Site Congestion in Recreation Choice Models:
A Generated Regressors Approach to Beach Site Selection

PAUL HINDSLEY[1] , CRAIG E. LANDRY[2,*], OKMYUNG BIN[2], and HANS VOGELSONG[3,**]

*[1] Program in Coastal Resources Management, East Carolina University, Greenville, NC, USA; [2] Department of Economics, Brewster Building, East Carolina University, Greenville, NC 27858; [3] Department of Recreation and Leisure Studies, East Carolina University, Greenville, NC, USA; *Author for correspondence (e-mail: landryc@ecu.edu)*

July 2007

[**] Special thanks to Christopher Ellis for providing the data. Thanks are due to seminar participants at the Southern Economic Association's 2006 Annual Meeting and LSU's Center for Natural Resource Economics and Policy's 2nd National Forum on Socioeconomic Research in Coastal Systems.

Site Congestion in Recreation Choice Models:
A Generated Regressors Approach to Beach Site Selection

Abstract. Site congestion has received limited attention in revealed preference studies. Difficulties of incorporating congestion in site choice models include simultaneity bias and complications associated with obtaining data on expectations. Using site choice data from an onsite sample of North Carolina beach goers, we attempt to address these issues by generating exogenous proxies for anticipated congestion with the help of instrumental variables. We find statistical support for including congestion in our random utility models, and a concave (downwards) relationship between congestion and utility. Results suggest that, in accord with theory, failure to account for congestion can lead to bias in RUM parameter estimates.

Key words Random Utility Model, travel cost, congestion, generated regressor, beach access

JEL classification*:* C35, C81, D12, H41, Q26

Site Congestion in Recreation Choice Models:
A Generated Regressors Approach to Beach Site Selection

1. Introduction

Recreation demand models have a rich history in natural resource economics as a tool for exploring consumer preferences for recreation and informing regulatory policy and management decisions. The basic random utility model (RUM) makes use of a utility maximization framework with a particular specification that provides a simple, closed-form econometric solution and an insightful characterization of individual preferences for attributes of recreation sites. Site congestion is an attribute that has received limited attention in the empirical revealed preference literature. As an attribute of the recreation site, congestion presents formidable challenges for empirical modeling. Simultaneity is inherent in the relationship between site choice and congestion—congested sites will appear popular because many individuals choose to visit them, but they are congested simply because they are popular. Endogeneity of congestion will lead to upward bias in the RUM congestion coefficient and bias in the coefficients of covariates that are correlated with congestion. In addition, there are measurement issues associated with site congestion; *ex post* congestion is typically observed by researchers, but *ex ante* anticipations of congestion likely influence behavior.

In this article, we explore the role of congestion in a RUM of beach site selection. We review how previous research has examined the issue of congestion in recreation site choice models and offer a simple approach for addressing site congestion with intercept data and small choice sets. Our data pertain to barrier island sites on the North Carolina coast. As the data are gathered onsite, our estimation strategy must adjust for endogenous stratification. In addition, congestion is only observed at the chosen site. We use weather conditions to generate exogenous approximations of congestion levels at all sites in the choice set and invoke the assumption of rational expectations to utilize our congestion forecasts as approximations of anticipated congestion—the amount of congestion that individuals envision at the various sites when deciding which site to visit. Standard errors of the RUM are adjusted to reflect the sampling distribution of the first-stage generated regressor equation. Our results suggest that, in accord with theory, failure to account for congestion in estimation can lead to bias in RUM parameter estimates.

This paper proceeds as follows. Section 2 reviews the literature on congestion in recreation demand models. Section 3 describes the endogeneity problem as related to congestion and some of the existing approaches for dealing with endogeneity in RUMs. Section 4 presents the data, and section 5 describes the chosen method. Section 6 reports our empirical results, and section 7 concludes.

2. Recreation Demand and Site Congestion

Congestion at recreation sites induces interdependency among individual preferences analogous to the non-additivities in demand explored by Leibenstein (1950).[1] Physical space at any recreation site must be limited; congestion sets in when some individuals perceive other users of the site in such a way that it influences their own utility of consumption attributable to visitation at the site. Under such conditions, individual demand for recreation is not additive in the conventional sense, as demand for individual

[1] The "bandwagon" effect—in which utility from consumption is an increasing function of others' consumption—and the "snob" effect—in which utility from consumption is a decreasing function of others' consumption—are both examples of non-additivity in market demand.

i at any price can depend on the aggregate demand of the other $j \neq i$ users of the site (Fisher and Krutilla 1972). Congestion may also be expressed as affecting individual i's cost of visiting the site (Jakus and Shaw 1997), in which case the price of visit is a function of aggregate demand.

Early studies on congestion in recreation demand focused on efficiency of market outcomes and optimal management strategies (Fisher and Krutilla 1972; Cicchetti and Smith 1973; Anderson and Bonsor 1974; McConnell and Duff 1976; Freeman and Haveman 1977; McConnell 1988). Generally, excessive congestion (or "crowded" conditions) is characterized as a negative attribute in site choice models, though aversion to congestion may be heterogeneous. Findings indicate that consumer surplus measures will be biased if congestion effects are present and not accounted for; the bias will be upward if congestion is a negative attribute. This is further complicated by the premise that some individuals prefer more congested areas to those with a lower density of users (Eroglu and Harrel 1986**;** Anderson, Kerstetter, and Graefe 1998).

From an empirical perspective, congestion presents formidable challenges. If congestion affects site choice and it is missing from estimation, misspecification can occur in the form of omitted variable bias—the researcher can expect bias in parameter estimates for all covariates that are correlated with the omitted congestion variable. Since congestion is likely to be driven by the relative desirability of site attributes, the analyst can expect some degree of correlation between congestion levels and site attributes, the implication being a high potential for bias in parameter estimates that are relevant for policy analysis. Including congestion levels in empirical analysis is not straightforward, however. In site choice models, a congestion measure is also likely to cause simultaneity bias because the congestion measure, which is treated as an exogenous variable in the model, is a function of the choices of other individuals—site congestion will be endogenous because it is determined simultaneously with site choice. Thus, we have potential for bias even if congestion is included in the model.

Empirical recreation demand analyses of congestion have primarily made use of stated preference methods (Cicchetti and Smith 1973; McConnell 1977; Walsh et al. 1983; Boxall, Rollins, and Englin 2003). As recognized by Timmons and Murdock (2007), stated preference models avoid simultaneity bias by hypothetically varying the degree of congestion independent of other parameters. While arguably capable of providing insights into recreation behavior, this approach may suffer from any number of biases that can plague stated preference studies. Only a handful of researchers have attempted to address congestion in the context of revealed preference models (Schuhmann and Schwabe 2004; Boxall, Hauer, and Adamowicz 2005; Timmins and Murdock 2007).

