A Test of Choice Set Misspecification
for Discrete Models of Consumer Choice

J.R. DeShazo*
Department of Policy Studies, UCLA

Trudy Ann Cameron
Department of Economics, UCLA

Manrique Saenz
Central Bank of Costa Rica
November 5, 2001

ABSTRACT

We develop and evaluate a test of choice set misspecification for a multinomial logit choice model. This test determines whether the choice set designated by the researcher mistakenly assigns relevant substitutes to the numeraire good. We develop this test by generalizing the traditional McFadden-type conditional logit model to evaluate whether the traditional model is conditioned on an overly restrictive set of substitution possibilities. The test has a convenient feature: while it requires information on potentially relevant, but omitted, substitute goods, it does not require the researcher to observe consumers' choices among these omitted potential substitutes if they select the numeraire good (which contains these omitted substitutes). A comparison of the traditional multinomial logit choice model and our more general model suggests that choice set misspecification may produce biased parameters that distort welfare estimates to a consequential extent.

Corresponding author: JR DeShazo, deshazo@ucla.edu or UCLA School of Public Policy, 3250 Public Policy Building, Los Angeles, CA 90049

JEL Classification: C25, C42, C51, L15, Q26

1. INTRODUCTION

Researchers employ discrete choice models in fields as diverse as marketing, transportation, environmental economics, health economics and industrial organization. To implement these models empirically, the researcher must distinguish those goods that are explicitly subject to the discrete choice (i.e. the designated choice set) from the so-called "numeraire" good. In principle, the empirical model should conform to the maintained hypothesis that individual preferences over the designated set are separable from those over the items that make up the composite numeraire good. However, few tasks in empirical modeling require as much subjective judgment or involve as much uncertainty as specifying the designated choice set. A mis-specified choice set can produce flawed parameter estimates that misrepresent the degree of substitutability between goods, since it fails to acknowledge the the availability of substitutes in the choice set actually considered by the consumer ([Stopher, 1980 #272], [Williams, 1982 #271]). The absence of a formal statistical test to evaluate alternative choice set specifications has further exacerbated the difficulty of choice set specification. We develop one such formal test and examine its performance in an empirical example.

Researchers may employ our test to address several types of uncertainty that arise when specifying a designated choice set that captures the extent of the market (from the perspective of a given individual consumer). Consider a researcher who is specifying a designated set of *m* differentiated goods, W = $\{X_1,..X_i,..X_m\}$, where each good, X_i, is composed of a vector of **n** attributes $\{x_{i1},.,x_{ij},..x_{in}\}$. The researcher may be uncertain about set of attributes (x_{ij}) to associate with each good *as well as* the overall set of goods in the individual's choice set (X_i). The researcher may also be uncertain about the relevant range of attribute levels within an individual's choice set. For example, researchers have been unsure when to exclude goods because they are "too distant" or "too expensive" to have been deemed part of an individual's market. Our test may help to resolve the uncertainty associated with these specification choices.

To make the problem more concrete, let $W = X \cup Z$ be the true set of all relevant substitutes for the good or goods being studied. The goods in set X comprise the researcher's designated choice set while the goods in Z are in some ways similar to X and relevant to the consumer, but they are incorrectly aggregated by the researcher into the "No Purchase" (NP) option.[1] Let Y be the set of all genuinely separable goods that make up the intended numeraire good from the perspective of the researcher. The true utility function is assumed to be additively separable in $W= X \cup Z$ and Y, but not in X and Z.

Often, the researcher will assume that utility is separable in X and $Z \cup Y$. In order to test this separability assumption, we propose a more general random utility model. Once the attributes of the relevant Z-type alternatives have been identified, this model allows the researcher to model the numeraire good, Y, and the goods in Z as *different* alternatives. We can use this new specification to address the following questions:

1) Do individuals appear to interpret a numeraire good as including the possibility of consuming other goods, in Z, that are close substitutes for the goods in X?
2) Does the degree to which this happens differ across consumers?
3) How does allowing Z and Y to represent *different* alternatives affect estimated preferences over the attributes of goods in X? Are the differences between the traditional and the more general model likely to be economically significant?

In revealed preference (RP) settings, miss-specification occurs in at least two ways when the researcher fails to recognize the substitution possibilities for an omitted, but closely related, class of goods (Z). First, the researcher may fail to collect choice information for that set. Second, the researcher may fail to incorporate that set of substitutes into the W, thus incorrectly identifying the numeraire good as $Z \cup Y$.

Researchers in the RP literature have explored the importance of correctly specifying the set W from two angles. Some have sought to show how parameter estimates may be affected by the researcher's specification of the choice set (Stopher, 1980; Williams and Ortuzar,1982; Swait and Ben-Akiva,1987; Parsons and Hauber, 1998). Others have tried to identify W explicitly and allow it to vary systematically across individuals (Boxall, et al., 1996, Swait, 2001). The most sophisticated examples of this approach formally model the determination of the set W as an endogenous part of the choice process (Haab and Hicks (1997), Ben-Akiva and Boccara (1995) and Cascetta and Papola (2001)). The test that we present in this paper should be viewed as complementary to this strand of literature.

The endogeneity of choice sets may be an unavoidable feature of any revealed-preference choice context. In stated preference (SP) choice scenarios, however, it is sometimes assumed that the researcher has the ability to specify the relevant choice set exogenously. The present paper calls into question even this assumption. In an SP setting, researchers give individuals a specifically constructed set of alternatives W, and an NP option, which is assumed to represent the choice of Y.[2] Formally, the problem is that researchers interpret the individual's selection of

---

[1] We deem this problem to be more interesting than the case where the researcher specifies the choice set too broadly relative to the individual's true choice set.

[2] This "No Purchase" option may alternatively be called a "status quo" option, an "opt out" option, or a "none-of-the-above" option. Such an option is advocated on the rationale that forcing a selection from only the designated choice set is unrealistic for the consumer (see Ruby, et al. (1998)). Freeman (1991) demonstrates that a "status quo" alternative is necessary in stated preference (SP) experiments. Morey (2000) considers the conditions under which excluding the status quo or a none-of-the-above option from the choice set will and will not bias stated preference estimates. Boyle, et al. (2001) demonstrates empirically that the presence or absence of an opt-out alternative in stated preference surveys can make a difference to preference estimates.

the NP option as equivalent to choosing Y, while the individual interprets her choice of the NP option as a choice of either Z or Y.

We illustrate our test of choice set misspecification using SP data, but our approach can be easily applied to RP data. Methodologically, our innovation is equivalent to introducing a weight on Z, which allows us to estimate the degree to which individuals in a sample considered goods in Z to be viable substitutes for X. If the estimated weight is zero, then our model becomes the standard multinomial logit based on choice between W and Y, where W equals Z. If the weight equals one, then the individual behaves as if he had been explicitly given Z, and thus faced a choice between X, Z and the NP option.

Unfortunately, when the individual selects the NP option, the researcher typically does not have any information about exactly which Z or Y good was purchased (for RP data, or imagined to be most preferred, for SP data). Fortunately, to use our test, researchers do not need information on what individuals actually intended to choose when selecting the NP option. Economic intuition would suggest the total WTP estimate for a good is overestimated when the presence of close substitutes in set Z is overlooked. The demand for goods in X will be more elastic when the goods in Z are available than when they are not.

We can provide an imperfect assessment of this prediction that welfare estimates will differ with miss-specified substitution possibilities. This assessment is imperfect because we need to assess the effect on expected utility of eliminating a good completely from the choice set, rather than eliminating it only from set X, as would be the usual simulation with the traditional model. However, without information on individuals' actual choice of goods from Z, we may estimate the welfare changes that are conditioned on only changes in the availability of goods in X, rather than $X \cup Z$.

We estimate a traditional model and two new alternative models using an available data set concerning foreign tourists' demands for recreational trips and tour packages in Costa Rica. In preview, our findings reveal that individuals give measurable levels of attention to real substitutes from outside of the hypothetical stated-preference designated choice set. In addition, different population segments (socioeconomic groups) appear to assign significantly different weights to these outside options. Our available data set does not have sufficient information to permit rigorous welfare calculations for the entire population of potential international tourists who might visit Costa Rica. However, we use our estimated utility function parameters to calculate point estimates of what we will term "conditional" willingness to pay (WTP) for different types of recreational sites, given that the international traveler has already incurred the cost of getting to Costa Rica. For each, we test the null hypothesis (implied by the traditional specification) that the weight associated with the Z-type alternatives equals zero. As a result, the traditional model, which may be conditioned on a poorly specified choice set, may overestimated the value of some of the designated goods and the numeraire good.

