

## **Estimation of travel mode choice for domestic tourists to Nha Trang using the multinomial probit model**

**Vo Van Can\***

\* Tromsø University Business School at University of Tromsø and Accounting-Finance  
Faculty at University of Nha Trang

### **Abstract**

The purpose of this study is to examine how the characteristics of domestic tourists and attributes of travel modes influence the tourists' modal choice to Nha Trang, Vietnam by applying the multinomial probit model. The analysis is based on primary data surveyed from tourists visiting Nha Trang in March, 2011. A total of 402 valid samples were used from 554 initial samples. The study provides several important findings concerning tourists' modal choice. Travel time per kilometer, per-kilometer travel cost to income ratio, mode quality variables, and income are key elements in explaining the tourists' modal choice decision. In addition, tourists with a lower income tend to be more sensitive to change in per-kilometer cost. Furthermore, the high-income tourists are much more likely to choose plane or train rather than coach. Understanding the tourists' modal choice behavior may help tourism transport companies to develop appropriate marketing strategies.

**Keywords:** multinomial probit; travel mode; choice; domestic tourist.

## 1. Introduction

Nha Trang city is a well-known tourist destination in Vietnam. The number of both domestic and foreign arrivals to Nha Trang has increased consistently in recent years. In particular, there has been a substantial growth in domestic tourism. In the period 2005–2010, the annual growth in domestic tourism was 15%. In conjunction with the increase in the number of the tourists, the number of modes to meet tourists' travel demands has also increased. Vietnamese tourists can choose one of various transport modes such as coach, train, plane, sharing a car, driving their own car and others for their trips arriving in Nha Trang. Of these modes, coach, train and plane are the most popular forms of transport for tourists. However, an important question is why a tourist prefers this travel mode to alternative travel modes. Answering this would enable the tourism transport industry to predict with greater certainty which travel modes will be chosen by tourists. This may thereby help tourism transport companies to develop appropriate marketing strategies to meet tourists' modal choice demand.

Nature of modal choice decision-making process amongst many modes is a discrete choice analysis (Dow and Endersby, 2004; Train, 2009). This analysis is based on the economic theories of random utility assuming that a traveler chooses the mode with the highest utility (Bhat, 1998; Bolduc, 1999; Train, 2009). Since discrete choice analysis was introduced to solve transportation related problems, hundreds of studies have applied it in various fields (Bolduc, 1999). These studies have focused on analysing the behavior of the decision-making process such as modal choice (Bhat, 2000; Bolduc, 1999; Cohen and Harris, 1998; Commins and Nolan, 2011; Dissanayake and Morikawa, 2005; Ewing et al., 2004; Habib, 2012; Train, 1980), choice of car type (Choo and Mokhtarian, 2004; McCarthy, 1996), choice of tourist destinations (Fesenmaier, 1988; Nicolau and Más, 2006; Train, 1998). However, using discrete choice analysis, especially multinomial probit model, for modal choice of tourists to a destination has not been paid much attention by researchers, to the author's knowledge. In the case of tourists' modal choice, data on a dependent variable fall into one of several mutually exclusive and unordered alternatives. Such data are normally modelled through the multinomial logit (MNL), nested MNL (NMNL), and multinomial probit (MNP) models, amongst others (Swait and Adamowicz, 2001).

Of the above models, the MNL is the simplest and most popular model (Baltas, 2007; Horowitz, 1991; Train, 2009). The important assumption of this model is that the random

components of the utilities of various alternatives are independently and identically distributed (IID). This hypothesis creates several advantages in applications because calculation of choice probabilities, presentation, and interpretation of the parameters are easy (Baltas, 2007; Train, 2009). However, because the IID assumption is often violated (i.e., the similarities amongst alternatives in a choice set are ignored) in practice, the applications of the MNL are limited (Daganzo and Sheffi, 1977; McFadden, 1974). To overcome part of this limitation, researchers have proposed an application of the NMNL model in which the random components of the utility function are correlated within each of the groupings (McFadden, 1978; Williams, 1977). Nevertheless, in situations where there is panel data with unobserved heterogeneity or situations with random preference variation, the NMNL model cannot apply (Horowitz, 1991).

The MNP model partially overcomes the above-mentioned problems. This model assumes that the random components of utility functions have a multinomial normal distribution with zero mean and nonzero covariances between alternatives (Chintagunta, 1992; Train, 2009). These assumptions allow a much more flexible pattern of error correlation, do not require the specification of a nesting structure, and are therefore more realistic (Dow and Endersby, 2004). However, to benefit from the MNP model, researchers have to pay a high price in term of complex computation of the choice probabilities of multidimensional integrals (Bolduc, 1999; Chintagunta, 1992; Horowitz, 1991; Weeks, 1997). Over time, researchers have improved the method and found ways to reduce the computational burden of the integrals. Amongst the researchers, McFadden (1989) proposed a method of simulated moments for estimating the parameters of MNP models. In addition, together with the substantial progress in information technology, several applications for implementing the MNP model are available on the market. Thus, the MNP has become increasingly popular in studies on individual's choice behavior, such as branch choice (Chintagunta, 1992; Paap and Franses, 2000), modal choice for working trips (Bolduc, 1999; Johansson et al., 2006), electoral research (Alvarez and Nagler, 1998; Dow and Endersby, 2004). According to McCulloch and Rossi (1994), the appeal of the MNP approach lies mainly in the relaxation of the IID property.

In this study, the MNP model was chosen to examine how characteristics of domestic tourists and attributes of transportation means affected the tourists' modal choices. The tourists referred to here were individuals coming from the three main tourist markets in

Vietnam, namely Hồ Chí Minh city, Đà Nẵng and Hà Nội. The study has only focused on the holiday trips. Business trips and trips for the purpose of visiting friends, relatives and others were not included in the analysis. Because travel for purposes other than holidays is driven by different needs (Gössling and Hall, 2006), these decisions were not included. For example, the trip and modal choice of a business traveler could be decided by the employer. Alternatively, the data used for this study came from a survey of 554 tourists who were visiting Nha Trang. However, the number of valid samples were 402, accounting for 72.56% of the total sample interviewed. The most important variables that affect the tourists' modal choice are income, travel time per kilometer, per-kilometer travel cost to income ratio, and variables reflecting quality of modes. This study is part of an effort to enhance understanding of the tourists' modal choice behavior to Nha Trang.

