Chapter III

Research Framework and Theoretical Background

This chapter provides detailed descriptions of the research framework that will be used to achieve the research objectives as described in the previous chapter. In general, this chapter explains several main tasks that performed in this study. Theoretical background for model development and ridership analysis was discussed as well. Lack of mass transit system and other causes of lower-than-expected ridership are discussed to shed light on potential improvement strategies. Detail description on data collection and preliminary analysis was discussed in the next chapter.

3.1 Research Framework

The critical objective of this research is to shed more light on the role of walking accessibility on transit ridership as well as how station characteristics are translated into walking accessibility. As defined in the first chapter, accessibility can be seen as the interaction result between transit elements and individuals who attempts to use the system [31]. Therefore, how elements on the walkways and station can support accessing trip that lead to the use of mass transit is the key to prescribing public policies then can help to attract more passengers. With these considerations in mind, the research framework can be designed as shown in the Figure 3.1.

The literature review had illustrated that many factors influence transit use, and walking accessibility is one major factor. However, those factors are examined mostly in developed countries. Detailed investigation need to be carried out to shed more light on such factors in developing countries, such as those in South East Asia where there are new development of mass transit systems. Although there may be several shared factors in developed and developing worlds, specific characteristics might be quite distinct and lead to different policy prescriptions. The purpose of the comparative study is to evaluate the effectiveness of mass transit system in the study areas, namely Bangkok and Manila, comparing with the system in the established city that is Sapporo.

