

8. A comprehensive model to capture the preference for mass rapid transit in Dhaka

Annesha Enam and Charisma F. Choudhury

1. BACKGROUND

Dhaka, the capital of Bangladesh and one of the fastest growing megacities of the world, is already subjected to acute traffic congestion on a regular basis. Increasing the physical capacity to relieve congestion is however not feasible since already more than 70 per cent of the area is built up (Bari and Hasan, 2001). This has recently prompted the Government to prioritize the introduction of mass rapid transit (MRT) options like bus rapid transit (BRT) and Metro Rail in the city. Planning these MRT options however requires rigorous mode choice models that can be used to predict ridership and quantify the willingness-to-pay (WTP) of the travellers. Though Dhaka is an old city (dating back to the 16th century), very few travel demand models have been developed for the city so far. Among the previous studies, four step travel demand models were developed in Dhaka Metropolitan Area Integrated Transport Study (DITS, 1993), Strategic Transport Plan (STP, 2005) and Dhaka Urban Transport Network Development Study (DHUTS, 2010) as well as Habib (2002) and Hasan (2007). However, in each case, the mode choice components were simplified and have grave limitations. In DITS, the mode choice model was simplified to a binomial choice model between private and public modes. In Habib (2002), a multinomial logit (MNL) model structure was adapted for the mode choice but the calibration results were counterintuitive with positive sign of the coefficients for time and cost parameters. In STP (2005), which is the most extensive travel demand model for Dhaka in recent years, a wide-scale household interview survey has been conducted for the first time. In the mode choice component of STP (2005), only two modes were considered i.e. public transport (PT) and individualized motorized vehicles (IMV). In the IMV group, cars and taxis were grouped together, overlooking their very different attributes (running cost, avail-

ability, accessibility, etc.) and the non-motorized vehicles (rickshaws) were not considered in spite of the fact that 37 per cent of the person trips in Dhaka were made by rickshaw, as reported in the same study (STP, 2005). Further, the STP model adapted pre-set rules for determining choice-sets and ignored the heterogeneity among respondents. In Hasan (2007), a rule based choice model was adapted for car and an MNL model was adapted for the choice among rickshaw, auto-rickshaw, taxi and bus. Hasan's model was based on STP data but the level-of-service (LOS) variables were updated using a supplemental survey (for cost) and outputs of the software EMME/2 (for travel time). The potential measurement errors introduced in this process have however been ignored. In DHUTS (2010), a two step sequential mode choice model has been developed. The first step involves a choice among walking, rickshaw and other modes (which includes private car, bus and auto rickshaw) based on origin–destination (O–D) shortest path distance. The second step deals with the choice among private car, bus and auto rickshaw using travel cost as the only explanatory variable.

As evident from the description above, the existing mode choice models for Dhaka are based on pre-set rules, ad-hoc choice-sets and network derived LOS values (without any correction for measurement errors). Further, the models are not robust enough to account for the new MRT, particularly since the LOS of MRT will vary significantly from the current modes. It may be noted that the limitations of the available datasets played a key role in the deficiencies of the previously developed mode choice models. This has prompted the current research, where we present a comprehensive mode choice model that overcomes the limitations of the previous models and is robust enough to capture the preferences for the proposed MRT modes.

In this chapter, the STP data have been explored in detail, the key modelling issues have been identified and modelling approaches have been proposed to account for the data limitations. The improvements from the proposed approaches have been demonstrated by comparing the value of time (VOT) values. The rest of the chapter is organized as follows: A short description of the revealed preference (RP) data, highlighting the main limitations of the data, and a description of the stated preference (SP) data (collected as part of this research) are presented first. This is followed by a description of the model framework. In the subsequent section, the estimation results of all the model components are presented, followed by the VOT comparisons. A summary of the findings and directions for future research are presented at the end.

2. DATA

In this research, the household interview survey (HIS) conducted as part of the Strategic Transport Plan (STP) Study in 2005 has been used as the main data source. This is supplemented by a detailed SP survey. The limitations of the RP data and the features of the SP data are elaborated below.

2.1 RP data

In the STP HIS more than 6000 households (STP, 2005) were interviewed and a huge amount of information was collected regarding the location and type of the households, socio-economic characteristics of their members, vehicle ownership of the households, daily trip information of the respondents and also some attitudinal information about the respondents. The socio-economic information data are extensive, with age, gender, education level, employment status, occupation, driving license status, address of the worksite and educational institutions. The daily trips of all members of the households were reported which revealed information more than 30,000 trips. The trip diary consisted of the origin–destination locations, start and end time between origin and destination (but not for each modal segment of the trip chain), and purpose and transport modes of each trip segment. The attitudinal part comprised questions on reasons behind choosing the current mode of travel and existing problems of the current travel modes, and asked for suggestions for the improvement of the traffic situation, to name a few. More details of the data are available in STP (2005). Detailed exploration of the data revealed three key limitations:

- (1) Unobserved choice-sets
- (2) Unobserved LOS variables like travel time of unchosen mode and travel cost
- (3) Absence of preference indications for MRT

These are elaborated below.