Jakus and Shaw (1997) review some of the difficulties in addressing congestion in revealed preference studies. They characterize four types of congestion: actual, expected, anticipated, and perceived. Actual and perceived congestion are *ex post* measures, actual congestion being the congestion level measured by an outside observer (i.e. someone not in the sample), and perceived congestion an individual's subsequent assessment of congestion at a site. Expected and anticipated congestion are *ex ante* measures, expected congestion being the mean of a distribution, and anticipated congestion being an individual's own estimate of the congestion they would experience if they decide to visit a site. The Random Utility Model presumes that site choice stems from a rational assessment of site attributes, such that anticipated congestion is most likely to be the relevant measure for RUM estimation. Obtaining information on individuals' *ex ante* assessments, however, can require extraordinary effort on the part of the researcher.

Recent revealed preference studies liken the generation of congestion to a Nash bargaining model, where individuals make recreational site choices based on their expectations of the location that other individuals may choose (Boxall, Hauer, and Adamowicz 2005; Timmins and Murdock 2007). This concept was originally illustrated by Cesario (1980), who described a resorting mechanism in which individuals shift their use from more congested sites to less congested sites conditional on site attributes. In equilibrium, individuals' expectations of congestion are confirmed by the actual behavior of others. Thus, theory provides a mechanism by which actual congestion may provide a reasonable approximation of anticipated congestion—in equilibrium the two measures should be equal.

Schuhmann and Schwabe (2004) use various measures of actual and expected congestion as a proxy for anticipated congestion in recreational fishing in North Carolina. Their actual congestion measure is that observed by surveyors, while expected congestion is a measure of central tendency from previous time periods. They find a stable concave (downwards) relationship between utility and congestion, with "optimal" congestion varying with the measure of congestion employed and by user type. Thus, empirical evidence suggests that moderate levels of congestion may enhance utility associated with a site. Schuhmann and Schwabe do not consider the possible endogeneity of actual congestion.

Boxall, Hauer, and Adamowicz (2005) collect *ex post* assessments of anticipated congestion for wilderness visitors in Canada. They develop a model of congestion forecasting that posits the formation of anticipated congestion as the result of a bounded yet rational assessment of information on site characteristics, individual characteristics, and previous congestion levels of a subset of the universe of sites. They identify this formulation as more realistic than assuming perfect forecasting across all sites. They estimate competing site choice models using the *ex post* assessment of anticipated congestion levels (or its modal value where missing at the individual level) and using a forecasted congestion measure that serves as an instrument. They find that the forecasted congestion model performs better in terms precision and plausibility of parameter estimates. Their results indicate the mean effect of congestion is indeed negative (i.e. a "bad"), but users exhibit heterogeneous preferences for congestion with 11% exhibiting a positive effect (i.e. a "good").

Timmins and Murdock (2007) use an endogenous sorting model to account for site congestion. This method works well in light of the difficulties associated with finding suitable instruments. An adequate instrument must correlate with congestion, but exhibit redundancy in the site choice model. Timmins and Murdock invoke equilibrium conditions with regard to expected share of individuals (i.e. congestion) across sites and use a contraction mapping to identify fixed site effects, which are then decomposed into covariate effects via a smoothed GMM quantile estimator. One of the advantages of their approach is that the use of instrumental variables is relegated to the linear part of estimation, in which the properties of this approach are well understood. Their results confirm that failure to account for congestion is likely to lead to omitted variable bias. In their study, they find that a failure to account for congestion leads to an understatement of welfare attributable to loss of a site by roughly ½.

We propose a generated regressors approach for estimating the impacts of congestion in a model of recreational beach choice in North Carolina. Our method's strengths lie in its ability to deal with a small sample of sites and low cost of data collection. The approach of Timmins and Murdock (2007) requires a large number of sites be sampled in order for there to be sufficient data for the second stage regressions; our approach works with a choice-based sample with information on congestion at only the chosen site and with a small number of sites in the choice set.

We control for endogeneity of congestion measures with an instrumental variables approach assuming *ex post* observed congestion is a valid measure of *ex ante* anticipated congestion. Our generated regressors approach addresses the lack of data on congestion across all sites for all choice occasions. Bias from intercept sampling, otherwise known as choice based sampling, is corrected using Weighted Exogenous Sample Maximum Likelihood methods (Manski and Lerman 1977).[2] We explore specifications of a site-duration choice model within a RUM framework both with and without measures of congestion.

3. Endogeneity in RUM

Endogeneity occurs when observed explanatory variables are correlated with error terms, resulting in misspecification when relying on estimation procedures such as RUMs, which assume independent errors (Louviere et al. 2005). This can be represented

[2] Other researchers have addressed the potential for bias due to choice based sampling by allocating sampling proportions according to population proportions in their sampling design (Schuhmann and Schwabe 2004; Boxall, Hauer, and Adamowicz 2005).

mathematically in RUMs, as follows: the utility that individual *n* obtains from choosing site *i* can be decomposed into a deterministic portion of utility, represented by $V(\cdot)$, and a random component of utility, represented by e_{ni}. The utility function is:

$$U_{ni} = V(tc_{ni}, x_i, s_n) + e_{ni} \tag{1}$$

where tc_{ni} represents travel and time costs for individual *n* to site *i*, x_i represents observed attributes of site *i*, and s_n represents characteristics of individual *n*. Endogeneity occurs in this context when the random component of utility (e_{ni}) is correlated with observed site characteristics or travel costs. In discrete choice models, this correlation is assumed to be zero or constant across alternatives (Train 2003). Violation of this assumption results in biased parameter estimates. Endogeneity can be depicted with the decomposition:

$$e_{ni} = \xi_i + \varepsilon_{ni} \tag{2}$$

where ξ_i represents error specific to site *i* and ε_{ni} represents individual-specific idiosyncratic taste for sites which is assumed to be independent across individuals and sites, as well as identically distributed. Timmins and Murdock (2007) interpret ξ_i as reflecting unobservable site characteristics that induce a mechanical correlation between the level of congestion and e_{ni}.

When a congestion measure is included as a site attribute, the utility associated with site selection can be depicted as:

$$U_{ni} = V(tc_{ni}, x_i, c_{ni}, s_n) + e_{ni}, \tag{3}$$

where e_{ni} follows (2), and c_{ni} represents the anticipated congestion (in the sense of Jakus and Shaw). In equilibrium, anticipations of congestion are confirmed at the site level and we have:

$$c_i = c(m, \text{Prob } m \text{ chooses } i\text{: } V(tc_{mi}, x_i, c_i(\text{Prob } m \text{ chooses } i), s_m) + e_{mi} > V(tc_{mj}, x_j, c_j(\text{Prob } m \text{ chooses } j), s_m) + e_{mj}; \xi_i) \quad (4)$$

For $m \neq n$ ($m = 1, \ldots, n-1, n+1, \ldots, M$) individuals visiting site i. Actual congestion is simultaneously determined by all users' site choice processes, which within the probabilistic structure of RUM reflects the probability of other m individuals choosing site i. In this formulation, any observed congestion measure is likely correlated with ξ_i, which represents unobserved site characteristics that influence sorting across sites. To correctly estimate the impact of congestion on site choice, it is necessary to break this correlation. In our model, we assume that congestion on and between sites exists in an equilibrium state, thus allowing us to use a fitted value of congestion estimated using observed congestion (c_i) as a proxy for anticipated congestion (c_{ni}).

