CHAPTER 4

4.1 INTRODUCTION

As observed from the previous chapter, the stochastic disaggregate models of travel demand seemed to be the best feasible option for depicting travel-demand in a behavioral format. However, the transport survey so far conducted for the two selected small and medium cities of West Bengal revealed application of the policy insensitive aggregate models for estimating travel demand. Hence, the application of the stochastic disaggregate models for these two cities seemed a justifiable venture. Following such trend, a specific model of the said category is considered for application in the two selected cities to have a more behavioral and policy-sensitive travel demand estimation of them. Before proceeding further, the ground for the usage of a specific model for the two cities along with their field of application needs to be highlighted. The following part is an effort in that direction.

4.2 MAPPING THE DOMAIN OF THE ATTEMPTED TRAVEL DEMAND ESTIMATION

As can be grasped from the proceeding pages, the framework for travel demand estimation is a complex procedure involving interactions among innumerable factors. For example, trips to a particular area is strongly influenced by the employment opportunity of that region, or the acquisition of the means of transportation to a particular destination depend among other things on the availability of transportation facilities and so on. The complexity in the simultaneous assessment of these factors may lead to a position of intractability. Hence, demarcating the domain of analysis, to serve the purpose of narrowing the negative externalities, prevailing in an urban transport network, can be a worthy effort.

A major concern in urban travel analysis is the prediction of link flows i.e. mode choice, on the available transportation infrastructure, which serves as a benchmark for transport policy planners to assess the level of deviation from the norms by the policies pursued. A new set of policy prescriptions to ward-off the ill effects associated with such prediction can then be thought of. The focal position, mode-choice determination play in urban travel demand estimation can be gauged from the very high majority of models developed in the disaggregate set-up satisfying that end.
4.3 THE IMPORTANCE OF MODE CHOICE MODELS IN DISAGGREGATE TRAVEL DEMAND ESTIMATION

The initial application of disaggregate modeling techniques in transportation was made for the choice of travel mode by Warner in 1962, Lisco in 1967, Lave in 1969, McGillivray in 1972, Peat, Marwick and Mitchell in 1973. A considerable number of researchers also investigated the performance of mode-choice models for work-trips. Works by Atherton and Ben Akiva (1976), Ben Akiva and Richards (1975), Parody and Train (1976), Richards (1975), Pratt and DTM (1976) or by Cambridge Systematics in 1975 deserves special mention. The most famous of the behavioral travel demand models—the study by Charles River Associates for the Southwestern Pennsylvania Region Planning Commission on the city of Pittsburgh was an extension of the mode choice models to predict other choices as that of frequency, destination along with mode for shopping travel.

Apart from the precedence, the above studies created in backing mode-choice models, a recent thrust towards appropriate transportation system management [especially with low capital policies for improving the efficiency of urban transportation system focusing attention on public transportation (as this study also have brought out) and similar high-occupancy vehicle like bus] are intended to encourage shift towards discrete mode-choice models. For example, certain management strategies like the rise in petrol / diesel prices, parking-fees, imposing tolls on expressways, bypasses and highways or the sanitation of lanes as bus-bays all have ramifications on urban travel demand with profound impact on mode-choice.

In a multi-modal urban transport network for modeling mode-choice model in a behavioral set-up, the common model in vogue is the Multinomial Logit Model.

The Multinomial Logit Model is found to be the simplest and the most efficient of the available disaggregate models in terms of amenability to mathematical manipulation and easier estimation and interpretation of parameters. Hence, it is favoured widely as also in this study.

4.4 DISAGGREGATE MODELS—THE WORKING FRAMEWORK

Travel decisions of a typical trip-maker are choices among discrete sets of alternatives e.g. frequencies, destinations, modes and routes of travel. The disaggregate travel demand
models ensure a convenient behavioral and mathematical framework for portraying these choices. Mathematical representation of such a framework is as follows.