2.1.1 Unobserved choice-sets

Assessing the choice-set correctly is a prime requirement for estimating travellers' final choice of mode, because any misspecification of the choice-set is likely to lead to an erroneous prediction of the traveller's final choice. But specifying the choice-set correctly has been a very difficult task in developing countries like Bangladesh where the transport network data

are not structured. For example, para-transits like human-haulers (and even in some cases traditional transits like buses) operate beyond their permitted routes, which makes it impossible to infer the accessibility to these public transport modes from certain zones using transport network data. Further, local experience suggests that the family structure and socio-economic status of the travellers play a vital role in dictating the choice-sets. This fact has further been supported by the estimated choice-set generation model where the coefficients to the socio-economic variables (for example, household income) have been found to be significant. Multi-use of the same chauffeur driven car also adds complexity in the car availability since neither car-ownership nor possession of a driving license has a direct correlation with car availability. For example, a person belonging to a household where no member has a driving licence (a common proxy for car availability in the rule based choice-sets) can still have car in the choice-set. On the other hand, the affluent people who have access to a car are very often unaware of what public transport options exist, mainly because of the poor travel-information system. Therefore, a car or bus is often 'not chosen' in spite of having higher LOS as it is not in the choice-set in the first place. This makes the task of determining the choice-set correctly even more essential.

2.1.2 Unobserved LOS variables

High quality LOS data are also essential for accurate estimates of the mode choice model. Disregarding the measurement errors in the LOS variables can lead to inconsistent estimates of the parameters (Walker et al., 2008). But, in the context of Dhaka, the available data from STP HIS has only the stated travel times of the respondents/travellers for the chosen mode. The travel times of the unchosen modes and the fares of all the modes are missing in the data set. This data has been supplemented by Hasan (2007), who has determined the distances between different traffic analysis zones (TAZ) from the network assignment using EMME/2 and has used this distance to arrive at the travel times of different modes. Hasan (2007) has also conducted a small scale HIS to calculate an average fare for different modes. But the measurement errors introduced in this manner have been ignored in the study.

2.1.3 Absence of preference for MRT

Dhaka is probably the only mega city with no well-structured public transport system. The LOS of the current bus and paratransits are not comparable to those of BRT or Metro Rail. Therefore, the RP data collected during the STP HIS are not comprehensive enough to capture the preference of the travellers for the proposed new modes.

2.2 SP data

In order to address the above mentioned limitations of the existing data, a small scale SP survey has been conducted (data collection period: December 2009 – January 2010), collecting 1016 samples. Respondents have been presented with mode choice scenarios that included BRT and Metro Rail alongside their current modes. Three attributes (travel time, travel fare and frequency) were used to describe the new alternatives. Instead of using the absolute values for travel times and travel costs, these attributes for the two new modes were presented as the incremented/decremented values from the current travel costs and times. The profiles of the BRT and Metro Rail have been generated separately with the statistical software SPSS using full factorial design. Four levels of travel time and travel cost each and five levels of frequency resulted in 80 different profiles ($4^1 \times 4^1 \times 5^1$) for BRT and Metro Rail. These 160 profiles when combined with each other resulted in 3160 (${}^{80}C_2$) different choice-sets. Extreme/dominant combinations have been eliminated from the choice-sets using the following two criteria.

- (1) In every retained choice-set, BRT levels of service attributes were at most one level better than the corresponding Metro Rail level of service attributes. For example, in the SP options for medium trips, the four levels of the attribute 'reduced travel time' for BRT were 10, 15, 18 and 20 minutes respectively and those for Metro Rail were 10, 15, 20 and 25 minutes respectively. So, in a particular choice-set, if the 'reduced travel time' for BRT is 20 minutes (level 4) the corresponding value for the Metro Rail is constrained to be 20 minutes (level 3) or better;
- (2) In every choice-set the ratio of 'reduced travel cost' to 'reduced travel time' has always been greater in Metro Rail than in BRT.

These two constraints were imposed from an underlying assumption that the level of service of the proposed Metro Rail routes will be better than those of the proposed BRT routes in general and consequently, the value of time (VOT) for the Metro Rail will be greater than that of BRT.

The survey questionnaire included an explicit question on the respondent's available modes. Beside the mode choice information in hypothetical scenarios, the survey provided information regarding the socio-economic characteristics of the respondent (e.g. age, gender, occupation, education, household income, car ownership, chauffeur availability, etc.) as well as attributes of the trip in question (e.g. purpose, duration, distance travelled, etc.). The major portion of the data was collected by roadside face-to-face random surveys, i.e. the survey data was collected by a roadside interview among randomly chosen persons. This was supplemented by random

telephone surveys. The detailed survey methodology and data description are available at Enam (2010).

3. PROPOSED METHODOLOGY

The proposed model framework addresses the data limitations where the choice-set of the respondents (in case of RP data), attributes of modes and the utility of different modes are not directly stated by the respondents and therefore are unobserved to the analyst. The observed variables include the socio-economic characteristics of the respondents, measured values of the trip attributes from network analysis and the revealed and stated choices of the respondents in RP and SP surveys respectively. The analyst has to rely on the measurement relationships among the observed and unobserved variables to arrive at the unobserved variables. In this process error terms are associated with the unobserved variables. However, these error terms are assumed to be independently and identically (iid) gumbel distributed leading to logit forms.

The integrated model consists of the following model components:

- (1) Choice-set generation;
- (2) Measurement error correction;
- (3) Revealed mode choice; and
- (4) Stated mode choice.

