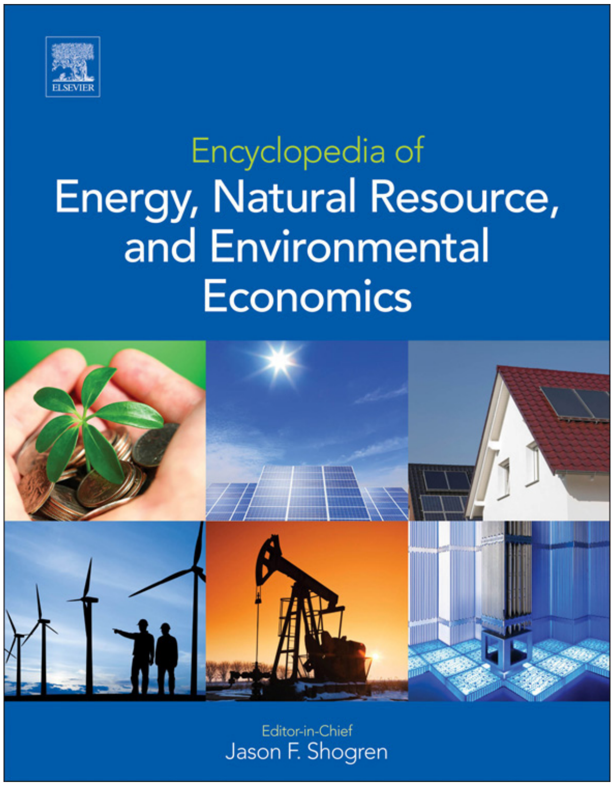

Parsons G.R. (2013) Travel Cost Methods. In: Shogren, J.F., (ed.) Encyclopedia of Energy, Natural Resource, and Environmental Economics, volume 3, pp. 349-358 Amsterdam: Elsevier.

Travel Cost Methods

GR Parsons, University of Delaware, Newark, DE, USA

Introduction

Economists have been concerned with measuring the economic value of recreational uses of the environment for over 50 years. This has been largely motivated by benefit–cost analyses of environmental policies and damage assessments where legal rules call for valuation in circumstances where some harm has been done to the environment. Benefit–cost analysis under the Clean Water Act (USA) is a good example of the former wherein values due to improved water quality, such as better fishing and swimming, are needed to judge the regulatory impact of water quality standards. Damage assessment under the Oil Pollution Act (USA) is a good example of the latter where analysts seek values of lost recreational uses, such as beach closures or lower hunting quality, to establish the size of compensable payments.

Economists have used both revealed and stated preference methods to estimate recreational use values. The travel cost model (TCM) is the primary revealed preference method used in this context. It has been in existence since Harold Hotelling suggested the method in his now famous letter to the director of the National Park Service in 1949. By most accounts, it is one of the success stories in valuing environmental goods. It enjoys broad application in policy settings, and the scholarly literature addressing the theory and empirical method has grown considerably in the past three decades.

The basic insight underlying the TCM is that an individual's 'price' for recreation at a site, such as hiking in a park or fishing at a lake, is his or her trip cost (including out-of-pocket travel and time cost) of reaching that site. Viewed in this way, individuals reveal their willingness to pay for recreational uses of the environment in the number of trips they make and/or the sites they choose to visit. The actual measurement of these values entails some form of demand estimate for recreation trips and, in turn, measurement of consumer surplus.

This article provides an introduction to the TCM and how it is used to value recreation sites and the attributes of recreation sites. It is intended to be comprehensive but will not develop the theory or econometrics underlying each model or address all of the methodological inquiries pertaining to the validity and robustness of the model. It is instead a statement of the current practice in practical terms.

It is common to classify TCMs into three groups: seasonal demand, site choice, and Kuhn–Tucker (KT). The lines between these models have blurred somewhat with the advance of research, but this classification still helps in organizing the material.

The earliest TCMs, dating back to the 1960s, are seasonal demand models that work much like a demand curve for any consumer good. Trip cost is treated as the 'price' of the good and the number of trips taken over a season is treated as the 'quantity demanded.' The simple observation that the closer one lives to a site (lower price), the more trips one takes (higher quantity demanded) is taken as a simple downward-sloping demand curve. To estimate such a curve, one gathers cross-sectional data on the number of trips taken by people living at different distances from the site. Then, by regressing the number of trips on the measured trip cost, a demand relationship is revealed, from which conventional measures of surplus may be derived. Seasonal demand models have proved to be useful for valuing the opening and closure of a site and computing per trip values for use in transfer studies. It is especially useful if policies are focused on a single site or on only a few sites that serve as substitutes for one another.

Site choice models, however, are now the most commonly used form of the TCM. These were introduced in the mid-1980s and were motivated by a need for models that allow for valuation of changes in site quality (e.g., improved water quality at lakes) and to consider recreation demand where the number of sites is large. These models were built during the rapid expansion of random utility theory in the 1980s and 1990s that eventually led to Daniel McFadden's Nobel Prize for work in this area. His Nobel Prize lecture actually included an application using a TCM.

Site choice models consider an individual's choice of visiting one site from among many possible sites on a given choice occasion. While approaches exist for accommodating seasonal demand in site choice models, at their heart, they are based on the choice of a single site during a single choice occasion. Which site a person visits is assumed to be a function of the attributes of the sites (size, quality, access, etc.) and the trip cost of reaching the site. Individuals reveal their relative values of site attributes in the sites they choose. Because trip cost is one of the attributes, an individual's choice reveals the relative values for the attributes of sites in money terms. To estimate a site choice model, one gathers data on the actual site choices made by individuals. Then, usually using some form of multinomial logit in the context of a random utility model where trip cost and site attributes serve as arguments in the utility function, a probabilistic choice model is estimated. As these parameters are hypothesized to come from individual utility functions, they readily accommodate welfare analysis. Site choice models have been useful for valuing quality changes in site attributes, closure of one or more sites in a region, and addition of new sites. Given the proliferation of software that can accommodate discrete choice random utility models and their ability to address many policy issues in a defensible and easy-to-understand way, site choice models have come to dominate the TCM literature.

