Searching for a Model of Multiple-Site Recreation Demand that Admits Interior and Boundary Solutions

Edward R. Morey, Donald Waldman, Djeto Assane, and Douglass Shaw

For most recreation demand data sets, different individuals visit different subsets of the available sites. Interior solutions (i.e. individuals who visit all recreational sites) are not the norm. Boundary solutions (i.e., individuals who do not participate, or who visit some, but not all, of the sites) predominate. We critique eight demand models in terms of their ability to accommodate boundary solutions. Three models are recommended for consideration when there are multiple sites and the data set includes a significant number of boundary solutions: a repeated nested-logit model, a multinomial share model, and a Kuhn-Tucker model.

Key words: boundary solutions, interior solutions, Kuhn-Tucker models, multinomial share models, repeated nested-logit models.

Consider a sample of individuals who may or may not participate in a site-specific recreational activity such as fishing. Each individual has more than one site to choose from and, through the course of the season, will choose how many trips to take to each of the available sites. Individuals may choose no trips. Assume there are complete trip records for the season for each individual in the sample, and consider the problem of modeling the number of trips each individual takes to each site.

While the problem has many aspects, we concentrate on the fact that, for most recreation demand data sets, individuals choose different subsets of the available alternatives. Interior solutions (i.e., individuals who visit all recreational sites) are not the norm. Boundaries (i.e., individuals who do not participate, or who visit some, but not all, of the sites) predominate. In this paper we consider how best to model this aspect of recreation demand.

Concern for this topic is not new. Hanemann lays out the problem of boundaries, but restricts his analysis to cases where the individual is constrained to consume only one of the alternatives. Hanemann notes for demand theory in general that "This is a restrictive assumption, but it is employed in almost all of the theoretical and empirical literature on discrete/continuous choice" (p. 543), and "One would like to allow for a more general corner solution in which the consumer may select any subset of the brands—not necessarily one of them, but not necessarily all of them" (p. 560). The problem of boundaries in recreational demand is extensively discussed in Bockstael, Hanemann, and Strand (chap 8). A characterization of the problem that is consistent with constrained utility-maximizing behavior, but deterministic, is presented in Kling. See Bockstael, Hanemann, and Kling; Bockstael, McConnell, and Strand; and Loomis and Ward for general surveys of recreational demand modeling.

In the multiple-site context, concern about boundaries has led to development of repeated discrete-choice models of recreational demand and the multinomial share models of site

Edward Morey and Donald Waldman are associate professors in the Department of Economics, University of Colorado, Boulder. Djeto Assane is assistant professor in the Department of Economics, University of Nevada, Las Vegas, and Douglass Shaw, formerly senior economist with RCG Hager Bailey, Inc., is associate professor in the Department of Agricultural Economics, University of Nevada, Reno.

The authors would like to thank two editors and the referees for their comments, and Bill Breffle and Trudy Cameron for insightful discussions.

choice.[1] For single sites, the only boundary solution that is observed is nonparticipation (i.e., zero trips).[2] In the single-site context, concern for the influence of nonparticipation on demand and benefit estimation has motivated the estimation of tobit models and modified tobit models (Bockstael, Hanemann, and Strand; Kling; and Smith), other truncated models of recreational demand (Shaw), and the recent count models of recreational demand (Smith; Creel and Loomis; Hellerstein; Hellerstein and Mendelsohn).[3]

The intent of this paper is to critique existing and proposed models of recreational demand in terms of their ability (or inability) to successfully accommodate boundaries. Seven existing models and one new model are considered. The purpose is to leave the reader with a better understanding of what the theoretical constraints (both economic and statistical) and data constraints (proportion of boundary observations) mean for the estimation of recreational demand.

A General Model of Participation and Site Choice

To make multisite demand concrete, assume that during the season the angler chooses from a vector of J alternatives where x_j is the chosen number of trips to site j, $j = 1, 2, \ldots, J - 1$, and x_J is the number of days the individual chooses not to fish. Alternatively, the J alternatives in the model could all be sites if the number of trips is exogenous. We will adopt the first interpretation, but all the issues and models discussed are also relevant in the more restrictive case where nonparticipation is not modeled. Boundary solutions dominate because few individuals visit all of the sites.

Individual preferences, defined over the season, can be represented by the direct random utility function

$$(1) \quad U = U(\mathbf{x}, \mathbf{a}, \boldsymbol{\varepsilon})$$

where $\mathbf{a} \equiv [a_{kj}]$, with a_{kj} the magnitude of attribute k the individual associates with alternative j, for $k = 1, 2, \ldots, K$. These magnitudes are observed by both the individual and the investigator.

Variable $\boldsymbol{\varepsilon}$ is a vector of influential determinants of the preference ordering that are known to the individual, but are random variables from the investigator's perspective. The elements of $\boldsymbol{\varepsilon}$ may or may not be correlated across sites and/or trips.

Notice that the season is the decision period. This is less restrictive than assuming each day or week in the season is an independent choice occasion. The preferences, equation (1), are sufficiently general to encompass the preferences assumed in many popular recreation demand models (e.g., discrete choice models, a single-site demand function, systems of site-specific demand functions, count models, tobit models, etc.).

Given a seasonal budget, B, the individual maximizes equation (1) subject to the constraint $B \geq \mathbf{p}'\mathbf{x}$, where p_j is the cost of a trip to site j, $j = 1, 2, \ldots, J - 1$, and p_J is the cost of a day at home. The resulting demand functions are

$$(2) \quad x_j = x_j(B, \mathbf{p}, \mathbf{a}, \boldsymbol{\varepsilon}) \qquad j = 1, 2, \ldots, J.$$

These demands are random variables from the investigator's perspective; their expectation is denoted $E[x_j] \equiv \hat{x}_j$.

