72.
- Behnood, A., Mannering, F. (2017) “Determinants of Bicyclist Injury Severities in Bicycle-Vehicle Crashes: A Random Parameters Approach with Heterogeneity in Means and Variances”, *Analytic Methods in Accident Research*, 16, pp. 35–47.
- Bhat, C. R. (1998) “Accommodating Variations in Responsiveness to Level-of-Service Variables in Travel Mode Choice Modelling”, *Transportation Research Part A: Policy and Practice*, 32 (7), pp. 495–507.
- Bhat, C. R. (2000) “Incorporating Observed and Unobserved Heterogeneity in Urban Work Travel Mode Choice Modelling”, *Transportation Science*, 34 (2), pp. 228–238.
- Bhat, C. R., Eluru, N., and Copperman, R. B. (2000) “Flexible Model Structures for Discrete Choice Analysis”, In: Hensher, D. A., Button, K. J. (eds) *Handbook of Transport Modelling*, Elsevier Science, Oxford.
- Bhat, C., Zhao, H. (2002) “The Spatial Analysis of Activity Stop Generation”, *Transportation Research Part B: Methodological*, 36 (6), pp. 557–575.
- Bhattacharya, S., Alberini, A., Cropper, M. L. (2007) “The Value of Mortality Risk Reductions in Delhi, India”, *Journal of Risk and Uncertainty*, 34 (1), pp. 21–47.
- Breffle, W. S., Morey, E. R. (2000) “Investigating Preference Heterogeneity in a Repeated Discrete-Choice Recreation Demand Model of Atlantic Salmon Fishing”, *Marine Resource Economics*, 15, pp. 1–20.
- Chamberlain, G. (1980) “Analysis of Covariance with Qualitative Data”, *The Review of Economic Studies*, 47 (1), pp. 225–238.
- Daziano, R. A., Bolduc, D. (2013) “Incorporating Pro-Environmental Preferences Towards Green Automobile Technologies Through a Bayesian Hybrid Choice Model”, *Transportmetrica A: Transport Science*, 9 (1), pp. 74–106.
- Domencich, T. A., and McFadden, D. (1975) *Urban Travel Demand: a Behavioral Analysis*, North-Holland Publishing Company.
- Eboli. L., Mazzulla, G. (2008) “Willingness-to-Pay of Public Transport Users for Improvement in Service Quality”, *European Transport*, 38, pp. 107-118.
- Greene, W. H., Hensher, D. A., Rose, J. (2006) “Accounting for Heterogeneity in the Variance of Unobserved Effects in Mixed Logit Models”, *Transportation Research Part B: Methodological*, 40 (1), pp. 75–92.

- Hensher, D. A. (2001) "Measurement of the Valuation of Travel Time Savings. *Journal of Transport Economics and Policy*, 35, pp. 71–98.
- Hensher, D. A., Rose, J. M., Ortúzar, J. de D., Rizzi, L. I. (2009) "Estimating the Willingness to Pay and Value of Risk Reduction for Car Occupants in the Road Environment", *Transportation Research Part A: Policy and Practice*, 43 (7), pp. 692–707.
- Hensher, D. A., Rose, J. M., and Greene, W. H. (2005) *Applied Choice Analysis: A Primer*, Cambridge University Press.
- Hensher, D. A., Rose, J. M., and Greene, W. H. (2015) *Applied Choice Analysis*, 2<sup>nd</sup> ed. Cambridge University Press.
- Hensher, D. A., Louviere, J., Swait, J. (1998) "Combining Sources of Preference Data", *Journal of Econometrics*, 89 (1–2), pp. 197–221.
- Hess, S., Hensher, D. A. (2010) "Using Conditioning on Observed Choices to Retrieve Individual-Specific Attribute Processing Strategies", *Transportation Research Part B: Methodological*, 44 (6), pp. 781–790.
- Horowitz, J. L. (1980) "The Accuracy of The Multinomial Logit Model as an Approximation to the Multinomial Probit Model of Travel Demand", *Transportation Research Part B*, 14, pp. 331-342.
- Horowitz, J. L. (1981) "Sampling Error, Specification and Data Errors in Probabilistic Discrete Choice Models", In: *Applied Discrete Choice Modelling*, Croom Helm.
- Karuppanagounder, K., Balakrishnan, S. (2016) "Estimation of Willingness to Pay Values for Road Traffic Crash Risk Reduction in India", *Proceedings from the TRB 95<sup>th</sup> Annual Meeting Compendium of Papers*.
- Lapparent, M. De., Axhausen, K. W., Frei, A. (2013) "Long Distance Mode Choice and Distributions of Values of Travel Time Savings in Three European Countries", *European Transport*, 53.
- Louviere, J., Hensher, D. A., and Swait, J. (2000) *Stated Choice Methods: Analysis and Applications*, Cambridge University Press.
- Massiani, J., Rosato, P. (2008) "Using Conjoint Analysis to Investigate Preferences of Inhabitants for the Future of a Greyfield Area: An Application to the Old Port in Trieste", *European Transport*, 39, pp. 59-81
- McFadden, D., Train, K. (2000) "Mixed MNL Models for Discrete Response", *Journal of Applied Econometrics*. 15 (5), pp. 447–470.
- Ortúzar, J. de D., and Luis, G. W. (2011) *Modelling Transport*, John Wiley and Sons.
- Revelt, D., Train, K. (1998) "Mixed Logit with Repeated Choices: Households' Choices of Appliance Efficiency Level", *Review of Economics and Statistics*, 80 (4), pp. 647–657.
- Rizzi, L. I., Ortúzar, J. de D. (2003) "Stated Preference in the Valuation of Interurban Road Safety, Accident Analysis and Prevention", 35, pp. 9–22.
- Sillano, M., Ortúzar, J. de D. (2005) "Willingness-to-Pay Estimation with Mixed Logit Models: Some New Evidence", *Environment and Planning A*, 37 (3), pp. 525–550.
- Train, K. E. (1998). *Recreation Demand Models with Taste Differences Over People*", *Land Economics*, 74 (2), pp. 230-239.
- Train, K. E. (2001) *A Comparison of Hierarchical Bayes and Maximum Simulated Likelihood for Mixed Logit*, Cambridge University Press.
- Van den Berg, V., Kroes, E., Verhoef, E. T. (2008) "Choice of Season Cards in Public Transport: A Study of a Stated Preference Experiment", *European Transport*, 40, pp. 4-32.

*Acknowledgements*

The authors extend sincere gratitude to the Centre for Transportation Research, Department of Civil Engineering, National Institute of Technology Calicut for providing necessary funding to carry out the research work. Also, the help rendered by Ms Geethu Ashok, Post Graduate student of the Traffic and Transportation Division of the institute during the initial formulation of the work is greatly acknowledged.