Let a trip maker has before him a set of J mutually exclusive alternative, of which one has to be chosen. Let C denote the set of all available alternatives and for each $j \in C$, let $z_j$ denote the vector of attributes of the individual expressed through an utility function $U(z_j)$. For any two distinct alternatives $k$ and $j \neq k$ in C, an individual selects the alternative $j$ to $k$ subject to the condition that $U(z_j) > U(z_k)$, i.e. on the concept of utility maximization. At this juncture, it is worth mentioning that though the representation of the assumed travel choice preferences through an utility function have yet been empirically tested in principle but certain types of non-compensatory and sequential decision rules along with constraint on choices that remain encompassed in the utility maximization principle are not found to have contributed against travel behaviour.

The impracticability in analyzing the whole spectrum of attributes that characterize an alternative (as some of the attributes are of qualitative nature or there may be idiosyncrasies of the typical trip-maker) leads to degrees of uncertainty in the proper selection of choice. This inherent uncertainty is dealt with by dividing the utility function into deterministic and random components. While the deterministic component is a function representing the observed attributes of individuals and alternatives (e.g., income, automobile ownership, household size, etc.) and accounts for the systematic effects of these attributes on choice, the random component deals with effects of the unobserved attributes (e.g., health, social status, occupation, other schedule commitments beside personal inclinations of the trip-maker). Thus, the utility $U_j$ of alternative $j$ is mathematically expressed as

$$U_j = V(x_j, \theta) + \xi(x_j, \theta)$$

where,
- $x_j$ is the vector of observed attributes of the individual for alternative $j$,
- $\theta$ is the vector of constant parameters that need to be estimated,
- $V(.)$ is the deterministic part of the utility function,
- $\xi$ is the random term associated with the alternative $j$ such that its probability distribution is dependent upon $x_j$ and $\theta$. 
The inclusion of random components renders probabilistic estimate of a travel choice rather than a deterministic prediction of the same. If \( x \) denote the matrix \((x_1, x_2, \ldots, x_j)\) then the probability that a randomly selected trip-maker chooses alternative \( j \in C \) is given by

\[
P(j \mid x, \theta, C) = \text{Prob}(U_j > U_k \text{ for all } k \in C, k \neq j) = \text{Prob}[V(x_j, \theta) + \xi_j > V(x_k, \theta) + \xi_k \text{ for all } i \in C, \forall k \neq j]
\]

The probability of selecting an alternative by a typical trip-maker is obtained through an explicit functional relation between the choice probabilities and the deterministic components of utility given the probability distributions of the random utility components which are assumed to be known.

### Multinomial Logit Model

Based on the various probability distributions of the random utility components of the choice function, various disaggregate models are being conceived of. The most famous and the simplest of those probability distributions taken up assumes the random variable \( \xi_k \) \((k=1, 2, \ldots, J)\) to be independently and identically distributed with zero mean and independent of \( x \) and \( \theta \) along with the difference between any two terms, say, \( \xi_1 \) & \( \xi_2 \) following a Logistic distribution. \( \xi \) follows a distribution function of the type—the Gumbel type I extreme value distribution;

\[
F(\xi) = e^{-e^{\xi}}
\]

Assuming the random components to have the above probability distribution, the choice probabilities are related to the deterministic components of the utility function through the well-known Multinomial Logit Model (MNL) represented as

\[
P(j \mid x, \theta, C) = \frac{\exp\{V(x_j, \theta)\}}{\sum_{k=1}^{n} \exp\{V(x_k, \theta)\}}
\]

where, \( x_j, \theta, j \) represents the elements as specified above, and developed by Daniel McFadden (1973) on the basis of the relationship between the probability distribution between the random component and the choice probabilities as also pointed out by Marschak in 1960 (See Appendix A.1 for deduction).

Various other assumptions of the probability distribution of the random component \( \xi \) like the multivariate normal distribution with zero mean and an arbitrary variance covariance matrix or the Generalized Extreme Value Distribution (GEV) leads to Multinomial Probit Model (Daganzo, 1979; Hausman & Wise, 1978) and the GEV models respectively.
4.5 THE ESTIMATION PROCEDURE

So as to determine the probability estimates of the choice alternatives, through the deterministic component of the choice function, a need to estimate the value(s) of the constant parameter \( \theta \) is essential. These are estimated by fitting the appropriate model to data consisting of observations of the choices and values of the observed attributes for a random sample of individuals.

The model thus fitted with observation of choices along with observed values of the variables is then calibrated for the parameters through the Maximum likelihood approach.

