CORRECTING FOR ON-SITE SAMPLING IN RANDOM UTILITY MODELS

Klaus Moeltner and J. Scott Shonkwiler

This study demonstrates how the joint distribution of a set of conditional trip counts to a system of recreation-sites can be adjusted for on-site sampling. An econometric approach is proposed that addresses both the size-biased distribution of the sampled visits and the weighted distribution of reported visits to ancillary destinations in a multivariate random utility framework. Estimation results indicate that uncorrected models produce biased estimates of trip counts and welfare measures. The empirical application examines jet skiing in the Lake Tahoe region.

Key words: Multi-site system, On-site sampling, Random parameters, Random utility models

Running Head: On-Site Sampling in Random Utility Models

Klaus Moeltner is assistant professor, and J. Scott Shonkwiler is professor, Department of Resource Economics, University of Nevada, Reno. This research was supported in part by the Nevada Agricultural Experiment Station, publication # 51042915. We would like to thank three anonymous reviewers for their helpful comments.

On-site sampling can be an economical way to gather information on users of recreational amenities, such as hiking trails, campsites, or beaches. When only a small segment of the population of interest is expected to visit the site under investigation, on-site sampling may often be the only cost-effective strategy to collect sufficient visitation data for statistical analysis. However, in most efforts to study recreation demand researchers are ultimately interested in visitation levels and welfare effects associated with the "average" user in the underlying population.

This raises the problem that compared to such a representative individual, recreationists intercepted on-site are likely more avid participants in the activity of interest, and more frequent visitors to the examined site. In statistical terms, their visitation behavior is driven by a different probability distribution from the one specified by the researcher for the general population. Patil and Rao have studied this issue in a general context. They label distributions associated with observations collected "by nature" weighted distributions. If the probability of inclusion in a given sample is commensurate to the size of the variable to be recorded, the weighted distribution is also called a size-biased distribution. This applies to our context, where the notion of size is related to visitation frequency, i.e. the total number of visits over the research period for a given individual intercepted on-site. More avid participants are more likely to be included in an on-site sample. Thus, their visitation behavior follows a size-biased version of the statistical visitation model stipulated for the general population of users.

Shaw and Englin and Shonkwiler (1995a) examine this problem in the context of recreation demand. They refer to size-biased sampling as "endogenous stratification".

They find that ignoring size-bias in on-site sampling leads to inconsistent estimation results and misleading welfare measures. They also illustrate how to correct econometric models for size-biased sampling for a variety of count data specifications. A detailed discussion of specification issues related to on-site sampling is also given in Haab and McConnell (chapter 7). However, these existing applications focus implicitly or explicitly on the demand for only a single recreation-site for a specific respondent.

In many circumstances, researchers are interested in visitation demand for several sites, especially if sites are close substitutes for a given outdoor activity. In that case, an efficient and economical strategy to gain visitation information for the entire recreation system is to elicit information on trips to *all* sites in the system from a visitor interviewed at a given destination. Such an approach, while cost effective, will extend weighted sampling problems to all sites in the system, since the individual's visits to the other sites are likely correlated with visits to the site of interception. For example, if more visits to the sampled site are associated with fewer visits to the remaining sites for a given person, demand for the other sites may be underestimated relative to the true demand for these sites by a representative user from the underlying population.

In this study we propose an econometric approach that addresses both the size-biased distribution of the sampled visits and the weighted distribution of visits to ancillary destinations in a multivariate random utility framework. We illustrate the relationships between weighted distributions and the underlying latent distribution specified for the general population, and compare estimation results and welfare effects of corrected and uncorrected specifications. We apply our theoretical framework to the estimation of the demand for jet skiing in the Lake Tahoe region of the central Sierra Nevada.

Model Formulation

The proper statistical approach for addressing multivariate size-biased sampling is presented in Jain and Nanda. We specialize their results to systems of count data. Assume a given recreation area includes sites *j=1...J*. The joint probability mass function (pmf) of trips to the *J* destinations for a member of the general population is denoted as

$$p(\mathbf{y}) = prob(Y_1 = y_1, Y_2 = y_2, \cdots, Y_J = y_J) \tag{1}$$

where Y_j is the random variable "trips to site j" and y_j denotes an actually observed trip count to site *j, j=1...J*. Assume further that a given respondent is intercepted at site *k* and asked to report trips to all sites in the system for the research period of interest. As illustrated in Jain and Nanda the size-biased joint pmf, p($\mathbf{y}^s$), is then given by

$$p(\mathbf{y}^s) = \left(\frac{y_k}{E[Y_k]}\right) \cdot p(\mathbf{y}) \qquad y_{j \neq k} = 0,1,2,\cdots \qquad y_k = 1,2,\cdots \tag{2}$$

where superscript *s* indicates that on-site sampling is taking place at one of the sites in the system, *E* denotes the expectation operator, and p($\mathbf{y}$) is the original joint pmf given in (1).

The moments of the *marginal* pmf for the site at which sampling occurs can then be derived as

$$\begin{aligned} &E[Y_k^s] = E[Y_k]^{-1} \sum_{y_k} \sum_{\mathbf{y}_{j \neq k}} y_k^2 p(\mathbf{y}) = E[Y_k] + \frac{V[Y_k]}{E[Y_k]} \qquad \text{and} \\ &\mathrm{V}[Y_k^s] = E[Y_k]^{-1} \sum_{y_k} \sum_{\mathbf{y}_{j \neq k}} (y_k - E[Y_k^s])^2 y_k p(\mathbf{y}) = \frac{E[Y_k^3]}{E[Y_k]} - \frac{(V[Y_k])^2}{(E[Y_k])^2} - 2V[Y_k] - (E[Y_k])^2 . \end{aligned} \tag{3}$$

The analogous expressions for non-intercept sites in the system are

$$E\left[Y_j^s\right]=E\left[Y_k\right]^{-1}\sum_{y_k}\sum_{y_j}\sum_{\mathbf{y}_{\neq j,k}} y_j y_k p(\mathbf{y})=\frac{E\left[Y_j Y_k\right]}{E\left[Y_k\right]}=E\left[Y_j\right]+\frac{COV\left[Y_j,Y_k\right]}{E\left[Y_k\right]} \quad \text{and}$$

(4) $$V\left[Y_j^s\right]=E\left[Y_k\right]^{-1}\sum_{y_k}\sum_{y_j}\sum_{\mathbf{y}_{\neq j,k}} y_j^2 y_k p(\mathbf{y})-E\left[Y_j^s\right]^2=\frac{E\left[Y_j^2 Y_k\right]}{E\left[Y_k\right]}-\frac{COV\left[Y_j,Y_k\right]^2}{E\left[Y_k\right]^2}- \\ -E\left[Y_j\right]^2-2COV\left[Y_j,Y_k\right],$$

where sampling takes place at site k, site j is one of the other destinations, and vector $\mathbf{y}_{\neq j,k}$ comprises all remaining sites in the system. It can be noted that under independence of Y_j and Y_k the expressions in (4) collapse to their corresponding uncorrelated moments. Both (3) and (4) indicate that moments associated with individual site counts under on-site sampling can be derived directly from the original moments of the corresponding marginal distributions. We will exploit these relationships in our empirical specification below.