The paper is structured as follows. Section 2 briefly reviews the traditional random utility choice model, and Section 3 presents our simple generalization. Section 4 describes our data and Section 5 explains our results. Section 6 concludes, with a discussion of possible directions for future research.

2. TRADITIONAL RANDOM UTILITY CHOICE MODEL

McFadden's conditional logit choice model will be familiar to most researchers who study choices among multiple alternatives. Each consumer has a true choice set consisting of a set of goods W plus a numeraire good, Y. The consumer is constrained to purchase either one good from the set W or none at all, spending the rest of her income (I) on the numeraire good.

The indirect utility for consumer i associated with selecting good j from set W and spending the rest of her income on the numeraire good is given by:

$$u_{ij}^{*} = z_{ij}\eta + \alpha(I_i - p_{ij}) + \varepsilon_{ij}^{*} \tag{1}$$

where

z_{ij} = set of attributes of good j for individual i

I_i = income of individual i

p_{ij} = price of good j for individual i

ε_{ij}^{*} = random error (extreme-value distributed)

The level of indirect utility associated with each alternative is typically normalized on the indirect utility from Y, the numeraire-only option:

$$u_{iY}^{*} = \alpha I_i + \varepsilon_{iY}^{*} \tag{2}$$

The normalization makes the *net* indirect utility associated with consuming the numeraire alternative, $u_{iY} = u_{iY}^{*} - u_{iY}^{*}$, equal to zero. The net indirect utility associated with the choice of good $j \in W$ becomes:

$$\begin{aligned} u_{ij} &= u_{ij}^{*} - u_{iY}^{*} \\ &= z_{ij}\eta - \alpha p_{ij} + (\varepsilon_{ij}^{*} - \varepsilon_{iY}^{*}) \\ &= x_{ij}\beta + \varepsilon_{ij} \\ &= v_{ij} + \varepsilon_{ij} \end{aligned} \tag{3}$$

where

$v_{ij} = x_{ij}\beta$

$x_{ij} = \begin{bmatrix} z_{ij} & -p_{ij} \end{bmatrix}$

$\beta = \begin{bmatrix} \eta & \alpha \end{bmatrix}$

In the traditional multinomial logit choice model, the probability that consumer i chooses good $j \in W$ and spends the rest of her income in the numeraire is:

$$p_{ij} = \frac{e^{x_{ij}\beta}}{1 + \sum_{k \in W} e^{x_{ik}\beta}} \qquad (4)$$

and the probability of consuming the numeraire only is:

$$p_{iY} = \frac{1}{1 + \sum_{k \in W} e^{x_{ik}\beta}} \qquad (5)$$

The traditional model is estimated by collecting data on all of the x_{ij} variables and on the consumer's choices. Let $c_{ij} = 1$ if individual i chooses alternative j, and let $c_{ij} = 0$ otherwise. The log-likelihood function to be maximized by an appropriate choice of the unknown utility parameters, β, is:

$$Log\ L = \sum_{i=1}^{n} \sum_{j \in W, NP} c_{ij} \log\left(P_{ij}\right) \qquad (6)$$

However, specifying all the relevant elements of W for each consumer might be extremely difficult due to a lack of information on the individual's true choice set. In the case of SP methods, the NP option may include not only the numeraire good, Y, but also some number of goods of type Z. We continue to normalize the utility differences on the level of indirect utility derived from the numeraire good, but now the relevant choice probabilities for goods of type X in the designated set must be modified. Specifically, there is a separate term for set Z in the denominator of the probability formulas:

$$p_{ij} = \frac{e^{x_{ij}\beta}}{1 + \sum_{k \in X} e^{x_{ik}\beta} + \sum_{k \in Z} e^{x_{ik}\beta}} \qquad (7)$$

where $j \in X$. Likewise, the true probability associated with the NP option is now the sum of the probability of consuming the numeraire good only, plus the probabilities associated with each of the goods in Z:

$$p_{i,NP} = \frac{1}{1+\sum_{k\in X} e^{x_{ik}\beta} + \sum_{k\in Z} e^{x_{ik}\beta}} + \frac{\sum_{k\in Z} e^{x_{ik}\beta}}{1+\sum_{k\in X} e^{x_{ik}\beta} + \sum_{k\in Z} e^{x_{ik}\beta}}$$

$$= \frac{1+\sum_{k\in Z} e^{x_{ik}\beta}}{1+\sum_{k\in X} e^{x_{ik}\beta} + \sum_{k\in Z} e^{x_{ik}\beta}} \tag{8}$$

3. A GENERALIZATION

We now describe a generalized specification for the Random Utility Model that can be illustrated with our available sample of stated preference (SP) survey data concerning recreational trips in Costa Rica by international tourists. In doing so, we argue that the researcher should allow for the possibility that the respondent chooses "None of the Above" if she prefers some trip that she knows or suspects is available in the real world over the short list of hypothetical alternatives being presented to her in the survey.

Instead of $u_{ij} = x_{ij}\beta + \varepsilon_{ij}$, the formula for the indirect utility function defined in equation (3), we assume a more-general function involving a strictly positive parameter γ:

$$u_{ij} = \ln(\gamma) D_{ij} + x_{ij}\beta + \varepsilon_{ij} \tag{9}$$

where

D_{ij} = a dummy variable, activated when $j \in Z$

x_{ij} = vector of attributes of site $j \in X$ or $j \in Z$

ε_{ij} = difference of extreme value error terms

The estimated value of the unknown parameter γ will vary with the degree to which individuals perceive goods in Z to be substitutes for good in X. Consider a good $j \in Z$ (such that $D_{ij} = 1$). An estimated value of $\gamma = 1$ implies that the net utility derived from this implicit alternative in Z is the *same* as that from an otherwise similar alternative in the designated set X. If the same level of utility is derived from the vector of attributes x_{ij} whether $j \in Z$ or $j \in X$, we would expect to have $\ln(\gamma) D_{ij} = 0$, implying that $\gamma = 1$.

Similarly, if the estimated value of γ will be less than one if the net utility derived from an implicit alternative $j \in Z$ (so that $D_{ij} = 1$) tends to be systematically *smaller* than the utility derived from an otherwise similar explicit alternative $j \in X$. In this case, we would expect $\ln(\gamma) D_{ij}$ to be negative, implying a negative estimate for $\ln(\gamma)$, since $D_{ij} = 1$, which further implies $\gamma < 1$.

When the goods in set Z are viewed by the consumer as just another variant of the numeraire good, Y, then the estimated value of γ should be zero. Why? Because the "net inclusive indirect utility" derived from all the goods in set Z *plus* the numeraire good Y should be zero (which is the same as the net inclusive indirect utility from the numeraire good in the traditional model). The inclusive indirect utility for a set of alternatives is generally defined as the exponential average of the systematic portions of utility derived from each of the constituent alternatives in that set. For the combined set of alternatives $Z \cup Y$, if it is true that $\ln\left[\left(\sum_{j \in Z} \exp\left(\ln \gamma + x_{ij}\beta\right)\right) + \exp(0)\right] = 0$, then $\gamma \sum_{j \in Z} e^{x_{ij}\beta} = 0$, which can be true in general only if $\gamma = 0$. Note that this circumstance implies a discontinuity at $\gamma = 0$, because $\ln(0)$ is then undefined in equation (9).