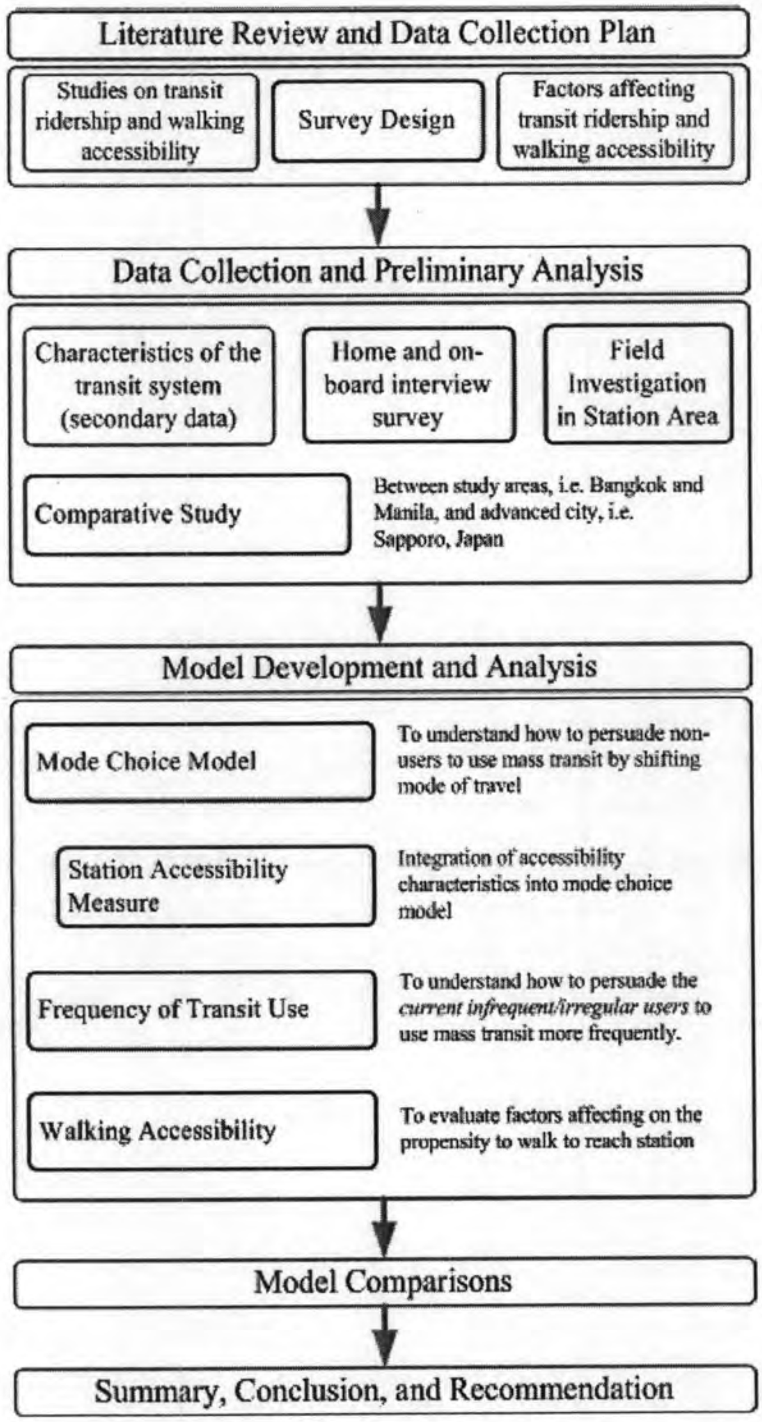

Figure 3.1 Research Framework

Model development was carried out in both study areas of Bangkok and Manila. In those models the roles of accessibility were examined by including variables referring accessibility, such as access distance, in the models. The objective of the evaluation of the frequency of transit use is to understand how to persuade the current users, who are infrequent or irregular users, to use mass transit more frequently. Ordered logit model with four outcomes was used to identify the probability of frequency of mass transit use. Those frequency outcomes were never use mass transit within a week, only once a week, between twice or thrice a week, and more than 4 times a week or almost everyday.

The objective of the development of mode choice model is to understand how to persuade non-users to use mass transit by shifting their modes to mass transit. The multinomial logit, MNL, was used in the model while the nested logit, NL, model was used in case of a violation on the assumption of independence from irrelevant alternatives, IIA. The integration of more accessibility parameters into the mode choice model was carried out by involving the Station Accessibility Measure in the model.

The development of walking accessibility model was intended to understand the role of factors that affect the propensity to walk to reach transit station. The binary logit model was employed with the outcomes of *walking* and *non-walking* mode.

3.2 Theoretical Background

3.2.1 Mode Choice Model

Figure 3.2 shows the general procedure for estimating mode choice model. The choice models were developed based on the results of interview survey. The models were estimated using the data with choice specific-attributes as well as individual-specific characteristics to explain the mode choice. The *multinomial logit* (MNL) model was used as the initial model. The application of *nested logit* (NL) model was carried out if the results from the estimated MNL model violate the IIA assumption. The theoretical model development can be explained as follow.

Suppose in the mass transit corridor, the i^{th} individual faces J choices of mode to travel from home to work while both home and work are located within the mass transit coverage area. The utility of choice $j, j \in J$, is

$$U_{ij} = \mathbf{z}_{ij}\boldsymbol{\beta} + \varepsilon_{ij} \tag{3.1}$$

where $\mathbf{z}_{ij}$ is vector of explanatory variable and $\boldsymbol{\beta}$ is the vector of estimable parameters.

The probability of that choice j being chosen:

$$\Pr(U_{ij} > U_{ik}) \text{ for all other } k \neq j. \tag{3.2}$$

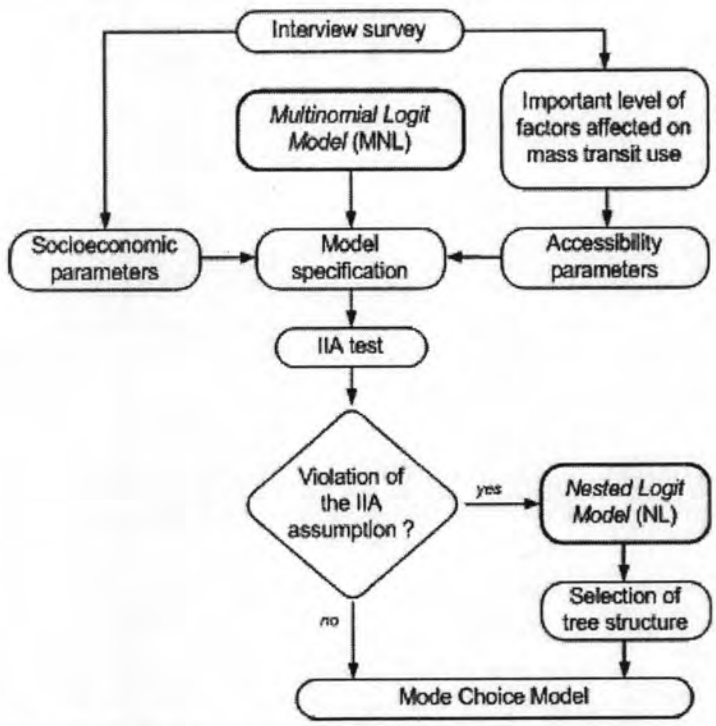

Figure 3.2 Procedure in Developing of Model Share Model

The model is made operational by a particular choice of distribution for the disturbances. Let Y_i be a random variable that the choice is chosen. If (and only if) the J disturbances are independent and identically distributed with Gumbel distribution, i.e.:

