ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF ARTS AND SOCIAL SCIENCES

WTP CALCULATION ACCURACY USING STATED PREFERENCE SURVEY DATA: COMPARISON BETWEEN LOGISTIC REGRESSION AND NEURAL NETWORK METHODS

M.Sc. THESIS

Gülce OPUZ

Department of Economics

Economics M.Sc. Programme

JULY 2020

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF ARTS AND SOCIAL SCIENCES

WTP CALCULATION ACCURACY USING STATED PREFERENCE SURVEY DATA: COMPARISON BETWEEN LOGISTIC REGRESSION AND NEURAL NETWORK METHODS

M.Sc. THESIS

Gülce OPUZ
(412171005)

Department of Economics

Economics M.Sc. Programme

Thesis Advisor: Asst. Prof. Dr. Christopher HANNUM

JULY 2020

<u>İSTANBUL TEKNİK ÜNİVERSİTESİ ★ SOSYAL BİLİMLER ENSTİTÜSÜ</u>

BİLDİRİMLİ TERCİH ARAŞTIRMASI VERİLERİNDE WTP HESAPLAMA DOĞRULUĞU: LOJİSTİK REGRESYON VE YAPAY SİNİR AĞ YÖNTEMLERİ KARŞILAŞTIRILMASI

YÜKSEK LİSANS TEZİ

Gülce OPUZ
(412171005)

Ekonomi Anabilim Dalı

Ekonomi Yüksek Lisans Programı

Tez Danışmanı: Dr. Öğr. Üyesi Christopher HANNUM

TEMMUZ 2020

Gülce-Opuz, a M.Sc. student of ITU Graduate School of Arts and Social Sciences student ID 412171005, successfully defended the thesis/dissertation entitled "**WTP CALCULATION ACCURACY USING STATED PREFERENCE SURVEY DATA: COMPARISON BETWEEN LOGISTIC REGRESSION AND NEURAL NETWORK METHODS**", which he/she prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor : **Assist. Prof. Dr. Christopher HANNUM** ....
Istanbul Technical University

Jury Members : **Assist. Prof. Dr. Shourjo CHAKRAVORTY** ....
Istanbul Technical University

Assist. Prof. Dr. Christopher HANNUM ....
Istanbul Technical University

Assist. Prof. Dr. Metin PİŞKİN ....
Istanbul Zaim University

Date of Submission : 14 June 2020
Date of Defense : 14 July 2020

To my family,

FOREWORD

First, I would like to thank my advisor Dr. Christopher Hannum for his kindness and contributions throughout the thesis.

Also, I am thankful to Dr. Quentin Stoeffler for all of his support in my academic career. I had the opportunity to progress in my career thanks to our studies with him. I will always be grateful to Dr. Stoeffler.

I would also like to thank my dear friend Sena Tömekçe, for her support and good times we had. Prof. Dr. Neslihan Özkan's support in my studies was very valuable, I am also thankful to her for her time and guidance.

Finally, I am grateful to all ITU Economics faculty members for their help during my studies and assistantship at ITU.

JUNE 2020 Gülce OPUZ

TABLE OF CONTENTS

Page

ABBREVIATIONS

DCE	**:** Discrete Choice Experiment
DC	**:** Discrete Choice
ANN	**:** Artificial Neural Network
SP	**:** Stated Preference
WTP	**:** Willingness to Pay
WTA	**:** Willingness to Accept
RUM	**:** Random Utility Model
IIA	**:** Irrelevance of Independent Alternatives Assumption
CV	**:** Compensating Variation
EV	**:** Equivalent Variation
AME	**:** Average Marginal Effects
GW	**:** Generalized Weights

SYMBOLS

$\boldsymbol{U_i}$	**:** Expected Utility
$\boldsymbol{y}$	**:** Expected output of ANN model
$\boldsymbol{d}$	**:** Actual output of ANN model
$\boldsymbol{\eta}$	**:** Learning rate of ANN model
$\boldsymbol{\beta}$	**:** Coefficient of Logistic Regression model
$\boldsymbol{P}$	**:** Expected output of Logistic Regression Model

LIST OF TABLES

Page

LIST OF FIGURES

Page

WTP CALCULATION ACCURACY USING STATED PREFERENCE SURVEY DATA: COMPARISON BETWEEN LOGISTIC REGRESSION AND NEURAL NETWORK METHODS

SUMMARY

Valuing environmental recreational sites, which is also referred as non-market valuations, is widely used in environmental economics studies. Survey datasets are used to assess the price value of the site to an individual, which is referred as maximum willingness to pay (WTP), in order to understand demand for the sites. WTP calculations are used to decide on recreational site alterations, therefore an accurate WTP calculation is very important on the decision. The most common type for recreational site surveys is stated preference, where respondents are given two or more alternatives to choose one option out of them. Logistic regression is used in binary choice datasets because linear models don't work with binary outputs. Thus, traditional logit models are used most commonly with random utility models for stated preference discrete choice datasets. Neural networks on the other hand are improved logistic regression models, they are combination of logistic regressions with a learning process. Therefore, in order to calculate more accurate willingness to pay estimates, neural network models which are improved logistic regressions, could be used instead of simple traditional logit models. With this research, we wanted to test the superiority of neural networks on traditional logit in terms of willingness to pay estimation accuracy. In order to compare WTP calculation accuracies, we calculated out-of-sample prediction accuracies of the two methods and compared marginal effects and coefficients of covariates. Marginal effects of covariates would acquire us the effect of covariates on the probability of choosing one alternative, therefore they indicate the variable importance on WTP calculations. Coefficients on the other hand are used in WTP calculations because they represent the strength of particular covariate on the site choice.

We examined two different articles that use stated preference survey datasets which are conducted for assessing the demand for recreational sites. First one is the survey data from Neher et al. (2019) that consists of boaters' choice of river alternatives. Boaters are given two hypothetical river alternatives with different levels of attributes including choosing neither option. For choosing neither alternative option, we assigned attribute levels as zero for all attributes. Since the data consists of only alternative attributes, conditional logistic regression is implemented. Both conditional logit and neural network methods are fitted on the randomly selected 80% of the normalized dataset, which stands for "training" data, and prediction accuracies were tested on the remaining data as test sample for 30 times. As a result, on average neural network model couldn't outperform traditional logit in terms of out-of-sample prediction accuracy. Traditional logit gave approximately 1% higher accuracy measure. Reason behind this might because the dataset perfectly satisfies all conditions of logit assumptions and maybe because of having neither option with zero attribute values in the dataset.

Second dataset is from Morey et al. (2002). Bikers are given two hypothetical biking sites and they are expected to choose one site alternative. From this data, Morey et al. (2002) try to assess WTP by calculating compensating variation, which is found by the amount of money that needs to be added or subtracted from the budget so that maximum utilities in the first and latter state would be same after an alteration in the biking site. Since there are only two choices with individual and alternative attributes together, we used simple logistic regression. Same training and test process with the first data is implemented 10 times. Results showed that neural network model's out-of-sample prediction accuracy was approximately 2.3% higher than traditional logit. Because neural network model gave higher accuracy measure on this dataset, we further continued on calculating and comparing coefficients and marginal effects of both methods. Results showed that out of 11 covariates, 8 variables' marginal effects are statistically significantly different at 1% level. In terms of coefficients, 5 variables are different at the same confidence level. This indicates that for high number of covariates, coefficients and average marginal effects of two methods significantly differ from each other. In conclusion, we found evidence that neural network gives more accurate WTP calculation and since for high number of covariates, coefficients and average marginal effects of two methods significantly differ, recreational site change decisions would be different with using neural network and logistic regression methods.

BİLDİRİMLİ TERCİH ARAŞTIRMASI VERİLERİNDE WTP HESAPLAMA DOĞRULUĞU: LOJİSTİK REGRESYON VE YAPAY SİNİR AĞ YÖNTEMLERİ KARŞILAŞTIRILMASI

ÖZET

Piyasa dışı değerlemeler olarak da adlandırılan çevresel doğal alanların (park, sahil, dağ vb.) değerlendirilmesi, çevre ekonomisi çalışmalarında yaygın olarak kullanılmaktadır. Anket çalışmaları, bu tür sahalara olan talebi anlamak için vatandaşların doğal alanlara olan maksimum ödeme isteğini (WTP) değerlendirmek amacıyla kullanılmaktadır. WTP hesaplamaları doğal alan değişikliklerine karar vermek için kullanılır, bu nedenle doğal alan değişiklik kararlarında doğru bir WTP hesaplaması çok önemlidir. Doğal çevresel alan değerlendirme anketleri içinde en yaygın olan tür, katılımcılara iki veya daha fazla farklı özelliklere sahip doğa alanları alternatifleri verilip, katılımcıların bu alternatifler içinden bir seçim yapması istenen anketlerdir. Bu anket türleri bildirimli tercih olarak adlandırılmaktadır. İkili seçim veri kümelerinde lojistik regresyon kullanılır, çünkü doğrusal modeller ikili çıktılarla çalışmamaktadır. Bu nedenle, bildirimli tercih anketlerinin değerlendirilmesinde en yaygın olarak kullanılan modelleme lojistik regresyondur. Yapay sinir ağları ise lojistik regresyon modellerinin geliştirilmiş halidir. Yapay sinir ağları birden fazla lojistik regresyonun öğrenme süreciyle bir araya gelmesiyle oluşmaktadır. Bu nedenle, daha doğru bir WTP hesaplaması için basit bir lojistik regresyon yerine lojistik regresyonun geliştirilmiş hali olan yapay sinir ağları kullanılabilmektedir. Bu araştırma ile, yapay sinir ağlarının geleneksel lojistik regresyon üzerindeki üstünlüğünü iki farklı bildirimli tercih araştırma datası üzerinde WTP hesap doğruluğu açısından test ettik. WTP hesaplama doğruluklarını karşılaştırmak için, iki yöntemin de model dışı tahmin doğruluklarını hesapladık ve değişkenlerin iki yöntemdeki marjinal etkilerini ve katsayılarını karşılaştırdık. Değişkenlerin marjinal etkileri, bu değişkenlerin bir alternatif seçme olasılığı üzerindeki etkisini vermektedir, bu nedenle WTP hesaplamasına katılacak olan değişkenler marjinal etkilere bakılarak kararlaştırılmaktadır. Öte yandan, değişken katsayıları ise WTP hesaplamalarında direkt olarak yer almaktadır, çünkü bu katsayılar o değişkenin alternatif seçimi üzerindeki gücünü temsil etmektedirler.