The final TCM is the KT model. Application of this model came into practice in the early 2000s. KT models are seasonal, but they are set in a probabilistic framework that shares many of the properties of site choice models. In a sense, KT models bring together the strengths of seasonal and site choice models in a unified model. However, it has proven to be computationally more cumbersome than the simpler seasonal demand and site choice models. While its rise has been slow, its use will no doubt increase in the years ahead.

 http://dx.doi.org/10.1016/B978-0-12-375067-9.00002-4

Encyclopedia of Energy, Natural Resource, and Environmental Economics, (2013), vol. 3, pp. 349-358

Like any area of economic research, there are numerous theoretical and empirical issues surrounding application of the TCM. These include, among other things, measuring trip cost (especially the time cost component), dealing with multiple-purpose trips, incorporating time interdependence, treatment of congestion, and defining sites and choice sets. Each of these issues is briefly addressed following a presentation of the three models outlined earlier.

Finally, an area of increasing interest in TCM research is the use of state preference data in combination with revealed preference travel cost data, for example, data on how people report they would change trips if water quality were improved in response to a hypothetical survey question on this matter. This enables an analyst to explore unobserved, yet policy-relevant, changes to recreation sites. Given the more than adequate coverage of this topic elsewhere in this volume in connection with several valuation models, it is only briefly addressed in this article.

Seasonal Demand Models

Introduction and Theory

The TCM in its 'seasonal demand' form is the traditional variety. It considers demand for use of a site over an entire season. It treats trips to a site as the 'quantity demanded' and trip cost as the 'price.' This gives a conventional demand function

$$r_n = f(p_n, ps_n, z_n) \qquad [1]$$

where r_n is the number of trips taken by individual n to the site during the season, p_n is the price or trip cost for individual n to reach the site (travel and time cost), ps_n is the vector of trip cost substitute sites, and z_n is the vector of individual characteristics believed to influence the number of trips taken in a season (e.g., age and income). This form may be readily derived from conventional utility theory or household production theory. It is usually done using separate budget and time constraints to explicitly show the opportunity cost of time in trip cost.

Consumer surplus for access to the site for a season is

$$cs_n = \int_{p_n^0}^{p_n^*} f(p_n, \mathbf{ps}_n, z_n)\mathrm{d}p_n \qquad [2]$$

where p_n^0 is the current trip cost to the site and p_n^* is the 'choke price' – the trip cost at which demand for trips goes to zero for individual n. If the site were lost, cs_n is the loss in welfare – sometimes called 'access value.' Analysts sometimes report mean *per trip* values that take the form cs_n/r_n. Exact Hicksian measures of surplus may also be derived as usual.

Estimation

In estimation, an analyst gathers cross-sectional data on a sample of individuals for a given season: number of trips taken to the site, trip cost, and other relevant demand shifters. Then, via the spatial variation in trip cost (people living at different distances from the site have different 'prices' for trips), one estimates an equation like eqn [1].

The choice of functional form for estimation has been the subject of inquiry since the methodology was first developed. Earlier functional forms were continuous – linear, log-linear, log-log, etc. Most modern forms are from the family of count data models – Poisson, negative binomial, zero-inflated, hurdle, and so on. Count data models are designed for analyses with a nonnegative integer-dependent variable and are quite versatile for handling truncation, large number of zero trips in the data, and preference heterogeneity. This has made them popular for seasonal demand function estimation. A Poisson model is the simplest form. An individual's probability of making y trips to a site in a given season is

$$pr(r_n = y) = \frac{\exp(-\mu_n)\mu_n^y}{y!} \quad \text{where } y = 0, 1, 2, \ldots, \quad \text{where}$$
$$\mu_n = \exp\left(\beta_p p_n + \beta_{ps} ps_n + \beta_z z_n\right) \qquad [3]$$

μ_n is the expected number of trips taken by a person n. It serves as the 'demand' expression for eqn [1], and the Poisson model puts it in probabilistic form. The parameters in eqn [3] are estimated by maximum likelihood where each person's probability of taking the number of trips actually taken is used as an entry in the likelihood function. Seasonal consumer surplus for a person n in the Poisson form is $\hat{c}s_n = \hat{r}_n / -\hat{\beta}_p$, where 'hatted' values denote estimates. Per trip consumer surplus is a simple constant $\hat{c}s_n/\hat{r}_n = 1/-\hat{\beta}_p$.

An undesirable feature of the Poisson model is an implicit constraint that the mean and variance of r_n are equal. If upon testing the data fail to support this assumption, it is common to use a negative binomial model to relax this constraint. As testing usually shows that equality of mean and variance does not hold, the negative binomial form is frequently used.

On-Site and Off-Site Samples

One of the more important decisions an analyst makes when estimating a seasonal demand model is whether to gather data on-site or off-site. Often, the number of visitors to a particular recreation site is a small fraction of the general population. If so, sampling the general population may require a large number of contacts to form a reasonable sample size of recreational users. On-site data have the advantage that every individual 'intercepted' has taken at least one recreation trip. In this way, gathering data on-site is often cost-effective. However, there are at least two disadvantages of on-site data: endogenous stratification and truncation. As individuals taking more trips over a season are more likely to be drawn for inclusion in the sample, there is oversampling in direct proportion to the number of trips one takes over a season (e.g., a person taking two trips is twice as likely to be sampled as a person taking one trip). Estimation that fails to account for this effect will give biased parameter estimates for the general population. At the same time, the analyst never observes individuals taking zero trips in the season, so there is no direct observation at the 'choke price' on the demand function, which is important in the computation of consumer surplus. Both endogenous stratification and truncation are easily corrected econometrically. One can show in the Poisson form that simply using $y-1$ instead of y in estimation in eqn [3] corrects for both effects. The correction is somewhat more complicated in more complex forms such as negative binomials, but it is possible there as well.

Off-site sampling has the advantage that nonparticipants are observed. This presumably makes for more accurate estimation of the choke price and hence estimation of consumer surplus. In some models, nonparticipants get special treatment, wherein the analyst estimates a two-part model: First, the decision to take a recreation trip or not (the participation decision) and second, the number of trips to take (the frequency decision). These models are also members of the count data family of models and are known as hurdle and zero-inflated Poisson models. The model uses, more or less, a simple bivariate choice model for the participation decision and a Poisson model of one form or another for the frequency decision. If one believes that participation and frequency are governed by different decision processes, these models are beneficial. A zero-inflated count model applied in our context has the form

$$pr(r_n = 0) = \varphi_n(p_n, ps_n, z_n) + (1 - \varphi_n(p_n, ps_n, z_n))\exp(-\mu_n)$$
$$pr(r_n = y) = (1 - \varphi_n(p_n, ps_n, z_n))\frac{\exp(-\mu_n)\mu_n^y}{y!} \quad [4]$$

The first equation is the probability of observing a person take zero trips in the season, and the second equation is the probability of observing a person taking one or more trips in the season where μ_n is the same as shown in eqn [3] and φ_n is a simple bivariate logit model. The first term in the first equation models the participation decision, the probability of a person being someone who engages in the type of recreation under study at all. The second term in the first equation captures those who engage in the type of recreation under study but happen to not make a trip in the current season, the probability of being a participant but taking no trips. The second equation is the frequency decision for those taking at least one trip, the probability of taking r_n trips in the season given that that person is a participant. Again, estimation is by maximum likelihood wherein probabilities from eqn [4] are loaded for each observation according to actual choices. Seasonal and per trip consumer surplus in a zero-inflated model have the forms $(1 - \hat{\varphi}_n)\hat{\mu}_n/-\beta_p$ and $1/-\beta_p$, where the 'weighting' term, $(1 - \hat{\varphi}_n)$, accounts for participation.

