Response to the feedback comments and changes to the preliminary version

2 Content

Section 3.2:
- Paragraph 1:

1. Added ‘where J_n is the number of alternatives in the choice set to an individual denoted n’, to define.
2. Changed the wording of ‘Consequently, the difference in the disturbances $\varepsilon_n = \varepsilon_{in} - \varepsilon_{jn}$ follows the logistic distribution.’ to ‘A property of the Gumbel-distribution is that the difference in the disturbances $\varepsilon_n = \varepsilon_{in} - \varepsilon_{jn}$ follows the logistic distribution (Ben-Akiva and R.Lerman, 1985)’

3 Exposition and Structure

- Errors such as using ‘analysis’ instead of ‘analyse’ have been changed, for example on page 5; ‘Lerman and Manski (1979) analysis the sampling. . . ’ changed to ‘Lerman and Manski (1979) analyse the sampling. . . ’.
- Throughout the paper when referring to tables and sections, ‘see table...’ and ‘see section. . . ’, have been changed to the ‘see Table...’ and ‘see Section. . . ’
- Reordered the reference and appendix sections.
- Removed articles that are not cited in the paper from the reference list.
- The reference to the article in section 5.1 was in the list of the references. However, was cited and entered in the reference list as in accordance stated by the authors on the paper. This now has been changed in line with the rest of the paper.

Overall, no major changes have been made. To correct the repetition and to improve the clarity, sections would need to be rewritten or extended, however, as the paper has page restraint and has marked highly given its current content and structure this has been left and noted for future projects.

Discrete Choice Models and Application with an Empirical Study: Students Mode of Transport to Lectures

Serena Patel
24718319
University of Southampton

January 2013

1 Abstract

Many researchers recognise the power of discrete choice models to determine decision makers' behaviour. This paper outlines the basic discrete choice models, the linear probability model, the logit model and the probit model, focusing on how the binary logit model can be used in practice to predict the probability of a particular option being selected. Through two published examples of real life applications, the use of the discrete choice models in determining brand credibility affect on a consumers choice of purchase, and distinguishing health workers preferences to treatments have been addressed. A simple empirical study has been conducted to determine how distance and other attributes affect students' choice of travel mode to university for lectures, identifying problems that arise with fitting sample data to discrete choice models. The study illustrates that the basic discrete choice models are not a good fit for the small sample data used, but identifies that increasing distance traveled reduces the probability a student chooses to walk to lectures.

2 Introduction

Choice models are developed to model decisions taking into account influencing factors and are applied throughout society. Without restriction choice can be very complex. Consider the decisions each individual makes throughout a single day, what food to consume, what to wear, what TV pro-gramme to watch and there are so many more. We can restrict choice such that a decision maker chooses from a set of discrete alternatives and model the behaviours. Choice models, most often, are applied to consumer choice theory in the economical and market sense, such that organisations analyse the behaviour of consumers to investigate the impact and benefits of introducing new products. Such case is the choice of mode of transport. In the UK, the Department of Transport have expert analysts developing choice models and interpreting the effect of new public travel modes and the potential demands.

There is vast amount of academic literature on discrete choice models dating back to the 1920's when the concept of random utility was developed by Thurstone (1927). Ben-Akiva and Lerman (1985) and Train (1986) review discrete choice models with applications to travel demand investigating practicality of model assumptions to real sample data. Researchers such as Daniel L. McFadden have produced a list of papers (Daniel L. McFadden, 2011) using and developing discrete choice models for different areas of study. For example, Mcfadden (1982) review economic applications (Ben-Akiva and Lerman, 1985) and McFadden (1974) review urban travel demands.

This paper aims to convey the basic ideas of discrete choice modelling with simple descriptions and there uses. In the first part of this paper, basic discrete models are outlined, including binary and multinomial discrete choice models, where the choice of two alternatives or more than two alternatives, respectively, are available in the choice set. Theoretical models are applied to real data and used in practice such that behavioural interpretations can be made. Particular examples from previous studies that identify the significance of factors that influence health staff attitudes and brand credibility are described in Section 3. Section 4 explores how to fit sample data to a binary discrete model to determine the probability of an alternative being chosen. The second part of the paper uses real sample data collected from 20 university students to investigate how distance and other factors may affect University of Southampton students choice of mode of transport to travel to University for lectures. The data has been modelled to basic discrete models with interpretations of the results. With later discussion of the potential impact of a free bus pass given to all students on travel mode choices and whether the cost of doing so would be beneficial.

3 Basic Discrete Models

In this section we introduce three basic discrete choice models, the linear probability model, the logit model and the probit model. Firstly, we to distinguish the components that contribute to the development of the models. Discrete choice models attempt to assign probability to an alternative being selected based on reliable information of decision makers' behaviour when presented with a set of choice set of alternatives. For an individual denoted n, the choice set, C_n must be exhaustive, finite and mutually exclusive, such that for the individual all the possible alternatives they could select are considered in the model and such that two alternatives can not both be chosen. When modelling choice we assume that the decision maker will select the alternative which achieves the greatest utility[1] (Ben-Akiva and Lerman, 1985). Utility, U_{in}, can be split into two components and can be expressed as,

$$U_{in} = V_{in} + \varepsilon_{in}, \quad \forall i \in C_n, \tag{3.1}$$

where V_{in} is the deterministic (or observable) utility, and ε_{in} is the disturbance[2] (or unobservable) utility. The deterministic utility can be divided further into attributes and socioeconomic characteristics that contribute to the utility level. Attributes are factors that influence a decision maker's choice and socioeconomic characteristics are features of the decision maker that may contribute to their decision, for example, in selecting a mode of transport for a commuter, cost and travel time are attributes, whereas the gender and age of the decision maker are socioeconomic characteristics (Train, 2003). Therefore, the deterministic utility, V_{in}, can be expressed as a function of the observable attributes and socioeconomic characteristics,

$$V_{in} = \sum_{k=1}^{K} \beta_k x_{kin}, \tag{3.2}$$

where x_{kin} is the k^{th} observed attribute or socioeconomic characteristic with the unknown coefficient parameter β_k, of the decision maker, n, to select the alternative i. For the rest of the paper, we term both attributes and socioeconomic characteristics as attributes and we denote $\boldsymbol{x} = (x_1, \ldots, x_k)$ and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)$.