In addition to the traditional instrumental variable (IV) method, a number of potential approaches are available for dealing with endogeneity in choice models (Louviere et al. 2005).[3] In this paper, we generate exogenous approximations of site congestion (a variant of the more traditional instrumental variables approach) to break the correlation between congestion and site attributes. This approach is not ideal, as the properties of IV estimation are not clear in nonlinear models due to the fact that proof of consistency of IV relies on the expectation operator (which is a linear operator). Nonetheless, the traditional IV approach fits our data limitations—a choice-based sample

[3] For example, the BLP method (Berry, Levinsohn, and Pakes 1995), the control function approach (Villas-Boas and Winer 1999; Blundell and Powell 2001; Petrin and Train 2006), and the dual approach (Matzkin 2004) are potential candidates.

with information on congestion at only the chosen site and with a small number of sites in the choice set.[4] While we cannot prove consistency of our estimates, our results exhibit a pattern consistent with successfully controlling of endogeneity. In any event, our approach is arguably better than doing nothing to address site congestion (as in many previous studies on recreation site choice).

4. Data

Our study utilizes visitation information from seven beach sites in North Carolina, collected between July 2, 2003 and November 2, 2003. Data were gathered onsite at Cape Lookout National Seashore, Hatteras Island, Fort Macon State Park, Pea Island National Wildlife Refuge, the Rachel Carson National Estuarine Research Reserve, Topsail Island, and Wrightsville Beach. Figure 1 exhibits the spatial distribution of sites in coastal North Carolina. These sites were chosen in an effort to adequately represent the wide variety of beach recreation sites found on North Carolina's barrier islands.

The data were collected onsite at different times of the day and on different days of the week, approximately ten days per month. Each beach was surveyed at least once every third week on alternating days of the week, so as to acquire the most representative sample possible. Due to the onsite sampling procedure, corrections for endogenous stratification were necessary to achieve consistent results (Manski and Lerman 1977).

Table 1 provides summary statistics for individuals in our sample and for the seven beach sites. Distance to each site is calculated using the individual hometown and

[4] With a larger number of sites in the choice set, the BLP approach can be utilized (as shown by Timmons and Murdock (2007)). With information on congestion levels across all sites on all choice occasions, the control function approach could be explored. The dual approach can be employed if exogenous variation in the endogenous variable can be found (Train and Winston forthcoming).

beach zip codes. Travel costs are measured as the sum of pecuniary and time costs. Pecuniary travel costs reflect fuel and vehicle wear-and-tear (approximated as \$0.485×[round trip distance to each site]). Opportunity costs of the travel time are estimated as a fraction of annual income (approximated as $\frac{1}{3}$ × hourly wage).[5] Average round-trip travel cost is around \$450. The average group size is 4.65 people and two-thirds of those intercepted were making a multiple-day trip to the beach.

Given the limited number of sites in our data, we must be judicious in our choice of site characteristics to be included. We include log-transformed shoreline length in our model in order to capture the scale of each site, and we express some of our site attributes in "per unit length" terms in order to reflect this standardization. Failure to address differences in site area could lead to aggregation bias, since several of our sites had the potential to be dissected into smaller sites. Haener et al. (2004) found that including variables which account for the size of aggregated sites allows model parameters to be equivalent across scales.[6] Average shoreline length is about 19.5 miles.

Because we are primarily interested in the relationship between site choice and measures of beach access, our RUM model incorporates a number of site characteristics which are likely to influence access. The average site has 2.6 access points with parking every mile and about 0.08 access points with off-road-vehicle (ORV) access every mile. Only Cape Lookout National Seashore and Hatteras Island allow ORV access. In addition to access variables we include the number of residential housing units as an indicator of the potential for rental accommodations, as well as a rough indicator of the

[5] We addressed missing income responses with an hourly wage rate of \$5.15, the minimum wage in N.C. As a result, this measure is a conservative estimate of the true value.

[6] Some sites may have the potential to be affected by aggregation bias. The variables indicating shorelength and the number of access points with parking should address this bias.

level of development on site. The average site has 2,026 residential housing units, and 24% of the sites in our data set can only be accessed via ferry. The small choice set limits our ability to model other amenities that may drive site choice because of correlation with included access variables. We, however, explore the potential for omitted variable bias in our analysis.

Actual (observed) congestion was recorded by a surveyor at the time surveys were administered. Before interviewing individuals, the surveyor counted all people that were visible at the site from a central vantage point using a handheld clicker counter. The surveyor then administered surveys to individuals in that particular area. At times, the surveyor walked some distance from the access point to administer the survey. If traveling sufficiently far, the surveyor would make an additional person count to reflect varying congestion along the shore. Average congestion was 55 people, with a standard deviation of 75, a minimum of 1, and maximum of 440.

5. Methods

We propose that it is possible to employ a generated regressors approach to address missing congestion data, while at the same time using instruments to purge endogeneity of congestion in our site choice model. We assume that recreational users are constrained utility maximizers who choose the site $i \in I$ which gives the highest level of satisfaction via equation (3). We assume beachgoers exhibit rational expectations in considering the congestion levels at potential beach sites. Actual site congestion is simultaneously determined by each recreational user's optimal choice via equation (4).

In order for our generated congestion forecasts to function as suitable instruments, we must impose: $E(z'\varepsilon) = 0$, where z is a vector of instrumental variables (including the exogenous variables in (3)) and ε is the idiosyncratic error from equations (2) and (3); and our instruments should be partially correlated with endogenous congestion once the other exogenous variables are controlled for (Wooldridge 2001). While the latter condition can be tested in a first-stage linear IV regression, the former condition, in general, cannot be tested as ε is unobservable. Moreover, in nonlinear models the expectation $E(z'\varepsilon) = 0$ does not necessarily apply because the expectations operator is a linear operator. Nonetheless, as we show below, our results appear consistent with this restriction. To identify the congestion effect within an instrumental variables framework, we require instruments that explain congestion at the site, but not choice between sites. Suitable instruments include exogenous demand shocks (i.e. occurrences that alter demand for recreation over choice occasions).

Our first stage equation produces a proxy for anticipated congestion as a generated regressor. This approach is necessary because congestion is observed at chosen sites, but not at alternative sites in the choice set. The generated regressor addresses simultaneity bias by the inclusion of instruments which explain beach congestion, but not site choice. Our semi-log functional form imposes non-negativity on the congestion measure in the first stage. The second stage of estimation is a discrete site choice model in which fitted congestion at each site is included as a covariate. In equilibrium, anticipated and experienced levels of congestion will be equal. Our first stage forecast of *ex post* actual congestion serves as a proxy for *ex ante* anticipated congestion.

