The Journal of Choice Modelling 13 (2014) 37–48

Contents lists available at ScienceDirect

The Journal of Choice Modelling

journal homepage: www.elsevier.com/locate/jocm

Statistical properties of consideration sets

Richard T. Carson [a,b,*], Jordan J. Louviere [a,c]

[a] The Institute for Choice, The University of South Australia, Australia
[b] Department of Economics, University of California, San Diego, United States
[c] The School of Marketing, The University of South Australia, Australia

A R T I C L E I N F O

Available online 19 February 2015

Jel classification:
D1
M31
C35

Keywords:
Multi-stage choice models
Sample selection bias

A B S T R A C T

The concept of a consideration set has become a central concept in the study of consumer behavior. This paper shows that the common practice of estimating models using only the set of alternatives deemed to be in the set considered by a consumer will usually result in estimated parameters that are biased due to a sample selection effect. This effect is generic to many consideration set models and can be large in practice. To overcome this problem, models of an antecedent volition process that defines consideration will effectively need to incorporate the selection mechanism used for inclusion of choice alternatives in the consideration set.

1. Introduction

The concept of a consideration set is well-established in marketing (e.g., Kotler, 2003; Howard and Sheth, 1969; Roberts and Nedungadi, 1995; Roberts and Lattin, 1997; Chiang et al., 1999; Erdem and Swait, 2004; Salisbury and Feinberg, 2012). A search (August 29, 2013) of Google returned over a hundred thousand hits for "consideraticon set", and assessment of the first 200 suggested that only a small fraction were not relevant. A search of Scholar Google returned over 8000 documents, and an assessment of the first 200 listings suggested that all these documents were relevant. Even taking duplicate and irrelevant documents into account, these searches clearly indicate that the concept of a consideration set is pervasive in marketing and related fields. This paper looks at statistical issues associated with the typical use of the consideration set concept to truncate the set of goods from which a consumer is assumed to choose. For applied researchers the main message of the paper is that the selection process that determines what goods are in the consideration set will almost always need to be successfully modeled in order to obtain consistent estimates of the choice process.

As a stylized example of the situation we address in this paper, consider a purchaser of a new good who is later surveyed and asked what alternatives were considered when making the purchase. The person provides a consideration set of alternatives that were the mostlikely to be chosen rather than all the alternatives explicitly or implicitly considered. Another stylized example occurs when a researcher has panel data on an individual's purchases in a particular category over time and the researcher forms the consideration set by including all products in the category that were ever purchased rather than the larger set of alternatives in the category which may well have been considered in some fashion. In both examples, the problem we identify and discuss in this paper occurs when consideration sets used by researchers differ from those actually used by individuals. This problem arises (inter alia) because consumers typically will not reveal (e.g., recall and report) alternatives having a relatively low probability of being chosen, in either revealed or stated preference contexts.

* Corresponding author at: The Institute for Choice, The University of South Australia, Australia.
E-mail address: rcarson@ucsd.edu (R.T. Carson).

http://dx.doi.org/10.1016/j.jocm.2015.01.001
1755-5345/

More generally, many research applications that focus on understanding and modeling consumer preferences often formally classify goods into two groups: 1) goods a consumer would consider purchasing and 2) goods a consumer would not consider purchasing (probabilistically or deterministically). Each good typically is represented as a bundle of attributes like price, size, color, quality and brand name. It also is common to ask consumers to answer survey questions about each good classified as being "in" a consideration set to elicit extra preference information. Such questions take many forms, such as: a) reporting any goods in the category ever purchased, b) directly identifying the consideration set (i.e., listing it or checking all that apply from a list), c) ranking or rating each "considered" good", d) asking which one of the "considered" goods will be chosen next or was chosen most recently or last, and/or e) many other similar possibilities (i.e., questions that try to identify a consumer's most preferred option – see, e.g., Narayana and Markin, 1975; Reilly and Parkinson, 1985; Brown and Wildt, 1992; Horowitz and Louviere, 1995). The assumed consideration set, together with measured attributes/features of alternatives within it and the choice(s) actually made often are used to estimate statistical choice models. Such a modeling strategy effectively assumes a higher order antecedent volition process leading to the identified consideration set that has no direct link to choice processes within the consideration set. In this paper we show that although it is common in commercial and academic research to use such questions to identify consideration sets to proxy a person's choice set and to estimate choice models conditional on these sets, using such consideration set measures raises statistical issues related to selection bias (Heckman, 1979; Vella, 1998) that can have substantive implications for the way in which model estimates are interpreted and applied.

The literature on consideration sets has evolved in several distinct directions:

1. In one dominant stream, consideration sets are seen as endogenous quantities to be estimated from consumer panel data and/or from consumer choice experiment data. Examples include a) Gensch (1987), Andrews and Srinivasan (1995) and Swait (2001a; 2001b), which is a small sample of researchers who proposed and estimated two stage models of consideration and choice; b) Roberts and Lattin (1991), who developed a model of how consumers form a consideration set at a particular point in time; and c) Chiang et al. (1999), who developed a model of consideration set formation and choice that allows for heterogeneity in the parameters of both processes. Thus, this research stream focuses on drivers of consideration and choice.
2. A second major research stream treats consideration sets as exogenous quantities defined by some type of direct measurement process. Examples include a) Narayana and Markin (1975), who classified brands into "inept" and "inert; b) Wright and Barbour (1977), who coined the term "consideration set" and suggested that brands "known" to consumers can be classified into acceptable and unacceptable; and c) Horowitz and Louviere (1995), who used aided and unaided recall questions to measure which brands were/were not considered.

Many variations on the above two research streams exist, such as Fotheringham's (1988) treatment of consideration sets as 'fuzzy' and Yee et al.'s (2007) practical computational way to look at many non-compensatory rules that could be used to form choice sets.[1] A common feature of most of these more technical papers on consideration sets is that they try to nest the standard neoclassical model of consumer choice as a special restricted case. Thus, a statistical specification issue that impacts the standard model is relevant to many more complex models.

The concept of a consideration set is valuable for modeling consumer choice in fields like marketing because it allows more flexibility in variables and degrees of influence that can occur at different points in decision making processes. This paper is agnostic with respect to whether the standard neoclassical model is correct or adequate. In particular, we leave that debate aside, and instead focus on a more basic question, namely what happens if a typical statistical analysis is performed using only a consideration subset of known alternatives, however defined, in the standard neoclassical model.[2] This is the most straightforward case. Yet, a motive for using consideration sets often seems to be the belief that parameters and/or variables that drive consideration fundamentally differ from those that drive choice. Hence, estimating a model that includes all available choices will lead to biased parameter estimates and misleading views of consumer behavior. We believe this may well be true in many situations of interest to researchers. However, we show that one cannot avoid sample selection issues by simply assuming that there is an antecedent volition process driving consideration that differs and is not linked to the one driving choice. In fact, selection problems are generally worse in this case.

Some researchers who used consideration sets tried to correct for selection bias (e.g., Ben-Akiva and Boccaro, 1995; Paap et al., 2005; von Haefen, 2008), but our review of the literature in marketing suggests that this is not typical of empirical practice. For example, in an early use of consideration sets, Krishnamurthi and Raj (1985) defined a household's consideration set as any brand in the category purchased during a prior 52 week period. Of course, households might have purchased other brands if (for example) there was a large price decrease in one or more brands not purchased in the previous year.[3] A well-known paper by Allenby and Ginter (1995) uses a heteroscedastic logit framework to flexibly fit price and promotional parameters, which nicely illustrates one way in which consideration sets are used.[4] They eliminated all

[1] See also the special issue of *International Journal of Research in Marketing* on consideration sets edited by Roberts and Nedungadi (1995).