The method consists of selecting the value of the parameters that aptly fits the choice by combining well with the applied observations. The objective is accomplished by constructing a likelihood function \( L(\theta) \) and determining the estimated value of \( \theta \) i.e. \( \hat{\theta} \) by maximizing the logarithm of the likelihood function, i.e. the log likelihood function, given by

\[
\log L(\theta) = \sum_{n=1}^{N} \log P_c(\theta, a_n),
\]

where,

- \( \theta \) represents the parameter(s) to be estimated
- \( a_n \) is the attribute vector of the \( n^{th} \) observation
- \( c_n \) being the choice of the \( n^{th} \) observation
- \( N \) is the total number of observations.

Various standard optimization procedures are in vogue to obtain the maximum of \( \log L(\theta) \). Among the commonest ones are steepest ascent, variable-metric or the Newton-Raphson methods. The last one is widely used towards meeting the objective in Multinomial Logit model as it is considered the fastest search method. Various computer packages including EALimDep Version 7.0 econometric package used in the present analysis utilizes it through a process of iteration. In it \( \theta_n^{*} (\theta_n^{*} = \theta + \eta^{*} \Delta \theta) \), the parameter value for which \( \log L(\theta_n) \) is maximum is chosen through approximation of \( \log L(\theta) \) by a quadratic function with gradient and the matrix of partial second order derivatives equal to those of \( \log L(\theta) \) at the current point and setting \( \theta_n^{*} \) equal to the maximum of the quadratic function.
the method to work, the matrix of partial derivatives of second order needs to be negative
definite at each visited point, since otherwise the quadratic function does not have a
unique finite maximum value. The condition is automatically achieved in the MNL model
where the log likelihood function is strictly concave. For other models, the parameter
value 0 is chosen such that it is sufficiently close to the optimum.

4.6 APPLYING THE MODE-CHOICE MODEL—THE REQUIREMENTS

For the model to have practical application especially for the two cities where they are
wished to be applied, certain intricate details of the model e.g. the explanatory variables
with which the model works, the form of the parameters, the form of the choice function
and lastly the data requirements are to be examined. Also these features are also to be
appropriately administered with the conditions that are prevalent for the two cities where
they are to be applied. The following part elucidates the above requirement.

4.7 THE KEY ATTRIBUTES IN MODE-CHOICE

VARIABLES

Mode-choice models in a disaggregate set-up are based on the usual underlying postulates
of utility maximization and hence are behavioral in their characteristic. Hence, individual
data of the potential trip-makers used to explain his / her mode-choice are categorized into
two types—socio-economic demand and level of service or supply. The important socio-
edconomic demand variables used to explain the mode-choice behaviour are

(a) income
(b) car-ownership
(c) age and role in household
(d) household size
(e) residential location
(f) profession

The major level of service or supply variables on the other hand are

(a) travel time (in-vehicle / access-time / waiting time / transfer time)
(b) travel cost
(c) various qualitative variables like comfort, reliability, safety, etc
In a proper mode-choice analysis, the need is not only to select significant variables but those that provide an opportunity to analyze the pursued planning policy in order to determine their deficiencies. Besides, the availability of authentic information about the variables also shapes their inclusion. Keeping these two objectives into consideration, the two supply variables of travel time and travel cost are chosen as the ideal explanatory variables to conduct the present study.

**WAYS TO CHOOSE THE EFFICIENT MODE**

The behavioral models of mode-choice allow two distinct ways by which to choose the models from. They are

(a) mode-abstract postulate where the potential trip-maker forced with a choice of modal alternatives perceive the attributes themselves rather than the mode being considered. Hence, any two distinct mode with the same level of service attributes are treated as a single alternative. The postulate has its basis in Lancaster's theory of abstract commodities. In it, the consumer's decision to consume depend upon the characteristics of the available goods and services rather than the goods themselves. In urban mode choice analysis, the mode-abstract assumption implies the same choice function with the parameters same for all the alternatives.

