

UC San Diego

Recent Work

Title

Statistical Properties of Consideration Sets

Permalink

<https://escholarship.org/uc/item/9qt9g4p4>

Authors

Carson, Richard T.
Louviere, Jordan J.

Publication Date

2006-07-01

2006-07

UNIVERSITY OF CALIFORNIA, SAN DIEGO

DEPARTMENT OF ECONOMICS

Statistical Properties of Consideration Sets

By

Richard T. Carson
Department of Economics
University of California, San Diego

Jordan J. Louviere
School of Marketing
University of Technology, Sydney

DISCUSSION PAPER 2006-07
July 2006

STATISTICAL PROPERTIES OF CONSIDERATION SETS

Richard T. Carson
Department of Economics
University of California, San Diego

Jordan J. Louviere
School of Marketing
University of Technology, Sydney

Draft
July 2006

Abstract

Consideration sets have become a central concept in the study of consumer behavior. Frequently, consumers are asked to split choice alternatives into those that they would consider and those that they would not. Information on alternatives not in the consideration set is then typically not used in subsequent analysis. This practice is shown to lead to biased estimates of preference parameters. The reason for this is shown to be a form of sample selection bias.

JEL Codes: D1, M31, C35

Keywords: choice models, random utility, sample selection bias

Statistical Properties of Consideration Sets

Abstract

Consideration sets have become a central concept in the study of consumer behavior. Frequently, consumers are asked to split choice alternatives into those that they would consider and those that they would not. Information on alternatives not in the consideration set is then typically not used in subsequent analysis. This practice is shown to lead to biased estimates of preference parameters. The reason for this is shown to be a form of sample selection bias.

1. Introduction

The concept of a consideration set is well-established in marketing (e.g., Kotler, 2003; Roberts and Nedungadi 1994; Chiang, Chib and Narasimhan 1999). In applied work related to understanding and modelling consumer preferences, consideration sets are often used to classify goods into two groups: 1) those that a consumer would consider purchasing and 2) those that a consumer would not consider purchasing (probabilistically or deterministically). Each good that is so classified typically can be decomposed into bundles of attributes like prices and features, such as price, size, color, quality and brand name. For each good classified as being “in” a particular consumer’s consideration set, it is common to ask the consumer additional questions to obtain preference information, such as ranking or rating each “considered” good, or which of the “considered” goods would be chosen or were chosen most recently (i.e., measures that are used to identify the consumer’s most preferred option – see, e.g., Narayana and Markham 1975; Reilly and Parkinson 1985; Horowitz and Louviere 1995). Observed consideration sets, attribute/feature measures and preference information often are used to estimate statistical (choice) models.

Indeed, the literature seems to have evolved two streams of research involving consideration sets, namely one in which consideration sets are treated as endogenous quantities to be estimated from consumer panel data and/or from consumer choice experiment data, and a second in which the consideration sets are given exogenously as quantities

defined by some measurement process. Examples of the former include Roberts and Lattin (1991), who developed model of how consumers form a consideration set at a particular point in time; Andrews and Srinivasan (1995), who developed a two stage model of consideration and choice; and Chiang, Chib and Narasimhan (1999), who developed a model of consideration set formation and choice that allows for heterogeneity in both processes. Examples of the latter include Narayama and Markham (1975), who classified brands into “inept” and “inert”; Wright and Barbour (1977), who coined the term “consideration set” and suggested brands that are “known” to consumers could be classified into acceptable and unacceptable; and Horowitz and Louviere (1995), who examined aided and unaided recall questions to measure which brands were/were not considered.

Our review of the large and growing literature associated with the definition, measurement and use of consideration sets in marketing suggests that there has been surprisingly little research into the statistical implications of the common empirical practice of defining a consideration set in some fashion and then estimate a preference or choice model based on alternatives deemed to be inside the consideration set.¹ Thus, the purpose of this paper is to rigorously investigate the statistical implications of using consideration set information in statistical model estimation, and to describe the consequences of using such information to estimate statistical models and derive policy inferences from the results.