$$F(\varepsilon_{ij}) = \exp\left(-e^{-\varepsilon_{ij}}\right) \tag{3.3}$$

then

$$\Pr(Y_i = j) = \frac{e^{\mathbf{z}'_{ij}\boldsymbol{\beta}}}{\sum_{j=1}^{J} e^{\mathbf{z}'_{ij}\boldsymbol{\beta}}} \tag{3.4}$$

In the case of data consists of choice-specific attributes in addition to individual-specific characteristics, the appropriate model is,

$$\Pr(Y_i = j \mid \mathbf{z}_{i1}, \mathbf{z}_{i2}, \ldots, \mathbf{z}_{iJ}) = \frac{e^{\boldsymbol{\beta}'\mathbf{z}_{ij}}}{\sum_{j=1}^{J} e^{\boldsymbol{\beta}'\mathbf{z}_{ij}}} \tag{3.5}$$

The model express in (3.5) has been called by different names, including *conditional logit model* [36] or *multinomial logit* (MNL) *model* [37] or *categorical logit model* (statistic software of Stata).

Note that the odd ratios in the conditional logit model are independent of other irrelevant alternatives. This property is convenient as regards to estimation, but it is not realistic restriction to place on consumer behavior. The property of the logit model where P_j/P_k is independent of the remaining probabilities is called the *independence from irrelevant alternatives, IIA*. Hausman's specification test can be conducted to test the assumption of IIA [36]. It posits that if a subset of the choice set truly is irrelevant, omitting it from the model altogether will not change parameter estimates systematically. The statistic for the test is,

$$\chi^2 = \left(\hat{\boldsymbol{\beta}}_s - \hat{\boldsymbol{\beta}}_f\right)' \left[\hat{\mathbf{V}}_s - \hat{\mathbf{V}}_f\right]^{-1} \left(\hat{\boldsymbol{\beta}}_s - \hat{\boldsymbol{\beta}}_f\right) \tag{3.6}$$

where s indicates the estimators based on the *restricted subset*; f indicates the estimators based on the full set of choices; $\hat{\mathbf{V}}_s$ and $\hat{\mathbf{V}}_f$ are the respective estimates of the asymptotic covariance matrices. The statistics has a limiting chi-squared distribution with K (the number of choices in restricted model) degrees of freedom. If the independence from irrelevant alternatives hypothesis is rejected, Greene [36] suggested grouping the alternatives into subgroups that allow the variance to differ across the groups while maintaining the IIA assumption within the groups. This specification results is a *nested logit model* and will be subsequently discussed.

There are specific scale parameters for each trunk, limb, branch, and twig or elemental alternative, in the *nested logit model*. The notation recommended by Louviere et al., 2000 (as cited in [37) is as follow:

- trunk: $\tau_{(l)}$ → scale parameter for the l^{th} trunk.
- limb: $\gamma_{(i,l)}$ → scale parameter for the i^{th} limb of trunk l,
- branch: $\lambda_{(j|i,l)}$ → scale parameter for the j^{th} branch of limb i of trunk l, and
- twig or alternatives: $\mu_{(j|i,l,k)}$ → scale parameter for the k^{th} elemental alternative of branch j of limb i of trunk l.

Note that the scale parameters must be equal for each alternative present within any given branch. Thus, the subscript notation k is redundant and it is omitted for now on. An example of tree structure for two branches and five choices can be seen in Figure 3.3.

