ESTIMATING, CALIBRATING AND VALIDATING DEMAND FOR FEEDER SERVICES DURING INCIDENT DETECTIONS.

Syahriah BACHOK
PhD student
Transport Systems Centre
University of South Australia
GPÓ Box 2471, Adelaide
South Australia 5001, Australia
Phone +61 8 8302 1778
Fax +61 8 8302 1880
E-mail: syahriah.bachok@postgrads.unisa.edu.au

Abstract:

This paper discusses the outcomes of stated choice surveys undertaken on-board suburban commuter rail services. Public transport users were provided with hypothetical information regarding delay durations and alternative modes disseminated via different media. Users’ responses, particularly the stated demand for supplementary feeder services, were analysed for the purpose of developing sub-modal split models during incident detections. The stated preference surveys were preferred because currently there exists neither an integrated public transport information system nor an integrated feeder service for the public transport in the Study Area.

Relevant parameters such as socio-demographic and trip characteristics have been estimated using multinomial Logit models. The analysis utilised statistical and multinomial logit software packages such as Microsoft Excel, SPSS and NLogit (LIMDEP). The employed mode choice models included variables such as age, gender, vehicle ownerships, household characteristics, trip purposes, travel time, travel costs and trip length. Inclusion of non-conventional (latent) factors namely “repeated or historical behavioural (mode) choices”, “receiving delay information during previously interrupted trips” and the concept of loss aversion have been considered for the models. Results from the estimation, calibration and validation runs are presented and compared to produce most suitable models representing public transport travellers’ mode switching behaviour in the events of rail incidents.

Keywords: feeder services, multinomial Logit modelling, incident, information.
1. INTRODUCTION

Improvements in public transport are often measured collectively by several methodologies including stated preference surveys. It is often difficult to single out one particular improvement and its impact on ridership or patronage rates. This paper discusses the potential impacts of improving information system, in particular, one that could be provided during incidents or after an incident being detected.

This paper deals with the outcomes of stated choice surveys undertaken on-board suburban commuter rail services. Public transport users were provided with hypothetical information regarding delay durations and alternative modes disseminated via different media. Users’ responses, particularly the stated demand for supplementary feeder services, were analysed for the purpose of developing sub-modal split models during incident detections. The stated preference surveys were preferred (Hensher, 1996) because currently there exists neither an integrated public transport information system nor an integrated feeder service for the public transport in the Klang Valley region of Malaysia, the Study Area.

Hypothetical scenarios are utilised to extract rail passengers’ decision making process during rail service disruptions. The stated preference surveys conducted were aimed at rail passengers with available and accessible choices of alternative modes when the events occurred. Cluster sampling and systematic selection of respondents in the samples were adopted. These users were approached at termini or stations with possible interchanging or transferring activities so that choices of mode would include those prescribed by the information as well as passengers’ own alternative vehicles.

It was timely that modelling efforts of feeder services demand were conducted. The potential demand for the currently disintegrated feeder bus services could be fed into the planning and dispatching of these vehicles (as part of emergency and crowd management). Integration of rail and bus feeder services provided greater mobility options for travellers. Higher riding habits of public transport system were evident in the United States where multi-destination bus-and-rail integration was introduced (Brown and Thompson, 2008).

The research attempted to model the changes in mode decision making through models development, parameter estimates, models calibration and validation. The experiences of estimating, calibrating and validating the dataset will be discussed. The issues and challenges faced by the analyst in the three exercises are also elaborated. The conclusions drawn reflect the uniqueness of the case study which may be applicable to other developing cities with similar public transport services.

2. AIM AND OBJECTIVES

The research focuses on behavioural changes among public transport passengers as a result of integrated information provision. The information prescribed the alternative modes and routes in the event of an intercity train derailment. The information also contained the reason of the delay and its duration. Two objectives of the research are to capture the mode switching behaviours among rail users upon provision information on alternative transport modes, and attempt to model the mode choices for the continuing journeys which have been disrupted. To ascertain the impacts of
information provision on mode choice decision making amongst the existing rail users, the study adopted primary data collection approach. In particular, the demand for bus feeder services is to be modeled. The behavioural changes among rail passengers as a result of integrated information provision are captured utilising structured pen-and-paper questionnaire.