In the framework, socio-economic characteristics of the travellers collected from the SP survey are used to model the choice-set and to probabilistically predict the unobserved choice-sets of the respondents in the RP data. The measurement errors of the network derived (Hasan, 2007) mode attributes are addressed by regression analysis. The predicted choice-set values (i.e. the probabilities for each observation to have a particular choice-set given the universal choice-set) and the corrected LOS data are used to deduce the utility parameters with the revealed choice of the respondents. The RP data is also augmented with SP data (in which the choice-sets consisted of the proposed new modes as well) to establish the utility parameters of the combined model. The four model components are briefly described in the following sub-sections.

3.1 Choice-set generation model

As described in Section 1, in the previous mode choice models developed for Dhaka City, different deterministic rules have been used for specifying

the choice-sets. Previous research has shown that incorrect representation of choice-sets and the imputation of choice-sets by the use of logical rules can lead to biases in parameter estimation and errors in forecasting due to misallocation of the alternatives (Williams and Ortuzar, 1979).

In classical economic theories mode choice has been modelled as a two stage sequential process: (1) generation of the choice-set, and (2) selection of the final choice of mode (Manski, 1977). The choice-set generation process has been considered either as deterministic (e.g. Ben-Akiva and Lerman, 1974; Train, 1980, etc.) or probabilistic (e.g. Wermuth, 1978; Swait, 1984) depending upon the degree of confidence in the information at hand. The probabilistic approach of choice-set generation process has produced several different methodologies starting with the logit captivity model which has been independently developed by Ben-Akiva (1977) and Gaudry and Dagenais (1979). Other approaches of probabilistic choice-set generation model include independent availability logit model (Swait, 1984) and parameterized logit captivity model (e.g. Swait and Ben-Akiva, 1985). However, none of the above mentioned choice-set generation models were directly applicable in the context of Dhaka, where the RP data had missing LOS as well and was not a reliable input for estimating the choice-set generation component in the above manner.

In the proposed approach, a choice-set generation model has been developed using the SP survey data to predict the choice-set probabilistically for the choice context of Dhaka using socio-economic characteristics, origin–destinations and trip purposes. In the SP survey, an explicit question was included on the availability of different modes as perceived by the respondents. Each respondent was presented with a list of typical travel modes in Dhaka and asked what modes were available to them for this particular trip. The answer of the respondents explicitly revealed the modes considered by the respondents for the trip in question. These stated choice-sets along with the comprehensive socio-economic and trip related data of the travellers have been used to form their choice-sets in RP data.

For the mode choice context and the population of Dhaka city the universal choice-set consists of bus/tempo, car, CNG/taxi and rickshaw. In the extreme cases, the traveller is captive to a single mode or considers all modes in the choice-set. Thus, the total number of non empty subsets of this universal choice-set is fifteen. It should be mentioned that, during the estimation of the model three choice-sets have been eliminated due to the absence of a considerable number of observations, which resulted in twelve choice-sets.

The candidate socio-economic attributes and mode characteristics to affect the choice-set generation process are monthly household income (HHI), gender, education, occupation, age, trip purpose, travel duration,

etc. To quantify the effect of these variables in the choice-sets of an individual, an explorative analysis has been done where the correlation with the choice-sets was found to be highest for income, age and gender of the traveller and duration and purpose of the trip.

Based on the findings of the exploratory analysis, a discrete choice modelling technique has been used to estimate the utility parameters of different choice-sets. A linear utility function is associated with each choice-set. The utility of a choice-set i of individual n can be expressed as in equation (8.1).

$$U_{in} = \beta_i X_{in} + \epsilon_{in}, \forall i \in C_n; \quad (8.1)$$

where, X_{in} = socio-economic characteristics of the individuals and attributes of different modes; β_i = coefficient of X_{in} ; ϵ_{in} = random error term \sim Gumble (0,1) and C_n = universal choice-set or the choice-set determined deterministically for individual n . The twelve choice-sets for which the utility equations have been estimated are:

- | | |
|---------------------------------------|------------------------|
| 1. bus/tempo, car, cng/taxi, rickshaw | 2. bus/tempo, car |
| 3. bus/tempo, car, cng/taxi | 4. bus/tempo, cng/taxi |
| 5. bus/tempo, cng/taxi, rickshaw | 6. bus/tempo, rickshaw |
| 7. bus/tempo | 8. car, cng/taxi |
| 9. car, cng/taxi, rickshaw | 10. car, rickshaw |
| 11. cng/taxi, rickshaw | 12. rickshaw |

The methodology is further elaborated in Enam and Choudhury (2011).

3.2 Measurement error correction

As mentioned in Section 2, the LOS variables are partially unobserved in the STP HIS data. This data has been supplemented by Hasan (2007) through network analysis without accounting for the measurement errors introduced in that manner. It has been acknowledged that the measurement errors could have resulted from the following two major sources:

- (1) The distances are the zone to zone distances, i.e. they are not the actual distance between the origin and destination of the traveller. Though the zone to zone distances may suffice for the needs of aggregate analysis (e.g. trip distribution and trip assignment) such

assumptions may not be sufficient for the models that are estimated at disaggregate level;

- (2) The traffic on the streets of Dhaka is heterogeneous in nature, and due to congestion as well as chaos due to weak lane-disciplines, the vehicles do not necessarily run at their free flow speeds. In fact the speeds of the vehicles are not only a function of the mode but also depend on the road characteristics (e.g. width, surface quality, traffic mix and presence of non-motorized traffic). Since a single speed for a mode has been used for the calculation of the travel time ignoring these sources of heterogeneity, it might have introduced some errors.