Kullandığımız iki bildirimli tercih araştırması anket verisinden ilki Neher et. al (2019)'a aittir. Bu data, rafting sporcularının nehir alternatifleri seçimlerinden oluşmaktadır. Sporculara farklı seviyelerde niteliklere sahip iki varsayımsal nehir alternatifi verilmiştir ve sporculardan bu iki alternatiften birini veya hiçbirini seçme seçenekleri sunulmuştur. Çalışmamızda hiçbir seçeneği seçmeme alternatifi için tüm nehir özellikleri sıfır olarak atanmıştır. Bu veri setinde yalnızca nehir özellikleri yer aldığından, koşullu lojistik regresyon uyguladık. Hem koşullu lojistik regresyon hem de sinir ağı yöntemleri, normalleştirilmiş veri kümesinin içinden rastgele seçilen %80 veri seti üzerine uygulanır, bu da "öğrenme" verileri anlamına gelir ve tahmin doğruluğu uygulanan modelin geriye kalan %20'lik veri üzerinde test edilmesiyle hesaplanır. Çalışmamızda bu hesaplamayı 30 rastgele seçilmiş veri üzerinde

gerçekleştirdik. Sonuç olarak, yapay sinir ağı modeli, örnek dışı tahmin doğruluğu açısından geleneksel lojistik regresyondan daha iyi performans gösteremedi. Geleneksel lojistik regresyon yaklaşık %1 daha yüksek tahmin doğruluğu verdi. Bunun nedeni, veri kümesinin lojistik regresyon varsayımlarının tüm koşullarını mükemmel bir şekilde karşılaması olarak açıklanabilir. Ayrıca, veri kümesinde sıfır özellik değerlerine sahip olan iki seçeneğin de seçilmemesi alternatifinin yer alması nedeniyle de olabilir.

Araştırmada kullandığımız diğer data seti Morey et al. (2002) makalesine aittir. Bisiklet sporcularına iki varsayımsal bisiklet alanı alternatifi verilip katılımcılardan alternatiflerden birini seçmeleri beklenmektedir. Morey et al. (2002), WTP'yi, bisiklet sahasındaki bir değişiklikten sonra birinci ve ikinci durumdaki maksimum faydaların aynı olacağı şekilde bütçeye eklenmesi veya çıkarılması gereken para miktarı ile bulunan telafi edici varyasyonu hesaplayarak bulmaktadır. Bu veri setinde iki seçenek bulunduğundan ve hem katılımcıların bireysel özellikleri hem de alternatiflerin nitelikleri birlikte yer aldığından, basit lojistik regresyon kullanılmıştır. Neher et. al (2019) data seti üzerinde gerçekleştirilen tahmin doğruluğu hesaplama yönteminin aynısı bu data üzerinde de 10 kez uygulanmıştır. Sonuçlar, yapay sinir ağı modelinde örnek dışı tahmin doğruluğunun geleneksel lojistik regresyondan yaklaşık %2.3 daha yüksek olduğunu göstermiştir. Yapay sinir ağı modeli bu veri seti üzerinde daha yüksek doğruluk ölçümü verdiğinden, her iki yöntemin katsayılarını ve marjinal etkilerini hesapladık ve karşılaştırdık. Sonuçlar 11 farklı değişken içinden 8'inin marjinal etkilerinin %1 düzeyinde istatistiksel olarak önemli ölçüde farklı olduğunu göstermiştir. Katsayılar açısından ise aynı güven düzeyinde 5 adet değişkenin farklı sonuç verdiğini gözlemledik. Bu, çok sayıda değişken için, katsayıların ve iki yöntemin ortalama marjinal etkilerinin birbirinden önemli ölçüde farklı olduğunu göstermektedir. Sonuç olarak bu çalışmada, yapay sinir ağlarının daha doğru WTP hesaplaması verdiğine dair kanıt bulduk ve iki yöntem çok sayıda eş değişken için farklı katsayı ve ortalama marjinal etki verdiğinden, doğal alanlarda yapılacak olan değişiklik kararlarının yapay sinir ağı ve lojistik regresyon yöntemleri ile farklı olacağını tespit ettik.

1. INTRODUCTION

As environment started to gain more importance, environmental and ecological studies had pace in recent years, economists study more on these issues. An important branch of environmental economics consists of non-market valuations, where economists try to assess the willingness to pay and contingent valuation for natural sites such as rivers, parks etc. In order to conduct these evaluations, surveys are conducted on high number of people, where respondents are been asked to make choices on different attribute alternatives. These surveys can be done in several formats. In this research, we study one of the most common types, stated preference (SP) choice surveys, where respondents are given same site with two or more different attributes and expected to make a choice between those alternatives, sometimes with choosing neither option.

SP datasets may include both individual characteristics (age, gender, marital status etc.) and the matter of interest's characteristics, regarding the scope of the research. On the decision behaviour of respondents, their individual attributes should also be taken into account before suggesting a change in the matter of interest. With this method, researchers try to estimate respondents' utility functions by their preferences in a set of choices (Kroes & Sheldon, 1988). For SP data, the most commonly used utility model is random utility. With random utility model (RUM), overall utility is expressed as a function of individual and interest's attributes, added to a random component. The assumption is that the individual would choose the alternative that gives the highest utility.

Stated preference methods in surveys are important in literature since these can be applied in various studies, not only in environmental and ecological, but also for transportation (Kroes & Sheldon, 1988) and other related non-market valuation studies. Along with contingent valuation and willingness to pay estimations, these studies can be used in demand analysis and forecasting.

For analysing discrete choice stated preference data, traditional methods are multinomial or mixed logit for more than two choices and logistic regression for presenting only two choices.

Logistic regression tries to fit a model that has a binary or dichotomous outcome, unlike linear regressions (Hosmer et al., 2013) and network models are improved logistic regressions where combination of multiple logistic regressions create a single model with a learning process. Thus, instead of using traditional simple logistic regression on stated preference datasets for calculating WTP, we propose that an improved version of traditional models, artificial neural network, would give higher accuracy on WTP calculations. Novel approach of this study is to understand whether using artificial neural network (ANN) models would be sufficient instead of traditional logistic regression methods, with a comparison in terms of their out-of-sample prediction accuracy, marginal effects of determinants and their coefficients for an accurate willingness to pay calculation. The primary goal of this research is to get a more accurate estimation of willingness to pay. We believe that since ANN models are suitable for predictions, it would give more accurate WTP estimations. We test this hypothesis in this research by comparing out-of-sample prediction accuracies of traditional logit and neural network models. Since neural networks are fit on the data with learning process, prediction accuracy on the given data would be higher than traditional logit. However, when the model would encounter new data, it will fail to give accurate results. Therefore, in this research, we used out-of-sample prediction accuracies of both neural network and logit models.

With the emerge of big data, ANN methods are widely used in econometric analysis, especially when the research focus is on prediction rather than summarization (Varian, 2014). Prediction is an important instrument for SP models because economists try to make predictions of behavioural responses on a change of a non-market value. Thus, we expect that ANN methods might predict more accurately on SP datasets.

Neural network and logit models differ in terms of different aspects. According to Daly (1982), when attributes have non-linear properties, logit models can't converge on maximum likelihood estimation with Newton's method. In addition, logit models cannot perform well when different alternatives are correlated with each other, resulting in generating a covariance matrix that has non-zero values on non-diagonal. A neural network model can overcome these problems by generating arbitrary covariance matrix, also neural network models have the ability to work on non-linear relationships (Bishop, 1996). Also, conditional logit has some drawbacks in willingness to pay estimation context (Johnson et al., 2000). According to Johnson et al. (2000), tests demonstrate that conditional logit models cannot satisfy irrelevance of independent alternatives assumption (IIA), that is assuming that the ratio of probabilities for any two alternatives is independent from the levels in another alternative. In addition, conditional logit doesn't use the correlations within each individual's series of choices.

Following section shows the studies in literature on the comparison of ANN and logit models for the analysis of SP data. Section 3 describes data that is used in this research and in section 4, analysis methods are shown in detail. Section 5 and 6 present results and conclusions of this research.

2. LITERATURE REVIEW

In this research, we used two different discrete choice stated preference survey datasets. When individuals are given hypothetical discrete choice sets that are expected to make a single choice, the model is called stated preference choice model (Adamowicz et al., 1994). Datasets that we use in this research are surveys that have been conducted in order to assess the demand of recreational sites by calculating willingness to pay (WTP) for visiting the sites. Survey questions consist of river site and biking site choices from given alternatives.

In order to analyse stated preference models, random utility is frequently used (Adamowicz et al., 1994). In our research, we also used random utility model for both datasets. In a random utility model, assumption is, individuals choose the alternative that gives the highest utility among the all alternatives that have systematic utility component (v) and random component (ε).

Willingness to pay estimation can be done in various methods. First one, which is used in the dataset from Morey et al. (2002), is by finding compensating variation (CV). Compensating variation measures the amount of money that compensates after a change in fee or another attribute in the recreational site that holds individual's utility constant (J. Zhao & Kling, 2004). Morey et al. (2002) describe CV per-ride as the amount of money that needs to be extracted from budget in order to make the utility of latter state equal with the utility of the initial state. D. McFadden (1999) also uses same notation and names the amount of money that needs to be extracted from the budget in the case of restoration of the site for equivalate utilities before and restored state as equivalent variation (EV). These concepts are used for same purpose as measuring willingness to pay or willingness to accept (WTA) in the case of compensating deuteriation in the site by the increased amount of money in the income so that maximum utility before and after the deuteriation would be same. D. McFadden (1999) proposes that mean willingness to pay calculation can be done by averaging income reductions (increase) in the case of deuteriation (restoration) for compensating variation measurements.

In Morey et al.'s (2002) research, according to logit model estimation from stated preference choice dataset, two alternatives of biking sites, W and S are simulated as being asked to individuals to make a choice between them. For individual i, that has systematic component of random utility as v, mathematical representation of CV modelling is as follows:

$$E(U_i^0) = \ ln\left(exp(v_{iW}^0) + exp(v_{iS}^0)\right) \tag{1.1}$$

$$E\left(U_i^1(c)\right) = \ ln\left(exp(v_{iW}^1) + exp(v_{iS}^1)\right) \tag{1.2}$$

$$E\left(U_i^1(c)\right) = \ E(U_i^0) \tag{1.3}$$

Where c is the amount of money that is subtracted from the budget in the latter state. Willingness to pay is calculated from this measurement. In order to calculate impact of a quality change, W. Adamowicz et al., (1994) also uses the same approach with Morey et al. (2002) with the following equation:

$$CV = \ -\frac{1}{\alpha}\left\{ln\left(\sum_i exp(v_{n0})\right) - ln\left(\sum_i exp(v_{n1})\right)\right\} \tag{2}$$

Where α represents marginal utility of income and v_{n0} stands for systematic component of random utility in the initial stage and v_{n1} represents latter stage for individual n. From the above equation, they calculate the impact of quality change of a river on the decision of boaters.

Other than compensating variation, another way of calculating willingness to pay is implemented from discrete choice experiment (DCE) surveys. In these types of datasets, individuals are expected to make a choice between alternatives. According to the number of choices, model fit can be done with simple or multinomial logit, mixed logit or random parameter logit. In Neher et al.'s research, (2019) random utility model is fitted to conditional logit. Mean willingness to pay is found by:

$$MWTP_k = \frac{\beta_{1k}}{\beta_2} \tag{3}$$

Where k represents kth element of x variables, being specific to the attribute level and β_2 is the coefficient of the cost of alternatives. This expression can also be interpreted as willingness to pay for an improvement in x_k (Hole, 2007). Since model is logistic regression, it is assumed that β coefficients are normally distributed therefore estimated WTP is ratio of two normally distributed variables (Hole, 2007).