3. METHODOLOGY

Several stated preference surveys including pilot and actual surveys have been undertaken in 2007. Passengers on-board a suburban commuter rail services originating from stations with transfers facilities or alternative modes were approached to fill in a questionnaire set. Some 537 completed questionnaires were used as valid responses. Hypothetical scenarios regarding different delay durations and information media were presented. The surveys were conducted with rail users, making mode choice decisions under nine scenarios of disseminated information and travel times. The questionnaire consisted of 40 questions. It was divided into three sections. The first section dealt with the trip characteristics. The second section was on the hypothetical scenarios of delay duration and information media. The last section was on the socio-demographic characteristics. Users were given about 15 minutes to complete the surveys on-board the train.

Passengers were found to have difficulties in exacting the travel time, cost and distances. Most of the respondents were not able and willing to express or disclose this information for a variety of reasons. To simplify this, respondents were asked to consider the current trip attributes when answering the questions. Many highlighted the unreliability of public transport services i.e. rail and buses. Taking cognisance of the burdens and the cognitive efforts required to complete the forms, the following assumptions have been put to respondents while answering the questionnaire.

The alternative mode “rail” and “wait” will be associated with full delay duration but at no extra cost. The alternative mode “feeder” and “other buses” will have the costs similar to the current fare rate for feeder buses servicing other rail-based public transport system in the Study Area. The route will be directed to cover areas being effected by the derailment and destined to the nearest unaffected stations. Buses will be associated with half the delay time. “Private and other vehicles” category was assumed to have the full costs from egress station to final destination, in addition to the rail fare. This category was not affected by the delay.

Assumptions of travel time and travel costs were made based on the existing situations. Feeder services linking major landuses to rail-based public transport in Klang Valley were rated at RM2 per trip. Average speed for each alternative, access or egress mode was adopted from studies in Peninsular Malaysia. In the absence of local evidence, Singaporean, Indonesian and Thailand’s case studies (i.e. average speed of car motorcycle and walking trips) were adopted and adapted. There were not many users with bicycles or trishaws as access or egress modes. Boat rides from islands nearby the seaport were assumed to have an average speed as posted on the website and as reported by respondents. Costs were based on circumstances in Malaysia. Fuel price was set at RM2 per liter, and parking at rail stations at RM3 per entry. However, additional costs such as road tolls, wear and tear, maintenance and repair costs were not factored in under the hypothetical scenarios. A point worth
noting is that the majority of respondents have access to a motorcycle which is a relatively cheap transport mode and an easy alternative to various other modes.

4. MODELS DEVELOPMENT

Selection of mode choice models have been contributed by the nature and circumstance of the journeys made by travellers. In this research, public transport users, in particular, rail passengers made their journeys by a suburban commuter rail services. When train services are disrupted, users have to make choices such as alternative mode and routes to complete their trips. The mode choices, under the hypothetical scenarios included among others, rail (waiting on-board for services to resume), supplementary feeder buses and other private vehicles. Based on these, a multinomial Logit (MNL) model was considered appropriate to represent the mode choices (Hensher, Rose and Greene, 2005). In the analysis, the factors or independent variables are assumed to be personal profiles, trip and public transport services characteristics. MNL model subscribes to the notion that a mode will be selected if the utility derived from consumption is the greatest or the disutility is the least. Hence, the following equation of random utility maximisation is adopted:

$$C_{ij} = a + \beta_1 t_{ij} + \beta_2 t'_{ij} + \beta_3 t^p_{ij} + \beta_4 t^n_{ij} + \delta$$

where

- $C_{ij}$ = the (dis)utility derived from a trip made from origin zone $i$ to destination zone $j$ by a particular mode $C$ (say, a feeder bus)
- $t$ = observed factors or parameters associated with the mode (determinant or independent variables)
- $a$ = constant (alternative specific constant)
- $\beta$ = coefficient of parameters
- $\delta$ = unobserved parameters

The attempt of the research is to model the proportion ($P_{ij}$) of users or trip made using feeder buses. Under MNL, the proportion of trips by feeder is given by this algorithm:

$$P^1_{ij} = \frac{T^1_{ij}}{T_{ij}} = \frac{e^{\beta_{1ij}}}{e^{\beta_{1ij}} + e^{\beta_{2ij}} + e^{\beta_{3ij}}}$$

where:

- $T^1_{ij}$ = number of trips made from zone $i$ to zone $j$ using feeder (mode 1)
- $T_{ij}$ = total trips made from zone $i$ to zone $j$

For the research purpose, it was assumed that the case study area was a single zone.