For estimation purposes, an alternative and often more convenient way to represent the solution to the individual's utility maximization problem is the system of share functions, where s_j is the proportion of choice occasions that alternative j is chosen

$$(3b) \quad s_j \equiv \frac{x_j}{\sum_{l=1}^{J} x_l} = s_j(B, \mathbf{p}, \mathbf{a}, \boldsymbol{\varepsilon}) \quad j = 1, 2, \ldots, J$$

with expected shares $E[s_j] \equiv \hat{s}_j$. Demand functions and share functions should be viewed as simply two different ways to express the solution to the utility maximization problem, rather than as different models. Modeling and estimation requires specification, or derivation, of a probability rule for either the vector of demands or vector of shares. A probability rule is a mathematical function that indicates how the

---

[1] Numerous authors have proposed and/or estimated repeated discrete-choice models of recreational demand. See Feenberg and Mills, Caulkins, Bishop, and Bouwes (1984, 1986), Bockstael, Hanemann, and Strand, Carson, Hanemann, and Wegge, Morey, Shaw, and Rowe, Morey and Rowe, and Morey, Rowe, and Watson. For multinomial share models see Morey (1981, 1984, 1985) and Morey and Shaw.

[2] The only other boundary in the single-site context is all of the season at the site, and no time at home. This boundary is not observed.

[3] There is, in addition, an extensive literature on the estimation of single-site demand when the data set excludes nonparticipation, that is, when boundary solutions are excluded from the sample. For a review of the issues see Bockstael, Hanemann, and Strand (chap. 3), and Bockstael et al.

random demands (shares) are distributed. Probability rules for continuous random variables are called density functions, and for discrete random variables are called mass functions. Denote a random variable that is continuous over part(s) of its domain but discrete over other parts of its domain a mixed random variable, and denote probability rules for mixed random variables continuous/discrete probability rules.

It is often easier to work with probability rules defined over shares than probability rules defined over demands because the restrictions on shares (constrained to the zero-one interval, sum to one) are simpler to impose than the restrictions on demands (budget exhaustion and nonnegativity).

Modeling interiors and boundaries requires a probability rule for the shares that fulfills all the properties of shares, includes boundaries in its domain, and, in addition, admits the possibility of observing a significant number of boundaries. In the next section, we describe the properties such a probability rule must have.

We are concerned with modeling boundaries and interiors in situations where there are three or more alternatives in the choice set (nonparticipation and at least two sites).[4] In this multiple-site framework, nonparticipation (zero trips) is simply one of the many possible types of boundary solutions. In contrast, when there is only one site in the choice set, the problem is one of modeling the number of trips each individual takes to the single site, and the only observed boundary solution is nonparticipation (zero trips). Much research has been done to incorporate this single boundary solution (nonparticipation) into single-site models of recreational demand.[5] In contrast, in this paper we are concerned with the problem of modeling recreational demand with multiple sites, where it is important to estimate the allocation among sites, and where the data set contains significant numbers of different types of boundary solutions.

Properties of the Probability Rule

The choice of the probability rule is dictated by the restrictions the theory and constraints impose on observed demands. Specifically, all shares, s_i, must be bounded by zero and one ($0 \le s_i \le 1$) and sum to one, so only $J - 1$ of the shares can independently vary; that is $s_i = 1 - \sum_{l \in J} s_l$, $l \neq i$. The probability rule for the J shares, $f(s_1, s_2, \ldots, s_J)$ is therefore defined over a $J - 1$ unit simplex, with s_J defined as the residual $s_J = 1 - \sum_{j=1}^{J-1} s_j$.[6] For example, if there are three alternatives (two sites and nonparticipation), the probability rule $f(s_1, s_2, s_3)$ is defined over the two-dimensional $\{s_1, s_2\}$ right triangle, where $s_3 = 1 - s_1 - s_2$. See figure 1. A probability rule with positive density or mass outside of the unit simplex is inconsistent with the concept of shares. Choice of an appropriate probability rule is therefore restricted to those whose domain is restricted to the unit simplex. An observed share vector in the interior of this simplex is an interior solution; those on the boundaries are boundary solutions. Hanemann denotes solutions where only one of the observed shares is positive and the rest zero as extreme corner solutions.

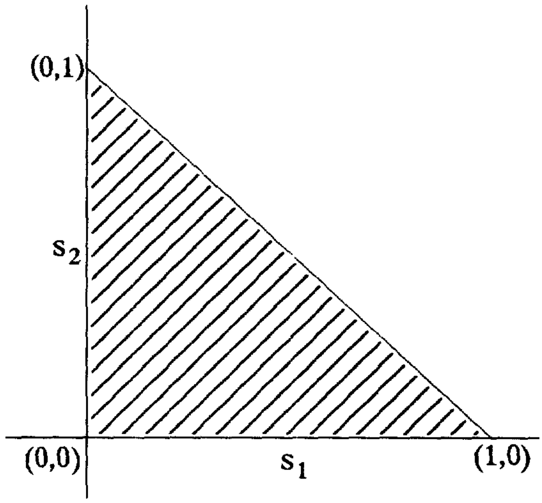

Figure 1. The Unit Simplex ($J = 3$)

If all of the observed shares in a data set are positive, it is defensible to consider only probability rules that restrict themselves to the interior of the unit simplex. However, observed boundaries are inconsistent with such probability rules. Data sets that contain numerous

[4] Or, if nonparticipation is not modeled, at least three sites.

[5] See, for example Bockstael, Hanemann, and Strand, Kling, Shaw, Smith; Creel and Loomis; Bockstael et al.; Hellerstein, and Hellerstein and Mendelsohn. These models are sometimes applied to data sets with multiple sites, but in estimation the trips are grouped together and not modeled by site, that is, the number of trips are estimated, but not their destinations.

[6] A $J - 1$ unit simplex is a right isosceles triangle in $J - 1$ dimensional space whose legs are unity. Dual to this unit simplex for the shares, the probability rule for the demand functions is defined over a $J - 1$ right triangle whose length in the jth direction is the maximum number of times the individual can afford to choose that alternative.