Using the generalized indirect normalized utility function in equation (9), the probability associated with a good in set X is given by:

$$\mathrm{P}_{ij} = \frac{e^{\ln(\gamma)D_{ij} + x_{ij}\beta}}{1 + \left(\sum_{k \in X} e^{\ln(\gamma)D_{ij} + x_{ik}\beta}\right) + \left(\sum_{k \in Z} e^{\ln(\gamma)D_{ij} + x_{ik}\beta}\right)} \tag{10}$$

and the probability associated with the NP option is given by:

$$\mathrm{P}_{iNP} = \frac{1 + \left(\sum_{k \in Z} e^{\ln(\gamma)D_{ij} + x_{ik}\beta}\right)}{1 + \left(\sum_{k \in X} e^{\ln(\gamma)D_{ij} + x_{ik}\beta}\right) + \left(\sum_{k \in Z} e^{\ln(\gamma)D_{ij} + x_{ik}\beta}\right)} \tag{11}$$

Since $D_{ij} = 0$ for $\mathrm{j} \in \mathrm{X}$ and $\mathrm{D}_{ij} = 1$ for $\mathrm{j} \in \mathrm{Z}$, and since $e^{\ln(\gamma)} = \gamma$, these formulas can be simplified. The probability that individual i chooses trip j as it is explicitly offered in the survey is given, for all $j \in X$ by:

$$\mathrm{P}_{ij} = \frac{e^{x_{ij}\beta}}{1 + \left(\sum_{k \in X} e^{x_{ik}\beta}\right) + \gamma\left(\sum_{k \in Z} e^{x_{ik}\beta}\right)} \tag{12}$$

The probability that individual i chooses the NP option is given by:

$$\mathrm{P}_{iNP} = \frac{1 + \gamma\left(\sum_{k \in Z} e^{x_{ik}\beta}\right)}{1 + \left(\sum_{k \in X} e^{x_{ik}\beta}\right) + \gamma\left(\sum_{k \in Z} e^{x_{ik}\beta}\right)} \tag{13}$$

where:

x_{ik} = row vector of attributes for the k$^{\mathrm{th}}$ trip

$k \in X$ = a trip option explicitly presented in the survey

$k \in Z$ = a trip option assumed to be contained in the NP option

β , γ = unknown parameters to be estimated simultaneously by maximum likelihood.

Our new specification, therefore, subsumes the model with choice probabilities as in (7) and (8) as well as the traditional model with choice probabilities as in (4) and (5). We allow the terms for the relevant-but-implicit Z alternatives within the NP option to have a different weight in the probability formulas (the new parameter γ, instead of just unit weight or zero weight). If $\gamma = 1$, the trips contained in set Z are apparently considered by the consumer to be as feasible as any of the trips in X, and the consumer has similar knowledge about both groups of options. If $\gamma = 0$, then trips that are not offered explicitly in the survey are apparently not considered at all by the consumer (i.e. they are indistinguishable from the numeraire good). An estimated value of γ such that $0 < \gamma < 1$ implies that the consumer recognizes the goods in Z as substitutes for X, but on a less-than-equal footing.

Is γ a simple constant? Or does it vary systematically with individual characteristics? The effect of existing real recreation opportunities available in the market on the individual's responses to the artificial survey choice scenarios may depend on the amount of information that she has about such opportunities. If tourists have very limited information about other sites actually available, or if they do not interpret the NP option as containing such choices, the existence of other real options ought not to have a strong effect on the individual's choices.

In order to identify population heterogeneity in access to (or utilization of) information on relevant-but-implicit alternatives, we explore the possibility that γ may be a function of socio-demographic characteristics. Variables such as age, education, and income might influence the level of information that the individual possesses at the time of the survey. Hence, we generalize the probabilities described by (12) and (13) by replacing the scalar-valued γ with a systematic varying parameter $\gamma_i = \gamma_0 + \sum_m \gamma_m s_{im}$, where s_{im} is a vector of socio-demographic variables.[3]

4. DATA

Our data are drawn from a survey carried out in 1998, involving a sample of 1,033 international tourists in Costa Rica. Each respondent is asked to choose one alternative from a designated set of thirteen tour packages and one-day trips plus a "None of the Above" option.[4] Each tour package has a brief description of the sites to be visited, the length of stay at each, and the price of the trip or tour package.[5] In each stated-preference exercise, the set of trip packages (goods) presented to each individual was randomly drawn from a universal experimental choice set. The designated choice sets presented to respondents each contained thirteen tour packages, randomly drawn (in a stratified manner) from the universal opportunity set.[6] Our SP analysis evaluates the demand for the 27 underlying specific sites that visitors actually consumed for between one and three days, alone or in combination with other sites, over a total trip length that ranges from four to twelve days.

---

[3] The preference parameters themselves could also be heterogeneous across different socioeconomic groups. We do not pursue such a model here, except to note that suppressed preference heterogeneity could manifest itself as heterogeneity elsewhere in the model, as is sometimes the case with heteroscedasticity in the errors.

[4] The fourteen packages were actually presented in groups: first seven short trips, then six long trips, and then the respondent was asked to choose among the most-preferred short and most-preferred long trips. In this paper, we collapse the explicit two-stage choice process into an implied one-stage choice among fourteen total alternatives. Detailed hierarchical choice analysis is reserved for a separate paper.

[5] An appendix, available from the authors, presents a sample choice set including these short and long trips.

[6] We constructed this universal set of trips by collecting 239 actual tour package itineraries on the Costa Rican market in 1998. From these tour packages we identified six general types (or themes) of multiple day trips and four general types of one-day trips. Based on actual market data we then set bounds on (1) number of sites in the package, (2) length of stay per site, and (3) price per package for each type of tour or one-day (one-site) trip.

As a practical matter, what real-world trip alternatives should we include in the NP set? We will let this set contain 9460 trips or tours, actually available, which vary in their complement of sites visited and the length of stay at each. Approach allows for the case in which an individual chooses the NP option because she prefers a trip different (perhaps in price but maybe also in attributes) from any of the representative trips offered in the survey. A vast number of possible tourist trips in Costa Rica could be constructed, so we render the selection finite by employing some basic restrictions on the set: We obtain all the possible site and length-of-stay combinations given the 27 most commonly visited sites and the constraints just described.[7] Fortunately, since all foreign visitors are construed to be starting from the same "origin" (namely, the international airport in the capitol city of San Jose), the transportation and lodging costs associated with each of the 9460 possible packages are not individual-specific, but the same for all visitors. For the 9460 relevant-but-implicit alternatives that make up Z, we use the real levels of these same attributes accompanied by the real costs of these trips.[8] The price variable merits particular attention if any type of welfare estimation is to be pursued.

We classify the 27 actual sites used in the choice scenarios into seven "site types" according to their main feature: volcano, beach, forest, river, golf, fishing trip, and island. The basic specification that we use mimics the one used in Saenz (2000). In terms of the groups of variables used in this illustration, the utility derived by an individual from alternative j (namely, equation (1)) will be given by:

---

[7] Each trip can include at most four types of sites, and at most three sites within each type. Hence the maximum number of sites visited on a trip is twelve. The maximum length of stay at each site type is three days, and the maximum length for a complete trip is sixteen days.

[8] The price that was stated on the survey for each hypothetical trip or tour package was randomly generated. The lower and upper bounds of its distribution are based on market data for similar types of trips. The price associated with the real trips included in the NP option is the real out-of-pocket cost of the trip within Costa Rica. This variable is calculated as the sum of the transportation and lodging costs. The transportation cost is obtained assuming that the individual rents a car and takes the fastest route available to visit all the sites included in the trip.

$$
\begin{aligned}
u_{ij}^{*} &= \eta_1 TYPE_{ij} + \eta_2 TYPE\_INTERACT_{ij} + \eta_3 DAYS\_AT\_TYPE_{ij} \\
&+ \eta_4 DAYS\_AT\_TYPE^2 + \eta_5 DAYS\_AT\_TYPE\_INTERACT_{ij} \\
&+ \alpha(I_i - INTL\_TRIP\_COST_i - P_j) + \varepsilon_{ij}^{*}
\end{aligned} \tag{14}
$$

where $TYPE_{ij}$ is a vector of dummy variables capturing the seven possible main features of any site visited on trips or tour packages j, and $TYPE\_INTERACT_{ij}$ is a subset of all possible interactions among site type dummies, which will be nonzero only when tour j involves visits to more than one site type. $DAYS\_AT\_TYPE_{ij}$ is a vector of count variables, each of which captures the total number of days at sites with each of the seven main features. $DAYS\_AT\_TYPE\_INTERACT_{ij}$ is a subset of all possible interactions among total days at site types with each main feature. Again, it is nonzero only for tours involving more than one site type. The cost of getting to Costa Rica, $INTL\_TRIP\_COST_i$ will differ across travelers, but will be identical across alternative trips within Costa Rica for each traveler, since they all begin their domestic Costa Rican trip segments at the international airport in San Jose. In calculating net utilities normalized on the NP option, income, I_i, and $INTL\_TRIP\_COST_i$ will drop out, given this linear specification. Differences relative to the NP option in the attributed levels will equal their absolute levels under the assumption that the NP option has no trips and therefore has zero values for trip types and days. Our empirical version of equation (3) is thus:

$$
\begin{aligned}
u_{ij} &= \beta_1 TYPE_{ij} + \beta_2 TYPE\_INTERACT_{ij} + \beta_3 DAYS\_AT\_TYPE_{ij} \\
&+ \beta_4 DAYS\_AT\_TYPE^2 + \beta_5 DAYS\_AT\_TYPE\_INTERACT_{ij} \\
&+ \beta_6(-P_j) + \varepsilon_{ij}
\end{aligned} \tag{15}
$$