Relevant parameters such as socio-demographic and trip characteristics have been estimated using multinomial Logit models. Three software packages were utilised to analyse and model the mode selections. The objective was not to assess the capability
of each package but rather to complement and verify each other’s results. The analysis utilised statistical and multinomial Logit software packages such as Microsoft Excel, SPSS and NLogit (LIMDEP). The employed mode choice models included variables such as age, gender, vehicle ownerships, household characteristics, trip purposes, travel time, travel costs and trip length. Inclusion of non-conventional (latent) factors namely “repeated or historical behavioural (mode) choices”, “receiving delay information during previously interrupted trips” and the concept of loss aversion have been considered for the models.

Raw data entered as SPSS and MsExcel data. They included nine dependent variables. These are choices made by respondents on different occasions or scenarios. The percentages of respondents opting for feeder were disproportionate as in Table 1.

Table 1: Percentage of respondents selecting feeder by media type and delay duration.

<table>
<thead>
<tr>
<th>Feeder choice (%)</th>
<th>Media of information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delay duration</td>
<td>Audio</td>
</tr>
<tr>
<td>60min</td>
<td>30.54</td>
</tr>
<tr>
<td>45min</td>
<td>25.30</td>
</tr>
<tr>
<td>30min</td>
<td>14.71</td>
</tr>
</tbody>
</table>

The dataset has 34 independent variables. Some of these variables are run through factor analysis (principle component analysis), others were interpolated as input to estimate other variables. Derived variables or estimated variables included time, cost and distances. A GIS-based map produced as hard copy was made a reference for determining the shortest distance between access and egress points (from and to the last known stations where passengers alighted and boarded the trains). In the absence of the routes or if areas were outside the map’s boundary, the Google Earth maps will be referred to. The map and relevant queries can also be accessed at www.naviandmap.com. The shortest path was assumed, but passengers may not necessarily follow this path, for reasons such as comfort, convenient, familiarity and safety.

Figure 1: Microsoft Excel independent variables dataset
The dataset is produced in both MsExcel and SPSS spreadsheet. Microsoft Excel produced the dataset as in Figure 1. The delay durations were selected to be 30 minutes, 45 minutes and 60 minutes. During the survey period, the rail services run at headways of 15 minutes during peak hours. Information media were divided into audio, text and visual representations of delay information and alternative mode availability. The dependent variables are those mode choices selected by respondents under the scenarios, based on a factorial design.

SPSS produced two different windows for the data set, one being variable view (Figure 2) and the other is data view (Figure 3). Excel and SPSS produced uniform attributes for the selected alternative mode by each respondent. Independent variables including socio-demographic characteristics (SDCs), trip characteristics and mode choice attributes are presented in one window.

![Figure 2: Variable view in SPSS](image)

![Figure 3: Data view in SPSS](image)

NLOGIT/LIMDEP may utilise a converted/imported dataset from most spreadsheet programs including MsExcel. The reproduction is presented as a text file and can be
edited through a data editor. NLOGIT produced the choice set and the selected alternative mode with respective attributes separately. In Figure 3 shown here, the attributes were travel time including delays and travel costs of the continuing journeys.

5. PARAMETER ESTIMATES

All the 34 independent variables have undergone the principle component analysis. All the variables were assessed for their correlations, normality and other factor analyses. Sieving techniques, trial and error, combinations of factors and multicollinearity tests were carried out. Finally, only eight variables were selected for model estimation. The variables were gender, age, year of education, income, number of accessible vehicles, access cost, egress cost and revealed preference on mode switching in previous delayed trips. Relative costs of alternative mode and relative travel time of alternative mode are specific to respective alternative modes.

The results are presented for three different software analyses. Excel provides linear regression analysis (Norusis, 2005). This limits the results of estimation to modelling only numerical/interval dependent variable along a linear equation. The signs and degree of coefficients are also doubtful because negative values are associated with number of vehicles accessible to a traveller, while travel time and travel costs are represented by positive signs. Parameters significantly contributing to the estimated models are age, income, egress cost and revealed changes of mode choice. Time and costs of alternative mode are not significant.