Our generated regressor approach is inspired by empirical methods which utilize observed variables to proxy for unobserved site quality attributes. The method has similarities with techniques used to calculate expected fish catch, commonly applied within the recreation fishing demand literature. McConnell, Strand, and Blake-Hedges (1995) model recreational anglers' expected catch using a count process. Expected catch is assumed to be a function of some combination of individual characteristics (skill set, experience, etc.), trip characteristics (trip length, time of year, etc.), and site characteristics. Fitted measures allow observed catch conditional on these characteristics to act as a proxy for expected catch. In a very similar way, our congestion measure utilizes observed congestion to model unobserved anticipated congestion.[7]

We can depict the specific form of the mean level of anticipated congestion as

$$\ln Q_{ih} = \alpha_1 x_i + \alpha_2 z_{ih} + \upsilon_{ih}, \tag{5}$$

where Q_{ih} is the number of individuals observed at site i on day h, x_i represents characteristics of site i, z_i is a vector of instrumental variables for site i on day h, and υ_{ih} is the error term for site i on day h. We use a linear projection of anticipated congestion, rather than a count model as is common in dealing with expected catch, in order to maintain orthogonality between $\hat{Q}_{ih}$ and υ_{ih}. One of the most difficult tasks is finding viable instruments for our model; the instruments must be factors that help explain congestion, but not choices between sites. We employ weather variables specific to the

[7] One major difference between expected catch and expected congestion lies in the importance of individual characteristics in the anticipation of quality. Fishing success, while partly determined by external factors, is heavily influenced by the skill of the fisherman. In the case of beach recreation, the characteristics of a particular user arguably have little influence on resulting site congestion. In equilibrium individuals sort across sites such that utility is maximized and anticipated congestion equals realized congestion (Cesario 1980; Timmins and Murdock 2007). Thus in equilibrium, individual characteristics have no impact on the level of congestions other than how preferences for congestion influence sorting behavior.

choice occasion (i.e. day). The weather variables are collected from three coastal weather stations which collectively cover all seven sites in our dataset. Each weather station covers a large geographical area, encompassing multiple beach sites. We argue that there is little variation in weather between sites on any given day, so these variables are not likely to influence choice between sites; however, factors such as temperature, precipitation and convective storms are likely to influence the level of congestion at a site on any given day.

As some individual and site attributes are unobservable to the researcher, the RUM focuses on the probability of choice. The probabilistic statements in our basic model are represented by the conditional logit model, where V_{ni} is a linear function and each random component, ε_{ni}, is assumed to be independently, identically distributed extreme value (McFadden 1974). A benefit of using the conditional logit model is the fact that it results in the following closed form expression:

$$P(n \text{ chooses beach site } i) = P(i_n) = \frac{\exp(U_{ni})}{\sum_J \exp(U_{nj})}, \tag{6}$$

which is globally concave in parameters.

In estimation it was necessary to correct for potential bias from on-site sampling (also known as choice-based sampling). On-site sampling was chosen because it was both convenient and cost effective. Numerous authors have identified the key issues related to choice-based samples (Manski and Lerman 1977; Manski and McFadden 1981; Cosslett 1981; Imbens 1992). These studies show that data collected from onsite sampling leads to parameter estimates influenced not only by the value of the site attributes, but also by the sampling intensity applied by the researcher. This means that if the sampling proportions are different than the population proportions, the parameter estimates will be inconsistent with the true values. We correct for choice based sampling using weighted exogenous stratification maximum likelihood estimation (WESMLE) as determined by Manski and Lerman (1977).[8] WESMLE is a pseudo-maximum likelihood estimation procedure in which a weight is computed as the population proportion divided by the sample proportion. Table 2 gives the sample and population proportions used in estimation of the WESMLE procedure.[9] The WESML estimator is commonly used when the distribution of site choices made by recreational users in the population is known but the marginal distribution of exogenous characteristics, specifically user characteristics, is not necessarily known. Our estimator is:

$$\max_{\beta} \sum_{n=1}^{N} w(i_n) \ln P(i_n \mid z_n, \beta) \tag{7}$$

where $w(i) = \dfrac{H(i)}{Q(i)}$ is Manski and Lerman's weight, with *H(i)* representing the population proportions and *Q(i)* representing the sample proportions of site visitation.[10]

Without corrections, our estimation procedure is inherently flawed due to the loss of information between the two stages. In general, the coefficient estimates for generated regressors will be consistent, but the standard errors and test statistics will be invalid because the inclusion of anticipated congestion in a second stage regression fails to account for the sampling variation of the fitted value (Wooldridge 2001). As a result, we

---

[8] Special thanks to David Brownstone for the Stata code for the correction of the WESMLE covariance matrix.

[9] Population proportions were determined by using existing visitation projections in area land use plans. In these equations, visitation is a function of parking capacity, hotel capacity, rental capacity, and ferry schedules.

[10] Manski and Lerman (1977) show that this estimator is consistent for β but not efficient. We use Cosslett's (1981b) weight, which is slightly different than the weight shown above, but is more efficient. Specifically, $w(i_n) = \dfrac{r(s)N}{n_s}$, where n_s represents the number of observations in each strata (i.e. site), $N = \sum n_s$ for $\forall s \in S$, and *r(s)* represents the qualification factor (determined by utilizing the population proportions for site choice).

utilize a paired bootstrap procedure to estimate the standard errors within the WESMLE procedure. The paired bootstrap allows us to approximate the effect of the first step estimation on the asymptotic variance-covariance of the second step. The process consists of random draws from the sample with replacement. Each new sample is then used to estimate the recreation site choice model. This is repeated so to develop a distribution of coefficient estimates, which are then used to calculate the standard errors.[11]

6. Results

The results for our first stage congestion regression models are presented in table 3. As requisite in IV estimation, we include site characteristics from the RUM in our first stage equation. In addition, the forecast includes a dummy variable indicating weekend or holiday visitation. Our instruments are temperature (in degrees Fahrenheit), a dummy variable indicating occurrence of a thunderstorm, and precipitation (in inches). In addressing endogenous variables via IV, it is necessary to test for the appropriateness of instruments. Our specification exhibits a joint F statistic of 21.46 for the instruments precipitation, temperature, and thunderstorms which is well above the threshold suggested by Staiger and Stock (1997).

We next test if the instrumental variables are orthogonal to the error process in the IV equation. We test our subset of instruments using the Hansen C test. The Hansen C statistic tests if a subset of a model's overidentifying restrictions appear to be satisfied through the calculation of two Sargan-Hansen statistics, where the first Sargan-Hansen

[11] For more information on the empirical applications of bootstrap procedures, see Horowitz (2001) or Brownstone and Valletta (2001).

statistic is calculated from the full model and the second statistic comes from a model without the instruments (Baum, Schaffer, and Stillman 2003). The difference between the two Sargan-Hansen statistics is distributed χ^2_{df}, with *df* = number of instruments, and is used to test the null hypothesis that instrument orthogonality conditions are satisfied. Our initial specification does not exhibit orthogonality between instruments and the error term; the p-value for the Hansen C test is less than 0.0001, indicating that the null (orthogonality of instruments) should be rejected at any conventional level of statistical significance.

