To Include or Exclude Attributes in Choice Experiments: A Systematic Investigation of the Empirical Consequences

Jordan J. Louviere, University of Technology, Sydney
Towhidul Islam, University of Guelph

Abstract

We investigate the stability of brand and price attribute tradeoffs in choice experiments for two product categories under conditions in which seven other attributes are systematically included/excluded according to a master experimental design. Choice model estimates from 64 different experimental conditions were converted to willingness-to-pay (WTP) units to control for error variance differences between conditions and allow inter-attribute utility comparisons. Statistical analysis of the resulting 64 sets of WTP estimates for all attributes (except price, which is used in the WTP calculations) demonstrates that even after controlling for variance differences, there are large, systematic differences in utility estimates depending on which attributes are included/excluded in choice experiments. Implications for future research are noted.

Keywords: Willingness-to-pay (WTP), missing information, discrete choice experiments

Introduction

This paper reports on a systematic investigation of the empirical consequences of including/excluding attributes/features in discrete choice experiments. (aka "choice-based conjoint"). Academics and practitioners commonly suggest that the number of attributes varied in conjoint and choice experiments should be limited (eg, Carson, et al 1994), and various methods have been proposed to limit the number of attributes varied for any one respondent, such as the partial profile approach proposed by Chrzan and Elrod (1995) and hybrid conjoint proposed by Green (1984), to name only two. Many other researchers arbitrarily limit the number of attributes by using ad hoc methods to define "the most salient attributes", such as the qualitative approaches discussed by Louviere (1988) and Louviere, Hensher and Swait (2000).

A separate literature in psychology and marketing has investigated the process by which humans make decisions when there is "missing information" (e.g. Huber and McCann 1982; Johnson and Levin 1985; Simmons and Lynch 1991). While there is no consensus in this research area, it is fair to say that there is evidence to suggest that inferences about effect sizes and attribute significance depends on the nature and composition of attributes varied. The latter evidence obviously begs the question as to whether results of conjoint and choice experiments systematically differ depending on the attribute information included/excluded in experiments. This paper investigates this issue for two product categories (pizza and fruit juices), and finds differences associated with including or excluding attributes in experiments.

Choice Experiments

Discrete choice experiments (DCEs) were pioneered by Louviere and Woodworth (1983), and while this approach often is termed "choice-based conjoint", DCEs only superficially resemble traditional conjoint analysis. For example, as Louviere (2001) notes, DCEs have an underlying behavioral theory (random utility theory or RUT) that supports many forms of statistical models ranging from simple (eg, conditional logit) to complex (eg, mixed logit, multinomial probit). Unlike traditional conjoint analysis, RUT provides a theoretically sound, comprehensive approach to modeling the structure and evolution of markets, different choice states (choose now, delay, never choose), and choice of volumes, trial and repeat, etc.

A choice experiment consists of one or more options that are described by different levels of a set of experimental attributes. Thus, DCEs can be simple, such as presenting options described by different combinations of attribute levels one-at-a-time, asking respondents to state whether they would/would not take some action (eg, buy/not buy); or DCEs can be complex, such as presenting attributes of several competing brands. Regardless of whether DCEs are simple/complex, all DCEs require decisions about which attributes to include/exclude. Thus, potentially serious issues are associated with the empirical consequences of including/excluding attributes. However, little attention seems to have been directed towards the potential biases that might result from such decisions or the extent to which different results obtain and/or different inferences are made about attribute effects and, the distribution of choice probabilities.

The purpose of this paper is to systematically investigate differences in model parameters and choice probabilities associated with inclusion or exclusion of various attributes in DCEs.

Research Approach

To rigorously examine inclusion/exclusion of attributes in DCEs, we developed an overarching experiment to vary presence/absence of a number of attributes. Prior to designing the experiment we undertook qualitative and quantitative research to identify potential lists of attributes that influence choices of delivered pizza and fruit juice products. Following recommendations made by Louviere (1988) and by Louviere, Hensher and Swait (2000), we identified a long list of potential attributes, and selected the nine (for pizza) and ten (for fruit juice) most salient attributes identified for each product category. Because brand and price are fundamental features of virtually all DCEs, we excluded brand and price from the overarching experiment for both pizza and juice ("Juice Type" is also excluded from Juice) , but included them in the DCEs.