[2] Including alternatives that an agent is unaware of in the set from which the agent is assumed to have chosen can also create substantial statistical problems not pursued in this paper.

[3] Hauser and Wernerfelt (1990) show this is a common way to define consideration sets.

[4] Baltas and Doyle (2001) note that a common practice in using scanner panel research is to drop smaller brands, and that "such sample selection

small brands comprising 13% of consumer purchases in the category before model analysis.[5] Punj and Brookes (2002) studied new economy car segment purchases, and measured consideration sets with a survey question asking respondents to name up to six cars that they "seriously considered buying other than the one they purchased". All of the preceding ways to define consideration sets truncate the set of alternatives available to consumers.

Besides such formal (operational) definitions and/or uses of consideration sets that can be associated with sample selection issues, informal uses of consideration sets also suggest sample selection issues may be common in academic and applied research. For example, potential respondents often are screened out of focus groups and surveys based on not having used or considered a limited set of brands in a category.[6] Likewise, marketing research analyzes of scanner panel and other types of purchase data often are performed only on a subset of brands assumed to be the ones most likely to be considered along with a key brand of interest. Implicit in many ad hoc ways to define consideration sets is the objective of reducing the number of goods that researchers and/or respondents must deal with to a more manageable number.

Thus, the purpose of this paper is to rigorously investigate and discuss the implications of using consideration sets in statistical model estimation. The contribution of the paper is twofold: 1) we discuss sample selection bias issues and their implications for statistical model estimation and inference that apply to many situations of interest to academics and practitioners in marketing and other fields (e.g., transportation, environmental economics), and 2) we explicitly discuss selection bias issues in the context of using consideration sets to define choice sets in choice model and related applications. Specifically, we consider two non-trivial cases: a) a single underlying process drives consideration and choice, and b) there is a two stage process, with each stage governed by a different process. We examine the implications of using consideration sets based on research results in the econometrics literature dealing with sample selection and truncation issues (e.g., Maddala, 1987; Vella, 1998; Heckman and Robb, 2000; Greene, 2003).

We begin our discussion by assuming the existence of some idealized utility index (Deaton and Muellbauer, 1980).[7] We next consider a case where one only has a simple discrete indicator of the most preferred alternative available. We rely on basic assumptions about translating consumer preferences into a statistical framework to illustrate the main points, and we note that these generalize to more complex and realistic models. To anticipate our results, we briefly summarize them before beginning a more rigorous treatment.

1. If one estimates statistical models from datasets that include only "considered" options, one will obtain biased estimates of preference parameters unless those models take explicit account of the influence of the selection process.
2. If one allows different processes to drive consideration and choice, in most realistic cases the situation becomes worse relative to a single underlying utility function, not better.
3. Even if one could obtain consistent estimates of the preference parameters, limiting model analyzes only to "considered" options produces biased estimates of confidence intervals for the preference parameters.
4. Even if one ignores issues related to the consistency of preference parameter estimates and associated confidence intervals, the information content associated with a randomly chosen option in a consideration set generally will be less than a randomly chosen option not in a consideration set.
5. Goodness-of-fit measures like McFadden's pseudo R^2 will tend to be lower (often much lower) because a statistical choice model will do a better job predicting which alternative was chosen out of a complete set of options than it will from a subset of options that are much closer in utility space.

The preceding five results are immediate consequences of the fact that using consideration set information to measure consumer choice sets is a type of truncation of the dependent variable. All of these results either are direct consequences of the underlying statistical mechanism at work or can be readily inferred from it. In this sense, we are not breaking any new ground. More specifically, the statistical issues that arise from using consideration set information to estimate choice models are equivalent to issues long studied by labor economists concerned with sample selection bias (e.g., Heckman, 1979; 1990). Indeed, we were surprised to find that sample selection bias seems to have received relatively less attention in marketing

(*footnote continued*)
practices" are inappropriate when a complete picture of the market is required.

[5] Of course, one can combine omitting smaller brands with omitting alternatives that are never or rarely purchased. For an example, see Siddarth et al. (1995).

[6] We did an informal survey of firms that collect primary marketing research data in a major metropolitan area and found that all employed this type of screening and routinely used it for clients.

[7] For instance, this index could be a willingness to pay (WTP) measure, which has the advantage of being continuous, having a true "zero" point and a scale (dollars) interpreted the same way by all consumers. However, it is difficult to elicit or observe WTP directly, so discrete indicators of WTP (e.g., whether a good was purchased or not at a particular price) often are obtained instead. Such indicators can be mapped back to WTP measures and marginal WTP with respect to an attribute of the good; this often is termed a part-worth, and is a common quantity of interest in applied marketing research. Its drawback is that it allows for the possibility of negative WTP due to the normal error component, a specification that makes sense in some contexts, but not others where WTP should be non-negative. In this case, it is possible to allow for a spike at zero (Kriström, 1997) which can be modeled as a separate process (Werner, 1999). Cameron and James (1987), Train and Weeks (2005) and Sonnier et al. (2007) show how choice models can be transformed from preference space to WTP space.

than many other areas of applied microeconomics, biomedical research (e.g., Hernán et al., 2004 and sociology (Winship and Mare, 1992) despite its ubiquitous nature.[8]

2. Continuous case

We begin with the continuous dependent variable case because it is easier to see the issues and helps make our major point readily transparent. That is, we assume, for simplicity and transparency, that utility for the j-th good can be represented by the linear relationship:

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

where there are $i=1, 2, \ldots, n$ individuals, X_j is a vector of attributes of the jth good, β is a vector of k preference parameters (i.e., part-worth utilities) 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 critical to this discussion; one can make assumptions about different underlying processes that give rise to it, ranging from unobserved individual characteristics to 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 $j=1, \ldots, k$ goods, and the correlation between error terms for different goods is zero.

Although researchers use different ways to measure consideration sets, in one way or another almost all ask respondents to divide a set of goods into two subsets and report an indicator of U_{ij} for the subset actively considered. A key insight in our discussion of this issue is obtained by noting that a good will not be contained in the consideration set if:

$$\beta X_j + e_{ij} \leq C_{ij}, \tag{1}$$

where C_{ij} is referred to as a censoring or cutpoint. Frequently C_{ij} is assumed to be zero for all consumers as choices of goods for which a consumer has negative utility usually are unobserved; however, it may be more appropriate in many cases to have C_{ij} vary with the good and/or consumer. Thus, a good can fail to enter a consideration set for three reasons: 1) an individual perceives the X_j vector to be undesirable, 2) the value of ε_{ij} is sufficiently small, and/or, more generally, 3) a linear combination of the two preceding reasons.