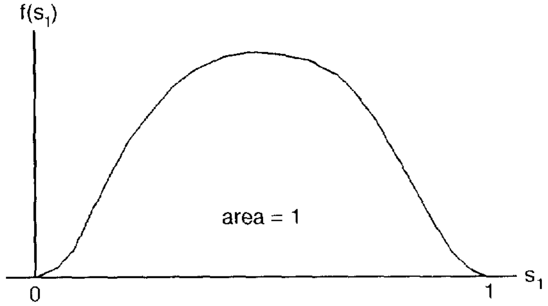

Figure 2. A continuous distribution of the Unit Simplex ($J = 2$)

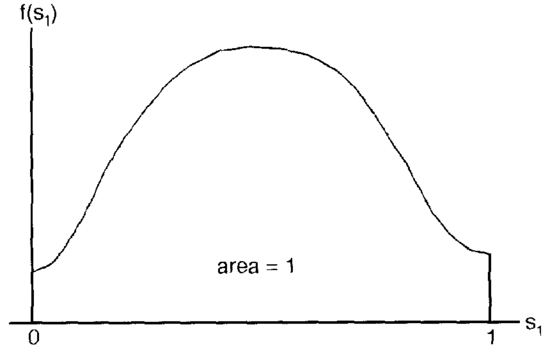

Figure 3. A continuous distribution on the Unit Simplex ($J = 2$) with positive density at zero and one

boundary solutions require probability rules that have positive mass on the boundaries of the unit simplex. The stochastic modeling of shares that can repeatedly assume the same value (e.g., zero) is a difficult, nontraditional problem in statistics. The problem is easy to appreciate: for continuously distributed shares, the probability of many shares with the same value is zero.

Some examples will help clarify these points. A continuous distribution defined over the unit interval (the $J = 2$ case because it is easy to graph) is the beta distribution (a special case of the Dirichlet distribution), shown in figure 2. This would be an appropriate stochastic assumption for a data set that included only interior solutions; since $f(s_1) = 0$ for $s_1 = 0$ and $s_1 = 1$, it is inconsistent with an observed share of zero or one. Next consider a density, illustrated in figure 3, that is consistent with a data set that contains one or at most a few zeros or ones, but would be inconsistent with a data set that included a significant number of boundary solutions (or a significant number of like observations of any one value). In figure 3, there is a positive density at the boundaries $[f(0) > 0$ and $f(1) > 0]$. There are many discrete distributions consistent with a significant number of boundary solutions, such as the one shown in figure 4. A problem with using a discrete distribution is that numerical values for shares not explicitly part of the distribution have zero probability. A mixed distribution, the one shown in figure 5 for example, can accommodate any observed share value and a significant number of boundary solutions.

The Search For a Probability Rule Consistent with Observed Boundaries

The difficulties associated with the task of finding a probability rule to model numerous boundary solutions is made even more difficult if shares are to be consistent with constrained utility-maximizing behavior. Seven existing demand models, and one new multisite demand model are critiqued in terms of their ability (or inability) to accommodate boundaries. This section divides models of interiors and boundaries into three categories: models that assume shares are continuously distributed random variables, models that assume shares are discrete random variables, and models that assume shares are mixed random variables.

Shares that are Continuous Random Variables

First, consider models that assume shares are continuously distributed random variables. In this category, the traditional approach is to assume that $J - 1$ shares, or demands, are multivariate normally distributed, and that the expected values of these shares solve a deterministic constrained utility maximization problem. However, the normality assumption is inconsistent with the properties of both share functions and demand functions (Woodland). Specifically, normally distributed shares are not restricted to the unit simplex, so there is a posi-

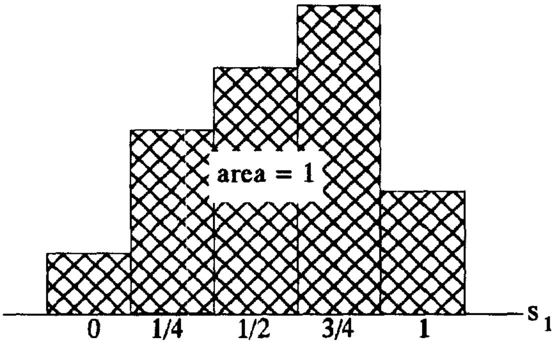

Figure 4. A discrete distribution on the Unit Simplex (J = 2)

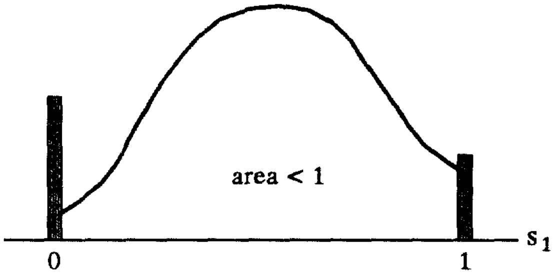

Figure 5. A mixed discrete/continuous distribution on the Unit Simplex (J = 2)

tive probability of shares both greater than one and less than zero. In addition, because a normally distributed random variable has positive density, but not positive probability at zero and one, it admits the possibility of boundaries, but it is inconsistent with a significant number of boundaries. In spite of this inconsistency, shares (or demands) are assumed normally distributed in much empirical work. Normality is defended as a reasonable approximation for data sets that include only interior solutions (Woodland), but cannot be justified for data sets that include a significant number of boundaries.[7] These conclusions also apply to other density functions that are not restricted to the unit simplex.

Because density functions for shares must be restricted to the unit simplex, Woodland has suggested the use of the Dirichlet distribution

$$(4) \qquad f(s_1, s_2, \ldots, s_J; \hat{x}_1 \hat{x}_2, \ldots, \hat{x}_J) = K \prod_{j=1}^{J} s_j^{(\hat{x}_j - 1)},$$

$$\hat{x}_j > 0 \text{ and } 0 < s_j < 1 \ \forall J, \text{ and } \sum s_j = 1$$

where

$$K = \frac{\Gamma\left(\sum_{j=1}^{J} \hat{x}_j\right)}{\prod_{j=1}^{J} \Gamma(\hat{x}_j)},$$

and $\Gamma(.)$ is the Gamma function.

The Dirichlet distribution, unlike the normal distribution, is consistent with both shares summing to one and being bounded by zero and one ($0 < s_j < 1$). However, the Dirichlet distribution only has positive density in the interior of the unit simplex, so boundaries are inadmissible. Therefore, the Dirichlet distribution is suitable for systems of share equations if all of the observations are interior solutions, but not if the data set includes boundaries. In fact, because there is no density at zero or one, the Dirichlet cannot be used to estimate data sets that include zero shares.