Now that we have specified behavioral assumptions, we derive the basis of the basic probability choice models of an alternative being chosen. As a decision maker chooses the alternative with

[1] Utility is an economic satisfaction measure commonly used in economics after Marschak (1960) derived the idea of Random Utility Models (RUM) from Thurstone's (1927) development of psychological stimuli (Ben-Akiva and Lerman, 1985).

[2] The disturbance is the random unknown element contributing to the utility.

the greatest utility, we write the probability that a decision maker chooses an alternative i from a set of choices C_n as,

$$P_n(i) = P(U_i > U_j),\ \forall j \in C_n,\ j \neq i. \tag{3.3}$$

This shows that the overall value of utility does not generally matter in this case, as it is the comparison of the individuals preference of the alternatives that is considered, therefore we assume that the behaviours of individuals are monotonic, continuous and complete (Mylene, et al., 2011). By substituting the utility split into deterministic and disturbance variables we write,

$$P_n(i) = P(V_{in} + \varepsilon_{in} > V_{jn} + \varepsilon_{jn}),\ \ \forall j \in C_n,\ j \neq i, \tag{3.4}$$

$$P_n(i) = P(V_{in} - V_{jn} > \varepsilon_{jn} - \varepsilon_{in}),\ \ \forall j \in C_n,\ j \neq i. \tag{3.5}$$

Therefore, we say that the probability of i being selected is the probability that the observable difference between the alternatives is greater than the difference of the disturbances. It is the distribution of the disturbances that are unknown and give rise to variations of discrete choice models from assuming that the difference of the disturbances, $\varepsilon_n = \varepsilon_{in} - \varepsilon_{jn}$ follow different distributions. We now consider the three basic discrete choice models.

3.1 The Linear Probability Model

The linear probability model (LPM) is the simplest discrete choice model, with $\varepsilon_n = \varepsilon_{in} - \varepsilon_{jn}$ uniformly distributed between two values $-L$ and L. The binary LPM for the probability alternative i being chosen over alternative j is expressed as,

$$P_n(i) = \begin{cases} 0 & \text{if } V_{in} - V_{jn} < -L, \\ \frac{V_{in} - V_{jn} + L}{2L} & \text{if } -L \leq V_{in} - V_{jn} \leq L, \\ 1 & \text{if } V_{in} - V_{jn} > L, \end{cases} \tag{3.6}$$

where $-L < \varepsilon_n = \varepsilon_{in} - \varepsilon_{jn} < L,\ \ L > 0$, and V is as in (3.2), (Ben-Akiva and Lerman, 1985). The LPM is generally helpful to calculate rough estimates of the attribute coefficients, as the ordinary least square estimator (OLS) method can be used and calculated with ease; however, the model has a number of flaws to being a discrete choice model. Firstly, when using the OLS, heteroscedasticity may arise as we assume the variance of ε_n to be constant which is unlikely to be the case using real discrete choice data, and also fails the OLS criterion that assumes the ε_n is normally distributed. Other problems include the forecasting of probabilities of i outside the range used to estimate $\boldsymbol{\beta}$ (Börsch-Supan, 1987), as the model predicts that there is zero probability outside $[-L, L]$, otherwise a probability of less than 0, or greater than 1 could be given, clearly dissatisfies probability axioms. This may be impractical for researchers who are investigating the impact of demand due to the introduction of a new product.

The main problem of the linear probability model appears to be the restrictions associated with the rigid shape of the models function. Next we consider the logit and probit models which both have softer S shaped functions that are transformations of the linear probability model.

3.2 The Multinomial Logit Model

The multinomial logit model was originally developed by Luce (1959) (Train, 2003) based on choice probabilities, and later derived by Marschak (1960). It is expressed as,

$$P_n(i) = \frac{e^{V_{in}}}{\sum_{j \in C_n} e^{V_{jn}}}, \tag{3.7}$$

which is reduced to the binary logit model with two alternatives in the choice set when $J_n = 2$, where J_n is the number of alternatives in the choice set to an individual denoted n (Ben-Akiva and R.Lerman, 1985). Each ε_{in} is assumed to be independently, identically and Gumbel-distributed (follows the extreme value distribution). A property of the Gumbel-distribution is that the difference in the disturbances $\varepsilon_n = \varepsilon_{in} - \varepsilon_{jn}$ follows the logistic distribution (Ben-Akiva and R.Lerman, 1985). This model may be intuitive for a discrete choice model, as it is assumed decision makers select alternatives based on utility maximisation, and the extreme value distribution shifts probability to the extreme values, such that it has fatter tails, and larger choice probability for very small and very big values (Börsch-Supan, 1987). The logit model is conveniently computable by estimating the K unknown parameters, $\boldsymbol{\beta} = \{\beta_1, \beta_2 \ldots \beta_K\}$, through the maximum likelihood estimator method (see Section 4). This method also allows relative ease when large numbers of alternatives or observations are considered.

The multinomial logit model can incorporate taste variations[3], however is limited to accounting for the observable utility variables, but not the unobservable (Train, 1986). Other problems of the model arise from the assumption that the disturbances are independent, as it leads to assuming that all the attributes and alternatives are independent too. The problem is well-known and called 'Independence of Irrelevant Alternatives' (IIA), (Börsch-Supan, 1987). Consider the blue-red bus paradox (Ben-Akiva and Lerman, 1985), where the probability of an individual to travel by a red bus or by car is 0.5 each. By introducing a new alternative, a blue-bus that is identical to the red-bus other than the colour, you would expect the probability that an individual chooses the red or blue bus to be 0.25 each and for the car to have probability of 0.5. The logit model however would consider the new blue-bus alternative as an independent choice assigning the probability of $\frac{1}{3}$ to each alternative. Therefore, when modelling choice with the logit model, the attribute and alternatives should be independent of one another. There have been several adaptions of the model to overcome IIA problem, such as the nested logit model. In recent years the nested logit model has also been developed further to handle relaxed assumptions. Schuessler and Axhausen (2007), describe these models in further detail, however this is beyond the scope of this paper.