The rest of the study is organized as follows: Section 2 introduces the structure of the MNP model. Section 3 delineates the selection of variables included in the model. Section 4 describes the survey and the sample. Section 5 carries out a data analysis while Section 6 presents the results of the analysis of the tourists' behavior in relation to travel mode choice. Finally, Section 7 provides conclusions and implications.

## 2. The multinomial probit model

A travel mode choice process of a tourist from a choice set can be described by a utility function. According to McFadden (1974), this function includes two components. The first component is the observed utility of the alternative. It depends on both specific attributes of the mode and traits of the tourists. The second component is a random term that depicts the effects of the unobserved attributes and characteristics on the utility (Horowitz, 1991). Suppose that the choice set includes  $j$  modes ( $j=1, \dots, m$ ). Based on the studies of Dow and Endersby, 2004, Geweke, et al., 1994, and McFadden, 1974, the utility that tourist  $i$  ( $i=1, \dots, n$ ) obtains from mode  $j$  in the set may be written as follows:

$$U_{ij} = \beta' X_{ij} + \gamma'_j Z_i + \varepsilon_{ij}$$

where  $U_{ij}$  is the tourist  $i$ 's utility for the mode  $j$ ;  $X_{ij}$  is a vector of observed attributes relating to the travel mode  $j$  as perceived by the tourist  $i$  and  $Z_i$  is a vector of traits of the tourist  $i$ . The coefficients of  $\beta$  and  $\gamma_j$  are vectors of fixed parameters of the variables  $X_{ij}$  and  $Z_i$ , respectively and the prime denotes transposition. The final component,  $\varepsilon_{ij}$ , is unknown term.

Tourists will choose the mode that provides the greatest utility. The probability that tourist  $i$  will choose mode  $m$  is the probability that the utility of mode  $m$  exceeds the utility of all other available modes in the set. Therefore,

$$p_{im} = P(U_{im} > U_{i1}, U_{im} > U_{i2}, \dots, U_{im} > U_{ip})$$

$$= P\{(\varepsilon_{im} - \varepsilon_{ij}) > [\beta'(X_{ij} - X_{im}) + Z_i(\gamma'_j - \gamma'_m)] \mid j \neq m\}$$

The MNP model is obtained by assuming that errors are of a multivariate normal distribution with mean zero and covariance matrix  $\Sigma$  where  $\Sigma$  is not restricted to being a diagonal matrix (Chintagunta, 1992; Dow and Endersby, 2004; Geweke, et al., 1994). Let

$\varepsilon_{ij}^* = \varepsilon_{im} - \varepsilon_{ij}$  and  $V_{ij}^* = \beta'(X_{ij} - X_{im}) + Z_i(\gamma'_j - \gamma'_m)$  with  $\forall j \neq m$ , then the probability is:

$$P(\text{mode} = m \mid \beta, \gamma, X, Z) = \int_{V_{i1}^*}^{\infty} \int_{V_{i2}^*}^{\infty} \dots \int_{V_{ip}^*}^{\infty} f(\varepsilon_{i1}^*, \varepsilon_{i2}^*, \dots, \varepsilon_{ip}^*) \partial \varepsilon_{i1}^*, \partial \varepsilon_{i2}^*, \dots, \partial \varepsilon_{ip}^*, j \neq m$$

where  $f(\cdot)$  is the probability density function of the multivariate normal distribution.

### 3. Variables

#### 3.1. Dependent variable

The dependent variable is a discrete variable that represents a choice from a set of modal choice. The set must meet the three criteria: finite, mutually exclusive, and collectively exhaustive (Ben-Akiva and Koppelman, 1974; Train, 2009). There are several different ways to come to Nha Trang. However, only coach, train, and plane are included in the present study. Other modes, such as car and motorbike, were not included in the choice set. Car was excluded because it is not a very popular transport mode for holiday. In addition, the distances between the cities are too great for motor biking to be an unviable alternative.

#### 3.2. Independent variables

In a comprehensive review of studies published in the period 1962–1979 on modal choice by Barff et al. (1982), independent variables were divided into two main categories. While the first category expresses mode attributes, the second category reflects travelers' characteristics. Many studies use both categories (e.g., Commins and Nolan, 2011; Ewing et al., 2004; Koppelman et al., 1993). Some only use mode attributes (e.g., Kraft and Kraft,

1974; Recker and Golob, 1979) or just traveler characteristics (e.g., Ben-Akiva and Richards, 1976; Lerman and Ben-Akiva, 1976)<sup>4</sup>. In addition, dummy variables are also often added to the model specification in specific (Bolduc, 1999). As independent variables, travel time, travel cost, income, comfort, convenience, safety, auto ownership, and demography are commonly used. Based on this, the following independent variables were included in the current study.

- City dummies

Distance probably has an important effect on modal choice. Specifically one would expect a traveler from a distant city to be less likely to choose coach compared to plane or train. Since all travelers in this study comes from one of three cities, a dummy for two of them, Hồ Chí Minh city and Đà Nẵng, is used. Hà Nội was then picked as the reference city. The dummies might pick up city specific traits in addition to distance. However, a preliminary model with distance in addition to city dummies revealed that using both variables caused singularity, so the dummies are mainly measuring the effect of distance.

