The Socio-Economic Marine Research Unit (SEMRU)
National University of Ireland, Galway

Working Paper Series

Working Paper 12-WP-SEMRU-07

Panel Travel Cost Count Data Models for On-Site Samples that Incorporate Unobserved Heterogeneity with Respect to the Impact of the Explanatory Variables

Stephen Hynes and William Greene

For More Information on the SEMRU Working Paper Series
Email: stephen.hynes@nuigalway.ie, **Web:** www.nuigalway.ie/semru/

SEMRU *Working Paper Series*

Panel Travel Cost Count Data Models for On-Site Samples that Incorporate Unobserved Heterogeneity with Respect to the Impact of the Explanatory Variables

Stephen Hynes* and William Greene

Abstract

In this paper, we examine heterogeneity in the trip preferences of recreationists by applying a random parameters negative binomial model and a latent class negative binomial model to a panel data set of beach users at a site on the west coast of Ireland. This is the first such attempt in the literature to account for heterogeneity with respect to the impact of the chosen explanatory variables in contingent behaviour travel cost models of demand where the researcher also must account for the fact that the sample data has been collected on-site. The analysis also develops individual consumer surplus estimates and finds that estimates are systematically affected by both the random parameter and latent class specifications. There is also evidence that accounting for individual heterogeneity improves the statistical fit of the models and provides a more informative description of the drivers of recreationalist trip behaviour.

Keywords: *Contingent behaviour, travel cost, count data, heterogeneity, latent class, random parameter, endogenous stratification, truncation, negative binomial, consumer surplus.*

JEL Classification: Q26, Q51

*Corresponding Author

This work was funded through the Beaufort Marine Research Award, which is carried out under the Sea Change Strategy and the Strategy for Science Technology and Innovation (2006-2013), with the support of the Marine Institute, funded under the Marine Research Sub-Programme of the National Development Plan 2007–2013.

Author Contact Details: Stephen Hynes, Socio-Economic Marine Research Unit, J.E. Cairnes School of Business and Economics, National University of Ireland, Galway, Email: stephen.hynes@nuigalway.ie; William Greene, Economics Department, Stern School of Business, New York University, Email: wgreene@stern.nyu.edu

1. Introduction

Users of a recreational site such as a beach or a forest park tend to be diverse and have different reasons for wanting to visit such sites. In the discrete choice recreational demand literature this has been a well recognized fact since Train (1998) and now the vast majority of published work involving the estimation of destination choice random utility models allow for the mixing of taste intensities in the population of interest. This recognized heterogeneity across recreational groups using a site such as a beach or forest park (and indeed even within particular recreational groups) has not been given the same treatment in count data travel cost models of recreation demand as it has in the discrete choice literature. This is especially true in on-site contingent behavior models of recreational demand. In this article we therefore analyze revealed and contingent recreational trip decision making of a group of beachgoers, comparing two empirical panel on-site count data modeling approaches that account for unobserved preference heterogeneity across individuals, namely the random parameter (RP) model and the latent class (LC) model. It is argued that these approaches can be used to better understand the factors that influence the frequency of recreational trips and in estimating the welfare impacts resulting from changes in site quality.

There have been several attempts in the literature to combine the travel cost model revealed preference method and stated preference contingent valuation approaches to non-market valuation in the form of the contingent behaviour model. This is done with the objective of measuring the welfare impact of a hypothetical change in implicit price or in environmental quality (Whitehead et al., 2008). Usually, this variation in site or environmental quality is obtained through a stated change in hypothetical visits. Examples of the use of the Contingent Behaviour TCM approach in recreational demand modelling include Grijalva et al. (2002), Hanley et al. (2003), Christie et al. (2007), Martınez-Espineira and Amoako-Tuffour (2008) and Beaumais and Appéré (2010)[1].

Preference heterogeneity is an element that has also been previously been incorporated into cross-sectional count data models. Random parameters count data models for example have previously been applied to model the frequency of accidents on roadway segments (Anastasopoulos and Mannering, 2009) and to model the demand for off-road vehicle recreation (Holmes and Englin, 2010). In terms of using latent class count models to account for heterogeneity with respect to the impact of the slope coefficients Wedel et al. (1993) presented a latent class Poisson model that accounted for heterogeneity in both the base mean event rate and the regression coefficients. Scarpa et al. (2007) also examined the existence of latent classes in the total demand for recreational days in the Eastern Italian Alps by applying finite mixing to a zero-inflated cross-sectional count demand model. Elsewhere, Baerenklau (2010) used a latent class approach to incorporate unobserved heterogeneity into an aggregate count data framework in an effort to control for endogenous spatial sorting in zonal recreation models. Panel count data models that incorporate unobserved heterogeneity, with respect to the impact of the explanatory variables, have also been previously developed (see for example Wang et al., 1998). Furthermore, statistical packages such as Nlogit (Greene, 2007) and Latent Gold Choice (Vermunt and Magidson, 2005) now contain standard commands that allow the researcher to readily incorporate a discrete mixture distribution into panel count data models.

[1] For an in-depth review of the contingent behaviour modeling literature the interested reader should see Whitehead et al. (2008b).

It should also be noted that Egan and Herriges (2006), Beaumais and Appéré (2010) and Moeltner and Shonkwiler (2010) have previously developed panel contingent behaviour models that account for the on-site sampling issues of endogenous stratification and truncation[2]. The error term added to the parameterized mean function of the Poisson models used by the aforementioned authors can be interpreted as capturing unobserved heterogeneity. These studies still however do not account for the presence of unobserved heterogeneity via the slope coefficients. Therefore, while the many studies that applied contingent behavior count-data models have added to our knowledge of factors affecting recreational trip frequencies, to date no travel cost model exists for panel data that simultaneously accounts for the on-site sampling issues of endogenous stratification and truncation and the presence of unobserved heterogeneity via slope coefficients for the explanatory variables. This gap in the literature is filled in this paper with the specifications of both RP and LC on-site contingent behavior models.

In what follows we first (section 2) present our extension of Englin and Shonkwiler's specification to a panel data negative binomial count data model that corrects for endogenous stratification and truncation and also allows for unobserved heterogeneity in the population via both latent class and a random parameter specifications. Section 3 provides a description of the recreational beach site used in the application of our models and includes a brief description of survey design and data collection procedures. Our estimation results are then presented in section 4. Finally, the paper concludes with a discussion of its major findings and their implications for recreational demand modeling.

2. Methodology

In a contingent behavior study of recreational demand, each person i in the data set yields two responses. The first is the number of trips (y_{i1}) they have made to the recreational site under current conditions in the previous 12 months (response or scenario t = 1), and the second observation is how many trips (y_{i2}) the person says they would make if a specified change in recreational opportunity at the site occurs due to some hypothetical change in site quality or facilities (response or scenario t = 2). These trip counts are limited to non-negative integers and the distribution of trips tends to be positively skewed from zero, thus preventing the use of a standard ordinary linear regression model (Cameron and Trivedi 2005). Due to the multiple trip observations per individual that occurs in a contingent behaviour analysis the researcher can also employ a panel data modelling approach.

