World Applied Sciences Journal 22 (4): 500-505, 2013
ISSN 1818-4952

DOI: 10.5829/idosi.wasj.2013.22.04.419

The Effects of Coding on the Analysis of Consumer Choices of Public Parks

[1]*Bakti Hasan-Basri* and [2]*Mohd Zaini Abd Karim*

[1]Department of Economics and Agribusiness, Universiti Utara Malaysia
[2]Othman Yeop Abdullah Graduate School of Business, Universiti Utara Malaysia

Abstract: The purpose of this paper is to determine whether the use of different data coding give different results in the estimation of consumer choice model. The results of the analysis indicate that both dummy and effect coding produce similar results in terms of the model goodness of fit and coefficient of price. However, the estimated coefficients are different, leading to different value of WTP. The estimation model that used dummy coding seems to produce better results based on the total number of significant coefficients. Hence, WTP calculation in the case of dummy coding is more reliable. Based on the results, the use of dummy coding is preferred in the case where the estimation model does not include intercept. The finding suggests that the interpretation of estimates using different coding should be done with caution as it gives different results which can leads to different policy implications.

Key words: Effect coding % Dummy coding % Choice Experiment % Willingness to Pay % Conditional Logit model % Public Parks

INTRODUCTION

Choice Experiment (CE) technique has been applied by many analysts for various purposes. One of the purpose is to place a monetary value for non-market goods such as environmental quality [1] and [2]. However, the use of the technique has been expanded to value goods that are traded in the market. These include valuing public value on health services [3]; food quality [4]; and theme parks [5].

Popularly known as Choice Modelling (CM), CE has several advantages compared to other stated-preference (SP) approach technique, Contingent Valuation Method (CVM). CE is more appealing relative to the CVM for several reasons. First, it forces respondents to trade-off changes in attribute levels against the costs of making these changes. Second, it allows respondents to choose a *status quo* option rather than opt for changes. Third, it uses an econometric model that is consistent with the theory of probabilistic choice. Finally, it permits the benefits of environmental goods be estimated through measures of both compensating and equivalent surplus [6].

Application of CE involves respondents choosing one of several alternatives that are available on a given choice card. These alternatives refer to various hypothetical scenarios that might be used to portray an environmental good or service. These alternatives usually consist of possible combinations of various attributes and in order to portray a wide range of scenarios different levels of attributes are employed. Since it was introduced to the literature by [7], the technique has evolved in terms of *inter alia* its application, analysis and methodologies.

To date, a couple of methodologies issues in CE have been discussed in valuation literature include the analysis of status quo [8]; experimental design [9]; heterogeneity in tastes [10]; and designing choice cards [11]. However, less attention is given to issues of data coding on the CE data. To the best of our knowledge, only [3] did a study on this coding analysis. The study, however, focused more on model that includes the intercept. The purpose of this paper is to determine whether the use of different data coding give significantly different results of the estimated model. Our study differs from the [3] by focusing on estimating model that has no intercept. This is important

Corresponding Author: Mohd Zaini Abd Karim, Othman Yeop Abdullah Graduate School of Business, Universiti Utara Malaysia, 06010 Sintok, Kedah, Malaysia.
Tel: +604-9283528, Fax: +604-9286346.

as it is suggested that estimating a model using CE method with unlabeled experiment should not exclude the intercept [12][1].

The rest of the paper is organized as follows. Next section describes the CE method. The following section discusses the methodology. After that, the next section presents and discusses the empirical results. In the final section, we conclude the paper.

Choice Experiment (CE) and Model Specification: The underpinning theories behind CE technique are theory of value and random utility theory. Theory of value explains that consumers' utilities are actually based on the characteristics or attributes (or a combination of the attributes) of goods [13]. In other words, the utility received from the consumption of goods is no longer subject to the goods *per se* but to the attributes possessed by the goods. Random utility theory explains when consumers make a choice between alternatives the choice is based on an assumption about the highest utility that they can receive from it [14].