The model fit which was represented by the $r$-square ($R^2$) was 0.05, a relatively low value. Parameter estimates were unsatisfactory because of the low model fit. There are two issues with regard to estimating in MsExcel. First, the estimation or regression analysis was made for scale/interval variables while the actual dataset consisted of nominal/categorical dependent variables. Thus, the analyst was unable to estimate the parameters for the models. Second, while calibration process can be
conducted for results from other analysis tool, outputs from MsExcel produced calibration results were with negative probability values. This was taken as illogical and the process was abandoned.

Hence, the analyst looked for a second opinion in the SPSS software package. Using the same dataset, the estimation outputs are produced by SPSS as in Figure 5.

![Figure 5: SPSS estimation outputs](image)

To estimate the parameters, SPSS program required the reference category to be the base comparison category against the alternative groups. However, it is capable to estimate parameters for nominal or categorical dependent variables.

The model fit is, however, unsatisfactory. The $R^2$ was 0.10, slightly better than that of MsExcel but still a relatively low value. Using “feeder” as the reference category, the parameters to be considered in the model, i.e. having $p$-values less than 0.05, were age, egress cost, revealed mode change (feeder-rail referenced group) and vehicle access (feeder-other referenced group). At this juncture, only 45% of the feeder category can be estimated, a relatively low figure if the objective is to model this particular mode choice. Again, time and costs of alternative mode are not significant.

The signs and size of the coefficients also raised some concerns. For example, rail trip costs and rail users’ egress cost have, positive values, instead of the conventional wisdom that costs posed some disutility for the respective mode. SPSS estimation output version was in the form regression analysis on nominal/ordinal variables, which was applicable to MNL. Hence, it was possible to estimate the feeder choice probability. However, it was cautioned that estimation processes using SPSS would require a reference category (base group), i.e. needing category referencing such as feeder to estimate and later calibrate the models. Therefore, it was thought that specialised software could be utilised to improve the parameter estimates exercises.

Dataset for NLOGIT/LIMDEP can be imported from Excel dataset files and converted as a text file for estimation runs. NLOGIT/LIMDEP produced specific cost
and time attributes for each alternative mode. However, cautions need to be taken when dealing with SDCs and alternative specific constants (ASCs) which are uniform throughout the dataset. Estimating runs, therefore made them redundant, specifically for the last “choice” in the dataset, i.e. alternative mode of private vehicles and others category.

![Figure 6: NLOGIT/LIMDEP estimation outputs](image)

Four main parameters were analysed in isolation. The outputs showed that parameters with significant contribution were income and revealed changes of mode choice for rail group. Similar to the previous two analysis tools, total travel time and costs of alternative mode are not significant. The model fit was also disturbingly low ($R^2$ was 0.027). It was worse because no significant parameter can be estimated for “feeder” choice.

Another run to include all ten variables produced a slightly better result. The $R^2$ was 0.04. Parameter with significant contribution to feeder choice was found to be vehicle access. Despite the poor results, the estimation runs provided valuable findings. There are similarities in the factors contributing to the selection of “feeder” mode choice. Factors like age, income, vehicle access, egress cost and revealed mode change were found repeatedly in the different analysis packages.

It seems that specific attributes of the alternative modes such as travel time and travel costs are insignificant in explaining the behavioural changes of rail users. None of the outputs produced by the three softwares indicated that the two variables were significant in determining mode choices.

Chi-square value is 31.15240, critical value at 95% confidence level for 10 degrees of freedom (for LIMDEP) is 18.31 and for 22 degrees of freedom (for SPSS) is 33.93. Since the chi-square value is greater than the critical value, therefore $H_0$ that stated that the specified model is no better than the base comparison model was rejected.
Estimating, calibrating and validating demand for feeder services during incident detections.

Instead, it can be concluded that the alternative model can significantly represent the mode switching behaviour of passenger from rail to feeder buses by the parameters estimated as contributors, albeit the low model fit.

Table 2: Parameter estimations by delay and information scenarios by analysis tools.