There are no multivariate density functions with positive density over the entire unit simplex, including boundaries, and restricted to the unit simplex. They must be created by starting with a density function that includes the unit simplex and truncating to zero those parts of the density function outside of the unit simplex. For example, the density that is outside of the unit simplex could be uniformly added to the density over the unit simplex.[8] This approach was pursued by Morey (1984, footnote 4) with a multivariate normal distribution. This class of density functions, called truncated/uniformly-added density functions, has two desirable properties with respect to share equations: it restricts the shares to the unit simplex and admits observed boundaries. However, truncated/uniformly-added density functions, like other density functions, associate zero probability with observing a significant number of boundaries (see, for example, figure 3). Therefore, like all models that assume the shares are con-

[7] Data sets with significant numbers of observed boundaries are historically quite recent. When demand systems were mostly defined in terms of broad aggregates such as food, clothing, and housing, everyone consumed some of each category. It is the emerging prevalence of micro data sets where there is information on each individual's consumption of a large number of individual goods that has created the empirical problem of boundaries.

[8] Formally, define some continuous random variable ω. If its density function $f(\omega)$ is defined over R, and the density is truncated to zero below $\omega = a$, then the truncated/uniformly-added density function is $f(\omega)/[1 - F(a)]$ where $F(\cdot)$ is the CDF of ω.

tinuously distributed random variables, they are inconsistent with data sets that contain a significant number of observed boundaries. However, these models can be estimated with such data sets, and may produce acceptable results if there are only a few boundaries

Shares that are Discrete Random Variables

Assuming that trips to the J sites take only nonnegative integer values, observed shares will be discrete, rather than continuous, random variables. This assumption is a simple, but restrictive, way to accommodate observed boundaries.

Given that trips can only be taken in nonnegative integer units, the mass function for the trips can be represented by a multinomial distribution

$$(5) \quad f(x_1, x_2, \dots, x_J; T, \theta_1, \theta_2, \dots, \theta_J) = \frac{T!}{\left(\prod_{i=1}^{J} x_i!\right)} \prod_{j=1}^{J} \theta_j^{x_j}$$

where $\theta_1, \theta_2, \dots, \theta_J$ are parameters of the distribution, and T is either fixed or an additional parameter. The multinomial distribution restricts the shares to the unit simplex, and it can attach positive probability to any vector of shares consistent with the integer nature of trips, including those that involve boundaries. This makes the multinomial model well suited for modeling and for estimating with data sets that include significant numbers of boundaries. The multinomial is restrictive in that it requires trips to site k be negatively correlated with trips to site i, for all k and i.

The only modeling issue is specification of θ and T to make these parameters consistent with constrained utility-maximizing behavior, and to make the model empirically tractable. If T is the number of independent choice occasions during the season and θ_j is the per-choice occasion probability that alternative j will be chosen, then the multinomial is a repeated discrete-choice model of alternative selection (e.g., repeated logit or nested-logit models of participation and site choice).[9] If $T = \sum_{j=1}^{J} x_j$ and $\theta_j = \hat{s}_j$, the model is the multinomial share model of Morey (1981, 1984, 1985) and Morey and Shaw.

Repeated discrete-choice models of participation and site choice assume the season can be divided into a discrete number of choice occasions, T, where θ_j is the per-choice occasion probability that alternative j is chosen. These probabilities are derived by specifying conditional indirect utility functions for each of the alternatives (the $J - 1$ sites and the nonparticipation alternative) as a function of the budget, the cost of the alternative, characteristics of the alternative, and a random component. These models are random utility models. If these random terms are independent drawings from an extreme value distribution, the model is a repeated logit model of participation and site choice. If the drawings are from a less restrictive form of the generalized extreme value distribution, the model is a repeated nested-logit of participation and site choice. Two restrictive separability assumptions in both the repeated logit and nested-logit models are that (*a*) the deterministic component of utility from alternative j is not a function of the attributes of the other alternatives, or even of the attributes of that alternative on other choice occasions; and (*b*) the random components are statistically independent across choice occasions. The repeated logit model of participation and site choice imposes the additional separability restriction of complete statistical independence (i.e., all random components uncorrelated).

A drawback of discrete-choice RUMs of site choice is how they achieve consistency with constrained utility maximization over the season. We are uncomfortable with assuming that the season consists of a finite number of independent choice occasions, where the probability of visiting site j on a given trip is independent of where the individual went on prior trips and where the individual might go on future trips. However, because of this restrictive assumption, repeated nested-logit models are both easy to estimate and consistent with constrained utility-maximizing behavior. This makes them attractive for data sets that contain significant numbers of boundary solutions.

Alternatively, one can adopt the discrete-choice multinomial model of site choice and assume $\theta_j = \hat{s}_j$ (the expectation of the share for site j), where T is the observed number of trips plus days at home ($T = \sum_{j=1}^{J-1} x_j + x_J$), and where the expected shares are derived by maximizing expected seasonal utility, $E[U(\mathbf{x}, \mathbf{a}, \boldsymbol{\varepsilon})]$, subject to the constraints $B \geq \mathbf{p'x}$ and $E[\boldsymbol{\varepsilon}] = 0$. The multinomial share model associates positive

[9] See Carson, Hanemann, and Wegge, Morey, Shaw, and Rowe, Kling and Thomson, Morey and Rowe, and Morey, Rowe and Watson

probability with observed interiors and boundaries. This model is consistent with maximizing utility subject to the seasonal budget constraint, but the expected share, $\hat{s}_j$, is not constrained by the integer nature of trips: the integer constraint is accommodated by the stochastic specification, the multinomial. The multinomial share model is therefore consistent with constrained utility-maximizing behavior.

In addition, the multinomial share model, unlike the repeated discrete choice models, does not impose strong separability across choice occasions. Given this, and the ease with which the multinomial share model can be estimated, it is a viable candidate for estimation when the data contain significant numbers of observed boundaries. This is particularly true when it is difficult to justify the assumption that the probability of visiting a site, on a given choice occasion, is independent of where the individual has been or might go.

Shares that are Mixed Random Variables

First, consider out-of-simplex truncation where a density function for some $J - 1$ dimensional random variable $\tilde{s}$ is chosen such that $\tilde{s}$ includes the unit simplex as a subset. The probability rule for s is created by truncating to zero those parts of the density function for $\tilde{s}$ that are outside of the unit simplex, and adding the density of $\tilde{s}$ that was outside of the unit simplex to the boundaries of the unit simplex. Typically, out-of-simplex truncation is defined as

(6) $s_j = 0 \text{ if } \tilde{s}_j \le 0$ and

$$s_j = \frac{\tilde{s}_j}{\sum_{l \in M} \tilde{s}_l} \text{ if } \tilde{s}_j > 0 \text{ where } M \equiv \{j\text{: } \tilde{s}_j > 0\}.$$

This type of truncation is referred to as out-of-simplex truncation because all of the density outside of the unit simplex is truncated.