Though there has been a vast literature on measurement error in economic theory or even in medical science, most of this is not directly related to measurement errors in discrete choice modelling or in transportation planning. Brownstone et al. (2001) and Steimeitz and Brownstone (2005) have developed and demonstrated multiple data imputation techniques to overcome the measurement errors in travel data. Ben-Akiva et al. (2002) have acknowledged that this technique is applicable only when validation data are available. Walker et al. (2008) have modelled travel time as a latent variable to overcome the consequences of measurement error for the work trip context of Chengdu, China. There are other works also which focus on the improvement of level of service measurements (e.g. Ortuzar and Ivelic, 1987; Ortuzar and Willumsen, 2001).

In the present study, a relationship is developed between the stated travel time (calculated from the stated departure and arrival time of the work trips by the travellers) by the respondents and the travel time obtained from network assignment and assumed speed to explore potential systematic variations between the two and determine necessary correction factors. It may be mentioned that, in this study, the stated travel time has been treated as a better estimate of the travel time of different modes than the network derived travel time, on the ground that the sources of errors of the available network derived travel time are supposed to be greater than those of the stated travel time. In this regard a regression analysis has been performed to develop a relationship among the true and measured travel times between the O-D pairs. The stated travel times by the travellers are considered as the true measure of the travel time while the calculated travel times are used as the measured travel times with errors. The proposed relationship of the true and measured travel times can be expressed as in equation (8.2).

$$T_{true} = \alpha + \beta * T_{measured} + \epsilon \quad (8.2)$$

Where T_{true} is the true value of the travel time i.e. the stated travel time by the traveller; $T_{measured}$ is the measured value of the travel time; α , β are the systematic components to be estimated by regression analysis; ϵ is normally distributed random error term with mean zero and variance σ^2 ; i.e. $\epsilon \sim N(0, \sigma^2)$. In this analysis, the hypothesis of the presence of systematic components of error is tested through statistical tests. It may be noted that α indicates the fixed component of the measurement error (if any), β indicates the systematic scale difference between the true and the measured values (if any) and ϵ represents the random part of the error. The methodology is further elaborated in Enam and Choudhury (2011).

The basic difference between the approaches proposed in this work and that followed by Walker et al. (2008) is that the travel time has been treated here as an observed variable with some measurement errors, while in the latter case travel time has been treated as a latent variable.

3.3 RP and SP mode choice model development

In the present research, the universal choice-sets of the RP and the SP data source are different. In case of SP choices, the choice-sets consisted of the set {current mode, BRT and Metro Rail}. For example, for a ‘current bus/tempo user’, the choice-set comprises bus/tempo, BRT and Metro Rail, for a ‘current car user’, the choice-set comprises car, BRT and Metro Rail, etc. Therefore, the choice-sets are observed variables in case of SP data.

In the RP data, the exact choice-sets are however unobserved. Therefore, the choice-set generation model and measurement error correction factors are incorporated into the mode choice model. In order to do that, the estimated parameters of the choice-set generation model (Enam and Choudhury, 2010) are used to calculate the probability of the different choice-sets for each individual and the estimated measurement error correction factors (Enam and Choudhury, 2010) are multiplied to the derived travel time to deduce the corrected travel time.

In order to incorporate the choice-set probabilities into the mode choice model, equation (8.3) has been used.

$$P_{in} = P_n(i|C_1)*P_n(C_1) + P_n(i|C_2)*P_n(C_2) + \dots + P_n(i|C_c)*P_n(C_c) \tag{8.3}$$

where, P_{in} is the probability of individual n choosing a mode i ; $P_n(i|C_c)$ is the probability of choosing mode i given choice-set C_c and $P_n(C_c)$ is the probability of C_c being the choice-set of the individual n and C_c is the c^{th} choice-set of individual n .

$$P_n(i|C_c) = \frac{e^{V_{in}}}{\sum_{j=1}^M e^{V_{jn}}} \quad (8.4)$$

Where, V_{in} is the deterministic utility portion of mode i and M is the total number of modes in the choice-set C_c .

$$V_{in} = \beta_i X_{in}, \forall i \in C_c; \quad (8.5)$$

where, X_{in} = attributes of different modes and socio-economic characteristics of the individual n and β_i = coefficient of X_{in} to be estimated by maximum likelihood estimation technique.

Linear utility functions similar to equation (8.5) have been used for SP mode choice. Equations (8.6) and (8.7) have been used for the calculation of the likelihood for RP and SP observations respectively.

$$LL^{RP} = \sum_1^{n=N^{RP}} \sum_1^{i=M^{RP}} y_{in}^{RP} \log(P_{in}^{RP}) \quad (8.6)$$

$$LL^{SP} = \sum_1^{n=N^{SP}} \sum_1^{i=M^{SP}} y_{in}^{SP} \log(P_{in}^{SP}) \quad (8.7)$$

where, LL^{RP} and LL^{SP} are the log-likelihood of the RP and SP observations respectively, N is the number of observations, M is the number of modes in each case and y_{in} is an indicator variable, the value of which is 1 if individual n chooses mode i and 0 otherwise.