The variables used in the econometric estimation are described in Table 1.The bottom panel of Table 1 presents the number of sampled tourists that chose each type of package or trip

---

The lodging cost is calculated on the basis of the average price of a luxury hotel at the site visited. This was the type in the SP exercise. As outlined above, the choice set presented to each individual includes a set of multiple-day tour packages and a set of one-day tour packages, and each tour package is characterized by the type(s) of sites that it includes. The actual physical sites presented to the individual are randomly drawn from a pool of real sites of the appropriate type. For example, the themes "One-day trip #1 to volcano" and "One-day trip #2 to volcano" present different draws of a site from a set of volcanoes. The NP option is chosen by about 12% of the sample, and this is the third-highest choice frequency in the table. (The only packages chosen more frequently than the NP option are long trips involving several types of sites.) [9]

5. RESULTS AND IMPLICATIONS

We present a complete set of results for each of three different exploratory models. We first estimate the traditional specification, which restricts to zero the coefficient γ on the D_{ij} dummy variable. Then we estimate Model 1, which frees up γ and allows the Z-type trips to be implicit in the NP option (implying the possibility of the consumer considering a much wider variety of relevant alternatives than just the designated set presented to her in the survey). Model 2 is more general still, in that it allows γ to vary systematically with a number of respondent characteristics, allowing different types of consumers to devote different amounts of attention to these Z-type trips in making their choices.

Our primary goal in this paper is to compare the indirect utility parameter estimates that result from our two new alternative specifications to the traditional multinomial logit random

---

of lodging specified in the packages presented to the respondents.

[9] Given that our sample consists of people who are already consuming a trip or tour package in Costa Rica, this high percentage choosing the NP option suggests that individuals may be interpreting this option to include the possibility of choosing other trips available in reality.

utility specification. Our ability to consider welfare estimates is more limited. The cost of international travel to reach Costa Rica is not considered in our models (and it drops out of the specification because of the linear-in-net-income functional form). Thus, we are not in a position to treat our sample of international travelers arriving in Costa Rica as representative of the entire population of potential foreign tourists who might visit Costa Rica. Instead, our sample can at best be argued to be representative of foreign tourists who have already selected Costa Rica as their destination, and we will limit our welfare calculations to estimates of their *conditional* WTP for sites, given that they have already reached Costa Rica. These estimates might be interpreted as incremental WTP for excursions within Costa Rica, beyond any WTP to reach Costa Rica itself.

In discussing our results, we focus first on the key new parameter in our models, the weight, γ, attached to the relevant-but-implicit trip alternatives that are lumped into the NP option. We view the parameter γ (or its systematically varying analog) as a simple "nuisance" parameter, whose presence is necessary in our models so that the desired indirect utility-difference parameters, β, may be estimated with less bias. We therefore compare the indirect utility parameter estimates in the standard and alternative specifications. These capture the marginal indirect utilities associated with each of the trip attributes. Finally, to illustrate the possible practical consequences of failing to accommodate $\gamma \neq 0$, we will explore the different implications of the traditional model and our generalized models for point estimates of the incremental willingness to pay (WTP) for the set of *real* trip and tour options in Costa Rica, given that travelers have already selected Costa Rica as their vacation destination. We will also examine the implied WTP for access to sites of different types (volcanoes, beaches, forests, etc.).

5.1 Indirect Utility Parameter Estimates

Table 2 presents the coefficient estimates and asymptotic t-test statistics for the traditional model and two alternative specifications. In alternative Model 1, the single γ coefficient is statistically significant and positive. This implies that, on average, a positive weight is placed on the relevant-but-implicit trip alternatives in set Z. The traditional model is nested within this more general specification and corresponds to the case where γ equals zero. Therefore, in terms of how well the observed choices can be explained, the traditional specification is statistically rejected in favor of our new alternative.

We conjecture that respondents' level of information about alternative trip combinations and prices (in Z) will vary across individuals. In alternative Model 2, generalizing γ to be a systematically-varying function of socio-demographics allows the weight on the Z-type terms in the choice probabilities to differ across population segments. It appears that the weight is greater for respondents who are single or divorced, and smaller for educated and high-income individuals, although collinearity between these variables may make it difficult to discriminate between education and income effects.[10,11]

The coefficients for trip attributes obtained under our two alternative specifications exhibit important differences relative to those obtained under the traditional one. In the

---

[10] A possible explanation for the effect that marital status and income have on this weight is the likely difference in costs associated with taking the offered trips on your own, versus taking them as offered in the survey. Individuals who are single or divorced may be more flexible to changes in plans halfway through a trip than married tourists who are more likely to travel with their spouse or family. Therefore, single tourists might be attracted to taking the trip on their own (at its real cost), even if this implies more uncertainty about how things might turn out in their choices of transportation, lodging, etc. High income tourists might be less informed than others because their opportunity cost of time is higher and they might have invested less time in learning about Costa Rica before leaving their home country. The cost of getting additional information to take a trip on their own (or look for a tour operator who will tailor a tour package to their preferences) might make the trips implicit in NP less attractive

[11] One could also argue that high income tourists have lower marginal utility of income and therefore are willing to pay more for a planned trip being offered to them.

traditional model, the linear terms involving dummies for each individual site type (Volcano, Beach, Forest, River, Golf, Angling, and Island) are individually significantly negative, but they become insignificant under our alternative specifications. In the traditional model, the relative popularity of the NP option in relation to one-day trips is expressed via low estimated marginal indirect utility values (in particular, by significant negative coefficients on the *TYPE* dummies). In our alternative specification, the relative popularity of the NP option might be explained by the utility value from the relevant-but-implicit trip options contained in NP. Negative coefficients on the site-type dummies are no longer needed to explain the observed choices.

In moving from the traditional specification to our more-general models, the coefficients on all of the $TYPE\_INTERACT$ variables involving the site-type dummies decrease substantially (i.e. the coefficients on (Beach)*(Forest), (Beach)*(Volcano), (Forest)*(Volcano), and (Volcano)*(River)). These coefficients partly capture substitutability. These differences can have important implications concerning the degree to which the WTP for one type of site will depend on the presence of an opportunity to visit other types of sites during the same trip.[12]

5.2 Purging Welfare Estimates of the Influence of Relevant-but-Implicit Alternatives

This section considers point estimates of the conditional WTP for access to each of the main site types in the real Costa Rican choice set of 9460 tour packages or trips, given that a traveler has reached Costa Rica. These estimates are compared for the traditional and alternative

[12] In a fully quadratic specification of a direct utility function, the coefficients on interaction terms characterize the shapes of the level curves of the function. These level curves will be ellipses if all of the second derivatives of the function have a common sign. A positive coefficient on an interaction term signifies a positive axis to the ellipse and a negative coefficient signals a negatively sloped axis. When viewed from the origin, as for a set of indifference curves, the level curves thus become increasingly sharply curved as the coefficient on the interaction term increases from negative to positive. In the indirect utility function in this paper, however, we have interactions both in the presence of a site type and in the number of days at each site type, so the intuition is not as clear.