We explored possible misspecification due to correlation among subgroups in our sample. Consider subgroups that vary by group size, or the size of the traveling party ("group" variable in table 1). By choosing to visit a site, groups contribute to the overall site congestion, and this effect varies with the size of the group. Larger groups induce more congestion than do smaller ones, and sites that attract large groups will tend to exhibit greater variation in congestion levels. Thus, variability in congestion could systematically vary with group size. We cluster our robust standard errors by group size in order to account for this subgroup correlation. Utilizing this specification with intra-cluster correlation, we find the Hansen C test fails to reject the null hypothesis that temperature, precipitation, and the thunderstorm dummy variable are proper instruments (p-value=0.1443). Table 3 shows the Hansen C test results for the semi-log congestion specifications with and without clustering.

Having some confidence in the validity of our instruments, we turn to the IV regression. Results are presented in the last column of table 3. The parameter estimates for the IV equation are reduced form estimates, but the coefficients on the instruments are nonetheless intuitive. We see that sites are less congested on rainy days and on days with thunderstorms. Also, as the temperature increases, the congestion forecasts are higher. As expected, congestion is considerably higher on weekends and holidays (as exhibited by the "weekend" variable).[12] The mean and range of forecasted congestion during weekend and weekdays/holidays for each beach site is depicted in table 4.

The second stage of estimation uses the fitted congestion measures as an exogenous variable in the RUM of site choice and trip duration. We compare a basic site choice model to a specification which includes the fitted congestion measure. We also include two specifications with site specific constants.[13] These specifications should give some indication of how well our expected congestion measure performs. If we adequately address the endogeneity in this congestion measure, our coefficient on expected congestion should be similar in the models with site constants and without.[14] If our expected congestion measure does not break the correlation with unobserved site attributes, the coefficients for the models with and without site specific constants should differ.

To produce correct standard errors with two stage estimation, we utilize a paired bootstrapping procedure. We take 475 random draws with replacement from our sample; within each iteration, we run our WESMLE procedure, saving the coefficient estimates. We repeat this procedure 1000 times, so to develop a distribution of parameter

[12] Hypothesis tests (t-tests) indicate significant differences between mean weekday and weekend/holiday congestion for each site at the 0.001 level.
[13] Likelihood ratio tests indicate that models with site specific constants have better fit than those without, but including site specific constants limit analysts' ability to measure policy relevant questions relevant to individual site attributes that do not vary within site.
[14] Murdock (2006) has a detailed discussion of site specific constants in recreation site choice models.

coefficients. We then calculate our standard errors from the distribution of 1000 bootstrapped coefficients.

In our second stage site choice model, individuals chose between single- and multi-day trips at the seven different sites. We approach the estimation of the RUM with an emphasis on beach access among the seven sites. Access points with parking provide an amenity in that beach goers have more options in accessing the beach and can be more confident that they will be able to find a parking spot once they arrive. ORV access points give users more ability to spread out along a beach and enable users to bring more beach gear onsite. The ferry-only dummy variable indicates sites only accessible via ferry (either public or private). Ferries can limit access due to schedule limitations and additional travel cost (both time, and in the case of private ferries, money). Importantly, these attributes will almost certainly be correlated with congestion levels, and thus one cannot hope to accurately estimate the parameters for these covariates without accounting for congestion.

We also include a variable indicating the number of residential houses onsite and a variable for the natural log of shore length in miles. Residential housing provides a rough estimate of the number of rental opportunities as well as an idea of the level of development on the site. Shore length helps account for potential aggregation bias. Results for our site choice models are displayed in table 5.

Models 1 and 2 present our access attributes specification that corresponds with the IV regression in column 3 of table 3, model 2 including a quadratic specification of anticipated congestion. Models 3 and 4 provide robustness checks of our specification, by including site-specific constants and covariates that vary within site—travel cost, the multi-day indicator, and congestion. Likelihood ratio tests indicate all models are jointly significant at p-values less than 0.001. All parameter estimates are statistically significant at p-values less than 0.05, except for "housing" in model 2, the Topsail Island and Wrightsville Beach indicators in model 3, and the Topsail Island indicator in model 4. Signage of the estimates generally conforms to expectations. Travel costs have a negative effect on utility, all else being equal. The number of access points with parking per mile, ORV access points per mile, and shoreline length each has a positive effect on utility. The dummy variable for exclusive-ferry access is negative, and the multi-day trip indicator is positive.

Similar to the findings of Schuhmann and Schwabe (2004), our model supports a quadratic congestion effect in the representative beach user's utility functions.[15] Figure 2 depicts this relationship for the model without site-specific constants (model 2) and the model with site-specific constants (model 4). The difference in optimal congestion between the two models is negligible—61 people for model 2 and 62 people for model 4. A t-test of the coefficient value for model 2 indicates that there is no statistically significant difference when compared with coefficient value for model 4. This suggests that our instruments perform well in the first stage regression.

Corresponding to the notion that failure to include congestion in the model leads to omitted variable bias, we witness interesting changes in parameters estimates across the models. When forecasted congestion is included, there is a change in the coefficients for parking, ferry-only access, and ORV access variables. This suggests a correlation of congestion with these measures. When the congestion measure is included, the

[15] A likelihood ratio test indicates that the quadratic form has better fit than the linear form ($\chi_1^2 = 17.52$).

coefficient on access points with parking subsequently increase, suggesting that failure to account for congestion leads to other attributes reflecting congestion effects. Parameter estimates for attributes which intensify congestion, such as access points with parking, will be attenuated in estimation (biased toward zero). On the other hand, parameter estimates for attributes that allow users greater ability to avoid congestion, such as ORV access and ferry only access, will be upward biased in estimation. T-tests indicate that downward bias in the parking variable is statistically significant with 99% confidence (t-value = -8.61) and upward bias in the ferry only access variable is statistically significant with 95% confidence (t-value = -1.89). We did not find a statistically significant upward bias for ORV access points (t-value = 0.72).

Our next step involves the calculation of welfare estimates for changes in amenities at these seven sites. Table 6 depicts these results. When we compare the basic site choice model (model 1) to the congestion model (model 2), we find that failure to include the congestion measure understates the welfare effect of access points with parking and overstates the welfare effect of ORV access points. In a comparison of welfare measures, the percentage difference in welfare loss for a 25% change in access points with parking (per mile) is roughly 19% greater for the specification with congestion. On the other hand, the percentage difference in welfare loss for a 25% change in ORV access points (per mile) is roughly 9% less for the specification with congestion. Examining the welfare effect of expected congestion, we find a 9% greater effect in the model without site specific constants in comparison to the model with site specific constants. As indicated by the confidence intervals, however, none of these welfare differences are statistically significant.

7. Discussion

Despite the prevalence of literature that identifies the level of congestion at recreation areas as an import site attribute, the difficulty associated with incorporating congestion in revealed preference models of recreational demand has discouraged most scholars from exploring congestion effects. The primary difficulty can be traced to the complexity of endogenous regressors in nonlinear models. We describe congestion as a measure that is simultaneously determined by multiple agents. This process produces bias in probability models of site choice because the simultaneous determination of congestion is likely caused by unobserved attributes at the sites themselves. Our solution is to apply a generated regressor approach with instrumental variables similar to that used in the development of proxies for expected catch and other unobserved *ex ante* environmental quality measures. With the application of weather variables and weekend/holiday characteristics that help explain expectation for site congestion without explaining the choice between sites, we attempt to meet the necessary orthogonality conditions for consistent estimation.