The general utility function for Level 1 is

$$V_J = \mu_{(j|i,l)}\beta_{0J} + \sum_{m=1}^{M} \mu_{(j|i,l)}\beta_{mJ} f\left(\mathbf{X}_{mJ}\right) \tag{3.7}$$

where m is the number of alternative attributes. For Level 2, in case where no variables are defined in the branch level, the utility function is:

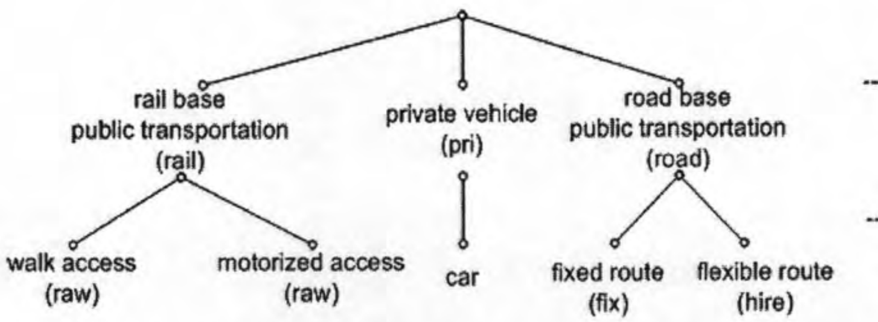

Figure 3.3 Tree Structure of Two Branches and Five Choices

$$V_{(j|i,l)} = \frac{\lambda_{(j|i,l)}}{\mu_{(j|i,l)}} \times IV_{(j|i,l)} \tag{3.8}$$

where the *inclusive value*, IV, of the j^{th} branch: $IV_{(j|i,l)} = \ln \sum e^{(\mu_{(j|i,l)} V_j)}$.

There are various ways to determine IV such as restricting the scale parameter of μ or λ by normalizing them to one, allowing unrestricted the scale parameter of μ or λ, or just fixing the IV to be any arbitrary number. When scale parameters are not normalized it is called as non normalized nested logit (NNNL) model. In this research, when only the scale parameter of μ is normalized the model called random utility model 1 (RU1), whereas when only λ is normalized, the resulting model is called random utility model 2 (RU2).

According to the tree structure in Figure 3.3, the specification of the utility function and calculation of the probability of each mode can be described as follow. Let the total travel time (*ttime*) and the total travel cost (*tcost*) be general attributes of the alternative modes, i.e. *raw*, *ram*, *car*, *fix*, and *hire*. The parameter of car ownership (*owncar*) and distance from home to station (*acdist*) are the specific attributes to the alternative modes.

The utility function of Level 1 (mode of *hire* is a base mode):

$$\begin{aligned} V_{raw} &= \mu_{raw}\beta_{0raw} + \mu_{raw}\beta_T ttime + \mu_{raw}\beta_C t\cos t + \mu_{raw}\gamma_c owncar + \mu_{raw}\gamma_a acdist \\ &= \mu_{raw} \cdot \mathbf{V}_{raw} \end{aligned}$$

$$\begin{aligned} V_{ram} &= \mu_{ram}\beta_{0ram} + \mu_{ram}\beta_T ttime + \mu_{ram}\beta_C t\cos t + \mu_{ram}\gamma_c owncar + \mu_{ram}\gamma_a acdist \\ &= \mu_{ram} \cdot \mathbf{V}_{ram} \end{aligned}$$

$$\begin{aligned} V_{car} &= \mu_{car}\beta_{0car} + \mu_{car}\beta_T ttime + \mu_{car}\beta_C t\cos t + \mu_{car}\gamma_c owncar + \mu_{car}\gamma_a acdist \\ &= \mu_{car} \cdot \mathbf{V}_{car} \end{aligned}$$

$$V_{fix} = \mu_{fix}\beta_{0fix} + \mu_{fix}\beta_T ttime + \mu_{fix}\beta_C t\cos t + \mu_{fix}\gamma_c owncar + \mu_{fix}\gamma_a acdist$$
$$= \mu_{fix} \cdot \mathbf{V}_{fix}$$

$$V_{hire} = \mu_{hire}\beta_T ttime + \mu_{hire}\beta_C t\cos t$$
$$= \mu_{hire} \cdot \mathbf{V}_{hire}$$

According to the IID assumption, $\mu_{(j|i,l)}$ is equal for all elemental alternatives within same branch. Thus, $\mu_{fix} = \mu_{hire}$ and $\mu_{raw} = \mu_{ram}$

In the case of *random utility model 1* (RU1), the scale parameter, μ, is normalized, $\mu_{(j|i,l)} = 1$. Then, $\mu_{bus} = \mu_{van} = \mu_{raw} = \mu_{ram} = \mu_{car} = 1$.