Valuing Quality Changes and Multiple Site Models

Although the strength of seasonal demand models is not in the valuation of site attributes, such as water quality or acres of open space, there are empirical approaches using the model for this purpose. The most popular approach uses contingent behavior response data in combination with the trip data. For example, in addition to asking respondents to report total number of trips over a season, one also asks them to report the total number of trips they would have taken 'if the expected catch rate of fish at the site had been ½ the current rate' or 'if the width of the beach had been twice its current width.' Then, using the newly created contingent trip count, a second TCM under the new hypothetical conditions is estimated. The area between this new demand curve and the original – the difference in consumer surplus estimates in the two conditions – is an estimate for the value of the attribute change. These demand models are typically estimated simultaneously and often with logical parameter and error term restrictions. This approach accepts the condition of 'weak complementarity' in the behavior model – that a person receives utility from the attributes of a site only by visiting the site.

It is also possible to observe demand function shifts using actual trips to several different sites that vary in attribute quality. 'Pooling' or 'stacking' multiple sites in this way allows the analyst to enter site attributes in eqn [1], estimate parameters for site attributes, and then use that estimate to calculate welfare as described in the contingent behavior case above. These cross-sectional models have variously been called varying parameter models, pooled models, and stacked multiple site models. The basic problem underlying these models is that they fail to integrate site choice across the set of sites under consideration in a meaningful way. For this reason, these models have largely fallen out of favor, although applications do still appear from time to time in the published literature.

The final form of the seasonal demand model to consider is a multiple-site model in which a system of count demand equations allowing for substitution across sites and utility theoretic restrictions is developed. There are a number of applications along these lines, but this has largely been confined to settings with a few sites and has focused on access values instead of valuing site attributes. For the most part, these models have given way to site choice random utility models and KT models as a way of integrating many sites into the decision model and to conduct welfare analysis for changes in site attributes. The KT model is really the state-of-the-art demand system model and is presented in the section 'Kuhn–Tucker Models.'

Site Choice Models

Introduction

The most commonly used TCM in the literature today is a model of recreation site choice based on the random utility theory. Known as the random utility maximization (RUM) model, it has proven to be quite versatile for measuring access value (e.g., opening or closure of one or more recreation sites) and quality changes at one or more sites (e.g., improved water quality, wider beaches, and increased bag rate for hunting). Its appeal hinges on its ability to handle many sites and substitution among sites in a plausible and easily estimable way. The behavioral basis underlying the model is also easily understood and intuitive making it all the more attractive for policy analysis and damage assessment.

Theory

The time frame in a site choice model is a single choice occasion, usually a day, in which an individual makes one recreation trip. The individual is assumed to face a set of S possible sites for a trip. Each site i $(i = 1, 2, \ldots, S)$ is assumed to give individual n $(n = 1, 2, \ldots, N)$ some utility U_{in} on a given choice occasion. The utilities are assumed to be a function of the trip cost of reaching the site and attributes of the site, such as natural amenities, water quality, size, and access. As before, trip cost includes travel and time cost.

Letting p_{in} be trip cost, q_i and $\tilde{q}_i$ be vectors of site attributes that may or may not share some of the same terms, and $\tilde{z}_i$ be a vector of individual characteristics, *site utility* for person n at site i is

$$U_{in} = \beta_{tc}p_{in} + \beta_q q_i + \beta_{qz}\tilde{q}_i\tilde{z}_n + \varepsilon_{in} \quad [5]$$

Typically, site utility is linear as shown or some form with a nonlinear transformation of characteristics (such as the log of beach width) that maintains linearity. The coefficient on trip cost is the marginal utility of income as it describes how site utility changes with a decrease in income if a trip is taken. Site utility often includes site-specific constants in the vector q_i capturing 'average' differences across sites missed by the vector of attributes. The vector $\tilde{z}_n$ is interacted with $\tilde{q}_i$ to capture individual heterogeneity. For example, 'boat ramp' as a site attribute in $\tilde{q}_i$ might be interacted with 'boat ownership' as an individual characteristic in $\tilde{z}_n$, if one believes that boat ramps only matter to people who own a boat. Individual characteristics cannot be entered alone because they are invariant across sites. Individual characteristics may, however, be interacted with a site-specific constant if there is a compelling reason that an individual characteristic affects a person's proclivity to visit one site over another. The error term, ε_{in}, captures site attributes and individual characteristics that influence site choice but are unobserved by the analyst.

On any given choice occasion, an individual is assumed to choose the site with the highest site utility giving *trip utility* of the form

$$V_n = \max(U_{1n}, U_{2n}, \ldots, U_{Sn}) \quad [6]$$

Trip utility, V_n, is the basis for welfare analysis in the RUM model. It is used to value loss or gain of sites (access values) and changes in site attributes. For example, consider an oil spill that closes sites 1 and 2. Trip utility with the closures becomes

$$V_n^{\text{closure}} = \max(U_{3n}, U_{4n}, \ldots, U_{Sn}) \quad [7]$$

where sites 1 and 2 have been dropped from the choice set. Trip utility declines from V_n to V_n^{closure}.