specifications. For the first line in the body of Table 3, labeled "All Trips," we calculate a point estimate of the overall WTP associated with the availability of the whole set of 9460 trips that are actually obtainable in Costa Rica. It is calculated as the normalized inclusive utility value from all available real choice options, divided by the negative of the marginal utility of income, α, in order to monetize this utility difference:

$$(-1/\alpha)\ln\left(1+\sum_{j=1}^{9460}\exp(x_{ij}\beta)\right) \tag{16}$$

This is the estimated maximum WTP to avoid the utility loss that would accompany the simultaneous elimination of all real trips, given that the international traveler is already in Costa Rica. We conduct these simulations based on choice set of $Z \cup X$ rather than simply X by letting parameters of the preference function, β (which includes α) and γ allowed to take on whatever values the choice data dictate.[13]

These WTP estimates, of course, assume that the estimated indirect utility parameters are the same as their true-but-unknown parameter values. At this juncture, many researchers use simulation methods to generate sampling distributions for the point estimates of WTP, calculated from this nonlinear function of data and the maximum likelihood estimates of the parameters α and β, which can be assumed to be jointly distributed asymptotically normal random variables. The procedure involves repeated sampling from the presumed joint normal distribution of these maximum likelihood parameter estimates. (See Krinsky and Robb (1986) and Krinsky and Robb (1990).) However, our main point of concern in this paper concerns the

[13] The formulas for calculating welfare effects in discrete choice models are derived in Small and Rosen (1981), with the particular formula for logit-based models appearing on page 127 of that paper.

effect of our proposed generalizations upon the utility parameters. Thus, these WTP point estimates are offered only to illustrate the possible magnitudes of the welfare implications.[14]

What about WTP for access to each site *type* among the 9460 actual trips or tours in the real market? To simulate the absence of particular site types, the dummy variable for that site type in the $TYPE_{ij}$ vector is counterfactually set equal to zero, as is the corresponding $DAYS\_AT\_TYPE_{ij}$ variable. These variables are also set equal to zero where they enter into interaction or squared terms. Let v_{ij} denote the systematic portion of the real utility difference associated by individual i with trip or tour j, where j=1,…,9460 real trips and tour packages, and let v_{ij}^{S} signify the counterfactually simulated version of this expression in the absence of a particular site type, S. Then the WTP formula is given by:

$$WTP(S) = \frac{1}{\alpha}\left[\ln\left(1+\sum_{j=1}^{9460}\exp(v_{ij})\right)-\ln\left(1+\sum_{j=1}^{9460}\exp(v_{ij}^{S}\right)\right] \qquad (17)$$

The effect of an incorrect characterization of the NP option is greatly evident in the coefficients on the trip attributes displayed in Table 2. The slope coefficient estimates for specific trip attributes display potentially important differences when our new specifications are used.

Table 3 also shows our estimates of WTP to retain all trips that contain a certain site *type* using the coefficients from the traditional and alternative specifications. Observe that the WTP

---

[14] We have explored simulated confidence intervals for our WTP estimates, but they are rather wide. The results are discussed in detail in an Appendix available from the authors. This appears to be due to collinearity among a number of the attributes, and imprecision in our estimate of the coefficient on the price variable. The parameter variance-covariance matrix upon which the WTP simulations are based exhibits negative correlation between the price coefficient and a small number of the forest-related x_{ij} variables. As a result, the variability in the ratios that constitute the WTP formula is large. In spite of this undesirable property of our particular data set, we have demonstrated the potential importance of our generalized model. Future researchers, with access to better samples of choice data, should certainly pursue confidence intervals for fitted WTP in any thorough analysis where the welfare implications are crucial.

estimates are qualitatively smaller for each site type under our alternative specifications but especially for forests and volcanoes (44% and 25% lower, respectively). This is consistent with the fact that the coefficients on the site-type dummy variable interaction terms are all lower in the alternative models. Site types are closer substitutes for each other according to our alternative specification estimates, and, therefore, the consumer surplus from each site type individually is now lower, even though the WTP for the overall set of 9460 real tour packages or trips is not much changed.

Comparing Models 1 and 2, the fitted WTP estimates differ very little. This reflects the insensitivity of the basic utility parameter estimates to the choice of a constant or systematically varying γ parameter. While there is statistically significant heterogeneity in the size of the γ parameter, this heterogeneity appears to be unimportant to sample average fitted WTP estimates.

6. CONCLUSIONS AND DIRECTIONS FOR FUTURE RESEARCH

The test of separability proposed in this paper will help researchers judge whether they are using the information from these surveys in an appropriate manner. If $\gamma = 0$ cannot be rejected, the usual assumption that there are no relevant-but-implicit alternatives may be justified. If $\gamma = 0$ is rejected, a maintained hypothesis of the traditional Random Utility choice model is violated and the welfare implications of utility parameters estimated using a traditional specification might be distorted. In particular, failure to acknowledge the role of relevant-but-implicit alternatives can lead to incorrect inferences about the degree of substitutability between goods.

We have illustrated our new test in a context where, contrary to the researcher's intention, individuals may take into account real world substitutes when choosing their preferred alternative in a SP survey. Such a possibility is typically ignored in the literature on demand estimation based on SP data. We show that ignoring these real world alternatives has the potential to produce important biases in the indirect utility function parameter estimates derived from SP data, and thereby to distort willingness to pay estimates for site attributes. We observe apparently non-trivial differences in point estimates of the marginal willingness to pay estimates for most site attributes between the traditional model and our proposed alternative specifications.

Our evidence, furthermore, suggests several issues that should be addressed in future research. First, our new model is a simple generalization of the conventional random utility model. It involves a new parameter (or set of parameters) for which the zero hypothesis can readily be rejected. But the nature of this generalization is relatively arbitrary. It would be appealing, for example, if one could show that it was equivalent to a nested-logit model with censoring of the specific choice at the most disaggregate level on the NP node.

Second, we have observed that the larger the number of alternatives in the relevant-but-implicit category, the greater will be the number of terms in $\sum_{j \in Z} \exp\left(x_{ij}\beta\right)$ and the smaller will be the estimated value of the parameter γ. It would be helpful to formalize our expectations for the size of γ as a function of the number of alternatives entertained for the set Z.

Third, the relevance of real world alternatives is likely to depend on the information that is possessed about them by each respondent. We use only a set of sociodemographic attributes to proxy for heterogeneity across individuals in such information. Direct questions about familiarity with the relevant real market may be feasible in some other applications. Finding variables to measure this level of information, and making the weight given to implicit real world alternatives a function of these variables, could help quantify this dependence.

Fourth, constructing the set of relevant real world alternatives considered by an individual is a difficult task. Our evidence suggests that, in this case, the results are not very sensitive to the actual range of relevant-but-implicit alternatives to be included.[15] However, more research on this subject is necessary. Exogenous information on the characteristics that are subjectively preferred by the individual might also help the researcher define the set of real alternatives that should be included in the model.

Fifth, our empirical specifications have maintained an assumption of homogeneous preferences over site types and durations-at-sites as attributes of trip packages. Heterogeneity has only been explored as it influences the size of the weight associated with the exponentiated inclusive utility attached by respondents to the set of relevant-but-implicit alternatives. Richer models with heterogeneous preferences should certainly be entertained with larger and more detailed data sets.

On a different front in the SP literature, one that is related to substitutability issues in multiple discrete choice models, researchers have begun to employ random parameters logit models in modeling SP choices (see for example Train (1998), Revelt and Train (1998), Layton and Brown (2000), McFadden and Train (2000), Brownstone, et al. (2000), Hensher (2001), and Hanley, et al. (2001)). These models allow for quite general forms of heterogeneity across the sample in the degree of substitutability between attributes. Our sixth point is that respondents whose preferences exhibit different apparent degrees of substitutability among the attributes represented in the alternatives in the designated choice set could be respondents who are paying more or less attention to the array of real alternatives that the researcher has inadvertently lumped into the NP option. There may be some observational equivalence between these two

[15] A smaller set of implicit-but-excluded alternatives, consisting of the same 13 trips offered to the individual, but at their market prices, yields qualitatively similar findings.

models. However, if differing apparent degrees of substitutability are just an artifact of consumers' differing degrees of recognition of relevant-but-implicit alternatives outside the explicit SP choice set, this could be an important behavioral explanation for the phenomenon of apparently heterogeneous preferences among the alternatives in the designated set.

Finally, our approach may have applications to assessing the "extent of the market" for specific commodities, for specific types of consumers. Subsets of real alternatives could be assigned different γ parameters as weights, and each of these γ_m parameters could vary systematically with individual characteristics. This would allow assessment of which subsets of relevant-but-implicit alternatives are attended-to most closely by different types of people.