Following Greene (2003, pp. 780–790), we first define certain functions of a truncated normal distribution in terms of the quantity: $\alpha_{ij}=(C_{ij}-\beta X_{ij}/\sigma)$:

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 readily can be shown that:

$$E(U_{ij}|U_{ij} > C_{ij}) = \beta X_{ij} + \sigma\lambda(\alpha_{ij}), \tag{2}$$

using the formula for a truncated normal and that

$$\partial E(U_{ij}|U_{ij} > C_{ij})/\partial X_{ijk} = \beta_{jk}[1 - \delta(\alpha_{ij})],$$

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

One also can show that $\mathrm{VAR}(U_{ij} \mid U_{ij}>C_{ij})=\sigma^2[1-\delta(\alpha_{ij})]$ will be biased toward zero, which in turn biases confidence

[8] Clearly there are exceptions, such as Thomas (2001). Yet, in so far as random utility choice models and sample selection effects are two major themes in microeconometrics (for which Dan McFadden and James Heckman, respectively, won the Nobel Prize in 2000), although McFadden's contributions have had a major influence on the marketing research literature, Heckman's influence has been much smaller. But, citation counts from the business and economics literature as a whole clearly show that Heckman's work has been equally, if not more, influential. Specifically, we did not find a cite to Heckman's classic 1979 paper on selection bias in any key marketing papers on consideration sets except for Degeratu et al. (2000); but their treatment of selection bias involved their respondent sample, not their choice sets. Other marketing papers exploring aspects of selection effects include Winer (1983) who looks at selection bias due to participants dropping out of panels; Krishnamurthi and Raj (1988) on the need to model the discrete choice of whether to purchase a product is linked to the continuous choice of how much to purchase; Moe and Schweidel (2012) who look at section effects in terms of how prior posts influences who posts on product review boards, and Salisbury et al. (2013) who show selection effects can occur in market experiments even when random assignment at the initial stage is possible. Many marketing processes involve multiple stages with selection likely to occur at each stage (Wachtel and Otter, 2013).

intervals downward. Further, although a constant term in an OLS regression for positive U_{ij} recenters 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 ε_{ij} are homoscedastic. One can obtain consistent estimates of β using several different techniques if all the C_{ij} are known (Greene, 2003; Maddala, 1987).

One also can show that statistical procedures that use information (i.e., X's) about goods that do not enter consideration sets (i.e., they are censored) are more efficient than procedures that only use observations for which inclusion in the consideration set is observed. Several factors underlie gains in efficiency from using information about numbers of goods not entering consideration sets. To begin, consider a case where there are no attributes of the goods. Here the number of goods not entering consideration sets helps to define properties of the error distribution. Now consider a case where goods have attributes. Recall that for this case, $\beta X_j+\varepsilon_{ij}\le C_{ij}$, or equivalently, $\beta X_j\le C_{ij}-\varepsilon_{ij}$. As the ε_{ij} are random normal variables, the X_j that fulfill this condition will differ from the X_j included in consideration sets. Thus, exclusion of the latter X_j reduces variation in the design (estimation) matrix, and so increases confidence intervals for the preference parameter estimates. It is important to note that this effect works in the opposite direction of the bias in the estimate of the variance term previously noted; consequently, in any empirical application, directional bias in confidence intervals is unknown.

If consumers have the same preferences for some attribute values, which is implied if one estimates a single β vector for all consumers, or if the preferences of most consumers fall in a reasonably narrow range (which is empirically implied by results from many random coefficients models), the X_j associated with the consideration set will be "similar" in many respects, further reducing variability in the design matrix.[9] In this case replacing a randomly chosen good in the consideration set with one not in that set almost always will improve the precision of preference parameter estimates.

Another intuitive result is that commonly-used summary goodness-of-fit statistics like McFadden's (1974) pseudo R^2 will tend to be biased downward. That is, such measures reflect how well a fully parameterized model fits data relative to a "null" model. A fully parameterized model almost always will be better at predicting that options outside consideration sets have lower probabilities of being chosen relative to options in consideration sets. This occurs because the "outside" options are in a less desirable part of the X space, while particular options chosen from a (consideration) set of options tend to be in more desirable parts of the X space. The more successful consideration sets are at producing small sets of options close in utility space, the more pronounced will be the decrease in apparent explanatory power of models.[10] This may not be of concern if one is only interested in how well a model discriminates between options in the consideration set, but it will mislead analysts if their aim is to assess how well a model captures the underlying parameters of a utility function.

The above discussion of consideration sets closely parallels the econometrics literature on issues related to the effects of selecting samples based on the value of the dependent variable. All the issues noted above have been shown to be potentially correctable, but it may be difficult to satisfy the assumption that all C_{ij} are known to analysts but exogenous from the ith consumer's perspective. The simplest variants of sample selection models assume C_{ij} is a known constant for all individuals, which typically is zero. This might make sense for the case of consideration set questions that ask for a set of goods for which consumers have positive willingness to pay, but this is not a common way to measure consideration sets. If C_{ij} is unknown but is the same for all individuals, or it is the same for exogenously identifiable groups of individuals, or if C_{ij} only varies with j (e.g., C_{ij}=the lowest price of the jth good), there are ways to consistently estimate the preference parameters (e.g., see Carson and Sun, 2007). If C_{ij} is unknown and varies across individuals, the key quantity previously described for correcting sample selection bias, $\alpha_{ij}=(C_{ij}-\beta X_{ij}/\sigma)$, is unidentified because neither α_{ij} or C_{ij} are observed. In this case, each individual's C_{ij} is absorbed into a) the estimate of the vector of preference parameters, β, giving the appearance of a distribution of preference parameters even if all individuals actually share the same vector of β's, and b) the estimate of σ, introducing heteroscedasticity where none previously existed. In turn, the heteroscedasticity biases estimates of β (see Hurd, 1979). Ioannatos (1995) gives a way to estimate the distribution of individual cutpoints, and to integrate this distribution out to obtain consistent estimates of slope parameters of interest; this approach requires strong normalizing assumptions on variance/covariance parameters.

Now we turn our attention away from difficulties due to individual unobserved cutpoints, and instead focus on whether selection problems disappear if there is a two stage process where a first stage determines which alternatives are in consideration sets and a second stage determines choice within consideration sets. As before, we also assume that one can

[9] The following example should help to illustrate this issue. Let our neoclassical consumer with substantial prior knowledge of available options, and positive but low search cost, eliminate from the set of possible options to be more closely examined any option with a price greater than some amount; knowing that options with a price above this amount are highly unlikely to be her most preferred option. Further, let our consumer accurately report her consideration set. Estimation of the price coefficient within the consideration set may well be insignificant due to the limited range or price inside the consideration set even though price is perhaps the dominant factor in the purchase decision. Estimating a choice model with the full set of options would clearly reveal the price effect.

[10] As an example, consider someone who needs to purchase a new oven for their single family home who is in an appliance store. After wandering around the store, they ask the salesman to tell them about four ovens. None of the ovens in this identified consideration set are high-end ovens for aspiring chefs with very desirable attribute levels and high prices nor are there any ovens with low levels of most attributes and low prices aimed primarily at apartment and condo dwellers. The person has implicitly considered both of these feasible alternatives. The presumed consideration set, however, now spans a much narrower range of prices and has a substantially truncated range of attribute levels. While a narrowing of the range of attributes levels in the consideration set is a reasonably general result, doing so in terms of it having more desirable attributes levels is of course subject to interpretation. For instance, for those for place heavy weight on price this may mean removing goods with the most desirable levels on other attributes from the consideration set, and, if an attribute is related to the time of purchase, the outside good will of necessity occupy most of the temporal attribute's space.

represent the second stage with a continuous variable indicator of utility like WTP_{ij}, and a discrete indicator is observed for whether a good is in a consideration set. There is a large econometrics literature for this case, which deals with what are called hurdle model problems (Mullahy, 1986). If one process drives both the generation of the dependent variable and the selection mechanism, the appropriate mechanism is a single hurdle model. If a different process drives the data generating process for the dependent variable and the selection mechanism, the appropriate mechanism is a double hurdle model. An early example of a double hurdle process is due to Cragg (1971), who used a probit model to determine if there was a positive expenditure, and a truncated normal to model positive expenditures. Cragg's model nests the single hurdle Tobit model as a special case. Heckman's selection model (Heckman 1976, 1979) can be viewed as a generalization of Cragg's model whereby both processes are linked via their error terms. Double hurdle models were later developed for two step models in which several different data generating processes can drive each step (e.g., Blundell and Smith, 1994; Smith, 2002; Greene, 2003). It is important to note that most consideration set models that posit a two step process fall under this framework.