In particular, we examine the statistical implications of using consideration set information in light of research results in the econometrics literature dealing with sample selection and truncation issues (e.g., Maddala, 1987; Greene, 2000). To begin our discussion, we consider the existence of some idealized utility index (See Deaton and Muellbauer, 1980), such as (Hicksian) maximum willingness to pay or net willingness to pay if price is included

¹ There are, of course, a number of competing theoretical and statistical models for determining whether an alternative is or is not in a consideration set (see for example the papers in the special issue on consideration sets edited by Roberts, J.H. and P. Nedungadi, 1995). This paper is not intended to enter that debate but rather to address the much simpler question of what happens if a standard statistical analysis is performed only using the consideration set, however, defined.

as a variable.² We then consider the case where only a simple discrete indicator of the most preferred alternative is available. In both cases, we rely on very simple assumptions about the translation of consumer preferences into a statistical framework in order to illustrate the main points that generalize to more complex and realistic models.

More generally, an examination of consideration sets is warranted because of potential confusion in the use of terms and application of measurement methods. In particular, the term “choice set” is well-established in the probabilistic discrete choice modeling literature (e.g, Ben-Akiva and Lerman 1985) and in many studies of consumer choice behavior (e.g., Louviere, Hensher and Swait 2000). That is, if a consumer faces a choice among J options, a choice set is some subset of all possible combinations of options, or $2^J - 1$ combinations (one subset is null). Traditionally, the choice set for any particular individual is given exogenously by observation or measurement (for an exception, see Swait 2001) or is fixed by the researcher as in many stated choice experiments (see, e.g., Louviere, Hensher and Swait 2000). The notion of a consideration set implies that individuals actively or passively reduce the size of choice sets, such that a consideration set, denoted C_n for the n -th individual, is $< J$, the total number of options available for choice in a particular context. The consideration set literature is dominated by the view that one can observe or measure such sets for different individuals by “asking them” (in various ways) which of the J options would be “considered”. Leaving aside issues of the potential lack of incentive compatibility associated with the way in which such questions are asked (Carson, Groves and Machina, 1999), it is important to ask whether the use of consideration set information to define choice sets and specify models is consistent with the underlying statistical theory associated with

² The willingness to pay measure is ideal because it is continuous, has a true “zero” point and a scale “dollars” that is interpreted in the same way by all respondents. In practice, willingness to pay may be difficult to obtain so another almost continuous variable such as ratings of the different alternatives on some scale may be elicited. Such a measure does not circumvent any of the issues raised here.

statistical models of consumer preferences and/or poses problems of estimation bias. This paper focuses on these latter issues.

To anticipate our results, we provide a brief summary before beginning a more rigorous treatment: 1) If one estimates statistical models from datasets in which only “considered” options are included in the estimation, this will result in biased estimates of the preference parameters. 2) Even if one could obtain consistent estimates of the preference parameters, limiting the model analysis to “considered” options produces biased estimates of confidence intervals for the preference parameters. 3) Even if one ignores issues related to the consistency of the estimates of the preference parameters and their confidence intervals, the information content associated with including a randomly chosen option in a consideration set generally will be less than a randomly chosen option that is not in a consideration set. These three results are immediate consequences of the fact that using consideration set information to measure consumer choice sets represents a type of truncation of the dependent variable. Specifically, the statistical issues posed by using consideration set information in model estimation are equivalent to those that have been studied by labor economists related to sample selection (Heckman, 1979; 1990).

2. Continuous Case

It is useful to begin with the continuous dependent variable case because it is easier to see the issues, and helps to make our major point readily transparent. That is, we assume that willingness to pay (WTP) for the j -th good can be represented as:

$$WTP_{ij} = \beta X_j + \varepsilon_{ij},$$

where there are $i=1, 2, \dots, n$ individuals, X_j is a vector of attributes of the j th good, β is a vector of preference parameters to be estimated, which for simplicity is assumed to be the same for all individuals, and ε_{ij} is a random component. The random component is the critical

element in this discussion, and one can make different assumptions about various underlying processes that give rise to it, ranging from unobserved individual characteristics to pure random components associated with optimization errors. We make a very simple assumption, namely that ε_{ij} is normally distributed with mean zero and standard deviation σ for all of the $j=1, \dots, k$ goods, and the correlation between error terms for different goods is zero.

