An application of Multinomial Logit Model (MNL) on tourist destination choices

S Nurlaela1,2
1 Urban and Regional Planning Department, Faculty of Architecture Design and Planning
2 Institut Teknologi Sepulung Nopember
Email: sitnurlael0@gmail.com

Abstract. This paper provides an analysis of multinomial logit model (MNL) on tourist destination choice. The choice set consisted of the Tropical North Queensland (TNQ), competed with the California Beach, the Cornwall Beach, and “No Travel” choice. The MNL model was used to identify the best scenario that gave the TNQ a market leadership in accordance with good budget performances. A scenario was developed by testing variable price combined with dummy variable of increasing family facilities; then testing the scenario based on the combination of willingness to pay (WTP) and elasticity analysis. Findings reported the best scenario was a combination of policy on the price and family facilities which made TNQ lead the markets and gave the highest revenues. Furthermore, the elasticity analysis was conducted to understand the impact of policy changes on the model outputs. The result showed that the market share and revenue of TNQ under the chosen scenario still exceed the performances of other scenarios. It was worth noting that these additional revenues raised from the chosen scenario ($71,662) cannot exceed the cost for family facility improvements; otherwise there was no policy visible than “do nothing” scenario.

Keywords: Choice modelling, Multinomial logit model (MNL), tourist destination choice, the Willingness to pay (WTP), the Elasticity analysis.

1. Introduction
In a seamless world, tourist destinations competed globally. Tropical North Queensland (TNQ) faced such of competition, especially from those with similar product characteristics. There were three tourist destinations considered in this paper, particularly tourist destinations with similar characteristics between the Tropical North Queensland (TNQ) in Australia, California Beach in USA (CaB) and Cornwall Beach in Jamaica (CoB). The choice set included “no choice” or none alternative, which was identified as “to prefer stay home rather than travel for holiday at a distance”.

To attract more tourists to TNQ and win the competition among competitors, an analysis in tourism demand was conducted by a discrete choice model. The model specified the probability of one alternative being chosen by consumers (i.e. tourists) among a choice set. This amount of probability represents a market share of the chosen alternative. The economic theory suggests that tourists will choose a destination that maximize their utility.

The MNL model could reveal factors in which affecting the choice decisions. The policies would be consisted of some actions to interfere these factors in a way that the model can maximize the utility
values. Therefore, the model would maximize the probability of TNQ being chosen within the choice set. The aim of this paper was to identify the best scenario in which the performance of the scenario was not only measured by the market share, but also by the monetary performance.

Figure 1. The images of tourist destinations
Source: Google Image

2. Multinomial logit model (MNL)

The review on the MNL model presented here was adapted mainly from [1], [2], [3], [4] and [5]. Logit model is defined as a discrete outcome model derived in multinomial form (three or more outcomes); by assuming the disturbances are Weibull distributed (Gumbel extreme value type I) [1, p. 440].

MNL model is one family of logit model in which the decision maker faces more than 2 (two) choices. It has a non-linear form of probability based on the utility theory; or in the form of logistic distribution function [3], which the defined utility function is in a linear form. The objective of MNL model was to estimate a function that determines the outcome probabilities [1].

The choice set is defined as a discrete outcome model, in which the utility function that will be maximized by the model often in a form of price and income as arguments, i.e. the form of indirect utility [1, p. 317] was written in the form as follows:

\[ y_m = \frac{\partial V}{\partial p_m} / \frac{\partial V}{\partial Inc} \]  (1)

Where:
\( y_m \) is the utility maximizing demand for good \( m \),
\( V \) is the indirect utility, \( p_m \) is the price of good \( m \),
\( Inc \) is the decision maker’s income.

However, by applying the utility framework to the discrete outcome, the model becomes the utility determining the choices (as opposed to the model determining the outcomes), in which the form is:

\[ T_{in} = \beta_i X_{in} \]  (2)

Where:
\( T_{in} \) is a linear function that determines discrete choice \( i \) for observation \( n \),
\( \beta_i \) is a vector of estimable parameters for discrete choice \( i \),
\( X_{in} \) is a vector of the observable characteristics or covariates that determine discrete outcomes for observation \( n \).

The derivation of discrete outcome model into the model of discrete choice implies that the model is compensatory, i.e. change in factors that determine the function of utility for each discrete choice do not matter as long as the total value of the function remains the same [1, p. 317].