Thus, it is important to know whether a separate process that governs selection into a consideration set can help the analyst. Consider the simplest case of having to predict whether an individual good is in or out of a choice set. Assume an underlying latent variable $S_{ij}=f(Z_{ij}, \theta)+\mu_{ij}$, $f(\cdot)$ an arbitrary function of a set of observed variables, Z, typically has some overlap with X, θ a vector of parameters that can be estimated in principle, and μ_{ij} a normal mean zero error term. The indicator variable, I_{ij}, for S_{ij}, takes on the value of 1 (good j is in the consideration set) if it is larger than a constant C, and equals 0 otherwise. This framework is sufficiently general to incorporate a first stage in which (for instance) consumers may use non-compensatory rules that differ from those used to make a choice from within a consideration set. Unfortunately, this case offers little help from a sample selection bias perspective because selection on C remains. Indeed, following Greene (2003, p. 783), it is possible to rewrite the two parts of equation (2) with this new selection equation involving I_i and show that the expected value of β depends on the role of X in the selection equation for I_i and the ratio of the variances of ε and μ times ρ the correlation between the two error terms. In particular, the expected value of β_{jk} estimated from only choices in the consideration set is given by:

$$\partial E\left(U_{ij}|I_{ij}=1\right)/\partial X_{ijk} = b_{jk} - \partial f(\bullet)/\partial X_{ijk}[\rho\sigma_{\varepsilon}/\sigma_{\mu}]\delta_i(a_u), \tag{3}$$

where δ_i is equal to $\lambda_i^2-\alpha_i\lambda_i$ and α_μ is equal to $f(\cdot)/\sigma_\mu$.[11] Thus, the estimate of β_{jk} is equal to itself plus a fairly complex term reflecting the influence of X_{jk} on the equation defining consideration sets and the interaction between the error terms of the two equations.[12]

It is useful to examine (3) more carefully to see whether special cases exist in which no bias results from using a two stage procedure that assumes the two stages are not linked. There is one obvious case. If consumers use a single binary attribute Z_s to place goods in a consideration set so that all goods with one level of the attribute are in, and all goods without that level are out, and this process has no error, then four special (parametric) special cases result: 1) The Z_s parameter is not identified in models estimated only using alternatives in a consideration set because all alternatives share this attribute value, $\partial f(\cdot)/\partial Z_{ijs}$, or its more general non-pointwise differential equivalent equals zero, the correlation coefficient of the error terms in the two equations equals zero, and σ_μ equals zero. Technically, this case is indeterminant because setting σ_μ equal to zero involves division of a positive number by zero. However, allowing a very small amount of random error is enough to give consistency in the estimates of the identified β_k. 2) Another special case is ρ equal to zero, which is not surprising as this requires no selection on the idiosyncratic error component in the choice model. The possibility that ρ is zero is testable; yet, while such "probabilistic independence" might hold in a particular empirical application, this seems to be an implausible assumption on which to base a reliable consideration set procedure.[13] 3) A more likely possibility is that $\partial f(\bullet)/\partial X_k$ equals zero for a particular X_k, which would occur for example if an attribute influences consideration but not choice. This also is testable, with two strong caveats, namely that this condition must hold for all attributes of interest in the choice equation when examining marginal tradeoffs and for all attributes in the choice equation when calculating summary statistics and the correct specification must be known for both stages of the process. 4) There also is no bias if there is no error in the choice model, but this case is unlikely to be of empirical interest as perfect predictions of choice are already possible without consideration of any possible selection rule. Going in the opposite and perhaps more interesting direction, as the predictive power of the model governing selection into the consideration set becomes increasingly small relative to the predictive power of the choice model, selection bias disappears. Effectively what happens in this case is that the particular alternatives in the consideration set become randomly chosen.

[11] Here we conveniently assumed that $f(\cdot)$ is differentiable with respect to X at the point of evaluation. This may not be true, particularly for some types of non-compensatory functions. In such cases, it may be possible to achieve only set-wise rather than pointwise identification.

[12] It is possible to set out plausible cases in the model defined by equation (3) where the estimate of β_{jk}, without the sample selection correction, is arbitrarily far from its true value. To see this, consider the case where a consumer tends to eliminate alternatives from consideration that have high prices but tends to pick the higher priced alternatives in the consideration set. Let the correlation between error terms in the two equations have the expected positive sign, and let the standard deviation of the error component in the choice equation be large relative to the standard deviation of the error component in the consideration equation. In this case, the estimate of β_{jk} is likely to be negative and quite large in magnitude; this likely will conform to a researcher's prior while the true value of β_{jk} is small and positive conditional on the consideration set selection mechanism.

[13] We found no direct empirical tests of this proposition in the consideration set literature. Gensch (1987) recognized that both stages are linked but did not write down the complete likelihood for his model; hence, he failed to see the full nature of the linkage.

A more general framework that allows for a sequence of levels for different attributes to influence inclusion in the consideration set where there is a (modest) stochastic component for choosing the order of attributes levels used to eliminate the alternatives does not provide a clean separation. This is not surprising due to the equivalence of this model and certain representations of generalized extreme value models (Batley and Daly, 2006). The correct statistical approach for a two stage process models the sequential nature of the first stage and a neo-classical RUM-based second stage choice model builds on the simultaneous equation framework of Manski and McFadden (1981).[14]

It is ironic that consideration set models often are driven by a belief that $\beta_{jk} \neq \partial f(\bullet)/\partial X_{jk}$, so that estimating a single model that forces the two parameters to be equal will bias the estimate of β_{jk}. Even if this belief is indeed correct,[15] a typical approach that estimates only the model on the consideration set or two unrelated equations that predict consideration and choice will not yield an unbiased estimate of β_{jk}, as equation (3) clearly shows, except in the highly unlikely circumstances noted above. An unbiased estimate requires one to correctly take into account correlations in error terms from the equation driving the consideration set and the equation driving choice, as well as the impact of X_{jk} on both equations. While this makes an analyst's job harder, modeling the system as a whole, and particularly the connection between the two equations, may well be where most of the interesting action occurs from a marketing or policy perspective.[16] The case where the entire set of alternatives is relevant and the parameters governing consideration and choice are the same is the benchmark that alternative models of consideration and choice represent different processes need to beat in order to be useful to decision makers.

3. Discrete case

The discrete case is largely analogous to the continuous case because bias in the estimated parameters of the latent variable and its summary statistics carry over to estimation using only an indicator variable of U_{ij}. However, there are a few important differences that can be seen by noting the usual representation of the utility of the jth good in a random utility model (RUM):

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

where in simple specifications 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. Different distributional assumptions such as the normal or extreme value, lead to different statistical models. We note that an i.i.d. assumption effectively is the same as the continuous case, and a strong link between both models is well-known in the literature (e.g., Cameron and James, 1987).[17] Thus, it is straightforward to show that all issues in the continuous case carry over to the discrete case, although three noteworthy differences arise.