Although researchers use different types of questions to obtain consideration set information, in general they effectively ask individuals to divide a set of goods into two subsets and report WTP_{ij} for the subset that will be actively considered. A key insight in our discussion of this issue is obtained by noting that a good will be/not be in a consideration set if:

$$\beta X_j + \varepsilon_{ij} \leq C_i,$$

where C_i is referred to as the censoring or cut point. Thus, a good can fail to enter a consideration set for three reasons: 1) an individual perceives the X_j vector as undesirable, 2) a sufficiently small value of ε_{ij} , or more generally, 3) a linear combination of the two preceding reasons.

Following Greene (2000, pp. 897-905), we first define certain functions of a truncated normal distribution in terms of the quantity: $\alpha_{ij} = (C_i - \beta X_{ij})/\sigma$:

1. A term known as an inverse Mills ratio, also known as a hazard function:

$$\lambda(\alpha_{ij}) = f(\alpha_{ij})/[1 - F(\alpha_{ij})],$$

where $f(\bullet)$ and $F(\bullet)$ are pdf and cdf of the normal distribution respectively.

2. The quantity

$$\delta(\alpha_{ij}) = \lambda(\alpha_{ij})[\lambda(\alpha_{ij}) - \alpha_{ij}],$$

can be shown to take on values between 0 and 1 for all values of (α_{ij}) .

Given these two definitions, it can be readily shown that

$$E(WTP_{ij} | WTP_{ij} > 0) = \beta X_{ij} + \sigma \lambda(\alpha_{ij}),$$

and that

$$\partial E(\text{WTP}_{ij} \mid \text{WTP}_{ij} > 0) / \partial X_{ij} = \beta[1 - \delta(\alpha_{ij})],$$

rather than β . Because $\delta(\alpha_{ij}) < 1$, the absolute value of the estimate of β will be attenuated toward zero. The bias generally increases in C_i , which governs the fraction of observations omitted from estimation.

It also is possible to show that $\text{VAR}(\text{WTP}_{ij} \mid \text{WTP}_{ij} > 0) = \sigma^2[1 - \delta(\alpha_{ij})]$ also will be biased toward zero, which implies that confidence intervals will be too small. Although the constant term in an OLS regression on the positive WTP_{ij} will recenter the residuals to have zero mean, the presence of the α_{ij} term in the variance expression shows that heteroscedasticity is present even if the original e_{ij} are homoscedastic. One can obtain consistent estimates of β using several different techniques if it is reasonable to assume that all of the C_i are known (Greene, 2000, Maddala, 1987).

It also is possible to show that procedures that use information (i.e., X 's) about goods that do not enter consideration sets (i.e., they are censored) in addition to the observed values of WTP_{ij} are more efficient than those that use only the observations for which WTP_{ij} is observed. Several factors underlie the gains in efficiency from using information about the number of goods that do not enter consideration sets. To begin, consider the case in which there are no attributes of the goods. In this case the number of goods that do not enter the consideration set helps define the properties of the error distribution. Now consider the case where the goods have attributes and recall that $\beta X_j + \varepsilon_{ij} \leq C_i$ or equivalently those for which $\beta X_j \leq C_i - \varepsilon_{ij}$. Since the ε_{ij} are random normal variables, the X_j that fulfill this condition will differ from the X_j included in the consideration set. Thus, exclusion of the latter X_j will reduce the variation in the design (estimation) matrix, and consequently increase the size of the confidence intervals for the preference parameters. Moreover, if most consumers have clear preferences for some attribute values, the X_j associated with the consideration set will be

“similar” in many respects, which will further reduce the variability of the design matrix. If this the latter obtains, replacing a randomly chosen good in the consideration set with one not in the consideration set almost always will improve precision of preference parameter estimates.

The preceding discussion of consideration sets closely parallels the econometrics literature on issues related to the effects of selecting samples based upon the value of the dependent variable. All the issues noted above have been shown to be potentially correctable, although it may be difficult to satisfy the assumption that all C_i are known. The simplest variants of sample selection models assume that C_i is a known constant for all individuals, which typically is zero. The latter might make sense for consideration set questions that ask for which goods consumers have positive willingness to pay, but this is not a common way to elicit consideration set information. If C_i is unknown but is the same for all individuals, there are ways to consistently estimate the preference parameters (see Carson, 1989). However, for quasi-continuous variables like rating scales that are commonly used in academic and applied marketing research, it is well-known that different people use the same rating scales in different ways. In this case, the entire statistical model is unidentified because each individual's C_i is absorbed into the estimate of the preference parameters, resulting in the appearance of a distribution of preference parameters even if all individuals actually share the same vector of β . This problem can be circumvented only if there is a reliable way to estimate the C_i used by each individual. In the latter case a second equation is required to predict the C_i that involves the restriction that some variables in this second equation cannot appear in the original preference specification, or that the error terms of the two equations are uncorrelated.