Following the work of Shaw (1988), Grogger and Carson (1991), Englin and Shonkwiler (1995) and Greene (2008) we assume that, based on such data, a panel data count model of recreational demand can be estimated using a negative binomial distribution for the dependent count variable. As with Englin and Shonkwiler (1995) we also adjust our modeling strategy to control for the fact that our data were collected on-site. Unique in the literature we also adjust our random effects panel data negative binomial model corrected for on-site sampling to allow for the mixing of taste intensities over, firstly, continuous

[2] Endogenous stratification and truncation are two other important issues of relevance for on-site collected contingent behaviour models when the data has been collected on-site. Truncation refers to the fact that on-site data contains information on active visitors only and is therefore truncated at positive demand for trips to the site (Shaw, 1988 and Englin and Shonkwiler, 1995). Secondly, an on-site survey is subject to the problem of endogenous stratification where due to the method of data collection the likelihood of being sampled depends on the frequency with which an individual visits the site.

value distributions using a random parameters modeling framework and secondly, over a finite group of taste segments in the population using a latent class modeling framework. Our starting point for a panel of trip data, i=1,…,N individuals and t =1,…T_i responses (here, T_i = 2) for that individual, is the standard negative binomial model for count data that allows for overdispersion in the responses;

$$P(y_{it} \mid \mathbf{x}_{it}) = \frac{\Gamma(y_{it}+1/\alpha)}{\Gamma(1/\alpha)\Gamma(y_{it}+1)}\left(\frac{1/\alpha}{\lambda_{it}+1/\alpha}\right)^{1/\alpha}\left(\frac{\lambda_{it}}{\lambda_{it}+1/\alpha}\right)^{y_{it}} \qquad (1)$$

where λ_{it} = exp($\boldsymbol{\beta}'\mathbf{x}_{it}$) is the conditional mean function and $1/\alpha$ is the overdispersion parameter. (For convenience at this point, observation subscripts will be omitted.) The vector $\mathbf{x}$ represents the set of explanatory variables reported for each individual i. It is a $k\times1$ vector of observed covariates and $\boldsymbol{\beta}$ is a $k\times1$ vector of unknown slope parameters. The scalar α and the vector $\boldsymbol{\beta}$ are structural parameters to be estimated from the observed sample. Larger values of α imply greater overdispersion. The model reduces to the Poisson when $\alpha = 0$.

The density that applies to the observations obtained on site was shown by Shaw (1988) to equal:

$$P(y \mid \mathbf{x},\ on\ site) = \frac{yP(y \mid \mathbf{x})}{\sum_{s=1}^{\infty} P(s \mid \mathbf{x})}. \qquad (2)$$

For the negative binomial model in particular, the result [see Englin and Shonkwiler, 1995, p. 106, (9)] is

$$P(y \mid \mathbf{x},\ on\ site) = \frac{y\Gamma(y+1/\alpha)\alpha^{y}\lambda^{y-1}(1+\alpha y)^{-(y+1/\alpha)}}{\Gamma(1/\alpha)\Gamma(y+1)}, y = 1,2,... \qquad (3)$$

Departing from this point there are two possible ways of accounting for individual heterogeneity in the parameters in the panel negative binomial model adjusted above for endogenous stratification and truncation (equation 3) where unobserved heterogeneity in the distribution of y_{it} is assumed to impact the mean λ_{it}. These are the *random parameter* (RP) modeling approach and the *latent class* (LC) modeling approach.

The RP count model generalizes (3) by allowing the coefficients of observed variables to vary randomly over people rather than being fixed. Conditional on individual preferences for taking trips to the site the probability of observing an individual taking y trips in a given period is still generated by a negative binomial process, but the marginal probability across individuals requires integrating over a distribution of preferences which needs to be specified by the analyst. The multivariate normal and its transformations are of particular appeal in this context because of their computational tractability. The general form of the RP model for the negative binomial regression in this case is given by:

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \Gamma\mathbf{v}_{it} \qquad (4)$$

where $\boldsymbol{\beta}$ is the fixed means of the distributions for the random parameters, Γ is a lower triangular or diagonal matrix which produces the covariance matrix of the random parameters as $\boldsymbol{\Gamma\Gamma}'$ (the matrix of standard errors that scale the heterogeneity factors) and

$\mathbf{v}_{it}$ is the unobservable $K \times 1$ latent random term in the ith observation in β_i (K being the number of parameters). Although not done in this particular application the researcher could also add $\Delta\mathbf{z}_i$ to the right hand side of equation (4) where $\mathbf{z}_i$ is a set of M observed variables which do not vary over time and which enter the means and Δ is the coefficient matrix, $K \times M$, which forms the observation specific term in the mean.
The probability $P(y|\mathbf{x},\mathbf{v},\ on\ site)$ is the term that enters the log likelihood that is maximized to obtain the estimates of $\boldsymbol{\theta} = (\alpha,\boldsymbol{\beta},\boldsymbol{\Gamma})$. However, because $\mathbf{v}_{it}$ is unobserved, equation (4) cannot be used to define a likelihood function, and it is necessary to integrate this latent random term out. This is done by specifying a parametric probability density function, $g(\mathbf{v}_{it})$, that has a known distribution. Given this structure, the log-likelihood for the random parameters panel count data model is then given by:

$$\log L(\alpha,\beta,\Gamma) = \sum_{i=1}^{N} \log \int_{\mathbf{v}_i} \prod_{t=1}^{T_i} p(y_{it} \mid \mathbf{x}_{it}, \mathbf{v}_i, \alpha, \beta, \Gamma, on\ site) g(\mathbf{v}_i) d\mathbf{v}_i \qquad (5)$$

where $P(y_i|\mathbf{x}_i, \mathbf{v}_{it}, (\beta_i,\alpha)$ on site) is given in (3) with $\lambda|\mathbf{v}_{it} = \exp(\beta_i'\mathbf{x}_{it})$, β_i is given in (4) and $g(\mathbf{v}_{it})$is the mixing function. Each element of $\mathbf{v}_{it}$ has mean zero and variance one and may be distributed as normal, uniform or triangular.