Probability rules for shares created by out-of-simplex truncation have desirable boundary properties. Specifically, shares are restricted to the unit simplex and the probability rule admits data sets that include significant numbers of observed boundaries. A problem with out-of-simplex truncation is that the proportional expansion of the shares that are positive is not, in general, consistent with maximizing $U(\mathbf{x}, \mathbf{a}, \boldsymbol{\varepsilon})$, subject to $B \ge \mathbf{p}'\mathbf{x}$ and $\mathbf{x} \ge 0$.

While out-of-simplex truncation could be applied to any multivariate density function defined over the unit simplex and some or all of the space adjacent to it, all examples we are aware of assume $\tilde{s}$ is multivariate normal. For example, if $J = 2$ and normality is assumed, and only negative shares are truncated to zero, the model is the classic tobit model (Tobin).[10] If $J = 2$, normality is assumed, and both negative shares and shares greater than one are truncated, the model is a double-truncated Tobit model. Out-of-simplex truncation of the normal has been extended to a system of three share equations by Wales and Woodland who estimated family budget shares for beef, lamb, and other meats in meat consumption in Australia.

Wales and Woodland's observed shares are drawn from an out-of-simplex truncated normal, where $E[\tilde{s}]$ are the shares resulting from maximizing expected seasonal utility, $E[U(\mathbf{x}, \mathbf{a}, \boldsymbol{\varepsilon})]$, subject to the constraints $B \ge \mathbf{p}'\mathbf{x}$ and $E[\boldsymbol{\varepsilon}] = 0$. If $\tilde{s}$ are interpreted as desired shares, this model is consistent with constrained utility-maximizing behavior in terms of the expected desired shares, but provides no explanation for why desired shares are normally distributed, or why the observed shares and the discrete/continuous nature of their distribution is consistent with constrained utility-maximizing behavior.

To provide a rationale for truncating shares that are random variables, a distinction must be made between observed shares and desired shares. A distinction results if the individual's desired demands were obtained by maximizing utility subject to some but not all of the constraints. Ignoring some of the constraints leads to desired shares that are impossible to achieve if the ignored constraints are binding. In such cases, the individual's observed shares will result from a truncation of the desired shares where the truncation takes account of the additional constraints. An important consideration is whether the truncation rule is consistent with constrained utility-maximizing behavior.

Consider trying to explain out-of-simplex truncation of the normal distribution, where the $\tilde{s}$ are the individual's desired shares, and where these desired shares are distributed multivariate normal with means equal to the shares derived by maximizing expected seasonal utility, $E[U(\mathbf{x}, \mathbf{a}, \boldsymbol{\varepsilon})]$, subject to the constraints that $B \ge \mathbf{p}'\mathbf{x}$ and $E[\boldsymbol{\varepsilon}] = 0$. Normally distributed desired shares imply that an individual will sometimes have desired shares that are negative or

---

[10] This model extended to a system of goods by Amemiya (1974).

greater than one. Since it is impossible to choose shares that are negative or larger than one, the individual adjusts by truncating negative shares to zero and proportionately adjusting the other shares. This is not convincing. If it is believed that desired shares are normally distributed, negative shares obviously must be truncated to zero, but assuming positive desired shares are converted to observed shares by proportionately expanding them is, at best, an approximation. Nonetheless, it is difficult to imagine that desired shares could be normally distributed. If the individual is rational, desired shares, like observed shares, must be restricted to the unit simplex.

That desired shares must be restricted to the unit simplex led us to consider a model where desired shares have a Dirichlet distribution and observed shares are obtained by truncating the desired shares with a minimum-share rule. We refer to this new model as the truncated-Dirichlet share model.[11] We assume a two-stage decision process: at the first stage, desired demand is determined, and at the second stage, desired behavior is truncated into observed behavior. Assume the desired shares, $\tilde{s}$, have a Dirichlet distribution with expected desired shares obtained by maximizing $E[U(\mathbf{x}, \mathbf{a}, \boldsymbol{\varepsilon})]$, subject to the constraints $B \geq \mathbf{p}'\mathbf{x}$ and $E[\boldsymbol{\varepsilon}] = 0$. At the second stage, the observed shares, s_j, are determined from the desired shares, $\tilde{s}_j$, by the following truncation rule:

$$(7) \qquad s_j = 0 \text{ if } \tilde{s}_j \leq s_{\min}(j) \text{ and}$$

$$s_j = \frac{\tilde{s}_j}{\sum_{l \in S} \tilde{s}_l} \text{ if } \tilde{s}_j > s_{\min}(j) \text{ where}$$

$$S \equiv \{j : \tilde{s}_j > s_{\min}(j)\}.$$

Put simply, if $\tilde{s}_j$ falls below some minimum level, $s_{\min}(j)$, the individual chooses not to visit site j. This truncation rule is referred to as minimum-share truncation.

Minimum-share truncation of the Dirichlet is indefensible within a pure framework of constrained utility-maximizing behavior. It is at best a "rule of thumb" that approximates the individual's decision process. However, we find it a more plausible model than the truncated-normal model. Like the truncated-normal model, the truncated-Dirichlet model restricts the shares to the unit simplex and admits the possibility of data sets with significant number of observed boundaries. Unlike the truncated-normal model, it also appropriately restricts the desired shares to the unit simplex. As a result of the minimum-share truncation rule that maps desired shares into observed shares, the likelihood function necessarily involves the evaluation of some line and surface integrals of the Dirichlet function, in addition to evaluation of the usual probability density functions.

While we have estimated a truncated Dirichlet share model (Morey, Waldman, Assane, and Shaw), we do not advocate this model or refinements to it. Minimum share truncation of the Dirichlet is preferable on theoretical grounds to the out-of-simplex truncation of the normal, but minimum-share truncation of the Dirichlet is not completely consistent with constrained utility-maximizing behavior, and the model is significantly more difficult to estimate than either the repeated nested-logit model or the multinomial share model. Difficulty of estimation is on par with the next model—a Kuhn-Tucker model—which is consistent with constrained utility-maximizing behavior.