For an accurate willingness to pay estimation, variable selection is extremely important. Researches that use logit models for estimation, marginal effects measure the change in choice probability by one unit (percent) change in an input variable (X. Zhao et al., 2020). Therefore, marginal effects in logit models are important on determining which predictors to use in willingness to pay estimations. According to Adamu et al. (2015), results of coefficients and marginal effects of logit models have different interpretations for willingness to pay estimations. The sign of the coefficient indicates the direction of the variable and willingness to pay; positive (negative) coefficient is an indicator that willingness to pay would increase (decrease) with an increase in the given variable. In addition, the size of the coefficient provides the strength of the variable in willingness to pay calculation, as shown in Equation 3.

In this research, we used artificial neural networks and logistic regression for WTP calculations and compared out-of-sample accuracies of the two methods. In literature, similar studies had conducted. One of them is from McNelis (2005), where authors compared the two methods in terms of their out-of-sample prediction accuracies for predicting credit card defaults. Their results showed that logit and neural network gave almost same accuracy measures. Similarly, Bell (1997)'s comparison on bank failure predictions gave almost identical results, neural network and logit performed almost same. Whereas, in their research Shi et al. (2012) found evidence that neural network outperformed logistic regression significantly for predicting surgery survivals.

Even though traditional logistic regression methods are more frequently used for predicting probabilities of choices in stated preference models, some researchers used artificial neural network models instead of traditional methods and compared their results (De Carvalho et al., 1998) (Jeng & Fesenmaier, 1996) (Beuthe et al., 2008). Neural network models have advantage over non-linear relationships (Bishop, 1996), but they lack in some aspects of willingness to pay estimations. According to X. Zhao et al. (2020), since demand depends on the random utility modelling, with one attribute substituting the other may give same individual utility measure and this feature cannot be captured by machine learning methods. In addition, neural network modelling being used in forecasting demand in these types of ecological studies has a drawback in terms of finding relationships between covariates and desired output. Therefore, Olden & Jackson (2002) proposes a method that uses connection weights in the model for acquiring contribution of input variables on the output. They suggest summation of each connection weights across input to output including hidden layers and calculating relative contribution in percentages of each covariate in order to assess which variables to include in demand estimations. Since neural network models have both advantages and disadvantages over traditional logit models, a detailed research on comparison of two models in discrete choice surveys for willingness to pay concepts have been studied by various researchers.

In their research, De Carvalho et al. (1998) uses one real and two synthetic discrete choice stated preference data in order to compare logit and neural network model with traditional backpropagation algorithm in terms of their predictive performance on forecasting travel demand. Their goodness of fit results showed that neural network model couldn’t perform better than logit model, its result is slightly worse than logit results. Authors claim that reason behind this would be that the data being constructed to fulfil all logit assumptions.

Yang et al. (1993) uses neural network models to predict route choice behaviour, where with the given alternatives, drivers choose between freeway and side road. Results show that neural network model correctly predicts 75% of the data. Similarly, Islam et al. (2016) uses neural network models in order to analyse customer satisfaction surveys for evaluating public transport services. They ask respondents to rank the satisfaction level for different services. They use three different neural network models; Generalized Regression Neural Network (GRNN), Probabilistic Neural Network (PNN) and Pattern Recognition Neural Network (PRNN). Their results show that GRNN and PNN outperformed PRNN, by having around 75% prediction accuracy.

Jeng & Fesenmaier (1996) also compare neural network and logit models to predict discrete choice models. They use backpropagation algorithm with sum of squared errors in the neural network model. Importance ranking of variables from logit model is found by the absolute values of the coefficients, the most important variable having the highest value. For neural network model, same importance ranking is found by part-worth utility, as an estimate for relative contribution of each variable to the total variance of the output variable in percentage share. Logit and neural network comparison shows that in terms of out-of-sample prediction accuracy, neural network model outperforms logit by having 85%, whereas logit model has only 57% accuracy. In addition, with respect to variable importance on the decision, there are significant differences between logit and neural network models.

In their research, Beuthe et al. (2008) compare conditional logit and neural network models on predicting rank preferences of transportation alternatives. For variable importance ranking, in conditional logit model researchers use marginal effects, and for neural network modelling they use below equations:

$$U(z_i, \theta) = \sum_{j=1}^{n} w_j \left[1 + exp\left(-\alpha_j - \beta_j z_{ij}\right)\right]^{-1} \tag{4.1}$$

$$W_j = \frac{\left|u_j\left(z_j^{max}\right) - u_j\left(z_j^{min}\right)\right|}{\sum_{j=1}^{n}\left|u_j\left(z_j^{max}\right) - u_j\left(z_j^{min}\right)\right|} \tag{4.2}$$

Where w_j represents the neural weight of attribute j and all neuron weights are constrained as they sum up to 1. w_j values are multiplied by the logistic regression function and summation of each n number of attributes gives the expected utility of the respondent. W_j stands for the effect of attribute j as a parallel approach to marginal effects of conditional logit.

3. DATA

In this research, we have used two different discrete choice survey data that are conducted in order to assess demand for recreational sites. One is from Neher et al. (2019) and other one is from Morey et al. (2002).

3.1 Grand Canyon Boaters Discrete Choice Survey Data

First data is from Neher et al. (2019). In their study, Neher et al. (2019) calculates willingness to pay (WTP) of Grand Canyon boaters with variations in flow level, having sandbars or not, level of fluctuation of the river and cost of the trip. Descriptive statistics are shown in Table 1.

Table 1 : Descriptive statistics for boater choice survey.

Covariate	Observations	Mean	St. Dev.	Min	Max
Decision	7,200	0.333	0.471	0	1
Flow	7,200	13,720.830	14,270.990	0	40,000
Fluctuation	7,200	0.288	0.453	0	1
Sandbars	7,200	0.378	0.485	0	1
Bid	7,200	766.159	981.854	0	2,600
Flow2	7,200	391.894	572.517	0	1,600

In the survey, individuals are given 2 different choices of boating trip characteristics, Trip A and Trip B. They have three different choices: Trip A, Trip B, or neither. There are four different boat trip characteristics:

- Flow: Average river flow level
- Sandbars: River having sand bars or not
- Fluctuation: River flow having fluctuation or not
- Bid: Your individual trip costs increased by: ____$

The given amounts for trip cost increase are $0, 90, 275, 1200$ and 2600 dollars. Average river flow level consists of $5000, 13000, 22000$ and 40000. Sand bars and Fluctuation are binary variables, being 1 when the boating area has sandy parts for sandy beaches and being 1 when there is fluctuation in river flow, zero otherwise for both variables. Flow2 stands for $(Flow/1000)^2$.

Each question that is asked to respondents is referred to as scenario. Therefore, each scenario has 2 alternatives with choosing neither option, in total making 3 options. In each scenario; flow, fluctuation and bid amounts are given different for calculating WTP of the respondents. There are 409 respondents in the data, each is given 6 different scenarios, that makes 2454 scenarios in total. Out of these, 54 of them were left blank by the respondents therefore were have taken out them from calculations. Decision indicates which alternative is chosen: 1 for alternative A, 2 for alternative B and 3 for choosing neither. %23 of respondents chose alternative A, %44 chose alternative B and %33 chose neither alternatives.

3.2 Biking Site Choice Discrete Choice Survey Data

This survey data is from Morey et al. (2002), where they study on the effect of alteration in bike site characteristics on the choices of bikers. In the survey, 1390 number of respondents are given 2 alternatives of biking sites that have different attributes, and respondents are expected to choose one alternative. Descriptive statistics are represented below:

Table 2 : Descriptive statistics for biker choice survey.

Covariate	Observations	Mean	St. Dev.	Min	Max
choice	1,390	0.547	0.498	0	1
dist	1,390	-1.672	5.711	-14	14
dist*vfc	1,390	-4,457.842	20,240.610	-43,400	40,600
vfc	1,390	-208.201	949.232	-1,800	1,800
fee	1,390	-1.792	3.592	-7	4
vfc*peaks	1,390	-1,184.173	3,027.388	-8,400	6,400
peaks	1,390	-0.334	1.544	-3	3
str	1,390	-4.561	9.123	-21	14
hiker	1,390	-0.565	0.606	-1	1
experience	1,390	1.960	0.670	1	3
mont biker	1,390	0.648	0.478	0	1
age	1,390	29.783	10.003	10	69

Site characteristics consist of:

- Total miles of the trails
- The miles of single track
- Total vertical feet of climbing
- Total number of peaks
- Usage of hikers or pedestrians in the trial
- Entry fee

Since there are two alternatives in the survey, all site characteristics represented in Table 2 are differenced values, being as (Alternative A – Alternative B). Total miles of the trial options are: 7, 14 or 21 miles. Total miles of single track are: 0, 7, 11 and 14 miles. Vertical feet of climbing are: 400, 1200 and 2200. Number of peaks options are 1, 2 and 4. Proposed entry fees are 1 and 5 dollars.

Along with trial characteristics, individual attributes were also expected to answered by the respondents. These are:

- Age
- Experience (novice, intermediate, or advanced)
- Considering oneself as a mountain biker or not (1 for mountain biker, 0 otherwise)

Regarding experience, 25% of respondents classify their biking experience level as novice, 55% as intermediate and 20% as advanced.

4. METHODS

For calculating non-market valuation and decision predictions using stated preference survey datasets, neural network models can be used instead of traditional logistic regression methods. For testing its usability, first we calculated out-of-sample prediction accuracies of both methods and compared them. If out-of-sample prediction accuracy of neural network model exceeds traditional regression models, we conclude that neural network model can be used instead of traditional methods and continue further by analysing how willingness to pay estimation varies with neural network model than traditional logistic regression methods.

4.1 Conditional Logistic Regression

Conditional logistic regression is first offered by McFadden (1974). How it differs from multinomial logit is that with conditional logit, the emphasis is on characteristics of alternatives whereas its individual's characteristics in multinomial logit (Hoffman & Duncan, 1988). Empirically, if researcher believes individual features are more determinative on the decision, then multinomial logit is more appropriate; whereas if individuals make choices depending on the characteristics of the alternatives, conditional logit is more suitable for the analysis. Probability models for multinomial logit and conditional logit are as follows:

$$P_{ij} = \frac{e^{X_{ij}\beta_j}}{\sum_{k=1}^{J} e^{X_{ij}\beta_k}} \tag{5}$$

$$P_{ij} = \frac{e^{Z_i\alpha}}{\sum_{k=1}^{J} e^{Z_i\alpha}} \tag{6}$$

Where Equation 5 represents conditional logit and Equation 6 is multinomial logit. X_{ij} are characteristics of alternative j on individual i and Z_i are attributes of individual i.

For both conditional and multinomial logit models, likelihood function is same which is shown as:

$$log\ L = \sum_i \sum_j y_{ij} P_{ij} \tag{7}$$

Where y_{ij} becomes 1 for the chosen alternative, 0 otherwise. P_{ij} values differ for multinomial and conditional logit regressions. For the surveys that don't include individual attributes, using conditional logit is more appropriate.