The fitted congestion measures improve the estimation of recreational demand models. We find that failure to include congestion in the model leads to downward bias for site attributes with positive correlation with congestion and upward bias for attributes with negative correlation. As the number of access point with parking spots increases, congestion generally increases. Controlling for congestion allows more accurate understanding of influence of these types of amenities on site selection. Other amenities, such as ORV access points, give beach goers more of an opportunity to spread out along the beach, thus diminishing congestion. While we do find a smaller coefficient for ORV access when we control for congestion, the difference is not statistically significant.

We find that congestion has a nonlinear effect on beach users' utility. At low levels, users appear to be positively affected by the presence of other beachgoers. As congestion levels increase, however, this affect begins to negatively impact beach users. This finding coincides with the importance of user interaction and space availability on beach experience. It is possible that some users experience some type of bandwagon affect when going to the beach. The affect of congestion may only become negative when other users begin to infringe upon each others personal space or when beach visitors' crowding norms are exceeded. This nonlinear relationship deserves more attention, but data limitations inhibit a deeper investigation.

Research extensions could investigate this bandwagon effect in an attempt to better understand optimal levels of congestion. A deeper understanding would allow managers to develop strategies to better manage congestion over time and space to maximize user satisfaction and minimize amenity degradation. Studies have shown that the character of use is as important as actual density levels in determining perceived congestion levels (Graefe, Vaske, and Kuss 1984; Shelby and Heberlein 1986). For example, a relatively small number of people who are participating in an activity that conflicts with other users (i.e. loud music, active games, etc.) may cause visitors to feel more crowded than a larger number of people who are participating in more passive recreational activities. This suggests that managers can influence perceptions of congestion through policies, rules, and zoning.

We are unable to account for preference heterogeneity for congestion among users. Previous recreational demand models have identified preference heterogeneity for congestion using stated choice methods (Michael and Reiling 1997; Boxall, Rollins, and Englin 2003), but we are only aware of two revealed preference study that illustrates this type of heterogeneity (Boxall, Hauer, and Adamowicz 2005; Timmons and Murdock 2007). We hypothesize that heterogeneous users may respond differently to congestion according to personal characteristics or their primary beach activity. For example, some sunbathers may regard beach visitation as a social event and surfers may desire an audience. Anderson, Kerstetter, and Graefe (1998) referred to this as the "functional density" of a site, rather than utilizing the negative connotations associated with the term "crowding". Other users, such as anglers or those seeking peace and quiet, are likely to have a lower tolerance for congestion. Data limitations hinder our ability to further explore preference heterogeneity. While investigation of congestion preferences via mixing distributions seems a worthwhile endeavor, the difficulties associated with applying such methods to endogenously stratified samples renders this approach beyond the scope of our study.[16] We leave further investigations of preference heterogeneity for congestion to future research.

8. Conclusions

This study addresses issues related to congestion as a site attribute in revealed preference models of site choice. Typically, congestion is ignored because the difficulty of incorporating endogenous variables in nonlinear models—one usually cannot ensure the

[16] We attempted to test for heterogeneity using interaction terms by primary beach activity. Our lack of identification may have been the result of our small choice sets.

necessary conditions for IV hold in a nonlinear setting, and it is often difficult to find adequate instruments. We examine recreational beach site choice using intercept data by employing a two stage estimation procedure. The first stage generates an exogenous proxy for anticipated congestion at all sites in the choice set by projecting observed congestion onto site characteristics and variables describing weather conditions (which serve as instruments). The second stage site choice/duration model includes anticipated congestion as a covariate and is estimated by Weighted Exogenous Stratified Maximum Likelihood Estimation (WESMLE) in order to correct for choice-based sampling.

RUMs that include congestion exhibit a better statistical fit. Results suggest that failure to include congestion measures leads to bias in parameter estimates for covariates that are correlated with congestion. This bias impacts welfare estimates, potentially under- or over-valuing site amenities. We find a quadratic relationship between congestion and utility. At low levels of congestion, utility responds positively to increases in users; as space becomes more limited, congestion appears to negatively impact utility. Our data limits our ability to investigate preference heterogeneity for congestion. We leave this investigation to future studies.

References

Anderson, M., D.L.Kerstetter, and A.R.Graefe, (1998), ‘The Effects of Festival Attributes upon Perceptions of Crowding’, in H.G. Vogelsong, comp. ed., *Proceedings of the 1997 Northeastern Recreation Research Symposium* (pp. 182-185). Bolton Landing, NY: US Forest Service.

Anderson, F.J. and N.C. Bonsor, (1974), ‘Allocation, Congestion, and the Valuation of Recreational Resources’, *Land Economics* 50, 51-57.

Baum, C., M.E. Schaffer, and S. Stillman, (2003), ‘Instrumental Variables and GMM: Estimation and Testing’, *Stata Journal* 3(1), 1-31.

Berry, S., J. Levinsohn, and A. Pakes, (1995), ‘Automobile Prices in Market Equilibrium’, *Econometrica* 63, 841-890.

Blundell, R.W. and J. Powell, (2004), ‘Endogeneity in Semiparametric Binary Response Models’, *Review of Economic Studies* 71, 655-679.

Boxall, P., K. Rollins, and J. Englin, (2003), ‘Heterogeneous Preferences for Congestion During a Wilderness Experience’, *Resource and Energy Economics* 25, 177-195.

Boxall, P., Hauber, and W. Adamowicz, (2005), ‘Modeling Congestion as a Form of Interdependence in Random Utility Models’, Staff Paper 05-01, Department of Rural Economy, University of Alberta.

Brownstone, D. and R. Valletta, (2001), ‘The Bootstrap and Multiple Imputations: Harnessing Increased Computing Power for Improved Statistical Tests’, *Journal of Economic Perspectives* 15 (4), 129-141.

Cesario, F. (1980), ‘Congestion and the Valuation of Recreation Benefits’, *Land Economics* 56 (3), 329-338.

Cicchetti, C.J. and V.K. Smith, (1973), ‘Congestion, Quality Deterioration, and Optimal Use: Wilderness Recreation in the Spanish Peaks Primitive Area’, *Social Science Research* 2, 15–30.

Cicchetti, C.J. and V.K. Smith, (1976), *The Costs of Congestion: An Econometric Analysis of Wilderness*. Cambridge, MA: Ballinger.

Cosslett, S. (1981), ‘Maximum Likelihood Estimator for Choice-based Samples’, *Econometrica* 49(5), 1289-1316.

Eroglu. S., and G. Harrell, (1986), ‘Retail crowding: Theoretical and strategic implications’, *Journal of Retailing* 62(4), 346-364.

Freeman, A. and R.H. Haveman, (1977), ‘Congestion, Quality Deterioration, and Heterogeneous Tastes’, *Journal of Public Economics* 8, 225–232.

Fisher, A. and J. Krutilla, (1972), ‘Determination of Optimal Capacity of Resource-Based Recreation Facilities’, *Natural Resources Journal* 12,: 417-444.