3.3 The Probit Model

The probit model is similar to that of the logit model, however it is derived from assuming the disturbances, ε_{in}, follow the standard normal distribution $N(0, \sigma^2)$. This leads to the binary probit model expressed as,

$$P_{in} = \Phi\left(\frac{V_{in} - V_{jn}}{\sigma}\right), \tag{3.8}$$

where Φ denotes the standardised cumulative normal distribution (Ben-Avika and Lerman, 1985). The multinomial probit model, however, is not easily written in closed form as the disturbances of the alternatives have joint multivariate normal distributions. The probit model is more flexible than the logit model, because the IIA assumption is relaxed as the model allows for correlation patterns between the disturbances. This is due to the fact that the model does not depend on the individual σ_i's, but on σ (Ben-Akiva and Lerman, 1985). The probit model also allows for taste variations in both the observable and unobservable utility variables as the coefficients can change with the attributes (Daganzo, 1979). In the case of the binary probit model the parameters are relatively easy to calculate, and are quite precise as the normal distribution has relatively thin tails, however, with increasing the number of alternatives the computation of the parameters becomes increasingly difficult as each iteration of integration needs to be calculated (Börsch-Supan, 1987). This is overcome by the vast number of computer programs that allow the ease of such calculations.

[3] Taste variations reflect that each decision maker can value attributes differently to an alternative (Train, 1986). For example, a commuter may value the time attribute for a mode of transport higher than a student, who may be more concerned about the price.

There are positives and negatives to both the probit and logit models. In past years, the logit model is commonly used in practice due to the ease of computation and interpretation. In the following sections we will focus on the use of the logit model.

4 Binary Logit Model in Practice

In this section we describe how a binary logit model may be fitted in practice where real data is observed, and can be used to determine the probability of an alternative being chosen.

In a study, sample data is collected to represent the population of interest, as surveying all individuals of large populations can be time consuming, costly and may not be feasible. There are many sample strategy's to consider when performing a choice model experiment. There are three main sampling methods; random sampling[4], where the sample is randomly selected from the whole population, exogenous sampling, where samples are chosen such that a range of the attributes are considered, and choice-based sampling, where the samples are selected from the alternatives such that each alternative is represented. Lerman and Manski (1979) analyse the sampling methods in further detail. Samples collected can be based on revealed preference data or stated preference data. Revealed preference data is collected on the basis of real situations where the decision makers actually select an alternative, whereas stated preference data is the alternative that decision makers would select given a hypothetical situation (Train, 2003).

Assuming we have conducted an experiment survey from N individuals to formulate sample data with two alternative choices, A and B, with deterministic utilities V_A and V_B respectively, we have a binary situation, and can fit the data to a binary model. If K attributes, $\boldsymbol{x} = x_1, x_2, \ldots, x_K$, were observed which influence the decision makers choice, we can use (3.2) and write,

$$V_{An} = \sum_{k=1}^{K} \beta_k x_{Ank}, \quad \text{and} \quad V_{Bn} = \sum_{k=1}^{K} \beta_k x_{Bnk}. \tag{4.1}$$

From the sample data we have a set of data with the $\boldsymbol{x}$ variables[5] and the decision makers choice of A and B both known[6]. It is then left for the unknown parameters $\boldsymbol{\beta} = \{\beta_1, \beta_2, \ldots, \beta_K\}$ to be estimated. If we assume the random sampling method has been adopted we can use the maximum likelihood estimator method.

The likelihood equation is a joint function of the observed variables and the unknown parameters. For a discrete binary choice model the likelihood function is expressed as,

$$L = \prod_{n=1}^{N} P_n(A)^{V_{An}} P_n(B)^{V_{Bn}}. \tag{4.2}$$

For a binary logit model such that $P_n(A) + P_n(B) = 1$,

$$P_n(A) = \frac{1}{1 + e^{-\beta x_n}}, \quad \text{and} \quad P_n(B) = \frac{e^{-\beta x_n}}{1 + e^{-\beta x_n}}, \tag{4.3}$$

(Ben-Akiva and Lerman, 1985). The maximum likelihood estimators $\hat{\boldsymbol{\beta}} = \{\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_K\}$ are found by maximising the value of the likelihood equation (4.3). As the equation is a product,

[4] Random sampling is based on the IIA assumption presented in Section (3.2) and therefore it is specific to the logit model (Börsch-Supan, 1987).

[5] The same attributes can be applied to both choices such that if an attribute does not affect one of the attributes the coefficient will be zero eliminating it from the equation.

[6] We are assuming that the two choices are available to all N decision makers, which in practice may not be feasible as for example, a commuter may not choose bus if there is no service in their area.

we can take the logarithm to obtain the log-likelihood equation,

$$logL = \sum_{n=1}^{N} logP_n(A)^{V_{An}} + \sum_{i=1}^{N} logP_n(B)^{V_{Bn}}, \tag{4.4}$$

$$\text{so, } \; logL = \sum_{n=1}^{N} V_{An} logP_n(A) + \sum_{n=1}^{N} V_{Bn} logP_n(B). \tag{4.5}$$

(4.5) can be differentiated with respect to each β_k and set to zero (4.6),

$$\frac{dlogL}{d\boldsymbol{\beta}} = 0, \tag{4.6}$$

to find the value of each $\hat{\beta}_k$ when (4.5) is maximised. Under most conditions the likelihood function is globally concave and therefore yields a unique value for each $\hat{\beta}_k$, which is the estimate of the parameter β_k. Sometimes finding the maximum likelihood estimate parameters can be quite difficult, therefore mathematical computer programs, such as SAS, can be used. The Logistic Procedure in SAS can be executed by using the 'PROC LOGISTIC' command. The output displays a number of tables, of which the one labeled 'Analysis of Maximum Likelihood Estimates' shows the estimates to the unknown parameters, as well as the standard error, Wald chi-square statistics and the chi-square p-value of the estimates. The model can be tested for 'goodness of fit' to the data by test statistics against the null hypothesis. The likelihood ratio test statistic is chi-square distributed given by $-2(logL(\mathbf{0}) - logL(\hat{\boldsymbol{\beta}}))$ where, $logL(\mathbf{0})$ is log likelihood equation when parameters $\boldsymbol{x}$ are zero, and $logL(\hat{\boldsymbol{\beta}})$ is the maximum value of the log likelihood equation (Cramer, 1991). The SAS output provides this test statistic. Ben-Akiva and Lerman (1985), state that the likelihood ratio may not be a good test as it can often reject the null hypothesis at low significant levels. We can use the Wald chi-square test statistic[7] and corresponding p-values to test the individual estimate statistics against the null hypothesis.[8]