- Travel time per kilometer

In behavioral research related to travel mode choice, one of the key variables often used is travel time (Ben-Akiva and Richards, 1976; Bolduc, 1999; Kraft and Kraft, 1974; Lave, 1969; McConnell and Strand, 1981; Quarmby, 1967). Travel time reflects the total time from original location to destination rather than terminal-to-terminal (Kraft and Kraft, 1974). However, this time is normally detailed into many different components in order to reflect more clearly the effects of each type of travel time on modal choice behavior. For example, Quarmby (1967) divided overall travel time into “travel time” and “excess travel time”, while Koppelman and Bhat (2006) and Rassam et al., (1971) separated that time into “in-vehicle travel time” and “out-of-vehicle travel time”. Bolduc (1999) broke travel time of all the modes down into “walking time”, “in vehicle time”, and “waiting time”, but the concept of “waiting time” was not used for car. Actually, walking time and waiting time are also travel time outside of mode. Thus, in this study travel time would be divided into on-mode and out-of-mode travel time.

---

<sup>4</sup> Summaries of mode attributes and traveler characteristics of the above-mentioned models are contained in Barff et al. (1982).

On-mode travel time is “the sum of the time spent traveling, and dwell time during which all passenger boardings and alightings occur” (Zorn et al., 2012, p.757). It reflects the actual duration that a passenger spent on a specific mode from the departure to arrival terminal. Thus, this time only reflects part of the whole travel time that a tourist spent for the journey. Alternatively, out-of-mode travel time is the time not spent inside the mode, including access time, waiting time, and egress time (Koppelman and Bhat, 2006). It can be calculated as the difference between the sum of travel times and on-mode travel time (Kraft and Kraft, 1974). In the empirical studies on modal choice behavior, several studies indicate that on-mode and out-of-mode travel time in particular, and travel time in general have important effects on modal choice decision of travelers (see, for example, Bolduc, 1999; Commins and Nolan, 2011; Johansson et al., 2006; Habib, 2012; Swait and Ben-Akiva, 1987). Furthermore, travelers are much more sensitive to out-of-mode time than to on-mode time and therefore a minute of out-of-mode time will generate a higher disutility than a minute of on-mode time (Koppelman and Bhat, 2006).

In order to compare travel time across different modes for different cities, travel time is normalized by distance. Such a normalization also removes potential multicollinearity between travel time and the city dummies. Since slow transport is usually considered a disadvantage, the expected sign of this variable is negative.

- Per-kilometer travel cost to income ratio.

Travel cost is also a variable widely used in modal choice models (Kraft and Kraft, 1974; Lave, 1969; Quarmby, 1967). Transportation costs are closely connected to distance. Thus, in order to easily compare travel cost between cities, travel cost is normalized by distance. In addition, the cost of a mode appears very differently for a person with high income. Hence, costs are represented as fraction of the budget by dividing it with income. High costs are normally considered a disadvantage, so the expected sign for this variable is negative.

- Modal service quality and price attributes

According to Litman (2008, p.43), “a minute spent in unpleasant conditions waiting for a bus may seem like an hour, while an hour spent working, resting, or conversing while traveling on a comfortable bus or train may seem pleasant or even delightful”. This refers to an important effect of modal service quality on the decision-making process. Apart from the

mode quality, reasonability of pricing element against mode quality is also considered as a determinant affecting modal choice of travelers. A mode with better service quality and a reasonable price will naturally be much more likely to be chosen by travelers. The modal service quality and reasonability of travel price are normally rated subjectively through travelers' experience or perception of modes' service attributes. These attributes may include safety, comfort, convenience, punctuality, reliability, and others (Johansson et al., 2006; Koppelman et al., 1993; Litman, 2008).

The effects of such attributes on modal choice process of travellers also have been discussed in more detail by Barff et al. (1982). Although these variables play an important role in explaining the individual's mode-choice behavior, the effects can be different in the various situations. For example, the study by Johansson et al., (2006) show that comfort, flexibility, and environmental preferences have significant effects on mode choice, but convenience and safety are insignificant. An analysis similar by Koppelman et al., (1993) also finds crucial effects of three service attributes, namely "convenience and independence," "timely service", and "inexpensiveness" on mode choice. However, both "ease of close parking" and "safety and comfort" are not found to have a significant effect on ridesharing propensity. Koppelman et al., (1993, p.348) conclude that "there seems to be general agreement that attitudinal factors are more important in determining ridesharing propensity than socio demographic attributes or the standard travel time and cost variables used in transportation mode choice models."

#### - Income

In behavioral research on modal choice, income is the most frequently used variable and has been found to be a strong significant correlate of travelers' modal choice (Barff et al., 1982; Lave, 1969). If travel costs are very different amongst modes, the effect of modal choice decision will depend on the size of the travelers' income; if travel costs are similar, the effect will be less (Bhat, 1998; Lave, 1969; Quarmby, 1967). Therefore, higher income travelers may be more willing to pay an extra premium to satisfy their preferences than lower income travelers are. They are more likely to accept a higher travel cost for their trip to benefit from the additional convenience afforded by a particular mode relative to other modes. In contrast, those who have a low-income level commonly choose travel modes with low prices like coach. This has been confirmed from the empirical results on intercity mode choice by Bhat (1995). His findings indicate that the high-income travelers favor plane travel



in comparison with the other modes, and the low-income travelers are likely to choose train travel. However, the effects of income on work trip mode choice in suburban areas are not significant (see, Koppelman et al., 1993).

- Demographic variables

Many demographic variables such as gender, age, occupation, and education level of individuals are also examined in modal choice models. They have been found to have statistical significance in some cases, but they are often more useful for analysing market segmentation (Barff et al., 1982). In this study, three demographic variables were used, namely age, gender, and occupation.

#### **4. Sample and survey**

##### ***4.1. Research design***

Based on the previous literature review and on distributions of the tourism experts, the questionnaire was designed in order for the information from the survey to fit the objectives of this study. However, the length of the questionnaire as well as the accuracy associated with the responses of respondents was also taken into account.

