Equation Chapter 1 Section 1 Mixed Poisson Regression Models with Individual Panel Data from an On-site Sample

by

Kevin Egan
Department of Economics, University of Toledo

Joseph Herriges†
Department of Economics, Iowa State University

February 11, 2004

* The authors would like to thank Scott Shonkwiler, Jeff Englin, and the participants of the 2003 Western Regional Project meetings (W-1133) for many helpful comments on an earlier draft of the paper, as well as two anonymous referees and the Co-Editor for their many helpful suggestions. Any remaining errors are, as usual, our own.

† Corresponding author: Joseph A. Herriges, 260 Heady Hall, Department of Economics, Iowa State University, Ames, IA 50011-1070. Phone: (515)294-4964, Fax: (515)294-0221, Email: jaherrig@iastate.edu.
Mixed Poisson Regression Models with Individual Panel Data from an On-site Sample

Abstract

The purpose of this paper is to consider the problem of controlling for on-site sampling in the context of a system (or panel) of demand equations. Specifically, in the context of recreation demand, we are concerned with the situation in which survey respondents are asked to provide information not only about the actual trips to a specific site (observed behavior), but also their anticipated trips (either under current conditions or given price and quality changes). A multivariate Poisson-log normal (MPLN) model is used to jointly model the observed and contingent behavior data and to correct for on-site sampling.

Keywords: Count data model, on-site sampling, recreation
I. Introduction

Cost considerations often drive analysts to rely upon intercept (or on-site) surveys to collect information about recreation demand at a site (or sites) of interest. This guarantees that survey respondents will include users of the resource in question. Unfortunately, the sampling procedure also comes at a cost of both truncation (excluding non-users) and endogenous stratification (over sampling those individuals who are more frequent users of the site). As a result, the sample is no longer representative of the broader population. Failure to correct for on-site sampling will result in biased estimates of recreation demand and any corresponding welfare estimates.

There have been a number of papers in the literature focused on controlling for intercept sampling in recreation demand analysis. Shaw (1988) develops a correction for both the truncation and endogenous stratification problems in the case of a single site Poisson count data model. Englin and Shonkwiler (1995) subsequently extended Shaw's correction to the case of the Negative Binomial (NB) count data model. The advantage of the NB model is that it allows for overdispersion (i.e., the situation in which the conditional mean number of trips is less than the conditional variance of trips), a common characteristic of recreation demand data. The limitation of both of these efforts is that they are focused on a single demand equation.

The purpose of this paper is to consider the problem of controlling for on-site sampling in the context of a system (or panel) of demand equations. Specifically, we are concerned with the situation in which survey respondents are asked to provide information not only about the actual trips to a specific site (observed behavior), but also their anticipated trips (either under current conditions or given price and quality changes).
The latter trip data, typically known as contingent behavior data, has been used to study the impact of changing environmental conditions (See, e.g., Rosenberger and Loomis, 1999; Whitehead et al., 2000; and Grijalva et al., 2002). Unfortunately, if the observed and contingent behavior (CB) data are collected through an on-site survey, the sampling problems become more complex. The observed behavior data are, as before, subject to truncation and endogenous stratification. While the contingent behavior data are not directly impacted, they are incidentally truncated and endogenously stratified. That is, while the sampling does not exclude individuals who anticipate zero trips in the future, they are less likely because the sampling procedure has excluded individuals who took zero trips in the past and oversampled individuals who, at least in the past, frequently took trips and thus are characterized by a general participation avidity. As a result, it is important to model the observed and contingent behavior data in a panel data framework, controlling for correlation between these data sources and the sampling mechanism used.

In this paper, the multivariate Poisson-log normal (MPLN) model is used to jointly model the observed and contingent behavior data and to correct for on-site sampling. Aitchison and Ho (1989) first suggested the MPLN model but did not include regressors in their analysis. Munkin and Trivedi (1999) estimate a bivariate PLN model. The advantage of the MPLN specification is the fact that, as Shonkwiler (1995) notes, "...only the multivariate Poisson-lognormal distribution can both reproduce an arbitrary correlation structure and account for overdispersion." We modify the MPLN model to control for on-site sampling.

The resulting model is used to analyze survey data collected on-site at Clear Lake in north central Iowa. Specifically, the survey data included observed trips for 2000 and
contingent behavior trips for 2001 under both current prices and two sets of price increases. We find a substantial bias results if the sampling procedures are ignored, overstating both the average number of trips to the site (by a factor of 14) and the welfares associated with the recreational opportunities at Clear Lake.

Finally, while the focus of this paper is on the spillover effects that on-site sampling can have for follow-up contingent behavior questions, the problems extend to the multi-site setting as well. Individuals intercepted at any one site are going to be biased towards those who have avidity towards trips to that site. Their reported patterns of trips to substitute sites will not, in general, be representative of the population as a whole. In developing the on-site sampling correction in the context of CB surveys, we also discuss the issues associated with extending the model to a multi-site setting.

II. Correcting for On-Site Sampling

It has long been recognized that, while on-site (or intercept) surveys provide a convenient mechanism for insuring that a sample includes site users, the resulting sample is no longer representative of the population as a whole. This section provides an overview of the corrections developed for the single-site setting. These corrections are then extended for the multivariate scenario.

A. The Univariate Model

Shaw (1988) was the first to recognize the complex set of problems that characterize on-site samples in recreation demand analysis. In addition to the count nature of the data (i.e., non-negative integers), he notes that on-site surveys exclude those who do not visit the site (truncation) and over sample those who frequent the site.
regularly (endogenous stratification). His correction for these problems, based on the Poisson regression model, is both intuitive and easy to implement.

Shaw (1988) begins by assuming that population trips to the single site of interest follow a univariate Poisson distribution. That is,

\[
f(y|x) = \frac{\exp(-\lambda)(\lambda)^y}{y!}, \quad y = 0, 1, 2, \ldots
\]  

(1)

where \( y \) denotes the number of trips taken by an individual,

\[
\lambda = E(y|x)
\]

\[
= \exp(\beta'x)
\]  

(2)

denotes the expected number of trips for an individual with characteristics vector \( x \), and \( \beta \) denotes the unknown parameters of the distribution to be estimated.

In correcting for the on-site sampling, Shaw assumes that visitors taking \( y \) trips are \( y \) times more likely to be sampled than someone who takes only one trip. The distribution of trip data collected onsite (\( \tilde{y} \)) is then the product of the population distribution and the odds (relative to an average individual) of being included in the sample; i.e.,

\[
g(\tilde{y}|x) = \frac{\tilde{y}}{E(y|x)} f(\tilde{y}|x) \quad \tilde{y} = 0, 1, 2, \ldots
\]  

(3)

Patil and Rao’s (1978) Theorem 1 implies that the corresponding mean for the on-site sample is then given by:

\[
E(\tilde{y}|x) = E(y|x) + \frac{V(y|x)}{E(y|x)}.
\]  

(4)
Thus, the degree of bias in using an on-site sample mean to measure the population mean depends upon the degree of overdispersion in the population (i.e., the ratio of $V(y \mid x)$ to $E(y \mid x)$). This makes intuitive sense as an increased variance in the underlying trip data implies that there will be more extreme, high frequency users, dominating the on-site sample and pulling the on-site trip average up. At the other extreme, as the distribution of trips degenerates to a constant (i.e., $V(y \mid x) \to 0$) the on-site sample and the population means will converge.