<table>
<thead>
<tr>
<th>Scenarios</th>
<th>Excel parameters</th>
<th>LIMDEP parameters (df10=18.31)</th>
<th>SPSS(feeder-rail) parameters (df22=33.93)</th>
<th>SPSS(feeder-other) parameters</th>
<th>Coef.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visual60</td>
<td>$R^2=0.05$</td>
<td>$R^2=0.04$</td>
<td>$R^2=0.10$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Age 0.007</td>
<td>$\chi^2=45.52$</td>
<td>Age 0.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Income 0.00006</td>
<td></td>
<td>Egress Cost 0.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Revealed mode 0.25</td>
<td>Vehicle access 0.17</td>
<td>Revealed mode -1.02</td>
<td>Vehicle access -0.16</td>
<td></td>
</tr>
<tr>
<td>Text60</td>
<td>$R^2=0.06$</td>
<td>$R^2=0.05$</td>
<td>$R^2=0.12$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Age -0.096</td>
<td>$\chi^2=54.22$</td>
<td>Age 0.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Income 0.00006</td>
<td></td>
<td>Income 0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Revealed mode 0.34</td>
<td></td>
<td>Revealed mode -1.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Audio60</td>
<td>$R^2=0.06$</td>
<td>$R^2=0.05$</td>
<td>$R^2=0.12$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Income 0.00007</td>
<td>$\chi^2=54.04$</td>
<td>Income 0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Revealed mode 0.38</td>
<td>Time delay 0.013</td>
<td>Revealed mode -1.20</td>
<td>Age 0.02</td>
<td></td>
</tr>
<tr>
<td>Visual45</td>
<td>$R^2=0.04$</td>
<td>$R^2=0.04$</td>
<td>$R^2=0.09$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Income 0.00007</td>
<td>$\chi^2=42.86$</td>
<td>Income 0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Revealed mode 0.17</td>
<td>Vehicle access 0.17</td>
<td>Revealed mode -0.88</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Text45</td>
<td>$R^2=0.05$</td>
<td>$R^2=0.05$</td>
<td>$R^2=0.11$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Income 0.00007</td>
<td>$\chi^2=49.77$</td>
<td>Income 0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Revealed mode 0.26</td>
<td></td>
<td>Revealed mode -0.91</td>
<td>Year education -0.10</td>
<td></td>
</tr>
<tr>
<td>Audio45</td>
<td>$R^2=0.05$</td>
<td>$R^2=0.05$</td>
<td>$R^2=0.11$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Income 0.00007</td>
<td>$\chi^2=48.70$</td>
<td>Income 0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Revealed mode 0.26</td>
<td>Time delay 0.014</td>
<td>Revealed mode -1.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Visual30</td>
<td>$R^2=0.05$</td>
<td>$R^2=0.05$</td>
<td>$R^2=0.10$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Income 0.00006</td>
<td>$\chi^2=42.71$</td>
<td>Income 0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Year education 0.033</td>
<td>Vehicle access 0.23</td>
<td>Year education -0.10</td>
<td>Income 0.000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Revealed mode 0.26</td>
<td>Time delay 0.014</td>
<td>Revealed mode -1.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Text30</td>
<td>$R^2=0.03$</td>
<td>$R^2=0.06$</td>
<td>$R^2=0.11$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Income 0.00004</td>
<td>$\chi^2=38.20$</td>
<td>Income 0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Year education 0.033</td>
<td></td>
<td>Year education -0.16</td>
<td>Revealed mode -0.70</td>
<td></td>
</tr>
<tr>
<td>Audio30</td>
<td>$R^2=0.03$</td>
<td>$R^2=0.06$</td>
<td>$R^2=0.12$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Income 0.00004</td>
<td>$\chi^2=39.87$</td>
<td>Income 0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gender -0.95</td>
<td></td>
<td>Gender -0.96</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cost rail 1.045</td>
<td></td>
<td>Cost Rail 1.018</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Greene, 2005).
Parameter estimations (for feeder only) for other scenarios are depicted in Table 2. The Table presents the findings from all the scenarios, with their respective $R^2$, Chi-square and parameter estimates. The results were in stark contrast to those in the available body of literature. They are so unconventional that the analyst failed to make any worthwhile conclusion at this point.

An advantage of LIMDEP is that datasets can be imported from spreadsheets e.g. *.*XLS and its capability to undertake regression analysis on nominal/ordinal variables. It was quite challenging to prepare dataset and input command/run estimation because of the need for a full grasp of command writing. Nevertheless, it has an advantage of estimations with alternative specific constant, and specific mode attributes (time and cost variables). Estimation and calibration are, thus, possible.

6. CALIBRATION AND VALIDATION

Results from the estimation, calibration and validation runs are presented and compared to produce most suitable models representing public transport travellers’ mode switching behaviour in the events of rail incidents. The calibration can be done in all analysis tools. The outputs presented here were calculated in MsExcel. The calibration process was undertaken using two-thirds of the observed data, randomly selected. The probabilities of trips made using feeder services, i.e. $P(\text{feeder60V})$, were calibrated at different levels of egress cost (RM0-66) for Excel and SPSS parameter estimates and also at different levels of vehicle access (0-17) for LIMDEP and SPSS.