4.1.1 Conditional logistic regression on boater choice survey data

Conditional logit is used in the paper from Neher et al. (2019) for analysing the decisions in Discrete Choice (DC) Stated Preference survey responses. In the DC survey, individuals have three choices; Trip A, Trip B or choosing neither trips. Each trip has different levels of attributes. Individuals are offered six different scenarios that include different Trip A and Trip B options. In total, it is expected from respondents to make 6 different choices. With the conditional logit model, individuals choose the option that maximizes their utility function which is in the format of below equation:

$$U_{ij} = V\left(X_{ij}; \beta^i\right) + e_{ij} \tag{8}$$

Where U_{ij} is the utility for individual i for the choice option j and X_{ij} represents the attributes of j perceived by individual i. Survey questions don't include individual attributes, therefore utility function doesn't contain individual characteristics. e_{ij} stands for the stochastic error component.

We use Equation 5 to find predicted probabilities from the utility function shown in Equation 8. Then, we measured the accuracy of the out-of-sample prediction of the conditional logit model. In order to eliminate non-linearity in the model, squared Flow variable is added to the regression. Decision has to be 1 for the chosen alternative and 0 for the other two. Result of the conditional logit model is shown in Table 3:

Table 3 : Conditional logit on boater choice.

	Dependent variable:
	Scenario
Flow	0.038***
	(0.007)
Flow2	-0.001***
	(0.0002)
Fluctuation	-0.002
	(0.059)
Sandbars	-0.107*
	(0.056)
Bid	-0.0003***
	(0.00003)
Observations	2,400
Note:	**p<0.1; **p<0.05; ***p<0.01*

Flow variable is found by dividing actual flow level by 1000, and Flow2 is its squared value. In order to make a robust comparison with neural network model, we used same dataset for each analysis. Thus, we normalized the data so that each variable would have values between 0 and 1, 0 being the minimum and 1 being maximum for each variable respectively. Normalization is done with the following equation:

$$x_{i,N} = \frac{x_i - \min(x)\forall(x)}{\max(x)\,\forall(x) - \min(x)\forall(x)} \tag{9}$$

Where $x_{i,N}$ stands for the normalized input x for ith observation. For finding accuracy measure of this model, we divided the data as train and test being randomly selected 80% and 20% of whole sample respectively. After dividing the data, we fit trained conditional logit model on test sample to find out-of-sample prediction accuracy. Estimated probabilities of conditional logit model are found by extracting log-odds as probabilities scaled on logit model. The probability of choosing alternative m is represented as:

$$P_i(m) = \frac{\exp[U(X_{im}|\beta)]}{\sum_{ij}\exp[U(X_{im}|\beta)]} \tag{10}$$

Since these probabilities belong to groups of three choices, we expect that an individual would choose the alternative that gives the highest predicted probability out of three choices.

4.1.2 Logistic regression on biker choice survey data

Another data that was used in this analysis is from Morey et al. (2002). The paper evaluates survey questions asked to bikers for assessing the effect of alterations in biking sites on their choices. In order to understand the effect of introducing entrance fee, number of trails and opening biking site to hikers and pedestrians on bikers' decision, consumer surplus is calculated for alteration in site characteristics and compensating variations from the alterations are found for evaluating willingness to pay.

For this data, random utility model (RUM) is used for the analysis, that is represented as:

$$U_{ij} = V_{ij} + \varepsilon_{ij} \tag{11.1}$$

Where V_{ij} stands for function of trail and individual attributes and ε_{ij} is the error term.

$$V_{ij} = V(Z_j, S_i) \tag{11.2}$$

S_i are attributes of individual i and Z_j represents the characteristics of site j. Individual specific attributes are age, experience and bikers considering themselves as mountain biker or not. Alternative specific attributes are distance, vertical feet climbing (vfc), miles of single track (str), entrance fee, trial having hiker and pedestrian permission, number of peaks, distance of trial, distance*vfc and vfc*number of peaks. Since there are only two choices, simple logistic regression is used for probability prediction. Result of whole sample is presented in Table 4:

Table 4 : Logit on biker choice.

	Dependent variable:
	Choice
dist	1.147*
	(0.670)
dist*vfc	-2.724**
	(1.246)
vfc	0.922
	(0.865)
fee	-0.491
	(0.371)
vfc*peaks	2.507***
	(0.815)
peaks	-0.403
	(0.467)
str	0.980**
	(0.497)
hiker	-1.249***
	(0.234)
experience	0.064
	(0.172)
mont_biker	-0.119
	(0.127)
age	0.403
	(0.343)
Constant	-0.544*
	(0.309)
Observations	1,390
Note:	*$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$*

For alternative specific attributes, their difference is taken and coded as a single variable; i.e. dist variable represents the difference of distances in both alternatives (Alternative A – Alternative B).

After normalizing the data with Equation 9, out-of-sample prediction accuracy is calculated by dividing the data into randomly selected 80% and 20%, called as train and test samples, respectively. Above logistic regression model is fit on train sample, then the resulted model's prediction accuracy is calculated by implementing the model on the test sample. This process is also called "k-fold cross validation" where k denotes the number of test and train samples. If k equals 10, that means the process will be implemented 10 times and average out-of-sample accuracy will be calculated from resulting accuracy measures of these 10 runs.

For this dataset, the logit model is implemented on randomly selected train samples and predicted probability accuracies are calculated by implementing same model on test samples for 10 times. If predicted probability is higher than 0.5, we conclude that predicted choice will be Alternative A, Alternative B otherwise. In whole sample, 55% of the respondents choose Alternative A and the model correctly predicts 62% of whole sample.

4.2 Artificial Neural Network (ANN)

In this research, along with traditional logistic regression models, we used single and multi-layer ANN models to analyse the prediction accuracies on discrete choice stated preference surveys implemented for biking site and boating river choices. A simple ANN model starts with input variables (x) as signals represented as nodes and hidden layers. A single layer neural network illustration is represented as:

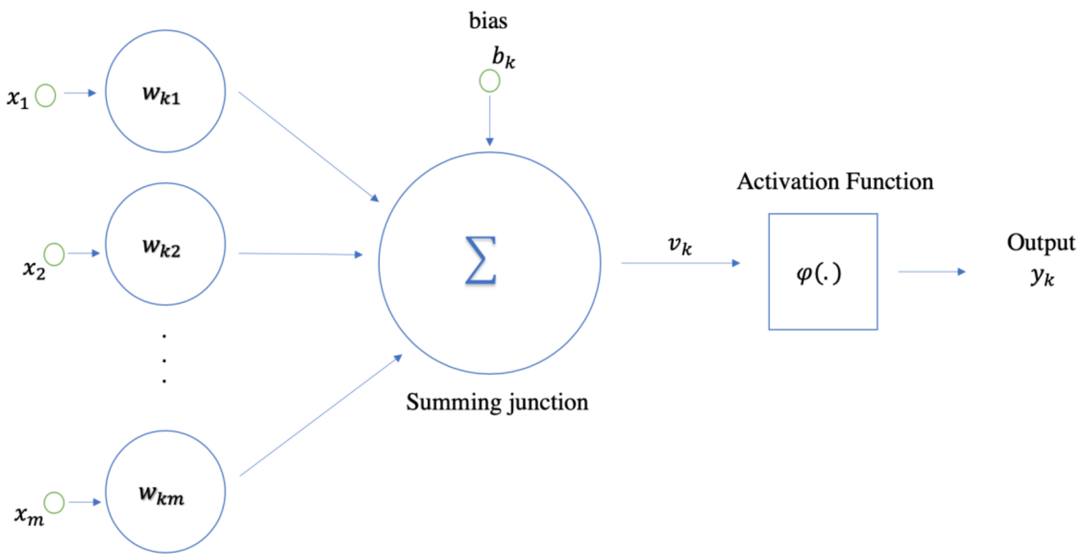

Figure 1 : An ANN model example.

Mathematical representation of Figure 1 is as follows:

$$u_k = \sum_{j=1}^{m} w_{kj} x_j \qquad (12.1)$$

$$y_k = \varphi(u_k + b_k) \tag{12.2}$$

Where $u_k + b_k = v$

$$\varphi(v) = \frac{1}{1 + \exp(-v)} \tag{12.3}$$

Where w_{kj} represents the weight of node j to node k in corresponding layers, that is initialized as random numbers, x_j is the input signal j. b_k is the bias for node k, represented as blue nodes in the figure. The summation at Equation 12.1 goes into activation function φ, which is simple logistic function in our models. This process is called forward propagation (Haykin, 1999). In terms of econometrical terminology, signals are variables, weights can be referred as coefficients β_i where $i = 1, \ldots, m$ and biases are β_0 values, without a learning process. Thus, a single node neural network would be same as running traditional logistic regression model on the variables, because input variables would be used in one logistic regression and predicted output would be calculated from this model. The biggest difference of ANN's from traditional logit is the "learning process" where weight justification is done with respect to the error of the model.

Weight justification in our biking site model is done with traditional backpropagation algorithm and error calculated by taking "sum of squared error" as:

$$e_k(n) = d_k(n) - y_k(n) \tag{13}$$

$d_k(n)$ denotes the desired, or real output and $y_k(n)$ is the output that was created by the model. Error is simply calculated by taking the difference. Aim of ANN model is to converge to $d_k(n)$. According to the error $e_k(n)$, model alters weights w_{kj} and runs again. The change in weights is done with a process called "backpropagation".

Models we use in the research are multilayer neural network models, thus there exists one or more hidden layers. For C number of output neurons total error is shown as:

$$\varepsilon(n) = \frac{1}{2}\sum_{j \in C} e_j^2(n) \tag{14.1}$$

The averaged squared error of all C neurons gives the "cost function" (loss function):

$$\varepsilon_{av}(n) = \frac{1}{N}\sum_{j \in C} \varepsilon(n) \tag{14.2}$$

For minimizing the cost function, weight correction is done by taking partial derivative of the cost function of node n to the corresponding weight that is assigned. According to the chain rule:

$$\frac{\partial \varepsilon(n)}{\partial w_{ji}(n)} = \frac{\partial \varepsilon(n)}{\partial e_j(n)} \frac{\partial e_j(n)}{\partial y_j(n)} \frac{\partial y_j(n)}{\partial v_j(n)} \frac{\partial v_j(n)}{\partial w_{ji}(n)} \tag{15.1}$$

Since:

$$\frac{\partial \varepsilon(n)}{\partial e_j(n)} = e_j(n) \tag{15.2}$$

$$\frac{\partial e_j(n)}{\partial y_j(n)} = -1 \tag{15.3}$$

$$\frac{\partial y_j(n)}{\partial v_j(n)} = \varphi_j'(v_j(n)) \tag{15.4}$$

For multi-layer network v_k becomes $v_k = \sum_{j=1}^{m} w_{kj} y_j$ therefore:

$$\frac{\partial v_j(n)}{\partial w_{ji}(n)} = y_j(n) \tag{15.5}$$

$\frac{\partial \varepsilon(n)}{\partial w_{ji}(n)}$ equation becomes $-e_j(n)\, \varphi_j'(v_j(n)) y_j(n)$, which is called as "gradient descent". Weight justification is done by multiplying learning rate η with gradient descent. Thus,

$$\Delta w_{ji}(n) = \eta \frac{\partial \varepsilon(n)}{\partial w_{ji}(n)} \tag{16}$$

Backpropagation algorithm stops when $\Delta w_{ji}(n)$ becomes a very small decimal number.