Using the Poisson distribution in equation (1), Shaw derives the on-site sampling distribution to be

$$g(\tilde{y} \mid x) = \frac{\tilde{y} \exp(-\lambda)(\lambda)^{\tilde{y}}}{\tilde{y}!}, \quad \tilde{y} = 0, 1, 2, \ldots$$

$$= \begin{cases} 
0 & \tilde{y} = 0 \\
\exp(-\lambda)(\lambda)^{\tilde{y}-1}/(\tilde{y}-1)! & \tilde{y} = 1, 2, \ldots
\end{cases}$$

(5)

The form of the on-site sample’s distribution is convenient since it can be estimated using standard statistical packages designed to estimate a Poisson regression model. The only change required for on-site sampling is to replace $\tilde{y}$ with $\tilde{y} - 1$ as the dependent variable.

One limitation of Shaw’s model is, like all Poisson models, it imposes the assumption of equidispersion; i.e.,

$$\lambda = E(y \mid x) = Var(y \mid x).$$

(6)

In practice, however, recreation demand data typically exhibit overdispersion with the conditional trip variance exceeding the conditional trip mean. Following the logic of Shaw, Englin and Shonkwiler (1995) extend the on-site corrections to the negative
binomial model. Specifically, if population trips are characterized by the negative
binomial distribution

\[ f(y|x) = \frac{\Gamma(y + \alpha^{-1}) \lambda^{y} \alpha^{y} (1 + \alpha \lambda)^{-(y + \alpha^{-1})}}{\Gamma(y + 1) \Gamma(\alpha^{-1})}, \quad (7) \]

then the on-site sample will be characterized by the distribution

\[ g(\tilde{y}|x) = \frac{\tilde{y} \Gamma(\tilde{y} + \alpha^{-1}) \tilde{\lambda}^{\tilde{y} - 1} (1 + \alpha \tilde{\lambda})^{-(\tilde{y} + \alpha^{-1})}}{\Gamma(\tilde{y} + 1) \Gamma(\alpha^{-1})}. \quad (8) \]

In this case the mean and variance for the on-site sample are

\[ E(\tilde{y}|x) = \lambda + 1 + \alpha \lambda \quad (9) \]

and

\[ Var(\tilde{y}|x) = \lambda \left(1 + \alpha + \alpha \lambda + \alpha^{2} \lambda\right), \quad (10) \]

allowing for overdispersion and reducing to Shaw’s Poisson model as \( \alpha \to 0 \).

**B. The Multivariate Setting**

The results of the previous section apply only to the univariate setting. However, there are many examples in practice where a system of counts must be modeled. This is the case, for example, if intercept surveys are conducted at several sites simultaneously or if trip data are gathered at a single site for a series of years or under a series of hypothetical or actual scenarios. Laitila (1999) has addressed the former problem using independent Poisson distributions for each site and conditioning on the total number of trips taken. In this paper, we focus our attention on the latter problem. As noted above, the latter scenario has arisen in recent years, as recreation demand surveys frequently ask not only for information on past trips (observed behavior), but also inquire as to changes...
in trip behavior in future years and under hypothetical changes to the recreation site of interest (contingent behavior). We begin this section by reviewing the multivariate count data models and then develop corrections to those models for on-site samples.

1. **Multivariate Count Data Models**

   The simplest extension of the univariate Poisson count data model to the multivariate setting is to assume that trip data follow independent Poisson distributions. Specifically, if \( y_j \) denotes the number of trips that an individual would take (or has taken) under scenario \( j \), then the joint conditional distribution for the vector of trips \( y = (y_1, \ldots, y_J) \)' is given by

   \[
   f(y | x) = \prod_{j=1}^{J} \frac{\exp(-\lambda_j) (\lambda_j)^{y_j}}{y_j!}, \quad y_j = 0, 1, 2, \ldots \tag{11}
   \]

   where

   \[
   \lambda_j = E(y_j | x_j) = \exp(\beta_j' x_j) \tag{12}
   \]

   and \( x = (x_1, \ldots, x_J) \)'.

   The problem with the model in (11) is that the assumption of independence is unlikely to hold in practice. Individuals who have taken a large number of trips in the past (say \( y_1 \) ) are also likely to take a large number of trips in the future or under proposed changes to the site being studied (i.e., \( y_2, \ldots, y_J \) ). There have been a number of multivariate count data models developed in the literature to allow for correlation across counts for the same individual. Most of these models are mixed Poisson specifications.
that allow for a common shared source of unobserved heterogeneity in the counts for a
given individual. Mixed Poisson models begin by assuming that there is an unobserved
factor (or factors), \( v_j = \exp(\epsilon_j) \), associated with trips taken under scenario \( j \). If \( v_j \) were
known, then the corresponding trips would follow a standard Poisson process, with

\[
f(y_j | x_j, v_j) = \frac{\exp(-\tilde{\lambda}_j)(\tilde{\lambda}_j)^{y_j}}{y_j!}, \quad y_j = 0, 1, 2, \ldots
\]  

and

\[
E(y_j | x_j, v_j) = \tilde{\lambda}_j = \lambda_j v_j = \exp(\beta_j' x_j + \epsilon_j),
\]

With the \( v_j \) (or equivalently \( \epsilon_j \)) being unobserved, the relevant distribution for \( y_j \)
becomes

\[
f(y_j | x_j) = \int \cdots \int \prod_{j=1}^{J} \frac{\exp(-\tilde{\lambda}_j \exp(\epsilon_j))(\tilde{\lambda}_j \exp(\epsilon_j))^{y_j}}{y_j!} h(\epsilon_j) \, d\epsilon_1 \cdots d\epsilon_J, \quad y_j = 0, 1, 2, \ldots
\]

where \( h(\epsilon_j) \) denotes the pdf for \( \epsilon_j \). Thus, the distribution of the trip vector, \( y \), becomes
a mixture of Poisson distributions. There are two consequences of this mixing process.
First, the equidispersion assumption in equation (6) will no longer apply to the individual
trip data (i.e., the \( y_j \)’s). Second, allowing for correlation among the \( \epsilon_j \)'s across scenarios
(\( j \)) for a given individual will induce correlation among the corresponding \( y_j \)'s for that
individual.