(b) mode-specific assumption in which the choices are influenced by the level of service attributes that varies from one alternative to another. Hence, to a potential trip-maker, the modal choice gets influenced by the characteristics of the specific mode apart from the

In quantitative terms, the difference between a mode-abstract and a mode-specific model is in the variation of the choice function. While in the mode-abstract model, the parameters of the choice function, remain independent of the available modes, in the other, it is not. For example, in an area where there are two modes say A and B, and where the modal choice is based upon “travel-time” (travel-time of the A\textsuperscript{th} mode=\(T_A\) and travel-time of the B\textsuperscript{th} mode=\(T_B\)) and “travel cost” (travel cost of the A\textsuperscript{th} mode=\(C_A\) and travel cost of the B\textsuperscript{th} mode=\(C_B\)), by mode abstract specification of choice model, the choice function would be of the form
V(i)=aT_i + pC_i where,
α and β are constant parameters with i=A or B
In mode-specific formulation, the choice function would be represented as
V(i)=α_i T_i + β_i ψ_i where,
α_i and β_i are constant parameters, different for each of the two modes and ψ_i is a dummy variable / constant term that captures the effect of mode (that are not captured by the variable common to the modal alternatives.).

As mode-abstract choice function attempts to quantify mode characteristics rather than the mode themselves, it is seen as consistent with the travel demand analysis built on the framework of micro-economic theory. From the computational point of view, the requirements of a lower number of model parameters entail some amount of pragmatic advantage. However, in areas, where the modal characteristics play a prominent role in its choice, the abstract mode choice looses its edge to mode-specific function.

From the above discussion, the mode-abstract choice function seemed more appealing as an introductory approach in predicting modal choice in a framework for constructing the choice functions that is needed for the analysis.

4.8 DETERMINING THE FORM OF THE CHOICE FUNCTION

The computations associated with parameter estimation are simplified greatly when the deterministic component of the utility function is linear and additively separable in nature.
\[ V(X_j,\alpha) = \theta f(x_j) \] where, \[ \theta f(x_j) = \theta_1 x_1 + \theta_2 x_2 + \ldots + \theta_n x_n, \quad j = 1, \ldots, n; \]
f(.) is a known vector valued function of the attributes of various modes. The notion have hence been adopted in the present work. Such assumption found favour as a result of the large number of applications using the MNL model based on similar pattern.

4.9 THE FINAL FORM OF THE CHOICE FUNCTION APPLIED IN THIS STUDY

The conclusion that emerged about the inclusion of variables, form of parameters and the choice function lead towards final shape of the choice or the deterministic form of the utility function that was applied for choosing the appropriate mode for the two cities as
\[ V(i) = aT_i + bC_i \]
Where, $V(i)$ is the choice function of the trip maker for each of the two cities,

$T_i$ is the travel-time incurred by the $i^{th}$ trip-maker for each of the two cities.

$C_i$ is the travel-cost incurred by the $i^{th}$ trip-maker for each of the two cities.

### 4.10 ACQUISITION OF DATA

As explained above, the objective of this empirical effort lied in ascertaining the adequacy of the present policy pursued through modal choice models of travel demand analysis. In keeping with the objectives all types with distinct decision of mode choice are applied in the mode-choice models along with their attributes. Hence, the data preparation effort needs developing an accurate data-base. The initial step in obtaining and preparing data for the test estimation of the demand models lied in selection of appropriate urban areas that provide authentic data on individual trip records as well as “well developed” data on travel times and costs.

A source of authentic data on the above two criteria have been the “Traffic and Transportation Studies” of two urban areas of West Bengal e.g. Haldia Municipal Area and Howrah Municipal Corporation Area by Rail India Technical and Economic Services Limited (RITES). However, their data are mostly of aggregate nature and hence remained unsuitable for usage in the present analysis.

### 4.11 NEED FOR TRANSFORMING THE AVAILABLE DATA

Thus, steps toward transforming those data to suit the present demand through procedures based on statistical averaging followed by certain degrees of approximation were attempted.