In a simple example that only consists of two goods in a market, *i* or *j*, the behavioral model is therefore, choose good *i* if and only if $U_{in} > U_{jn}$. In random utility terms, the probability that consumer *n* chooses good *i* (P_{in}) is shown in (1).

$$
\begin{aligned}
P_{in} &= P_r\,(U_{in} > U_{jn}) \\
&= P_r\,(V_{in} + e_{in} > V_{jn} + e_{jn}) \\
&= P_r\,(e_{jn} - e_{in} < V_{in} - V_{jn})
\end{aligned} \quad (1)
$$

The probability of choosing good *i* can be obtained by specifying assumptions on the distribution of the error terms, e_n. Note that, basically, the error terms are assumed to be distributed independently and identically (iid) with a Gumbell (or Type 1 extreme-value) distribution [15]:

$$f(e) = \exp\,(-\exp(-\mu e)) \quad (2)$$

Following (2), [14] showed that the selection of good *i* can be expressed in terms of a logistic function where the error terms are assumed to be distributed as Gumbell, with a scale factor . The logistic distribution can be generalized to the case of three or more parks and the function can be expressed as a Conditional Logit (CL) model.

The scale factor, however, cannot be identified in the estimation model because its value is confounded with the vector of utility parameters and it is therefore assumed that $\mu = 1$. The probability of choosing good *i* in the CL model is shown in (3):

$$P_{in} = \frac{\exp(V_{in})}{\sum_{j\in Jn} \exp(V_{jn})} \quad (3)$$

The CL is preferred to analyze CE data because its ease of estimation compared to other models such as multinomial probit [16]. In CE, qualitative attributes with multiple levels are coded with effect coding [5, 17] and [18] or dummy coding [19] and [20]. Both coding applies the same rule where with *L* levels of qualitative attribute, the total numbers of variable that must be generated has to be equal to L-1. This has to be done to avoid the dummy trap. However, each coding use different numerical value for the base level. Dummy coding use zero while effects coding use -1. When qualitative level is present, both coding use one.

Effect coding has several advantageous compared to dummy coding. Among others, the effect coding is able to orthogonalise the attribute effects with the intercept unlike dummy coding where it confounds with the intercept. This confounding occurs as we are indefinite whether the estimated coefficient for intercept measures average utility or utility at the basic level. An illustration of utility function with one attribute at three levels is used to further explain the confounding. Suppose that the attributes can be classified as high, medium and basic (status quo) levels. An indirect utility function of this illustration can be expressed as follows:

$$V_{im} = \$_0 + \$_1 D_{j1} + \$_2\, D_{j2}$$

where V_{im} refers to the response of respondent *i* on choice card *m*, $\$_0$ to $\$_2$ are the estimated coefficients and D_{j1} and D_{j2} are the dummy variables. Note that, only two dummy variables are included with three levels of qualitative attribute. The role of intercept (i.e. $\$_0$) is to capture the average effect on utility for all factors that are not included in the model [16]. By using dummy coding value, the utility of respondent *i* for average effect on utility and utility at basic level are indifferent at $\$_0$. On the other hand, this is not the case in effect coding as the utility at basic level is estimated by $\$_0 - \$_1 - \$_2$ while the average effect on utility remains the same at $\$_0$.

[1]Experiment in CE can be conducted in two ways, whether using a labeled alternative or unlabeled alternative. They differ in terms of information from the heading or title of each alternative whether it is informative or uninformative to decision maker. If the title is uninformative or generic, it is defined as an unlabeled experiment, vice versa.

Table 1: Attributes and their levels

Attribute	Level	Items
Park's Amenities	Basic	Toilet; prayer hall; picnic area and baby change
	Higher	Basic facilities plus café and shuttle bus
Recreational Facilities	Basic	Relaxation or enhancing social interaction such as children playground; jogging; reading and taking picture
	Medium	Minimum facilities plus outdoor sporting such as exercise station; bicycle hiring and boat hiring
	Higher	Medium facilities plus with adventure games such as wall climbing; flying fox and hanging bridge
Interpretation	Basic	Interpretive signs on the plants and animals
	Medium	Interpretive signs on the plants and animals and pamphlet
	Higher	Interpretive signs on the plants and animals; pamphlet; staffed stations and video presentation at information centre
Natural Attractions	Basic	A simple activity that people can engage in nature-based like viewing natural scenery
	Medium	Activities that involve people to enjoy natural scenery view and watch animal displays
	Higher	Activities that include natural scenery view; animals' displays and nature-based activities (e.g. planting)