Maximum likelihood estimation of the random-parameters negative binomial model as specified in equation (5) cannot easily be computed analytically due to the required integration of the function over the distribution of the random parameters. Therefore, the model is estimated using a simulation-based maximum likelihood method where the estimated parameters are those that maximize the simulated log likelihood function while allowing for heterogeneity in the parameter estimates. Halton draws are used to provide a more efficient distribution of draws for numerical integration than purely random draws (Greene, 2007; Bhat, 2003 and Train, 1998). We therefore define an unbiased simulator of $P(y_i|\mathbf{x}_i, \mathbf{v}_{it}, (\beta,\alpha)$ on site) as $P^*(y_i|\mathbf{x}_i, \mathbf{v}_{it}, (\beta,\alpha)$ on site). Then, a simulated maximum likelihood estimator for the parameter vector θ is given by:

$$\log L^S(\alpha,\beta,\Gamma) = \sum_{i=1}^{N} \log \frac{1}{R} \sum_{r=1}^{R} \prod_{t=1}^{T_i} p(y_{it} \mid \mathbf{x}_{it}, \mathbf{v}_{ir}, \alpha, \beta, \Gamma, on\ site) \qquad (6).$$

A large number R of simulated draws (r = 1, …, R) from the distribution of $\mathbf{v}_i$ replace the integral in equation (5) with a simulated average, which is approximately equivalent when the number of simulated draws is sufficiently large. Maximization of the simulated maximum likelihood function produces the vector of mean values for the random parameters $\boldsymbol{\beta}$ and the matrix Γ which once again is the lower triangular or diagonal matrix which produces the covariance matrix of the random parameters that describe the distribution of $\boldsymbol{\beta}$ across the sample of individuals.

The second modeling approach used to account for unobserved heterogeneity with respect to the impact of the explanatory variables is the latent class model. Within this modeling framework, the heterogeneity with respect to the impact of the explanatory variables is accommodated by the latent sorting of individuals into groups. The continuous distribution is approximated by estimating the location of what Greene (2008) refers to as the "support points" and the mass probability in each interval. We interpret this discrete approximation as producing a sorting of individuals into C classes, c= 1,…C. The analyst does not observe directly which class, c = 1,…., C, generated observation $y_{it} \mid c$ and class membership must be estimated. The latent class model, in generic form, conditioned on the particular class can therefore be written as:

$$P(y|\mathbf{x}, \textit{on site, class} = c) = F(y|\mathbf{x}, \beta_c, \alpha_c). \qquad (7)$$

It should be noted that there is a separate dispersion parameter in each class as well. The unconditional *prior* probabilities attached to the latent classes are given by:

$$\pi_c = \mathrm{Prob}(\textit{class} = c) = \frac{\exp(\tau_c)}{\Sigma_{q=1}^{C}\exp(\tau_q)}. \qquad (8)$$

The logit formulation for the probabilities is a convenient parameterization that allows the prior class probabilities to be constrained to the unit interval and to sum to one. The normalization $\tau_C = 0$ is imposed because only C-1 parameters are needed, with the adding up restriction, to specify the C probabilities. With this structure, there is a one to one correspondence between the set of parameters, $(\tau_1,\ldots,\tau_{C-1},0)$ and the set of class probabilities, $(\pi_1,\ldots,\pi_{C-1},1-\Sigma_{c=1}^{C=1}\pi_c)$. For an individual observation, the unconditional probability is averaged over the classes,

$$P(y\,|\,\mathbf{x},\ \textit{on site}) = \sum\nolimits_{c=1}^{C}\pi_c P(y\,|\,\mathbf{x},\ \textit{on site},\ \textit{class} = c). \qquad (9)$$

While individuals are observed more than once in the sample we make the assumption that conditional on the class membership, which does not change for the person, the trip choices are made independently. It should be noted that there is correlation induced across choices in that the observed variables, $\mathbf{x}_i$ are correlated across visits and, as well, since the class membership is fixed, the individuals preferences, embodied in β_c are also common across visits. However, we have not assumed that there are unobserved factors that are omitted from the model and which are correlated across visits. With these assumptions, the joint probability of the T_i trip choices by individual i is given by

$$P(y_{i1},\ldots,y_{iT_i}\,|\,\mathbf{x}_{i1},\ldots,\mathbf{x}_{iT_i},\beta_c,\alpha_c,\textit{on site},\textit{class} = c) = \prod\nolimits_{t=1}^{T_i}P(y_{it}\,|\,\mathbf{x}_{it},\beta_c,\alpha_c,\textit{on site},\textit{class} = c)\ (10)$$

The log likelihood for the panel of data is obtained by using the joint probability in (10) to form the log likelihood in (11);

$$\log L = \sum\nolimits_{i=1}^{N}\ \log\left\{\sum\nolimits_{c=1}^{C}\pi_c\prod\nolimits_{t=1}^{T_i}\ P(y_{it}\,|\,\mathbf{x}_{it},(\beta_c,\alpha_c),\ \textit{on site},\ \textit{class} = c)\right\} \qquad (11)$$

The log likelihood is maximized to obtain the estimate of

$$\theta = [(\beta_1,\alpha_1),(\beta_2,\alpha_2),\ldots,(\beta_C,\alpha_C),(\tau_1,\ldots,\tau_C)].$$

It should be noted that the approach adopted in this study of adjusting for truncation and endogenous stratification in both the observed and contingent observations distribution is different from that in Egan and Herriges (2006) and Beaumais and Appéré (2010) where the observed behavior data are assumed truncated to zero and endogenously stratified but the contingent behavior data are not (i.e. the on-site sampling correction is only specified through observed data in their case). Even though our second observation for each person is the hypothetical number of trips they would make under changed site conditions, we argue that the problem of endogenous stratification and truncation still holds. The respondent is still someone who has a higher likelihood of being included in the sample due to their frequency of use. Also, Moeltner and Shonkwiler (2010) showed that on-site sampling issues persist even for past season trip reports if the respondent is intercepted on-site this season. The authors labelled this effect “avidity carryover". We argue that a similar effect could apply to the hypothetical trip observations, if we interpret them as “future season trips". In that case the contingent behavior data as well as the observed behavior data should be assumed truncated.

For consumer utility maximization subject to an income constraint, and where the number of trips are a nonnegative integer, Hellerstein and Mendelsohn (1993) showed that the expected value of consumer surplus, $E(CS_{it})$ derived from count models can be calculated as $E(CS_{it}) = E(y_{it}|x_i)/\beta_{pi} = \hat{\lambda}_{it}/(\beta_{pi})$ where y_{it} is the number of trips to the beach for individual i under conditions t, and λ_{it} is the underlying rate at which the number of trips occur, such that one would expect some number of trips in a particular year, i.e. λ_{it} is the mean of the random variable y_{it}. The coefficient, $ß_{pi}$ is the individual price (*i.e.* travel cost) coefficient. The per-trip $E(CS_{it})$ is simply equal to $-1/ß_{pi}$. The change in the consumer surplus resulting from an improvement in the coastal amenities is then given by

$$\Delta E(CS_i) = \Delta E(y_{ij}|x_i)/\beta_{pi} = (\hat{\lambda}_i^* - \hat{\lambda}_i)/\beta_{pi} \tag{12}$$

where $\hat{\lambda}_i$ is the expected number of trips before any improvements are made to the coastal amenities *(t = 1)* and $\hat{\lambda}_i^*$ is the expected number of trips after improvements are made to the coastal amenities *(t = 2)*. This suggests that the change in consumer surplus for individual i can be calculated by dividing the change in the predicted number of trips to the beach site by the coefficient of the travel cost variable.