The main difference in the continuous and discrete choice cases is that instead of directly estimating the β parameters, one estimates β/s where s is a scale factor. Loss of information on scale is a consequence of discrete information about preferences, which in turn raises the possibility that truncation bias might "cancel out" under the common practice of looking at ratios of estimated attribute parameters to investigate marginal tradeoffs. This issue deserves further exploration as there is an existing result for Tobit models that shows that all parameter estimates are biased by a constant factor under certain conditions (Cheung and Goldberger, 1984). This condition requires a normal error term and a multivariate normal X matrix. While a multivariate normal X matrix is unlikely in practice, in many applications design matrices are reasonably symmetric. For a simple linear model and standard i.i.d. logistic or normal errors Cheung and Goldberg's constant proportionate bias result may sometimes approximately hold in practice, and Ruud (1986) shows that there are more general conditions under which it will hold. This is important because a model in which all parameters are biased by a constant factor will predict the same choice behavior as a true model, but there will still be adverse impacts on confidence intervals and summary statistics.

Another difference between continuous and discrete choice cases is that instead of a cutpoint for inclusion in consideration sets being expressed in terms of the utility index, the inclusion cutpoint is expressed in probability terms, such that options with a sufficiently high probability of being chosen are included in consideration sets. Again, it is noteworthy that the key theoretical condition that drives the result in the continuous case still holds. That is, a low probability will be indicated for goods with undesirable X_j and/or small ε_{ij}. The issue of individual C_i's is perhaps even clearer here as different

[14] Considerable more flexibility with respect to needing to observe actual choice sets is possible, if one makes the assumption that an individual's utility from an alternative has a true random component (Dagsvik, 1994).

[15] It is always useful to allow the possibility that the two parameter vectors are statistically equivalent.

[16] Natural starting points that borrow from the econometrics literature on labor supply issues (e.g., Lee, 1995) can be used to test various motivations (e.g., Hauser and Wernerfelt, 1990) previously proposed to underlie a consumer giving differential attention to various alternatives. Simulation-based approaches including modern Bayesian methods can make estimating some of the desired statistics much more feasible.

[17] Further, the notion of a consideration set strongly resembles the structure of a nested logit model, a primary tool used to relax the i.i.d. assumption. Here Cardell (1997) cautioned against a 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 even to small amounts of heterogeneity in consumer preferences, although all the parameters, including the inclusive values linking the different levels, are technically identified and consistently estimated in large enough samples.

individuals making probabilistic cutoffs for inclusion in consideration sets in different ways will introduce a form of heteroscedasticity even if there was none to begin with.

It also should be noted that Horowitz and Louviere (1995) showed that one can consistently estimate parameters of a RUM model by using information on which vectors of X_j are in a consumer's consideration set and which are not in cases where there is one underlying utility function. In particular, each good in a consideration set is known to be preferred to each good not in a consideration set; hence, the latter defines a potentially large set of preference relations among different goods. Of course, this is exactly what one would expect given the underlying structure of a RUM model. Any information that one alternative is preferred to another provides useful information to estimate the underlying preference parameters. This approach differs from those that use consideration sets because no sample selection issues apply. However, the approach is inefficient as it does not use all of the available data (i.e., the actual choice made). If different processes drive consideration and choice, the Horowitz and Louviere approach potentially can be used to obtain consistent estimates of the parameters of the consideration equation as input to an augmented choice equation.

4. Illustrative Monte Carlo simulation

Now we consider how large the sample selection effects that we discussed are likely to be in practice, providing Monte Carlo evidence on this issue.[18] To help sharpen the focus of our Monte Carlo experiments we concentrate on only two key factors: 1) the nature of the underlying utility function (linear and quadratic), and 2) the approximate number of options included in a consideration set (4, 5, 6, or 7). Most other aspects of the experiments are held constant: 1) there are always 200 consumers, 2) these consumers can choose from 25 options,[19] 3) the same orthogonal main effects matrix (design) for X, with three factors, each of which takes on five levels -2, -1, 0, 1, and 2, is used to construct utility functions in all experiments, 4) all models estimated are simple conditional logit models, 5) underlying utility functions, including error components, are consistent with that specification, and 6) all experiments have 1000 replications. We recorded parameter estimates, estimates of standard errors for those parameters and pseudo-R^2 measures of goodness-of-fit associated with this experiment.

Let us first consider a simple linear specification

$$U_{ij} = .2X_{1ij} + .1X_{2ij} - .3X_{3ij} + \varepsilon_{ij},$$

where $i=1, 2, \ldots, 200$, $j=1, 2, \ldots, 25$, and ε_{ij} is an i.i.d. extreme value random variable with a standard deviation of 1, which effectively sets the scale of the correctly specified model. The reference case uses all 25 options. The other four cases use different cutpoints relative to U_{ij} to define consideration sets. Consideration sets with approximately 7, 6, 5 and 4 options were defined by cutpoints of $Y>.25$,.40,.55, and .70, respectively. Thus, different individuals have different numbers of options in their consideration sets.[20] It is important to note that the sample selection mechanism in this case is much simpler than Heckman's (1979). Specifically, since our cutpoints are deterministic, there is no correlation between error components in the selection and choice models; and as there are no predictor variables in the selection model, there cannot

Table 1
Linear specification.

Parameter	Mean with all options	Mean $Y>.25$	Mean $Y>.40$	Mean $Y>.55$	Mean $Y>.70$
$\beta1$	.5226	.3942	.3695	.3438	.3191
$\beta2$	.3653	.2308	.2133	.1951	.1783
$\beta3$	−.8487	−.5802	−.5436	−.5056	−.4694
$\sigma1(\beta1)$	.0707	.0736	.0751	.0776	.0820
$\sigma2(\beta2)$	.0695	.0711	.0723	.0745	.0787
$\sigma3(\beta3)$	.0790	.0856	.0879	.0915	0973
$\rho2$	.2210	.1161	.1033	.0901	.0778
Average number of options	25	7.21	6.10	5.08	4.12

[18] For convenience the results obtained here are from a Monte Carlo experiment. As shown earlier the direction of the observed effects can be derived analytically and, in many cases, the (expected) magnitude of the effect can also be determined analytically.

[19] This leads to consideration sets with 16-28% of the original choices. Seven alternatives is a fairly large consideration set relative to what often is in the literature. Below an average of four alternatives in the consideration set, we occasionally got consideration sets comprised of only one alternative.

[20] It is possible to define choice sets in such a way that each individual has the same number of choices, but this implicitly requires different cutpoints for each individual, which in turn induces a form of heteroscedasticity. This case is also more favorable to finding sample selection effects than the one we examine, and is particularly so when the individual cutpoints have to be estimated (Nelson, 1977).

be shared regressors or even correlation between regressors in the two models. It is well known (Heckman, 1979; 1990) that as soon as these assumptions are relaxed it is possible to get much more extreme results including sign reversals on the parameter estimates of the choice model.