A similar expression can be generated for a change in site quality at one or more sites. Suppose the water quality at sites 2 and 3 is improved through some regulation. If so, trip utility for person n becomes

$$V_n^{\text{clean}} = \max(U_{1n}, U^*_{2n}, U^*_{3n}, U_{4n}, \ldots, U_{Sn}) \quad [8]$$

where U^*_{2n} and U^*_{3n} denote the now higher utility due to the improved quality. In this case, trip utility increases from V_n to V_n^{clean}. In both cases, the change in utility is monetized by dividing the change by the coefficient on trip cost $-\beta_p$, which is our marginal utility of income, in eqn [5]. This gives the following compensating variation (also equivalent variation) measures for changes in trip utility

$$w_n^{\text{closure}} = (V_n^{\text{closure}} - V_n)/-\beta_p \text{ and}$$
$$w_n^{\text{clean}} = (V_n^{\text{clean}} - V_n)/-\beta_p \quad [9]$$

These are changes in welfare on a *per trip per person* basis.

Estimation

Because the error term, ε_{in}, on each site utility is unknown to researchers, the choice is treated as the outcome of a stochastic process in estimation. By assuming some explicit distribution for the error terms in eqn [5], each person's probability of visiting a site can be expressed in some form. The simplest is to assume that the error terms are independently and identically distributed (iid) type 1 extreme value random variables. This gives a closed form expression, a multinomial logit, for the choice probabilities. Each person's probability of choosing any site k from the set of S sites in the multinomial logit from is

$$pr_n(k) = \frac{\exp\left(\beta_{tc}p_{kn} + \beta_q q_k + \beta_{qz}\tilde{q}_k\tilde{z}_n\right)}{\sum_{i \in S} \exp\left(\beta_{tc}p_{in} + \beta_q q_i + \beta_{qz}\tilde{q}_i\tilde{z}_n\right)} \quad [10]$$

The parameters are estimated using data on actual site choices and maximum likelihood with the logit probabilities in eqn [10] – so a person's entry into the likelihood function is the probability of visiting the site actually chosen on a given choice occasion. Because researchers proceed as though choices are the outcome of a stochastic process, *trip utility* in eqns [6]–[8] is also stochastic. *Expected trip utility* is used as an estimate of V_n in empirical work. Using the assumption of iid type 1 extreme value distributions for the error terms gives each individual's *expected trip utility* as

$$E(V_n) = E\{\max(U_{1n}, U_{2n}, \ldots, U_{Sn})\}$$
$$= \ln\left\{\sum_{i=1}^{S} \exp(\beta_p p_{in} + \beta_q q_i + \beta_{qz}\tilde{q}_i\tilde{z}_n)\right\} + C \quad [11]$$

$i = 1$, where C is some unknown additive constant. It is a manifestation that the absolute level of utility is unmeasurable, and as it is shared and constant across all expected trip utilities, it is of no practical relevance in welfare analysis. $E(V_n)$ is often referred to as the 'log-sum' and is the empirical form of V_n used in welfare analysis. The steps in such an analysis are straightforward: estimate the parameters of site utility, use the parameters to construct expected trip utilities with and without some resource change using eqn *[11]*, and finally compute per trip losses per person substituting $E(V_n)$ for V_n in eqn [9] and with the estimates of $-\beta_p$ used to monetize the change in expected utility (note that C discussed above drops out when one differences the equations). In some cases, rarely, however, researchers will consider site utilities that are nonlinear in trip cost, allowing for nonconstant marginal utility of income and empirical forms of eqn [9] that are not a closed form. In this case, welfare is calculated using numerical methods.

One of the major drawbacks of the multinomial logit model is the restrictive way in which substitution occurs. Since site substitution is the pathway through which welfare effects are captured, it is important to handle it in as realistic a way as possible. The multinomial logit model assumes that the closure or decline in quality at one or more sites leads to a proportional increase in the visitation to all other sites – their shares remain in fixed proportion. This property, known as the independence of irrelevant alternatives, is usually unrealistic. For this reason, economists have turned almost entirely to alternative forms that allow for more realistic patterns of substitution. This is achieved, at least in a stochastic sense, by allowing for correlated error terms across the sites in eqn [5]. There are two common methods that allow for such correlation: nested and mixed logit. These forms dominate the travel cost random utility model literature. Both are generalizations of the multinomial logit model outlined above and follow the same steps outlined there.

The nested model has been used since the introduction of travel cost RUM models. Nested models place sites that share unobserved characteristics in a common 'nest' under the assumption that they serve as better substitutes for one another than sites outside the nest. It also renders a closed form choice probability similar to the conditional logit model, but it includes a new parameter for each nest that captures the degree of substitution among the sites within that nest. Researchers often nest sites by proximity (e.g., grouping sites in the same region together), resource type (e.g., grouping lakes and rivers in separate nests), and purpose (e.g., grouping trout and bass fishing trips separately). The option of purpose actually expands the choice model to include than just site choice.

The mixed logit model (or random parameters logit) is a more flexible approach for enriching the patterns of substitution. It allows for more, and overlapping, substitution structures and can also be easily configured to mimic what a nested logit model does. The mixed logit model induces correlation among the error terms by allowing parameters on the site attributes to have random components or dispersion terms. The random components then become a part of the error term in site utility, and this, in effect, causes correlation among site utilities. Consider an example: A site utility includes a dummy variable for 'state park' to distinguish community-run beaches from state-level park beaches. As state parks are likely to share unobserved attributes, one might estimate the model with a random parameter on 'state park'. If so, the estimated dispersion term, the variance, on 'state park' would cause all the sites sharing the attribute to be correlated as their shared error component would move in concert. The larger the dispersion, the greater the degree of correlation and hence substitutability among the state park sites. This, in turn, is implicitly captured in the welfare analysis.

Because models are estimable with a large number of random terms, the possible patterns of correlation are almost endless, making mixed logit an obvious choice for the recreation applications with welfare analysis. Unlike the multinomial and nested logit models, the mixed logit model does not yield a closed-form choice probability. Instead, it uses simulated probability methods to solve for choice probabilities that present themselves as integrals and yields estimates for the mean and dispersion of designated parameters. In some circumstances, the disturbance term is used to interpret the degree of taste heterogeneity in the sample. In this case, the greater the dispersion, the greater the unobserved heterogeneity in the sample. Welfare analysis with mixed logit uses a log-sum term exactly like eqn [11], but it requires the use of a simulated log-sum as some or all of the parameters in the equation vary by some known distribution. Due to its flexibility and now widespread presence in standard econometric packages, mixed logit has become extremely popular in travel cost random utility model applications.