The estimated parameters can be interpreted in terms of odds. Odds are the chances of an outcome occurring, for example odds are used in betting on races where there may be 2:1 odds of a contender winning. In this binary case the odds of A being selected is, $P(A)P(B)$. This is related to the estimates as each estimate represents the change in the log of the odds of a unit change in the attribute. The estimates represent the change in log of odds opposed to directly odds as the estimates have been fitted using logits.

Once $\hat{\boldsymbol{\beta}}$ has been calculated, they can be substituted into (4.3) to give,

$$P_n(A) = \frac{1}{1 + e^{-\hat{\boldsymbol{\beta}}\boldsymbol{x}_n}}, \quad \text{and} \quad P_n(B) = \frac{e^{-\hat{\boldsymbol{\beta}}\boldsymbol{x}_n}}{1 + e^{-\hat{\boldsymbol{\beta}}\boldsymbol{x}_n}}. \tag{4.7}$$

Therefore, for an individual, n, with K known attributes, denoted $\boldsymbol{x}$, the probability of A and B can be determined. Note that for a binary probability model the probability of either A or B can be calculated and subtract the answer from 1 to find the other probability. Similarly, the maximum likelihood estimator method can also be used to find the estimates of unknown parameters in the multinomial logit and the binary probit model.

[7] The Wald chi-square statistic is a test statistic that is approximately chi-square distributed, similar concept to the t-statistic which follows the t-distribution. The null hypothesis is rejected if the p-value $< \alpha$ where α is the level of significance. Note that this is for a one sided test.

[8] The null hypothesis is the claim that the parameter in consideration is insignificant at α significance level and is not significantly different from zero.

The probability of a particular alternative being chosen can be useful in real life scenarios. Consider a situation where a university contemplate distributing bus passes to all students. If the probability is low that students would chose to use the bus pass, executing the proposed idea may be wasteful. This scenario is discussed further in Section 6. In the next section two choice experiments are described showing how discrete choice models are be used in practice.

5 Applications

Discrete choice models are applied in research to a variety of fields. An obvious application is in marketing where a consumer chooses between different products, however discrete choice models have a wider application. The first example described considers the use of discrete choice models to determine health workers preferences towards potential changes in treatments for malaria in pregnancy, and the second example describes the influence of brand credibility to consumers choice of purchases.

5.1 'Evaluating Health Workers Potential Resistance to New Interventions: A Role for Discrete Choice Experiments' - Mylene, et al., 2011

Mylene, et al., (2011) conduct a study in Ghana, (Ashanti region), that distinguished potential issues regarding health workers attitudes to a new intervention for malaria in pregnancy. A discrete choice experiment using stated preferences was designed and fit to a logit model, with the potential issues posed as attributes. The discrete choice experiment overcomes the controlling issues when conducting a randomised controlled trial[9], as in the controlled trial the effectiveness differs to that of the operational effectiveness. Since the sample is collected based on stated preference, a discrete choice experiment can be executed while a controlled trial is in action. Under the randomised controlled trial, some patients were screened for malaria, and if a positive result was observed the patient was treated. This is known as intermittent screening and treatment (IST). The patient was treated with either the current drug used, sulphadoxine-pyrimethamine in intermittent preventive treatment (SP-IPT), or with an alternative drug artesunate-amodiaquine (AS-AQ), while other patients were not screened and given three doses of the current drug SP-IPT. All health workers, in 7 of 21 districts, to regionally represent the population were invited to complete a questionnaire on a particular day to select, based on personal opinion, between SP-IPT and IST, with 6 attributes of two levels assessed. The authors fit the observed stated preference data using Nlogit 4.0, and found that the 6 attributes considered were dominated by the health improvements of the new intervention. Through the fitted discrete choice model, predictions of health workers preference of alternative interventions compared to the current methodology were able to be prepared for analysis. This allowed the authors to recognise the responses of different health workers to the approaches and type of drug used. The predictions indicate that staff working at the health facilities for a longer period of time, and midwives would be more resistant to a change in intervention method than other antenatal care staff. However, if maternal and foetal health improvements were observed by the new intervention, there would be little difference in preferences between the health worker professions, who all showed positive attitude to the new intervention. The authors noted the importance of identifying factors that could affect the delivery of new interventions, which through the discrete choice experiment, insight to the attributes can be found. The authors also distinguish the limitations of using discrete choice modeling, such that the model is for simple data and cannot express the full meaning of some attributes. An attribute of the amount of workload was considered to be either 'overloaded' or 'able to cope', which with an alternative intervention, new tasks may be required that the description of workload is not justified or detailed enough. For the purpose of the study, the discrete choice model was

[9] In a randomised controlled trial, the current treatment is used, with the alternative treatment used on some randomly selected patients.

appropriate as the design was controllable allowing attributes that may not currently exist to be considered and forecast potential effects of a new intervention policy.