Based on the discussions above, assuming that all tourists face the same set of universal alternatives, the dependent variable was a set of modal choices for the tourists. It consisted of three alternatives: coach, train, and plane. They were coded as follows: 1 = coach, 2 = train, and 3 = plane. In the model, each tourist must choose one, and only one, of the three travel modes. The study contrasted train choice relative to coach choice, and plane choice relative to coach choice. Coach is chosen as the reference alternative because the number of tourists choosing this mode normally accounts for the highest rate of the three modes.

With regard to alternative-specific regressors, all on-mode travel time, total travel time, and travel cost of both the chosen mode and the competing modes in the set were included in the questionnaire. In addition, based on the studies by Johansson et al., (2006), Koppelman et al., (1993), and Litman (2008), the attitudinal variables for mode quality, namely safety, comfort, punctuality, price/quality, and luxury were included in the questionnaire. The tourists were only asked about the mode that they chose for these five variables. The attributes of the rival modes would be inferred from those tourists' responses. This procedure was based on a hypothesis that if a respondent agree with the attribute being better for the

chosen mode, the attribute was not as good for the competing modes. For example, if a tourist agreed with the statement: “The mode that you chose for your trip to Nha Trang is safer than the other modes”, it meant that the other modes was not considered as safe as the mode chosen. Suppose that tourists chose coach for their trip and agreed that this was chosen for safety. The above procedure would then mean that train and plane were viewed as less safe by the tourists than coach. This assumption was made in order to reduce the number of questions and interview time. Incomplete interviews were a slight problem, and additional questions would have made it worse. The sacrifice for this simplicity is that the travelers were not allowed to rate other travel modes as equal to that chosen.

All five mode-service attributes were measured using a two-point scale agree or disagree with the statements in the instrument and coded as follows: 1 if agree and 0 if disagree. Such a scale is “much shorter and more convenient to administer and score” (Komorita and Graham, 1965, p.987) and better than a multiple-choice scale (Ankeny, 1991). Furthermore, Schutz and Rucker (1975, p.323) conclude that “number of available response categories, at least within the 2- to 7-point range, does not materially affect the cognitive structure derived from responses to that scale.”

As case-specific regressors, monthly individual income, age, gender, and occupation of the tourist were used. The gender was coded as 1 if female and 0 if male; male was the base category. The age was classified based on the study by Dow and Endersby (2004) using a three-category variable: age 18-30; age 31–45, and age 46-60 with the third category regarded as the reference category. Given the study by Alvarez and Nagler (1998), a three-category variable for the tourists’ occupation (business sector staff, state sector staff, and other sector staff) was also applied in which other sector staff group regarded as the base category. In addition, city dummy variables were coded as 1 if Hồ Chí Minh city, 2 if Đà Nẵng, and 3 if Hà Nội.

#### ***4.2. Sampling and surveying method***

Using an above-designed questionnaire, the data collection method was implemented through a survey of face-to-face interview including two stages. The first stage was a pilot survey in order to improve the clearness, validity, and content of the questionnaire. In this stage, the questionnaires were delivered to twenty tourists. After receiving their feedbacks, a final standard version of the questionnaire was used in the next survey stage. The formal

survey was conducted with the support of three well-trained research students. The individual dataset was obtained through the direct interviews with the tourists visiting Nha Trang. Only individuals were interviewed. For tourists travelling as part of a tour or group, including households, only one decision-making person was interviewed, and just asked monthly income of decision-making individual (not the tourists group).

A convenient sample of 554 tourists was gathered in March 2011 at several well-known tourist sites in Nha Trang, such as Vinpearl Land, Diamond Bay, the PoNagar Towers, Phù Đổng Park, the Hòn Chồng tourism area, the Long Sơn Pagoda, and others. The respondents came from Hồ Chí Minh city, Đà Nẵng and Hà Nội. Tourists from these three cities were chosen because on the one hand those are three key tourist markets of Nha Trang, on the other hand all these cities have the same modal choice set.

## **5. Data analysis**

### ***5.1. Profile of respondents***

Table 1 presents descriptive statistics on the demographic profile of the tourists. Of the 554 cases collected, a block of 142 cases was rejected due to missing values, outliers, and other reasons. As a result, the number of usable cases in this study is 402, giving a sample rate of 72.56%. With the number of cases being 402 (corresponding to  $402 \times 3 = 1206$  observations), one dependent variable with three alternatives for a ratio is 134:1. Thus, the minimum sample size for multivariate statistics was satisfied because several researchers have suggested a minimum observation-to-predictor ratio of 10 to 1, with a minimum sample size more than 50 (Peng et al., 2002).

The age level of the tourists was from 18 to 58 and the average age in the sample was about 35. Of three groups, the 31–45 age range occupied the highest proportion (48.3%), while the first and the third accounted for 31.6% and 20.1%, respectively. Of the 402 tourists, there were 171 male tourists (approximately 43%) while female tourists accounted for slightly above 57%. About the occupation of tourist, of the 402 respondents, 111 tourists worked in state offices, 138 tourists worked in business sector, and the remaining was persons in other fields. Finally, monthly individual income of the tourists in the sample is also displayed in Table 1. The income was very different amongst the tourists, with a spread from VND 1.5 million to 50 million, and the average income VND 8.37 million.

Table 1. Demographic profile of respondents

Demographic variables	Category	Frequency	Percentage (%)
Age	18–30 years old	127	31.6
	31–45 years old	194	48.3
	Over 45 years old	81	20.1
Gender	Female	231	57.4
	Male	171	42.6
Occupation	Business sector staff	138	34.3
	State sector staff	111	27.6
	Other sector staff	153	38.1
Income	Monthly income (million VND)	8.37	-

### 5.2. Travel mode chosen by tourists

Frequencies and percentages for the departure places and modes chosen by the tourists are reported in Table 2. Of the 402, the number of Hồ Chí Minh tourists occupied the highest market share (53.2%), followed by the Hà Nội tourists (26.1%), and finally the Đà Nẵng tourists (20.6%). The unbalance proportions of the sample might be related to the distance from these cities to Nha Trang. In fact, the distance between Hồ Chí Minh city and Nha Trang is the shortest (say, 435 km) whereas the distance between Hà Nội and Nha Trang is the longest (say, 1053 km).