In this paper, we will focus our attention on one such mixed Multivariate Poisson
model, the Multivariate Poisson-Lognormal distribution (MPLN).\(^7\) The MPLN model
was introduced by Aitchison and Ho (1989) and derives its name from the fact that the vector $\nu_i$ is assumed to follow a multivariate lognormal distribution, or equivalently that $\varepsilon_j$ follows a multivariate normal distribution; i.e.,

$$\varepsilon_{ij} \sim N(0, \Omega).$$  \hspace{1cm} (16)

Substituting this distributional assumption into (15), we then have that

$$f(y_{ij} | x_j) = \prod_{j=1}^{J} \frac{\exp\left(-\frac{\hat{\lambda}_j}{y_j!}\right) \exp\left[-\frac{1}{2} \varepsilon_j^t \Omega^{-1} \varepsilon_j\right]}{(2\pi)^{J/2} |\Omega|^{1/2}} d\varepsilon_j, \hspace{0.5cm} y_j = 0, 1, 2, \ldots$$  \hspace{1cm} (17)

The trip means and variances become

$$E[y_{ij} | x_j] = \hat{\lambda}_j \exp\left(\frac{1}{2} \sigma_j^2\right) \equiv \delta_j$$  \hspace{1cm} (18)

and

$$Var[y_{ij} | x_j] = \delta_j + \exp\left(\sigma_j^2\right) - 1 \delta_j,$$  \hspace{1cm} (19)

where $\sigma_j^2 = Var(\varepsilon_j | x_j)$. Thus, equidispersion results only if $\sigma_j \to 0$. Correlation among the trips emerges because

$$Cov[y_{ij}, y_{ik}] = \delta_j \left[\exp\left(\sigma_{jk}\right) - 1\right] \delta_k, \hspace{0.5cm} j \neq k,$$  \hspace{1cm} (20)

where $\sigma_{jk}$ denotes the $(j,k)^{th}$ element of $\Omega$. One of the attractive features of the MPLN specification is that it does not restrict the sign of this correlation. The correlation between trips for two distinct scenarios $j$ and $k$ can be positive, negative or zero and depends directly upon the sign of the corresponding $\sigma_{jk}$. The downside of the MPLN specification is that, at the estimation stage, the pdf in (17) requires integration over a $J$-dimensional integral. However, either standard numerical procedures or simulation
techniques can be used to address this problem as long as the number of scenarios, $J$, remains relatively small.

An alternative to the MPLN model is the Multivariate Poisson Gamma (MPG) specification. In this case, it is assumed that there is a single unobserved factor, $u$, shared by all trip scenarios for the same individual; i.e.,

$$v_j = u \forall j$$

and that $u$ follows a $\text{gamma}(\alpha, \alpha)$ distribution with a mean of 1 and a variance of $\alpha^{-1}$.

Substituting this assumption into (15) yields

$$f(y_j | x_j) = \frac{\Gamma\left(\sum_{j=1}^{J} y_j + \alpha\right) \alpha^{\alpha\left(\sum_{j=1}^{J} \lambda_j + \alpha\right)} \prod_{j=1}^{J} \frac{\lambda_j^{y_j} \alpha}{y_j!}}{\Gamma(\alpha) \prod_{j=1}^{J} \prod_{j=1}^{y_j} \lambda_j^{y_j}}$$

The corresponding means and variances are given by

$$E\left[y_j | x_j\right] = \lambda_j$$

and

$$V\left(y_j | x_j\right) = \lambda_j + \alpha^{-1}\left(\lambda_j\right)^2.$$ 

Thus, the degree of overdispersion is a decreasing function of $\alpha$. The covariance between trip responses for a given individual becomes

$$\text{Cov}\left[y_j, y_k\right] = \alpha^{-1}\lambda_j \lambda_k.$$ 

One advantage of the MPG specification is the closed form nature of the count probabilities in equation (22), avoiding the need for numerical or simulation based integration when estimating the model. However, unlike the MPLN, the MPG imposes
considerable structure on the correlation among the counts, requiring the correlations to always be positive and driven by the single parameter $\alpha$.

2. **Controlling for On-Site Sampling**

The problem of on-site sampling emerges for the application we are considering because the first of the trip scenarios, $j=1$, corresponds to current trips to the site in question. Thus, $\tilde{y}_1$ is the truncated version of the population variable $y_1$, excluding observations with $y_1 = 0$, and endogenously stratified, with the sample over representing individuals that frequently visit the site. The relationship between the mean of $y_1$ and the on-site sample for $\tilde{y}_1$ has the same form as equation (4). If we were only interested in observed trip behavior, then on-site corrected versions of the univariate Poisson, Negative Binomial (both described in the previous section), or the univariate PLN model could be applied. However, individuals visiting the site are asked not only about their actual trip taking behavior to the site, but also about how often they plan to visit the site in future years and under a variety of possible changes to the site, generating a vector of on-site trip counts $\tilde{y}_{11} = (\tilde{y}_1, \tilde{y}_2, \ldots, \tilde{y}_J)'$. The contingent behavior trips $\tilde{y}_{1-} = (\tilde{y}_2, \ldots, \tilde{y}_J)'$, while not directly truncated or endogenously stratified, are impacted by the on-site nature of the survey through the correlation between $y_1$ and $y_{1-}$. Analogous to Moeltner and Shonkwiler’s (2005) multi-site situation, the on-site contingent behavior trips have means that are related to their population counterparts as follows:
\[
E(\tilde{y}_j \mid x_i) = E(y_j \mid x_i) + \frac{\text{Cov}(y_i, y_j \mid x_i)}{E(y_j \mid x_i)}
\]
\[= E(y_j \mid x_i) + \text{Corr}(y_i, y_j \mid x_i) \frac{V(y_j \mid x_i)}{E(y_j \mid x_i)} \]^{1/2} \quad (26)

The second line of equation (26) emphasizes that the bias from using an onsite sample for contingent behavior responses depends upon (1) the correlation between observed trips and the contingent behavior responses; (2) the degree of overdispersion in the contingent behavior data; and (3) the relative variability between the actual trips and the contingent trips.