The overarching presence/absence experiment therefore targets seven attributes for pizzas and juices. The overarching experiment allows us to vary presence/absence of the target attributes systematically and independently. Each target attribute is present/absent; hence, all possible combinations are given by a 2^7 factorial, which is too large to fully study. Instead, we based the experiment on a fraction of the complete factorial that has the property that all main effects and two-way interactions are orthogonal to one another (known as a "resolution 5" design). Thus, each treatment combination defines the attributes that are present and included in each DCE.

The overarching design contains 32 treatment combinations representing different experimental conditions. Each condition defines a DCE, so there are 64 total DCEs for both pizzas and juices. To make each DCEs as comparable as possible, we framed them as simple, one-at-a-time, "would you buy/not buy this offering" binary response tasks (See, e.g., Louviere, Hensher and Swait 2000, pp. 101-110). We also used the same 8 x 4^8 orthogonal main effects plan (OMEP) to construct the attribute level combinations in each DCE. The 8-level column can be transformed into a 2^3 or a 2 x 4 or several other possibiliites, and each 4-level column can be transformed into a 2 x 2. This property of regular fractions allows us to "mix and match" columns in the OMEP to create all the DCEs required. We randomly assigned 18 subjects to each DCE to obtain a total of 18 x 64 subjects (= 1152 total). A website was developed to administer the DCEs over the web and insure that subjects were randomly assigned to each condition. Respondents first answered some questions about the importance of each attribute and few personal characteristics questions (not considered in this paper).

Results

The results are contained in Table 1, which summarises the statistical results for the 32 experimental conditions for both the products. The tables contains two sets of results: 1) the mean implied willingness-to-pay (in dollars) for each attribute level, and the T-statistics; and 2) the estimated effects of the presence/absence of each of the experimental attributes on the willingness-to-pay (WTP) estimates. WTP measures are not commonly reported in academic and commercial DCEs in marketing; hence, they bear some discussion.

WTP is well-known in economics, where it is based on and derived from welfare theory. WTP measures compensating variation, which is the number of units of one attribute that will substitute ("compensate") for a change made to a second attribute (or changes made to more than one attribute). Thus, one needs a reference attribute that is (at least in principle) continuous, such as price. Thus, WTP is defined as the utility parameter associated with each attribute level divided by the utility estimate of price. Each attribute is measured in units of utilities per unit (or difference), and price is measured in utilities per dollar. Thus, the ratio of attribute utility estimate to price utility estimate is in attribute units (or differences) per dollar.

Not only is WTP directly interpretable, but it also has another important property for the tests that we wish to conduct. That is, all discrete choice models confound the utility parameters with the variance of the error component, such that one estimates β/σ_ε, not β per se (β is a utility estimate, σ_ε is the standard deviation of the error). As Swait and Louviere (1993) note, the confound does matter a great deal when one wants to compare model results from different sets of data, which is the case here. Specifically, one must control differences in error variability between data sets to make legitimate comparisons. Because WTP is the ratio of two choice model estimates, σ_ε is a common denominator, and hence WTP cancels the variability, and allows us to control for differences in data sets without resort to complex statistical methods.