$$V_{raw} = \beta_{0raw} + \beta_T ttime + \beta_C t\cos t + \gamma_c owncar + \gamma_a acdist = \mathbf{V}_{raw}$$

$$V_{ram} = \beta_{0ram} + \beta_T ttime + \beta_C t\cos t + \gamma_c owncar + \gamma_a acdist = \mathbf{V}_{ram}$$

$$V_{car} = \beta_{0car} + \beta_T ttime + \beta_C t\cos t + \gamma_c owncar + \gamma_a acdist = \mathbf{V}_{car}$$

$$V_{fix} = \beta_{0fix} + \beta_T ttime + \beta_C t\cos t + \gamma_c owncar + \gamma_a acdist = \mathbf{V}_{fix}$$

$$V_{hire} = \beta_T ttime + \beta_C t\cos t = \mathbf{V}_{hire}$$

In the case of *random utility model 2* (RU2), the scale parameter, λ is normalized, with equal μ under the same branch. Thus, $\mu_{raw} = \mu_{ram} = \mu_{rail}$;$\mu_{fix} = \mu_{hire} = \mu_{road}$; and $\mu_{car} = \mu_{pri}$.

$$V_{raw} = \mu_{rail} \cdot \left(\beta_{0raw} + \beta_T ttime + \beta_C t\cos t + \gamma_c owncar + \gamma_a acdist\right) = \mu_{rail} \cdot \mathbf{V}_{raw}$$

$$V_{ram} = \mu_{rail} \cdot \left(\beta_{0ram} + \beta_T ttime + \beta_C t\cos t + \gamma_c owncar + \gamma_a acdist\right) = \mu_{rail} \cdot \mathbf{V}_{ram}$$

$$V_{car} = \mu_{car} \cdot \left(\beta_{0car} + \beta_T ttime + \beta_C t\cos t + \gamma_c owncar + \gamma_a acdist\right) = \mu_{car} \cdot \mathbf{V}_{car}$$

$$V_{fix} = \mu_{road} \cdot \left(\beta_{0fix} + \beta_T ttime + \beta_C t\cos t + \gamma_c owncar + \gamma_a acdist\right) = \mu_{road} \cdot \mathbf{V}_{fix}$$

$$V_{hire} = \mu_{road} \cdot \left(\beta_T ttime + \beta_C t\cos t\right) = \mu_{road} \cdot \mathbf{V}_{hire}$$

Since no variables are specified in the branch level, the utility functions in the level 2 are:

$$V_{(j|i,l)} = \frac{\lambda_{(j|i,l)}}{\mu_{(j|i,l)}} \times IV_{(j|i,l)}$$

In RU1:

$$\mu_{rail} = 1 \rightarrow V_{rail} = \frac{\lambda_{rail}}{1} \times IV_{rail} = \lambda_{rail} \times \ln\left(e^{\mathbf{V}_{raw}} + e^{\mathbf{V}_{ram}}\right)$$

$$\mu_{pri} = 1 \rightarrow V_{pri} = \frac{\lambda_{pri}}{1} \times IV_{pri} = \lambda_{pri} \times \ln\left(e^{\mathbf{V}_{car}}\right) = \lambda_{pri} \times \mathbf{V}_{car}$$

$$\mu_{road} = 1 \rightarrow V_{road} = \frac{\lambda_{road}}{1} \times IV_{road} = \lambda_{road} \times \ln\left(e^{V_{fix}} + e^{V_{hire}}\right)$$

In RU2:

$$\lambda_{rail} = 1 \rightarrow V_{rail} = \frac{1}{\mu_{rail}} \times IV_{rail} = \frac{1}{\mu_{rail}} \times \ln\left(e^{\mu_{rail} V_{raw}} + e^{\mu_{rail} V_{ram}}\right)$$

$$\lambda_{pri} = 1 \rightarrow V_{pri} = \frac{1}{\mu_{pri}} \times IV_{pri} = \frac{1}{\mu_{pri}} \times \ln\left(e^{\mu_{pri} V_{car}}\right) = \mathbf{V}_{car}$$

$$\lambda_{road} = 1 \rightarrow V_{road} = \frac{1}{\mu_{road}} \times IV_{road} = \frac{1}{\mu_{road}} \times \ln\left(e^{\mu_{road} V_{fix}} + e^{\mu_{road} V_{hire}}\right)$$