We adopt the Dirichlet multinomial (DM) distribution first derived by Mosimann, and recently applied in the context of recreation demand analysis by Shonkwiler and Hanley to specify the the joint pmf for the general population, p(**y**). As described in the latter study, this distribution arises when the cell probabilities in a multinomial pmf, say $\pi_1,\pi_2,\cdots\pi_J$ follow a Dirichlet distribution with parameters $\alpha\theta_1,\alpha\theta_2,\cdots\alpha\theta_J$. A detailed derivation of the DM distribution is given in Shonkwiler and Hanley. The advantage of the DM distribution over the multinomial (MN) distribution is its ability to capture overdispersion in the data. This phenomenon occurs when seemingly identical units of observations (individuals or zones of origins) differ substantially in their number of visits to one or more sites in a given recreation system. The DM models such variability in trip counts through a dispersion parameter (α). As shown in Shonkwiler and Hanley, as α approaches infinity or, equivalently, $1/\alpha$ goes to zero, the DM collpases to the MN. This allows for convenient testing of the presence of overdispersion for a given sample of recreationists.

Under the DM the joint pmf for a representative visitor from the underlying population is thus given as

$$(5)\qquad p(\mathbf{y}\mid\alpha,\boldsymbol{\theta},Y)=\frac{Y!\,\Gamma(\alpha)\prod_{j=1}^{J}\Gamma(y_j+\alpha\theta_j)}{\Gamma(Y+\alpha)\prod_{j=1}^{J}\left(y_j!\,\Gamma(\alpha\theta_j)\right)}\qquad \theta_j\in S_U,\, j=1\cdots J,\qquad \alpha>0\,.$$

The notation is as follows: y_j is the number of trips to site j as above, Y is the (presumed exogenous) total number of trips to the recreation system taken by the individual during the research period, Γ denotes the mathematical gamma function, α is the dispersion parameter, the θ_j-terms can be interpreted as single-occasion choice probabilities, and S_U is a *J-1* dimensional unit simplex defined as $S_U=\{(\theta_1,\theta_2,\cdots,\theta_J):\theta_j>0,\sum_{j=1}^{J}\theta_j=1\}$. The closer α is to zero, the more diverse is the combined visitation pattern of all users from what would be expected under multinomial sampling. The marginal distribution of y_j is Beta-binomial, and can be written as

$$(6)\qquad p(y_j\mid\alpha,\theta_j,Y)=\frac{Y!\,\Gamma(\alpha)\Gamma(y_j+\alpha\theta_j)\Gamma(Y-y_j+\alpha(1-\theta_j))}{\Gamma(Y+\alpha)y_j!\,\Gamma(\alpha\theta_j)(Y-y_j)!\,\Gamma(\alpha(1-\theta_j))}\quad .$$

The moments of this marginal pmf take the form of

$$
(7)\qquad
\begin{aligned}
&E\left[y_j \mid \alpha,\theta_j,Y\right]=Y\theta_j\\
&V\left[y_j \mid \alpha,\theta_j,Y\right]=\rho Y\left(\theta_j-\theta_j^2\right)\\
&Cov\left[y_j,y_k \mid \alpha,\theta_j,Y\right]=-\rho Y\theta_j\theta_k \qquad \text{where } \rho=\left(\frac{Y+\alpha}{1+\alpha}\right).
\end{aligned}
$$

As indicated by (7), these moments reduce to those of the binomial distribution, i.e. the marginal pmf of the MN, as ρ approaches one, or, equivalently, as α approaches infinity.

Applying the relationships described in equations (2) through (4), we can now examine how the MN and DM distributions can be adjusted for on-site sampling. As before, we assume sampling takes place at site *k*. The on-site corrected joint pmf's for the two specifications are given as

$$
(8)\qquad
\begin{aligned}
&p\left(\mathbf{y}^s \mid \boldsymbol{\pi},Y\right)=\left(\frac{y_k}{\pi_k Y}\right)\cdot\frac{Y!\prod_{j=1}^{J}\pi_j^{y_j}}{\prod_{j=1}^{J}y_j!} \qquad \text{and}\\
&p\left(\mathbf{y}^s \mid \alpha,\boldsymbol{\theta},Y\right)=\left(\frac{y_k}{\theta_k Y}\right)\cdot\frac{Y!\Gamma(\alpha)\prod_{j=1}^{J}\Gamma\left(y_j+\alpha\theta_j\right)}{\Gamma(Y+\alpha)\prod_{j=1}^{J}\left(y_j!\Gamma\left(\alpha\theta_j\right)\right)},
\end{aligned}
$$

respectively.[1] The moments of the associated marginal pmf's for the site of interception (= site k), and remaining sites (= sites $j \neq k$) take the following form for the multinomial distributions:

$$
(9)\qquad
\begin{aligned}
&E\left[y_k^s \mid \pi_k,Y\right]=1+\pi_k(Y-1)\\
&E\left[y_{j\neq k}^s \mid \pi_j,Y\right]=\pi_j(Y-1)\\
&V\left[y_k^s \mid \pi_k,Y\right]=\pi_k\left(1-\pi_k\right)(Y-1)\\
&V\left[y_{j\neq k}^s \mid \pi_j,Y\right]=\pi_j\left(1-\pi_j\right)(Y-1)
\end{aligned}
$$

The analogous expressions for the DM distribution are

$$
\begin{aligned}
&E\left[y_k^s \mid \alpha,\theta_k,Y\right] = \rho + \theta_k\left(Y-\rho\right)\\
&E\left[y_{j\neq k}^s \mid \alpha,\theta_j,Y\right] = \theta_j\left(Y-\rho\right)\\
(10)\qquad &V\left[y_k^s \mid \alpha,\theta_k,Y\right] = \theta_k\left(Y-\rho\right)\left(\frac{3Y+\alpha-\alpha\rho\theta_k-4}{2+\alpha}+2\left(1-\rho\right)\right)+\\
&\qquad\qquad + \frac{2\left(Y-1\right)\left(Y-2\right)}{\left(1+\alpha\right)\left(2+\alpha\right)} + \left(1-\rho\right)\left(\rho-2\right)\\
&V\left[y_{j\neq k}^s \mid \alpha,\theta_j,Y\right] = \theta_j\left(Y-\rho\right)\left(Y+\alpha-\alpha\rho\theta_k\right)/\left(2+\alpha\right) \qquad \text{where } \rho = \frac{Y+\alpha}{1+\alpha}.
\end{aligned}
$$