3. Discrete Case

The discrete case is analogous to the continuous case with some important exceptions, which can be seen by noting that the usual representation of the utility of the j^{th} good in a random utility model (RUM) is:

$$U_{ij} = V_{ij} + \varepsilon_{ij},$$

where in simple formulations V_{ij} is usually parameterized as βX_j , and ε_{ij} is independent (of other options) and identically (drawn from the same error distribution) distributed error term, with different distributional assumptions such as the normal or extreme value, which leads to different statistical models. It should be noted that this IID assumption is effectively the same as the assumption made in the continuous case, and the strong link between the two models is well-known in the literature (Cameron and James, 1987).³ Thus, it is straightforward to show that all issues in the continuous case will obtain, although three differences arise that are worthy of note.

The first difference is that instead of directly estimating the parameter β , one estimates β/s where s is a scale factor. This loss of information on scale is a consequence of obtaining discrete information about preferences, and raises the interesting possibility that the truncation bias might “cancel out” under the common practice of looking at ratios of estimated attribute parameters to investigate marginal tradeoffs. This interesting issue deserves further exploration, but existing results suggest that the condition under which this is likely to happen is when a design matrix is multivariate normal (Cheung and Goldberger, 1984), a condition that is highly unlikely in real applications.

³ Further, the notion of a consideration set bears a strong resemblance to the structure of a nest logit model, which is one of the primary tools used to relax the IID assumption. Here, Cardell (1997) cautions against the limited information maximum likelihood approach to estimating nested logit models (e.g., partitioning into different sets and separately estimating) because it collapses the variance of the X space and is not robust to even a small amount of heterogeneity in consumer preferences.

The second difference is that instead of a cut point for inclusion in consideration sets being expressed in terms of willingness to pay, the cut point for inclusion now can be expressed in probability terms, such that options that have a sufficiently high probability of being chosen are included in consideration sets. Again it is noteworthy that the key condition that drove the result in the continuous case still holds. That is, a low probability will be indicated for goods with undesirable X_j or small ε_{ij} , which makes the issue of individual C_i 's even clearer. In particular, different individuals will make probabilistic cutoffs for inclusion in consideration sets differently.

The third and final point is that Horowitz and Louviere (1995) show that it is possible to consistently estimate the parameters of the RUM model by using information on which vectors of X_j are in consideration sets and which are not. In particular, each good in a consideration set is known to be preferred to each good not in a consideration set; the latter defines a potentially large set of preference relations between different goods, but no sample selection issues are involved in this approach. Further, if one asks individuals to indicate their most preferred option in their consideration set, this provides additional preference relationships between goods, which information can be exploited in estimation (see, e.g., Hensher, Louviere and Swait 1999; Louviere, Hensher and Swait 2000).

4. Discussion and Conclusions

Our results suggest that failure to use information on options not included in consideration sets leads to a form of truncation that results in sample selection bias. That is, the sample is selected on the basis of the value of the dependent variable, which can be influenced either by attributes of goods or by error components. In this case, more complex econometric estimation procedures would be indicated in order to obtain consistent parameter estimates. However, if one uses individual cut points to include or exclude goods in

consideration sets it will be difficult if not impossible to use these more complex procedures in the continuous case. The discrete choice case is more promising because one naturally can exploit information on X_j both in and out of consideration sets.

The assumption, implicit or explicit, lying behind the use of consideration sets is that consumer decision-making with respect to alternatives in and out of the consideration set is somehow different. The analysis presented here suggests this assumption will be difficult to test empirically without some type of structural model. Any procedure that effectively divides alternatives into those more likely and less likely to be chosen must operate at least in part on the basis of the magnitude of the ϵ_{ij} . Any consideration set procedure will also tend to narrow the included range of attributes. This will reduce the ability to statistically identify the true underlying functional form for the preference/choice model within the consideration set. This is because within a restricted range of the attribute levels, local linearity (*i.e.*, the systematic component V_i equals $X_i\beta$) is likely to be a good approximation, even though the actual preference function over a larger X space might be well-characterized by some more general smooth but non-linear function $f(X_i, \beta)$.