Then, the probability of each alternative modes can be calculated as follow.

$$\Pr(raw) = \Pr(raw \mid rail) \cdot \Pr(rail) = \frac{e^{V_{raw}}}{e^{V_{raw}} + e^{V_{ram}}} \times \frac{e^{V_{rail}}}{e^{V_{rail}} + e^{V_{pri}} + e^{V_{road}}}$$

$$\Pr(ram) = \Pr(ram \mid rail) \cdot \Pr(rail) = \frac{e^{V_{ram}}}{e^{V_{bus}} + e^{V_{van}}} \times \frac{e^{V_{rail}}}{e^{V_{rail}} + e^{V_{pri}} + e^{V_{road}}}$$

$$\Pr(car) = \Pr(car \mid pri) \cdot \Pr(pri) = \frac{e^{V_{pri}}}{e^{V_{rail}} + e^{V_{pri}} + e^{V_{road}}}$$

$$\Pr(fix) = \Pr(fix \mid road) \cdot \Pr(road) = \frac{e^{V_{fix}}}{e^{V_{fix}} + e^{V_{hire}}} \times \frac{e^{V_{road}}}{e^{V_{rail}} + e^{V_{pri}} + e^{V_{road}}}$$

$$\Pr(hire) = \Pr(hire \mid road) \cdot \Pr(road) = \frac{e^{V_{hire}}}{e^{V_{fix}} + e^{V_{hire}}} \times \frac{e^{V_{road}}}{e^{V_{rail}} + e^{V_{pri}} + e^{V_{road}}}$$

Some discussion related to the development of the *nested logit* models for the case of Manila in this research are as follow. Greene [36] recommended that in order to specify the nested logit model, it is necessary to partition the choice set into braches. However, there is no well-defined testing procedure for selecting appropriate tree structures.

Koppelman and Bhat [38] gave a good example in the selection of the tree structure for the case of BART Area (San Francisco). However, in the end, the decision of the appropriate choice of structure must be backed by the researchers' own justification.

Fillone [28] carried out *nested logit model* for urban transportation in Metro Manila. He performed *multinomial logit model* and *nested logit model* with 2 and 3 level. In that study, the statistical software program NLOGIT 3.0 was used to develop the model.

3.2.2 Station Accessibility Measure

There are many aspects of station characteristics can be considered influential mass transit use and the propensity to walking to access station. The role of these factors particularly in South East Asia was not yet explained clearly in literature reviewed. As found in the field investigation of stations and the vicinity, it is difficult to compare one aspect with another in order to evaluate their relative roles. For instance, both of walking distance and provision of escalator can influence the effort of walking to access station. The station accessibility measure was developed in attempt to develop a combine framework for the various station characteristics.

The multi criteria analysis technique (MCA) can be employed to develop station accessibility measures. One advantage of MCA is that it can combine various aspects that differ in units of measurement and characteristics with distinct assessment method. The other advantage is that, in the process of evaluation, various different factors can be combined into an assessment model with logical soundness, transparency, and ease to use. Similar works related to evaluation of quality level, for instance, is presented in Krambeck [39] who developed walkability index for cities around the world. The author used 22 indicators which were derived into 45 variables. To obtain clear distinction among the cities, a weighting analysis related to components and indicators was used. The weight factors were developed in different ways by using the point of view of transportation experts and stakeholders.

Zhu et al. [40] developed a multi-criteria framework for accessibility analysis for housing development from the buyer's perspective in Singapore. The researchers employed the multi-criteria analysis of characteristics which are of concerns by most respondents in selecting housing location and then rank them in terms of accessibility. GIS application was employed to calculate accessibility level of each parameter. Another work is by Tindale-Oliver [41] who applied the multi-criteria analysis to evaluate the proposed bus rapid transit corridors for Pinellas County, Florida. There were five objectives used to select the best corridors, while each objective has different criteria and measurement method. Threshold levels were calibrated for each criterion. Weight factors were then developed from government officials and stakeholders' opinion.