Table 1 presents the results of the Monte Carlo simulations. If we take the first column, (the simple logit model using all alternatives) as the best that can be achieved, examination of each row for the three parameters in the experiment reveals that each consideration set model formed by our different thresholds are biased, and the bias increases as the number of options in a consideration set decreases. The standard errors of the three parameters increase in size as the consideration set size decreases. Together with the shrinking of parameters toward zero, this implies that researchers are less likely to infer that parameter estimates are significantly different from zero.[21] Pseudo-R^2 declines by almost half for cases with about 7 options in consideration sets, and declines another 33% for consideration sets with around 4 options.[22]

The parameter estimates in Table 1 for the case involving systematically selected subsets of alternatives are badly biased, and increasingly so as consideration sets get smaller. Yet, the ratio of any two parameter estimates from any columns yields approximately the same estimate. While not quite the exact result of the Cheung and Goldberger (1984) multivariate normal design matrix for the Tobit model case where parameters are biased by the same proportionate amount, it is reasonably close. In turn, this implies that if the true data generating process is a simple linear specification, the estimated parameter vector conditional on a restricted consideration set is likely to predict an individual's choices reasonably well. The orthogonal design matrix helps with proportionate shrinkage of all coefficients toward zero, but this result in limited exploration seems fairly robust to alternative design matrices. We conjecture that this is a key reason why statistical problems with consideration sets have not been noted earlier. In this case, the main damage is done to the statistical fit of the model with respect to standard errors and overall goodness of fit statistics.[23]

Now we change the specification to a simple quadratic, increase the standard deviation of the error term to 2 because we increased the systematic component of the model (which will reduce all pseudo R^2's relative to Table 1) and use the same design matrix

$$U_{ij} = .2x1 + .1x2 - .1x3 + .1x1sq + .2x2sq - .3x3sq + \varepsilon_{ij}.$$

Table 2
Quadratic specification.

Parameter	Mean with all options	Mean $Y > .5$	Mean $Y > .75$	Mean $Y > 1$	Mean $Y > 1.25$
$\beta1$	.2738	.1846	.1746	.1644	.1535
$\beta2$	.1012	.0809	.0784	.0751	.0727
$\beta3$	−.1791	−.1121	−.1042	−.0951	−.0865
$\beta4$	.1427	.0912	.0857	.0796	.0730
$\beta5$	.2810	.1856	.1748	.1633	.1521
$\beta6$	−.4430	−.2909	−.2737	−.2563	−.2390
$\sigma1$	.0614	.0643	.0654	.0674	.0706
$\sigma2$	.0500	.0532	.0543	.0560	.0586
$\sigma3$	.0812	.0870	.0888	.0917	.0966
$\sigma4$	.0529	.0575	.0589	.0609	.0641
$\sigma5$	.0520	.0565	.0579	.0599	.0630
$\sigma6$	.0623	.0670	.0685	.0709	.0747
$\rho2$	.1378	.0866	.0814	.0764	.0717
Average number of options	25	7.07	6.06	5.09	4.18

[21] Three factors drive down the estimated t-statistics for the regressors in our simple choice model: (a) reduction in the number of implicit binary comparisons between the selected alternative and other alternatives, (b) bias toward zero of the estimated parameters on our choice model regressors, and (c) shrinkage in variability that the cutpoints induce on the attribute levels which feeds through the elements of the variance-covariance matrix. While (b) and (c) will occur in many, if not most cases, in contrast to (a), these effects are dependent on the exact nature of the selection mechanism. In Table 1, the dominant factor is (b), and while this likely often will be the case, this is not a general result. By construction in the correct model, estimates of the off-diagonal elements of the variance–covariance matrix should be close to zero: and this result did not appreciably change as we shrank the number of alternatives in the estimated model. This result is unlikely to carry over to situations where regressors are correlated.

[22] We focus on pseudo R^2, which is a measure of in-sample goodness-of-fit, but other measures such as those that emphasize out-of-sample misclassification rates may be of more interest in empirical applications; such measures also tend to show substantial declines with degrees of truncation. It is important to note that models that do not have an appropriate correction for sample selection bias may predict quite well in hold-out samples, if the (implicit) sample selection criteria do not change in the hold-out sample, and the model predicts badly in new situations.

[23] As an aside, note that as the ratios of estimated coefficients are similar across different datasets in Table 1, a conditional logit model that simply stacked the different datasets and allowed each to have a different scale factor would provide a reasonable fit. This suggests that one possible reason for different data sources yielding different estimates may lie in how the choices sets used for empirical estimation were defined.

This produces the same issues as consideration sets decrease (see Table 2): parameter estimates are biased increasingly toward zero, standard errors increase in size and pseudo-R^2 measures fall. Moreover, these results quickly reverse the linear case result that ratios of coefficients are approximately invariant with consideration set size. For example, the ratio of the $x1$ to $x2$ parameters, $\beta 1/\beta 2$, is 2.71 using all the options, 2.28 for consideration sets with (approximately) 7 options, 2.22 for 6 options sets, 2.19 for 5 option sets and 2.11 for sets with about 4 options. The ratio of the $x2/x3$ coefficients, $\beta 2/\beta 3$, is now $-.57$ for all options and becomes more negative: $-.72$, $-.75$, $-.79$, and $-.84$, as the consideration set becomes progressively smaller. Obviously, models estimated from consideration sets will predict very different outcomes than the same (correct) quadratic specification estimated from a full set of options. This occurs because the sample selection involved in consideration sets progressively reduces the part of the design matrix that supports one side of the quadratic, making it harder to estimate. It is further confounded by selection on the error terms, and the two effects interact to produce non-proportionate bias.[24] Many other common non-linear utility functions also will have the sort of non-proportionate bias in parameter estimates as a quadratic specification. Unfortunately with small consideration sets, it may be hard to support richer specifications, and local linearity will often be a reasonable approximation in many cases if the locale shrinks enough. Hence, finding that a linear model fits reasonably well for a particular definition of a consideration set should not be taken as providing strong evidence that a richer utility function does not underlie consumers' choices.

The range of potentially interesting Monte Carlo simulations that could be done using variants of the two stage model represented by Equation (3) is very large. Because equation (3) adds the possibility of selection directly related to the correlation of the error components and/or the relationship between regressors in the selection and choice models, the basic insights from Tables 1 and 2 largely carry over. This should be expected as the more general mechanism that defines what is in the consideration set has replaced the simple truncation point. The bias tends to be smaller (larger) as the correlation between error terms decreases (increases), the influence of X_k on the selection mechanism decreases (increases), and the magnitude of the standard deviation of errors in the selection equation increases (decreases) relative to that from the choice model. This last factor can have a large impact on bias in cases where membership in consideration sets is quite sharp and choice from consideration sets quite diffuse. Thus a weak consideration set model may result in less bias in the choice parameters, although empirically one might expect to see the correlation coefficient increase making the role of idiosyncratic components in both equations more important.

We have identified some situations where the effects of naively using consideration sets are less likely to be harmful and some situations where the effects are more likely to be large and more pernicious. Clearly there is much more that could be done to expand the dimensionality of our Monte Carlo experiment and we encourage other researchers to do so. We believe that the most fruitful direction for such work is to gather a set of antecedent volition processes that defines consideration, which have empirical support in particular contexts and then to look at the implications of failing to couple this process with that involving choice within the consideration set.

5. Discussion and conclusions

Our theoretical and Monte Carlo results suggest that failure to use information on alternatives not in consideration sets leads to a form of truncation of the dependent variable, which in turn produces sample selection bias. Sample selection leads to biased parameter estimates and confidence intervals, and reduces apparent goodness-of-fit. In most cases, the problem is self-inflicted, caused by eliminating alternatives by means of some mechanism that classifies them as not having been considered. The solution is straightforward – the antecedent volition process leading to the consideration set needs to be modeled in such a way as to include the selection effect so that consistent coefficient estimates from the choice model are obtained.