Finally, like the seasonal demand model, the major decision of on-site versus off-site data affects the econometrics used to estimate the model. Again, on-site sampling may be a cost-effective way of obtaining trip data but the data must be adjusted to account for the relative amount of time spent sampling at each site; otherwise, the choice probabilities will reflect in part (perhaps in large part) the relative extent of sampling at each site instead of the relative preferences for each site. To correct for on-site sampling bias in a meaningful way, one needs to know, or at least should be able to estimate, the proportion of trips to each site in the population. This population proportion can be used to adjust or weigh the sample choice probabilities in estimation. An alternative solution is to design a sampling strategy that applies equal sampling pressure across all sites, independent of their popularity. Off-site data are 'cleaner' in the sense that no such adjustment is needed.

Seasonal Forms

Site choice models, at their core, are occasion-based, centering on an individual trip. Oftentimes, analysts are interested either in the seasonal implications of a policy change or in resource changes that may engender changes in the number of trips taken (e.g., fewer fishing trips if catch rates decline). The site choice model described earlier disallows taking fewer or more trips over a season or no trip on a single choice occasion as the model is conditioned on taking a trip.

There are essentially two methods used to modify the basic choice model to make it seasonal and incorporate the possibility of adjusting the number of trips taken over a season: a repeated discrete choice model with a no-trip alternative and a linked seasonal demand model.

The repeated choice model simply adds a no-trip utility to the individual's choice. This typically takes the form

$$U_{0n} = \delta_0 + \delta_z z_n + \varepsilon_{0n} \quad [12]$$

where z_n is a vector of individual characteristics believed to influence whether or not a person takes a trip on a given choice occasion (z_n usually different from $\tilde{z}_n$). This might include age, family composition, years engaged in recreation, and so on. Each person now has $S+1$ choices on each choice occasion: visiting one of the S sites or taking no trip. The model is made 'seasonal' by simply repeating it for every choice occasion in the season, where the choice probabilities now include no-trip as one of the alternatives. The log-sum becomes an *expected choice occasion utility* instead of *expected trip utility* with the form

$$\begin{aligned} E(V_n) &= E\{\max(U_{0n}, U_{1n}, \ldots, U_{Sn})\} \\ &= \ln\left\{\exp(\delta_z z_n) + \sum_{i=1}^{S} \exp(\beta_p p_{in} + \beta_q q_i + \beta_{qz}\tilde{q}_i\tilde{z}_n)\right\} \\ &\quad + C_n \end{aligned} \quad [13]$$

Per trip welfare changes are calculated as before (see eqn [9] and discussion following eqn [*11*]) but become *per choice occasion per person.* Seasonal estimates of welfare change are simply per choice occasion values multiplied by the number of choice occasions in a season, $W_n = M \cdot w_n^{co}$, where w_n^{co} denotes the per choice occasion value and M the number of choice occasions.

Usually, an analyst will have data on trips over an entire season without knowing the specific date for each trip. If so, if a person took T_n trips, on M choice occasions, each of the T_n trips would enter the likelihood function as the probability of taking a trip and each of the $M - T_n$ no-trips would enter as the probability of taking no trip. This expands the data set considerably in estimation. In nested logit, the S sites are usually placed in a nest separate from the no-trip choice. In mixed logit, no-trip utility usually includes its own alternative-specific constant as shown in eqn [12] and will be treated as a random parameter. It is also desirable to allow for correlation of utilities across choice occasions in repeated models in estimation.

The alternative approach for introducing a seasonal dimension into a site choice model is a pseudo–seasonal demand model that 'links' the number of trips taken over a season with the expected trip utility from the site choice model. The linked model has the form

$$T_n = f\left(E(V_n)/-\beta_p, z_n\right) \qquad [14]$$

where T_n is the number of trips taken by a person n over the season, $E(V_n)/-\beta_p$ is the expected trip utility estimated in a site choice model (eqn [11]) divided by the estimated coefficient on trip cost from the same model, and z_n is a vector of individual characteristics. $E(V_n)/-\beta_p$ is the expected value of a recreation trip for person n on any given choice occasion.

One would expect the number of trips a person takes to increase with the expected value of a trip. In this way, the model can be used to predict how the number of trips over a season changes with changes in site characteristics or the loss/addition of sites in the choice set. For example, the expansion of designated open space at one or more recreation sites would increase predicted $E(V_n)$ from the site choice model, which, in turn, would increase the number of predicted trips in a linked model, thereby picking up seasonal adjustments in the number of trips taken. The linked model is typically estimated using a count data model, either stepwise or simultaneously with the site choice model. Seasonal changes in welfare in Poisson or negative binomial form can be shown to be $\Delta\hat{T}_n/\hat{\gamma}$ for person n, where $\Delta\hat{T}_n$ is the change in trips due the resource change and $\hat{\gamma}$ is the parameter estimate on expected trip utility in the linked model.

The model is admittedly ad hoc in the sense that it is not built from a consistent utility theoretic framework at the site choice and trip demand levels. Nevertheless, it has proved to be quite versatile and is usually easy to estimate. The repeated choice model can be written in a linked form as $f(.)=M\cdot(1-pr(\text{no-trip}))$, where M is the number of choice occasions in a season and pr(no-trip) is the probability of taking no trip in the repeated choice model. In this way, the two models, while ostensibly different, can be seen as simply different functional forms for the seasonal component of a site choice model.

Kuhn–Tucker Models

The KT model is the most recent of the travel cost models. The KT model in some ways brings together the best attributes of the seasonal demand and site choice models by modeling site choice and total trips over a season in a utility-consistent way. Although introduced into the recreation demand literature over 10 years ago, it has not seen particularly wide use, especially when compared with the site choice models mentioned in the previous section. This seems to be largely due to the complexity and computation difficulties one often encounters when estimating a KT model. Nevertheless, the KT model is at the cutting edge of travel cost demand modeling and is likely to see increased use in due time.