The deterministic utility portion for a mode in the combined model can be written as in equation (8.8).

$$V_{in} = \theta^{RP}(\alpha_i^{RP} + \beta X_i^{RP} + \omega Z_i) + \theta^{SP}(\alpha_i^{SP} + \beta X_i^{SP} + \delta W_i) \quad (8.8)$$

where, i is an alternative in choice-sets C^{RP} or C^{SP} , α s are data source specific alternative specific constants (ASCs), β is the utility parameter for the attributes common to both data sets, δ and ω are utility parameters for the unique attributes in each data set, and θ^{RP} and θ^{SP} are the scale factors associated with the error terms of the RP and SP data source respectively.

In the present study the scale factor of SP data source is set to one ($\Theta^{SP} \equiv 1$), and so the estimate of the scale factor of RP data source Θ^{RP} represents a relative scale with respect to the SP data scale.

It should be noted that, ideally, the parameters of all the four model components should be estimated jointly. The correlations among errors in between choice-set generation and revealed preference mode choice have however been ignored in this research and the parameters of the measure-

ment relationships in the choice-set generation model and RP-SP model have been estimated disjointly.

4. MODEL ESTIMATION

4.1 Choice-set generation model estimation result

As discussed in Section 3.1, the choice-set is likely to be affected by the attributes of the alternative modes as well as the socio-economic characteristics of the individuals. The estimation is started with monthly household income dummy and gender dummy, age dummy, trip purpose dummy and trip length dummy are added step by step. These variables are included on the ground of significant improvement in the goodness-of-fit (adjusted Rho-square) and the sign of the estimated parameters. The final estimation results of the choice-set generation model using BIOGEME (Bierlaire 2003) are presented in Table 8.1. The alternative for which a specific parameter is used has been identified in the same table in the left-most column.

The results indicate that all else being equal, the rickshaws have the smallest probabilities of being included in the choice-set. CNG/taxi on the other hand has the highest value of alternative specific constant (ASC) indicating higher likelihood of it being included in the choice-sets. However, it should be noted that the ASCs in SP studies are not representative of market shares; rather, they merely indicate the part of the utility unexplained by the explanatory variables (Bliemer et al., 2009).

Coefficients of income have different value and sensitivity for different income groups. The coefficient of income for the range of less than or equal to BDT¹ 20,000 monthly incomes is very significant (at more than 95 per cent confidence level) for the public and personalized public transport user group, which is intuitive for the context of the city since members of this income group generally have no or shared access to a car and are well aware of public transport availability. On the other hand, the coefficient of income for the range of greater than or equal to BDT 50,000 monthly incomes is statistically significant for the people whose choice-set includes car and does not include bus/tempo.

The dummy term introduced for female respondents indicates that female respondents have significantly high likelihood to consider the choice-sets which include car, CNG/taxi and rickshaw. The trip purpose dummy for educational and work trip exhibits significantly high preference of the user for the choice-sets which include bus/tempo and other personalized public transport. Two travel duration dummies are

Table 8.1 Choice-set generation model estimation results

Parameters	Estimated values (t-stat)	Alternatives in which the parameter is used
Constants:		
bus and tempo, $\alpha_{\text{bus_tempo}}$	0	1,2,3,4,5,6,7
car, α_{car}	0.344 (2.55)	1,2,3,9,10
cng and taxi, $\alpha_{\text{cng_taxi}}$	1.140 (12.65)	1,3,4,5,8,9,11
rickshaw, α_{rickshaw}	-1.150 (-13.60)	1,5,6,9,10,11,12
dummy: age between 18 to 25, β_{age}	0.555 (3.26)	4,5,6,7
dummy: female respondents, β_{gender}	0.912 (4.42)	8,9,10,11,12
dummy: income less than or equal to 20 000 BDT monthly, β_{income1}	2.280 (7.13)	4,5,6,7
dummy: income more than or equal to 50 000 BDT monthly, β_{income2}	0.430 (2.10)	1,8,9,10,11,12
dummy: trip purpose educational or work, β_{tp1}	1.090 (6.87)	4,5,6,7
dummy: trip duration greater than 45 minutes, β_{ttvlong}	1.400 (8.73)	4,5,6,7
dummy: trip duration 15 to 30 minutes, β_{ttvshort}	2.030 (6.07)	12
No. of Observations	756	
No. of Parameters	10	
LL at Zero	-1878.589	
LL at Convergence	-1372.242	
Adjusted Rho-Square	0.264	

introduced in the model. Both of them are highly significant. The long trip duration dummy (trip duration more than 45 minutes) indicates that public transport is more likely to be included in the choice-set for long trips, while the short trip dummy (trip duration 15 to 30 minutes inclusive) indicates that rickshaw is more likely to be included in the choice-set for short trips. The coefficient of age indicates that young commuters have a significant preference for the choice-sets which include bus/tempo.