5.2 'Brand Credibility, Brand Consideration, and Choice' - Erdem and Swait, 2004

Erdem and Swait (2004) investigate the affect brand credibility has on consumer choice. The authors conduct an empirical study in a North American university in which six different product classes were considered; athletic shoes, cellular telecommunication services, headache medication, juice, personal computers and hair shampoo. Stated preference data was collected for the classes, each with 5 brands presented to an individual, inquiring that if all prices are the same which brands would they consider buying, which brand they would most likely buy, and if they were to buy 10 items in the class at one time, how many of each brand would they buy. Information of the individuals' view of each brand regarding 'Expertise', 'Trustworthiness', 'Perceived Quality', 'Perceived Risk' and 'Information Costs Saved' were collected as attributes. The data was firstly fit to two binary logit models taking into account brand consideration with 'Yes' or 'No' as alternatives for each brand and secondly fit to two multinomial logit models taking into account the brand they would consider buying. For both the two binary and two multinomial logit models, the models were created for each class, and one with 'Trustworthiness' and 'Expertise' as attributes to evaluate the brand credibility, and the other for 'Perceived Quality', 'Perceived Risk' and 'Information Costs Saved'. The models allowed the authors to determine which attributes were statistically significant in the models for individuals' choices in each class of product. The study's findings indicate that across all product classes brand credibility through the analysis of 'Trustworthiness' and 'Expertise' are important to brand consideration. It was also found that 'Perceived Quality' is statistically significant across the product classes and has the most impact compared to 'Perceived Risk' and 'Information Costs Saved'. However, the authors' findings suggest overall 'Perceived Risk' and 'Information Costs Saved' have the most impact with higher uncertainty[10] of the product class. The discrete choice models used in this study allowed the authors to propose the significance in brand credibility's impact on consumer choice behaviour.

6 Application to University Students Choice of Transport Mode to Lectures

6.1 Introduction

This section presents an investigation to determine how distance affects Southampton of University students mode of transport choice to travel to lectures. Firstly, a binary logit model is considered with the alternative modes walk or not to walk, with later discussion of a multinomial logit model containing bus, bike or car choice modes available.

6.2 Data Source

Data was collected from a sample of 20 University of Southampton students in November 2013. The survey requested information on single trips to university for lectures using revealed preference data, so that students current behaviour towards their choice of travel mode to lectures is observed and can be analysed. The alternative choices have been defined from knowledge of current modes of transport around the University of Southampton area, with the choice set; 'Walk', or 'Not Walk' including, 'Bike', 'Car', or 'Bus'. The University of Southampton has a student population of around 23,000 (Higher Education Statistic Agency, 2011/12) therefore for the investigation we assume population to be infinite as the population is large compared

[10] Uncertainty indicates that the consumer is not very knowledgable of the class, for example, an individual does not know much about computers.

to the sample size (Ben-Akiva and Lerman, 1985). An exogenous sampling strategy has been adopted such that students invited for observation live a range of distances from the university campus. An online survey[11] was created and more than the required number of students were invited to complete the survey anonymously, in anticipation that not all students would complete the questionnaire, and the first 20 complete surveys to represent a range of distances traveled have been selected to analyse. The sample is mutually exclusive and independent as each observation is taken from a different individual student. The sample may not exhaust the population. This has been attempted by inviting a range of students across academic years and subjects to complete the survey, however sampling issues are discussed later in Section 7. The data collected and used throughout this section can be found in the appendix labeled Table 6. Each model in the following sections have been fitted using SAS 9.3 which uses the maximum likelihood estimation method (see Section 4.1).

6.3 'Walk' or 'Not Walk': Binary Models

Firstly, we consider a binary experiment in which we assume the individuals choose 'Walk', W, or 'Not Walk', Y, with distance, D, being the only attribute. Therefore we can fit the data to a binary logit model as suggested in section 3 as,

$$P_n(W) = \frac{e^{V_{Wn}}}{1+e^{V_{Wn}}}, \text{ where } \ V_{Wn} = \beta_0 + \beta_1 D_{Wn}, \quad n = 1, \ldots, 20.$$

equation For estimation purposes in the SAS, 'Not walk' is the reference variable[12]. The estimation results can be seen in Table 1. The table reports the estimate of the unknown parameters, the constant, $\hat{\beta_0}$, and the distance coefficient, $\hat{\beta_1}$.

Table 1: Estimates for Binary Logit Model: 'Walk' or 'Not Walk' with Distance Attribute

Parameter	DF	Estimate	Standard Error	t Value	Approx Pr > $\|\|t\|\|$
Intercept	1	3.496572	1.832560	1.91	0.0564
Distance	1	-2.202141	1.618637	-1.36	0.1737

The specific-alternative constant[13] estimate is positive implying that students prefer to 'Walk' relative to 'Not Walk'. The sign of the distance coefficient is negative, confirming our intuition that students are less likely to walk the further the distance they travel, and suggests that the logit of the probability of a student walking to lectures decreases 2.202 for each mile further from university. The t statistic for the distance parameter is -1.36 with p-value of 0.1737, therefore at a significant level of 5% the distance is not statistically significant implying that distance is not a significant attribute for a student's choice between 'Walk' and 'Not Walk'. However, this is an unreliable inference due to the small sample size. If two students are considered, one living 5 miles away from university, and the other living 0.1 miles away, it would be unrealistic to assume that the two different distances would not influence the probability for the student living further to not walk than the student living closer. The likelihood ratio of the model is 4.7499 with chi-squared p-value of 0.0293 therefore we can reject the null hypothesis at 5% significant level. This indicates that not all the parameters in the model are insignificant. This gives a contradicting result to the t test statistics given in Table 1. This again may be due to the small sample size, or suggest that the model is not a good fit to the data. In addition the standard errors for both the estimated parameters are relatively large which could indicate that the binary logit model may not be a good fit. The data has been fit to a binary probit model to see if the model is a better fit. The estimate results are in Table 2.

[11] The online survey was created and completed using, University of Southamptons' School of Psychologys' online survey tool 'iSurvey'. The participants accessed the survey directly given an url link.

[12] The reference variable is the variable that the estimates are relative to.

[13] The specific-alternative constant is the intercept parameter.

Table 2: Estimates for Binary Probit Model:'Walk' or 'Not Walk' with Distance Attribute

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	2.1375	1.0330	4.2820	0.0385
Distance	1	-1.3518	0.9481	2.0331	0.1539

The binary probit model results are similar to the binary logit model, with a negative estimate coefficient for the distance parameter and we would not reject the null hypothesis. However, the standard errors of the estimates are smaller than the binary logit model, and the likelihood ratio performs better with a test statistic for the model of 4.9222 with chi-squared p-value 0.0265, signaling that the probit model fits the data better than the binary logit model.