To reach statistically estimable probabilistic model, a disturbance term \( \varepsilon_{in} \) was added [1, p. 311], while [2] stated this as to following the random utility theory. There was a decomposition of utility that
decision maker obtains from alternative $i$ in $J_n$ (or $T_n$) into two parts, i.e. the part that is known by researcher, labeled as $V_{in}$ as a function of $f_i$; and another one is unknown part that is assumed to be a random variable and labeled as $\epsilon_{in}$ [4]. Then the utility based model becomes:

$$T_{in} = V_{in} + \epsilon_{in}$$

(3)

Where:

$V_{in}$ is a function that depends on the observed characteristics of the alternative as faced by the decision maker ($z_{in}$), the observed characteristics of the decision maker (label $s_n$), and a vector of parameter $\beta$ that is either known a priori by the researcher or estimated : $V_{in} = V(z_{in}, s_n, \beta)$ [4, p. 15].

Suppose utility function of $V_{in}$ follows the form of $\beta_i X_{in}$, then,

$$T_{in} = \beta_i X_{in} + \epsilon_{in}$$

(4)

Where:

$\epsilon_{in}$ is the disturbance term, due to several situations [1, p. 310]: (1) there are some variables have been omitted in the model, (2) the function form of linear maybe incorrectly specified, (3) there may involve the use of proxy variables as an approximate for missing variables in the data base, (4) variations in $\beta_i$ that are not accounted for, i.e. $\beta_i$ may vary across observations.

Although the utility function in the logit model is linear in $X$, the probability of the function is not linear [3]. The multinomial logit model is the logit model whereas the probability of the utility function is in the form of logistic distribution function [2]. The model is expressed as:

$$P_n(i) = \frac{e^{T_{in}}}{\sum_{j\in C_n} e^{T_{jn}}}$$

(5)

Where:

$j$ is the number of choice set, $n$ is the decision maker (observation), $i$ is the choice being decided by $n$, and $T$ is the utility function of individual/household $n$ toward choice $i$.

Following the theory that utility derived from random utility theory, the utility of each alternative will be divided into two components: (1) deterministic utility that is $V_{in}$ and (2) random component that is $\epsilon_{in}$.

Therefore, the choice $i$ will be chosen as the best alternatives when the utility of $i$ is greater than utility of any other alternatives in choice set, stated in [2, p. 101] as:

$$P_n(i) = Pr(T_{in} \geq T_{jn}, \forall j \in C_n, j \neq i)$$

(6)

$$P_n(i) = Pr(\beta_i X_{in} + \epsilon_{in} \geq \beta_j X_{jn} + \epsilon_{jn}, \forall j \in C_n, j \neq i)$$

(7)

$$P_n(i) = Pr(\epsilon_{jn} \leq \beta_i X_{in} - \beta_j X_{jn} + \epsilon_{jn}, \forall j \in C_n, j \neq i)$$

(8)

Usually, the proper probability mass function is $0 \leq P_n(i) \leq 1$, for all $i \in C_n$, whereas the total of probability value is one, or $\sum_{i\in C_n} P_n(i) = 1$.

Readers who are interesting in the derivation of probability function could refer to [2].

3. MNL model results and discussion

The data for this paper was sourced from http://www.library.unsw.edu.au/~thesis/adtd-ADFA/uploads/approved/adtd-ADFA20020624.165002/public/10chapter9.pdf. From the data, the MNL model was built based on some attributes as follows:

- Whether the tourist destination has a tropical climate (variable named as TROPICAL)
- Whether the tourist destination provides the family facility (variable named as FAMFAC)
Whether the tourist destination has un-spoil environment (UNSPOIL)
- Whether the tourist destination has a quiet development (QUIET)
- The personal characteristic of respondents or consumers in terms of age (AGE)
- The costs that a consumer would spend for 2 weeks holiday (PRICE).

The data key being used in the model is presented in table 1.

**Table 1.** Data key for the MNL model

| Variable Name | Variable Category | Data Range               |
|---------------|-------------------|--------------------------|
| Price         | Price of 14 days holiday | Destination attributes | AU$2,000 – $5,000 |
| Development   | Development/crowdedness situation | Destination attributes |                      |
| 0             | Quiet             | Destination attributes  |                      |
| 1             | Moderately busy   | Destination attributes  |                      |
| 2             | Very busy         | Destination attributes  |                      |
| Environment   | Environment condition | Destination attributes |                      |
| 0             | Un-spoil         | Destination attributes  |                      |
| 1             | Somewhat spoil    | Destination attributes  |                      |
| 2             | Very spoil        | Destination attributes  |                      |
| FamFac        | Which offers family facility | Attitudinal attributes |                      |
| 0             | No                | Attitudinal attributes  |                      |
| 1             | Yes               | Attitudinal attributes  |                      |
| Tropical      | Which has tropical climate | Attitudinal attributes |                      |
| 0             | No                | Attitudinal attributes  |                      |
| 1             | Yes               | Attitudinal attributes  |                      |
| Age           | Age of respondents | Demographic attributes  | 18 – 90 years       |
| <65           | Less than 65 years old | Demographic attributes |                      |
| >=65          | Equal or more than 65 years old | Demographic attributes |                      |