If the reason for using a consideration set is simply to reduce the number of possible choice alternatives considered by a survey respondent or to be analyzed by researchers, a better approach is to construct choice sets by including the alternative chosen by an individual plus a randomly chosen subset of other available alternatives (known as the "sampling of alternatives approach": McFadden, 1978; Parsons and Kealy, 2002). If one uses this approach, selection does not occur with respect to the magnitude of the error term or covariates in simple choice models. Stratified random sampling may be a reasonable way to deal with cases where a few brands comprise most of the market share but a sizeable number of brands each have small shares.[25]

However, an implicit or explicit assumption underlies the way that many researchers use consideration sets, namely that choices of alternatives in and out of consideration sets somehow differ. This reasonable assumption has motivated much

[24] Interestingly, the non-proportionate part of the bias is much more pronounced in the linear part of the model than the quadratic part, but that result is specific to the particular utility function we used where the quadratic component dominates the linear. Models with large interactions also can exhibit non-proportionate biases in parameter estimates. To fully examine this issue requires a design matrix that supports such a specification instead of the orthogonal main effects design that we used in our simulations.

[25] The assumption made here is that the data were generated by a multinomial logit model. Fox (2007) provides a semi-parametric form that is consistent under more general conditions but still requires an exchangeability assumption which is problematic with more general functional forms such as the random parameters logit model. Another promising way to deal with a large number of goods consistent with a consideration set concept is due to von Haefen (2008).

academic literature and is a potentially testable hypothesis. Nonetheless, separating the problem of determining what alternatives are in consideration sets from the problem of choice from alternatives in such sets introduces a new problem that few researchers have adequately dealt with. To wit, parameter estimates in choice equations will be biased unless one explicitly accounts for the sample selection issue. The size of this bias generally depends on the relationship between the idiosyncratic error components and the size of the impacts that variables driving consideration have on the choice part of the model.

A hypothesis that alternatives that are in or out of consideration sets differ in some way is difficult to test empirically without a structural model that formally nests the possibility that all options enter into an individual's choice problem in the same way. Thus, simply finding that parameter estimates differ in two independently estimated equations in a two stage model of consideration and choice is insufficient to conclude that two separate processes operate. Indeed, this statistical result is expected even if there is only one underlying process, as a selection effect will bias parameter estimates in the choice model which drops a subset of possible alternatives. One must control for influence of selection effects at the choice stage in order to make an informed judgment about whether processes driving consideration and choice differ. If selection into the consideration set is important, researchers will need to move to richer classes of models that formally link both stages.[26] The situations where jointly modeling consideration and choice result in large gains in predictability may well be ones where activation of the consideration process in different contexts results in the consumer choosing between very different sets of alternatives.[27]

In summary, the primary problem with mechanisms that seek to classify alternatives into more or less likely to be chosen is that they must at least partly operate on the basis of the magnitude of the utility for the alternative, and hence, its random component. Any consideration set classification mechanism will also tend to narrow ranges of attribute levels included in analyzes.[28] This, in turn, makes it harder to statistically identify true, underlying functional forms. It is important to note that this is not a difficulty with the consideration set concept per se but rather with typical empirical implementations. Correctly accounting for correlations between error components in a consideration set-choice framework and the induced selection effect will no doubt make estimation substantially more difficult than taking the ad hoc short cut of estimating a choice model using only alternatives thought to be the consideration set. Ultimately, however, models of the antecedent volition process that couple consideration and choice in a statistically appropriate manner are likely to help provide a much richer picture of consumer decision making processes.

References

Andrews, R.L., Srinivasan, T.C., 1995. Studying consideration effects in empirical choice models using scanner panel data. J. Mark. Res. 32, 30–41.
Allenby, C.M., Ginter, J.L., 1995. The effects of in-store displays and feature advertising on consideration sets. Int. J. Res. Mark. 12, 67–80.
Baltas, G., Doyle, P., 2001. Random utility models in marketing research: a survey. J. Bus. Res. 51, 115–125.
Batley, R., Daly, A., 2006. On the equivalence between elimination-by-aspects and generalized extreme value models of choice behavior. J. Math. Psychol. 50, 456–467.
Ben-Akiva, M., Boccaro, B., 1995. Discrete choice models with latent choice sets. Int. J. Res. Mark. 12, 9–24.
Blundell, R., Smith, R.J., 1994. Coherence and consistency in simultaneous models with censored or qualitative dependent variables. J. Econom. 64, 355–373.
Brown, J.J., Wildt, A.R., 1992. Consideration set measurement. J. Acad. Mark. Sci. 20, 235–243.
Cameron, T.A., James, M.D., 1987. Efficient estimation methods for"closed ended' contingent valuation surveys. Rev. Econ. Stat. 69, 269–276.
Cardell, N.S., 1997. Variance components structures for the extreme-value and logistic distributions with application to models of heterogeneity. Econom. Theory 13, 185–213.
Carson, R.T., Sun, Y., 2007. The tobit model with a non-zero threshold. Econom. J. 10, 197–212.
Chiang, J., Chib, S., Narasimhan, C., 1999. Markov chain monte carlo and models of consideration set and parameter heterogeneity. J. Econom. 9, 223–248.
Cheung, C., Goldberger, A., 1984. Proportional projections in limited dependent variable models. Econometrica 52, 531–553.
Ching, A., Erdem, T., Keane, M., 2009. The price consideration model of brand choice. J. Appl. Econom. 24, 393–420.
Cragg, J.G., 1971. Some statistical models for limited dependent variables with application to the demand for durable goods. Econometrica 39, 829–844.
Dagsvik, J.K., 1994. Discrete and continuous choice, max-stable processes, and independence from irrelevant attributes. Econometrica 62, 1179–1205.
Deaton, A., Muellbauer, J., 1980. Economics and Consumer Behavior. Cambridge University Press, New York.
Degeratu, A.M., Rangaswamy, A., Wu, J., 2000. Consumer choice behavior in on-line and traditional supermarkets: the effects of brand name, price and other search attributes. Int. J. Res. Mark. 17, 55–78.
Erdem, T., Swait, J., 2004. Brand credibility, brand consideration, and choice. J. Consum. Res. 31, 191–198.
Fox, J.T., 2007. Semiparametric estimation of multinomial discrete-choice models using a subset of choices. RAND J. Econ. 38, 1002–1019.
Fotheringham, A.S., 1988. Consumer store choice and choice set definition. Mark. Sci. 7, 299–310.
Gensch, D., 1987. A two-state disaggregate attribute choice model. Mark. Sci. 6, 223–239.
Greene, W., 2003. Econometric Analysis, 5th ed. Englewood Cliffs, NJ: Prentice Hall.
Hauser, J.R., Wernerfelt, B., 1990. An evaluation cost model of consideration sets. J. Consum. Res. 16, 393–408.
Heckman, J.J., 1976. The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. Ann. Econ. Soc. Meas. 5, 475–492.

[26] Such models exist; for example, Parsons et al. (2000) clearly recognized the sample selection problem involved in dropping options, and specified a straightforward nested model that allows parameter estimates to differ between options deemed as "familiar and not familiar". Swait (2001a) provides a flexible way to deal with endogenous choice set generation. Ching et al. (2009) provide a linked two stage model focused on resolving timing and uncertainty issues that may influence consumer decision making processes.

[27] Salisbury and Feinberg (2012), for instance, show that consideration sets tend to be larger when the task that consumers face is choosing one good from a set of options rather than when consumers are asked to choose multiple items (in the same class of goods) from a set of options.