REFERENCES

Ben-Akiva, M., and B. Boccara. "Discrete choice models with latent choice sets." *International Journal of Research in Marketing* 12, (1995): 9-24.

Boxall, P. C., et al. "A comparison of stated preference methods for environmental valuation." *Ecological Economics* 18, (1996): 243-253.

Boyle, K. J., et al. "A comparison of conjoint analysis response formats." *American Journal of Agricultural Economics* 83, (2001): 441-454.

Brownstone, D., D. S. Bunch, and K. Train. "Joint mixed logit models of stated and revealed preferences for alternative-fuel vehicles." *Transportation Research Part B-Methodological* 34, (2000): 315-338.

Cascetta, E., and A. Papola. "Random utility models with implicit availability/perception of choice alternatives for the simulation of travel demand." *Transportation Research Part C-Emerging Technologies* 9, (2001): 249-263.

Freeman, A. M. "Comment: Factorial Survey Methods and Willingness to Pay for Housing Characteristics." *Journal of Environmental Economics and Management* 20, (1991): 92-96.

Haab, T. C., and R. L. Hicks. "Accounting for choice set endogeneity in random utility models of recreation demand." *Journal of Environmental Economics and Management* 34, (1997): 127-147.

Hanley, N., et al. "Go climb a mountain: an application of recreation demand modeling to rock climbing in Scotland." *Journal of Agricultural Economics* 52, (2001): 36-52.

Hensher, D. A. "The valuation of commuter travel time savings for car drivers: evaluating alternative model specifications." *Transportation* 28, (2001): 101-118.

Horowitz, J., and J. Louviere. “What is the Role of Consideration Sets in Choice Modeling?” *International Journal of Research in Marketing* 12, (1995): 39-54.

Horowitz, J. L. “Modeling the Choice of Choice Set in Discrete-Choice Random- Utility Models.” *Environment and Planning a* 23, (1991): 1237-1246.

Krinsky, I., and A. L. Robb. “On Approximating the Statistical Properties of Elasticities.” *Review of Economics and Statistics* 68, (1986): 715-719.

Krinsky, I., and A. L. Robb. “On Approximating the Statistical Properties of Elasticities.” *Review of Economics and Statistics* 72, (1990): 189-190.

Layton, D. F., and G. Brown. “Heterogeneous preferences regarding global climate change.” *Review of Economics and Statistics* 82, (2000): 616-624.

McFadden, D. "Modeling the choice of residential location", in The economics of housing ed. D. McFadden, and J. Quigley, vol. 1. Cheltenham, UK, Edward Elgar, pp. 531-52.

McFadden, D., and K. Train. “Mixed MNL models for discrete response.” *Journal of Applied Econometrics* 15, (2000): 447-470.

Morey, E. “Forced-Choice and the Status Quo.” Manuscript. University of Colorado.

O'Neill, K., and D. Nelson (1981) An Assessment of Empirical Choice Set Formation, vol. XXII, pp. 72-79.

Parsons, G. R., and A. B. Hauber. “Spatial boundaries and choice set definition in a random utility model of recreation demand.” *Land Economics* 74, (1998): 32-48.

Parsons, G. R., and M. J. Kealy. “Randomly Drawn Opportunity Sets in a Random Utility Model of Lake Recreation.” *Land Economics* 68, (1992): 93-106.

Revelt, D., and K. Train. “Mixed logit with repeated choices: Households' choices of appliance efficiency level.” *Review of Economics and Statistics* 80, (1998): 647-657.

Roberts, J., and J. Lattin. “Development and Testing of a Model of Consideration Set Composition.” *Journal of Marketing Research* 28, (1991): 429-440.

Ruby, M. C., F. R. Johnson, and K. E. Mathews. “Just Say No: Opt-Out Alternatives and Anglers' Stated Preferences.” TER General Working Paper. Triangle Economic Research.

Saenz, M. , UCLA, 2000.

Small, K. A., and H. S. Rosen. “Applied Welfare Economics With Discrete Choice Models.” *Econometrica* 49, (1981): 105-130.

Stopher, P. “Captivity and choice in travel behavior models.” *Transportation Journal of ASCE* 106, (1980): 427-435.

Swait, J. “Choice set generation within the generalized extreme value family of discrete choice models.” *Transportation Research Part B* 35, (2001): 643-666.

Swait, J., and M. Ben-Akiva. “Incorporating Random Constraints in Discrete Models of Choice Set Generation.” *Transportation Research Part B-Methodological* 21, (1987): 91-102.

Train, K. E. “Recreation demand models with taste differences over people.” *Land Economics* 74, (1998): 230-239.

Williams, H., and J. Ortuzar. “Behavioral theories of dispersion and the mis-specification of travel demand models.” *Transportation Research B* 16, (1982): 167-219.

Table 1 - Description of Stated Preference Variables, Choice Frequencies

Variable	Description	Mean	Std. Dev.
N=14429 alternatives (1033 respondents x 13 tour packages or trips)			
Volcano	=1 if one or more volcanoes in package	0.325	
Beach	=1 if one or more beaches in package	0.379	
Forest	=1 if one or more forest sites in package	0.551	
River	=1 if one or more river sites in package	0.0922	
Golf	=1 if one or more golf resorts in package	0.0153	
Fishing	=1 if one or more fishing trips in package	0.0150	
Island	=1 if one or more island sites in package	0.0167	
Volcano days	Days at volcano site(s)	0.509	0.862
Beach days	Days at beach site(s)	1.189	2.016
Forest days	Days at forest site(s)	1.731	2.195
River days	Days at river site(s)	0.180	0.619
Price	= Price of the package in thousands of 1998 $US	$315.47	298.38
D_{NP}	Alternative-specific dummy for “No Purchase” option.		
N=1033 respondents			
Age	Age of respondent (1-6 categorical treated as cardinal)	3.20	1.53
Edu	Dummy for “has college education”	0.797	
Single	Dummy for respondent being single	0.512	
Income	Respondent’s (household) income (1998 $US/1000)	44.3	37.4
Male	Dummy for male respondent	0.611	

Package Theme Alternatives	Freq	%
“Beach and Tropical Forest”	100	9.7%
“Rainforest Vacation”	66	6.4%
“Sample of Costa Rica”	198	19.2%
“Beach Vacation”	55	5.3%
“Tropical Adventure Vacation”	130	12.6%
“Birding and Wildlife Vacation”	46	4.5%
One-day trip (miscellaneous) *	42	4.1%
One-day trip #1 to volcano	61	5.9%
One-day trip #2 to volcano	76	7.4%
One-day trip #1 to forest	49	4.7%
One-day trip #2 to forest	39	3.8%
One-day trip #1 to beach	28	2.7%
One-day trip #2 to beach	21	2.0%
None of the Above (NP)	122	11.8%

* This is a trip to a single destination. The type of destination included in each survey is randomly drawn from the following types: forest, river, golf resort, angling trip, and island