4.2 Measurement error correction estimation result

The regression analysis resulted in the estimation equation

$$T_{\text{true}} = -0.434 + 0.964xT_{\text{measured}}$$

Table 8.2 Final estimation results of regression analysis

Modes	Regression equation	Adjusted R square value	t-stats of the estimated β
Tempo and public transport (bus)	$T_{true} = 0.980 \times T_{measured}$	0.973	141.980
Taxicab and CNG	$T_{true} = 1.067 \times T_{measured}$	0.929	29.067
Private car/microbus and motor cycle	$T_{true} = 1.219 \times T_{measured}$	0.890	25.580
Rickshaw	$T_{true} = 0.780 \times T_{measured}$	0.873	54.267

with all the modes pooled together. The constant α has been statistically insignificant with a t-statistics value of -0.880 and consequently has been ignored for the consequent trials of the estimation. In the final estimation result the seven modes are grouped into four groups. The final estimation results of the analysis are presented in Table 8.2.

The estimation results indicate that as hypothesized there are some systematic measurement errors and these estimated parameters have been used to impute the travel times of the non-chosen alternatives while estimating the mode choice model.

4.3 Combined model estimation results

The combined RP-SP model has been estimated by GAUSS 10 software. The model estimation has been started with generic parameters (common to both data sources) for travel time and travel cost and SP data source specific generic waiting time parameter. In addition data source specific ASCs are also estimated. Market segmentation tests are done to test the systematic taste heterogeneity where Likelihood Ratio test is done to test the statistical significance of market segmentation. Based on the market segmentation test results taste heterogeneity is addressed in the final model by adding dummy parameters to the socio-economic attributes such as gender dummy, age dummy, trip purpose dummy and employment dummy. The estimation result of the final combined RP-SP model is presented in Table 8.3. It should be mentioned that all the parameters of the final model are common to both data sources except the ASCs of BRT and Metro and the coefficient to the NHB and HBO dummies which are specific to SP data (where home-based trips are those where home is either origin or destination and non-home-based trips are those where neither origin nor destination is home).

Table 8.3 Estimation results of the final combined RP-SP model

Parameters	Estimation results (t-stat)
Constants:	
Car, α_{car} , common in RP and SP	-1.5924 (-7.171)
Bus/tempo, $\alpha_{bus/tempo}$, common in RP and SP	0
CNG/taxi, $\alpha_{cng/taxi}$, common in RP and SP	-0.9622 (-7.136)
Rickshaw, $\alpha_{rickshaw}$, common in RP and SP	2.2598 (5.636)
BRT, α_{brt} , specific to SP	1.0443 (4.031)
Metro Rail, α_{metro} , specific to SP	0.4465 (1.684)
Travel time coefficient:	
Motorized modes, $\beta_{tt_motorized}$, common in RP and SP	-0.0348 (-4.547)
Non-motorized mode, β_{tt_rick} , common in RP and SP	-0.0752 (-7.682)
Travel cost coefficient:	
Bus/tempo and rickshaw, $\beta_{fare_bus/tempo,rick}$, common in RP and SP	-0.0709 (-4.182)
CNG/taxi and car, $\beta_{fare_cng/taxi,car}$, common in RP and SP	-0.0123 (-5.322)
BRT, β_{fare_brt} , specific to SP	-0.0524 (-7.926)
Metro Rail, β_{fare_metro} , specific to SP	-0.0379 (-6.485)
Generic WT parameter, β_{wt} , specific to SP	-0.0412 (-5.095)
dummy: full time employment, common in RP and SP	
Car, β_{emp}	1.3638 (9.330)
Other modes	0.0 (-)
dummy: income above 50,000 BDT/month, common in RP and SP	
Car, $\beta_{hincome}$	1.2851 (7.223)
Other modes	0.0 (-)
dummy: income below 20,000 BDT/month, common in RP and SP	
Bus/tempo, $\beta_{lincome}$	0.6181 (4.382)
Other modes	0.0 (-)
dummy: female gender, common in RP and SP	
Bus/tempo, β_{gender}	-1.0890 (-7.013)
Other modes	0.0 (-)
dummy: age greater than 40 years, common in RP and SP	
Bus/tempo, β_{age}	-0.3712 (-2.388)
Other modes	0.0 (-)
dummy: respondent is student, common in RP and SP	
Rickshaw, $\beta_{student}$	0.9901 (4.022)
Other modes	0.0 (-)
dummy: trip is non-home-based, specific to SP	
Car, β_{nhb}	0.7972 (3.352)
Other modes	0.0 (-)

Table 8.3 (continued)

Parameters	Estimation results (t-stat)
dummy: trip is home-based other, specific to SP	
Bus, β_{hbo}	-0.8556 (-5.749)
Other modes	0.0 (-)
RP scale factor relative to SP scale factor, Θ , multiplied with the ASCs and the travel time and cost parameters	3.6260 (4.908)***
No. of Observations	8459
No. of Parameters	21
LL at Zero	-13071.9
LL at Convergence	-7936.69
Adjusted Rho-Square	0.394

Note: *** t-stat with respect to one.

Estimated coefficients indicate that, all other things remaining the same, rickshaw is the most preferred mode, followed by BRT and Metro Rail, while car is the least preferred mode next to CNG/taxi.

The sensitivity of people towards travel time does vary significantly over motorized and non-motorized modes of transport, with higher value of coefficient for the travel time of rickshaw. Cost sensitivity also varies over different modes with highest absolute value of the parameter for bus/tempo and rickshaw, followed by BRT and Metro Rail. People's sensitivity for travel cost is least for car and CNG/taxi. The sensitivity for waiting time is found to be the same for all modes of transport.