The inclusion of intercept in the model estimation has been applied by many analysts whether in the generic alternatives (example, Alternative A and Alternative B) format (among others, [11]) or labeled alternatives (example, Botanical Park, Wild Park) format (among others, [2]). [12] Suggest to include the intercept for the labeled alternatives but not for the generic ones as the value corresponds to the intercept for the latter is meaningless in terms of its interpretation. There is nothing much that we can say on the value as the trade-offs in choice sets is between attribute levels that have no association with a particular label.

MATERIALS AND METHODS

The potential attributes to be used in the study were selected based on reviews on the related economic studies of outdoor recreation. After these potential attributes were identified, discussion with participants in focus group meeting was held to determine their suitability. Three focus group meetings with each session consists of 7-8 persons were conducted. The results from the meetings suggests the use of amenities, recreational facilities, natural attractions and information attributes. These attributes are depicted in various levels. The attributes of recreational facilities, natural attractions and information are described in three levels, while the attribute of amenities used two levels. Details of the attributes are explained in Table 1.

A software developed by Nguyen (accessible at http: //designcomputing.net/gendex/noa/) is used to develop the orthogonal main effects design (OMEP). To get a pair for each choice, the study employed the fold-over approach. Requesting respondents to answer all the eighteen generated choice cards would be burdensome for them and the worry is that their answers would be unreliable. Therefore, respondents were asked to give a response for six choice cards. Based on the pilot survey results, this amount are considered manageable.

In each card, respondents were presented with three unlabeled alternatives; Park A, Park B or Park C. Park C resembles the current situation of public parks in Malaysia. Unlabeled alternative parks (e.g. Park A and Park B) were portrayed with a combination of different attributes at various levels. Three levels of price are applied in this example; RM0, RM20 and RM35[2].

RESULTS AND DISCUSSION

From the face-to-face survey questionnaire technique, a total of 188 usable respondents on their preferences on recreational parks are used for this analysis. Each respondent is required to answer six choice cards producing a total of 1128 observations.

$$V = \$_1.\ \mathrm{Amen} + \$_2.\ \mathrm{Fac1} + \$_3.\ \mathrm{Fac2} + \$_4.\ \mathrm{Info1} + \$_5.\ \mathrm{Info2} + \$_6.\ \mathrm{Nat1} + \$_7.\ \mathrm{Nat2} + \$_8\ \mathrm{Pri} \quad (4)$$

where Amen is attribute of Park's Amenities at higher level; Fac1 is Recreational Facilities at medium level; Fac2 is Recreational Facilities at higher level; Info1 is Interpretation at medium level; Info2 is Interpretation at higher level; Nat1 is Natural Attractions at medium level; Nat2 is Natural Attractions at higher level; and Pri is price."

The effects of type of coding on Ces method can be determine form the estimation using the Conditional Logit (CL) model. Two CL models that used dummy coding value and effect coding value were estimated. The results of the estimation are presented in Table 2. The explanatory power for both models recorded an adjusted pseudo- R^2 value of 21%. To determine whether or not the coefficients are jointly zero, the Likelihood Ratio (LR) test is conducted.

[2]The exchange rate at the time of survey was US$1.00=RM3.80

Table 2: Conditional Logit and WTP estimation

	Dummy Coding		Effect Coding	
Variable	Coefficient	WTP[a]	Coefficient	WTP[a]
Amen	0.51792*** (0.08681)	6.20244***	0.25896*** (0.04341)	3.10122***
Fac1	1.05052*** (0.11548)	12.58078***	0.16026** (0.06665)	1.91926**
Fac2	1.62026*** (0.12212)	19.40378***	0.72999*** (0.07048)	8.74226***
Info1	0.20219** (0.10245)	2.42133**	0.05657 (0.06516)	0.67745
Info2	0.23467* (0.13908)	2.81031*	0.08905 (0.08482)	1.06643
Nat1	0.33570*** (0.11099)	4.02028***	0.10661 (0.06603)	1.27678
Nat2	0.35156*** (0.11571)	4.21022***	0.12247* (0.06868)	1.46672*
Pri	-0.08350*** (0.005002)	-	-0.08350*** (0.005002)	-
Log Likelihood function	-970.5034	-970.5034		
R-squared	0.21	0.21		