6.4 Four Alternative Modes: A Multinomial Logit Model

To further investigate distance on the choice of transport by students, the alternative choice set has been extended to four options with the alternative travel modes 'Walk', 'Bike', 'Bus', and 'Car'. Each of the 20 university students were asked which mode they currently use most and the data has been fit to a multinomial logit model. The alternative 'Walk' is used as the reference. The results are shown in Table 3.

Table 3: Estimates for Multinomial Logit Model: 'Walk', 'Bus', 'Bike', or 'Car' with Distance Attribute

Parameter	Mode	DF	Estimate	Standard Error	Wald Chi-Squared	Pr > ChiSq
Intercept	Bus	1	-4.3848	2.6791	2.6788	0.1017
Intercept	Car	1	-27.5728	297.7	0.0086	0.9262
Intercept	Bike	1	-3.9058	2.4847	2.4711	0.1160
Distance	Bus	1	2.3686	2.3079	1.0533	0.3048
Distance	Car	1	11.3486	146.9	0.0060	0.9384
Distance	Bike	1	1.9430	2.2127	0.7710	0.3799

The intercept estimates of all three modes are negative, which suggest that students prefer to walk than the other three alternatives when distance is evaluated at zero. The distance coefficient estimates are all positive, indicating that the logit of the probability for each of the alternatives increases relative to 'Walk' as distance increases. This supports the binary logit model results that an increase in distance causes the probability of a student selecting to walk to decrease. However, the coefficient for the variable in utility of the 'Car' alternative is greater than both 'Bus' and 'Bike'. This means that there is a greater increase in the logit of probability of a student choosing 'Car' than the increase in the logit of probability of 'Bus' or 'Bike' with an increase in distance. Hence, reflecting that the further a student travels to lectures the greater the influence they have of travelling by car opposed to other mode alternatives. The model also suggests that with the increase in distance, the probability increase of a student choosing 'Bus' is greater than the probability increase of choosing 'Bike'. We test the model against the null hypothesis using the likelihood ratio test statistic of 9.7882 with chi-square p-value 0.0205, which at 5% significance level we reject the null hypothesis and conclude that not all the parameters in the model are insignificant. However, for each estimate the chi-squared p-values are greater than 0.1 suggesting the null hypothesis can not be rejected. This again may be due to the small sample size causing the results to be unreliable.

6.5 Other Attributes

In the previous two models, the distance parameter has shown to be statistically insignificant to the models, however, both models show that with increasing distance students are less likely to walk. There are other attributes that may affect a students behaviour. Potential factors may include the amount of time taken to travel, the cost of the transport mode, the individuals gender and whether the individual has a bus pass or has a car available for use at university. The extra attributes have been included in the multinomial logit model with the estimate results shown in Table 4, with 'Walk' as the reference variable.

Table 4: Estimates for Multinomial Logit Model: Multiple Attributes

Parameter		Mode	DF	Estimate	Wald Chi-square	Pr > ChiSq
Intercept		Bus	1	-15.7571	0.0592	0.8078
Intercept		Car	1	-14.7047	0.0042	0.9483
Intercept		Bike	1	-8.4093	0.0076	0.9306
Distance		Bus	1	16.1132	0.0081	0.9281
Distance		Car	1	19.6061	0.0060	0.9384
Distance		Bike	1	27.0566	0.0428	0.8362
Time		Bus	1	-1.3763	0.0327	0.8566
Time		Car	1	-0.8809	0.0066	0.9353
Time		Bike	1	-1.4938	0.0463	0.8295
Cost		Bus	1	14.1309	0.0009	0.9766
Cost		Car	1	7.0417	0.0002	0.9892
Cost		Bike	1	-8.3634	0.0000	0.9953
Gender	Male	Bus	1	10.6353	0.0369	0.8476
Gender	Male	Car	1	1.3372	0.0001	0.9922
Gender	Male	Bike	1	0.8000	0.0002	0.9896
Bus pass	Yes	Bus	1	17.2126	0.0386	0.8441
Bus pass	Yes	Car	1	-0.0761	0.0000	0.9998
Bus pass	Yes	Bike	1	0.8372	0.0000	0.9920
Car owner	Yes	Bus	1	3.9306	0.0001	0.9943
Car owner	Yes	Car	1	2.7052	0.0002	0.9875
Car owner	Yes	Bike	1	-2.1046	0.0000	0.9983

We have consistency regarding students preference to walk as shown by the negative parameter estimates of the intercept. Likewise, holding all other attributes constant, with the increase in distance the probability increases that a student will select a different mode of transport than 'Walk'. The 'Time' coefficient estimates are all negative which implies that the longer time taken to travel, the increase in probability that the student selects 'Walk'. The increase in time causes the probability that 'Bike' or 'Bus' is chosen, decreases more than the probability of 'Car' being selected relative to 'Walk'. The increase in cost of travel indicates that a student has more likely chosen to select the alternatives 'Bus' or 'Car' as the corresponding estimates are positive. If a student is Male, the logit of the probability to travel my bus increases by 10.64 relative to walking than for a female, implying, that a male individual has a higher probability of choosing 'Bus' than a female. The estimate of the coefficient of 'Bus pass' suggests that owning a bus pass increases the probability of choosing to travel by bus greater than the increase in choosing to bike, and decreases the probability that the student travels by car. The attribute 'Car owner' does not lead to results that may be intuitively predicted, as the model suggests a student with a car has a greater probability increase to select 'Bus' than the probability increase to select 'Car'. This may indicate that for a student owning a car does not indicate that they will use it to travel to university for lectures. The reader should note that the estimates for each mode of transport are relative to the alternative 'Walk'.

Despite interpreting the estimates, the model may not be statistically significant. The likelihood ratio chi-square test statistic is 33.0338 with a chi-square p-value of 0.0165, therefore we reject the null hypothesis at a 5% significance level. However, the Wald chi-squared values and p-values of all the variables for each mode of transport fail to reject the null hypothesis. Again, this may be due to the small sample size.