In this research, the development of station accessibility measure was adopted the work done by Krambeck [39] and Tindale-Oliver [41]. There are two main components (objectives) were considered in the development of station accessibility score, i.e. station access and station facilities with different sets of variables employed to explain each component. Station accessibility score is defined as the average value of access score and facility score. Those scores are expressed as follow.

$$Access\ Score = \sum_{i=1}^{k} w_i * r_i \tag{3.9}$$

$$Facility\ Score = \sum_{j=1}^{l} w_j * r_j \tag{3.10}$$

where: $k; l$ = no. of criteria for station access and station facility

$w_i; w_j$ = weight representing relative importance of related criterion

Note that:

$$w_i = 100 * \frac{sum_i}{\sum sum_i}$$

$$sum_i = n_{1i} + 2 \cdot n_{2i} + 3 \cdot n_{3i} + 4 \cdot n_{4i}$$

$n_{1i \ldots 4i}$ = number of respondents who rated the factor *i* as *not important* (valued as 1), *somewhat important* (2), *important* (3), and *very important* (4), respectively.

Similar formula for w_j.

$r_i; r_j$ = rating point related to observed variables.

Station access score focuses on components related to trip to reach the station gate such as parking facilities, bus stop, provision of escalator and elevator, etc. On the other hand, station facility score focus on the facilities provided inside the station building such as toilet, chairs, security, and other station amenities. According to equation (3.9) and (3.10), the higher score indicates better conditions for using transit stations.

Unlike the studies which were adopted for the station accessibility score development, the weightings of the score are derived from travelers' point of view. In the interview survey, respondents were asked to rate the importance level of factors that affect of their behavior to use mass transit. The levels of '*not important*', '*somewhat important*', '*important*', and '*very important*' were used. The numbers of 1, 2, 3, and 4 were assigned to represent these levels as the factors' importance and the weight was derived by normalizing the value by the sum of factor's value.

The rating of 1, 3, and 5 were employed to articulate the threshold level of each variable. The rating of 5 is for the best condition the term of accessibility while the rating of 1 is for the worst one. The moderate condition is expressed by the rating of 3. Table 3.1 shows the criteria and rating system that were used in the development of the station accessibility measure.

Table 3.1 Criteria and Rating System in Station Accessibility Measures

Criteria	Description	Rating System		
		5	3	1
Related to Station Accessibility				
Road crossing	Number of road crossing on the way to station	≤ 4	5 to 14	more than 14
Escalator/Elevator	Presenting of escalator or elevator	at least one	-	none
Access Mode	Availability of other feeder modes	more than two transit routes per gate	-	not all gates have
Car Park	Car park facilities provided by mass transit authority	provided	-	not provided
Egress Mode	Number of availability transit modes from station	more than two transit routes per gate	-	not all gates have
Destination	Number of destination near station	more than two transit routes per gate	-	not all gates have
Related to Station Facility				
Security	Security precaution in station and train	present of sta. guard with enough number	present of sta. guard but not enough	no sta. guard
Crowdedness	Level of crowdedness in station area	not too crowded in peak hour	-	too crowded
Cleanliness	Level of cleanliness in station area	good condition	average	poor
Ticketing System	Number of ticketing machines	presented in every gate	-	not all gates have

It can be seen in the table that the threshold level of road crossing is 4 and 14 roads to cross. This number was derived from the studies of Singapore Mass Rapid Transit [8]. In that study, it was found that about 90% of the respondents walked less than the distance of 200 meters while about 10% walked for the distance of over 800 meters. It was revealed as well that the effort to cross road with two-lanes or wider equals to the effort to the walking distance of 55.4 meters. Therefore, the distance of 200 (equivalent to the effort to cross 4 roads) and 800 meters (to 14 road crossings) were used to define the threshold effort of walking due to cross the road.

3.2.3 Frequency of Transit Use

The objective of the evaluation is to investigate frequency of transit use of people who live within mass transit coverage service area. Based on the results in previous studies, infrequent mass transit user possess larger share of transit use and attracting them will generate greater ridership. The dependent variables constructed in the evaluation are categorical outcomes, i.e. never user (never use mass transit within a week), occasional (only once a week), frequent (between twice or thrice week), and regular (more than 4 times or almost everyday); therefore the ordinary least squared regression is not appropriate for this case. Since the outcomes have ordered from never to regular use, application of multinomial logit model will not be efficient because some information about the order of outcomes is not utilized. Hence, the ordered logistic model was used to evaluate the frequency of transit use.