Following the same logic as Shaw (1988) used in the univariate case, the distribution for the on-site sample vector of observed and contingent behavior trips is given by
\[
g_1(\tilde{y}_j \mid x_i) = \frac{\tilde{y}_i}{E(\tilde{y}_j \mid x_i)} f(\tilde{y}_j \mid x_i), \quad \tilde{y}_i = 1, 2, \ldots; \tilde{y}_j = 0, 1, \ldots \text{ for } j = 2, \ldots, J \quad (27)
\]
where the subscript 1 is used to denote the fact that the on-site sampling directly impacts the trips for scenario $j=1$. If the trips are independently distributed and each follow a Poisson process, then
\[
g_1(\tilde{y}_j \mid x_i) = \frac{\exp(-\lambda_i)(\lambda_i)^{\tilde{y}_j - 1}}{\tilde{y}_j - 1} \prod_{j=2}^{J} \frac{\exp(-\lambda_j)(\lambda_j)^{\tilde{y}_j}}{\tilde{y}_j!}, \quad \tilde{y}_i = 1, 2, \ldots; \tilde{y}_j = 0, 1, \ldots \quad (28)
\]
If the MPLN specification applies, however, then
\[
g_1(\tilde{y}_j \mid x_i) = \frac{\tilde{y}_i}{\delta_i} \int \cdots \int \frac{\exp(-\hat{\lambda}_j)(\hat{\lambda}_j)^{\tilde{y}_j}}{\tilde{y}_j!} \frac{\exp\left[-\frac{1}{2} \varepsilon' \Omega^{-1} \varepsilon \right]}{(2\pi)^{j/2} |\Omega|^{1/2}} d\varepsilon, \quad \tilde{y}_i = 1, 2, \ldots; \tilde{y}_j = 0, 1, \ldots (29)
\]
Using equations (4) and (26), the corresponding means for the on-site sample become:
\[
E(\tilde{y}_j | x_i) = \begin{cases} 
1 + \delta_1 \exp(\sigma_1^2) & j = 1 \\
\delta_j + \delta_1 \left[ \exp(\sigma_{1j}) - 1 \right] & j \neq 1
\end{cases}
\] (30)

If \( \sigma_i = 0 \), then \( \delta_i = \lambda_i \), there is no overdispersion associated with site 1, and the on-site sample mean is simply one more than the population mean. Likewise, if there is no correlation among the site alternatives (conditional on the observables), then the on-site sample mean reduces to the population mean of \( \delta_j \).

A similar correction applies for the MPG specification, yielding:

\[
g_i(\tilde{y}_i | x_i) = \frac{\tilde{y}_j \Gamma \left( \sum_{j=1}^f \tilde{y}_j + \alpha \right) \alpha^\alpha \left( \sum_{j=1}^f \tilde{\lambda}_j + \alpha \right)^{-\alpha - 1} \prod_{j=1}^f \frac{\tilde{\lambda}_j^{\tilde{y}_j}}{\tilde{y}_j !}}{\lambda_i \Gamma(\alpha) \left( \sum_{j=1}^f \tilde{\lambda}_j + \alpha \right)^{-\alpha - 1}, \tilde{y}_j = 1, 2, \ldots; \tilde{y}_j = 0, 1, \ldots} (31)
\]

While we have estimated the MPG model, the results were clearly dominated in our application by the MPLN specification in terms of a likelihood dominance criterion and the Akaike information criterion. In the empirical section below, we focus our attention exclusively on the MPLN specification, though the results from the MPG model are available from the authors upon request.

3. Extending the model to a multi-site setting

The above model is developed for the single-site situation, controlling for the direct impact on observed trips and the indirect effects on the follow-up contingent behavior questions. However, analysts are frequently interested as well in modeling the substitution patterns among visits to a series of available sites. Without contingent behavior follow-up questions, the translation of the model in the previous section to the multi-site, on-site sampling setting is trivial. If the intercepts occur at a single site, then
the model in the previous section applies directly, with the subscripts $j$ now denoting the various sites and site $j=1$ denoting the site of intercept. If the intercepts occur at a series of sites, then the correction must take into account the site at which the intercept occurs. For example, the MPLN distribution associated for individuals intercepted at site $k$ becomes:\(^{10}\)

$$g_k(\tilde{y}_i | x_i) = \frac{\tilde{y}_k}{\delta_k} \prod_{j=1}^{J} \frac{\exp\left(-\tilde{\lambda}_j\right) \tilde{y}_j^\lambda \exp\left[-\frac{1}{2} e^T \Omega^{-1} e\right]}{(2\pi)^{j/2} |\Omega|^{1/2}} \, d\epsilon_i, \quad \tilde{y}_k = 1, 2, \ldots; \quad \tilde{y}_j = 0, 1, \ldots, \quad j \neq k$$

There are two issues that should be noted, however, in using the above model for the multi-site setting. First, with multiple sites, the conditional means in equation (14) become a system of demand equations and the issue arises as to what restrictions should be placed on the price and income coefficients appearing in this system. A number of authors (e.g., Shonkwiler, 1999) suggest treating these expected demands as an incomplete demand system and imposing the associated integrability restrictions (see, e.g., LaFrance and Hanemann, 1984). There is, however, an issue as to whether such restrictions, which apply at the individual level, should be imposed on the average (or aggregate) demand system.

The second issue is how to conduct the on-site sampling itself; e.g., whether to survey just at a single site or split the sample and conduct on-site surveying at each of the sites in the choice set. While surveying at a single site is likely to be less costly, it seems ill-advised to use such an approach. Inferring trips to $J-1$ sites from the trip pattern of individuals who frequent a single site seems to be asking a lot of the structural model. We believe that a multi-site on-site survey analysis should, as in the case of Moeltner and Shonkwiler (2005), include intercepts at each of the sites.
Finally, extending the model to include both multiple sites and CB questions is similarly straightforward. In this case, the index $J$ should be equal to the number of sites times the number of CB scenarios included in the survey. It is important to note, however, that such an application requires obtaining data on trips, not only to the site being changed under the hypothetical scenario, but also on trips to the unchanged sites so as to model the substitutions that occur in response to the price or quality changes.

**III. Data and Model Specification**

The data used in our empirical application are drawn from an intercept survey of visitors to Clear Lake located in north central Iowa. Visitors’ names and addresses were collected on-site in the summer of 2000. These individuals were then mailed a survey in October, 2000. The survey asked respondents to provide four trip totals:

- **Observed Behavior (OB):** Their total number of trips to Clear Lake between November 1999 and October 2000.

- **Contingent Behavior (CB$_0$):** Their anticipated number of trips in 2001, given current travel costs.

- **Contingent Behavior (CB$_1$):** Their anticipated number of trips in 2001, given an increase in the total cost per trip of $B$. Specifically, individuals were asked: "Suppose that the price of visiting Clear Lake increases by $B$ per trip (due for example to gas prices, user fees, or equipment costs). How many times would you visit next year?" The value of $B$ was randomly assigned to each survey respondent and varied across individuals in the sample from $3$ to $15$, with a mean of $7.28$. 

• **Contingent Behavior (CB₂):** Their anticipated number of trips in 2001, given a price increase of $C per trip, where $C>B$. Again, the value of $C$ was randomly assigned to each survey respondent and varied across individuals in the sample from $7$ to $30$, with a mean of $16.91$.

In addition to gathering trip data, the survey also asked a series of contingent valuation questions, inquired as to the respondents’ attitudes towards water quality improvements, and gathered socio-demographic information.