The effect coding was created to allow for non-linear effects in the levels of attributes. In the estimation of the utility value, for example by regression model, the effect coding will made a different value of utility associated with each level of attributes being coded. By running the regression equation to model the utility, the effect coding and the non-linear marginal utility is described in the following graphs.
The utility equation for each of the choice set was developed as follows:

- **California** = 2.2684 – 0.00083Price + 0.2641Quiet + 0.1874Mbus – 0.4515Verybusy + 0.2641Unspoil + 0.0074Somep – 0.3385Veryspoil + 0.0285Age(<65) – 0.0285Age(>=65) + 0.443Famfac(1) – 0.443Famfac(0) + 0.386Tropical(1) – 0.386Tropical(0)
- **Cornwall** = 1.8055 – 0.00083Price + 0.2641Quiet + 0.1874Mbus – 0.4515Verybusy + 0.2641Unspoil + 0.0074Somep – 0.3385Veryspoil + 0.0285Age(<65) – 0.0285Age(>=65) + 0.437Famfac(1) – 0.437Famfac(0) + 0.82Tropical(1) – 0.82Tropical(0).
- **Tropical NQ** = 2.1831 – 0.00083Price + 0.2641Quiet + 0.1874Mbus – 0.4515Verybusy + 0.2641Unspoil + 0.0074Somep – 0.3385Veryspoil + 0.0285Age(<65) – 0.0285Age(>=65) + 0.488Famfac(1) – 0.488Famfac(0) + 0.88Tropical(1) – 0.88Tropical(0).
- **None** = - 1.1974 + 0.0285Age(>=65) – 0.0285Age(<65)

Within “do nothing” scenario, the initial market share of the TNQ was at 30%, similar to CaB, while CoB had 25% of market share, and the rest of the market (15%) choose not to travel.

3.1. Descriptive statistics
From descriptive statistics, the age of respondents ranged from 18 – 90 years in which the average age of respondents was 54 years. Variable price ranged from $2,000 to $5,000 in which the average price among the three alternatives was nearly similar ($3,500). Majority of the respondents viewed the TNQ
as a favorite destination based on the positive responses for TNQ related to the destination attributes. For example, more than a third respondents felt that TNQ has “quiet” and “un-spoil” situation, greater than the number of respondents that felt the same about the other two alternatives. However, only 47.47% respondents said that TNQ offers family facility, lower than Cornwall whereas 50.69% respondents said so.

Figure 3. Descriptive Statistics

3.2. The scenario analysis for PRICE
Suppose the tourist manager does nothing to the environment, to the development quality and the quantity of family facility. To capture more market share, they decreased the price of TNQ at about the willingness to pay or WTP of other attributes. For example, the willingness to pay for the environment
and the development attributes (UNSPOIL and QUIET) were about $318 or 8.97% of the average price of TNQ. Decreasing the price of TNQ at 8.97% of its current price or at the WTP of UNSPOIL and QUIET has increased the market share of TNQ around 4.10% and its revenue raised around $7,970. Meanwhile, decreasing the price of TNQ at the WTP for FAMFAC at around $588 or 16.58% of its current price has increased the market share for 7.79% and the revenue for $11,738.

3.3. The scenario analysis for QUIET, UNSPOIL and FAMFAC
Suppose the tourist manager wants to fix the price of TNQ and instead interferes to other attributes to give a better quality of tourist destination. Setting all of the development responses to QUIET, *cateris paribus*, has increased the market share of TNQ to 34% and an additional revenues of $31,500. Setting all of the environment responses to UNSPOIL has increased the market share of TNQ to 34.3% and an increase in revenues at $33,000. On the other hand, setting all family facility (FAMFAC) responses to YES, *cateris paribus*, has increased the market share to more than 38% with an additional revenues of more than $63,000.

The combination of scenario for PRICE and FAMFAC, i.e. decreasing the price around $588 and improve the family facility has produced a significant result in the TNQ market share which has jumped to 47%, i.e. an increase of 17.12% from the initial market share and its revenues rose high adding up to $71,600 for TNQ. Based on cross elasticity analysis, the additional shared of 17.12% for the TNQ came from the CaB that experienced market declined at about 7.3%, the CoB’s share also declined at about 5.76%, and the none alternative’s share declined about 4%.