[28] As Parsons et al. (2000) note, if a person's set of favorite beaches (consideration set) consists of only relatively nearby sites, distance to the beach will seem to be unimportant even if it is a key determinant of choice when looking at the actual set of options an individual faces.

Heckman, J.J., 1979. Sample selection bias as a specification error. Econometrica 47, 153–161.
Heckman, J.J., 1990. Varieties of selection bias. Am. Econ. Rev. 80, 313–318.
Heckman, J.J., Robb, R., 2000. Alternative methods for solving the problem of selection bias in evaluating the impact of treatments on outcomes. In: Wainer, H. (Ed.), Drawing Inferences from Self-Selected Samples, Lawrence Erlbaum, Mahwah, NJ.
Hernán, M.A., Hernández-Díaz, S., Robins, J.M., 2004. A structural approach to selection bias. Epidemiology 15, 615–625.
Horowitz, J.L., Louviere, J.J., 1995. What is the role of consideration sets in choice modeling. Int. J. Res. Mark. 12, 39–54.
Howard, J., Sheth, J., 1969. The Theory of Buyer Behavior. John Wiley, New York.
Hurd, M., 1979. Estimation in truncated samples when there is heteroscedasticity. J. Econom. 11, 247–258.
Ioannatos, P.E., 1995. Censored regression estimation under unobserved heterogeneity: a stochastic parameter approach. J. Bus. Econ. Stat. 13, 327–335.
Kotler, P., 2003. In: Marketing Management11th ed. Englewood Cliffs, NJ: Prentice Hall.
Krishnamurthi, L., Raj, P., 1985. The effect on advertising on consumer price sensitivity. J. Mark. Res. 22, 119–129.
Krishnamurthi, L., Raj, P., 1988. A model of brand choice and purchase quantity price sensitivities. Mark. Sci. 7, 1–20.
Kriström, B., 1997. Spike models in contingent valuation. Am. J. Agric. Econ. 79, 1013–1023.
Lee, L.F., 1995. Computation of opportunity costs in polychotomous choice models with selectivity. Rev. Econ. Stat. 77, 423–435.
Maddala, G.S., 1987. Limited Dependent and Qualitative Variables in Econometrics. Cambridge University Press, New York.
Manski, C., McFadden, D., 1981. Structural Analysis of Discrete Choice Data with Econometric Applications. MIT Press, Cambridge.
McFadden, D., 1974. Conditional logit analysis of qualitative choice behavior. In: Zarembka, P. (Ed.), Frontiers of Econometrics, Academic Press, New York.
McFadden, D., 1978. Modeling the choice of residential location. In: Karlquist, A., Lundquist, L., Snickars, F., Weibull, J.W. (Eds.), Spatial Interaction Theory and Planning Models, Amsterdam, North-Holland.
Moe, W.W., Schweidel, D.A., 2012. Online product opinions: incidence, evaluation, and evolution. Mark. Sci. 31, 372–386.
Mullahy, J., 1986. Specification and testing of some modified count data models. J. Econom. 33, 341–365.
Narayana, C.L., Markin, R.J., 1975. Consumer behavior and product performance: an alternative conceptualization. J. Mark. Res. 39, 1–6.
Nelson, F.D., 1977. Censored regression models with unobserved, stochastic censoring thresholds. J. Econom. 6, 309–327.
Paap, R., van Nierop, E., van Heerde, H.J., Wedel, M., Franses, P.H., Alsem, K.J., 2005. Consideration sets, intentions, and the inclusion of don't knows in a two-stage model for voter choice. Int. J. Forecast. 21, 53–71.
Parsons, G.R., Kealy, M.J., 2002. Randomly drawn opportunity sets in random utility model of lake recreation. Land Econ. 68, 93–102.
Parsons, G.R., Massey, D.M., Tomasi, T., 2000. Familiar and favorite sites in a random utility model of beach recreation. Mar. Resour. Econ. 14, 299–315.
Punj, G., Brookes, R., 2002. The influence of pre-decisional constraints on information search and consideration set formation in new automobile purchases. Int. J. Res. Mark. 19, 383–400.
Reilly, M., Parkinson, T.L., 1985. Individual and product correlates of evoked set size for consumer package goods. Adv. Consum. Res. 12, 492–497.
Ruud, P.A., 1986. Consistent estimation of limited dependent variable models despite misspecification of distribution. J. Econom. 32, 157–187.
Roberts, J.H., Lattin, J.M., 1991. Developing and testing a model of consideration composition. J. Mark. Res. 28, 429–440.
Roberts, J.H., Lattin, J.M., 1997. Consideration: review of research and prospects for future insights. J. Mark. Res. 34, 406–410.
Roberts, J.H., Nedungadi, P., 1995. Studying consideration in the consumer decision process: progress and challenges. Int. J. Res. Mark. 12, 3–7.
Salisbury, L.C., Feinberg, F.M., 2012. All things considered? The role of choice set formation in diversification. J. Mark. Res. 49, 320–335.
Salisbury, L.C., F.M. Feinberg, Y. Ying, 2013. When random assignment is not enough: accounting for intentional selectivity in experimental research, In: Proceedings of the Paper Presented at the International Choice Modeling Conference, Sydney.
Siddarth, S., Bucklin, R.E., Morrison, D.G., 1995. Making the cut: modeling and analyzing choice set restriction in scanner panel data. J. Mark. Res. 32, 255–266.
Smith, M.D., 2002. On specifying double hurdle models. In: Ullah, A., Wan, A., Caturvedi, A. (Eds.), Handbook of Applied Econometrics and Statistical Inference, Marcel-Dekker, New York.
Sonnier, G., Anslie, A., Otter, T., 2007. Heterogeneity distributions of willingness-to-pay in choice models. Quant. Mark. Econ. 5, 313–331.
Swait, J., 2001a. Choice set generation within the generalized extreme value family of discrete choice models. Transp. Res. Part B 35, 643–666.
Swait, J., 2001b. A non-compensatory model of incorporating attitude cutoffs. Transp. Res. Part B 35, 903–928.
Thomas, J.S., 2001. A methodology for linking consumer acquisition to consumer retention. J. Mark. Res. 38, 262–268.
Train, K., Weeks, M., 2005. Discrete choice models in preference space and willingness to pay space. In: Scarpa, R., Alberini, A. (Eds.), Applications of Simulation Methods in Environmental and Resource Economics, Springer, Boston.
Vella, F., 1998. Estimating models with sample selection bias: a survey. J. Hum. Resour. 33, 127–169.
von Haefen, R.H., 2008. Latent consideration sets and continuous demand systems. Environ. Resour. Econ. 41, 363–379.
Wachtel, S., Otter, T., 2013. Successive sample selection and its relevance for management decisions. Mark. Sci. 32, 170–185.
Werner, M., 1999. Allowing for zeros in dichotomous-choice contingent-valuation models. J. Bus. Econ. Stat. 17, 479–486.
Winship, C., Mare, R.D., 1992. Models for sample selection bias. Annu. Rev. Sociol. 18, 327–350.
Winer, R., 1983. Attrition bias in econometric models estimated with panel data. J. Mark. Res. 20, 177–186.
Wright, P., Barbour, F., 1977. Phased decision strategies: sequels to infinite screening. In: Starr, M., Zeleny, M. (Eds.), Multiple Criteria Decision Making, North-Holland, Amsterdam.
Yee, M., Dahan, E., Hauser, J., Orlin, J., 2007. Greedoid-based non-compensatory inference. Mark. Sci. 26, 532–549.