Of the 1,024 individuals intercepted at Clear Lake, 626 (or 62.7% of the deliverable surveys) returned a completed mail survey. In the analysis below, individuals were excluded from the final sample if they reported seasonal trips in excess of 52, allowing one trip per weekend. This resulted in 39 individuals being excluded from the sample. We also excluded households whose travel time was greater than five hours one way. Clear Lake is a unique natural lake in Iowa and does draw travelers from around the state. However, it is a regional attraction and the assumption is that anyone traveling from farther than five hours likely made the journey primarily for reasons other than to visit the lake. This excluded 19 additional households. Finally, for simplicity, a balanced panel was obtained by excluding visitors who did not answer all of the trip questions. The final sample size used in the analysis was $N=546$.

In the models estimated below, the average number of trips under scenario $j$ ($\lambda_j$) is assumed to be a function of the travel cost to Clear Lake, household income, and socio-demographic characteristics of the household. Specifically,

$$\lambda_j = \exp(\beta_{0j} + \beta_{ pj} P_j + \beta_{ qj} I_q + \delta_j z_j),$$  

(33)
where \( P_{ij} \) denotes the roundtrip travel costs from household \( i \)'s home to Clear Lake, \( I_i \) denotes household \( i \)'s annual income, and \( z_i \) is a vector of socio-demographic characteristics of the household, including:

- \( Male =1 \) if the survey respondent is male, =0 otherwise;
- \( Age = \) the age of the survey respondent;
- \( Age^2 \);
- \( School = 1 \) if the survey respondent has attended or completed some level of post-high school education; and
- \( Household = \) the total number of household members.

For observed trips (OB) and forecasted trips for 2001 (CB\(_0\)), travel costs were computed as $0.25 times the round-trip travel distance, computed using PCMiler, plus one third the respondent’s wage rate times their round-trip travel time. \( P_{ij} \) for CB\(_1\) and CB\(_2\) are computed in the same fashion, except that $B and $C are added to the travel costs, respectively.

Table 1 provides a summary of the data used in the analysis.\(^{11}\) There are a number of attributes of the raw trip data that are worth noting. First, for all four trip variables the unconditional mean number of trips in the sample is roughly the same order of magnitude as the corresponding unconditional standard deviation, indicating that the unconditional variance is eight to twelve times the unconditional mean. This suggests that overdispersion is likely to be a problem for all four trip variables in our sample and that a simple Poisson model for each trip variable will be inappropriate.\(^{12}\) Second, the observed number of trips (OB) is large, with households in the sample averaging over a dozen trips per year. This should not, however, be interpreted as indicative of the population as a
whole, but rather a reflection of the on-site sampling process. Households who frequent Clear Lake are more likely to be included in the sample precisely because they were more likely to be there when the intercepts occurred, hence inflating the sample average number of trips relative to the population’s average. Third, the observed trips (OB) are slightly higher (12.25) than the number of trips anticipated by the survey respondents for 2001 (11.65), suggesting relatively stable demand for visits to Clear Lake between 2000 and 2001. Fourth, and finally, the anticipated number of trips for 2001 decrease, as expected, with the total cost per trip, from an average number of trips just under 12 per year under current conditions (CB\(_0\)) to approximately 7.3 trips per year given an average cost increase of $17 per trip (CB\(_2\)). Thus, households appear to be responding to the hypothetical price increase at least in the direction expected.

Turning to the socio-demographic data, we find that the percentage of males (63%), average household income, and level of education are higher in the sample than in the Iowa population as a whole. This, in part, is also a consequence of the on-site nature of the survey process, as frequent recreationists are more likely to be included in the sample and these, in turn, are more likely to be males with a higher level of income and education.

In estimating the MPLN model using the Clear Lake data, several restrictions were imposed on the form of the \( \lambda_y \)'s (i.e. expected trips). First, we assume that the \( \beta \)'s in equation (33) are the same across the three contingent behavior trips, with trips changing only due to changes in the corresponding price levels. Second, we assume that the socio-demographic factors (other than income) impact the expected number of trips in
the same way for both observed trips and the three contingent trips. The resulting functional forms for the \( \lambda_{ij} \)'s are given by:

\[
\begin{align*}
\lambda_{ij} &= \begin{cases} 
\exp(\beta_{0,OB} + \beta_{P,OB} P_{i1} + \beta_{I,OB} I_{i1} + \delta' z_i) & j = 1 \\
\exp(\beta_{0,CB} + \beta_{P,CB} P_{j1} + \beta_{I,CB} I_{j1} + \delta' z_i) & j = 2, 3, 4.
\end{cases}
\end{align*}
\]

(34)

Finally, we also impose a restriction on the structure of the variance-covariance matrix for the MPLN model. Specifically, we assume that \( \Omega \) in equation (16) is given by

\[
\Omega = \begin{bmatrix}
\sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} \\
\sigma_{12} & \sigma_{22} & \sigma_{23} & \sigma_{24} \\
\sigma_{13} & \sigma_{23} & \sigma_{33} & \sigma_{34} \\
\sigma_{14} & \sigma_{24} & \sigma_{34} & \sigma_{44}
\end{bmatrix}
\]

\[
= \begin{bmatrix}
\sigma_{11}^2 & \rho_{OC}\sigma_{12} & \rho_{OC}\sigma_{13} & \rho_{OC}\sigma_{14} \\
\rho_{OC}\sigma_{12} & \sigma_{22}^2 & \rho_{OC}\sigma_{23} & \rho_{OC}\sigma_{24} \\
\rho_{OC}\sigma_{13} & \rho_{OC}\sigma_{23} & \sigma_{33}^2 & \rho_{OC}\sigma_{34} \\
\rho_{OC}\sigma_{14} & \rho_{OC}\sigma_{24} & \rho_{OC}\sigma_{34} & \sigma_{44}^2
\end{bmatrix}.
\]

(35)

This implies that the unobserved error component for the three contingent trips (CB0, CB1, and CB2) have the same correlations with each other (though different variances) and with the observed trip data.

IV. Results

Table 2 provides the estimates of the MPLN model. We present estimates both with and without the correction for on-site sampling. Several patterns emerge in the results. First, the price and income coefficients have the expected signs and are statistically significant at a one percent level for both observed and contingent behavior trips. All else equal, an increase in travel cost decreases the expected number of trips, whereas trips increase with income. Second, these coefficients (i.e., the \( \beta \)'s) do not differ substantially between the observed and contingent trips. However, the price
responsiveness is lower among the contingent trips than for the observed trips. Third, the price and income coefficients do not change substantially with the correction for on-site sampling, though they are generally smaller in size.

Turning to the socio-demographic characteristics, the results are less consistent across the corrected and uncorrected models. For the MPLN specification corrected for on-site sampling, all of the socio-demographic characteristics are statistically significant (at a 5% level or better) and have the expected signs. Men are found to take significantly more recreational trips to Clear Lake than women and the relationship between age and trips is quadratic, with the young and old taking more trips than middle aged individuals. Having attended college increases recreational trips, while a larger household size decreases trips. For the uncorrected specifications, the socio-demographic coefficients are generally less significant.