Graefe, A.R., J.J. Vaske, and F.R. Kuss, (1984), ‘Social Carrying Capacity: An Integration and Synthesis of Twenty Years of Research’, *Leisure Sciences* 6, 395-431.

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

Haener, M.K., P.C. Boxall, W.L. Adamowicz, and D.H. Kuhnke, (2004), ‘Aggregation Bias in Recreation Site Choice Models: Resolving the Resolution Problem’, *Land Economics* 80(4), 561-574.

Horowitz, J.L. (2001), 'The Bootstrap', in E.E. Leamer and J.J. Heckman, eds., *Handbook of Econometrics, Volume V*. Amsterdam: Elsevier Science Publishers.

Ibens, G.W. (1992), 'An Efficient Method of Moment Estimator for Discrete Choice Models with Choice Based Sampling', *Econometrica* 60, 1187-1214.

Jakus, P. and W.D. Shaw, (1997), 'Congestion at Recreation Areas: Empirical Evidence on Perceptions, Mitigating Behavior, and Management Preferences', *Journal of Environmental Management* 50, 389–401.

Leibenstein, H. (1950), 'Bandwagon, Snob, and Veblen Effects in the Theory of Consumer Demands', *Quarterly Journal of Economics* 65, 183-207.

Lewbel, A. (2004), 'Simple Estimators for Hard Problems: Endogeneity and Dependence in Binary Choice Related Models', Working Paper. Boston College.

Louviere, J., K. Train, M. Ben-Akiva, C. Bhat, D. Brownstone, T.A. Cameron, R. Carson, J.R. Deshazo, D. Fiebig, W. Green, D. Hensher, D. Waldman, (2005), 'Recent Progress on Endogeneity in Choice Modeling', *Marketing Letters* 16(3/4), 255-265.

Manski, C. and S. Lerman, (1977), 'The Estimation of Choice Probabilities from Choice Based Samples', *Econometrica* 45, 1977–1988.

Manski C.F. and D. McFadden, (1981), 'Alternative Estimators and Sample Designs for Discrete Choice Analysis', in C.R. Manski and D. McFadden, eds., *Structural Analysis of Discrete Data*. Cambridge, Massachusetts: MIT Press.

Matzkin, R. (2004), 'Unobservable Instruments', Working paper. Northwestern University.

McConnell, K. and V. Duff, (1976), 'Estimating Net Benefits of Recreation Under Conditions of Excess Demand', *Journal of Environmental Economics and Management* 2, 224-230.

McConnell, K. (1977), 'Congestion and Willingness to Pay: A Study of Beach Use', *Land Economics* 53(2), 185–195.

McConnell, K. E. (1988), 'Heterogeneous Preferences for Congestion', *Journal of Environmental Economics and Management* 15, 251–258.

McConnell, K. E., I. E. Strand and L. Blake-Hedges, (1995), 'Random Utility Models of Recreational Fishing: Catching Fish using a Poisson Process', *Marine Resource Economics* 10(3), 247–261.

McFadden, D. (1974), 'Conditional Logit Analysis of Qualitative Choice Behavior' in P. Zarembka, ed., *Frontiers of Econometrics*. Academic Press.

Michael, J. and S. Reiling, (1997). 'The Role of Expectations and Heterogeneous Preferences for Congestion in the Valuation of Recreational Benefits', *Agricultural and Resource Economics Review* October, 166-173.

Murdock, J. (2006),' Handling Unobserved Site Characteristics in Random Utility Models of Recreation Demand', *Journal of Environmental Economics and Management* 51(1), 1-25.

Petrin, A. and K. Train, (2006), 'Control Function Corrections for Omitted Attributes in Differentiated Product Markets', Working Paper, University of Chicago GSB.

Schuhmann, P.W. and K.A. Schwabe, (2004), 'An Analysis of Congestion Measures and Heterogeneous Angler Preferences in a Random Utility Model of Recreational Fishing', *Environmental and Resource Economics* 27, 429-450.

Shelby, B. and T. A.Heberlein, (1986), *Social carrying capacity in recreation settings*. Corvallis, OR: Oregon State University Press.

Staiger, D. and J.H. Stock, (1997), 'Instrumental variables regression with weak instruments', *Econometrica* 65(3), 557-586.

Timmins, C. and J. Murdoch, (2007), 'A Revealed Preference Approach to the Measurement of Congestion in Travel Cost Models', *Journal of Environmental Economics and Management* 53(2), 230-249.

Train, K. (2003), *Discrete Choice Methods with Simulation*. Cambridge, UK: Cambridge University Press.

Train, K. and C. Winston, 'Vehicle Choice Behavior and the Declining Market Share of U.S. Automakers', forthcoming *International Economic Review*.

Villas-Boas, J. and R. Winer, (1999), 'Endogeneity in Brand Choice Models', *Management Science* 45, 1324-1338.

Walsh, R.G., N.P. Miller and L.O. Gilliam, (1983), 'Congestion and Willingness to Pay for Expansion of Skiing Capacity', *Land Economics* 59(2), 195–210.

Wooldridge, J. (2001), *Econometric Analysis of Cross Section and Panel Data*. MIT Press.

Table 1: Descriptive Statistics

Individual Characteristics:

Variable	Description	Mean	Std. Dev
tc	Travel Cost (Implicit Costs + Explicit Costs)	450.8	459.28
nonwhite	Non White Respondents (Dummy)	0.15	0.36
age	Age of Respondent	41.28	12.12
hschool	Highest Level of Education is High School (Dummy)	0.34	0.47
married	Marriage (Dummy)	0.6	0.49
mday	Muliple Day Trip (Dummy)	0.67	0.47
local	Local Resident (Dummy)	0.18	0.39
kids	Kids in Group (Dummy)	0.4	0.49
group	Respondent Group Size	4.65	5.53
weekend	Weekend Observation (Dummy)	0.6	0.49
	N = 476 observations		

Site Characteristics:

Variable	Description	Mean	Std. Dev
parkp	# of Access Points with Parking per mile	2.60	3.93
orv	# of ORV Access Points per mile	0.08	0.16
ferry	Ferry-only Access (Dummy)	0.24	0.43
housing	Residential Housing Units	2025.88	2251.61
slength	Shore Length in miles	19.46	18.98
cong	Actual (Observed) Congestion	54.51	75.54
	I = 7 sites		
Instruments			
temp	Site Temperature (Degrees Fahrenheit)	83.7	4.88
precip	Precipitation in Inches	0.17	0.36
thstrm	Thunderstorms Forecast (Dummy)	0.08	0.26

Table 2: Sample and Population Proportions used in WESMLE

Island	Sample Proportion	Population Proportion
Topsail Beach	0.1497	0.2457
Wrightsville Beach	0.1559	0.2613
Hatteras Island	0.1726	0.2233
Pea Island	0.1310	0.0355
Rachel Carson	0.0936	0.0032
Cape Lookout	0.1497	0.1530
Fort Macon	0.1476	0.0777