Clearly, these moments are different from the original moments of the underlying marginal pmf's for both the site of sampling and the remaining destinations. Furthermore, if Y=1, i.e. a given person visits the system of sites only once, the variance terms collapse to zero, and the expectations reduce trivially to one for the site of interception, and zero for the other destinations. This is consistent with the fact that each sampled individual makes at least one visit to the system, and that our statistical framework is conditioned on the total number of visits to the set of destinations. Consequently, an observation of Y=1 (and thus ρ=1) implies that a given individual visited only the site of interception.

The remaining specification task entails the linking of the choice probabilities to an underlying utility-theoretic framework. We choose the random utility model (RUM) of McFadden due to its useful properties for analyzing discrete visitation data and its ability to generate exact per-trip welfare measures for site quality changes. We thus assume that person i's indirect utility for site j takes the linear form

$$
(11)\qquad U_{ij} = v_{ij} + \varepsilon_{ij} \quad ,
$$

where v_{ij} is parameterized to depend upon observed conditioning variables, and ε_{ij} is an idiosyncratic error term. Furthermore, we specify v_{ij} to include random parameters, i.e.

$$(12) \quad \begin{array}{l} v_{ij} = \beta_i' x_{ij} \quad \text{where} \\ \beta_i \sim mvn(\bar{\beta}, \Sigma) \end{array}$$

and *mvn* denotes the multivariate normal distribution. Combining this assumption with a generalized extreme value distribution for ε_{ij} yields the by now well known mixed logit (ML) expression for site choice probabilities (e.g. McFadden and Train), i.e

$$(13) \quad \text{Prob (person i chooses site j} \mid \mathrm{x_{ij}}) = \int_{\boldsymbol{\beta}_i} \left(\frac{\exp(\mathrm{v_{ij}} \mid \boldsymbol{\beta}_\mathrm{i})}{\sum_{\mathrm{k=1}}^{\mathrm{J}} \exp(v_{ik} \mid \boldsymbol{\beta}_\mathrm{i})} \right) f(\boldsymbol{\beta}_i)\, d\boldsymbol{\beta}_i$$

where $f(\boldsymbol{\beta}_i)$ refers to the probability density function (pdf) of random vector $\boldsymbol{\beta}_i$, and the dimensionality of the integral is commensurate with the number of random parameters in $\boldsymbol{\beta}_i$. The properties of the ML formulation have been discussed at length in the discrete choice literature (e.g. Train, 1998, 1999; McFadden and Train). Perhaps the most important feature of the ML is its ability to accommodate correlation patterns across choice probabilities. This eliminates the requirement of independence of irrelevant alternatives (IIA) that must be satisfied by standard conditional logit models. In addition, the random coefficients in (13) capture unobserved heterogeneity of preferences for site attributes. This is important in our application, as some of the site characteristics included in our empirical model have the potential to contribute to indirect utility in vastly different fashion across users. We apply (13) to specify π_j and θ_j, $j=1 \ldots J$, in our MN and DM framework, respectively.[2] As a result, the expressions in equations (5) through (10) are now to be interpreted as "conditioned on $\boldsymbol{\beta}_\mathrm{i}$". The unconditional counterparts can be derived by integrating over the random elements of $\boldsymbol{\beta}_i$, in analogy to (13).

To allow for a systematic examination of the effects of on-site correction and overdispersion, we estimate four different specifications for the joint pmf of trip counts: a multinomial model with and without on-site correction, and a Dirichlet-multinomial model with and without on-site correction. All four models employ the mixed logit formulation to parameterize site choice probabilities. Accordingly, we label the four models as "mixed multinomial logit" (MMNL), "on-site corrected mixed multinomial logit" (oMMNL), "mixed Dirichlet-multinomial" (MDM), and "on-site corrected mixed Dirichlet-multinomial" (oMDM).

As shown above, the four models are associated with different joint probability mass functions. Specifically, on-site corrected specifications have an added factor of observed over expected trip count for the sampled site in their likelihoods. The expectation term in the denominator of this factor, in turn, is a function of site probability, and thus of estimated parameters. This implies that the likelihood function for uncorrected models is misspecified. Consequently, estimated parameters will be biased. In contrast, omitting the dispersion parameter α in a multinomial logit model merely reduces efficiency. As discussed in Shonkwiler and Hanley, the multinomial logit specification will still generate consistent parameter estimates. In summary, if the oMDM is the correct specification, estimated parameters generated by the MMNL and MDM will be biased, and those flowing from the oMMNL will be consistent but inefficient. However, if the dispersion parameter is indeed significant (i.e. its inverse is significantly different from zero), the oMMNL will employ an incorrect expression to predict trip counts, as evident by a comparison of equations (9) and (10). In contrast, per-trip *welfare* measures generated by the oMMNL are consistent, as the underlying formula does not include the dispersion parameter (see equation (17) later in the text). Naturally, estimates for both trip counts and welfare produced by the MMNL and MDM will be biased, as they build on flawed model parameters. In addition, these uncorrected specifications also employ an incorrect expression for trip predictions as indicated above.

All models are estimated through simulated maximum likelihood (SML) (e.g. McFadden and Train). Conditional on $\boldsymbol{\beta}_i$, a given individual's contribution to the likelihood function (LHF), l_i, can be concisely expressed as

(14) $$l_i \mid \boldsymbol{\beta}_i = p(\mathbf{y}_i \mid Y_i) \mid \boldsymbol{\beta}_i$$

where $\mathbf{y}_i$ is a vector of person i's observed trip counts to sites 1 through J, Y_i denotes the total number of visits to the recreation system observed for person i, and $p(.)$ indicates the joint pmf associated with one of the four models listed above. Drawing R sets of β_i, estimating (14) for each set, and averaging over estimates yields an approximation of the unconditional LH component for person i,

(15) $$\tilde{l}_i = \frac{1}{R}\sum_{r=1}^{R} l_{ir} \qquad \text{where } l_{ir} = p(\mathbf{y}_i \mid Y) \mid \boldsymbol{\beta}_{ir} \quad .$$

Summing the log of (15) over all individuals produces the simulated log-LHF for the entire sample. We set R=1000 repetitions in our application.[3]

Data

The data for this analysis stem from an on-site survey of jet skiers implemented during the summer seasons of 2001 and 2002 at six lakes and reservoirs in the Tahoe region of the central Sierra Nevada. The survey was administered in-person by several interview teams on selected weekdays and weekends between the end of May and the end of August during both seasons. Interview days were approximately evenly distributed across lakes, with equal counts of specific days-of-the-week for each destination. Each respondent was asked to provide information on details for the trip of interception, information on the number of trips taken to the six lakes during the current season up to the interview date, and planned trips yet to be taken throughout the remainder of the current season.[4] Additional information important to this analysis includes technical details on the vehicle and jet ski used for the trip, as well as the gender and age of other household members present at the site. Overall, the survey effort generated 333 valid sets of responses, about evenly distributed over the two seasons.

