Ocean & Coastal Management 71 (2013) 170–175

Contents lists available at SciVerse ScienceDirect

Ocean & Coastal Management

journal homepage: www.elsevier.com/locate/ocecoaman

Estimating outdoor recreation demand with aggregate data: A revealed preference approach☆

Akbar Marvasti

NOAA, National Marine Fisheries Science Center, 75 Virginia Beach Dr., Miami, FL 33149, USA

ARTICLE INFO

Article history:
Available online 13 October 2012

JEL classifications:
Q26
C25

ABSTRACT

An aggregate travel cost model is employed to estimate the parameters of a single site recreation demand model for visiting a beach. Applying zonal data, the impacts of several variables which are likely to influence an individual's decision to take a trip to the beach are examined. The count nature of the dependent variable, distributional information (overdispersion), and heterogeneity of visitors' decisions are considered in employing three approaches for estimating the parameters of the demand function. The results suggest that the events, weather condition, season and weekend effect are among the most significant factors affecting the demand for the beach.

Published by Elsevier Ltd.

1. Introduction

This study uses aggregate data from the actual behavior of the public visiting a beach park in Galveston Island, Texas. The objective is to examine the effect of several factors such as travel cost, weather, and weekends on the demand for the beach. Recreation demand studies commonly use survey instrument and maximum likelihood techniques in the analysis of individual's choice among a set of alternatives by estimating the probability of visiting a site. For example, Provencher and Bishop (1997) use a fully dynamic structural model of the decision to visit a recreation site and apply their model to two fishing clubs on the shore of the Lake Michigan. Their estimated maximum likelihood model includes several explanatory variables such as expected fish catch and monetary cost of the trip as well as weather conditions such as wind speed and temperature. Interestingly, time cost measured by distance traveled to the site turns out to be an insignificant predictor of taking the trips.

Sampling problems often affect the results of recreation demand studies based on the survey instrument and their consideration in developing the estimation method improves the accuracy of the results. On-site interviews are associated with truncation and endogenous stratification problems (Shaw, 1988). For example, the omission of the non-users is a common problem arising from typical studies of recreation demand based on on-site interviews. Another problem in some recreation studies based on survey instruments is the dependency of the likelihood of being sampled on the individual's frequency of visiting the site. Englin and Shonkwiler's (1995) estimation of demand for hiking in the Pacific Northwest Cascade mountains uses negative binomial models to deal with the omission of the non-users and the endogenous stratification problems. To estimate the outdoor recreation demand for Iowa wetlands, von Haefen and Phaneuf (2003) compare continuous and count data (both negative binomial and Poisson) models. They show that continuous and count data models generate qualitatively similar policy inferences, though the estimated parameters differ, in spite of different behavioral and econometric assumptions made in the two methods.

Aggregate data avoids the sampling problems in survey data. Aggregate models typically use total number of visits by visitors to a recreation site along with explanatory variables such as per capita income and average cost of the trip. Using aggregate data has its own unique problems, however. The potential compositional heterogeneity in such data and linearity restrictions on the models has led to the concerns for the "problem of aggregation over individuals". Stoker (1993) argues that aggregate models require a process through which individual demands can be "recovered". According to Stoker, one approach to capture the underlying individual consumer behavior is the "representative consumer" model.

Several recent studies of recreation demand use aggregate count data for estimation. For example, Hellerstein (1991) estimates travel-cost recreation demand models with different functional forms using aggregate permit data for Bounty Waters Canoe Area. Fleischer and Tsur (2003), on the other hand, estimate aggregate

☆ The opinions expressed herein are those of author and not necessarily reflect the views of NOAA.

E-mail address: Akbar.Marvasti@noaa.gov.

0964-5691/$ – see front matter Published by Elsevier Ltd.
http://dx.doi.org/10.1016/j.ocecoaman.2012.09.006

recreation demand for multiple open-spaces sites in Israel – including beaches, urban parks, and national parks. They focus on government policies on land allocation and management of open spaces under increasing pressure from population growth and economic development. Their maximum likelihood estimate of the trip allocation model, based on data from a telephone survey, concludes that the public values the beaches more than the parks.

To reduce aggregation bias caused by assuming constant observations across zonal areas when averages or totals over subgroups of visitors are used, some recreation demand studies use zonal data to capture heterogeneity in visitors' characteristics. Moeltner incorporates zonal data based on zip codes to identify geographical origin of visitors to find their characteristics such as income, age group, and education level (Moeltner, 2003). Hellerstein (1995) employs Monte-Carlo simulations to compare bias from aggregation in zonal data with bias from narrowly specified individual-observation models. Focusing on the simple Poisson demand model, his results lead him to the surprising conclusion that, in fact, aggregate models can outperform individual-observation models. This is especially true when distribution information on the explanatory variables is used to control for errors-in-variables problems endemic to aggregate models.