Table 3: Instrumental Variables Congestion Equation Results

Variables	IV Test (no cluster)	IV Test (clustered)	IV Regression
tc	0.00011**	0.00011**	0.00015**
	(0.00008)	(0.00008)	(0.00010)
mday	-0.32361	-0.32609	-0.31984
	(0.10726)	(0.06217)	(0.08007)
parkp	0.18843	0.18840	0.18493
	(0.01638)	(0.02266)	(0.01431)
ferry	-0.29328	-0.29298	-0.34754
	(0.13521)	(0.13870)	(0.12707)
orv	2.16495	2.14331	1.46203
	(0.36443)	(0.18717)	(0.28457)
housing	0.00016	0.00016	0.00015
	(0.00003)	(0.00003)	(0.00003)
slength	0.25317	0.25355	0.33841
	(0.05647)	(0.03976)	(0.05042)
weekend	0.26414	0.03976*	0.30606
	(0.12469)	(0.15795)	(0.15640)
temp			0.04078
			(0.00721)
thstrm			-0.79694
			(0.24309)
precip			-0.51663
			(0.07915)
constant	1.63716	1.63775	-1.7719
	(0.17360)	(0.17873)	(0.64284)
Obs	476	476	476
R-squared	0.5583	0.5574	0.6114
Adjusted R-squared	0.5507	0.5498	0.6022
P-value for Hansen C Test	0.0000	0.1443	
Cluster	None	Group Size	Group Size

Dependent variable is natural log of observed congestion.
Robust standard errors are included in parentheses.
* Indicates coefficients that fail statistical significance at the 0.05 level.
** Indicates coefficients that fail statistical significance at the 0.1 level.

Table 4: Mean Forecasted Congestion Measures by Site

Island	Weekday Congestion	Range	Weekend - Holiday Congestion	Range
Topsail Beach	55.84	14.83 - 108.12	73.79	22.70 - 170.42
Wrightsville Beach	97.77	25.83 - 189.01	129.44	39.74 - 296.65
Hatteras Island	55.78	3.32 - 99.43	69.31	3.44 - 144.30
Pea Island	11.32	.68 - 19.95	14.08	.70 - 29.04
Rachel Carson	4.52	1.27 - 7.46	5.80	2.18 - 13.52
Cape Lookout	12.75	3.59 - 21.04	16.35	6.14 - 38.12
Fort Macon	6.78	1.91 - 11.18	8.69	3.26 - 20.26

Table 5: Random Utility Site Choice Model Results

	Model 1	**Model 2**	**Model 3**	**Model 4**
Travel Cost	-0.00668	-0.00663	-0.00677	-0.00675
	(0.00098)	(0.00095)	(0.00103)	(0.00101)
Multiple Day Trip (Dummy)	0.70097	0.56312	0.70055	0.56859
	(0.10522)	(0.11297)	(0.10524)	(0.11308)
Access Points with Parking/mile	0.13947	0.15924		
	(0.00230)	(0.02884)		
Ferry only Access (Dummy)	-0.47948	-0.38591		
	(0.04951)	(0.05233)		
ORV Access/mile	1.58271	1.45777		
	(0.17450)	(0.32059)		
Housing	0.00004	-0.00002**		
	(0.00001)	(0.00004)		
Ln(Shorelength)	0.57371	0.51809		
	(0.00741)	(0.02547)		
Expected Congestion		0.02082		0.02205
		(0.00967)		(0.00970)
Exp Congestion Squared		-0.00017		-0.00018
		(0.00005)		(0.00005)
Topsail Island (Dummy)			0.98066**	0.54095*
			(0.02628)	(0.34380)
Wrightsville Beach (Dummy)			1.07049**	1.02406
			(0.03031)	(0.43214)
Hatteras Island (Dummy)			1.50500	1.03635
			(0.06706)	(0.33444)
Pea Island (Dummy)			-0.72284	-0.82656
			(0.05246)	(0.06475)
Rachel Carson (Dummy)			-3.17189	-3.12336
			(0.01295)	(0.02571)
Cape Lookout (Dummy)			0.68249	0.56525
			(0.00153)	(0.05715)
LL	-1056.7968	-1040.5734	-1034.44	-1017.8708
LL(Null)	-1253.5522	-1253.5522	-1253.5522	-1253.5522
LR chi2(7)	393.51	425.96	438.22	471.36
Rho Squared	0.157	0.1699	0.1748	0.188

Standard Errors in parentheses

* Indicates coefficients that fail statistical significance at the 0.05 level.

** Indicates coefficients that fail statistical significance at the 0.1 level.

Table 6: Welfare Measures - Changes in Consumer Surplus

	Model 1	Model 2	Model 4
A 25% decrease in Access Points with Parking Per Mile	-$19.10 (-$25.27,-$12.92)	-$22.69 (-$33.66,-$11.73)	
A 25% decrease in ORV Access Points Per Mile	-$4.50 (-$5.75,-$3.25)	-$4.08 (-$5.93,-$2.23)	
A 10% increase in Congestion on site		-$5.40 (-$10.44,-$0.36)	-$4.90 (-$9.75,-$0.06)

* Bootstrapped 95% confidence interval in parentheses.

Figure 1
Map of the Seven Beach Areas in North Carolina

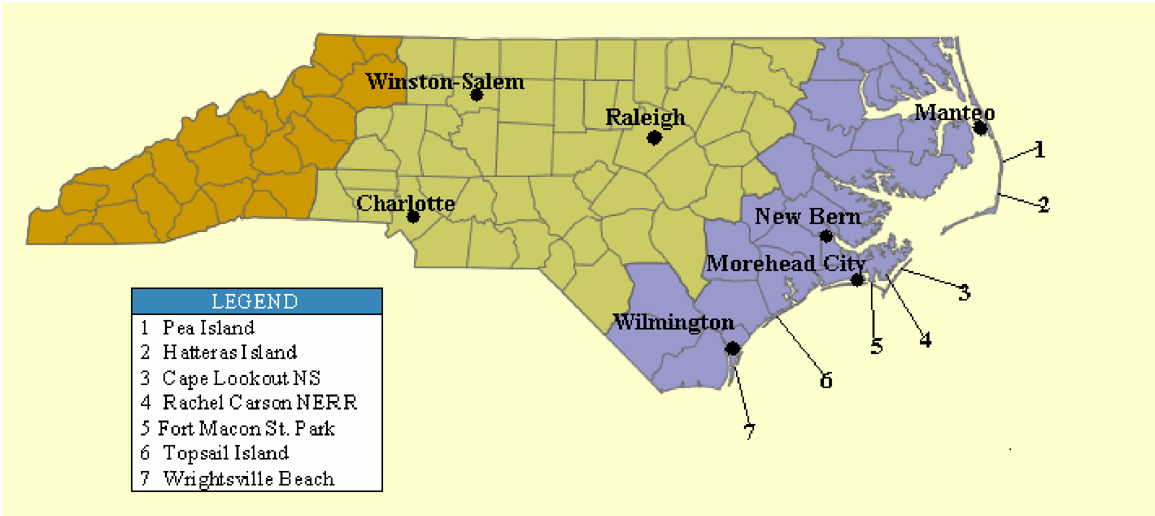

Figure 2

Graphical Relationship between Congestion and Utility