For the latent class model, the change in consumer surplus per trip is estimated by weighting the travel cost parameters from our Latent class model by the class probabilities in the NB Latent Class Model such that:

$$\Delta E(CS_i) = (\hat{\lambda}_i^* - \hat{\lambda}_i)/\sum_{c=1}^{c} \pi_c \beta_{cpi} \tag{13}$$

For the random parameters model, estimation of the change in consumer surplus per trip is approximated by simulation from draws of the estimated distributions for the travel cost random parameter.

It should also be noted that other features of the distribution of predicted trips beyond the expected value can be of interest too. For the RP model, quintiles have the advantage of being less sensitive to extreme values. In the results section we therefore report a number of percentile estimates along with the expected value from the CS distribution. It is also important to point out that the relevant comparison in welfare terms is between the number of predicted trips at the current level of coastal amenity provision at the beach site and the predicted number of trips at the improved level. Also, one cannot disaggregate benefit estimates into additional utility from those who take no extra trips to the beach and additional utility from those who visit most frequently. The beach travel cost study and the on-site collected dataset employed are described in the next section prior to the presentation of model results and welfare estimates.

3. Data and Study Background

The application of our model is to a data set generated from a survey that examined the possible welfare impact associated with the development of a coastal trail that connects two beach areas along the Galway Bay coastline in the west of Ireland. The data were generated from an on-site survey of visitors to Silverstrand beach approximately 7km outside of Galway city which is accessible by public road only. The beach was awarded a blue flag status in 2009 and is therefore required to comply with certain standards in terms of lifeguard safety and patrol as well as high water quality. The beach itself is only 300m long and has only limited facilities in the form of parking, benches, picnic tables and toilet facilities. Nevertheless it is a popular destination, particularly in the summer months for outdoor enthusiasts and is used heavily by the local urban community of Galway city and surrounding area as a recreational amenity. The beach caters to a wide range of uses including walking, swimming, sun-bathing, bird watching, kayaking and kite surfing.

Silverstrand beach was chosen as a site to investigate the issue of coastal access as a strip of privately owned agricultural land which has a cliff face at the waters edge prevents the access of recreationalists to a much larger area of beach and access along the shore to the nearby Salthill beach and promenade. If recreations could freely cross this section of agricultural land it would open up a coastal walk of over 4 miles. At present users of Silverstrand have no right to cross the private farmland to access the additional beach area. With this in mind respondents were asked a contingent behaviour question (see figure 1) in relation to how their usage of the beach facility would change if the length of beach at their disposal was increased through the opening up of a cliff walk that would give them access to an additional 1km of beach and also access along the shore to Salthill beach and promenade[3].

As part of the study, 146 personal interviews were carried out at the beach site. The questionnaire was piloted over a 2 week period in June 2009. This was followed by the main survey which took place at Silverstrand during the months of July and August 2009. Due to the non-response to certain questions in the main survey, 18 surveys were not deemed usable in the final analysis which resulted in a final sample of 128 individual responses being used for model estimation. The on-site interviews were conducted on both week days and weekends, during all daylight hours. The questionnaire solicited information on trips taken to the beach, activities undertaken, personal demographics, income, employment status, education, social relations and obligation free time. Each interview took approximately 20 minutes. Finally, attitudinal data were also collected from the respondents. The collection of both observed and contingent trip data points resulted in a panel data set of 256 observations.

[3] In particular, respondents were asked if the changes described on the card were implemented at the beach resource, would they change the number of trips they would take to the site over the next 12 months. This was followed up with an option of choosing 1. *no change in number of trips taken*, 2. *more trips* or 3. *fewer trips*. Finally the respondent was asked to state the increased (or decreased) number of trip if they had chosen option 2 or 3 (as is often the case in contingent behavior studies of this type no respondent chose option number 3).

Each respondent's travel cost was computed following the standard approach in the literature by considering the direct costs and the opportunity cost of travel. For each respondent *i* and each scenario *t*, the travel cost was calculated as:

$$TC_{it} = \left(\frac{Dist_{it} \times CostperKM}{Groupsize_i} \right) + (time\,(0.25 \times \left(\frac{Income_i}{2000} \right)))$$

where $Dist_{it}$ is the round-trip distance from the respondent's home to the site, *time* is the return travel time (in hours) from home to site, $CostperKM$ is the average petrol cost per mile (the Automobile Association of Ireland's calculation of €0.224/mile obtained from http://www.aaireland.ie/infodesk/cost_of_motoring.asp was used) and $Groupsize_i$ is the number of people that travelled to the site in the respondent's vehicle. Following Shaw and Feather (1999), the opportunity cost of travel time is included in the travel cost calculation as a proportion (0.25) of the hourly wage, where the hourly wage rate was taken as the respondents reported income divided by 2000, based on a 40 hour week for 50 weeks in a year. No allowance for on-site time was made in the travel cost calculation[4]. Relaxing/Sun bathing was highlighted as the main activity of 35% of all respondents in the survey followed by entertaining children (21%), swimming (13%), walking (11%) and other water sports (6%). Also, it is notable that 49% of respondents were male, 57% were in full-time employment and 63% had been educated up to degree level. Mean annual visits to the beach where each respondent was sampled were 11.76 (range 1-60). The day of the survey was the first ever visit to the beach for 7% of the sample and respondents spend on average 2 hours 31 minutes on site. A visit to the beach was the main purpose of the day's journey for 61% of the sample, and participants in the survey used the beach resource for, on average, 4.1 different recreational activities. Mean one-way distance travelled was 24 miles and respondents to the survey tended to be at the beach in groups of, on average, 2.2 persons (range 1 to 13). Further summary statistics associated with the sample are presented in table 1.

4. Results

Given the contingent behaviour scenarios described in Figure 1 and the model specifications described in Section 3 we present the results of a standard panel count data model accounting for endogenous stratification and truncation and the results of the 2 modelling approaches for our on-site sample that accounts for heterogeneity across observations with respect to the explanatory variables. In all models, the average number of trips undertaken by individual *i* under (the real or contingent) scenario *t* is assumed to be a function of the travel cost to the site, the travel cost to the respondent's next preferred substitute site, whether the respondent participates in a water sport while on-site, is a member of a recreation or environmental organisation, income, age, income whether the visit to the beach is by chance due to the respondent being in the area for other business and a 'Contingent Behaviour' variable, which indicates whether the visits we are explaining are actual, with current facilities, or hypothetical, with improved facilities. A further description of each of the independent variables is given in table 1.

[4] An in-depth discussion of the many issues that surround the calculation of the travel cost variable is beyond the scope of the article but for a good over view of the treatment of time and the specification of the travel cost variable in recreation demand models the interested reader is advised to see Hynes et al. (2009).