For boaters' choice survey ANN model, the algorithm is resilient backpropagation and error function is cross-entropy. Resilient backpropagation algorithm is first offered by Riedmiller & Braun (1993), as a faster backpropagation learning. The difference of resilient back propagation is that instead of using magnitude of derivate $\frac{\partial \varepsilon(n)}{\partial w_{ji}(n)}$ in Equation 16, the derivative's sign is used. If the sign doesn't change from previous iteration, the updated value (Δ_{ij}) is large, being higher than 1, whereas if the sign changes; which indicates a jump from minimum error because of large step, therefore update value is lessen to being between 0 and 1 in order not to skip the minimum error value.

After weight update value determination, weight update sign is determined by the sign of $\frac{\partial \varepsilon(n)}{\partial w_{ji}(n)}$. If sign of the derivative is negative, meaning decreasing error, update value is added to previous weight, otherwise; if error gradient is positive (increasing error) then update weight is subtracted in order to reach the minimum error value. When partial derivative changes sign, indicating minimum error is skipped, new weight becomes:

$$\Delta w_{ij}^{(t)} = -\Delta w_{ij}^{(t-1)}, \qquad if \ \frac{\partial \varepsilon(n)}{\partial w_{ji}(n)}^{(t-1)} * \frac{\partial \varepsilon(n)}{\partial w_{ji}(n)}^{(t)} < 0 \tag{17}$$

Where t denotes steps. In this case, weight is set to its previous state. In order not to repeat same process again, $\frac{\partial \varepsilon(n)}{\partial w_{ji}(n)}^{(t-1)}$ is set to zero.

For boater choice survey ANN model, instead of using sum of squared errors, we derived error function by cross entropy method that is shown as:

$$\varepsilon(n) = -\frac{1}{n}\sum_{x}[d * lny + (1-d) * ln(1-y)] \tag{18}$$

Cross entropy loss function is always positive and for the desired and actual values that are closer to each other (when $d = 0$ and y is almost 0 and $d = 1$ and y almost equals to 1), $\varepsilon(n)$ becomes smaller. Cross entropy method differs from sum of squared errors by its ability to prevent learning rate becoming slower (Nielsen, 2015). With cross entropy, error derivative becomes:

$$\frac{\partial\varepsilon(n)}{\partial w_j(n)} = -\frac{1}{n}\sum_{x}\left(\frac{d}{\varphi(v)} + \frac{1-d}{1-\varphi(v)}\right)\frac{\partial\varphi(v)}{\partial w_j(n)} \tag{19.1}$$

$$= -\frac{1}{n}\sum_{x}\left(\frac{d}{\varphi(v)} + \frac{1-d}{1-\varphi(v)}\right)\varphi'(v)x_j \tag{19.2}$$

$$= -\frac{1}{n}\sum_{x}\frac{\varphi'(v)x_j}{\varphi(v)\big(1-\varphi(v)\big)}(\varphi(v) - d) \tag{19.3}$$

From Equation 12.3, $\varphi'(v) = \varphi(v)(1-\varphi(v))$, therefore above equation becomes:

$$\frac{\partial\varepsilon(n)}{\partial w_j(n)} = -\frac{1}{n}\sum_{x}x_j(\varphi(v) - d) \tag{19.4}$$

Equation 19.4 shows that with cross entropy cost function, weight is modified by the error amount itself, that is $\varphi(v) - d$. Thus, if error amount is high, weight update will be large, or small otherwise, which acquires learning to be faster. Also, learning becoming slower after a point in sum of squared error function is caused by $\varphi_j'(v_j(n))$ term in the calculation becoming too small. Cross entropy avoids this problem by eliminating this term in the calculations.

4.2.1 ANN on boater choice survey data

We implemented a multi-layer neural network algorithm on wide format discrete choice stated preference survey data. Three different decision variables were created: l1 for choosing alternative A, l2 for choosing alternative B and l3 for choosing neither alternatives. The ANN model has 15 input signals as 15 different variables. Flow, fluctuation, sandbars and bid for each 3 alternatives, being 0 for choosing neither option. We also created $(Flow/1000)^2$, coded as Flow2. 1, 2 and 3 after the dots represent alternative A, alternative B and neither alternatives, respectively. The model is a multilayer neural network with two hidden layers each having 10 and 7 nodes. Activation function is logistic (sigmoid) and cross entropy cost function is used. Resilient backpropagation algorithm is used for learning process. The model before iterations is illustrated as:

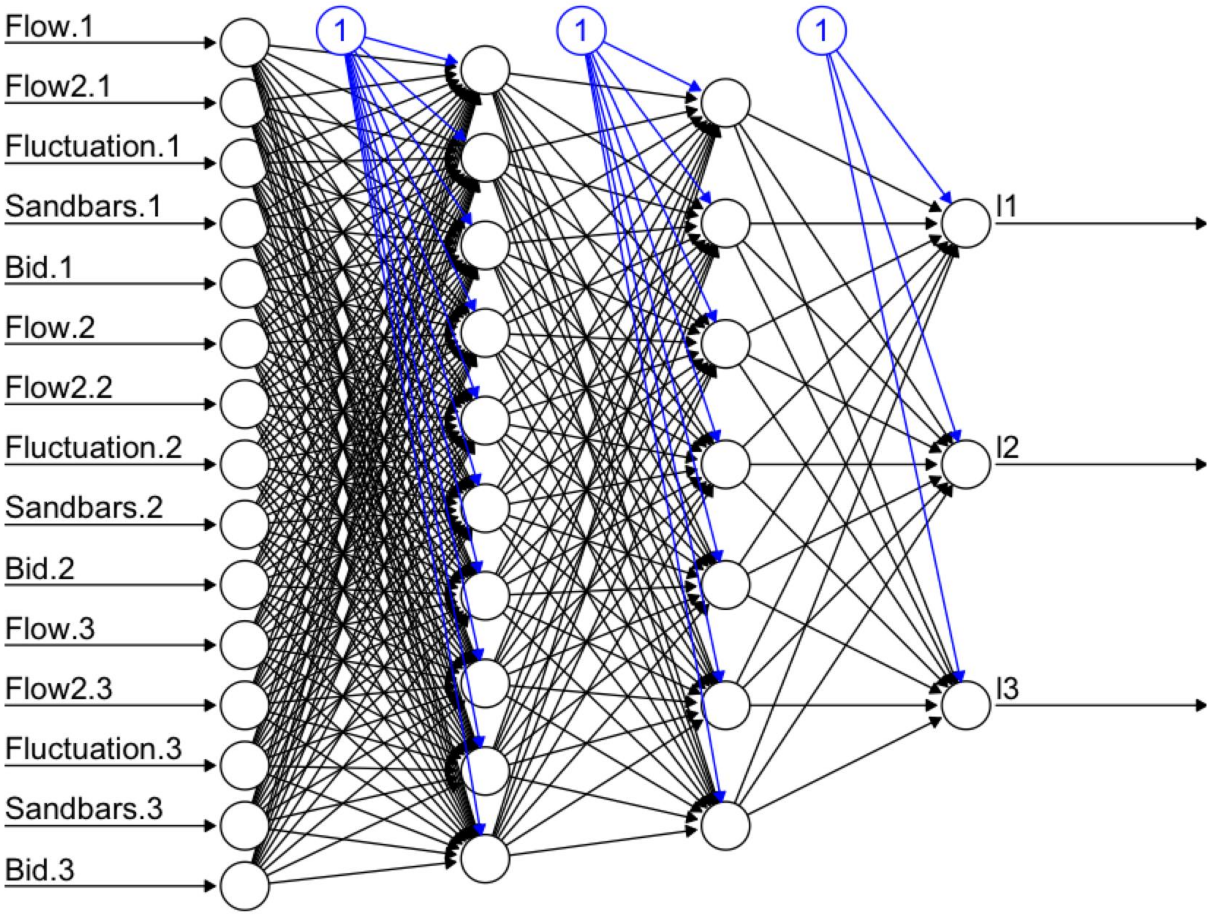

Figure 2 : ANN model for boaters' choice survey.

In Figure 2, blue nodes represent biases as b_k, k being 1 to 3 and variables stand for $x_i$values i being 1 to 15. Weights are updated according to resilient back propagation method and errors are calculated by cross entropy function.

First, data is normalized with Equation 9. After normalization, we divide the data into randomly selected test and train samples, 80% of whole sample for training sample set and 20% for test sample set. We fit the model into train sample and use that model on test sample to find the out-of-sample prediction accuracy. In order to reduce overfitting, learning stops at the iteration where the change in error, which is represented as threshold in neuralnet package of R (Günther & Fritsch, 2010), is less than 6%. Model is fitted for 30 runs and average out-of-sample prediction accuracy is calculated from these iterations by randomly selecting different samples that are trained in the model and tested for cross validation. Resulting accuracy is found by taking average of all accuracy levels for 30 iterations.

4.2.2 ANN on biker choice survey data

For biker site choice survey, we used single layer ANN model that has 4 hidden nodes with sum of squared error model, sigmoid activation function and traditional backpropagation algorithm where learning rate η being 0.001. For eliminating overfitting, threshold, which is percentage change in error at each iteration, is set to 6.5%. That means, neural network will stop learning when the percentage change in error reduction will be less than 6.5. Output variable, choice represents the probability of choosing alternative A. Model before iterations is as being showed in Figure 3:

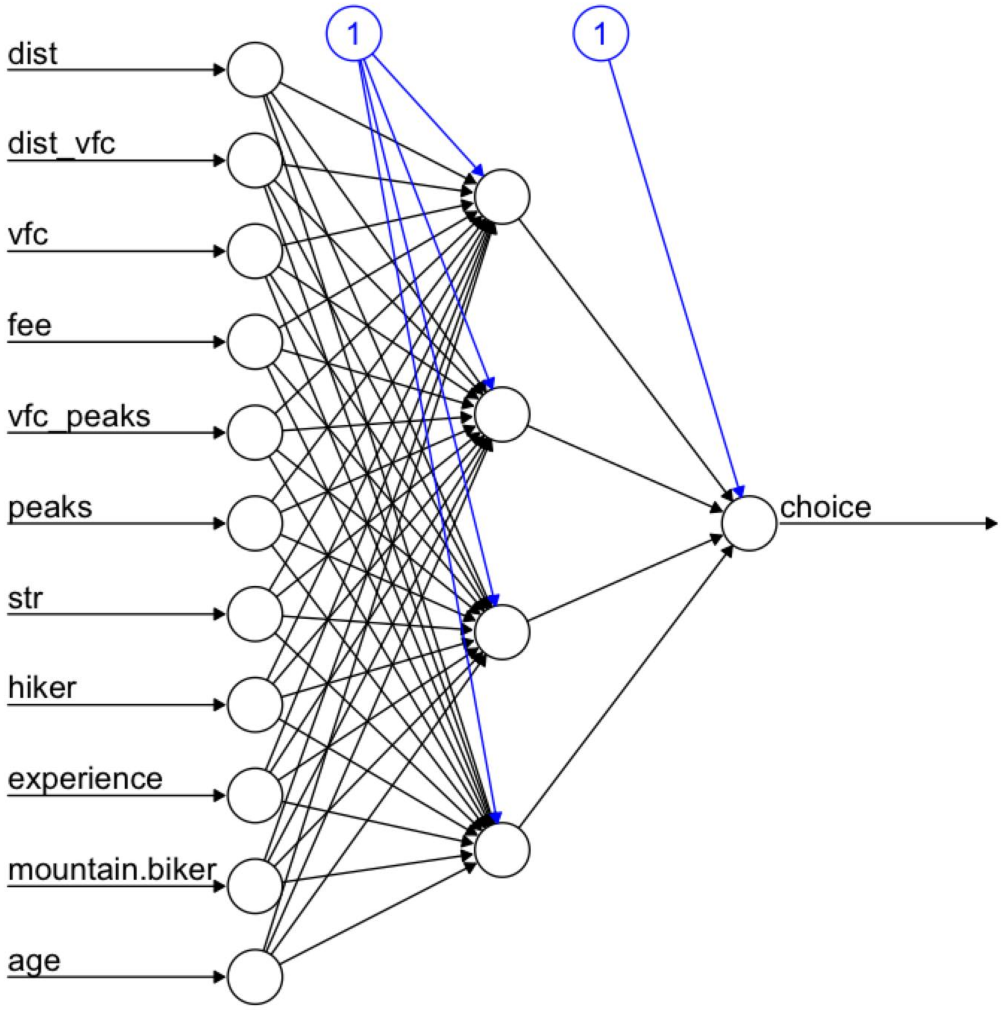

Figure 3 : ANN model for bikers' site choice survey.