With regard to the travel mode chosen, the number of the tourists choosing plane for the journey was the lowest while the figure of coach was the highest. For Hồ Chí Minh market, coach was the most chosen mode whereas, for Đà Nẵng market, the number of the tourists chose train with the highest proportion. In contrast to the two markets above, the majority of Hà Nội tourists chose plane for the trips. Thus, the demand of modal choice for tourism travel amongst the tourists in the three markets has an interestingly obvious difference. While the Hồ Chí Minh tourists seem to refer coach, the Đà Nẵng tourists seem to refer train, the Hà Nội tourists seem to refer plane for their tourism trip to Nha Trang.

Table 2. Frequencies and percentages for departure places and modes chosen

Mode	Hồ Chí Minh city		Đà Nẵng		Hà Nội		Total	
	Frequency	%	Frequency	%	Frequency	%	Frequency	%
Plane	35	16.4	17	20.5	50	47.6	102	25.4
Coach	115	53.7	30	36.1	22	21.0	167	41.5
Train	64	29.9	36	43.4	33	31.4	133	33.1
Total	214	53.2	83	20.6	105	26.1	402	100.0

### 5.3. Descriptive statistics of the alternative-specific regressors

Table 3 shows that there are significant differences in average out-of-mode travel times per kilometer. Amongst the modes, the mean out-of-plane spent time was the highest, up to over 0.22 minute/km, while these times for coach and train were much low, at about one-third the time of plane. The average out-of-coach spent time per kilometer was the lowest, probably because many coach companies have promotion policies that involve picking up the passengers at their homes and dropping off the passengers at the destinations. As a result, in many cases, the out-of-coach time was equal to zero. Contrast to the out-of-mode travel times, the average on-plane spent time per kilometer was the lowest of these modes, whereas those times for coach and train were similar to each other (approximately 1.21minutes/km) and many times greater compared with that of plane.

With regard to per-kilometer cost to income ratio, the data in Table 3 indicate that on average the ratio for coach was the lowest (0.08%), while the ratio for plane was 0.039%, approximately five times higher in comparison to the ratio of coach, and threefold higher against the ratio of train. Plane being the most expensive mode with the highest ratio, it has a cost-plane disadvantage. However, this drawback is compensated by travel time economy, safety, and other attributes.

Table 3. Descriptive statistics of the alternative-specific regressors

Alternative-specific variable	Mean	Std. dev.	Minimum	Maximum
Out-of-mode travel time per km (in min/km)				
Out-of-coach travel time per km	0.06	0.04	0.00	0.34
Out-of-train travel time per km	0.08	0.05	0.02	0.41
Out-of-plane travel time per km	0.22	0.11	0.06	0.57
On-mode travel time per km (in min/km)				
On-coach travel time per km	1.21	0.19	0.96	1.63
On-train travel time per km	1.21	0.19	0.94	1.60
On-plane travel time per km	0.12	0.01	0.10	0.13
Per-km travel cost to income ratio (%)				
Coach cost per km to income ratio	0.008	0.004	0.001	0.033
Train cost per km to income ratio	0.012	0.007	0.001	0.048
Plane cost per km to income ratio	0.039	0.026	0.005	0.218
Comfort (binary variable)			0	1
Safety (binary variable)			0	1
Punctuality (binary variable)			0	1
Price/quality (binary variable)			0	1
Luxury (binary variable)			0	1

Apart from out-of- mode travel time per kilometer, on-mode travel time per kilometer, and per-kilometer travel cost on income ratio, the remaining variables in Table 3 are binary variables reflecting the travel mode quality. This study applied the “asmprobit” command in the STATA/SE12.0 for multinomial probit regression.

## 6. Results

Before constructing the model, a frequency analysis and the Mahalanobis distance method were carried out to detect outliers and numerical problems that can occur in the multivariate analysis. The findings indicated that there were no outliers or numerical problems. Furthermore, a test of multicollinearity amongst the independent variables was also conducted through testing of tolerance and variance inflation factor (VIF) values. Calculation of these values is similar to multiple regression analysis. That means the diagnostic information for multicollinearity can be obtained by running an ordinary least squares regression model using the same dependent and independent variables that are being used in the probit regression model. Menard (2002, p. 76) states that “Because the concern is with the relationship amongst the independent variables, the functional form of the model for the dependent variable is irrelevant to the estimation of collinearity”. The findings of the test show that the tolerance values of all independent variables were close to 1 and the VIF values of these variables were quite small at less than 3.0 (see Appendix A), suggesting that there is no multicollinearity amongst the independent variables in this study.

Estimation results of the MNP model are displayed in Table 4. According to the results obtained, the relationship between the modal choice variable and the set of independent variables is significant. The MNP model is much better than only a constant model. Further, it has a relatively high pseudo- $R^2$  value of 0.3838, indicating that approximately 38.38% of the variation in the dependent variable (modal choice) can be explained by the estimated MNP regression model. Thus, it can be concluded that the MNP model fits the sample data well.

On the other hand, all of the alternative-specific variables are highly significant, with expected signs except the luxury. Of the case-specific variables, the income and the state-sector staff variables show significant effects for both the plane/coach and train/coach equations. The remaining variables have certain significant effects on the modal choice. The following subsections will analyse in more detailed the results.

Table 4. Estimated coefficients and the corresponding t-Statistics in the MNP model.