Table 2: Coefficient Estimates for Traditional and Generalized Specifications

	Traditional (γ zero)		Model 1 (γ constant)		Model 2 (γ varying)	
Variables	**Coefficient**	**asy. t-ratio**	**Coefficient**	**asy. t-ratio**	**Coefficient**	**asy. t-ratio**
TYPE						
Volcano	-1.444	(-2.98)**	0.0341	(0.05)	0.0243	(0.04)
Beach	-1.715	(-7.96)**	-0.4268	(-1.04)	-0.4309	(-1.01)
Forest	-1.202	(-6.44)**	0.1201	(0.30)	0.1153	(0.27)
River	-2.074	(-2.88)**	-0.6881	(-0.91)	-0.6804	(-0.89)
Golf	-1.969	(-3.35)**	-0.6643	(-1.00)	-0.6685	(-0.96)
Angling	-0.9353	(-2.51)**	0.3735	(0.75)	0.3758	(0.75)
Island	-1.196	(-3.03)**	0.1125	(0.22)	0.1071	(0.21)
TYPE_INTERACT						
(Beach)*(Forest)	1.880	(7.78)**	0.7584	(2.01)**	0.7601	(1.89)*
(Beach)*(Volcano)	1.405	(4.34)**	1.154	(3.84)**	1.155	(3.75)**
(Forest)*(Volcano)	0.0958	(0.33)	-0.9725	(-2.56)**	-0.9697	(-2.39)**
(Volcano)*(River)	1.583	(4.03)**	0.7835	(1.94)*	0.7773	(1.75)*
DAYS_AT_TYPE						
Volcano days	1.257	(2.09)**	1.020	(1.75)*	1.030	(1.80)*
Beach days	0.3441	(3.06)**	0.3539	(3.44)**	0.3548	(3.34)**
Forest days	0.2990	(2.51)**	0.2303	(2.16)**	0.2329	(2.10)**
River days	2.599	(3.10)**	2.002	(2.53)**	1.990	(2.51)**
DAYS_AT_TYPE^2						
(Volcano days)^2	-0.2240	(-1.57)	-0.1872	(-1.36)	-0.1907	(-1.42)
(Beach days)^2	-0.0161	(-1.26)	-0.0185	(-1.57)	-0.0185	(-1.52)
(Forest days)^2	-0.0149	(-1.15)	-0.0066	(-0.56)	-0.0068	(-0.56)
(River days)^2	-0.4886	(-2.52)**	-0.3515	(-1.90)*	-0.3494	(-1.90)*
DAYS_AT_TYPE_INTERACT						
(Beach days)*(Forest days)	-0.0242	(-1.78)*	-0.0245	(-1.94)*	-0.0246	(-1.92)*
(Beach days)*(Volcano days)	-0.0847	(-2.32)**	-0.0843	(-2.53)**	-0.0835	(-2.36)**
(Forest days)*(Volcano days)	-0.0235	(-0.77)	-0.0125	(-0.47)	-0.0128	(-0.45)
(Forest days)*(River days)	-0.0437	(-1.14)	-0.1017	(-2.69)**	-0.1012	(-2.51)**
(Volcano days)*(River days)	-0.2393	(-3.06)**	-0.1904	(-2.69)**	-0.1877	(-2.57)**
-(Price/ $1000)	0.8349	(3.34)**	0.9100	(3.74)**	0.9140	(3.71)**
γ PARAMETER TERMS						
($D_{NP}/10^5$)			2.13	(4.04)**	3.06	(2.45)**
($D_{NP}/10^5$) Age					0.0137	(0.06)
($D_{NP}/10^5$) Edu					-1.78	(-2.10)**
($D_{NP}/10^5$) Single					1.28	(1.75)*
($D_{NP}/10^5$) (Income/10^5)					-1.02	(-1.59)
($D_{NP}/10^5$) Male					0.476	(0.96)
Log-L	-2543.58		-2537.52		-2528.50	
N	1033		1033		1033	

* significant at 0.01 level; ** significant at 0.05 level; persistently insignificant variables dropped.

Table 3: WTP to Avoid Losing Access to a Particular Site-Type from the Real Costa Rica Choice Set; Traditional and Generalized Specifications

Type of Site	Traditional (γ zero)	Model 1 (γ constant)	*% diff wrt trad.*	Model 2 (γ varying)	*% diff wrt trad.*
All Trips	\$13,572	\$13,388	*-1.36%*	\$13,391	*-1.34%*
Beach	\$4,478	\$4,045	*-9.66%*	\$4,052	*-9.51%*
Forest	\$3,653	\$2,062	*-43.56%*	\$2,063	*-43.53%*
River	\$2,616	\$2,259	*-13.64%*	\$2,261	*-13.58%*
Volcano	\$2,801	\$2,101	*-25.01%*	\$2,104	*-24.88%*

Appendix: Sampling distributions for fitted WTP estimates

The body of the paper mentions the difficulty that our particular sample presents for the generation of distributions for the WTP estimates that capture the variances and covariances of the estimated maximum likelihood parameters that enter into the WTP formulas. This appendix provides the results of 1000 simulations for the WTP estimates presented in Table 4, for the traditional model, and describes the problems we have encountered. Analogous problems will afflict our alternative models, so we do not bother to report those results here.

It would have been more satisfying if our data had not exhibited such a high degree of multicollinearity between some of the variables. Fortunately, future studies can be designed with this pitfall in mind and should be able to achieve considerably greater precision in their simulated distributions for WTP values. This data deficiency clouds, but does not diminish, the significance of our finding that the traditional model is too restrictive.

Simulated Distributions of WTP

a.) Marginal distributions of simulated WTP values:

	Mean	Median	2.5 percentile	97.5 percentile
All Trips	$ 15351	$ 13291	$ 8103	$ 33466
Beach	5005	4326	2581	10989
Forest	4126	3442	1923	9317
River	2940	2530	1438	6658
Volcano	3201	2768	1526	6955

b.) Covariance matrix for simulated WTP estimates:

```
             | All trips   Beach   Forest    River   Volcano
-------------+------------------------------------------
All trips    |   1.0000
Beach        |   0.9966   1.0000
Forest       |   0.9917   0.9845   1.0000
River        |   0.9955   0.9932   0.9872   1.0000
Volcano      |   0.9937   0.9932   0.9805   0.9915   1.0000
```

Correlations between attribute variables

Correlations computed across 13429 alternatives = 1033 respondents * 13 non-numeraire alternatives with non-zero attribute levels. (NOTE: Working variable acronyms are used in this display. Other than the initial p=Price variable, the order of variables corresponds to that used in Table 2 in the body of the paper.) Correlations in excess of 0.5 in absolute value are in boldface.

```
(obs=13429)
```

	p	volc	bea	fore	river	golf	fish
p	1.0000						
volc	0.1077	1.0000					
bea	0.1361	-0.3785	1.0000				
fore	**0.5543**	-0.0436	-0.2597	1.0000			
river	0.1793	0.0748	-0.2490	0.1816	1.0000		
golf	-0.0889	-0.0865	-0.0975	-0.1383	-0.0398	1.0000	
fish	-0.0872	-0.0857	-0.0965	-0.1369	-0.0394	-0.0154	1.0000
isla	-0.0934	-0.0903	-0.1018	-0.1443	-0.0415	-0.0163	-0.0161
dayvol	0.2883	**0.8527**	-0.2859	0.1662	0.1545	-0.0738	-0.0731
daybea	0.3793	-0.2864	**0.7549**	-0.0394	-0.1880	-0.0736	-0.0729
dayfore	**0.7139**	0.0353	-0.1582	**0.7117**	0.1321	-0.0984	-0.0974
dayriver	0.1986	0.0898	-0.2268	0.2125	**0.9110**	-0.0362	-0.0359
dayvol2	0.3423	**0.6357**	-0.1904	0.2552	0.1755	-0.0550	-0.0545
daybea2	0.3696	-0.2378	0.5854	0.0098	-0.1457	-0.0571	-0.0565
dayfore2	**0.5917**	-0.0080	-0.1414	0.5034	0.0495	-0.0696	-0.0689
dayriv2	0.1893	0.0870	-0.1958	0.2077	**0.7863**	-0.0313	-0.0310
nexbf	0.4738	-0.0753	0.5296	0.3735	-0.1319	-0.0516	-0.0511
nexbv	0.2210	0.2830	0.2512	0.1550	-0.0625	-0.0245	-0.0243
nexfv	0.4743	**0.6498**	-0.1594	0.4067	0.2262	-0.0562	-0.0557
nexvr	0.1651	0.2946	-0.1596	0.1844	**0.6411**	-0.0255	-0.0252
daybf	0.4090	-0.1352	0.4604	0.3247	-0.1146	-0.0449	-0.0444
daybv	0.1881	0.2340	0.2077	0.1168	-0.0517	-0.0203	-0.0200
dayfv	0.4158	**0.5536**	-0.1648	0.3466	0.0739	-0.0479	-0.0474
dayfr	0.2014	0.0122	-0.1923	0.2222	**0.7722**	-0.0307	-0.0304
dayvr	0.1463	0.2590	-0.1403	0.1621	**0.5636**	-0.0224	-0.0222