The model in table 2 is the random effects panel negative binomial accounting for on-site sampling (henceforth referred to as the NB corrected model). The travel cost coefficient in this model is significant at the 5% level and has the expected negative signs. This indicates that, on average, as the cost of travelling to the beach site decreases, the number of trips made to the site increases. The 'travel cost to the nearest substitute site' and the 'Incidental visit to the beach' variable are also significant and have the *a piori* expected signs. The 'contingent behaviour' variable was, surprisingly, found to be insignificant in the NB corrected model. This finding would appear to suggest that the hypothetical trail that facilitates access to a further area of beach does not have a statistically significant effect on the number of planned trips to the site. Once we account for the unobserved heterogeneity in our sample however the 'contingent behaviour' variable in both the latent class and the random parameter corrected NB models are highly significant.

Table 3 presents the results of the negative binomial panel model that allows for unobserved heterogeneous with respect to the impact of explanatory variables on the number of trips taken by the latent sorting of individuals into *C* classes as well as accounting for the issue of on-site sampling (henceforth referred to as the LC model). In order to decide the number of classes, we used the information criteria statistics first developed by Hurvich and Tsai (1989). We report the Akaike information criterion (AIC), the Baysian information criterion (BIC) and the Hannan Quinn statistic for all models in tables 2 and 3. In terms of the LC model no one number of classes minimize each of the measures. The 3 class specification has the lowest score on 2 of the criteria while the 2 class specification is lowest for the BIC. As Scarpa and Thiene (2005) and Hynes et al, 2008) point out these statistics provide guidance on the number of latent classes to choose but this decision also requires the discretion of the researcher. We hence choose only to report in table 3 the latent class corrected NB model estimates for the 2 class model even though two of the information criteria statistic were lower for the 3 and 4-class models. We reject the 4 class model as one of its classes has a complete set of insignificant parameter estimates and also both the 3 and 4 class models displayed a high number of insignificant parameter estimates in at least one of their other classes.

As can be seen from table 3, the travel cost coefficients in both classes are negative and significant at the 5% level and, as mentioned above, the contingent behaviour variable is also significant in both classes(at the 90% level in class 1 and at the 99 % level in class 2). The travel cost variable would appear to have more or less the same influence in both classes which would suggest that both classes exhibit 'price' sensitivity to the same degree. The 'travel cost to the nearest substitute site' and the 'Incidental visit to the beach' coefficients are also significant at the 5% level and have the *a piori* expected signs in both classes of the model. In fact, all variables are now significant at the 95% level in at least one of the two class segments. It is also interesting to note that the income coefficient is now significant for the smaller group of recreationists likely to be represented by class 1. This coefficient was insignificant in the NB corrected model. Only by allowing for taste heterogeneity in the sample do we pick up in the importance of this characteristic for a certain portion of recreationalists using the site. It should also be noted that for this smaller segment participation in water sports has no influence on the number of trips made to the site whereas it has for class 2. The two-class model specification allocated 22% of respondents to class one and 78% to class two.

Finally, it should be noted that the LC model had a lower log likelihood value (in absolute terms) and a lower score on all of the information criteria statistics than the NB corrected

model indicating that the latent class structure provides a better fit for our on-site sampled data that when we assume a homogenous mean influence of the explanatory variables amongst our beach recreationists.
Table 4 presents the results of our Random Parameter On-site Negative Binomial Contingent Behaviour Model (henceforth the RP model). The RP model resulted in the best statistical fit (relative to the on-site standard and LC models) as judged by the value of the log likelihood and the information criteria statistics. A likelihood ratio test comparing the non-random (table 2) and random-parameters models resulted in a χ^2 statistic with 7 degrees of freedom and an associated *p*-value of virtually 0. This indicates that the random-parameters model is statistically superior to the standard on-site NB model. A similar finding was found for the LC model.

The RP model was estimated by specifying a normally distributed functional form of the parameter density function and using simulation-based maximum likelihood with 200 Halton draws (Bhat, 2003 demonstrated empirically that this number of draws can produce accurate parameter estimates). The constant term and the contingent behaviour dummy were specified as non-random parameters. Given that the contingent behaviour variable can be thought of as simply a time period dummy for the before site change and after site change scenarios it would not intuitively make sense to specify it as a random parameter.

The means of the random parameters were statistically different than zero at the 0.01 level or higher for all respondent characteristics. Indeed, bar the water sport participation variable, both the mean of the random parameter and the standard deviation of the random parameter were statistically different than zero at the 0.05 level or higher for all other respondent characteristics. In contrast, some of the parameters for respondent characteristics were not significantly different than zero (at conventional levels) in the non-random NB model or in one or other of the classes in the LC model (the member of a recreational or environmental organisation coefficient was insignificant in both classes in the LC model).

As with the LC model, the NB model with random parameters provides a richer description of trip preferences than the standard on-site NB model (Table 2). For example, the standard on-site NB model indicated that income did not have a statistically significant effect on the demand for trips. The estimate of the income effect in the RP model however provides much greater information in terms of its influence on trip demand. The random parameter estimate on income is now positive but the standard deviation estimate indicates that a positive income effect does not apply to all respondents. Indeed, it would appear from the results that approximately 6 percent of the respondents have a negative income effect.

It is also interesting to note that there is highly significant standard deviation parameter associated with the travel cost coefficient. There is much debate in the literature in terms of the method of calculating travel costs in recreational demand studies. The standard approach is to multiply the distance to the different sites with a per kilometre "out-of-pocket" cost, usually calculated on the basis of marginal vehicle operating costs with perhaps the addition of an estimate of the opportunity cost of leisure time. The most common practice in the treatment of the opportunity cost of leisure time is to value it at the gross wage rate or some fraction thereof. However, Hynes et al. (2009) argue that the opportunity cost of time could be greater than the hourly wage rate (Feather and Shaw

(1999) show that, for those on a fixed work schedule, it is possible for the value of leisure time to be greater than the wage[5]), less than the hourly wage rate (there may be an element of disutility of work time) or equal to the hourly wage rate (from the classical economic perspective that the opportunity cost of other activities equals the marginal wage rate).

Hynes et al. (2009) also showed how sensitive the welfare effects of changes in recreational site quality and access are to the specification of the "price" in travel cost models of recreation demand. By specifying TC as a random parameter as is done in the RP model we allow for the fact that some individuals may gain utility from the extra distance traveled or from the fact that the value of a person's leisure time may be worth more than what that person earns on an hourly basis. By specifying the TC coefficient as random we are therefore also able to reveal the sensitivity of consumer surplus values to the potentially wide distribution of the impact of the travel cost coefficient on trip demand. With this discussion in mind we next turn our attention to the estimation of consumer surplus (CS) from our alternative model specifications. Estimating the welfare effects of changes in the quality or supply of site facilities or environmental goods is the main objective of most contingent behaviour studies. We therefore consider the implications for welfare measures of controlling for the unobserved heterogeneity in our sample through the assigning of distributions in our model parameters. In particular, we compare the CS per trip (real behaviour), the estimates of the change in number of trips taken and the change in total CS per recreationalist as a result of the hypothetical extension to the beach being provided through the creation of an adjoining walking trail, across the alternative model specifications. The welfare results based on the standard NB, the RP and the LC models are shown in Table 5.