Alternative-specific variables	Coefficient	t-Statistic		
Out-of-mode travel time per km	-6.2672***	-4.38		
On-mode travel time per km	-1.9560*	-1.94		
Travel cost per km to income ratio	-4.8837***	-2.92		
Comfort	0.2249***	2.68		
Punctuality	0.2791***	2.98		
Safety	0.6454***	5.48		
Price against quality	0.3231***	3.14		
Luxury	-0.1643	-1.55		
			Plane/Coach	Train/Coach
Case-specific variables	Coefficient	t-Statistic	Coefficient	t-Statistic
Income	0.0890**	2.03	0.0916**	2.36
Gender				
Female	-0.4869*	-1.88	-0.3461	-1.58
Age				
18 - 30 years old	-0.3051	-0.79	0.4587	1.37
31 - 45 years old	0.0162	0.05	0.6750**	2.01
Occupation				
Business sector staff	-0.2159	-0.7	-0.0599	-0.25
State sector staff	0.7178**	2.4	0.4892**	1.95
Distance				
Hồ Chí Minh city	0.5560	1.07	-0.7200***	-2.75
Đà Nẵng	0.1773	0.28	-0.4249	-1.45
Constant	-1.2836	-0.75	-0.4691	-0.93

Number of observations: 1206; Number of cases: 402; Wald  $\chi^2(24) = 97.98$ ;

Probability  $> \chi^2 = 0.0000$ ; Log simulated-likelihood = -267.25;

Pseudo- $R^2 = 0.3838^5$ ; \*\*\*Significant at the 1% level;

\*\* Significant at the 5% level; \* Significant at the 10% level.

<sup>5</sup> The Pseudo- $R^2$ , is computed by using the formula  $1 - (L_N(\hat{\beta})/L_N(\bar{y}))$ , where  $L_N(\hat{\beta})$  denotes the maximized log-likelihood value, and  $L_N(\bar{y})$  denotes the value of the log likelihood in the intercept-only model (Cameron and Trivedi, 2009). Here  $L_N(\hat{\beta}) = -267.25$ , and estimation of an intercept-only obtains  $L_N(\bar{y}) = -433.71$ , so Pseudo- $R^2 = 0.3838$



### ***6.1. Role of travel time and travel cost on modal choice of tourists***

Table 4 shows that both coefficients of the out-of-mode per kilometer and the on-mode travel time per kilometer are negative and statistically significant, as expected. However, the effect of the out-of-mode travel time on choosing the travel mode is much stronger than that of the on-mode travel time. The higher negative coefficient on the out-of-mode travel time compared to the on-mode travel time indicates that the out-of-mode travel time is more onerous than the on-mode travel time (Bhat, 1998). These figures suggest that if the out-of-mode and on-mode travel time of a specific mode for the journey increase, then demand for that mode will decrease and demand for all the other modes will increase. For example, if out-of-coach travel time per kilometer for the journey to Nha Trang increases, then demand for the coach will go down and, in contrast, demand for train and plane will rise.

Alternatively, the results also confirm that the costs are important for modal choice. The magnitude of the coefficient is relatively large (i.e. -4.8837), so the effect is strong. The utility of a mode and the probability that it will be chosen decreases as the per-kilometer cost to income of that mode increases.

These results recommend that tourism transport firms can attract tourists by shortening travel time, especially waiting time outside of mode, and offering a reasonable price policy. The shorter the out-of-mode travel time, the more attractive that mode is. We can already see this happening with coaches in picking up and dropping off their tourists from the tourists' homes.

### ***6.2. Role of quality and pricing factors on modal choice of tourists***

The figures in Table 4 show that the service quality variables actually have important effects on modal choice of tourists. The four variables, namely comfort, punctuality, safer, and price against quality are statistically significant and their signs are suitable with the expectation, suggesting that the utility of tourists is increasing in modal service quality. However, the sign of the luxury variable is not as expected and not significant.

Of the four variables, the safety variable has the strongest impact, suggesting that safety of travel mode is one of the most crucial criteria when considering mode choice and this completely matches the reality. As a result, the safer the mode, the more likely it is that the mode will be chosen by tourists. On the other hand, in the condition that all other factors hold true, a mode that departs and arrives at a destination punctually, provides tourists with more

comfort, and offers a reasonable travel price with service quality is more likely to be chosen. Therefore, to augment potential demand from tourists, tourism transport companies should consider promoting a safe, comfortable, and punctual image of tourism travel modes.

### 6.3. Role of income on travel mode choice

As expected, high-income travellers are significantly less likely to choose coach. It can be seen that the tourists who chose the coach had the lowest average income (VND 5,679,000), while the tourists who had the highest average income (VND 12,721,000) chose the plane. The tourists who chose the train for their trip had a monthly average income of VND 8,410,000. Clearly, the differences in the monthly income of the tourists led to differences in modal choice decision. Tourists with higher income seem to like choosing plane or train than coach when going on tourism. Furthermore, the results of the MNP regression in Table 4 also support further that interpretation. As expected, the estimation coefficients of income in the plane/coach and train/coach equations are respectively 0.089 and 0.092, and both are statistically significant at the 5% level. This implies that an increase in income of the tourists could lead to an increase in the choice probability of plane or train relative to coach. In addition, as analysing the marginal effect of a change in income on the modal choice probabilities, the results continue to support the above explanatory.

Table 5. Marginal effects of a change in income on the modal choice probability of tourists

Mode	Probability (choice = mode)	Marginal effect	t-Statistic	Average monthly income (VND 1000)
Coach	0.4286	-0.0297	-2.72	5,679
Plane	0.1495	0.0063	1.10	12,721
Train	0.4219	0.0234	2.65	8,410

The figures in Table 5 show that the choice probability of coach and train is quite similar, slightly more than 42% while that of plane is relatively low, reaches 14.95%. Although the coach choice probability is the highest, an extra increase in monthly income of tourists will lead to a strongly reduce in the probability of the choice of coach and an increase in the probability of the choice of either train or plane. Therefore, the tourists' modal choice decision largely depends on their monthly income, and the higher income tourists are more likely to choose plane or train for their trip against coach.