Finally, turning to the parameters associated with the mixing distribution (i.e., the elements of $\Omega$) we note that the correlation among the trips is high, with both $\rho_{oc}$ and $\rho_{cc}$ estimated to be positive and close to one. All of the estimated $\sigma_j$’s are significantly different from zero, indicating overdispersion in the data.

The parameter estimates in Table 2 can be used to illustrate implications of the models in terms of trip behavior and the implied welfare gains associated with each trip. Table 3a provides estimates of the consumer surplus per trip calculated as $CS_j = \beta_{r,j}^{-1}$ for both observed trips ($j=1$) and contingent behavior trips for 2001 ($j=2, 3, 4$) using the parameters estimated in Table 2. Both the corrected and uncorrected models predict roughly the same consumer surplus per trip, ranging from $\$56$ to $\$80$. Correcting for the
on-site sampling leads to a somewhat larger surplus measure, with an increase of 10% for actual trips and an increase of 43% for predicted trips.\textsuperscript{16}

The big impact, however, from correcting for on-site sampling comes in the form of the predicted number of trips. Table 3b provides estimates of the population average trips. For the MPLN model this corresponds to $\delta_y$ in equation (18). As expected, there is a substantial difference between the average numbers of trips when the model is corrected for on-site sampling versus when it is not. Without this correction, average trips range from 15.8 to 15.1. These are somewhat larger than the sample averages reported in Table 1. However, correcting for the on-site sampling, we see a substantial drop in the estimated average number of trips in the population. For the MPLN model the average is reduced by more than two-thirds to under five trips per household. The estimates in Table 3b are based upon the household characteristics (i.e., age, income, education, etc.) found in the survey sample. However, these too are biased by the on-site sampling process. Table 3c recalculates the estimated average number of trips using observations on household characteristics for a random sample of Iowa residents, drawn as part of the 2001 Iowa Lakes Survey. The average number of trips per household drops further as a result to approximately one trip per household.

Finally, there are a number of restricted versions of the model of interest. The first constrains the parameters of the observed and contingent behavior trip functions to be the same; i.e., $\beta_{k,o} = \beta_{k,c}$, $k = 0, P, I$. The results are reported in column three of Table 4. In general, the resulting parameters are a compromise between the observed and contingent behavior parameters, but the restriction itself is clearly rejected using a likelihood ratio test with a p-value of less than 0.001.
The second restricted version of the model replaces the multivariate lognormal mixing distribution with a single lognormal variable (i.e., $\varepsilon_{ij} = \varepsilon_i \sim N\left(0, \sigma^2\right) \forall j$).

Essentially, we are restricting $\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4$ and $\rho_{oc} = \rho_{cc} = 1$. This mimics the structure of the MPG distribution, but uses a lognormal mixing distribution rather than a gamma one. While this model represents a boundary restriction on correlation parameters, making a standard likelihood ratio test problematic, the large reduction in the log-likelihood function suggests little support for this alternative specification.

V. Conclusions

On-site samples are frequently used in recreation demand analysis to insure that users of the site in question are represented in the sample. It has long been recognized that this results in a sample that is both truncated and endogenously stratified with respect to the respondents’ reported trips to the site. The correction procedures that have been previously developed focused on observed trip data alone (e.g., Shaw, 1988, and Englin and Shonkwiler, 1995). However, researchers are frequently incorporating contingent behavior questions into their recreation demand surveys as well, asking households to indicate their future trip plans and how their trips might change given price or quality changes to the site in question (See, e.g., Rosenberger and Loomis, 1999; Azevedo, Herriges, and Kling, 2003; and Grijalva, et al. 2002). While the contingent behavior trip responses are not directly truncated or endogenously stratified, they are impacted indirectly through their correlation with observed trips. The contingent behavior data, like its observed counterpart, will not be representative of the population as a whole. In this paper, we have presented an extension of Shaw’s (1988) correction to a multivariate setting using the MPLN model.
The empirical analysis, using data from an intercept survey at Clear Lake in northcentral Iowa, indicates that the failure to correct for on-site sampling procedures results in substantial bias in the estimated average number of trips to the site, both observed and contingent, overstating population trip levels by a factor of 14. The impact on the estimated consumer surplus per trip is somewhat small. We also reject the hypothesis that the observed and contingent trips follow exactly the same demand structure, but the differences, while statistically significant, appear to be minor.
Finally, it should be noted that the procedures proposed here and elsewhere in the recreation demand literature (e.g., Shaw, 1988, Englin and Shonkwiler, 1995; and Moeltner and Shonkwiler, 2005) rely heavily on the assumption that non-users are implicitly much like their user counterparts, simply coming from a different portion of the same underlying distribution. The procedures are essentially extrapolating back to non-users from users, while at the same time correcting for the over-sampling of frequent users. The solution is analogous to a Tobit type correction in the case of censored data. In the Tobit situation, one is assuming that the negative valued (censored) observations are just like the positive valued observations – they simply have an error term that is large enough to cause their latent variable to be negative. A similar approach is being used here. If non-users are fundamentally different from users, the correction will be inappropriate and only data on non-users will allow one to correct for this. Again, in the Tobit setting, this is analogous to the argument that a double-hurdle type model is needed.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>OB trips ($y_{12}$)</td>
<td>12.25</td>
<td>11.86</td>
<td>1</td>
<td>52</td>
</tr>
<tr>
<td>CB$<em>0$ trips ($y</em>{12}$)</td>
<td>11.65</td>
<td>11.76</td>
<td>0</td>
<td>50</td>
</tr>
<tr>
<td>CB$<em>1$ trips ($y</em>{13}$)</td>
<td>10.24</td>
<td>10.85</td>
<td>0</td>
<td>50</td>
</tr>
<tr>
<td>CB$<em>2$ trips ($y</em>{14}$)</td>
<td>7.34</td>
<td>9.12</td>
<td>0</td>
<td>50</td>
</tr>
<tr>
<td>Travel Cost ( $P_{12}$)</td>
<td>$55.28</td>
<td>$54.46</td>
<td>$5.40</td>
<td>$442.10</td>
</tr>
<tr>
<td>Travel Cost + $B$ ( $P_{13}$)</td>
<td>$62.56</td>
<td>$55.45</td>
<td>$8.40</td>
<td>$452.10</td>
</tr>
<tr>
<td>Travel Cost + $C$ ( $P_{14}$)</td>
<td>$72.18</td>
<td>$56.87</td>
<td>$12.40</td>
<td>$467.10</td>
</tr>
<tr>
<td>Household Income ( $I_i$)</td>
<td>$60,178</td>
<td>$38,064.46</td>
<td>$7,500</td>
<td>$200,000</td>
</tr>
<tr>
<td>Male</td>
<td>0.63</td>
<td>0.48</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Age</td>
<td>43.65</td>
<td>13.57</td>
<td>15</td>
<td>82</td>
</tr>
<tr>
<td>Education</td>
<td>0.74</td>
<td>0.44</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Number of Household Members</td>
<td>3.08</td>
<td>1.40</td>
<td>1</td>
<td>9</td>
</tr>
</tbody>
</table>
Table 2. Multivariate Poisson LogNormal Models
(Standard Errors in Parentheses)\(^a\)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Corrected for On-Site Sampling</th>
<th>Not Corrected for On-Site Sampling</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\beta_{0,OB})</td>
<td>0.57** (0.09)</td>
<td>2.05** (0.04)</td>
</tr>
<tr>
<td>(\beta_{0,CB})</td>
<td>0.46** (0.08)</td>
<td>1.68** (0.04)</td>
</tr>
<tr>
<td>(\beta_{P,OB})</td>
<td>-1.58** (0.07)</td>
<td>-1.73** (0.07)</td>
</tr>
<tr>
<td>(\beta_{P,CB})</td>
<td>-1.26** (0.05)</td>
<td>-1.80** (0.06)</td>
</tr>
<tr>
<td>(\beta_{1,OB})</td>
<td>0.96** (0.08)</td>
<td>1.09** (0.11)</td>
</tr>
<tr>
<td>(\beta_{1,CB})</td>
<td>0.95** (0.06)</td>
<td>1.32** (0.10)</td>
</tr>
<tr>
<td>Male</td>
<td>10.73** (4.06)</td>
<td>8.64 (5.54)</td>
</tr>
<tr>
<td>Age</td>
<td>-2.77** (0.69)</td>
<td>-2.09* (0.93)</td>
</tr>
<tr>
<td>Age(^2)</td>
<td>0.024** (0.007)</td>
<td>0.013 (0.011)</td>
</tr>
<tr>
<td>School</td>
<td>12.42** (4.06)</td>
<td>-3.43 (6.61)</td>
</tr>
<tr>
<td>Household</td>
<td>-2.96* (1.42)</td>
<td>-5.71** (1.76)</td>
</tr>
<tr>
<td>(\sigma_1)</td>
<td>1.23** (0.04)</td>
<td>0.97** (0.03)</td>
</tr>
<tr>
<td>(\sigma_2)</td>
<td>1.27** (0.04)</td>
<td>1.12** (0.03)</td>
</tr>
<tr>
<td>(\sigma_3)</td>
<td>1.25** (0.04)</td>
<td>1.15** (0.03)</td>
</tr>
<tr>
<td>(\sigma_4)</td>
<td>1.16** (0.04)</td>
<td>1.15** (0.03)</td>
</tr>
<tr>
<td>(\rho_{OC})</td>
<td>0.96** (0.005)</td>
<td>0.93** (0.006)</td>
</tr>
<tr>
<td>(\rho_{CC})</td>
<td>0.99** (0.002)</td>
<td>0.98** (0.004)</td>
</tr>
<tr>
<td>LogLik</td>
<td>-6,135.4</td>
<td>-6,133.6</td>
</tr>
</tbody>
</table>