### 4.12 LIMITATIONS OF THE TRANSFORMED DATA

The effect of such transformations yield values of the requisite variables that remained fixed for a finite set of individual rather than distinct values of variables for each of the surveyed trip-maker. In other words, while for individual-wise sample survey the travel-time and travel-cost for the actually chosen mode could have changed from individual to individual, the transformation of aggregate data into individual level led to same travel-time and travel-cost for the chosen mode for a group of individual. Further, from the transformed data only the travel-time and cost of the actually chosen mode could be approximately determined. However, for the Multinomial Logit model to work, the
perceived travel-times and costs for each of the available / competing modes by each of the trip-makers also need to be furnished. The deficiency is taken care of by substituting such values with the weighted averages of the travel-times and costs of the trip-makers who actually chose such models. Thus, such data loses its behavioral characteristics to a certain extent and induces certain amount of biasness in the estimated model coefficients and the probability estimates. The present analysis stands inclusive of these inherent limitations. Still then, the results are believed to portray the major features that are of significance to the policy planners.

4.13 GENERATING THE DATABASE REQUIRED FOR THE STUDY*

As specified above, the basic data source for the following analysis have been the aggregate household and traffic survey by RITES in their Traffic and Transportation Survey report on Haldia Town. Their informations are subjected to logical manipulation to yield the data necessary to execute the present analysis.

4.14 SHAPING THE RITES DATA TO ITS DESIRED FORM

Haldia Municipal Area

In the RITES report, a total of 1794 households were found to be interviewed in a way so as to provide all the wards in the municipality with equal weightage in the selection of households. These households were found to fall under six (6) income ranges. Since, the average household sizes are found to be 4.24 hence, the total family members for the households in each income class / category can be approximated by multiplying the number of households by the average sampled household size. Again, by multiplying the total family members in each income class by the per-capita vehicular trip rates, the overall vehicular trips generated by them are obtained. These generated trips can well be decomposed into their modal share components as per the percentage share of each mode (obtained by excluding the effect of walk mode from the total modal share of passenger trips) [AppendixC.1]

*Pain, Asis Kumar (Forthcoming: Apr.-Jun., 2004); “Stochastic Travel Demand Models in Mode-choice Estimation—a case study of work-trips in Haldia Municipal Area”, Indian Journal of Transport Management, Central Institute of Road Transport, Pune.
On the basis of an Origin-Destination (O-D) survey by RITES of all vehicular trips, the Per-capita vehicular trip rate is obtained as 1.217. The various ingredients in terms of purpose that comprises the Per-capita vehicular trip rate is given below in the Appendix (See Appendix A.5). Utilizing the Per-capita vehicular trip rate and the percentage share of existing modal usage, the total trips used by the sample population in each income class in respect of the existing modes are obtained. (Appendix C.2)

The O-D survey spoken of above, also surveyed 13206 trips of their in-vehicle travel-time. These trips were broken up among six ranges of travel-time (See Appendix B.6) so as to obtain a percentage share of trips for each range. These percentage values are applied to each of the modal trips so as to decompose the total trips by each of the modes in respect of their travel-time. (Appendix C.3)

After breaking the trips of each plying mode by their in vehicle travel-time, a further break-up of the trips in respect of travel-cost as a percentage, obtained from the O-D survey (Appendix B.7) are then applied to decompose the trips made by each plying mode in terms of their travel-time. (Appendix C.4)

Thus the number of trips for different ranges of travel costs for each range of travel-time for each of the income classes obtained. Since, the Per-capita Vehicular Trip rate is 1.217 (as obtained from O-D survey by RITES), hence the number of individuals that generate these trips are obtained by dividing each trip of the above-mentioned classification by that rate. Thus, a complete disintegration of the modal trips for different ranges of travel -costs and in-vehicle travel times and the individuals performing those trips are obtained. However, the presence of ranges of travel-time and cost results in the in-application of such enumeration as inputs for the Logit model analysis. The mean for each range of travel cost and travel-time (in-vehicle) is adopted as the appropriate singular-value for the accomplishment of the desired objective.

A further degree of realism is imparted by introducing the effect of waiting time to that of the in-vehicle travel-times of various plying modes. Allocation of such waiting-time for each mode is based on the examination of the ground situation in question. The waiting times allocation for various plying-modes are as in Appendix C.5.
The ultimate data to be used as inputs representing trip-makers incurring various travel-costs (single-valued) for each of the travel-time (single-valued in-vehicle travel-time + waiting-time) components are obtained as follows (Appendix C.6).

Limitation of the applied software program (EA LimDep Version 7.0) enabled using various fractions of the generated data (for each income category) to reach at the objectives aimed at. The various fractions of data usage for the various income classes are given as follows (Appendix C.7).