Another issue related to estimates of recreation demand is the complexity of the decisions made by individuals. Haab (2003), for example, has questioned traditional models of recreation demand, which assume independence of recreation trip decisions. He argues that the assumption of independence of recreation trips leads to bias in parameter estimates and measurement of welfare. Then, he develops three models of recreation demand. In his first model, the pure chance, he maintains the assumption of independence of recreation trip decisions. In his second model, the true contagion, he accounts for interdependence of trip decisions in a fully dynamic model consistent with the model used by Provencher and Bishop (1997). In his last model, the apparent contagion, he assumes a constant probability for trips across choice occasions, while unobserved differences, or heterogeneities, in probability across individuals sampled remains. Using data from a telephone survey of eastern North Carolina households, Haab tests his hypothesis of interdependence of recreation trips to Albemarle and Pamlico Sounds. His statistical results strongly reject both the traditional independence assumption model as well as the true contagion model, but they accept the intermediate model, which relaxes the assumption of temporal independence.

Unlike many beach recreational demand studies, which are based on survey instruments or deal with hypothetical decisions made by visitors, the analysis here is based on actual recreation decisions of the entire population of visitors to a public beach. The remainder of this paper is organized as follows. First, a simple theoretical framework for aggregate recreation demand is outlined and an empirical model is presented. Then, the data and the method of applying an unique zonal information are described. Next, aggregate demand for all individuals visiting a beach site is estimated with conditional demand and least squares techniques as alternative approaches. After presentation of the results and discussions of the findings, concluding remarks are made.

2. Methods

2.1. Model

In developing an aggregate travel cost model here, we begin with the specification of an individual behavior because aggregate behavior is the product of individual decisions. Consider individual i from jth visitor market in time t taking x_{ijt} number of trips to the beach which require travel cost ofP_{ijt}. The individual's demand is also influenced by two sets of shift vectors. One vector, A_{ijt}, is made of individual specific variables such as travel cost to other sites, advertisements, and income. The other vector, M_t, is made of a set of shift factors which are common across individuals such as weather condition, seasons, days of the week, and events. Thus, the individual demand for time period t can be specified as

$$x_{it} = f_i\,(P_{it},\, A_{it},\, M_t) \tag{1}$$

The aggregate demand is the sum of the population or

$$X_t = \sum_{i=1}^{N} f_i\,(P_{it},\, A_{it},\, M_t),\; i = 1,2,\ldots,N \tag{2}$$

the empirical model to be estimated uses the weighted mean values of travel cost vector as $\overline{P_t} = \sum_{j=1}^{J} w_j P_{ijt}$ and individual specific variable vector of $\overline{A}_t = \sum_{j=1}^{J} w_j A_{ijt}$, where w_j is the weight assigned to the jth visitor market. Thus, the broad aggregate model to be estimated is

$$X_t = \left(\overline{P}_t, \overline{A}_t, M_t\right) \tag{3}$$

however, the empirical specification of the beach aggregate demand function to be estimated here is defined based on the available data as follows:

$$\begin{aligned} X_t = {} & g_0 + g_1\,\mathrm{TVC1}_t + g_2\,\mathrm{TVC2}_t + g_3\,\mathrm{CL}_t + g_4\,\mathrm{RA}_t + g_5\,\mathrm{EV}_t \\ & + g_6\,\mathrm{FR}_t + g_7\,\mathrm{SA}_t + g_8\,\mathrm{SU}_t + g_9\,\mathrm{TEM}_t + g_{10}\,\mathrm{SE}_t + g_{11}\,\mathrm{INC}_t \\ & + g_{12}\,\mathrm{POP}_t + g_{13}\,\mathrm{ADV}_t + g_{14}\mathrm{Year02} + g_{15}\,\mathrm{Year03} + \varepsilon_t \end{aligned} \tag{4}$$

here, TVC1 is the real travel cost to Galveston, TVC2 is the real travel cost to a substitute beach location, CL is cloudy days, RA is rainy days, EV is event days, FR is Fridays, SA is Saturdays, SU is Sundays, TEM is the average daily temperature, SE is the season, INC is real per capita income, POP is the population, ADV is real advertising expenditures, Year02 and Year03 are fixed effect dummy variables for 2002 and 2003, respectively, and ε is the error term. The expected signs of the coefficients in Equation (4) are intuitive and are presented in column 2 of Table 2.

2.2. Data

Data used in the empirical analysis here is from the Apffel Park in Galveston, Texas. Therefore, the estimates are for a single site. The study covers the 2001–2003 period for the days the park was open. While the park is open approximately seven months of the year – between early March till the end of September – it considers the period of late May till late August as the season. A few times during the study period, storms and subsequent repairs have caused the park to close. The park was open 538 days during this period when 247,615 cars entered the park. The park has a capacity limit of approximately 6,700 parking spaces. However, only during one popular event day in the past did the capacity actually become a constraint. At the other extreme, zero value has also been observed only for one of the days that the park has been open. In a screened sample, this observation is treated as outliers and dropped from the data set. One observation is dropped because the capacity has constrained (been binding) the entry to the park only in one event day and the truncated regression method is likely to create real numerical problems. Thus, the number of observations in the screened sample is reduced to 537. Park officials believe that each vehicle has on the average three occupants. The dependent variable is then structured by multiplying the number of cars