The standard panel NB corrected model accounting for truncation and endogenous stratification results in a higher mean CS per trip estimate than the models that allow for unobserved heterogeneous with respect to the impact of explanatory variables on the number of trips taken. The distribution of CS estimates for the LC model varies across classes, with each class having a specific CS per trip estimate. The class weighted population estimate of per-trip consumer surplus for the LC model is estimated with 95% confidence to be between €16.93 and €27.21. With a mean CS per revealed trip estimate of €21.67 and €15.67 for class 1 and 2 respectively this model provides the most conservative mean CS estimates across all the reported models[6]. The RP model provides a distribution of CS values, which we have simulated with 10,000 normal draws and by ordering the results to identify percentiles. The average values of these are used as an approximation to equation (12). The mean CS per trip estimate in this case lies between the equivalent NB corrected model estimate and the weighted LC estimate. Having said that, the chosen percentile point estimates of CS per trip from the RP model provides a much wider distribution of estimates for our sample than either of the other two modeling approaches.

[5] There is also evidence that people are in corner solutions in the labour market where they are forced to work more hours than they would wish. For example, Feather and Shaw (1999) report that almost 50% of their respondents stated that they were "over-employed".

[6] While nothing in the construction of the latent class model assures that the consumer surplus measures in a two class model will bracket the result from a one class model (the NB corrected model) it is still interesting to note that the CS estimate in the NB corrected model does not fall between the 2 class estimates of the latent class corrected NB model. This may be an indication that the one class model is forcing an overestimate of the consumer surplus measure and that that controlling for heterogeneity in the population with respect to the impact of the chosen explanatory variables provides more reliable CS estimates.

To estimate the recreation benefits from the access improvements and the addition of the walking trail and additional beach area, the steps outlined in the methodology section were followed. To calculate the proportional change in recreationalist welfare from implementation of the coastal walking trail, we first take into account the stated change in trips to the beach site if the trail were to be put in place. Such a facility improvement would increase visits by an estimated 3.32 trips per year under the NB corrected model. This is the lowest predicted change in trips across all model specifications.
Even though the LC model provides the lowest mean CS per trip estimates it predicts the second largest change in the number of trips taken per individual as a result of the beach site changes being implemented (6.04 additional trips per person per annum). The RP model predicts the largest change in the number of trips taken per individual (6.28 additional trips per person per annum). The relatively low CS per trip estimate for the LC model means that the estimated total increase in consumer surplus from the beach facility improvements per person per year (the class weighted estimate) is only €0.82 higher than the estimate associated with the NB corrected model (€102.26 and €101.44 respectively). The RP model produces the largest mean estimate for the change in CS per person per year at €140.04. This is approximately 37% larger the equivalent estimate under the LC model.

Finally, while analysts tend to evaluate models based on the likelihood function, which says how well the model is predicting individual behavior conditional on the distributions that have been assumed, it is also worthwhile to evaluate the models based on how well the model is predicting aggregate behavior. On this basis, an examination of Figure 1 and 2 would suggest that our LC and RP models appear to produce latent revealed and contingent trip predictions that lie below that seen in the on-site sample, as one would expect, but can at the same time also predict the small (but deflated) number of higher frequency visitors evident in the sample. The standard NB on-site model does not appear to be able to accomplish that.

5. Discussion and Conclusions

In this paper, we presented an extension to Shaw's (1988) and Englin and Shonkwiler's (1995) count data models corrected for on-site sampling where we incorporated heterogeneity with respect to the impact of the explanatory variables. In doing so we contrasted two contingent behavior panel modelling techniques, namely, the random parameter negative binomial model and the latent class negative binomial model and applied them to revealed and contingent travel data obtained from a survey of visitors to a beach on the outskirts of Galway city in Ireland. We then derived welfare estimates relating to individual consumer surplus per trip and relating to changes in the quality of the beach site. The distribution of these welfare estimates were derived based on the alternative modeling strategies. What this article contributes to the literature is the development of panel on-site count data models that allow for the mixing of preference intensities with respect to trips taken in the relevant population.

While Egan and Herriges (2006), Beaumais and Appéré (2010) and Moeltner and Shonkwiler (2010) have previously developed count data panel models corrected for on-site sampling their approaches may still be inadequate and potentially misleading if the population of interest is heterogeneous with respect to the impact of the chosen explanatory variables. Adding a random parameters or latent class framework to these on-site panel models facilitates a much deeper analysis of the factors driving the decision to make a particular number of trips to a recreational site. It also highlights the fact that there are distinct segments of the population who make that decision based on different influences.

Also, allowing for heterogeneity in terms of the impact of the travel cost coefficient on the number of trips taken facilitates the notion often debated in the travel cost literature that some individuals may gain utility from the extra distance traveled and that the value of a person's leisure time may be worth more than what that person earns on an hourly basis. By specifying the travel cost coefficient as being heterogonous in terms of its impact on trip demand, as is accommodated in the LC and RP models, we can demonstrate the sensitivity of consumer surplus values to the potentially wide distribution of the travel cost coefficient.

For the application considered in this paper it was found that accounting for individual heterogeneity in parameter estimates greatly improved the statistical fit of the contingent behaviour models. It was also interesting to note that mean estimates of consumer surplus per trip were found to be lower in both the RP and LC models relative to the corresponding standard on-site NB model. This may be due to the effect of omitted explanatory variables in the standard NB corrected model, which are picked up on to some extent by accommodating heterogeneity of the parameters in the RP and LC models. Further research is needed in this regard.

Both the latent class and the random parameter approaches also generates additional information which is potentially very useful to recreational site managers, simply by identifying groups of users with particular demands. In this study, segments of the sample that contributed strongly to recreation demand at the beach site, such as higher income individuals, were identified in the random parameter models (and a smaller cohort of lower income individuals in one segment of the latent class model), whereas they were not evident in the standard on-site panel count data analysis. Understanding the degree of heterogeneity in recreation demand is therefore an important research area for both practitioners and site managers. Being able to identify different types of users within a count data modelling framework should allow such managers to better allocate resources between policy issues such as beach congestion and access by boat owners to the water.

An obvious question to ask is which approach to modelling unobserved preference heterogeneity in a count data setting is preferable? We would argue that RP approach may be more relevant in situation where the recreationalist being analysed are coming from a diverse population as may be the case with beach users. However it should be noted that while the RP approach identifies which attributes have significant levels of heterogeneity in preferences, and quantifies the degree of the spread of values around the mean the analyst must impose a distributional form on preferences, and welfare estimates can be quite sensitive to this choice. On the other hand, if the recreational groups of interest can be obviously catergorised such that the spread of preferences is what Hynes et al (2008) refer to as "lumpy", such that broad classes of people exist with rather similar values to each other, but rather different values to everyone else, then the latent class approach makes more sense as it is flexible across classes but imposes the homogeneity assumption within the classes. Therefore, if were interested in a contingent behavior analysis of a distinct group of recreationalists; for example, different types of rock climbers, or kayakers of different skill levels then the LC approach may be preferable.