The Kuhn-Tucker (K-T) model of interior and boundary solutions (Wales and Woodland; Lee and Pitt; Bockstael, Hanemann, and Strand) explicitly specifies the distribution function for $\boldsymbol{\varepsilon}$ in the random-utility function, $U(\mathbf{x}, \mathbf{a}, \boldsymbol{\varepsilon})$. This direct specification of the distribution of $\boldsymbol{\varepsilon}$ differentiates the K-T model from the out-of-simplex truncation and minimum-share truncation models. Those models explicitly specify a probability rule for the shares, but do not consider the dual restrictions on the random components in the utility function. In contrast, in the K-T model, after specifying an explicit distribution function for $\boldsymbol{\varepsilon}$, $f(\boldsymbol{\varepsilon})$, the conditions for the maximization of $U(\mathbf{x}, \mathbf{a}, \boldsymbol{\varepsilon})$ subject to $B \geq \mathbf{p}'\mathbf{x}$ and $\mathbf{x} \geq 0$ are derived.[12] The K-T conditions and $f(\boldsymbol{\varepsilon})$ implicitly define the probability rule for the shares. These probabilities are functions of the unknown parameters in $U(\mathbf{x}, \mathbf{a}, \boldsymbol{\varepsilon})$ and, for boundaries, are

[11] For an empirical application to recreational sites in the Adirondaks, see Morey, Waldman, Assane, and Shaw

[12] In dual space, Lee and Pitt first specify an indirect random utility function, $V(\mathbf{p}, B, \mathbf{a}, \boldsymbol{\varepsilon}) \equiv \max_x\{U(\mathbf{x}, \mathbf{a}, \boldsymbol{\varepsilon}) : B = \mathbf{p}'\mathbf{x}\}$, which ignores the nonnegativity constraints on demand. Desired demands (shares) are derived from $V(\mathbf{p}, B, \mathbf{a}, \boldsymbol{\varepsilon})$ using Roy's Identity. Some of the desired demand will be negative. Negative demands are truncated to zero. Observed shares for the consumed goods are obtained from the functions for the desired shares, evaluated at the market prices for the consumed goods and at the prices for the nonconsumed goods that would just make the demands for those nonconsumed goods zero

multiple integrals of the density of the transformation from $\boldsymbol{\varepsilon}$ to $\mathbf{s}$.

The parameters in $U(\mathbf{x}, \mathbf{a}, \boldsymbol{\varepsilon})$ are estimated by maximizing the likelihood function $L = \prod_{i=1}^{N} \text{prob}(s_{i1}, s_{i2}, \ldots, s_{iJ})$, where s_{ij} is individual i's observed share for alternative j. The shares resulting from the K-T model are mixed random variables on the unit simplex, and their implicit probability rule admits mass on the boundaries of that simplex. These properties make it an attractive model when the data set includes a significant number of observed boundaries. Its theoretical advantage over the other two truncation models is that its transformation of desired shares into observed shares is consistent with constrained utility-maximizing behavior.[13]

Consider the basic problem of maximizing $U(\mathbf{x}, \mathbf{a}, \boldsymbol{\varepsilon})$, equation (1), subject to the budget constraint $\mathbf{p}'\mathbf{x} \leq B$ and the nonnegativity constraints $x_j \geq 0 \,\forall j$ The K-T conditions are

$$(8) \qquad x_j > 0 \text{ if } \frac{\partial U(\mathbf{x}, \mathbf{a}, \boldsymbol{\varepsilon})}{\partial x_j} - \lambda p_j = 0 \text{ and}$$

$$x_j = 0 \text{ if } \frac{\partial U(\mathbf{x}, \mathbf{a}, \boldsymbol{\varepsilon})}{\partial x_j} - \lambda p_j < 0$$

where λ is the Lagrangian multiplier. These Kuhn-Tucker conditions are well behaved, whereas the demand functions are discontinuous at zero. It is therefore easier to work with the Kuhn- Tucker conditions directly, although the intent is still to derive each individual's probability rule for the observed shares, $\text{prob}(s_1, s_2, \ldots, s_J)$.

To derive these probability rules, define the variable $\eta_j = [\partial U(\mathbf{x}, \mathbf{a}, \boldsymbol{\varepsilon})/\partial x_j] - \lambda p_j, j = 1, \ldots, J$, where $\boldsymbol{\eta} \equiv [\eta_1, \eta_2, \ldots, \eta_J]$. The elements of $\boldsymbol{\eta}$ are deterministic from the individual's perspective but are random variables from the analyst's perspective. If $\eta_j = 0$, alternative j is chosen one or more times during the season, and if $\eta_j < 0$, alternative j is never chosen. The density function for $\boldsymbol{\eta}$, $g(\eta_1, \eta_2, \ldots, \eta_J)$, is determined from the $f(\boldsymbol{\varepsilon})$ by a change of variables that is a function of $U(\mathbf{x}, \mathbf{a}, \boldsymbol{\varepsilon})$. The density function $g(\eta_1, \eta_2, \ldots, \eta_J)$ is then used to derive the probability rule $\text{prob}(s_1, s_2, \ldots, s_J)$. To simplify the notation, the M alternatives (sites) the individual never chooses are listed first, in which case

$$(9) \quad \text{prob}[x_j = 0, j = 1, \ldots, M \text{ and } x_j = \bar{x}_j > 0, j = M+1, \ldots, J]$$
$$= \text{prob}[\eta_j < 0, j = 1, \ldots, M \text{ and } \eta_j = 0, j = M+1, \ldots, J]$$
$$= \int_{-\infty}^{0} \cdots \int_{-\infty}^{0} g(\eta_1, \eta_2, \ldots, \eta_M, 0, \ldots, 0) \cdot d\eta_1 d\eta_2 \ldots d\eta_M = \text{prob}[s_j = 0, j = 1, \ldots, M \text{ and } s_j = \bar{s}_j > 0, j = M+1, \ldots, J].$$