*Significant at 5% level; **significant at 1% level.
\(^a\)All of the parameters are scaled by 100, except the constants (which are unscaled), and the income coefficient (which is scaled by 100,000).
Table 3. Fitted Trips and Consumer Surplus Measures

<table>
<thead>
<tr>
<th></th>
<th>Corrected for On-Site Sampling</th>
<th>Not Corrected for On-Site Sampling</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MPLN</td>
<td>MPLN</td>
</tr>
<tr>
<td>a. Consumer Surplus Per Trip</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$CS_1$</td>
<td>63.5</td>
<td>57.8</td>
</tr>
<tr>
<td></td>
<td>(58.1, 69.3)</td>
<td>(53.1, 62.9)</td>
</tr>
<tr>
<td>$CS_2$</td>
<td>79.6</td>
<td>55.8</td>
</tr>
<tr>
<td></td>
<td>(73.6, 87.2)</td>
<td>(52.4, 60.0)</td>
</tr>
<tr>
<td>b. Fitted Population Trips</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$E[y_{i1}</td>
<td>x_{i1}]$</td>
<td>4.60</td>
</tr>
<tr>
<td></td>
<td>(3.75, 5.56)</td>
<td>(14.16, 17.88)</td>
</tr>
<tr>
<td>$E[y_{i2}</td>
<td>x_{i2}]$</td>
<td>4.56</td>
</tr>
<tr>
<td></td>
<td>(3.76, 5.53)</td>
<td>(13.39, 17.14)</td>
</tr>
<tr>
<td>c. Fitted Population Trips (corrected for population characteristics)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$E[y_{i1}</td>
<td>x_{i1}^p]$</td>
<td>0.88</td>
</tr>
<tr>
<td></td>
<td>(0.42, 1.68)</td>
<td></td>
</tr>
<tr>
<td>$E[y_{i2}</td>
<td>x_{i2}^p]$</td>
<td>1.05</td>
</tr>
<tr>
<td></td>
<td>(0.50, 1.99)</td>
<td></td>
</tr>
</tbody>
</table>