#### ***6.4. Effects of demographic variables and city dummies on modal choice of tourists***

The gender is a significant predictor of modal choice. The finding suggests that the female tourists prefer to travel by coach rather than by plane for the trip. In addition, between the age groups, age 18–30 and age 31–45, only the age 31–45 group has a significant effect on modal choice, suggesting that this group prefers travelling by train to travelling by coach. On the other hand, as regard to the occupation, the business group is not found to have a significant effect on the modal propensity. Conversely, the state group has significant effects on both the modal choice equations. The tourists who are working in state sector are more likely to choose plane or train for their tourism journey than coach.

With respect to the city, only the Hồ Chí Minh has significant impact on train/coach alternative while the effect of the Đà Nẵng variable is insignificant on both of the equations. The Hồ Chí Minh tourists prefer coach for their trip than coach, which is not surprising, being the city closes to Nha Trang.

### **7. Conclusions and implications**

This study has given a detailed picture of tourists' modal choice behavior for their journey to Nha Trang. The use of random utility models of discrete choice is a major advantage in travel demand modelling. Applying the MNP analysis for this study helped to overcome the limitations of the MNL analysis because it allows the relaxation of the IID property.

The objective of this analysis was to examine and explore the critical factors affecting the tourists' modal choice. In this analysis out-of-mode and on-mode travel time per kilometer, per-kilometer travel cost to income ratio, safety, comfort, punctuality, price/quality, and income are the variables having important contributions to explaining the tourists' modal choice decision. In addition, some of the categorical variables such as gender, age, and occupation of tourists as well as distance also have certain important effects on this decision.

The results of the regression indicate that shorter per-kilometer travel time, especially out-of-mode travel time, as well as lower per-kilometer travel cost of a specific mode would increase tourists' the demand. In addition, the service quality of mode is also a key element influencing the modal choice, and the tourists with a higher income tend to choose plane or

train. These findings might be helpful for tourism transportation firms in offering the relevant strategies with tourists' modal choice demand.

**Appendix A: Multi-collinearity diagnostics**

Label	Explanation	VIF	1/VIF
outtd	Out-of-mode travel time per km	2.91	0.3433
onttd	On-mode travel time per km	2.72	0.3678
cdi	Travel cost per km to income ratio	2.41	0.4155
age1	18–30 years old	2.06	0.4857
age2	31–45 years old	1.91	0.5231
hcm	Hồ Chí Minh city	1.86	0.5381
dan	Đà Nẵng	1.59	0.6276
inc	Income	1.52	0.6589
occsa	State sector staff	1.41	0.7116
occb	Business sector staff	1.28	0.7801
lux	Luxury	1.15	0.8715
pq	Price against quality	1.15	0.8720
saf	Safety	1.12	0.8913
gen	Gender	1.06	0.9470
pun	Punctuality	1.04	0.9658
com	Comfort	1.02	0.9772

## References

- Alvarez, R.M. and Nagler, J. (1998) When politics and models collide: Estimating models of multiparty competition. *American Journal of Political Science*, 42(1), 55–96.
- Ankeny, M.L. (1991) Evaluating end–user services: Success or satisfaction? *The Journal of Academic Librarianship*, 16(6), 352–356.
- Baltas, G. (2007) Econometric models for discrete choice analysis of travel and tourism demand. *Journal of Travel & Tourism Marketing*, 21(4), 25–40.
- Barff, R., Mackay, D. and Olshavsky, R.W. (1982) A selective review of travel-mode choice models. *Journal of Consumer Research*, 4, 370–380.
- Ben-Akiva, M. and Koppelman, F.S. (1974) Multidimensional choice models: Alternative structures of travel demand models. *Transportation Research Board*, 149, 129–142.
- Ben-Akiva, M. and Richards, M.G. (1976) A disaggregate multimodal model for work trips in the Netherlands. *Transportation Research Record*, 569, 107–124.
- Bhat, C.R. (1995) A heteroscedastic extreme value model of intercity travel mode choice. *Transportation Research Part B: Methodological*, 29(6), 471–483.
- Bhat, C.R. (1998) Accommodating variations in responsiveness to level-of-service measures in travel mode choice modeling. *Transportation Research Part A: Policy and Practice*, 32(7), 495–507.
- Bhat, C.R. (2000) Incorporating observed and unobserved heterogeneity in urban work travel mode choice modeling. *Transportation Science*, 34(2), 228–238.
- Bolduc, D. (1999) A practical technique to estimate multinomial probit models in transportation. *Transportation Research Part B: Methodological*, 33(1), 63–79.
- Chintagunta, P.K. (1992) Estimating a multinomial probit model of brand choice using the method of simulated moments. *Marketing Science*, 4, 386–407.
- Choo, S. and Mokhtarian, P.L. (2004) What type of vehicle do people drive? The role of attitude and life style in influencing vehicle type choice. *Transportation Research Part A: Policy and Practice*, 38(3), 201–222.
- Cohen, A. and Harris, N. (1998) Mode choice for VFR journeys. *Journal of Transport*

*Geography*, 6, 43–51.

- Commins, N. and Nolan, A. (2011) The determinants of mode of transport to work in the Greater Dublin Area. *Transport Policy*, 18, 259–268.
- Daganzo, C.F. and Sheffi, Y. (1977) On stochastic models of traffic assignment. *Transportation Science*, 11, 253–274.
- Dissanayake, D. and Morikawa, T. (2005) Household travel behaviour in developing countries: Nested logit model of vehicle ownership, mode choice and trip chaining. *Transportation Research Record*, 1805, 45–52.
- Dow, J.K. and Endersby, J.W. (2004) Multinomial probit and multinomial logit: A comparison of choice models for voting research. *Electoral Studies*, 23, 107–122.
- Ewing, R., Schroeder, W. and Greene, W. (2004) School location and student travel analysis of factors affecting mode choice. *Transportation Research Record: Journal of the Transportation Research Board*, 1895, 55–63.
- Fesenmaier, D.R. (1988) Integrating activity patterns into destination choice models. *Journal of Leisure Research*, 20(3), 175–191.
- Geweke, J., Keane, M. and Runkle, D. (1994) Alternative computational approaches to inference in the multinomial probit model. *The Review of Economics and Statistics*, 76(4), 609–632.
- Gössling, S., and Hall, C. M. (2006) Uncertainties in predicting tourist flows under scenarios of climate change. *Climatic Change*, 79(3–4), 163–173.
- Habib, K.M.N. (2012) Modeling commuting mode choice jointly with work start time and work duration. *Transportation Research Part A: Policy and Practice*, 46(1), 33–47.
- Hesseln, H., Loomis, J.B. and González–Cabán, A. (2003) The effects of fire on hiking demand: A travel cost study of Colorado and Montana. *USDA Forest Service Proceedings RMRS–P–29*, 177–186.
- Horowitz, J. (1991) Reconsidering the multinomial probit model. *Transportation Research Part B: Methodological*, 25(6), 433–438.
- Johansson, M.V., Heldt, T. and Johansson, P. (2006) The effects of attitudes and personality traits on mode choice. *Transportation Research Part A: Policy and Practice*, 40(6),