Guiliano [42] identified several factors to explain frequency of transit use. Their list of variables was modified to fit this research such as mode attributes (especially total travel time), demographic parameters (gender, age, education level, car ownership, etc.), and transit accessibility parameters (e.g. access distance). Under the discrete choice framework, the probability of an individual become frequent user can be written as follow.

$$\Pr(\textit{frequent}) = \frac{1}{1+e^{-f(d,\mathrm{S},\mathrm{T},\mathrm{A})}} \qquad (3.11)$$

where: d = deterrence parameters such as time or cost,
S = vector of traveler's socioeconomic characteristics,
T = vector of trip characteristics,
A = vector related to accessibility

It could be said that beside time and cost, socioeconomic factors, characteristics of trip and quality of accessibility between homes to station play important roles in making an individual become regular user of mass transit.

In the *ordered logit* model, the outcomes of never, occasional, frequent, and regular are used as dependent variables of the model. Let y* denotes an unobserved score that determines ordered the outcome of observation.

$$y^* = \mathbf{x}\boldsymbol{\beta} + \varepsilon \qquad (3.12)$$

where: $\mathbf{x}$ = vector of characteristics; attributes data of observation n
$\boldsymbol{\beta}$ = vector of estimable parameters

ε = random error

The ordered outcome can be observed as follow:

$y = 0$ if $y^* \leq 0$

$y = 2$ if $0 \leq y^* \leq \mu_1$

$y = 3$ if $\mu_1 \leq y^* \leq \mu_2$

...

$y = J$ if $\mu_{J-1} \leq y^*$

where the μ are unknown parameters to be estimated together with $\boldsymbol{\beta}$. Depending on assumption of error term, ε, the general form of probability model can be expressed as

$$\begin{aligned}\Pr(y=j) &= \Pr(\mu_{j-1} < y^* \leq \mu_j) \\ &= \Pr(\mu_{j-1} < \mathbf{x\beta} + \varepsilon \leq \mu_j) \\ &= \Pr(\varepsilon < \mu_j - \mathbf{x\beta}) - \Pr(\varepsilon \leq \mu_{j-1} - \mathbf{x\beta}) \\ &= \mathbf{F}(\mu_j - \mathbf{x\beta}) - \mathbf{F}(\mu_{j-1} - \mathbf{x\beta})\end{aligned}$$

where $\mathbf{F}(\bullet)$ is cumulative distribution function

By the assumption that the error terms are Gumbel distributed, the *ordered logit* model can be estimated and the probability of an individual belong to the categories of never users, occasional, frequent, and regular user can be written as follow.

$$\begin{aligned}&\Pr(y = 1) = \Pr(\text{never}) = \mathbf{F}(\mu_1 - \mathbf{x\beta}) \\ &\Pr(y = 2) = \Pr(\text{occasional}) = \mathbf{F}(\mu_2 - \mathbf{x\beta}) - \Phi(\mu_1 - \mathbf{x\beta}) \\ &\Pr(y = 3) = \Pr(\text{frequent}) = \mathbf{F}(\mu_3 - \mathbf{x\beta}) - \Phi(\mu_2 - \mathbf{x\beta}) \\ &\Pr(y = 4) = \Pr(\text{regular}) = 1 - \mathbf{F}(\mu_3 - \mathbf{x\beta}) \qquad (3.13)\end{aligned}$$

Maximum likelihood estimation (MLE) method was performed to estimate parameters of the model and the software package of Stata/SE 8.2 for Windows was employed.

3.2.4 Walking Accessibility

As discussed extensively in the literature, especially work done by [23], [11], and [8], the propensity to walk to a mass transit station depends on many factors, besides distance. These factors can be classified into users' socioeconomic characteristics and accessibility characteristics along path to reach station. In the evaluation of propensity to walk, role of station characteristics is examined, especially for the case of developing countries. Similarly to reviewed literature, the discrete choice framework was employed. The probability of an individual to walk to reach station is

$$\Pr(walk) = \frac{1}{1 + e^{-f(dist, \mathbf{S}, \mathbf{St})}} \tag{3.14}$$

where: dist = distance between home (originated point) and station,
S = vector of traveler's socioeconomic characteristics,
St = vector related to characteristics of station area,

Socioeconomic characteristics that influence users to walk to reach station may include gender, age, occupation, car ownership, income, and so on. The transit station characteristics may include land use of the surrounding area, as well as the street network characteristics, which can affect the choice of walking to the station.