Blue nodes represent bias values. Predicted choice being higher than 0.5 indicates that the prediction is Alternative A.

Normalized data by Equation 9 is divided to randomly selected 80% as training and 20% for testing in order to calculate out-of-sample prediction.

4.3 WTP Estimation Comparison: Logit and ANN Models

For an accurate willingness to pay estimation, both the choice of parameters and calculation of coefficients of those parameters are needed. Thus, in this research, we compared logit and neural network models in terms of parameter selection (variable importance) and coefficient estimations. For the datasets that have more accurate predictions with neural network model, we continued with finding and comparing marginal effects of logit and neural network for assessing variable importance, and coefficients of logit models and generalized weights of ANN models in order to compare willingness to pay estimations.

In traditional logistic regression models, marginal effects are used to assess the alteration of predicted probabilities with the changes in input covariates. As simple logistic (sigmoid) regression is shown on Equation 12.3, v is replaced by the utility function, $x\beta$. Thus, the probability of choosing alternative A becomes:

$$Pr(y=1) = \frac{1}{1+\exp(-x\beta)} \qquad (20.1)$$

In order to find marginal effects of x variables, we need to derive the change in the probability of choosing alternative A on the change of variable x_i,

$$\frac{\partial Pr(y=1)}{\partial x_i} = \frac{exp(x\beta)}{1+\exp(x\beta)^2}\frac{\partial x\beta}{\partial x_i} \qquad (20.2)$$

$$= Pr(y=1) * \; Pr(y=0) * \; \beta_i \qquad (20.3)$$

Average marginal effects are used in our analyses, as taking average of marginal effects at each observation.

For finding same average marginal effect measure of the neural network model, we first wrote functional form of the model being as:

$$y = f\left(\sum_{j=1}^{n} w_{kj}^2 * f\left(\sum_{j=1}^{m} w_{kj}^1 x_j^1 + b_k^1\right) + b_k^2\right) \qquad (21)$$

Where x_j^1 represents input nodes and b values are biases. f function is simple logistic regression:

$$f = \frac{exp(x)}{1+exp(x)} \qquad (22)$$

We first implemented neural network model, and calculated weights (w) and biases (b) values. In order to estimate the same marginal effect calculation with logit model, we calculated partial derivative $\frac{\partial y(x_i)}{\partial x_i}$, where i is the number of observations. Then, measured the average of marginal effects for all observations in order to equivalate average marginal effects of logit. In addition to marginal effects comparison, we also calculated and compared coefficient estimations of two methods, that are shown in Results section.

5. RESULTS

5.1 Boater Choice Survey Data Logit and ANN Prediction Accuracy Comparison

We regressed conditional logit on training sample and implemented the model on test sample to calculate prediction accuracy. In order to achieve more robust results, process is done on 30 different randomly selected test and train samples and prediction accuracies for both ANN and conditional logit model are calculated on the same datasets. Accuracy measures are as follows:

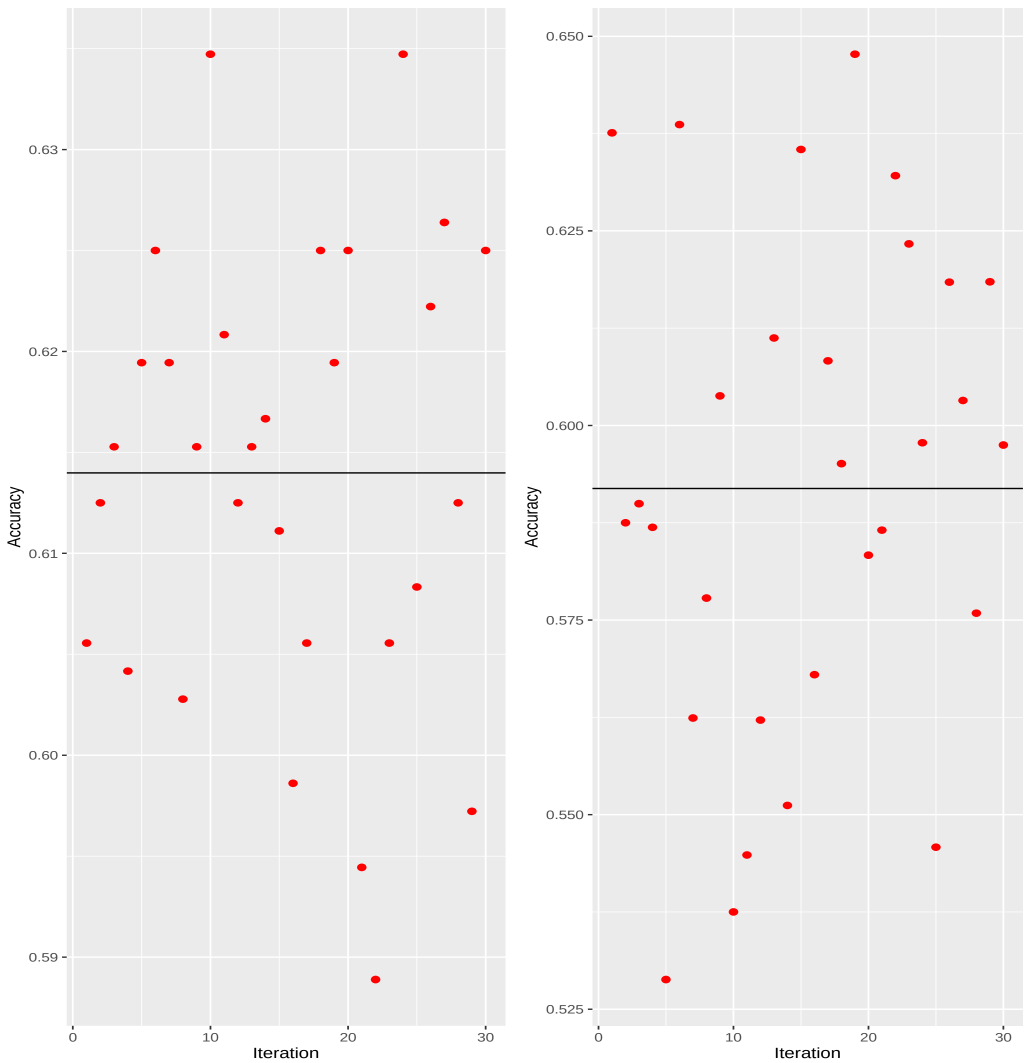

Figure 4 : Prediction accuracy comparison of boater choice survey.

First plot in Figure 4 belongs to conditional logit accuracy results and second plot is from ANN model accuracy. Black line shows average accuracy measures. It can be seen that average conditional logit out-of-sample accuracy is around 61% whereas for ANN model, average accuracy is almost 60%. Since ANN model out-of-sample prediction accuracy didn't give better results than conditional logit model, we didn't conduct further WTP estimation analysis using ANN model on boaters' choice survey data.

5.2 Biker Choice Survey Data Logit and ANN Prediction Accuracy Comparison

For biker choice survey data, out-of-sample prediction accuracies for 10 iterations are shown in Figure 5:

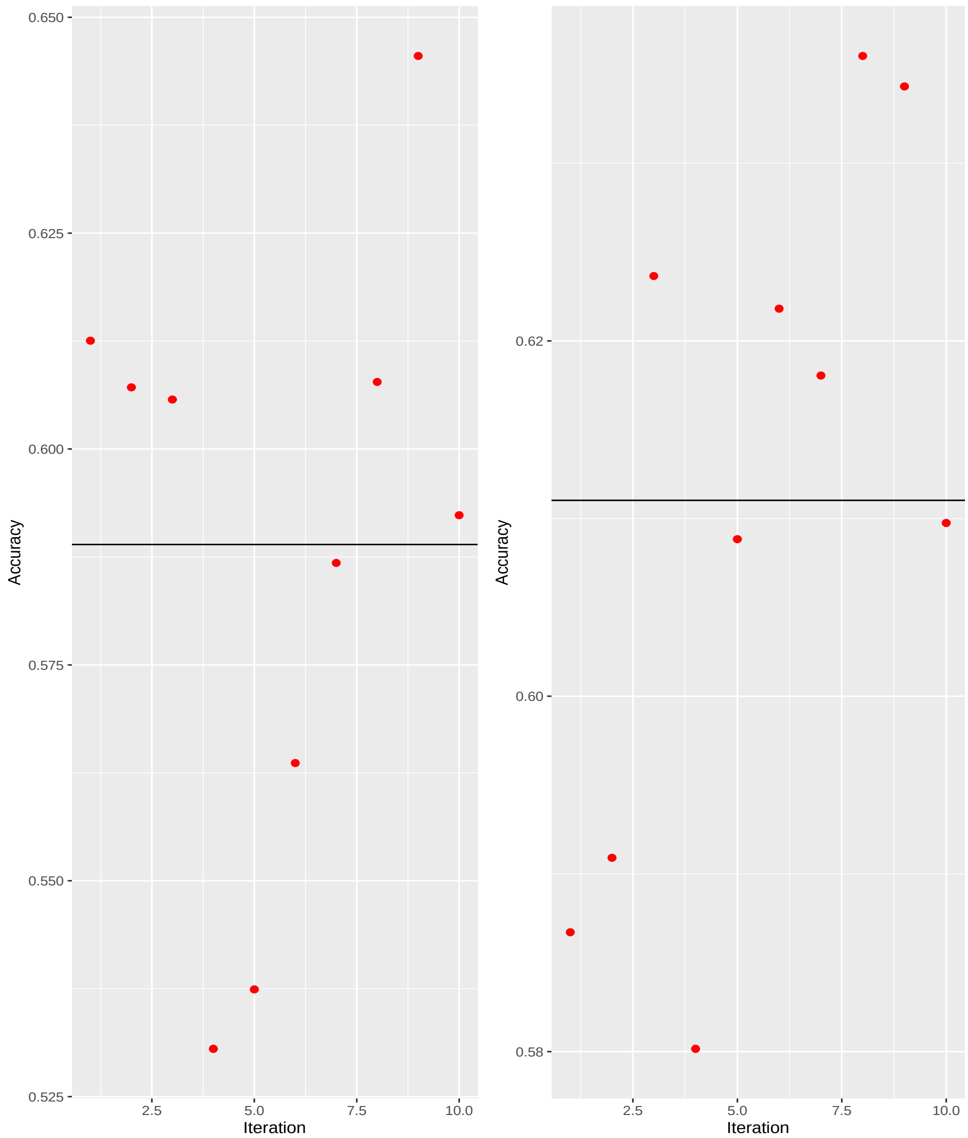

Figure 5 : Prediction accuracy comparison of biker choice survey.