In the KT model, individuals are assumed to maximize a seasonal direct utility function subject to a usual budget constraint. To ease notation, let us assume one site for now. An individual's choice is defined by

$$\max_{r,a}\{u(r,a;q,z,\varepsilon,\gamma)\} \text{ s.t. } p\cdot r + a = y, r \geq 0 \qquad [15]$$

where r is the number of trips taken to the site, a is a numeraire good with price one, $\boldsymbol{q}$ is a vector of site attributes, $\boldsymbol{z}$ is a vector of individual characteristics, ε is an error term, and $\boldsymbol{\gamma}$ is a parameter vector to be estimated. The use of the subscript n denoting individuals has been suppressed. In the budget constraint, there are p trip costs and y income. The KT first-order conditions for utility maximization then are

$$\frac{\partial u(r, y-p\cdot r;q,z,\varepsilon,\gamma)/\partial r}{\partial u(r, y-p\cdot r;q,z,\varepsilon,\gamma)/\partial a} \leq tc; \quad r \geq 0;$$
$$r\cdot[\partial u(r,y-p\cdot r;q,z,\varepsilon,\gamma)/\partial r - tc\cdot\partial u(r,y-p\cdot r;q,z,\varepsilon,\gamma)/\partial a] = 0 \qquad [16]$$

These are the usual complementary slackness conditions that allow for both corner (zero trips) and interior (nonzero trips) solutions. The trick to making the KT model operational is to select a form of the utility function that allows one to rewrite the conditions in eqn [16] as

$$\varepsilon \leq g(r,p,y,q,z,\gamma); \quad r > 0; \quad r\cdot[\varepsilon - g(r,p,y,q,z,\gamma)] = 0 \qquad [17]$$

Equation [17] makes the model empirical because it allows the analyst to write realizations from a data set (number of trips taken by respondents) in probabilistic terms that can be entered into a maximum likelihood function for estimation. For a given individual n, if $r_n=0$, then $\varepsilon_n<g_n$, whereas if $r_n>0$, $\varepsilon_n=g_n$. So, if £ has some known distribution, an explicit form for the probabilities for each observation can be used in estimation. Some applications, for example, have used iid type 1 extreme value error terms, such as those used in the multinomial logit. The estimated parameters $\hat{\gamma}$ are then used to construct the fitted direct utility functions for each individual, which, in turn, may be used to estimate exact measures of surplus for some hypothetical changes in site attributes. The analog to the expected trip utility in eqn [7] for the RUM model is the maximum indirect seasonal utility corresponding to the problem in eqn [15] or

$$v_n = \max\left\{v_{1n}(p,y,q,z,\varepsilon,\gamma)_1, v_{0n}(p,y,q,z,\varepsilon,\gamma)\right\} \qquad [18]$$

where v_{1n} is the indirect seasonal utility conditioned on taking trips and v_{0n} is the indirect seasonal utility of not taking a trip. Eqn [18] simultaneously defines whether or not a person is a participant and, if so, how many trips are taken over the entire season. In welfare analysis for a change in site attributes, compensating variation is just the value of ΔW_n that solves

$$\max\left\{v_{1n}(p,y,x,z,\varepsilon,\gamma)_1, v_{0n}(p,y,x,z,\varepsilon,\gamma)\right\}$$
$$= \max\left\{v_{1n}(p,y-\Delta W_n,x^*,z,\varepsilon,\gamma)_1, v_{0n}(p,y-\Delta W_n,x^*,z,\varepsilon,\gamma)\right\} \qquad [19]$$

As the error term ε is random, the welfare change ΔW_n is also random. Also, because each element in eqn [19] is itself a maximum value function, there is no closed form solution to ΔW_n like the log-sum. It must be solved numerically using repeated draws on the assumed distribution for the error term. Importantly, the outcome of this process allows each individual in the sample to adjust visits to the site and their frequency. When more than one site is included in the model, eqn [18] includes 2^S available *combinations of sites* that might be visited over the season (including no trip), and each combination has an optimal number of trips for its included sites. In this way, attribute changes and site closures can lead to adjustments in the sites visited and the number of trips taken to each site.

As noted at the outset, due to the computation complexity of the KT model, it has not been used as widely as expected. Also, to make it operational, the form of the utility functions used has been rather restrictive. Still, unlike the seasonal demand and site choice models, the KT model provides utility theoretic consistency between site choice and trip frequency and allows for substitution among sites in the traditional (through cross-price terms) and stochastic (through error term correlation) ways. The model has also been applied in settings with a large number of sites and with numerous site characteristics.

Issues and Complications

This section addresses several perennial topics that complicate estimation and, in some cases, interpretation and use of the results in TCMs. These include multiple-purpose and overnight trips, measuring time cost, intertemporal substitution, choice set formation, and congestion.

Multiple-Purpose and Overnight Trips

Sometimes the purposes of a trip will extend beyond recreation at the site. For example, a person may visit family and friends, or go shopping, or even visit more than one site on a recreation outing. In these instances, the trip is producing (i.e., the trip cost is buying) more than single-site recreation, and it is no longer clear whether the simple travel cost paradigm applies.

For this reason, researchers often confine their analysis to day trips where multiple purposes are less likely to occur. This is sometimes done by either focusing on trips made within a day's drive from a person's home (assuming that these will largely be day trips) or by identifying day trips through a survey question. In some cases in the survey, the analyst identifies single-purpose day trips or day trips where recreation is the primary purpose. Sometimes, 'other purposes' are handled as an attribute of a site. For example, nearby shopping may be variable in a beach choice model. This, in effect, expands the nature of the recreation experience.

Expanding the model to overnight trips is problematic for a number of reasons. There are more costs to estimate (e.g., lodging at all sites). Length of stay can vary significantly over the sample (e.g., some people stay one night, others for 2 weeks.) The relevant choice set is likely to be considerably larger. For example, for a household in the United States, the set of substitutes for a week-long beach vacation may include all beaches in the United States and even beyond. Also, if people use long trips as a 'getaway,' nearby sites with low trip cost may be undesirable. Greater trip cost then, at least over some range, would be viewed as a positive attribute, complicating 'price' in the simple TCM. Finally, many overnight trips will be multiple-purpose/multiple-site excursions wherein the individual transits from one site to the next obviously straining the TCM paradigm.

In cases where individuals visit multiple sites on a single trip, one of the more promising approaches is to redefine a site such that there are 'single-site' sites and 'multiple-site' sites and then proceed with the logic of the TCM. Trip cost would be recalculated for a site with multiple sites by accounting for the costs of visiting all the sites on one trip. In a site choice model, one can think of this as a portfolio choice problem wherein each person chooses a portfolio of sites on any given trip. Characteristics of the portfolio would simply be alternative-specific constants for each site in the portfolio.

Measuring Travel and Time Costs

Trip cost is measured as the sum of travel and time cost plus any other expenses necessary to make the recreation trip possible. Travel cost includes fuel and depreciation of the owner's vehicle. In some instances, analysts will ignore the depreciation costs as inconsequential. In either case, travel cost is typically measured using round-trip distance from home to site times some standard cost per mile of operating a vehicle. Distance from a person's home to the site is usually calculated using a standard over-road software such as PC Miler. It should be noted that this has to be done to all sites in a persons' choice set. An alternative is to use an individual's reported trip cost from a survey question. This has the advantage of being the 'perceived' cost but the disadvantage of measurement/reporting error in the survey and the complication of usually having it reported only for the site actually visited by the respondent.