*95% confidence bounds are provided in parentheses*
Table 4. Hypothesis Tests Using Multivariate Poisson-Lognormal Model
(Standard Errors in Parentheses)\(^a\)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Unrestricted</th>
<th>Consistency (\beta_{k,0} = \beta_{k,C}, k = 0, P, I)</th>
<th>Restricted Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\beta_{0,OB})</td>
<td>0.57**</td>
<td>0.48**</td>
<td>0.39*</td>
</tr>
<tr>
<td></td>
<td>(0.09)</td>
<td>(0.08)</td>
<td>(0.12)</td>
</tr>
<tr>
<td>(\beta_{0,CB})</td>
<td>0.46**</td>
<td>-1.35**</td>
<td>-1.82**</td>
</tr>
<tr>
<td></td>
<td>(0.08)</td>
<td>(0.07)</td>
<td>(0.08)</td>
</tr>
<tr>
<td>(\beta_{P,OB})</td>
<td>-1.58**</td>
<td>-1.58**</td>
<td>1.35**</td>
</tr>
<tr>
<td></td>
<td>(0.07)</td>
<td>(0.05)</td>
<td>(0.06)</td>
</tr>
<tr>
<td>(\beta_{P,CB})</td>
<td>-1.26**</td>
<td>0.95**</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td>(0.08)</td>
<td>(0.07)</td>
</tr>
<tr>
<td>(\beta_{1,OB})</td>
<td>0.96**</td>
<td>0.95**</td>
<td>1.39**</td>
</tr>
<tr>
<td></td>
<td>(0.06)</td>
<td>(0.08)</td>
<td>(0.07)</td>
</tr>
<tr>
<td>(\beta_{1,CB})</td>
<td>0.95**</td>
<td>0.95**</td>
<td>1.39**</td>
</tr>
<tr>
<td></td>
<td>(0.06)</td>
<td>(0.07)</td>
<td>(0.07)</td>
</tr>
<tr>
<td>Male</td>
<td>10.73**</td>
<td>8.44</td>
<td>32.07</td>
</tr>
<tr>
<td></td>
<td>(4.06)</td>
<td>(4.46)</td>
<td>(7.69)</td>
</tr>
<tr>
<td>Age</td>
<td>-2.77**</td>
<td>-2.82**</td>
<td>-3.81**</td>
</tr>
<tr>
<td></td>
<td>(0.69)</td>
<td>(0.71)</td>
<td>(1.18)</td>
</tr>
<tr>
<td>(Age^2)</td>
<td>0.024**</td>
<td>0.025**</td>
<td>0.033**</td>
</tr>
<tr>
<td></td>
<td>(0.007)</td>
<td>(0.007)</td>
<td>(0.013)</td>
</tr>
<tr>
<td>School</td>
<td>12.42**</td>
<td>13.25**</td>
<td>10.65</td>
</tr>
<tr>
<td></td>
<td>(4.06)</td>
<td>(4.05)</td>
<td>(8.00)</td>
</tr>
<tr>
<td>Household</td>
<td>-2.96*</td>
<td>-3.34*</td>
<td>2.27</td>
</tr>
<tr>
<td></td>
<td>(1.42)</td>
<td>(1.49)</td>
<td>(2.97)</td>
</tr>
<tr>
<td>(\sigma_1)</td>
<td>1.23**</td>
<td>1.22**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
<td>(0.03)</td>
<td></td>
</tr>
<tr>
<td>(\sigma_2)</td>
<td>1.27**</td>
<td>1.26**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
<td>(0.03)</td>
<td></td>
</tr>
<tr>
<td>(\sigma_3)</td>
<td>1.25**</td>
<td>1.24**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
<td>(0.03)</td>
<td></td>
</tr>
<tr>
<td>(\sigma_4)</td>
<td>1.16**</td>
<td>1.16**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
<td>(0.04)</td>
<td></td>
</tr>
<tr>
<td>(\rho_{OC})</td>
<td>0.96**</td>
<td>0.96**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.005)</td>
<td></td>
</tr>
<tr>
<td>(\rho_{CC})</td>
<td>0.99**</td>
<td>0.99**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.003)</td>
<td></td>
</tr>
<tr>
<td>LogLik</td>
<td>-6,135.4</td>
<td>-6,159.5</td>
<td>-6,278.3</td>
</tr>
</tbody>
</table>

\(\chi^2_{df=3} = 48.2^{**}\) \(\chi^2_{df=5} = 285.8^{**}\)

\(^*\text{Significant at 5\% level; \**significant at 1\% level.}\)

\(^a\text{All of the parameters are scaled by 100, except the constants (which are unscaled), and the income coefficient (which is scaled by 100,000).}\)
VI. References


VII. Footnotes

1 The literature has already shown a need for this research as evidenced by Englin et al. (2001), who acknowledge their inability to estimate population values since their panel data was collected on-site.

2 Indeed, in a parallel line of research, Moeltner and Shonkwiler (2005) recently developed a correction for on-site sampling in the context of a multi-site survey without contingent behavior follow-up. Their approach is based on the Dirichlet-multinomial (DM) extension of the standard RUM multinomial logit model. An advantage of the MPLN model developed here is the general nature of the variance-covariance structure.

3 As Moeltner and Shonkwiler (2005) note, however, the problem had been studied earlier in the general statistics literature in context of weighted distributions and sized-biased sampling distributions. In the case of on-site sampling, the “size” is the frequency of trips to a site. Patil and Rao (1978) study this problem in the context of univariate distributions, while Jain and Nanda (1995) extend this to the multivariate setting.

4 As Shaw (1988) notes, a number of authors recognized earlier the truncation issue associated with on-site surveys, including Smith and Desvousges (1985). The issue of truncation in recreation demand was further discussed by Creel and Loomis (1990) and Grogger and Carson (1991).

5 Shaw (1988) actually provides two solutions to the on-site sampling, one based on the Poisson regression model and a second based on a continuous regression model of trip data. We focus our attention here on the count data model, though the corrections could be adapted for the continuous setting.

6 Note that, while our focus is on J denoting trips under different scenarios (i.e., actual trips and contingent behavior trips) for a single site, the multivariate count data models discussed in this section apply as well to modeling trips to multiple sites, with J denoting the number of sites.

7 The MPLN model can be viewed as incorporating random individual effects. An alternative approach would be to allow for individual fixed effects. Hausman, Hall and Griliches (1984) develop a fixed effects model in the context of patents and R&D expenditures. Englin and Cameron (1996) apply their model in the recreation demand context.

8 The MPG specification was introduced by Arbous and Kerrich (1951) in a bivariate context and subsequently extended by Bates and Neyman (1952) and Nelson (1985). In the economics literature, Hausman, Hall and Griliches (1984) use the MPG model as a random effects model to capture correlation between patents and R&D expenditures.

9 See Winkelmann (2000, p. 196).

10 While the focus of the Clear Lake survey was on the single site and follow-up CB questions, the survey did collect limited data of visits to other sites. In an appendix, available from the authors upon request, we analyzed the visits to the four sites included in the survey using the MPLN model in equation (32).

11 As noted by one of the reviewers, it is often helpful, when comparing the results of a study to other settings, to know the types of activities engaged in at the site being studied. In the Clear Lake survey, individuals were asked to indicate the percentage of time spent in a variety of activities. The top six activities received the following percentages: recreational boating (27.1%), swimming/beach use (16.3%), fishing (15.2%), nature appreciation (12.5%), picnicking (12.4%) and camping (12.4%).

12 Overdispersion in the count data regression model corresponds to the variance of the count variable, conditional on the explanatory variables, being larger than the corresponding conditional mean. As Cameron and Trivedi (1998, p. 77) note, this is not the same as the sample (i.e., unconditional) variance exceeding the sample mean. They do suggest, however, that “…if the sample variance is more than twice the sample mean, then data are likely to remain overdispersed after inclusion of regressors.”

13 A more general specification allowing the demographic effects to differ between observed trips and contingent trips was estimated, but the differences between the OB and CB parameters were not statistically different as a group based on a likelihood ratio test.

14 The MPLN model was estimated using maximum simulated likelihood following Munkin and Trivedi (1999). Hess, Train and Polak (forthcoming) develop a new simulation technique using a Modified Latin Hypercube Sampling Method. We employ this technique using 1000 draws in the simulation. The authors would like to thank Kenneth Train for suggesting this method of simulation and also thank Stephane Hess for providing the gauss code and suggestions for implementation.

15 A 95% confidence band is calculated for each point estimate in Table 3 using the Krinky-Robb approach.
The fact that the consumer surplus per trip does not change much is, of course, the result of the fact that the price coefficients themselves do not change much. This suggests that the marginal impact of price is not substantially different across individuals with different levels of recreational intensities. The total consumer surplus associated with the site, however, is substantially different since the total number of trips is different.