Given the relatively small sample size it would be wise to take a cautious view as to how representative the estimated welfare results are of the population of beach users in the west of Ireland. Nevertheless the estimated models still demonstrate how controlling for unobserved heterogeneity with respect to the impact of the explanatory variables can have a significant impact on predicted trips taken and on welfare estimation. Finally, it is important to state that while the focus of the paper was on incorporating heterogeneity in the preferences for recreation within an on-site model of contingent behaviour, the developed modelling framework is just as applicable to cases where data has been collected on-site in relation to trips taken by the same individuals over repeat time periods or on an individual's trip activity to alternative sites over a fixed period.

References

Anastasopoulos, P., Mannering, F., 2009. A note on modeling vehicle accident frequencies with random-parameters count models. *Accident Analysis and Prevention* 41: 153–159.

Baerenklau, K.A., 2010. A latent class approach to modeling endogenous spatial sorting in zonal recreation demand models. *Land Economics* 86(4), 800-816.

Bhat, C., 2003. Simulation estimation of mixed discrete choice models using randomized and scrambled Halton sequences. *Transportation Research Part B* 37 (1), 837–855.

Beaumais, O., Appéré, G., 2010. Recreational shellfish harvesting and health risks: A pseudo-panel approach combining revealed and stated preference data with correction for on-site sampling. *Ecological Economics* 69 (12), 2315–2322.

Christie, M., Hanley N., Hynes, S., 2007. Valuing enhancements to forest recreation using choice experiments and contingent behaviour methods. *Journal of Forest Economics* 13 (2), 75-102

Egan, K., Herriges, J., 2006. Multivariate count data regression models with individual panel data from an on-site sample. *Journal of Environmental Economics and Management* 52, 567–581.

Englin, J., Shonkweiler, J., 1995. Estimating social welfare using count data models: An application to long-run recreation demand under conditions of endogenous stratification and truncation. *The Review of Economics and Statistics* 77(1), 104-112.

Greene, W. H., 2007. *NLOGIT 4 Econometric Modeling Guide*, Volume 2, Econometric Software, Inc. New York

Greene, W. H., 2008. *Econometric Analysis*, 6th ed., New Jersey: Prentice Hall

Grijalva, T., Berrens, R., Bohara, A., Shaw, W.D., 2002. Testing the validity of contingent behavior trip responses. American Journal of Agricultural Economics 84 (2), 401–414.

Grogger, J., Carson, R., 1991. Models for truncated counts. *Journal of Applied Econometrics* 6, 225-238.

Hanley, N., Bell, D., Alvarez-Farizo, B., 2003. Valuing the benefits of coastal water quality improvements using contingent and real behaviour. *Environmental and Resource Economics* 24(3), 273-285.

Hellerstein, D., Mendelsohn, R., 1993. A theoretical foundation for data models. *American Journal of Agricultural Economics* 75(3), 604-611.

Holmes, T., Englin, J., 2010. Preference Heterogeneity in a Count Data Model of Demand for Off-Highway Vehicle Recreation. *Agricultural and Resource Economics Review*, 39: 75–88

Hurvich, M., Tsai, C., 1989. Regression and time series model selection in small samples. *Biometrika* 76, 297-307.

Hynes, S., Hanley, N., Scarpa, R., 2008. Effects on Welfare Measures of Alternative Means of Accounting for Preference Heterogeneity in Recreational Demand Models. *American Journal of Agricultural Economics* 90 (4), 1011-1027.

Hynes, S., Hanley, N., O'Donoghue, C., 2009. Alternative treatments of the cost of time in recreational demand models: an application to white water rafting in Ireland. *Journal of Environmental Management* 90(2), 1014-1021.

Martınez-Espineira, R., Amoako-Tuffour, J., 2008. Recreation demand analysis under truncation, overdispersion, and endogenous stratification: An application to Gros Morne National Park. *Journal of Environmental Management*, 88(4), 1320-32.

Moeltner, K., Shonkwiler, J., 2010. Intercept and recall: Examining avidity carryover in on-site collected trip reports. *Transportation Research Part D: Transport and Environment* 15(7), 418-427.

Scarpa, R., Thiene, M. 2005. Destination choice models for rock-climbing in the North-Eastern Alps: a latent-class approach based on intensity of participation. *Land Economics* 81, 426-444.

Scarpa, R., Thiene, M., Tempesta, T., 2007. Latent class count models of total visitation demand: days out hiking in the eastern Alps. *Environmental & Resource Economics*, 38(4), 447-460.

Shaw, D., 1988. On-Site Sample's Regression: Problems of Non-Negative Integers, Truncation, and Endogenous Selection. *Journal of Econometrics* 37, 211-223.

Shaw, D., Feather, P., 1999. Possibilities for including the opportunity cost of time in recreation demand systems. *Land Economics* 75(4), 592-602.

Train, K., 1998. Recreation demand models with taste differences over people. *Land Economics* 74, 185-194.

Vermunt, J., Magidson, J., 2005. Latent Gold 4.0 User's Guide. Belmont, Massachusetts: Statistical Innovations Inc.

Wang, P., Cockburn, I., Puterman, L., 1998. Analysis of patent data—a mixed poisson-regression-model approach." *Journal of Business and Economic Statistics* 16 (1), 27–41.

Wedel, M., Desarbo, W., Bult, J., Ramaswamy, V., 1993. A Latent Class Poisson Regression Model for Heterogeneous Count Data. *Journal of Applied Econometrics* 8(4):397-411.

Whitehead, J., Pattanayak, S., Van Houtven, G., Gelso, B., 2008. Combining revealed and stated preference data to estimate the non-market value of ecological services: An assessment of the state of the science. *Journal of Economic Surveys* 22, 872-908.

Figure 1. Distribution of Actual Sample Trips and Estimated Revealed Trips from On-site NB Models .