***significant at 1%, ** significant at 5% and *significant at 10%; std. errors are in brackets

[a]WTP is calculated in RM

In both models we can reject the null hypothesis that the coefficients are jointly equal to zero at the 1% significance level (LR statistic is 532.18) suggesting that the specified models are appropriate.

To investigate whether the individual coeficient is significantly different from zero ($\$_j = 0$), we used the Wald statistics test. The results show that, in the case of dummy coding, all variables are significant at least at the 10% level. This does not holds for effect coding. Some of the estimates, for example, coefficients of Info1, Info2 and Nat1 are not statistically significant. For the dummy coding, in terms of the marginal utility, the highest contribution comes from Fac2 followed by Fac1, Amen, Nat2, Nat1, Info2 and Info. On the other hand, for the effect coding, the highest contribution comes from Fac2 followed by, Fac1, Amen and Nat2. It is noteworthy that the coefficient values of the higher level were greater than the coefficient values for the lower level (coefficient of Fac2 is higher than Fac1). This indicates that the marginal utility received by respondents from the higher level attributes are greater than the marginal utility received from the lower level. This follows the axioms of *non-satiation*, where the utility received by a consumer increases if the commodity used by the consumer increases.

A comparison between dummy and effect coding can also be examined from the willingness to pay (WTP) value for attributes. The WTP explains the amount of money that consumers' willing to pay in order to get an additional improvement in an attribute [21]. This WTP can be obtained by calculating the ratio of coefficients for the attributes (or levels) with the parameter of price (or cost). In the study, the WTP was calculated using the Wald procedure (Delta method) in Limdep 8.0. Before calculating the WTP, analysts have to make sure that both attributes used to calculate the WTP are statistically significant, otherwise the calculated WTP is meaningless [12].

The WTP value of Amen in the case of dummy coding shows that respondents are willing to pay RM6.20 to have an improvement in the attribute from basic to the higher level. The value is calculated by dividing the coefficient of Amen (0.51792) with the coefficient of price (0.08350). Since both coefficients are statistically significant, the WTP is significant as well. However, in the case of effect coding, although the coefficient of price is significant, the coefficient of Info1 is not statistically significant preventing from calculating its WTP.

Overall, the WTP calculated from dummy coding is higher than the one calculated from effect coding. The WTP of Amen in dummy coding is twice the value of its WTP in effect coding. This is because the attribute consists of two levels and one can anticipate the difference is due to the use of different coding. However, the WTP calculated in dummy and effect coding is similar in terms of their differences between levels. For example, the difference of WTP between Fac2 and Fac1 is similar for both coding (both has a value of RM6.82).

CONCLUSION

This paper examined the effects of coding for attributes of qualitative levels in CE. To date, two types of coding being used by analysts to investigate the non-linear effects in the levels of attributes, namely dummy coding and effect coding. With each coding applies different value, the estimated coefficients and

WTP is expected to be difference. As suggested in the literature, effect coding should be use in estimating model that includes intercept due to the confounding problem. This has been supported with empirical results by [3]. However, there has been less discussion on models that exclude intercept.

This paper applies both coding approach in the analysis of consumer preferences on type of public parks to be visited in future. The findings from the analysis indicate that both dummy and effect coding produce similar results in terms of the model goodness of fit and coefficient of price. However, since the coefficients estimates differ, the value of WTP also differs. In addition, the estimated model using dummy coding seems to produce better results than the estimated model of effect coding based on the total number of significant coefficients. Coefficients of Info1, Info2 and Nat1 which are significant when dummy coding is applied are not significant in the case of effect coding. Hence, the WTP calculation in the case of dummy coding is more reliable. Based on the results, ones are suggested to use dummy coding when the estimation model does not include intercept. The finding suggests that the interpretation of estimates using different coding should be done with caution as it gives different results which can leads to different policy implications.