For this analysis, we retain all individuals who (i) took no more than 40 trips to the system of sites in the given time frame and (ii) spent no more than one day at the site. The resulting data set comprises 187 completed questionnaires, yielding a panel of 187 x 6 = 1122 trip-counts for the recreation system. Table 1 summarizes some basic lake and trip characteristics for this sample. As can be seen from the table, visitors report a total of 1671 trips to the recreation system. The largest numbers of seasonal trips are observed for Lahontan and Boca reservoirs. Both destinations offer numerous easy access and launching points, generally free of charge. Distances from visitor origin to destinations are comparable across lakes, with means in the 50 to 70 mile range. Another noteworthy feature captured in the table is the ban on 2-stroke engines for all watercraft in effect at Tahoe for both survey seasons. The detailed information on jet skis collected in the survey allows for the identification of such models and thus of visitors who were affected by the ban.

Estimation Results

Estimation results for the four models described in the previous section are given in Table 2. All models share the same set of explanatory variables, i.e. price of access ("price", computed as round trip distance times per-mile driving costs and the opportunity cost of time[5]), lake elevation in 1000 feet ("elevation"), lake surface in units of 10 square miles ("surface"), an indicator variable equal to one if a given lake does not offer jet ski rental and the interviewed visitor does not own a jet ski ("no rent"), an indicator variable equal to one for the combined outcome of "respondent owns a jet ski banned at Tahoe" and "site = Tahoe" ("ban"), and a site specific constant for Donner Lake.[6] The latter is included to explicitly capture public access and parking associated constraints at this site due to the preponderancy of private property along its shoreline. To allow for heterogeneity in preferences for "elevation" and "surface", we specify these two attributes to be associated with random coefficients. All models are estimated using robust standard errors (White) to guard against misspecification of the variance-covariance matrix of the estimated parameters.[7]

Generally, all four models fit the underlying data well, with the majority of estimated parameters significant at 5% or higher. The inverse of the dispersion parameter for the DM distributions emerges as different from zero with high significance, indicating the presence of overdispersion in the data. The estimated standard deviations for the two random coefficients are also significant at the 1% level, confirming heterogeneity in preferences for these two lake characteristics. It should be noted that the random coefficients in the MNL models do not address overdispersion in trip counts. This is evidenced by the joint significance of dispersion effects *and* second moments for random parameters in the DM specifications. Intuitively, random parameters are limited in their capacity to model variability in trip counts since they are directly linked to observed regressors. Thus, their effect on-site-choice probabilities hinges on both the variability in their associated regressors across sites, and on the functional relationship between these explanatory variables and trip counts. On the other hand, overdispersion as captured in the DM models applies directly to the entire pmf of y_j and can thus capture variability in the dependent variable that is not directly related to included regressors.[8]

The "elevation"- variable is best interpreted as a proxy for both water and air temperatures, with higher altitude signifying lower magnitudes for both measures. Estimates for mean and standard deviation of the associated random parameter differ across models. The oMMNL and MDM models estimate the mean to lie arbitrarily close to zero, which implies that approximately 50% of visitors consider warmer temperatures a positive site feature, while the other half of jet skiers appears to prefer cooler temperature levels. In contrast, the estimated parameter mean is significantly positive for the MMNL model and significantly negative for the oMDM specification. Relating these mean values to their respective standard deviations translates into preferences for warmer temperatures for 64% of visitors according to the MMNL and 33% of visitors based on the oMDM. This dichotomy in tastes is likely related to needs and preferences for wet suits during the early and late parts of the season. In mid-summer, the pleasantly cooler air temperatures at high mountain sites are likely to become the main driver of this elevation effect.

A larger lake surface constitutes a negative attribute for the majority of visitors as indicated the negative and significant estimate for the parameter mean for the associated random coefficient in the first three models. Relating mean estimates to standard deviations reveals that between 65% and 75% of jet skiers prefer smaller lakes. For the oMDM, this proportion reduces to 50%, given that the estimated parameter mean is insignificantly different from zero. This moderate preference for smaller water bodies for our sample of recreationists is probably related to the enhanced opportunities offered by smaller lakes to explore various segments of the shoreline even during a shorter visit and potentially also to safety concerns, i.e. the ability to reach shore or to secure help in case of technical problems with the jet ski.

The remaining variables have the expected sign in all four models. A higher access price has a negative effect on visitation, as do lacking rental facilities, the ban on certain jet ski models at Tahoe, and the Donner Lake indicator.

Generally, there is substantial variability in the magnitude of estimates across models. Parameter estimates generated by the MMNL models are generally larger in absolute value than those produced by the MDM counterparts. The same holds for estimates produced by the MMNL vis-a-vis those generated by the oMMNL. Within the MDM pair, coefficients tend to be smaller in absolute magnitude for the on-site corrected estimator.

A series of specification tests was conducted to provide a more rigorous assessment of the appropriateness of the four models given the underlying data. Test results are shown in Table 3. The upper part of the table identifies the pair of models to be compared, the type of test, and statistical test results. The bottom half offers a verbal interpretation of these results. As mentioned above, the MDM nests the MMNL. The same holds for the oMDM with respect to the oMMNL. This allows for the application of a robust Wald test (White) for the null hypothesis (H_0) that $1/\alpha$ equals zero. As indicated in the table, this hypothesis is strongly rejected for both pairs of models. To test for the appropriateness of on-site correction, we employ Vuong's test for non-nested models. As can be seen from the table, the test clearly rejects the uncorrected specifications in favor of the on-site-corrected versions for both MMNL and MDM models. Overall, this series of tests identifies the oMDM as the specification that is most compatible with the underlying data.