507–525.

- Komorita, S.S. and Graham, W.K. (1965) Number of scale points and the reliability of scales. *Educational and Psychological Measurement*, 25(4), 987–995.
- Koppelman, F.S. and Bhat, C.R. (2006) A self instructing course in mode choice modeling: multinomial and nested logit models. Prepared For U.S. Department of Transportation Federal Transit Administration, pp.1-241. [http://www.ce.utexas.edu/prof/bhat/COURSES/LM\\_Draft\\_060131Final-060630.pdf](http://www.ce.utexas.edu/prof/bhat/COURSES/LM_Draft_060131Final-060630.pdf)
- Koppelman, F.S., Bhat, C.R. and Shofer, J.L. (1993) Market research evaluation of actions to reduce suburban traffic congestion: Commuter travel behavior and response to demand reduction actions. *Transportation Research Part A: Policy and Practice*, 27(5), 383–393.
- Kraft, J. and Kraft, A. (1974) Empirical estimation of the value of travel time using multi mode choice models. *Journal of Econometrics*, 2, 317–326.
- Lave, C.A. (1969) A behavioral approach to modal split forecasting. *Transportation Research*, 3(4), 463–480.
- Lerman, S.R. and Ben-Akiva, M. (1976) Disaggregate behavioral model of auto-ownership. *Transportation Research Record*, 569, 34–51.
- Litman, T. (2008) Valuing transit service quality improvements. *Journal of Public Transportation*, 11(2), 43–63.
- McCarthy, P. (1996) Market price and income elasticities of new vehicle demands. *The Review of Economics and Statistics*, 78, 543–547.
- McConnell, K.E. and Strand, I.E. (1981) Measuring the cost of time in recreation demand analysis: An application to sport fishing. *American Journal of Agricultural Economics*, 63, 153–156.
- McCulloch, R. and Rossi, P.E. (1994) An exact likelihood analysis of the multinomial probit model. *Journal of Econometrics*, 64, 207–240.
- McFadden, D. (1974) Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in Econometrics* (pp. 105–142). Academic Press, New York.

- McFadden, D. (1978) Modelling the choice of residential location. *Transportation Research Record*, 673, 72–77.
- McFadden, D. (1989) A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica*, 57, 995–1026.
- Menard, S.W. (2002) *Applied logistic regression analysis*. Series: Quantitative Applications in the Social Sciences. Saga University Papers, volume 106.
- Nicolau, J.L. and Más, F.J. (2006) The influence of distance and prices on the choice of tourist destinations: The moderating role of motivations. *Tourism Management*, 27(5), 982–996.
- Paap, R. and Franses, P.H. (2000) A dynamic multinomial probit model for brand choice with different long-run and short-run effects of marketing-mix variables. *Journal of Applied Econometrics*, 15(6), 717–744.
- Peng, C.Y.J., Lee, K.L. and Ingersoll, G.M. (2002) An introduction to logistic regression analysis and reporting. *The Journal of Educational Research*, 96(1), 3–14.
- Quarmby, D.A. (1967) Choice of travel mode for the journey to work: Some findings. *Journal of Transport Economics and Policy*, 3, 273–314.
- Rassam, P.R., Ellis, R.H. and Bennett, J.C. (1971) The n-dimensional logit model: Development and application. *Highway Research Record*, 369, 135–147.
- Recker, W.W. and Golob, T.F. (1979) A non-compensatory model of transportation behavior based on sequential consideration of attributes. *Transportation Research Part B: Methodological*, 13(4), 269–280.
- Schutz, H.G. and Rucker, M.H. (1975) A comparison of variable configurations across scale lengths: An empirical study. *Educational and Psychological Measurement*, 35, 319–324.
- Swait, J. and Adamowicz, W. (2001) The influence of task complexity on consumer choice: A latent class model of decision strategy switching. *Journal of Consumer Research*, 28(1), 135–148.
- Swait, J. and Ben-Akiva, M. (1987) Empirical test of a constrained choice discrete model: Mode choice in São Paulo, Brazil. *Transportation Research Part B: Methodological*,



21(2), 103–115.

Train, K. (1980). A structured logit model of auto ownership and mode choice. *The Review of Economic Studies*, 47, 357–370.

Train, K. (1998) Recreation demand models with taste differences over people. *Land Economics*, 74(2), 230–239.

Train, K. (2009) *Discrete choice method with simulation*, Second Ed. Cambridge University Press, New York.

Weeks, M. (1997) The multinomial probit model revisited: A discussion of parameter estimability, identification and specification testing. *Journal of economic surveys*, 3, 297–324.

Williams, H.C.W.L. (1977) On the formation of travel demand models and economic evaluation measures of user benefit. *Environment and Planning A*, 9(3), 285–344.

Zorn, L., Sall, E. and Wu, D. (2012) Incorporating crowding into the San Francisco activity-based travel model. *Transportation*, 39, 755–771.