The employment dummy and the high income dummy are positive in car which implies that, all else being equal, the people with non-flexible work schedules and/or people with household income more than BDT 50,000 per month are inclined to use the car over other modes of transport. The positive parameter for NHB dummy in car indicates, all else being equal, people prefer the car for NHB trips.

The negative age dummy and low income dummy in bus and tempo mode indicate that the traveller whose household income is less than BDT 20,000 per month and the traveller whose age is over 40 years are reluctant to use bus/tempo over other modes, all else being equal. Also the negative gender dummy implies female travellers have an additional dislike for bus and tempo modes compared to male travellers. The negative parameter for HBO dummy in bus indicates that, all else being equal, people are reluctant to use the bus in HBO trips compared to other modes. The positive student dummy in rickshaw indicates that

students have additional preference for rickshaw compared to other modes.

Finally, the scale parameter Θ was also found to be significantly different from 1 (t-stat 4.93) indicating a significant difference in the variance between the RP and SP data. Since the scale factor is normalized for SP data source and the estimated value is greater than 1, it indicates that the variance of the error term of the RP data is smaller than that of the SP data. Since the proposed BRT and Metro are yet to be implemented in Dhaka, it was not possible to do a ridership prediction/rigorous validation. Value of time (VOT) has been calculated from the estimated model parameters as an alternative to demonstrate the improvement of the current model from the previously developed mode choice models for Dhaka city. However, a prediction test of the RP portion of the model has been presented in Enam and Choudhury (2011).

5. COMPARISON OF VALUE OF TIME

VOT is a key concept in transport planning in terms of mode choice decisions of the users. In the previous models developed for Dhaka city, the estimated VOT has been found to cover a wide range. For example, according to STP (2005) the VOT for the low income group of people (HHI < 12,500 BDT/month) is as small as 37.26 BDT/hour and the VOT for the middle income group of people is as high as 2040 BDT/hour with 1937 BDT/hour for the high income group of people. Again, in the model developed by Hasan (2007) the VOT has been found to vary from 190 BDT/hour for non-home-based trips to 477 BDT/hour for home-based education trips. The VOT has been estimated to be 195 BDT/hour for home-based work trips and 298 BDT/hour for home-based other trips.

In the current study three choice models have been estimated with three different data sources (RP, SP and combined RP-SP). The VOT as obtained from these three models is presented in Table 8.4. From the VOT values, it can be inferred that the VOT obtained from the stated choice of people is higher than that revealed from their actual choices. This difference may be attributed to policy response bias, justification bias, etc. (Ben-Akiva 2008). VOT is a very important criterion for assessing the fare structure to be introduced for new proposed modes of transport and even for the introduction of new attributes into an already existing alternative. In most of these cases, one has to rely on the SP data if the data of the comparable LOS or related comparable attribute are not available. The combination of the SP data with the RP data can therefore increase the reliability of the findings in such cases.

Table 8.4 VOT from the current study

Data source	VOT (Tk/hour)
SP:	
Bus/tempo	35.10
Rickshaw	22.86
CNG/taxi	251.46
Car	251.46
BRT	72.06
Metro	100.74
RP:	
Bus/tempo	26.10
Rickshaw	32.46
Car	106.80
CNG/taxi	141.60
Combined RP-SP:	
Bus/tempo	29.46
Rickshaw	63.66
CNG/taxi	169.74
Car	169.74
BRT	39.84
Metro	55.08

The VOT of the car and CNG/taxi user group is substantially different from the other mode users, which indicates that the affordability level of car and CNG/taxi users is significantly higher than that of the other mode users in Dhaka city. This fact can be utilized by providing reliable and on-time services to these groups in order to attract them to the proposed BRT and Metro Rail. VOT values obtained from the current study are much more realistic than the VOT obtained from similar studies in the recent past (STP, 2005 and Hasan, 2007) which were unrealistically high. This improvement can be attributed to the up-gradation of the LOS variables and better specification of the choice-set compared to the previous models.

6. CONCLUSION

In this study an effort has been made to address the key limitations of the RP mode choice data for Dhaka, such as missing choice-set information of the respondents and the measurement errors in the LOS variables. An extensive SP survey has been conducted to overcome the lack of preference data for the proposed improved public transport modes for Dhaka.

Finally a comprehensive mode choice model has been developed by combining the RP and SP data sources, and it should be noted that it is the first choice model of its kind for Dhaka city. The developed model has explicitly addressed the systematic taste heterogeneity of the people by the introduction of dummy variables, which is a significant improvement from the previous models where it was only possible to capture the time and cost sensitivities.

From a practical point of view, these models are likely to provide better ridership predictions and more robust cost-benefit analysis. This was evident from the VOT results calculated from the estimated parameters of the mode choice models, where the values were much more reasonable compared to values obtained from previous choice models. It may be noted that though the methodologies proposed in this chapter have been formulated in the context of Dhaka with available data in mind, the methodologies adapted in this research can provide useful guidance for mode choice developments in other developing countries, which often face modelling challenges and a similar dearth of data.