Trip Predictions

Table 4 compares sample statistics and model predictions for the average number of trips to each site. Sample results are captured in the first set of columns. The column denoted "on-site" refers to trip counts for all visitors intercepted at a given destination. The column labeled “off-site”, in turn, captures trips reported to sites other than the location of interception. The “all” column depicts the average over all on-site or off-site trips associated with a given lake. The relevant table entries for actual sample results are mean and standard deviation (“std”) of trip counts for each site.

As is evident from the table, average seasonal trip counts associated with on-site data exceed off-site averages by an order of magnitude for all sites. Specifically, average trip counts corresponding to off-site locations are less than one trip per season for all destinations, while on-site averages range from close to three trips for site 2 (Donner Lake) to over 10 trips per season for site 6 (Topaz Lake). These sample results are consistent with our assumption that respondents have stronger preferences for the site at which they are intercepted than for other destinations in the system. This validates the choice of on-site corrected statistical specifications to appropriately model visitation behavior.

It should be noted that the sample averages captured in the “all” column are *not* indicative of latent user demand for the wider population. To elicit latent demand per site, estimated parameters from on-site corrected models need to be employed in the expression for expected visits commensurate with the stipulated underlying pmf for the population of interest. For the oMDM, for example, these expressions are

$$(16) \quad \begin{aligned} &E\left[y_{ij} \mid \alpha, Y_i\right] = \int_{\boldsymbol{\beta}_i} \rho + \left(\theta_{ij} \mid \beta_i\right)\left(Y_i - \rho\right) f\left(\boldsymbol{\beta}_i\right) d\boldsymbol{\beta}_i \quad \text{and} \\ &E\left[y_{ij} \mid \alpha, Y_i\right] = \int_{\boldsymbol{\beta}_i} \left(\theta_{ij} \mid \beta_i\right)\left(Y_i - \rho\right) f\left(\boldsymbol{\beta}_i\right) d\boldsymbol{\beta}_i \end{aligned}$$

for the site of interception and an alternate site, respectively. We simulate these terms by drawing R vectors of $\boldsymbol{\beta}_i$ from their estimated distribution, computing the expression under the integral in (16) for each vector and respondent, averaging the resulting conditional trip counts over draws for each respondent, and averaging over respondents for each site.[9] The same simulation procedure is employed to derive average visitation counts per site for the MMNL, oMNNL and the MDM models. For the MMNL and MDM we preserve the “all” label for their respective columns in the table, as these models do not distinguish between on-site and off-site counts, and, analogous to the sample statistics, simply pool on-site and off-site data with equal weights.

As for the actual sample, we report the mean and standard deviation of predicted trips over respondents for each model. In addition, the table captures the difference between the mean of trip predictions for the latent population flowing from a given model and the mean generated by the oMDM (“d”). As discussed above, expected trip counts for all models other than the oMDM are biased due to misspecified likelihood functions or incorrect expressions for visitation moments. To assess the significance of this bias, we relate the difference in means to its standard deviation (“std(d)”). The exact process of deriving this standard deviation is described in the Appendix. As can be seen from the table, all differences-in-means are substantially larger than their associated standard deviations, generally by a factor of five to eight. In other words, all differences are significantly different than zero, which provides an empirical confirmation of the theoretical bias in trip predictions associated with the MMNL, oMMNL, and MDM.

We also generate estimated average counts per site based on on-site and off-site data, respectively, to allow for an assessment of the corrected models’ ability to accurately reproduce sample results. The simulation process to produce these estimates is identical to the methodology described above, with replacement of the expression for conditional expectations in (16) with the analogous terms for on-site and off-site counts as given in (9) and (10). The resulting means and standard deviations are given in the columns denoted "on-site" and "off-site" for the oMMNL and oMDM models in the table. As evident from the table, the oMDM matches sample results much more closely than the oMMNL for both on-and off-site visits.

Welfare Analysis

From a policy perspective, relevant welfare measures in this application are those that assess the economic impact on jet skiers of potential bans on jet ski use at lakes and reservoirs in the central Sierra Nevada. As mentioned above, some older jet ski models are already banned at Lake Tahoe due to environmental considerations. Alternatively, water needs during drought periods may require drawing down water reserves at some of these reservoirs to levels that make it impossible to launch jet skis and other vessels, leading de facto to a temporary site closure for all motorized use. An increased awareness of the associated welfare changes to recreational users may aid water planners in deriving socially efficient water management strategies.

As shown in Shonkwiler and Hanley, per-trip welfare estimates for the DM model can be computed using the same "log-sum" formula (e.g. Hanemann) as typically employed for MNL models. In our application, per-trip compensating variation (CV) for a given visitor and site is given by

$$(17) \qquad E\left[CV_{ij}\right] = -\int_{\boldsymbol{\beta}_i} \frac{1}{\beta_p}\left(\log\left(\sum_{J-1}\exp\left(v_{ij}\right)\right) - \log\left(\sum_{J}\exp\left(v_{ij}\right)\right)\right) f\left(\boldsymbol{\beta}_i\right) d\boldsymbol{\beta}_i$$

where β_p is the price coefficient, and v_{ij} is the observed component of indirect utility as given in (11). The summation over *J-1* sites indicates that one destination has been eliminated from the system. We approximate the average over coefficient draws and visitors of this expression for each site through the same simulation process as outlined above for trip counts.

Table 5 captures the same statistical measures for per-trip CV as those employed in Table 4 for trip counts. In addition to the empirical standard deviation of estimated welfare results over respondents ("std"), the table also captures the asymptotic standard deviation of the mean to compare the efficiency of welfare estimates between the oMDM and the oMMNL ("std(mean)"). Specifically, we use this statistic to compute an asymptotic 95% confidence interval for the welfare mean. We then follow Moeltner, and Shonkwiler and Hanley by reporting a statistic that relates the point estimates of the mean to the spread of its confidence interval. This indicator of relative efficiency is denoted as "spread-over-mean" (s.o.m) in the table. The computation of the standard deviation and the confidence interval for the mean is described in the Appendix.

With few exceptions, mean welfare estimates produced by the oMDM specification are larger for than those generated by the other models. This difference is quite pronounced for most destinations. More importantly, all but two differences associated with the uncorrected models are well over three standard deviations above or below zero. Analogous to our findings for trip counts, this lends empirical evidence to the bias in parameters and thus welfare predictions associated with models that are unadjusted for on-site sampling. As noted above, welfare estimates produced by the oMMNL are, in theory, consistent. The table results support this notion to some extent, as differences in mean CV between the oMMNL and the oMMD are substantially less pronounced for all but the last site than analogous differences for the uncorrected models.