6.6 Transport Choice given a Bus Pass

So far we have distinguished that with the sample data none of the attributes appear to be statistically significant. We now examine if distance would be a significant attribute if students are given a free bus pass. Currently, a Uni-link bus pass is given to all students living in student accommodation. A survey was conducted with the same 20 students, inquiring that if they were given a bus pass, which mode of transport would they choose to travel to lectures. The data is fit to a multinomial logit model using stated preference data with 'Walk' as the reference variable.

Table 5: Estimates for Multinomial Logit Model: Free Buss Pass

Parameter	Mode	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	Bus	1	-5.0273	2.3613	4.5329	0.0332
Intercept	Car	1	-29.6579	267.3	0.0123	0.9116
Intercept	Bike	1	-4.5161	3.3925	1.7721	0.1831
Distance	Bus	1	4.4146	2.2003	4.0254	0.0448
Distance	Car	1	13.5116	111.6	0.0147	0.9036
Distance	Bike	1	2.3273	3.3007	0.4971	0.4808

The results shown in Table 5 indicate that there is no real change in students choice of mode to travel to university for lectures, and all the estimates of the intercept are negative implying students still prefer to walk with distance zero. Comparing Table 5 and Table 3 we can see that the coefficient parameter estimates are quite similar. Despite given a bus pass, an increase in distance, increases the probability of a student would prefer the alternative 'Car', greater than increasing a students preference to select 'Bus'. This implies that given a bus pass students would not change their choice behaviour. In addition, the Wald chi-square test statistic and p-values conclude that the intercept for 'Bus', the attribute distance for 'Bus' and attribute distance for 'Bike' are statistically significant at 5% significance level.

6.7 Results

All the models indicate that with the increase in distance, the probability increases that a student would choose an alternative mode of travel to lectures than walking. When considering multiple transport modes with distance (Section 6.4), traveling by car appeared favourable the further the distance. Both these two results were as expected.

The multinomial logit models (Section 6.4 and 6.5) appear statistically insignificant, therefore suggesting that none of the attributes considered have a major influence on students choice of travel mode. This is unexpected as you would naturally assume that at least cost and distance would affect the probability of a mode of transport being chosen. This may be the result of the influence in the model of including the other attributes, or, and most likely due to the small sample obtained. Despite this, through interpreting the estimates the model indicates that for a student, the increase in time traveled and cost, having a bus pass or is male, increases the probability that the student takes the bus, where as for a student who does not own a car or is female increases the probability that the student bikes.

In Section 6.6 we have concluded that the 20 university students surveyed would not dramatically change their travel behaviours to lectures. However, the model does indicate that with a free bus pass, distance is an influencing factor for students choosing the 'Bus' alternative as the variable is statistically significant. With a positive coefficient estimate of 4.4146, (see Table 5) it can be interpreted that with a free bus pass, students are more inclined to take the bus, oppose to walking, to lectures the further they have to travel.

7 Discussion

In this paper we have presented three basic models that discrete choice analysis is based on, outlining particular applications in pre-operational research, consumer brand marketing and choice of transport mode. In particular, we have studied the binary and multinomial logit model which assume the disturbance utility follows the logistic distribution.

Through an investigation we have applied models to determine the effect of distance, and other attributes, to students choice of mode when traveling to university for lectures. The results indicate the increase in distance influences students not to walk. Other researchers have examined similar studies. Nkegbe and Abful-Mumin (2012) investigate the choice of transport to university for students in Ghana who are non-residents of the university they attend using a multinomial logit model. They find that students tend to travel by bus than walk, bike or motorbike, however students who live closer to the university prefer to walk. The result is similar to the study in this paper despite the different location. Further study's could be conducted at other universities to determine if students in general prefer to walk to university the closer they live, and not just at the University of Southampton and the University studied in Ghana.

The reliability of the results in Section 6 may not be sufficient, as there are many flaws in the models derived. Firstly, the models can only be good as the data they are based on. The data collected may not be accurate, as students could have been unconcerned when filling out the survey and quickly estimating the distance or travel time, or alternatively may have not read the questions properly. Distances and times were checked using Google maps (2013) for rough accuracy of the data. If it appeared that an individuals result was extremely inaccurate the individuals data was not used and the next students survey results was considered. This could cause bias' in the data and would need further analysis of the sample to identify correctors that could have been used. In addition the sample collected is small, and may not fully represent the student population of the University of Southampton. It may be the case that the majority of students surveyed walk to lectures, however, there may be student groups who take the bus or bike who's distance traveled have not been considered. To reduce this problem, further research could be conducted using choice-based sampling with on-board surveys and at bike racks. There is much literature on correcting bias' data in sampling (Lerman and Manski, 1979). In future experiments it would be appropriate to use a larger sample size to increase the reliability of the data such that statistically significant variables can be identified.

The assumptions of the models may also cause limitations, as the sample collected was not a random sample and does not fully satisfy the assumptions to use the maximum likelihood estimation and fitting the binary and multinomial logit models we assume IIA holds and included variables in the models that may not be appropriately fitted or significant, which may have skew the results. Further tests, such as the Hausman test for IIA, could distinguish fitting problems, such that the data can be fit to a more appropriate model. We have shown in Section 6.3 through the probit model that other models may fit the data better. In addition, SAS PROC LOGISTIC command has been used to fit the model and estimate the coefficient parameters. The command uses the maximum likelihood estimation, which can generate inconsistent results if the sample data is quasi-complete, and if the data does not overlap the estimates may differ. Webb, Wilson and Chong (2004) investigate the problem further with suggestions of excluding a variable and looking at standard errors to check for this issue.

This study does not clearly identify the affect of distance, and other attributes, to the transport to lectures by students at University of Southampton. Further research with a larger sample would be valuable to increase the significance of the variables. In future studies developed models of the logit model could be examined such that the data can be fit appropriately to a model that optimally represents the population. It may be interesting to continue investigations of the affect on students travel behaviours given a bus pass and the change in significance of the attributes. This could also be extended to specifically study the difference in travel behaviours between those who live in student accommodation against those in private rented. Such a study could be useful to the University of Southampton accommodation services when deciding if a bus pass is useful to students travel to lectures for the distance they travel.