References

- Andrews, R.L. and T.C. Srinivasan (1995), "Studying Consideration Effects in Empirical Choice Models Using Scanner Panel Data," *Journal of Marketing Research*, 32, 30-41.
- Ben-Akiva, M. and S. Lerman (1985), *Discrete Choice Analysis* (Cambridge: MIT Press).
- Cameron, T.A. and M.D. James (1987), "Efficient Estimation Methods for 'Closed Ended' Contingent Valuation Surveys," *Review of Economics and Statistics*, 69, 269-276.
- Cardell, N.S. (1997), "Variance Components Structures for the Extreme-Value and Logistic Distributions with Application to Models of Heterogeneity," *Econometric Theory*, 13, 185-213.
- Carson, R.T. (1989), "The Tobit Model with a Non-Zero Censoring Threshold," paper presented at the Econometric Society Meeting, Atlanta, December.
- Carson, R.T. and T. Groves and M.J. Machina), "Incentive and Informational Properties of Preference Questions," Plenary Address, European Association of Environmental and Resource Economists, Oslo Norway, June 1999.
- Chiang, Chib and Narasimhan (1999), "Markov Chain Monte Carlo and Models of Consideration Set and Parameter Heterogeneity," *Journal of Econometrics*, 9, 223-248.
- Cheung C. and A. Goldberger (1984), "Proportional Projections in Limited Dependent Variable Models," *Econometrica*, 52, 531-53.
- Deaton, A. and J. Muellbauer (1980), *Economics and Consumer Behavior* (New York: Cambridge University Press).
- Greene, W. (2000), *Econometric Analysis*, 4th ed. (Englewood Cliffs, NJ: Prentice Hall).
- Heckman, J.J. (1979), "Sample Selection Bias as a Specification Error," *Econometrica*, 47, 153-161.
- Heckman, J.J. (1990), "Varieties of Selection Bias," *American Economic Review*, 80, 313-318.
- Hensher, D.A., J.J. Louviere, and J.D. Swait (1999), "Combining Sources of Preference Data," *Journal of Econometrics*, 89, 197-221.
- Horowitz, J.L. and J.J. Louviere (1995), "What is the Role of Consideration Sets in Choice Modeling," *International Journal of Research in Marketing*, 12, 39-54.
- Howard, J. and J. Sheth (1969), *The Theory of Buyer Behavior* (New York: John Wiley).

- Kotler, P. (2003), *Marketing Management*, 11th ed. (Englewood Cliffs, NJ: Prentice Hall).
- Louviere, J.J., D.A. Hensher and J.D. Swait (2000), *Stated Choice Methods: Analysis and Application* (New York: Cambridge University Press).
- Maddala, G.S. (1987), *Limited Dependent and Qualitative Variables in Econometrics* (New York: Cambridge University Press).
- Marayama, C.L. and R.J. Markham (1975) "Consumer Behavior and Product Performance: An Alternative Conceptualization," *Journal of Marketing Research*, 39, 1-6.
- Reilly, M. and T.L. Parkinson (1985) "Individual and Product Correlates of Evoked Set Size for Consumer Package Goods," *Advances in Consumer Research*, 12, 492-497.
- Roberts, J.H. and J.M. Lattin (1991), Developing and Testing a Model of Consideration Composition," *Journal of Marketing Research*, 28, 429-440.
- Roberts, J.H. and J.M. Lattin (1997), "Consideration: Review of Research and Prospects for Future Insights," *Journal of Marketing Research*, 34, 406-410.
- Roberts, J.H. and P. Nedungadi (1995), "Studying Consideration in the Consumer Decision Process: Progress and Challenges," *International Journal of Research in Marketing*, 12, 3-7.
- Swait, J. (2001), "A Non-Compensatory Model of Incorporating Attitude Cutoffs," *Transportation Research Part B*, 35, 903-928.
- Wright, P. and F. Barbour (1977), "Phased Decision Strategies: Sequels to Infinite Screening," in *Multiple Criteria Decision Making*, M. Starr and M. Zeleny eds. (Amsterdam: North-Holland).