An examination of the efficiency gains associated with the oMDM relative to the oMMNL reveals that the oMDM generates slightly tighter s.o.m values for four of the sites. We thus find only weak evidence that controlling for overdispersion in on-site corrected models increases efficiency of welfare predictions in this application.

Overall, per-trip welfare effects associated with seasonal closure of individual sites appear somewhat modest in magnitude when compared to per-trip welfare measures associated with related forms of motorized (non-fishing) water recreation reported in the literature (e.g. Englin and Shonkwiler, 1995b, Loomis and Walsh). This can be explained to a large extent by the fact that most of these existing estimates are related to single destination models that do not capture substitution effects, or where close substitutes are simply not available. In addition, it should be kept in mind that the welfare estimates reported in Table 5 refer to the representative visitor from the underlying population of users. In contrast to our individuals intercepted on-site, these average users likely consider several of the destinations in our system as close substitutes and likely perceive the closure of any given site as less damaging than would avid visitors to that location.

Conclusion

This study examines how the problem of size-biased sampling can be addressed in a multivariate random utility framework. Specifically, we demonstrate how the joint distribution of a set of conditional trip counts associated with a system of recreation-sites can be adjusted for on-site sampling. We further illustrate how this corrected model permits the recovery of consistent estimates for underlying population parameters and visitation counts. A comparison of corrected and uncorrected specifications reveals that uncorrected models are plagued by a misspecification of the joint probability mass function and the expressions for the conditional mean of trip counts and consequently generate biased trip and welfare estimates.

In addition, our results offer empirical evidence that the specification of random coefficients in the expressions for site choice probabilities, while useful to overcome IIA problems and capture unobserved heterogeneity in preferences, is not an appropriate strategy to address either size-biased sampling or overdispersion in a multivariate random utility model with on-site data collection. A more rigorous examination of the properties of mixed logit kernels in this context could constitute a fruitful and important avenue for further research.

Appendix: Derivation of Population Statistics for Tables 4 and 5

To preempt the need for site subscripts and without loss of generality this appendix focuses on a specific site in our system. The simulation process to derive the entries in Tables 4 and 5 can be segmented into the following steps:

1. For each respondent *i*, draw R=100,000 sets of coefficient vectors $\boldsymbol{\beta}_i$. Then compute the estimation measure of interest, say per-trip CV (using the integrand in equation 17), for each draw *r* and person *i*. This yields an *n* (= sample size) by *R* matrix of welfare estimates.

2. For each respondent, compute the mean of welfare estimates over all coefficient draws (i.e. compute the mean for each row in the above matrix). This yields an *n* by 1 vector of unconditional (over $\boldsymbol{\beta}_i$) means of per-trip CV estimates

$$\mathbf{CV} = \begin{bmatrix} cv_1 \\ cv_2 \\ \vdots \\ cv_n \end{bmatrix}.$$

3. Compute the empirical mean $\bar{c}v = \frac{\sum_{i=1}^{n} cv_i}{n}$ and standard deviation $s = \sqrt{\frac{\sum_{i=1}^{n} (cv_i - \bar{c}v)^2}{n-1}}$ for this vector. The resulting values correspond to the statistics labeled "mean" and "std" in the tables.

4. Invoke the Central Limit Theorem and approximate the asymptotic standard deviation for $\bar{c}v$ as $\sigma_{\bar{c}v} \cong \frac{s}{\sqrt{n}}$. Compute a 95% confidence interval for the population mean as $c.i. = \{\bar{c}v \pm 1.96 \cdot \sigma_{\bar{c}v}\}$, and spread over mean ("s.o.m." in Table 5) as $\bar{c}v$ over the width of $c.i.$

5. Denote one of the uncorrected models as model 1 and the oMDM as model 2 and assign the same subscripts to all related statistics introduced above. Compute the difference in means as $d = \bar{c}v_1 - \bar{c}v_2$ (entry "d" in the tables). Since vectors $\mathbf{CV}_1$ and $\mathbf{CV}_2$ are based on the same data and are thus correlated, derive the asymptotic standard deviation for d as $std(d) = \sqrt{\frac{s_1^2 + s_2^2 - 2s_{1,2}}{n}}$, where $s_{1,2}$ is the empirical covariance between vectors $\mathbf{CV}_1$ and $\mathbf{CV}_2$. This measure is denoted as "std(d)" in Tables 4 and 5.

Footnotes

[1] It should be noted that there exists a convenient transformation of the generic MN distribution into the size-biased MN. Just as Englin and Shonkwiler (1995a) suggest that the size-biased Poisson pmf could be reproduced by applying the standard Poisson to on-site data from which one trip is subtracted, the size-biased MN can be derived from the MN by reducing visits to site k, the location of sampling, by one. This yields

$$p(\mathbf{y} \mid \boldsymbol{\pi}, Y) = \frac{(Y-1)! \prod_{j=1}^{J} \pi_{j \neq k}^{y_{j \neq k}} \pi_k^{y_k - 1}}{\prod_{j=1}^{J} y_{j \neq k}! (y_k - 1)!}.$$

This expression corresponds precisely to the pmf of the sized biased MN given in (8).

[2] While both πj in the MN and θj in the DM distribution share the same role of denoting the probability of choosing site *j* at a specific choice occasion faced by a given individual, we will maintain the distinctive notation for the two distributions throughout the remainder of the text for clarity of exposition.

[3] The Matlab algorithm to implement these models is available from the authors upon request.

[4] Restrictions on survey length preempted collecting trip details for visits other than the one intercepted on-site. Our analysis thus rests on the implicit underlying assumption that a) respondents correctly estimate the number of future trips during the remainder of the season for each site, and b) relevant trip details (such as vehicle type and group composition) remain largely unchanged over all trips for a given respondent and

season. Assumption a) is supported by the extremely stable weather conditions (dry and warm) that govern the research region during the summer months, i.e. respondents are unlikely to be forced to cancel projected trips due to adverse weather events.

[5] We arbitrarily specify per-mile driving costs at \$0.3 for jet ski renters, and \$0.4 for jet ski owners to allow for a "load penalty". The opportunity cost of time is derived as travel time in hours times 1/3 of the hourly wage rate.

[6] We also estimated specifications that included interaction terms of site characteristics with visitor attributes, such as basic demographic information and secondary activities associated with the site visit (hiking, fishing, swimming), in addition to random parameters for elevation and surface. However, none of these terms was estimated to have a significant effect on visitation probabilities for our sample of respondents. This may well be different in other applications.