Note that the dimensionality of the integration is not the number of alternatives in the choice set, but rather the number of alternatives that the individual never chooses. Since in a model of participation and site choice, individuals always spend some of the season at home, $s_J > 0$ and the number of alternatives not chosen is the number of sites not visited. In general, the integral, equation (9), will not have a closed-form solution. Two special cases where equation (9) has a closed-form solution are the repeated nested logit model and the multinomial share model. Equation (9) is significantly simplified if we assume that $U(\mathbf{x}, \mathbf{a}, \boldsymbol{\varepsilon}) = U[u_1(x_1, \mathbf{a}_1, \varepsilon_1), u_2(x_2, \mathbf{a}_2, \varepsilon_2), \ldots, u_J(x_J, \mathbf{a}_J, \varepsilon_J)]$. An even stronger separability assumption is made in the repeated nested-logit model of participation and site choice.[14] An additional assumption of the repeated logit model of participation and site choice is that the $\boldsymbol{\varepsilon}_j$ are all statistically independent. A common starting point for K-T models is to assume $\boldsymbol{\varepsilon}$ is a multivariate normally distributed with unconstrained covariance matrix, an assumption that will guarantee that equation (9) does not have a closed-form solution.

How many alternatives can be included in a general K-T model of participation and site choice will depend on the specifics of the model, the specifics of the data set, and the patience of the analyst. Estimation of a K-T model with $\boldsymbol{\varepsilon}$ multivariate normally distributed

[13] If prices are constant, Ransom (pp. 355–59) has shown that there is little difference between the K-T model and the simultaneous equations model with limited dependent variables of Amemiya. If prices are different, then the Amemiya model has heteroskedastic errors. The primary differences lie in the parameterization (that is, in what is estimated), and the way in which randomness is incorporated into the model. The K-T model begins with utility functions that are random across individuals and derives the resulting demand functions, constrained by the budget so that zero consumption may result. In the Amemiya model, demand functions are assumed and (truncated) random errors are added to them.

[14] $u_j(x_j, \mathbf{a}_j, \varepsilon_j) = \bar{u}_j(x_{j1}, \mathbf{a}_{j1}, \varepsilon_{j1}) + \ldots + \bar{u}_j(x_{jt}, \mathbf{a}_{jt}, \varepsilon_{jt}) + \ldots + \bar{u}_j(x_{jT}, \mathbf{a}_{jT}, \varepsilon_{jT})$, where t indexes the choice occasions.

and a fairly general utility function is possible if the number of sites not visited does not exceed four or five for most individuals. In these cases, the analyst should give serious consideration to estimating a general K-T model specification, for it allows an unconstrained specification of the covariance of errors, and hence no unnatural economic constraints. The only empirical application of a general K-T model is Wales and Woodland.[15] They assumed that $\boldsymbol{\varepsilon}$ is multivariate normally distributed and $U(\mathbf{x}, \mathbf{a}, \boldsymbol{\varepsilon})$ quadratic. The quadratic specification implies that the marginal utility for commodity j has only ε_j as an additive random component.

Current estimation of such a general K-T model with enough sites so that many individuals in the sample fail to visit as many as ten sites would require more patience, and computer time, than most of us possess. Currently, accurate and speedy numerical integration over ten dimensions is nonexistent. One of the most exciting, but not yet proven, potential solutions is to circumvent the numerical integration problem completely by using instead Monte Carlo methods to calculate the probability associated with a boundary solution.[16] This research has focused on the high-dimensional multinomial probit model, but we know of no substantive examples to date.

Suggestions for Choice of Estimator and Directions for Future Research

If a data set does not include boundary solutions, the best models to use would be the Dirichlet model or the truncated/uniformly-added normal model; both models appropriately restrict the shares to the unit simplex. If the data set includes only a small proportion of boundary solutions (less than 10%), a good choice would be to assume the shares have a truncated/uniformly-added distribution because this distribution admits density on the boundaries of the simplex. However, all multiple-site recreation demand data sets contain a significant number of boundary solutions. For example, all of the observations in Hanemann's Boston beach data are boundaries; the same is true in Morey's skiing data [Morey (1981)], and the Atlantic salmon fishing data set (Morey, Rowe, and Watson). Few individuals visit all sites. If there are extreme corner solutions (i.e., individual who visits only one of the sites) but no interior solutions or other types of boundaries, we recommend the extreme corner model of Hanemann.

If the sample contains a significant number of boundary solutions, we recommend for both participation and site choice a repeated nested-logit model, a multinomial share model, or a more general K-T model. The critical issue for estimation is the number of sites unvisited for each individual in the sample. If this number is less than five for most individuals in the sample, the K-T model is tractable and should be used because it imposes the fewest a priori restrictions on preferences.

If the number of sites not visited is more than five for many individuals, we recommend against the K-T model until computers become sufficiently fast, or the method of simulated moments becomes operational. Until then, we suggest a repeated nested-logit model of participation and site choice or a multinomial share model of participation and site choice. Both models are consistent with constrained utility-maximizing behavior and can be estimated with data that include individuals who visit a small number of the sites. Both models are restrictive in that they require that trips to site k be negatively correlated with trips to site i, for all k and i, and that the number of trips be integers. While the repeated nested-logit model is better understood than the multinomial share model, the repeated nested-logit model, unlike the multinomial share model, restrictively assumes independence across choice occasions.

In summary, an ideal model for data with significant numbers of observed boundaries that is both tractable and consistent with constrained utility-maximizing behavior does not currently exist. Shares must be restricted to the unit simplex. In addition, shares must either be discrete random variables, or mixed random variables, and admit positive probability along the boundary of the unit simplex. Boundaries are inconsistent with shares that are continuously distributed random variables. A critical modeling decision is whether to assume shares are discrete random variables (the repeated nested-logit model and the multinomial share model) or mixed random variables (the general K-T model).

[Received May 1992;
final revision received August 1994.]

[15] Lee and Pitt (1986) specify a translog functional form for the indirect utility function that contains a normally distributed ε, and from this random-utility function they derive probability functions for vectors of observed shares, but they do not estimate a model.

[16] See the pathbreaking theoretical work in the method of simulated moments by McFadden and Pakes and Pollard.

References

Amemiya, T. "Multivariate Regression and Simultaneous Equation Models When the Dependent Variables are Truncated Normal." *Econometrica* 42(September 1974):999–1012.