Thus, Table 1 contain descriptive statistics for the WTP estimates for each attribute level. To the extent that all conditions measure the same effects, one expects that all WTP estimates will be within the 95% confidence bounds of the mean estimates. Although not obvious from Table 1, all attributes display WTP estimates that lie very far outside the 95% confidence bounds, which suggests that some conditions produced significantly different estimates of the attribute utilities. More detailed investigation of differences in WTP is given by a regression analysis of the variation in WTP across the experimental condition as a function of the overarching design matrix (presence/absence). The last nine columns (for each product) in Table 1 contain the regression results, with significant effects highlighted in bold. A few words of detail about this analysis are in order. We included the obtained R-Square value associated with each model in each condition because this measures the level of explained variability in each model/condition. We would expect that the R-Square measures would be uniformly non-significant across conditions. However, if there is additional unexplained variability that is associated with conditions, we would expect to find significant R-Square effects. As can be noted in the tables, there are significant R-Square effects, which suggest that not only did the response means differ, but the error variability also differed from condition-to-condition. As can be seen from pizza results in Table 1, crust and delivery time were especially affected. In the case of price, the regression analysis was performed on the 32 price utility estimates not WTP estimates, so these results show that price effects vary systematically with number of toppings and delivery time, and the R-Square effect shows that higher fits (less error variability) are associated with much greater apparent price sensitivity. For juices, WTP differences are associated primarily with the percentage of real fruit juice, and there is an interesting systematic effect associated with presence/absence of type of packaging on many of the WTP ratios. As with pizza, the price results reflect effects on price utility estimates, and one can see the same effect of R-Square as in pizza, namely that higher fits are associated with significantly more price sensitivity. There are fewer significant effects of R-Square on WTP estimates, suggesting that unobserved error variability associated with particular attributes is less of a problem for juices, but it is associated with fresh versus concentrate, percentage of real fruit juice and Tropicana brand. While we could estimate two-way interactions due to the presence/absence conditions, we do not consider that analysis here in the interests of simplicity and page limits.

Conclusions

Our empirical results show that there can be very significant differences in the utility estimates associated with particular attributes depending on what other attributes are included/excluded in DCEs. Although we do not include a detailed analysis of the unexplained variability, we can show that the lowest error variability is associated with full information, namely using all the attributes. A useful next step in this research program would be to use models from each condition to predict the choices in the other conditions, which would measure out-of-sample fits. Even better would be to obtain actual choice data, such as reported last (most recent) choices of products and/or suitable scanner data, and conduct external validity tests of the ability of the various models to predict real choices.

References

Carson, R. Louviere, J. Anderson, D., Arabie, P., Bunch, D., Hensher, R., Johnson, R., Kuhfeld, W., Steinberg, D., Swait, J. Timmermans, H. and Wieley, J., 1994. Experimental Analysis of Choice. Marketing Letters. 5(4), 351-367.

Chrzan, K. and Elrod, T., 1995. Choice-Based Approach for Large Numbers of Attributes. Marketing News. 29(1), 20.
Green, P. E., 1984. Hybrid Model for Conjoint analysis: An Expository Review. Journal of Marketing Research. 21, 155-169.

Huber, J. and McCann, J., 1982. The Impact of Inferential Beliefs on Product Evaluations, Journal of Marketing Research. 19, 324-333.

Johnson, R. D. and Levin, I. P., 1985. More than Meets the Eye: The Effect of Missing Information on Purchase Evaluations. Journal of Consumer Research, 12 (September), 74-81.

Louviere, J. and Woodworth, G., 1983. Design and Analysis of Simulated Consumer Choice or Allocation Experiments. Journal of Marketing Research. 20 (November), 350-367.

Louviere, J. J., 2001. What If Consumer Experiments Impact Variances as Well as Means? Response Variability as Behavioral Phenomenon. Journal of Consumer Research.28(3), 506-511.

Louviere, J. J., Hensher, D. A. and Swait, J., 2000. Stated Choice Methods: Analysis and Applications, Cambridge University Press.

Louviere, J., 1988. Conjoint Analysis Modeling of Stated Preferences: A Review of Theory, Methods, Recent Developments and External Validity. Journal of Transport Economics and Policy. 22(1), 93-120.

Simmons, C. J. and Lynch, Jr, J.G., 1991. Inference Effects without Inference Making? Effects of Missing Information on Discounting and Use of Presented Information. Journal of Consumer Research. 17, 477-491.

Swait, J. and Louviere, J. J., 1993. The Role of the Scale Parameter in the Estimation and Comparison of Multinomial Logit Models. Journal of Marketing Research. 30(August), 305-314.