[7] For the DM models, this could occur through a misspecification of the dispersion parameter itself. It should be noted that robust standard errors do not compensate for any misspecification of the conditional mean function of the elements of **y**. Such misspecification is likely the case in models that fail to employ size-biased distributions to analyze data collected through on-site sampling.

[8] In our case, random coefficients only affect the magnitude of choice probabilities π_j, $j=1...J$. These terms, in turn, enter expressions for trip variance in the MNL models as $\pi_j(1-\pi_j)$ (see for example equ. (9)). The maximum value that this expression can take is 0.25. However, variability in trip counts for our sample is much

more pronounced than $0.25Y$ for the MMNL, or $0.25(Y-1)$ for the oMMNL for many visitor-site combinations.

[9] In this simulation process, $\boldsymbol{\beta}_i$ also includes all non-random coefficients. Draws for these coefficients are taken from their empirical distribution following the procedures outlined in Krinsky and Robb. For the two random coefficients in our model, we first draw a set of r_1 vectors for mean and standard deviations from their respective *empirical* distributions. For each draw, means and standard deviations are combined to specify the *estimated* distribution for the random coefficients. For each of these distributions, r_2 subsets of random coefficients are generated, leading to a total of $R=r_1 + r_2$ sets of random parameters for the overall simulation process. In our application, we set both r_1 and r_2 equal to 100.

References:

Englin, J., and J. S. Shonkwiler. "Estimating Social Welfare Using Count Data Models: An Application to the Long-Run Recreation Demand under Conditions of Endogenous Stratification and Truncation." *The Review of Economics and Statistics* 77 (1995a): 104-12.

________________. "Modeling Recreation Demand in the Presence of Unobservable Travel Costs: Toward a Travel Price Model." *Journal of Environmental Economics and Management* 29 (1995b): 368-377.

Haab, T. C., and K. E. McConnell. *Valuing Environmental and Natural Resources: The Econometrics of Non-Market Valuation*. Cheltenham, U.K. and Northampton, Massachusetts: Edward Elgar, 2002.

Hanemann, W. M. "Welfare Analysis with Discrete Choice Models." *Valuing Recreation and the Environment*. J. Herriges, and C. Kling, eds., pp. 33-64. Northampton MA: Edward Elgar, 1999.

Jain, K., and A. K. Nanda. "On Multivariate Weighted Distributions." *Communications in Statistics - Theory & Methods* 24 (1995): 2517-2539.

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

Loomis, J. B., and R. G. Walsh. *Recreation Economic Decisions*. State College, Pennsylvania: Venture Publishing, Inc., 1997.

McFadden, D. "Conditional Logit Analysis of Qualitative Choice Behavior." *Frontiers in Econometrics*. P. Zarembka, ed., pp. 105-142. New York: Academic Press, 1974.

McFadden, D., and K. Train. "Mixed Mnl Models for Discrete Response." *Journal of Applied Econometrics* 15 (2000): 447-470.

Moeltner, K. "Addressing Aggregation Bias in Zonal Recreation Models." *Journal of Environmental Economics and Management* 45 (2003): 128-144.

Mosimann, J. E. "On the Compound Multinomial Distribution, the Multivariate ß-Distribution, and Correlations among Proportions." *Biometrika* 49 (1962): 65-82.

Patil, G. P., and C. R. Rao. "Weighted Distributions and Size-Biased Sampling with Applications to Wildlife Populations and Human Families." *Biometrics* 34 (1978): 179-189.

Shaw, D. "On-Site Samples' Regression - Problems of Non-Negative Integers, Truncation, and Endogenous Stratification." *Journal of Econometrics* 37 (1988): 211-223.

Shonkwiler, J. S., and N. Hanley. "A New Approach to Random Utility Modeling Using the Dirichlet Multinomial Distribution." *Environmental and Resource Economics* (2003): 401-416.

Train, K. E. "Recreation Demand Models with Taste Differences over People." *Land Economics* 74 (1998): 230-239.

___________. "Mixed Logit Models for Recreation Demand." *Valuing Recreation and the Environment*. J. Herriges, and C. Kling. eds., pp. 121-140. Northampton MA: Edward Elgar, 1999.

Vuong, Q. H. "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses." *Econometrica* 57 (1989): 307-333.

White, H. "Maximum Likelihood Estimation of Misspecified Models." *Econometrica* 50 (1982): 1-25.

Table 1. Sample Characteristics

Lake	Elevation (Feet)	Surface Area (Squ. Miles)	Shoreline (Miles)	Jet Ski Rental	Ban on 2-Stroke Engines	Distance* (Miles, One Way)	On-site Interviews	Trips
Boca	5700	1.5	15	no	no	54.0	31	370
Donner	5969	1.5	7.5	yes	no	58.0	51	184
Lahontan	4150	23.2	69	no	no	74.0	33	399
Stampede	5949	5.4	25	no	no	60.8	26	180
Tahoe	6230	190.8	75	yes	yes	63.6	18	224
Topaz	5012	4.4	25	yes**	no	77.9	28	314
Total:							187	1671

*Mean over all respondents.

**Rental at Topaz started in 2002.

Table 2. Estimation Results

	MMNL			oMMNL			MDM			oMDM		
Variable	Coeff.	S.E.		Coeff.	S.E.		Coeff.	S.E.		Coeff.	S.E.	
Price	-0.115	(0.012)	***	-0.114	(0.014)	***	-0.060	(0.007)	***	-0.045	(0.007)	***
No Rent	-4.296	(0.845)	***	-3.475	(0.934)	***	-4.061	(0.632)	***	-2.011	(0.622)	***
Ban	-3.935	(0.876)	***	-1.846	(0.704)	***	-1.575	(0.995)		-2.189	(1.029)	**
Donner	-1.623	(0.430)	***	-1.570	(0.398)	***	-0.878	(0.238)	***	-0.759	(0.329)	**
Elevation(100ft)	0.184	(0.041)	***	0.052	(0.063)		-0.018	(0.025)		-0.047	(0.022)	**
Surface(10 sq. m.)	-0.232	(0.043)	***	-0.194	(0.036)	***	-0.049	(0.029)	*	0.026	(0.029)	
1/α	-	-	-	-	-	-	0.689	(0.126)	***	1.034	(0.167)	***
St. Deviations												
Elevation	0.493	(0.072)	***	0.477	(0.113)	***	0.125	0.036	***	0.107	(0.038)	***
Surface	0.314	(0.042)	***	0.405	(0.050)	***	0.118	(0.033)	***	0.136	(0.043)	***
Lhf	-771.214			-675.565			-615.252			-459.917		

Notes: White-corrected standard error in parentheses.