First plot belongs to simple logit and second one is from ANN model. Black line shows average accuracy. Average logit accuracy is around 58.8% whereas ANN model gives 61.1% accuracy on average.

5.2.1 Biker choice survey WTP comparison between logit and ANN

ANN model of bikers' site choice survey data gave higher performance in terms of out-of-sample prediction accuracy comparing with traditional logit. Therefore, we further continued on this model by calculating and comparing willingness to pay estimations in terms of variable importance in order to decide which ones to include in WTP calculation, and their coefficients. First, we found average marginal effects of logit model and generalized weights of ANN model. Average marginal effects (AME) of the logit model of the full dataset are shown in Table 5.

Table 5 : Logit average marginal effects of biker choice survey.

Covariate	AME	SE	z	p	lower	upper
age	0.10	0.08	1.18	0.24	-0.06	0.25
dist_vfc	-0.64	0.29	-2.20	0.03	-1.21	-0.07
dist	0.27	0.16	1.72	0.09	-0.04	0.58
experience	0.01	0.04	0.37	0.71	-0.06	0.09
fee	-0.12	0.09	-1.33	0.18	-0.29	0.06
hiker	-0.29	0.05	-5.57	0	-0.40	-0.19
mont_biker	-0.03	0.03	-0.95	0.35	-0.09	0.03
peaks	-0.10	0.11	-0.86	0.39	-0.31	0.12
str	0.23	0.12	1.98	0.05	0	0.46
vfc_peaks	0.59	0.19	3.12	0	0.22	0.96
vfc	0.22	0.20	1.07	0.29	-0.18	0.62

From Table 5, p values of average marginal effects demonstrate that dist*vfc, hiker, str and vfc*peaks are statistically significant at 5% confidence level. This indicates that these covariates are significantly effective on the decision of bikers while choosing trial sites. Figure 6 shows comparison between average marginal effects of logit and neural network models:

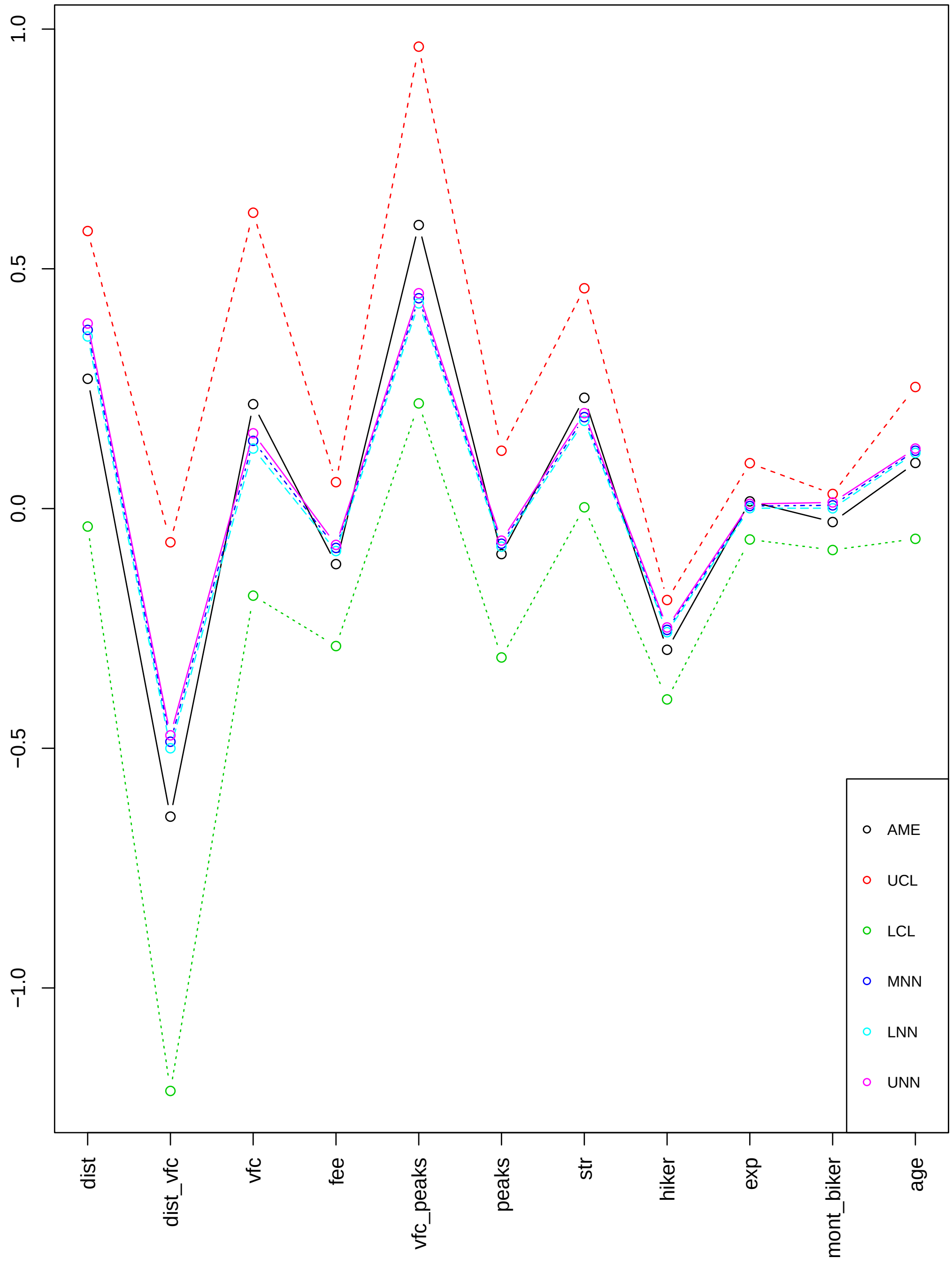

Figure 6 : Marginal effects of biking site choice dataset.

Upper and lower confidence intervals of neural network model is found by bootstrapping. Neural network is run with different randomly initialized weights for 10 times, which produces 13900 different observation of marginal effects. Bootstrapping and average marginal effects are calculated from these 13900 number of observations. AME shows the average marginal of logit estimation, UCL and LCL are upper and lower confidence levels for the corresponding AME. MNN represents average marginal effects of neural network model. LNN and UNN are lower and upper confidence levels for ANN model respectively. In addition to visualization, welch t-test showed that dist*vfc, fee, vfc*peaks, peaks, str, hiker, mont_biker and age have statistically significantly different average marginal effects for logit and neural network models at 1% confidence level.

After marginal effect comparison, we compared coefficient estimates of neural network and logit models. For calculation of neural network coefficients, we used generalized weights. Generalized weight calculations on neural network models were first offered by Intrator & Intrator (2001), as a parallel approach to coefficients of logit models for finding effect of each variable x_i on the output of ANN model, y. Mathematical representation of generalized weights can be shown as:

$$W_i = \frac{\partial \log\left[\frac{y(x)}{1-y(x)}\right]}{\partial x_i} \tag{24}$$

The generalized weight of variable x_i depends on all input variables, x. Small variance of W_i shows that x_i covariate mostly has linear effect on the output. In order to compare average marginal effects with generalized weights, we took average of W_i values of all observations. The similarity between generalized weights and coefficient of logit comes from following equation:

$$l = \log\left[\frac{p(x)}{1-p(x)}\right] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \ldots + \beta_n x_n \tag{25}$$

Above equation represent logistic regression, where l represents log-odds. From Equation 25, the derivative of log-odds with respect to covariates $\left(\frac{\partial \log\left[\frac{p(x)}{1-p(x)}\right]}{\partial x_i}\right)$ would acquire us β coefficients, which is the same equation of generalized weights calculation. Figure 7 demonstrates the comparison between coefficients and generalized weights:

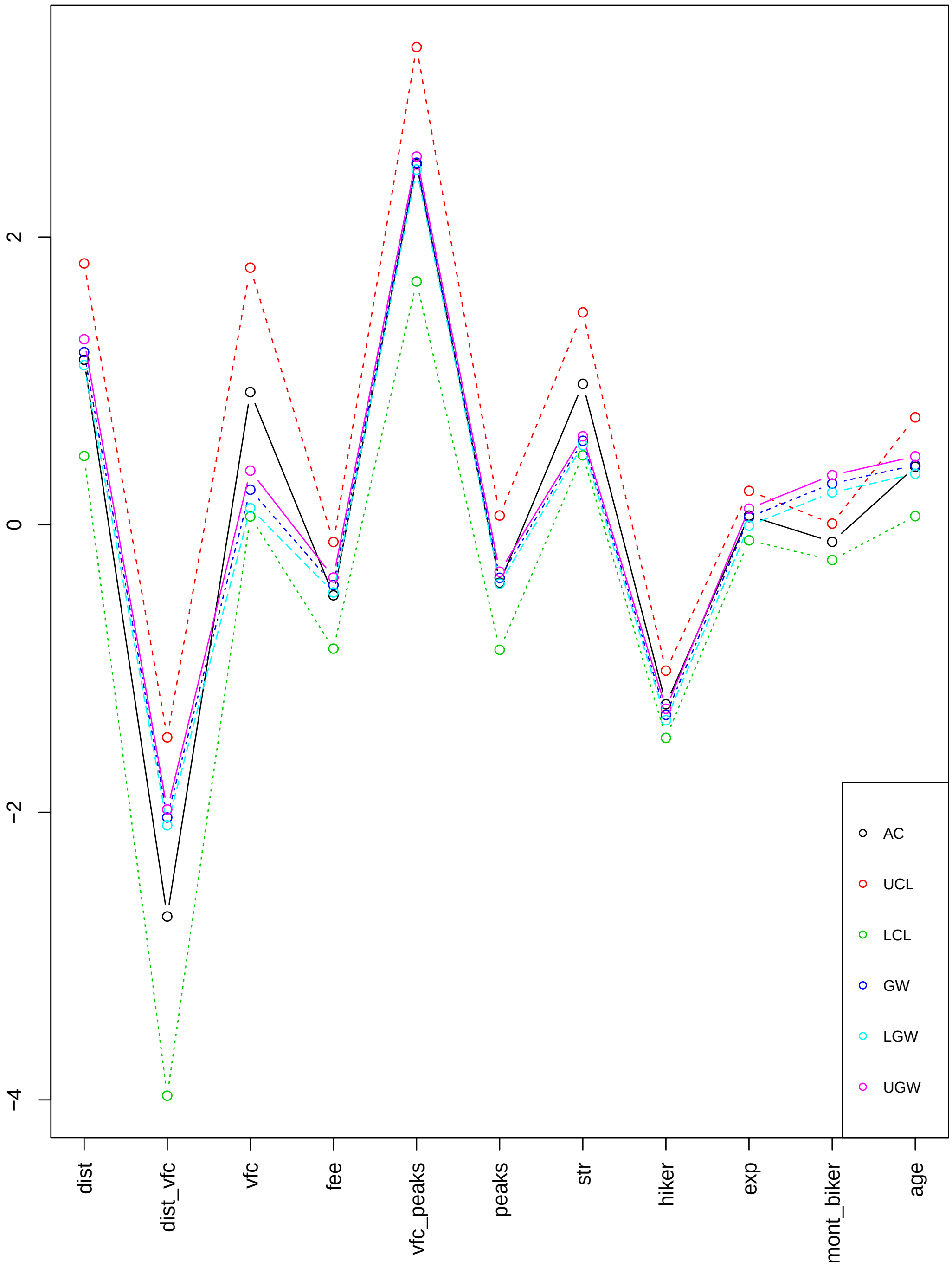

Figure 7 : Coefficients of biking site choice dataset.