NOTE

1. 1 US dollar = 71 BDT

REFERENCES

- Bari, M.F. and Hasan, M. (2001), 'Effect of Urbanization on Storm Runoff Characteristics of Dhaka City', Tsinghua University Press. XXIX IAHR Congress, Beijing.
- Ben-Akiva, M.E. (1977), 'Choice Models with Simple Choice-set Generation Processes', Working Paper, Department of Civil Engineering, MIT, Cambridge, Massachusetts.
- Ben-Akiva, M.E. (2008), 'Individual Choice Behavior: Theory and Application of Discrete Choice Analysis', *Case Studies Workbook*, MIT, USA.
- Ben-Akiva, M. and Lerman, S. (1974), 'Some Estimation Results of a Simultaneous Model of Auto Ownership and Mode Choice to Work', *Transportation*, **4** (4), 357–376.
- Bierlaire, M. (2003), 'BIOGEME: A Free Package for the Estimation of Discrete Choice Models', *Proceedings of the 3rd Swiss Transportation Research Conference*, Ascona, Switzerland.
- Ben-Akiva, M.E., McFadden D., Train, K., Walker, J., Bhat, C., Bierlaire, M., Bolduc, D., Boersch-Supan, A., Brownstone, D., Bunch, D., Daly, A., De Palma, A., Gopinath, D., Karlstrom, A. and Munizaga, A. (2002), 'Hybrid Choice Models: Progress and Challenges', *Marketing Letters*, **13**(3), 163–175.
- Bliemer, M., Rose, J., and Blokland, V. (2009), 'Experimental Design Influences

- on Stated Choice Outputs: An Empirical Study in Air Travel Choice', 12th Conference of the International Association of Travel Behavior Research, Jaipur, India.
- Brownstone, D., Golob, T.F., and Kazimi, C. (2001), 'Modeling Non-ignorable Attrition and Measurement Error in Panel Surveys: An Application to Travel Demand Modeling', in R.M. Groves, D. Dillman, J.L. Eltinge and R.J.A. Little (eds), *Survey Nonresponse*, New York: Wiley, 373–388.
- DHUTS (2010), 'Dhaka Urban Transport Network Development Study', Draft Final Report, Prepared by Katahira and Engineers International, Oriental Consultants Co. Ltd., and Mitsubishi Research Institute, Inc.
- DITS (1993), 'Greater Dhaka Metropolitan Area Integrated Transport Study', Prepared by PPK Consultants Declan International and Development Design Consultant (DDC), Dhaka.
- Enam, A. (2010), *Developing a Comprehensive Mode Choice Model to Capture the Preferences for Mass Rapid Transit in Dhaka*, M.Sc. Thesis, Department of Civil Engineering, Bangladesh University of Engineering and Technology, Dhaka.
- Enam, A. and Choudhury, C. (2011), 'Methodological Issues in Developing Mode Choice Models for Dhaka City, Bangladesh', *Transportation Research Record*, **2239**, 84–92.
- Gaudry, M. and Dagenais, M. (1979), 'The Dogit Model', *Transportation Research Record*, **13B** (2), 105–111.
- Habib, K.M.N. (2002), *Evaluation of Planning Options to Alleviate Traffic Congestion and Resulting Air Pollution in Dhaka City*, M.Sc. Thesis, Department of Civil Engineering, Bangladesh University of Engineering and Technology, Dhaka.
- Hasan, S. (2007), *Development of a Travel Demand Model for Dhaka City*, M.Sc. Thesis, Department of Civil Engineering, Bangladesh University of Engineering and Technology, Dhaka.
- Manski, C. (1977), 'The Structure of Random Utility Models', *Theory and Decision*, **8** (3), pp. 229–254.
- Ortúzar, J. de D. and Ivelic, A.M. (1987), 'Effects of Using More Accurately Measured Level of Service Variables in the Specification and Stability of Mode Choice Models', *Proceeding 15th PTRC Summer Annual Meeting*, 117–130, PTRC, London, England.
- Ortúzar, J. de D. and Willumsen, L.G. (2001), *Modeling Transport*, 3rd Edition, Chichester, Sussex, England: John Wiley and Sons.
- Steimetz, S.S.C. and Brownstone, D. (2005), 'Estimating Commuters Value-of-Time with Noisy Data: a Multiple Imputation Approach', *Transportation Research B: Methodological*, **36**, 865–889.
- STP (2005), 'Strategic Transport Plan for Dhaka', Prepared by Louis Berger Group and Bangladesh Consultant Ltd.
- Swait, J. (1984), 'Probabilistic Choice-Set Generation on Transportation Demand Models', Ph.D. Thesis, Department of Civil Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts.
- Swait, J. and Ben-Akiva, M.E. (1985), 'Constraints on Individual Travel Behavior in a Brazilian City', *Transportation Research Record, Journal of the Transport Research Board*, **1085**, 75–85.
- Train, K. (1980), 'A Structured Logit Model of Auto Ownership and Mode Choice', *Review of Economic Studies*, **XLVII**, 357–370.

- Walker, J., Li, J., Srinivas, S. and Bolduc, D. (2008), 'Travel Demand Models in the Developing World: Correcting for Measurement Errors', *Transportation Research Board Annual Meeting*, Washington DC, USA.
- Wermuth, M. (1978), 'Structure and Calibration of Behavioral and Attitudinal Binary Choice Model Between Public Transport and Private Car', *PTRC Summer Annual Meeting*, University of Warwick, England.
- Williams, H. and Ortúzar, J. (1979), 'Behavioral Travel Theories, Model Specification and the Response Error Problem', Working Paper 116, Institute for Transport Studies, The University of Leeds, England.