* Indicates significance at 10%. ** Indicates significance at 5%. *** Indicates significance at 1%.

n=1122 based on 187 day-visitors.

Table 3. Specification Tests

Test	First Model (Ho, Constrained)	Second Model (Ha, Unconstrained)	Type of Test	Test Distribution Name	D.o.f.	Test Stat.	P-value
	Presence of Overdispersion						
1	MMNL	MDM	Robust Wald	Chi-squ.	1	119.05	0.000
2	oMMNL	oMDM	Robust Wald	Chi-squ.	1	153.52	0.000
	Appropriatenes of On-site Correction						
3	MMNL	oMMNL	Vuong (non-nested)	t	1114	-7.77	0.000
4	MDM	oMDM	Vuong (non-nested)	t	1113	-14.31	0.000

Test	Results (in word)
1	Reject Ho: no overdispersion
2	Reject Ho: no overdispersion
3	On-site corrected model strictly better than uncorrected version
4	On-site corrected model strictly better than uncorrected version

Table 4. Predicted Trip Counts

Site		Sample			MMNL	oMMNL			MDM	oMDM		
		All	On-site	Off-site	All	Pop.	On-site	Off-site	All	Pop.	On-site	Off-site
1	Mean	1.98	9.74	0.44	2.07	1.86	6.01	1.01	1.90	1.48	9.40	0.43
	Std	4.73	6.98	1.73	3.04	2.72	3.31	1.84	2.78	2.01	5.50	0.72
	D	-	-	-	0.59	0.38	-	-	0.42	-	-	-
	Std(D)	-	-	-	0.09	0.06	-	-	0.07	-	-	-
2	Mean	0.98	2.86	0.28	1.13	0.90	2.27	0.51	1.09	0.76	2.69	0.28
	Std	3.30	5.67	1.14	2.28	1.90	3.11	0.92	2.11	1.21	4.09	0.33
	D	-	-	-	0.37	0.14	-	-	0.33	-	-	-
	Std(D)	-	-	-	0.09	0.06	-	-	0.07	-	-	-
3	Mean	2.13	9.49	0.56	1.76	2.21	5.60	1.44	2.07	2.41	8.47	0.78
	Std	4.84	7.56	1.55	2.87	3.29	5.34	2.14	3.72	3.63	6.65	1.19
	D	-	-	-	-0.65	-0.21	-	-	-0.35	-	-	-
	Std(D)	-	-	-	0.07	0.04	-	-	0.05	-	-	-

Note: D represents mean - mean(oMDM).

Table 4. Predicted Trip Counts (continued)

Site		Sample			MMNL	oMMNL			MDM	oMDM		
		All	On-site	Off-site	All	Pop.	On-site	Off-site	All	Pop.	On-site	Off-site
4	Mean	0.96	4.58	0.38	1.36	0.98	1.94	0.88	1.02	0.90	4.84	0.39
	Std	2.23	3.01	1.37	1.75	1.28	0.97	1.28	1.45	1.20	3.51	0.60
	D	-	-	-	0.46	0.08	-	-	- 0.12	-	-	-
	Std(D)	-	-	-	0.06	0.03	-	-	- 0.03	-	-	-
5	Mean	1.20	8.06	0.47	1.50	1.92	4.62	1.50	1.42	2.14	7.11	0.79
	Std	4.69	12.60	1.65	2.93	2.94	7.19	1.91	3.10	3.46	10.82	1.06
	D	-	-	-	-0.64	-0.23	-	-	- -0.73	-	-	-
	Std(D)	-	-	-	0.07	0.07	-	-	- 0.07	-	-	-
6	Mean	1.68	10.21	0.18	1.11	1.08	6.11	0.24	1.44	1.24	9.00	0.24
	Std	5.69	11.34	0.98	3.45	3.31	6.76	0.76	3.72	2.81	8.57	0.39
	D	-	-	-	0.26	0.68	-	-	- 0.20	-	-	-
	Std(D)	-	-	-	0.06	0.05	-	-	- 0.07	-	-	-

Note: D represents mean - mean(oMDM).

Table 5. Welfare Estimates

Site		MMNL	oMMNL	MDM	oMDM
1	Mean	2.84	2.67	4.04	4.20
	Std	3.28	2.92	4.47	3.47
	Std (mean)	0.24	0.21	0.33	0.25
	S.o.m.	0.33	0.31	0.32	0.24
	D	-1.36	-1.53	-0.16	-
	Std(D)	0.08	0.07	0.11	-
2	Mean	4.80	3.35	6.50	3.69
	Std	7.04	5.03	9.75	4.29
	Std (mean)	0.51	0.37	0.71	0.31
	S.o.m.	0.42	0.43	0.43	0.33
	D	1.11	-0.35	2.81	-
	Std(D)	0.23	0.11	0.41	-
3	Mean	6.22	8.43	6.02	8.88
	Std	8.78	10.23	11.27	11.53
	Std (mean)	0.64	0.75	0.82	0.84
	S.o.m.	0.40	0.35	0.54	0.37
	D	-2.65	-0.45	-2.86	-
	Std(D)	0.23	0.17	0.14	-

Notes: std(mean) represents the asymptotic standard deviation of the mean.
S.o.m. represents spread over mean (= mean divided by length of confidence interval).
D represents mean - mean(oMDM).

Table 5. Welfare Estimates (continued)

Site		MMNL	oMMNL	MDM	oMDM
4	Mean	1.36	0.97	1.78	2.31
	Std	1.14	0.76	1.65	1.58
	Std (mean)	0.08	0.06	0.12	0.12
	S.o.m.	0.24	0.23	0.27	0.20
	D	-0.95	-1.34	-0.53	-
	Std(D)	0.07	0.08	0.04	-
5	Mean	5.96	9.70	5.74	13.54
	Std	5.75	6.35	6.60	10.01
	Std (mean)	0.42	0.46	0.48	0.73
	S.o.m.	0.28	0.19	0.33	0.21
	D	-7.59	-3.85	-7.81	-
	Std(D)	0.42	0.41	0.33	-
6	Mean	2.87	2.73	4.70	4.37
	Std	7.36	6.95	9.11	6.87
	Std (mean)	0.54	0.51	0.67	0.50
	S.o.m.	0.73	0.73	0.56	0.45
	D	-1.50	-1.64	0.33	-
	Std(D)	0.11	0.11	0.19	-

Notes: std(mean) represents the asymptotic standard deviation of the mean.
S.o.m. represents spread over mean (= mean divided by length of confidence interval).
D represents mean - mean(oMDM).