Bockstael, N., K.E. McConnell, and I.E. Strand. "Recreation," *Measuring the Demand for Environmental Quality*. J.B. Braden and C.D Kolstad, eds. North Holland: Elsevier Science Publishers, 1991.

Bockstael, N., W.M. Hanemann, and C L. Kling. "Estimating the Value of Water Quality Improvements in a Recreation Demand Framework." *Water Resour. Res.* 23(May 1987):951–60.

Bockstael, N., W.M. Hanemann, and I E. Strand. *Measuring the Benefits of Water Quality Improvements Using Recreation Demand Models*, Vol. 2. Washington DC: U.S. Environmental Protection Agency, Office of Policy Analysis, 1986.

Bockstael, N., K.E. McConnell, I.E. Strand, and F. Arsanjani. "Sample Selection Bias in the Estimation of Recreation Demand Functions." *Land Econ.* 66(February 1990):293–302

Carson, R., M. Hanemann, and T Wegge. "Southcentral Alaska Sport Fishing Economic Study." Jones and Stokes Associates, Sacramento, CA, report prepared for Alaska Department of Fish and Game, 1987.

Caulkins, P., R Bishop, and N. Bouwes. "A Comparison of Two Travel Cost Models for Valuing Lake Recreation." Mimeo, AEA annual meeting, 1984.

____. "The Travel Cost Model for Lake Recreation: A Comparison of Two Methods for Incorporating Site Quality and Substitution Effects." *Amer. J. Agr. Econ.* 68(May 1986):291–97.

Creel, M , and J.B. Loomis. "Theoretical and Empirical Advantages of Truncated Count Data Estimators for Analysis of Deer Hunting in California." *Amer. J. Agr. Econ.* 72(May 1990):434–45.

Feenberg, D., and E. Mills. *Measuring the Benefits of Water Pollution Abatement* New York: Academic Press, 1980.

Hanemann, W.M. "Discrete/Continuous Models of Consumer Demand." *Econometrica* 52(May 1984):541–60.

Hellerstein, D.M "Using Count Data Models in Travel Cost Analysis with Aggregate Data." *Amer. J. Agr. Econ.* 73(August 1991):860–66.

Hellerstein, D.M., and R Mendelsohn. "A Theoretical Foundation for Count Data Models." *Amer J. Agr. Econ.* 75(August 1993):604–11.

Kling, C.L. "Comparing Welfare Estimates of Environmental Quality Changes from Recreation Demand Models " *J. Environ. Econ. Manage.* 15(September 1988):331–40.

Kling, C.L., and C J. Thomson. "The Implications of Model Specification for Welfare Estimation in Nested Logit Models " Paper presented at the W-133 meetings, Tucson, February 1994.

Lee, L.F., and M.M Pitt "Microeconometric Demand Systems with Binding Nonnegativity Constraints: The Dual Approach." *Econometrica* 54(November 1986):1237–42.

____. "Microeconometric Models of Rationing, Imperfect Markets, and Nonnegativity Constraints." *J. Econometrics* 36(September 1987):89–110.

Loomis, J.B , and F Ward "The Travel Cost Demand Model as an Environmental Policy Assessment Tool: A Review of the Literature." *W J. Agr. Econ.* 11(December 1986):164–78

McFadden, D. "A Method of Simulated Moments for Estimating Discrete Response Models Without Numerical Integration." *Econometrica* 57(September 1989):995–1026.

Morey, E.R "The Demand for Site-Specific Recreational Activities A Characteristics Approach." *J. Environ Econ. Manage.* 8(December 1981):345–71

____ "The Choice of Ski Areas. Estimation of a Generalized CES Preference Ordering with Characteristics." *Rev. Econ. Statist.* 66(November 1984):584–90.

____. "Characteristic, Consumer's Surplus and New Activities: A Proposed Ski Area." *J. Pub. Econ.* 26(March 1985):221–36.

Morey, E.R., and R.D. Rowe. "Assessment of Damages to Anglers and Other Recreators from Injuries to the Upper Clark Fork River Basin " Report prepared for the State of Montana, Natural Resources Damage Program, December 1993.

Morey, E.R , and W.D Shaw. "An Economic Model to Assess the Impact of Acid Rain: A Characteristics Approach to Estimating the Demand for and Benefits from Recreational Fishing " *Advances in Applied Microeconomics Vol. 5: Recent Developments in the Modelling of Technical Change, and Modelling Demand for and Valuation of Recreation Resources*. Link and Smith, eds. Greenwich CT JAI Press, 1990.

Morey, E.R., R.D Rowe, and M. Watson. "A Repeated Nested-Logit Model of Atlantic Salmon Fishing." *Amer. J. Agr. Econ.* 75(August 1993):578–92.

Morey, E.R., W.D. Shaw, and R.D. Rowe. "A Discrete Choice Model of Recreational Participation, Site Choice, and Activity Valuation When Complete Trip Data Are Not Available." *J. Environ Econ. Manag* 20 (March 1991):181–201.

Morey, E.R., W.D. Waldman, D. Assane, and W.D. Shaw. "Specification and Estimation of a Simple Model of Interiors, Corners and Boundaries." Discussion paper, Department of Economics, University of Colorado, 1992.

Pakes, A., and B. Pollard. "Simulation and the Asymptotics of Optimization Estimators." *Econometrica* 57(September 1989):1027–58

Ransom, M. "A Comment on a Consumer Demand System with Binding Non-Negativity Constraints." *J. Econometrics* 34(March 1987) 355–59

Shaw, D. "On-site Samples' Regression: Problems of Non-negative Integers, Truncation, and Endogenous Stratification." *J. Econometrics* 39(February 1988):211–23.

Smith, V.K. "Selection and Recreation Demand." *Amer. J Agr. Econ.* 70(February 1988):29–36.

Tobin, J "Estimation of Relationships for Limited Dependent Variables." *Econometrica* 26(January 1958):24–36.

Wales, T.J., and A.D. Woodland. "Estimation of Consumer Demand System with Binding Non-Negativity Constraints." *J. Econometrics* 21(April 1983):263–85.

Woodland, A.D. "Stochastic Specification and the Estimation of Share Equations." *J. Econ.* 10(August 1979).361–83.