	isla	dayvol	daybea	dayfore	dayriver	dayvol2	daybea2
isla	1.0000						
dayvol	-0.0770	1.0000					
daybea	-0.0768	-0.2142	1.0000				
dayfore	-0.1027	0.1793	-0.0432	1.0000			
dayriver	-0.0378	0.1614	-0.1712	0.1482	1.0000		
dayvol2	-0.0574	**0.9423**	-0.1410	0.2285	0.1766	1.0000	
daybea2	-0.0596	-0.1825	**0.9581**	-0.0326	-0.1328	-0.1234	1.0000
dayfore2	-0.0726	0.0831	-0.0784	**0.9419**	0.0591	0.1185	-0.0738
dayriv2	-0.0326	0.1466	-0.1478	0.1424	**0.9688**	0.1575	-0.1146
nexbf	-0.0539	-0.0149	**0.6204**	0.3024	-0.1201	0.0197	**0.5378**
nexbv	-0.0256	0.3358	0.1878	0.0989	-0.0570	0.3087	0.1068
nexfv	-0.0587	**0.8242**	-0.1287	0.3724	0.2332	**0.7887**	-0.1236
nexvr	-0.0266	0.3851	-0.1205	0.0389	**0.6358**	0.3760	-0.0934
daybf	-0.0468	-0.0854	**0.6121**	0.3377	-0.1044	-0.0446	**0.5685**
daybv	-0.0211	0.3436	0.2169	0.0520	-0.0471	0.3551	0.1618
dayfv	-0.0500	**0.8219**	-0.1377	0.4113	0.0788	**0.8552**	-0.1259
dayfr	-0.0321	0.0632	-0.1451	0.2252	**0.8885**	0.0826	-0.1126
dayvr	-0.0234	0.3967	-0.1059	0.0309	**0.6357**	0.4207	-0.0821

	dayfore2	dayriv2	nexbf	nexbv	nexfv	nexvr	daybf
dayfore2	1.0000						
dayriv2	0.0582	1.0000					
nexbf	0.1733	-0.1037	1.0000				
nexbv	0.0373	-0.0492	0.4431	1.0000			
nexfv	0.2221	0.2131	0.0781	0.4061	1.0000		
nexvr	-0.0374	0.5713	-0.0845	-0.0401	0.4534	1.0000	
daybf	0.2314	-0.0901	**0.8692**	0.2125	-0.0192	-0.0735	1.0000
daybv	0.0057	-0.0407	0.3504	**0.8267**	0.3207	-0.0331	0.1950
dayfv	0.2945	0.0725	0.0267	0.2715	**0.8520**	0.2110	-0.0205
dayfr	0.1401	**0.8789**	-0.1018	-0.0483	0.1177	0.3863	-0.0885
dayvr	-0.0351	**0.6170**	-0.0743	-0.0352	0.3986	**0.8791**	-0.0646

	daybv	dayfv	dayfr	dayvr
daybv	1.0000			
dayfv	0.2313	1.0000		
dayfr	-0.0399	0.0418	1.0000	
dayvr	-0.0291	0.2186	0.3876	1.0000

4. Correlations between parameter estimates (from parameter vcov matrix)

Estimated correlations in excess of 0.5 in absolute value are in boldface.

	p	volc	bea	fore	river
p	1				
volc	.04386375	1			
bea	.10798765	.1213127	1		
fore	.11938604	.15544626	.26833797	1	
river	.03631972	.04517431	.08329668	.12422356	1
golf	-.04383343	.03477286	.06229632	.07235165	.01701512
fish	-.07607608	.04489855	.09690253	.11316396	.03290066
isla	-.06528349	.04944517	.09239611	.10825904	.03016146
nexbf	-.1437686	-.06665376	-.25365659	-.34306742	-.06730818
nexbv	.0429013	-.12972521	-.21399092	-.02556067	-.0502816
nexfv	-.1137801	.09625844	-.05726738	-.14696097	.0174822
nexvr	-.01289792	-.12542679	-.12037569	-.16872987	-.1095511
dayvol	-.06624684	**-.96333544**	-.04077475	-.05263581	-.01204673
daybea	-.33325879	-.03249369	-.7269791	-.01224901	-.02063158
dayfore	-.41221431	-.04448761	-.09306342	**-.69302611**	-.05434739
dayriver	-.07799182	-.00643909	-.04746717	-.0788432	**-.95440647**
dayvol2	.00297087	**.90588617**	.015984	.00193433	-.01038038
daybea2	.16416992	.01631596	**.61919259**	-.07870304	.00497392
dayfore2	.24837509	.02265662	.06220332	**.62473119**	.0403515
dayriv2	.03344571	-.02041781	.03271665	.03703024	**.9138275**
daybf	.38108313	-.00205477	.14312699	.3217009	.03560592
daybv	.09292873	.21664755	.20775602	-.01564117	.04696083
dayfv	.23123834	.220712	.01504461	.24756754	-.00163851
dayfr	.19853527	.00846321	.01234775	.17196425	.25131945
dayvr	.11102788	.19726667	.07722196	.08129141	.23621393

	golf	fish	isla	nexbf	nexbv
golf	1				
fish	.0372625	1			
isla	.03478548	.05555955	1		
nexbf	-.04631184	-.07270968	-.0701345	1	
nexbv	-.0106079	-.01182284	-.01258204	-.3679869	1
nexfv	-.03586605	-.06174419	-.05464911	.1391679	-.43618574
nexvr	-.01991057	-.03100513	-.03080776	.0519049	.32016982
dayvol	-.00452859	.00371959	-.00432988	.01489496	.09880498
daybea	.01364151	.02594996	.02083042	-.05537967	.13115018
dayfore	.01159302	.02329805	.0172264	.02886109	.12017917
dayriver	-.00264068	-.00960463	-.00910232	.06065089	.01884238
dayvol2	.00566969	-.001437	.00648695	-.02088828	-.02515067
daybea2	-.00724035	-.01528899	-.01121442	.09203494	-.05367615
dayfore2	-.00388401	-.01060244	-.00613746	.04228574	-.12876241
dayriv2	.00487953	.01279189	.01126284	-.05123334	.0116202
daybf	-.0160456	-.02920887	-.02282071	**-.6401387**	.16637271
daybv	-.00218126	-.00801662	-.0071093	.13690782	-.70122688
dayfv	-.00639758	-.01031298	-.00997882	.04629963	-.01088159
dayfr	-.02473989	-.04119597	-.03183614	.02619692	-.11581835
dayvr	-.00034943	.00340653	.00320161	-.02821696	-.17225646

	nexfv	nexvr	dayvol	daybea	dayfore
nexfv	1				
nexvr	-.41584822	1			
dayvol	-.20279465	.08337664	1		
daybea	-.01224078	.03827799	.03814965	1	
dayfore	-.11990993	.15289837	.03537221	.10031706	1
dayriver	-.02181863	-.02683871	-.0060003	.0347949	.03984547
dayvol2	.21587647	-.02705406	**-.96262189**	-.00874179	.0171959
daybea2	.01312035	-.01001964	-.02263302	**-.94029003**	.0267819
dayfore2	.15805381	-.13569303	-.01214839	-.04664079	**-.95834498**
dayriv2	-.01372271	.07619885	.0298628	-.01629099	.01433866
daybf	.09312836	-.09956641	-.00821988	-.15488067	-.43577638
daybv	.08185196	-.13897107	-.18899918	-.2261918	.02157715
dayfv	-.49007725	.10212108	-.17725487	-.0319226	-.27366684
dayfr	.21290371	-.02060964	-.02478352	-.06962153	-.36163199
dayvr	.11314213	**-.69061912**	-.16080459	-.0740195	-.04491256

	dayriver	dayvol2	daybea2	dayfore2	dayriv2
dayriver	1				
dayvol2	.01811175	1			
daybea2	-.01467971	.01091666	1		
dayfore2	-.01730024	-.01952619	-.03671593	1	
dayriv2	**-.97018771**	-.02400092	.00996826	-.02367484	1
daybf	-.04767237	-.00256113	.00350507	.32702521	.01236781
daybv	-.02759435	.04893229	.15556904	-.04894628	-.00601995
dayfv	.02976831	-.01122621	-.0208924	.19031032	-.05165433
dayfr	-.3585475	-.01168096	.00439974	.30928018	.17988028
dayvr	-.18711396	.0291651	.04020856	.00676821	.07280914

	daybf	daybv	dayfv	dayfr	dayvr
daybf	1				
daybv	-.08936101	1			
dayfv	.03045879	.26331863	1		
dayfr	.21165086	.03939091	.03022792	1	
dayvr	.02631581	.28169031	.18888942	.26043746	1