Each of these above observations shows the travel-time and travel-cost of the actually chosen mode along with the perceived ones (travel time & cost that would be expended had the mode been chosen) for each of the competing modes of each individual concerned. As the perceived travel times and travel costs of each of the competing modes are unavailable for the observations, hence the shortfall is taken account of by substituting those perceived travel-time(s) and cost(s) with the weighted average values as obtained from each mode. (Appendices C.8 - C.13)

Howrah Municipal Area

A sample of 4000 households (2% of the total) was chosen by RITES from the total population to conduct a survey of various socio-economic characteristics of the population. It showed the average household size (of the chosen sample) to be 4.2. Also a break-up of the entire households (in percentage term) in terms of their monthly earnings (Appendix B.12) was obtained. An O-D survey on all types of trips that is undertaken within the Municipal Corporation Area provides information on travel characteristics. Utilizing this information, the required input-data for the mode-choice logit model is derived.

The percentage break-up of households in terms of their monthly earnings are applied to the sample households to obtain their income-wise break-up.
Product of the sample households in each income-class times the average household size provides the total family members of the households in each particular income-class. Again, multiplying these individuals by the Per-capita trip-rate gives the total trips performed by these individuals for each income-class (Appendix C.14). The Per-capita Trip Rate is estimated at 0.58 from the O-D survey.

From the O-D survey conducted by RITES, the distributions of passenger trips by various modes of transport (in percentage terms) are obtained (Appendix B.13). However, such decomposition takes account of the walk-mode besides other, which finds no relevance in the present analysis. Hence, its effect has to be absorbed. Excluding the effect of walk-mode from such percentage distribution of trips by the various modes provides a distribution of such trips by vehicular modes. The realigned distribution of the vehicular modes in percentage terms are shown in Appendix C.15

These percentage distribution of modes are utilized by breaking the generated trips of the sampled households of each income class in respect of modes (Appendix C.16)

The O-D survey also provides an assessment of the trips in terms of their in-vehicle travel-time and costs (See Appendix B.14) in percentage terms. It is taken for granted that the trips performed by the sampled households are also contained within the limits of those travel-times and costs. Hence, the percentage break-up applies well for the trips by the population of the sampled households. On this basis, the generated trips by each of the modes for the various income classes are broken down with respect of their travel-time (Appendix C.17)

The temporal break-up of the trips for each of the modes are then further dissociated in terms of their travel-costs and determined in Appendix C.18.

After the complete division of trips with respect of their in-vehicle trip-time, and then by travel-cost, an estimation of the individuals belonging to the sample household who perform those trips are obtained by dividing the trips by the PCTR. Such a procedure yields observation of individuals performing trips within a range of trip-time and cost by each mode. However, the in-vehicle travel-time and cost being in the range results in their inapplicability in the EALimDep software applied for the mode-choice models. Hence,
singular values of in-vehicle travel-times and costs are essential. Such values are obtained by taking average value of the respective range of trip-times and costs. Since, such average values of in-vehicle travel-time reflects only the in-vehicle time spent, hence, an addition of appropriate minutes (for each mode) that a trip-maker spends by waiting for each appropriate mode provides the appropriate total-time spent in making a trip. The appropriate waiting-time corresponding to each mode obtained from practical experiences are as follows (Appendix C.13).

The total individual observations of each income class for various singular values of in-vehicular travel-time (i.e. average in-vehicle travel-time + appropriate waiting-time) and travel-cost are then calculated in Appendix C.20.

The total individual observation for each income-class however still remains incompatible as inputs in the appropriate software set-up. It arise due to the input limitation of the software to contain the individual observations within specified limit (1000 observations). Such a limitation imposed on the total observations (as depicted in Table 5.14 above) leads to selecting various percentages of total observations for each income class (Appendix C.21).

Each observation comprises of the travel-time and cost of the chosen mode by the individual and the travel-time and cost of the other competing modes in front of him that would be incurred if the trip-maker chooses any of them. The absence of the travel-times and costs of the competing modes is accommodated by taking the weighted average of the travel times and costs of each of these modes obtained from their actual users (Appendices C.22-27).