UCL and LCL represent upper confidence level and lower confidence level for logit model respectively, which are also represented as “lower” and “upper” in Table 5. LGW and UGW are upper and lower confidence levels of generalized weights that are were calculated by bootstrapping the whole sample. From two sample t-test, we found that for dist*vfc, vfc, str, hiker and mont_biker variables coefficients and average generalized weights are statistically significantly different at 1% confidence level.

6. CONCLUSION

For more accurate willingness to pay estimations, neural network models might be used instead of traditional logistic regressions because neural network gives more accurate out-of-sample prediction accuracies when data has non-linearity. In addition, neural networks are designed to make more accurate predictions and they perform better with big datasets. Artificial neural networks are designed as an improved model of logistic regression. Thus, instead of using traditional logistic regression on WTP calculation, more accurate WTP estimation could be derived with an improved model, which is artificial neural network. WTP estimations are used for recreational site alteration decisions, for example decision of introducing an entrance fee. Therefore, more accurate WTP estimation means more accurate decisions on recreational sites. In this research, we compared traditional methods with neural network in terms of accuracy in willingness to pay estimations for discrete choice stated preference surveys.

Both datasets that we used are for estimating recreational site choice demand. In the first one, boaters evaluate river attributes and in the other one bikers evaluate biking site characteristics. In the boaters' choice survey, individuals are given two alternatives with a neither option which makes three alternatives in total, therefore we used conditional logit for its modelling. For neither option, we set all attributes to be zero. As a result, we couldn't find any evidence that neural network outperforming simple logistic regression in terms of predictive performance. Reason behind this might be because neither option has zero values for all attributes and neural network model couldn't learn properly from choosing neither option. In addition, another reason might be that the dataset perfectly satisfies all logit assumptions.

For bikers' choice survey dataset, respondents were given only two alternatives to choose from without a neither option. Thus, for this dataset we conducted simple logistic regression. Results demonstrated that neural network model fit better to the dataset, giving approximately 2.3% higher out-of-sample prediction accuracy. Since neural network model gave better results, we further continued on assessing the variables which are important on the decision of choosing a recreational site alternative. These variables would be used in willingness to pay estimations. We implemented two sample t-test in order to compare average marginal effect of each variable for neural network and logit models. Confidence intervals for generalized weights are calculated by bootstrapping. Results showed that dist*vfc, fee, vfc*peaks, peaks, str, hiker, mont_biker and age variables' marginal effects are statistically significantly different at 1% level. In terms of coefficients, for dist*vfc, vfc, str, hiker and mont_biker variables are different at the same confidence level. For discrete choice stated preference dataset that has more than 1000 observations with two choice alternatives, we found evidence that neural network performs better than traditional logistic regression in terms of out-of-sample prediction accuracy. In addition, we also found evidence that with neural network model, variable importance estimations (marginal effects) and their coefficients are significantly different than traditional logit model; which indicates the decisions of changes in recreational sites would be different with using neural network model.

REFERENCES

Adamowicz, W., Louviere, J., & Williams, M. (1994). Combining revealed and stated preference methods for valuing environmental amenities. *Journal of Environmental Economics and Management*. https://doi.org/10.1006/jeem.1994.1017

Adamu, A., Yacob, M. R., Radam, A., Hashim, R., & Adam, S. U. (2015). Economic Valuation of Ecotourism Resources in Yankari Game Reserve, Bauchi Nigeria. *Procedia Environmental Sciences*. https://doi.org/10.1016/j.proenv.2015.10.025

Bell, T. B. (1997). Neural nets or the logit model? A comparison of each model's ability to predict commercial bank failures. *International Journal of Intelligent Systems in Accounting, Finance & Management*. https://doi.org/10.1002/(sici)1099-1174(199709)6:3<249::aid-isaf125>3.0.co;2-h

Beuthe, M., Bouffioux, C., Krier, C., & Mouchart, M. (2008). A Comparison of Conjoint, Multi-Criteria, Conditional Logit and Neural Network Analyses for Rank-Ordered Preference Data. In *Recent Developments in Transport Modelling*. https://doi.org/10.1108/9781786359537-009

Bishop, C. M. (1996). Neural networks: a pattern recognition perspective. *Neural Networks*. https://doi.org/10.1.1.46.8742

Daly, A. (1982). Estimating choice models containing attraction variables. *Transportation Research Part B*. https://doi.org/10.1016/0191-2615(82)90037-6

De Carvalho, M. C. M., Dougherty, M. S., Fowkes, A. S., & Wardman, M. R. (1998). Forecasting travel demand: A comparison of logit and artificial neural network methods. *Journal of the Operational Research Society*. https://doi.org/10.1057/palgrave.jors.2600590

Günther, F., & Fritsch, S. (2010). Neuralnet: Training of neural networks. *R Journal*. https://doi.org/10.32614/rj-2010-006

Haykin, S. (1999). Neural Networks: A Comprehensive Foundation (3rd Edition). In *The Knowledge Engineering Review*.

Hoffman, S. D., & Duncan, G. J. (1988). Multinomial and conditional logit discrete-choice models in demography. *Demography*. https://doi.org/10.2307/2061541

Hole, A. R. (2007). A comparison of approaches to estimating confidence intervals for willingness to pay measures. *Health Economics*. https://doi.org/10.1002/hec.1197

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). Applied Logistic Regression: Third Edition. In *Applied Logistic Regression: Third Edition*. https://doi.org/10.1002/9781118548387

Intrator, O., & Intrator, N. (2001). Interpreting neural-network results: A simulation study. *Computational Statistics and Data Analysis*. https://doi.org/10.1016/S0167-9473(01)00016-0

Islam, M. R., Hadiuzzaman, M., Banik, R., Hasnat, M. M., Musabbir, S. R., & Hossain, S. (2016). Bus service quality prediction and attribute ranking: a neural network approach. *Public Transport*. https://doi.org/10.1007/s12469-016-0124-0

Jeng, J. M., & Fesenmaier, D. R. (1996). A neural network approach to discrete choice modeling. *Journal of Travel and Tourism Marketing*. https://doi.org/10.1300/J073v05n01_08

Johnson, F. R., Banzhaf, M. R., & Desvousges, W. H. (2000). Willingness to pay for improved respiratory and cardiovascular health: A multiple-format, stated-preference approach. *Health Economics*. https://doi.org/10.1002/1099-1050(200006)9:4<295::AID-HEC520>3.0.CO;2-D

Kroes, E., & Sheldon, R. (1988). STATED PREFERENCE METHODS. AN INTRODUCTION. *Journal of Transport Economics and Policy*.

McFadden, D. (1999). *Computing Willingness-to-Pay in Random Utility Models*. https://doi.org/10.4324/9780203074206.ch15

McFadden, D. L. (1974). Conditional Logit Analysis of Qualitative Choice Behavior. In *Frontiers in Econometrics*.

McNelis, P. D. (2005). Neural Networks in Finance. In *Neural Networks in Finance*. https://doi.org/10.1016/B978-0-12-485967-8.X5000-3

Morey, E. R., Buchanan, T., & Waldman, D. M. (2002). Estimating the benefits and costs to mountain bikers of changes in trail characteristics, access fees, and site closures: Choice experiments and benefits transfer. *Journal of Environmental Management*. https://doi.org/10.1006/jema.2001.0513

Neher, C., Bair, L., Duffield, J., Patterson, D., & Neher, K. (2019). Convergent validity between willingness to pay elicitation methods: an application to Grand Canyon whitewater boaters. *Journal of Environmental Planning and Management*. https://doi.org/10.1080/09640568.2018.1435411

Nielsen, M. (2015). *Neural Networks and Deep Learning* (Vol. 2018). Determination press.

Olden, J. D., & Jackson, D. A. (2002). Illuminating the “black box”: A randomization approach for understanding variable contributions in artificial neural networks. *Ecological Modelling*. https://doi.org/10.1016/S0304-3800(02)00064-9

Riedmiller, M., & Braun, H. (1993). Direct adaptive method for faster backpropagation learning: The RPROP algorithm. *1993 IEEE International Conference on Neural Networks*. https://doi.org/10.1109/icnn.1993.298623

Shi, H. Y., Lee, K. T., Lee, H. H., Ho, W. H., Sun, D. P., Wang, J. J., & Chiu, C. C. (2012). Comparison of artificial neural network and logistic regression models for predicting in-hospital mortality after primary liver cancer surgery. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0035781

Varian, H. R. (2014). Big data: New tricks for econometrics. *Journal of Economic*

Perspectives. https://doi.org/10.1257/jep.28.2.3

Yang, H., Kitamura, R., Jovanis, P. P., Vaughn, K. M., & Abdel-Aty, M. A. (1993). Exploration of route choice behavior with advanced traveler information using neural network concepts. *Transportation*. https://doi.org/10.1007/BF01307059

Zhao, J., & Kling, C. L. (2004). Willingness to pay, compensating variation, and the cost of commitment. *Economic Inquiry*. https://doi.org/10.1093/ei/cbh077

Zhao, X., Yan, X., Yu, A., & Van Hentenryck, P. (2020). Prediction and behavioral analysis of travel mode choice: A comparison of machine learning and logit models. *Travel Behaviour and Society*. https://doi.org/10.1016/j.tbs.2020.02.003

CURRICULUM VITAE

Name Surname	**: Gülce Opuz**
Place and Date of Birth	**: Istanbul – November 10, 1993**
E-Mail	**: opuz17@itu.edu.tr**

EDUCATION :

- **B.Sc.** : 2017, Koc University, Industrial Engineering
- **B.A.** : 2018, Koc University, Business Administration
- **M.Sc.** : 2020, Istanbul Technical University, Economics