The following equations are applied for the respective outputs.

For SPSS:

$$P(\text{feeder60V}) = \frac{1}{1 + e^{(0.026\text{age}_{\text{rail}} + 0.094\text{egresscost}_{\text{rail}} - 1.016\text{rev.mode}_{\text{rail}} + (-0.162\text{veh.access}_{\text{others}})}}}$$

For LIMDEP:

$$P(\text{feeder60V}) = \frac{1}{1 + e^{(0.0002\text{income}_{\text{rail}} - 0.85\text{rev.mode}_{\text{rail}} + (0.17\text{veh.access}_{\text{feeder}})}}}$$

By substituting other parameter coefficients with their mean ($\mu$) values, the following graphs of calibrated and validated models for the scenario of 60 minutes delay with visual information are compared. Validating is conducted by comparing the observed or actual mode switching, in particular the selection of feeder mode under different contributing parameters. The selected parameters are vehicle access and egress cost. The remaining one-third of the observed data is used for validating the calibrated models. Great variances and discrepancies are shown by the graphs of calibration and validation as demonstrated in seven figures presented here.

Figures 7 and 8 illustrate the relationship between the probability of selecting feeder and the number of vehicles accessible to a respondent, calibrated based on LIMDEP/NLOGIT and SPSS outputs respectively. Figures 9 and 10 depict the validation or more appropriately the invalidity of the models (cumulative and non-cumulative graphs respectively).
Syahriah BACHOK
Estimating, calibrating and validating demand for feeder services during incident detections.

Figure 7: NLOGIT/LIMDEP calibration output of vehicle access variable

Figure 8: SPSS calibration output of vehicle access variable

Figure 9: Validation output of cumulative feeder probability by vehicle access

Figure 10: Validation output of feeder probability by vehicle access
The probability of observed behaviour of selecting feeder decreased with the increasing number of vehicle accessible as an alternative to rail. Calibrated models from the parameter estimates by SPSS reflected this pattern correctly. It was also logical that as the number of alternative vehicle increases, users were more likely to switch mode from rail. The same cannot be said to the model calibrated by LIMDEP.

Figure 11: MsExcel calibration output of egress cost variable

Figure 12: SPSS calibration output of egress cost variable

Figure 13: Validation output of cumulative feeder probability by egress cost
Figures 11 and 12 show the relationship between the probability of selecting feeder and the egress cost of a respondent, calibrated based on MsExcel and SPSS outputs respectively. Figures 13 and 14 depict the validation or more appropriately the invalidity of the models (cumulative and non-cumulative graphs respectively).

The output from MsExcel was not appropriate for the dependent variables of nominal category. The results produced were against common sense that probability can carry a negative value. Thus, the calibration and validation processes are abandoned.

Under SPSS, the increase in egress cost would reduce the propensity of users selecting feeder services as an alternative mode for the continuing journeys. This was in contrast to the cumulative validated probability of the observed or actual mode switching behaviour. According to respondents, they will be more likely to change to a feeder bus in events of train derailment, given the increasing cost of egress trips.

There is a possibility that both trends are correct. The threshold is at the rate of RM2 which is the current bus fare. Interchanges and transfers with costs would be relatively unattractive to some respondents up to RM2 if their egress costs was less or comparable to the value. But as egress costs increases the propensity to switch mode would logically increase because the alternative feeder bus mode would be more attractive. However, further investigation may be needed to explain this variability.

Low model fitting results limited the capability to successfully calibrate the mode choices under all the scenarios. In particular, the exclusion of factors costs and time were inconsistent with many literature of mode choice modelling.

In other words, great discrepancies between the actual data and the modelled mode choices proved that the simulated data failed to represent what is going on in the real world. There are possible explanations for the anomalies. People may not behave as intended or as common a sense would be. Many respondents indicated that they would prefer waiting on the rail because buses were associated with risks such as waiting, transferring, unreliable scheduling and susceptibility to traffic congestion. They also reflected much of their behaviours to what had happened during the previous delayed trips. In this case, repeated or habitual behaviour in mode choice decision making can be one contributing factor.
7. DISCUSSION AND CONCLUSIONS

As far as the modelling exercises are concerned, the estimation runs were limited by model fit inadequacy, hence the lack of calibration capability. These conflicting findings could be partially explained by the differences in how parameter estimates are calculated. SPSS presented the probability of a mode selection by use of reference category against the other alternative categories. LIMDEP, however, produced the results in the utility equation of the alternative mode in question.