REFERENCES

1. Hanley, N., R.E. Wright and G. Koop, 2002. Modelling recreation demand using choice experiments: climbing in Scotland. Environmental and Resource Economics, 22(3): 449-466.
2. Morrison, M., J. Bennett, R. Blamey and J. Louviere, 2002. Choice modeling and tests of benefit transfer. American Journal of Agricultural Economics, 84(1): 161-170.
3. Bech, M. and D. Gyrd-Hansen, 2005. Effects coding in discrete choice experiments. Health Economics, 14(10): 1079-1083.
4. Lusk, J.L. and T.C. Schroeder, 2004. Are choice experiments incentive compatible? A test with quality differentiated beef steaks. American Journal of Agricultural Economics, 86(2): 467-482.
5. Kemperman, A.D.A.M., A.W.J. Borgers, H. Oppewal and H.J.P. Timmermans, 2000. Consumer choice of theme parks: A conjoint choice model of seasonality effects and variety seeking behavior. Leisure Sciences, 22(1): 1-18.
6. Bateman, I., R.T. Carson, B. Day, M. Hanemann, N. Hanley, T. Hett, *et al*., 2002. Economic valuation with stated preference techniques. Massachussets: Edward Elgar.
7. Louviere, J. and G. Woodworth, 1983. Design and analysis of simulated consumer choice or allocation experiments: An approach based on aggregate data. Journal of Marketing Research, 20: 350-367.
8. Scarpa, R., K.G. Willis and M. Acutt, 2007. Valuing externalities from water supply: status-quo, choice complexity and individual random effects in panel kernel logit analysis of choice experiments. Journal of Environmental Planning and Management, 50: 449-466.
9. Bliemer, M.C.J. and J.M. Rose, 2006. Designing Stated Choice Experiments: State-of-the-Art. Paper presented at the 11th International Conference on Travel Behaviour Research.
10. Scarpa, R. and M. Thiene, 2005. Destination choice models for rock climbing in the Northeastern Alps: A latent-class approach based on intensity of preferences. Land Economics, 81(3): 426-444.
11. Blamey, R., J. Bennett, J. Louviere, M. Morrison and J. Rolfe, 2002. Attribute causality in environmental choice modelling. Environmental and Resource Economics, 23(2): 167-186.
12. Hensher, D.A., J.M. Rose and W.H. Greene, 2005. Applied choice analysis: A Primer. Cambridge: Cambridge Univerisity Press.
13. Lancaster, K., 1966. A New Approach To Consumer Theory. Journal of Political Economy, 74: 132-157.
14. McFadden, D., 1973. Conditional Logit Analysis of Qualitative Choice Behaviour. In: P. Zarembka, (Ed.), Frontiers in Econometrics. New York: Academic Press.
15. Swait, J., 2007. Advanced Choice Models. In: B.J. Kanninen, (Ed.) Valuing Environmental Amenities Using Stated Choice Studies: A Common Sense Approach to Theory and Practice. The Netherlands: Springer.
16. Train, K., 2003. Discrete Choice Methods with Simulation. Cambridge: Cambridge University Press.
17. Lawson, S.R. and R.E. Manning, 2003. Integrating multiple wilderness values into a decision-making model for Denali National Park and Preserve. Journal for Nature Conservation, 11(4): 355-362.

18. Oh, C.O., R.B. Ditton and R. Riechers, 2007. Understanding Anglers' preferences for fishing tournament characteristics and policies. Environmental Management, 40(1): 123-133.
19. Bergmann, A., S. Colombo and N. Hanley, 2008. Rural versus urban preferences for renewable energy developments. Ecological Economics, 65(3): 616-625.
20. Rolfe, J. and J. Bennett, 2009. The impact of offering two versus three alternatives in choice modelling experiments. Ecological Economics, 68(4): 1140-1148.
21. Bennett, J. and V. Adamowicz, 2001. Some Fundamentals of Environmental Choice Modelling. In: J. Bennett and R. Blamey, (Eds.) The Choice Modelling Approach to Environmental Valuation. Cheltenham,UK: Edward Elgar.