8 Conclusion

This study outlines the use of basic discrete choice models, suggesting that they may be used in simple preliminary research studies before conducting more complicated experiments and models. Through the use of the logit model and applications we have identified that when modelling real data researchers should be cautious of the results due to assumptions made when developing and fitting models, as they may not be appropriate for the data. In the empirical application using sample data from 20 University of Southampton students, we have shown how discrete models can be interpreted and how distance and other attributes influence students choice of transport mode to university for lectures. Results through the binary logit model show that the further a student travels the less likely they will choose to walk. We do not conclude solid findings from any of the models due to the low significance of parameters and suggest future research described in the discussion should be considered to mitigate the flaws presented in this study.

9 References

Ben-Akiva, M., and Lerman, S. R., 1985 *Discrete Choice Analysis Theory and Application to Travel Demand*, Cambridge, The Massachusetts Institute of Technology.

Börsch-Supan, A., 1987. Economic Analysis of Discrete Choice *Lecture Notes in Economics and Mathematical Systems,* (296), Germany, Springer-Verlag Berlin - Heidelburg.

Cramer, J. S., 1991. *The LOGIT model: an introduction for economists.* Kent, Great Britain: Edward Arnold.

Erdem, T. and Swait, J., 2004. Brand Credibility, Brand Consideration, and Choice. *Journal of Consumer Research.* (31)1 pp 191-198, The University of Chicago Press. [Online] Available at: <http://www.jstor.org/stable/10.1086/383434> [Accessed 21 11 2013].

Daganzo, C. F., 1979. *Multinomial Probit The Theory and it's Application to Demand Forecasting.* London, Academic Press Inc. Ltd.

Google Maps, 2013. [Digital map] Available at: <https://maps.google.co.uk/> [Accessed 15 11 13]

Lerman, S. R. and Manski, C. F., 1979. Sample Design for Discrete Choice Analysis of Travel Behaviour: The State of the Art. *Transportation Research Part A: General,* (13)1, pp. 29 - 44.

Marschak, J., 1960. Binary Choice Constraints and Random Utility Indicators. *Mathematical Methods in the Social Sciences.* Stanford: Stanford University Press, pp. 312 - 329. [Online] Available at: <http://cowles.econ.yale.edu/P/cd/d00b/d0074.pdf> [Accessed 29 10 13].

Mcfadden, D., 1974. The Measurement of Urban Travel Demand. *Journal of Public Economics,*(3), pp. 303 - 328. [online] Available at: <http://emlab.berkeley.edu/reprints/mcfadden/measurement.pdf> [Accessed 29 10 13].

McFadden, D., 2011. *DANIEL L. McFADDEN.* [online] Available at: <http://elsa.berkeley.edu/~mcfadden/dlmcv10.html> [Accessed 19 11 2013].

Mylene, L., Paintain, L. S., Antwi, G., Jones, C., Greenwood, B., Chandramohan, D., Tagbor, H., Webster, J., 2011. *Evaluating Health Workers' Potential Resistance to New Interventions: A Role for Discrete Choice Experiments.* PLoS ONE 6(8): e23588. doi:10.1371/journal.pone.0023588. [e-journal] Available at: <http://www.plosone.org/article/info\%3Adoi\%2F10.1371\%2Fjournal.pone.0023588\#s2> [Accessed 13 11 2013]

Nkegbe, P. K., Abful-Mumin, N. Y., 2012. Choice of Transport Mode by Non-Resident University Students in Ghana. *International Journal of Business and Social Science,* (3)20 pp. 136 - 142. [Online] Available at: <http://ijbssnet.com/journals/Vol_3_No_20_Special_Issue_October_2012/15.pdf> [Accessed 21 11 2013].

Schlotzhauer, D. C., [No date]. *Some Issues in Using PROC LOGISTIC for Binary Logistic Regression.* [Online] Available at: <http://www.ats.ucla.edu/stat/sas/library/ts274.pdf> [Accessed 21 11 2013].

Schuessler, N. and Axhausen, K. W., 2007. *Recent developments regarding similarities in transport modelling.* paper presented at the 7th Swiss Transport Research Conference, Ascona, September 2007. [Online] Available at: <http://e-collection.library.ethz.ch/eserv/eth:30231/eth-30231-01.pdf> [Accessed 20 11 2013].

Thurstone, L. L., 1927. A law of comparative judgment. *Psychological Review,* 34(4), pp. 273 - 286.

Train, K. E., 1986. *Qualitative Choice Analysis.* Cambridge, Massachusetts: The Massachusetts Institute of Technology.

Train, K. E., 2003. *Discrete Choice Methods with Simulation.* 2^{nd} ed. Cambridge, United Kingdom: Cambridge University Press.

Webb, M.C., Wilson, J. R., and Chong. J., 2004. An Analysis of Quasi-complete Binary Data with Logistic Models: Applications to Alcohol Abuse Data. *Journal of Data Science,* (2), pp. 273 - 285.

10 Appendix

Table 6: Survey Data from 20 University Students

Observation	Mode	Distance (mile)	Time (min)	Cost (£)	Gender	Bus Pass	Car Owner
1	Walk	0.5	9	0	Female	No	No
2	Bus	0.9	14	0	Male	Yes	No
3	Car	3.9	25	2	Female	No	Yes
4	Bike	1.2	9	0	Male	No	No
5	Bike	1.0	7	0	Female	No	No
6	Walk	1.1	20	0	Female	No	No
7	Walk	0.5	11	0	Male	No	No
8	Walk	1.2	25	0	Female	No	No
9	Walk	1.5	30	0	Male	No	No
10	Walk	0.9	19	0	Male	No	No
11	Walk	0.1	3	0	Male	No	No
12	Walk	0.9	19	0	Male	No	Yes
13	Walk	0.8	17	0	Female	No	No
14	Walk	0.9	17	0	Female	Yes	No
15	Walk	0.6	11	0	Male	No	No
16	Walk	1.6	33	0	Male	Yes	No
17	Walk	1.1	24	0	Female	Yes	No
18	Walk	0.5	13	0	Male	No	No
19	Walk	0.4	8	0	Male	No	No
20	Bus	1.4	28	2	Female	Yes	No