3.4. The elasticity analysis
In addition, in terms of elasticity (cross elasticity analysis), the none alternative experienced the largest shifted of choice, that 27% of “none travel person” choose to travel to TNQ after the price of TNQ decreased at large amount $588. This also mentioned that people in this group were more elastic in price (1.63) compared to California (1.47) and Cornwall (1.39) respectively.

**Table 2.** The chosen scenario: decreasing the price of TNQ ($588 or 16.58% of the average price) and setting all of the famfac (family facility) responses to YES

| AGREGATION MODEL SHARE | AGREGATION BASE MODEL |
|-------------------------|-----------------------|
| ALTERNATIVES SHARE      | ALTERNATIVES SHARE    |
| CALIFORNIA 22.688%      | CALIFORNIA 30.00%     |
| CORNWALL 19.241%        | CORNWALL 25.00%       |
| TROPICALNQ 47.124%      | TROPICALNQ 30.00%     |
| None 10.946%            | none 15.00%           |
| TOTAL 100.00%           | TOTAL 100.00%         |

| REVENUE ANALYSIS BEFORE POLICY | AFTER POLICY | DIFFERENCES |
|--------------------------------|--------------|-------------|
| Average price $3,548.00        | $2,959.53    | -$588.47    |
| Market share 30.00%            | 47.12%       | 17.12%      |
| Population 217                 | 217          | 0           |
| % of Yes responses (FamFac) 52.53% | 100.00%     | 47.47%      |
| Revenue $230,974.80            | $302,637.33  | $71,662.53  |
| Increase of revenue            |              | $71,662.53  |
Table 3. The elasticity analysis of the chosen scenario

|                          | DIRECT ELASTICITY | CROSS ELASTICITY |
|--------------------------|-------------------|------------------|
| TROPICAL NORTH QUEENSLAND|                   |                  |
| % CHANGE IN PRICE OF TNQ | -16.59%           |                  |
| % CHANGE MARKET SHARE OF TNQ | 57.08%          |                  |
| ELASTICITY (e)           | -3.44             |                  |

|                          |                   |                  |
| CALIFORNIA               |                   |                  |
| % CHANGE IN PRICE OF TNQ | -16.59%           |                  |
| % CHANGE MARKET SHARE OF CALIFORNIA | -24.37%        |                  |
| ELASTICITY (e)           | 1.47              |                  |

|                          |                   |                  |
| CORNWALL                 |                   |                  |
| % CHANGE IN PRICE OF TNQ | -16.59%           |                  |
| % CHANGE MARKET SHARE OF CORNWALL | -23.04%       |                  |
| ELASTICITY (e)           | 1.39              |                  |

|                          |                   |                  |
| NONE                     |                   |                  |
| % CHANGE IN PRICE OF TNQ | -16.59%           |                  |
| % CHANGE MARKET SHARE OF NONE | -27.02%        |                  |
| ELASTICITY (e)           | 1.63              |                  |

4. Conclusion
The chosen scenario for TNQ developed in the MNL model gave the highest revenues and the highest market share. This scenario can compensate the loss of benefit for tourists due to no improvement in the environment and development quality, because the decrease of price in this scenario was exceed the WTP for QUIET and the WTP for UNSPOIL. Note, that the additional revenue raised from this scenario ($71,662) cannot exceed the costs for family facility improvements; otherwise there was no policy visible than “do nothing” scenario.

To conclude, MNL model is a useful analysis tool to examine the choice preferences. This paper applied this tool on tourist destination choices to identify the best possible scenario for the Tropical North Queensland in order to get a better position in the tourism market among the choice set.

References
[1] S.P. Washington, M.G. Karlaftis, F.L. Mannering. Statistical and Econometric Methods for Transportation Data Analysis. (second ed.). Chapman and Hall/CRC, Boca Raton, FL; 2011.
[2] Ben-Akiva, M.E., & Lerman, S.R. Discrete choice analysis: theory and application to travel demand. Cambridge, Mass: MIT Press; 1985.
[3] Gujarati, D.N. Basic econometrics. 2nd ed. New York: McGraw-Hill; 1988.
[4] Train, K. Qualitative choice analysis: theory, econometrics, and application to automobile demand. Cambridge, Mass: MIT Press; 1986.
[5] Hensher, D.A, J.M. Rose, and W.H. Greene. Applied choice analysis: a primer. Cambridge: Cambridge University Press; 2005.