The summary (Table 2) indicated $R^2$ of Excel are almost similar to $R^2$ of LIMDEP, with the exception of scenarios with audio-30 minutes delay and text-30 minutes delay. SPSS improves the $R^2$ by twice the values in Excel and LIMDEP. Similarities of contributing parameters are also evident. All software packages could not produce any model fit that is required (0.5 and above). Suffice at this stage to say that the dataset may be the contributor of this anomaly. Biasness was introduced during the conduct of surveys and estimation or interpolation of variables such as costs and time.

Validation exercise was also unsuccessful because results produced great disparities between the observed data and the modelled ones. In summary, four main variables are useful to explain some of the variations in the models. First, age variable provided positive coefficient for rail choices and negative for feeder choices. Older users will have greater propensity to select rail, but the younger have greater inclination for feeder services. Second, income variable has a negative coefficient for rail choices. The lower the income, the more likely it is to stick with rail and therefore less mode switching to feeder buses. It may also be true that the effects of increasing costs are not bearable for low income earners. Third, the egress cost variable has a positive coefficient for rail choices. The higher the egress cost, the more likely it is for users to remain with rail services and therefore lower demand for feeder buses. It may be that destinations with great distances cannot be comfortably and conveniently reached by buses. In addition, buses are more susceptible to traffic congestion. Rail has exclusive right of way, which provided greater confidence to users. Finally, revealed mode has a negative coefficient for rail choices but a positive one for feeder choices. Those who have previously switched travel modes, would not stay with rail. They would also have greater preferences to opt for feeder services. Time and costs are not necessarily important factors, which is in conflict with many findings of case studies elsewhere. As such, it is suggested that future works to be undertaken on analysing the sensitivity of the dataset towards delay duration and changes in information media. Both intervening variables can be presented as delta or unobserved factors to explain these behavioural changes.

There are other considerations like the current state of public transport services levels and information provision. Delays are so persistent in the existing services that it is easy for regular users to estimate delay duration and the causes; hence to contemplate on necessary actions. There are currently no alternative of equivalent quality and comparable costs to rail. Buses are worse off in incident events because of congestions. There are no differentiated services such as local, semi-express and express trains. The double track system is limited to the commuter-rail route, but priority is given to intercity services in most cases. No alternative or deviation is possible. Buses have very bad image, feeder buses are not sufficient and have low level of services. Ticketing is disintegrated; transfer costs are additional for rail users.
who choose to switch to buses. Private vehicles are more attractive. Most users are captive public transport users. Even though some have alternative access to private vehicles, the situation of congestion on the road further worsens the possible scenario. Rail services are seen better, especially light-rail like Ampang and Kelana Jaya lines.

The calibration substituted one parameter each at one time to determine which has greater contribution or made the models more sensitive. Validating, in this paper was concentrated on one of the nine scenarios. Both processes were not without their challenges, especially when the $R^2$ or model fit indicator was too low for estimation. Additional effort is also needed to better understand and utilise LIMDEP.

Finally, there are always possibilities to adjust the focus on literature relating to “unobserved factors/parameters” in the utility functions. There will be a departure from restricting the independent variables to only travel time and trip costs. Other factors can be revisited, such as historical or repetitive decisions, habit, risk aversion (users do not want to risk taking bus which is already susceptible to road traffic congestion) and familiarity with the existing system which can therefore assist users to estimate delay duration and when services are resuming to normal (based on personal experiences). Currently all respondents are being aggregated into a single group. Hence, an extended research on more disaggregated responses such as according to trip purposes or peak or off-peak period could also be conducted.

In conclusion, this paper has illustrated that modelling feeder buses demand, as part of an integrated public transport scheme, can be quite challenging. Consideration of parameters to be estimated, calibrated and validated proved to be an important elements in the overall exercise. Dataset of actual behaviour recorded may not be fully mimicked by the models developed as have been shown by this part of the research. However, it provides an insight into other possible determinants or treatments that can or have already been factored in by some models developed for case studies in developing cities.

REFERENCES