Measuring the time cost component is a thorny issue. In a world where everyone has a continuous labor-leisure budget constraint, the wage rate is an ideal value of a person's opportunity cost of time for a recreation trip. But, many (most?) individuals simply do not fit this prototype. If a person is a retired person, a student, a homemaker, an unemployed person, or is paid a fixed annual salary to work 40 h per week, there is no clear forgone wage and the opportunity cost of time is not so obvious. There are essentially two ways economists have handled this issue: using a 'wage-analogy' or inferring the value of time directly in the recreation choice.

The 'wage-analogy' is ad hoc but is the most common. One simply divides a person's annual income by the number of hours worked in a year (usually 2000) and uses this as a 'wage.' As people are not on the continuous labor-leisure budget constraint described earlier, this estimate is only loosely tied to theory. In the final calculation, analysts typically use one-third of this calculated wage as the estimated value of time. There is evidence from a number of sources that this is a reasonable adjustment. The mode choice literature in transportation studies, for example, supports this adjustment. Another reason given for using less than the full wage is that the trip to a recreation site itself may be of value. A nice ride through the country side, for example, may be a desirable part of the trip. For these reasons, albeit highly imperfect, the tradition of one-third of the wage continues to be used in applied work.

Inferring the value of time directly in a travel cost model is done by entering out-of-pocket travel cost and time separately in the model. In the random utility model in eqn [5], $\beta_p p_{in}$ is replaced with $\beta_{tc}tc_{in}+\beta_{tm}tm_{in}$, where tc_{in} is the out-of-pocket travel cost measured in money terms and tm_{in} is the simple round-trip time. In this way, the researcher is not explicitly placing an ex ante value on time. Instead, time is being accounted for in the analysis without prior explicit restriction and can be valued implicitly as β_{tm}/β_{tc}, the relative contribution of travel cost and time to a person's utility. Heterogeneity in the value of time can be accounted for in the usual way by interacting tm_{in} with attributes of individuals one believes may govern differences (e.g., income or not working full-time). Another strategy along these same lines is to identify in advance people who work for an hourly wage and have a flexible work schedule and then value their time directly with the wage while treating only those not working for an hourly wage, as previously described. The chief drawback of inferring time value within the analysis is that the tc_{in} and tm_{in} are often highly correlated as both are determined in part by one's distance from a site.

Despite some other creative efforts to value time, the two methods outlined continue to dominate applications in the literature. Neither is particularly satisfying, but there has been no compelling reason to abandon either in favor of a new approach.

Intertemporal Substitution

The bulk of the literature and most of the active research on TCMs ignore any dynamic aspect of decision making, yet it is hard to deny that it is important. Dynamic elements allow people to substitute sites over time and allow experiences early in the season or perhaps in the last season (such as a good catch rate of fish) to affect the choice of site and number of trips in the current season. Individuals may even base current site choices on expectations about future trips.

The repeated site choice model is, in principle, set up for just such an analysis as it considers an individual's trip choice day by day over a season. Nevertheless, few applications consider interdependence in this framework. The typical analysis treats each trip choice as independent of the previous and upcoming choices and takes no account of temporal characteristics (such as weather and day of the week). There are two good reasons for this. First, the data are more difficult to collect. To gather trip data by date of trip usually requires a diary to be maintained by the respondents for the recall to be accurate. This means repeating survey administration (perhaps, monthly throughout the season) or continual reminders to complete a diary sent early in the season. This increases the cost of the survey and leads to sample attrition. Second, there is an inherent endogeneity in trip choice over time. Unobserved factors affecting trips in period t are no doubt present in periods $t-1$ and $t+1$. If so, this feedback needs to be dealt with by purging the explanatory variables of any historical or future content (most notably, the lag of past trips to the site used as explanatory variables) of their endogeneity. The instrumental variables needed to make this possible have been elusive.

There have been a few efforts to build time interdependence into site choice models. As just noted, one way is to use a measure of past trips to a site as an explanatory variable in a TCM. For example, some have considered a simple dummy variable for whether or not a person has visited the site in the previous season or the current season as an explanatory variable in site utility in eqn [5]. A positive coefficient would imply 'habit formation' and a negative coefficient, 'variety seeking.' Other types of time-interdependent variables have or might include time since previous trip to a site, quality of experience on previous visits to a site, or known upcoming visitation plans. Research along these lines has been limited to a few exploratory studies and has largely ignored the issue of endogeneity; as such, it has not become a standard method in the literature. Most analyses continue to use data gathered without date-specific information and simply allow correlated errors over the season. Interestingly, seasonal demand models and the KT model implicitly estimate diminishing marginal utility of trips to individual sites *within a season* and, at least in this way, if it exists, implicitly capture the extent of habit-forming versus variety-seeking behavior in the sample.

A fully dynamic model where choices over the season are the result of solving a dynamic programming problem has been estimated, but given the computational difficulty, this has only been possible for a single-site discrete choice model ('go – don't go' each day of the season).

Finally, there have been efforts that combine contingent behavior data with trip choice data to infer intertemporal effects. For example, simply asking people what they might do if a site had been closed or a catch rate of fish at a site had been lower and then allowing for trip and site choice (substitution) in response to be in different time periods within or across seasons provides a data set with some time interdependence. These types of data have been exploited by allowing the response information to serve as alternative sites, but have fallen short because of not explicitly accounting for the dynamics inherent in the choice.

The evidence to date based on a small collection of studies and simple common sense is that accounting for intertemporal substitution and dynamics makes for better models of behavior and is likely to have a large impact on measures of welfare.

Choice Set Formation

When forming the relevant choice set for a multiple-site TCM study, the usual approach is to begin with the sites of policy significance and then expand it to include a reasonable set of substitute sites without reaching a number so large that estimation is infeasible. It is often driven by arbitrary political or geographic boundaries and, in some instances, leads to highly aggregated sites. In some cases, counties or even regions larger than counties can serve as individual sites. Most recent studies have used less aggregated sites, such as individual lakes, rivers, parks, or beaches. As a general rule, the more homogeneous the sites, the less the error faced in aggregation. In practice, it is best to err on the side of less aggregate sites.

The model being used can also be a factor in choice set size. Seasonal demand and KT models are nearly always estimated with fewer than ten or so sites, often with as few as three or four. It has, however, been shown that KT models can be estimated with a significantly larger number of sites. RUM-based site choice models are usually used when the number of relevant substitutes gets large. In some cases, this can be in the hundreds or even thousands. With certain restrictions applied to the model, it is possible to estimate a TCM using randomly drawn alternatives as a proxy for the full choice set, in effect, allowing for extremely large choice sets.

Most applications use choice sets defined by the analyst, but there is an ongoing debate about constructing and using choice sets defined by the respondents; that is, sites people are aware of and consider in making a choice. The difficulty with using choice sets formed by individuals is that the choice process that individuals use to form the narrow set of considered sites is part and parcel of the process of choosing the best site. Put differently, sites falling outside the set of considered sites are simply those with low utility and should be included in the analysis. If, on the other hand, individuals are unaware of sites, there is a reasonable argument for dropping them from the choice set in estimation.

Congestion

Congestion readily comes to mind when ones thinks of recreation sites. Given the growth in population and income and the decline in transit cost over time, recreation sites naturally see more use. In some cases, congestion can become a major policy issue. While the theory of incorporating congestion into TCMs is well understood, as is the idea of an efficient or optimal congestion at a site, it is difficult to incorporate its effect empirically, and there have been only a handful of studies that have attempted to do so. The difficulty is captured in a famous Yogi Berra quote about a favorite restaurant of his: "Nobody goes there anymore, it's too crowded." Observing many people at a site signals its desirability and hence high probability of visitation. At the same time, it may have gotten so popular that visitation is actually somewhat lower than it would have otherwise been but for congestion. How does one tease out the latter effect? An obvious start is to put some measure of congestion on the right-hand side of a seasonal demand or site choice model. But almost any measure one considers is correlated with excluded unobservables that influence individual demand – the same factors that effect individual demand (make a site desirable) also effect congestion. There is an inherent endogeneity problem, and sorting out the partial effect of congestion, without some cleaver instruments, is no easy task.

There are a few instances in the site choice random utility literature where instrumental variables have been successfully introduced to identify congestion effects. The instruments are still somewhat dubious (e.g., weather conditions, day of the week), but they have passed statistical tests. An alternative strategy, also appearing in the literature, is the use of some form of contingent behavior, combining stated and revealed preference data, where the stated component constitutes a response to some hypothetically introduced level of congestion. The effects of accounting for congestion in both the revealed preference/instrumental variable approach and the contingent behavior approach show that accounting for the effects of congestion are important to welfare analysis.

Conclusions

The TCM has been in use for over 50 years. It has grown in sophistication and use along with the growth in applied microeconometrics and applied welfare economics. It is the mainstay of nonmarket valuation for recreational uses of the environment.

This article has presented the three primary forms of the model in use today: seasonal demand, site choice, and KT. The seasonal demand models are the traditional forms and are best used in applications where there is a single or a few sites of interest, substitution outside this set is of limited relevance, and the focus is on access values. In some circumstances, cases can be made for using the model to value site attributes. However, the preferred model when the number of sites in question is large and/or quality changes at the sites are of interest is the site choice model using random utility theory. The site choice RUM model is the dominant model in the literature due to its flexibility and relative ease in application. The third model, the KT model, brings together the best of the seasonal and site choice models in a theoretically consistent way and may very well be the model of the future. However, because of its computational complexity, it has not been used widely.

The TCM certainly has its share of issues and complications. The major issues are essentially the same set confronting the model since its inception: measuring the value of time, dealing with multiple-purpose and overnight trips, accounting for intertemporal substitution, and forming the relevant choice set for estimation. Despite its flaws and blemishes, research and application of the TCM appear to be robust and poised for still more growth.

See also: **Allocation Tools:** Environmental Cost-Benefit Analysis; **Media:** Water Pollution from Industrial Sources; **Valuation Tools:** Benefit Transfer; Contingent Valuation Method; Hedonics.

Further Reading

Bockstael NE and Hanemann WM (1987) Time and the recreation demand model. *American Journal of Agricultural Economics* 69: 293–302.
Bockstael NE, Hanemann WM, and Strand IE (1986) Measuring the benefits of water quality improvements using recreation demand models. *Report to the U. S. Environmental Protection Agency*, *Cooperative Agreement CR-811043-01-0*. College Park, MD: University of Maryland.
Creel MD and Loomis JB (1990) Theoretical and empirical advantages of truncated count estimators for analysis of deer hunting in California. *American Journal of Agricultural Economics* 72(2): 434–441.
Habb TC and McConnell KE (2002) *Valuing Environmental and Natural Resources*. Cheltenham: Edward Elgar ch. 7 and 8.
Herriges JA and Kling CL (eds.) (1999) *Valuing recreation and the environment: Revealed preference methods in theory and practice*. Cheltenham: Edward Elgar.
Herriges JA and Kling CL (eds.) (2008) *Revealed Preference Approaches to Environmental Valuation*, vols. 1 and 2. Burlington, VT: Ashgate Publishing Company (Part IV).
Landry CE and Liu H (2009) A semi-parametric estimator for revealed and stated preference data – An application to recreational beach visitation. *Journal of Environmental Economics and Management* 57: 205–218.
McFadden D (2001) Economic choices. *American Economic Review* 91(3): 351–378.
Mendelsohn R, Hof J, Peterson G, and Johnson R (1992) Measuring recreation values with multiple destination trips. *American Journal of Agricultural Economics* 74: 926–933.
Morey E, Rowe RD, and Watson M (1993) A nested-logit model of Atlantic Salmon Fishing. *American Journal of Agricultural Economics* 75: 578–592.
Parsons GR (2003) The travel cost model. In: Champ PA, Boyle KJ, and Brown TC (eds.) *A Primer for Nonmarket Valuation*. Boston, MA: Kluwer Academic Publishers ch. 9.

Phaneuf DJ, Kling CL, and Herriges JA (2000) Estimation and welfare calculations in a generalized corner solution model with and application to recreation demand. *The Review of Economics and Statistics* 82: 83–92.

Phaneuf DJ and Smith VK (2005) Recreation demand models. In: Maler KG and Vincent JR (eds.) *Handbook of Environmental Economics*, vol. 2. North Holland: Elsevier.

Provencher B and Bishop RC (1997) An estimable dynamic model of recreation behavior with an application to great lakes fishing. *Journal of Environmental Economics and Management* 33: 107–127.

Shonkweiler JS and Shaw WD (1996) Hurdle count-data models in recreation demand analysis. *Journal of Agricultural and Resource Economics* 21(2): 210–219.