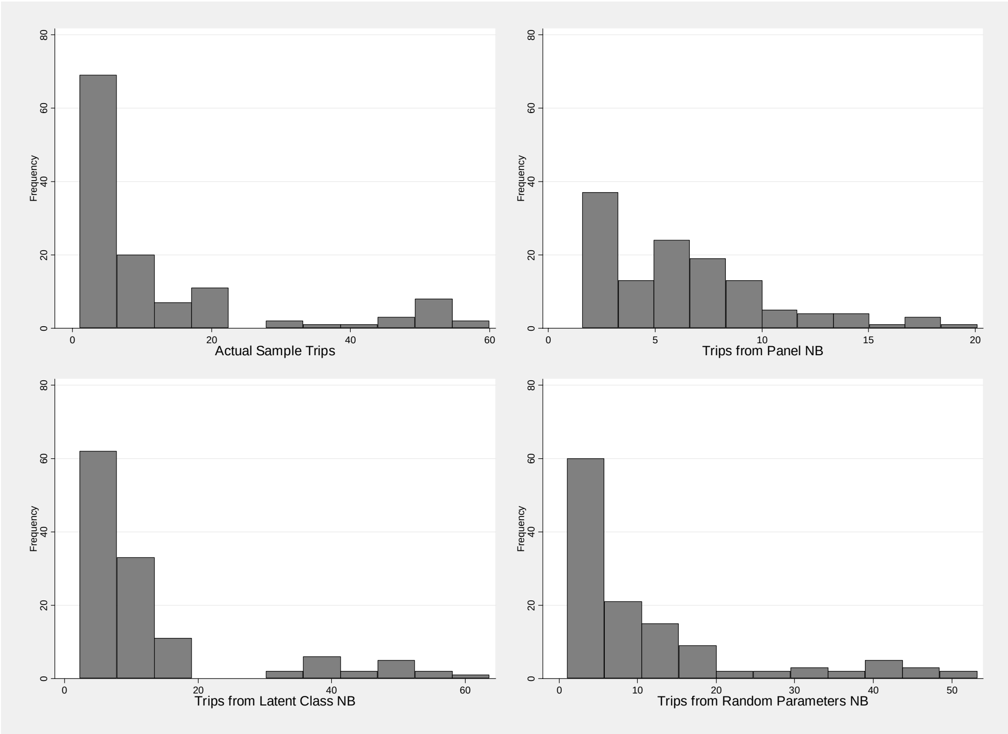

Figure 2. Distribution of Actual Sample Trips and Estimated Contingent Trips from On-site NB Models.

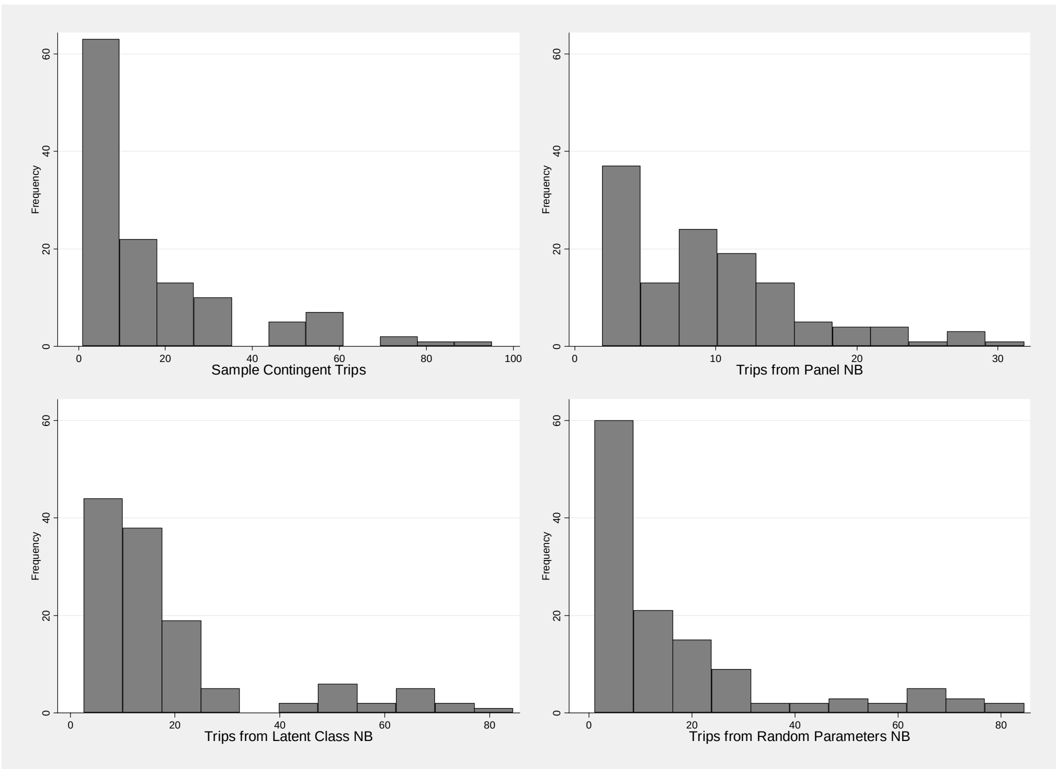

Table 1. Summary Statistics

Variable Name	*Description*	*Mean*	*Standard Deviation*
Actual trips	Number of trips respondent actually took to the beach in last 12 months	11.76	14.9
Hypothetical trips	Number of trips respondent would take in next 12 months if scenario implemented	17.31	19.23
Age	Age	41.06	13.68
Income	Gross annual income (€)	51,551	29,334
Incidental Visit to Beach	Dummy indicating whether trip to beach occurred by chance as happened to be in the area anyway (1) or was a planned trip to the beach (0)	0.39	0.49
Member of Recreation or Environmental Organisation	Dummy variable Indicating whether the respondent is an active member of a recreational organisation such as a kayak or surf club or an environmental organisation such as Birdwatch Ireland or Greenpeace	0.47	0.5
Travel Cost	Return travel cost from home to beach	15.28	17.43
Travel Cost Substitute Site	Return travel cost to the alternative site most frequently visited by respondent	13.77	15.32
Water Sport Participation	Dummy variable indicating whether trip to beach involved a water sport	0.15	0.36

Table 2. Negative Binomial Contingent Behaviour Model Adjusted for Truncation and Endogenous Stratification

	Negative Binomial Panel Count Model
Age	0.156*** (0.035)
Income	-0.003 (0.002)
Incidental Visit to Beach	-1.202*** (0.145)
Member of Recreation or Environmental Organisation	0.404*** (0.094)
Contingent Behaviour	0.481 (0.388)
Travel Cost	-0.033*** (0.009)
Travel Cost Substitute Site	0.032*** (0.009)
Water Sport Participation	0.553*** (0.145)
Constant	0.463 (0.420)
Scale Parameter	1.345 (1.584)
AIC	1735
BIC	1771
Log likelihood	-858

Standard errors are in parentheses. *** indicates significance at the 1% level, ** indicates significance at the 5% level and * indicates significance at the 10% level. The income variable has been rescaled by dividing by 1000.

Table 3. Latent Class Negative Binomial Contingent Behaviour Model accounting for Truncation and Endogenous Stratification

	Latent Class Negative Binomial Panel Count Model	
	Latent Class 1	*Latent Class 2*
Age	0.104*** (0.031)	0.215*** (0.052)
Income	-0.005** (0.002)	0.003 (0.002)
Incidental Visit to Beach	-1.481*** (0.197)	-0.680*** (0.221)
Member of Recreation or Environmental Organisation	-0.052 (0.101)	0.180 (0.114)
Contingent Behaviour	0.292* (0.172)	0.666*** (0.210)
Travel Cost	-0.047** (0.019)	-0.064*** (0.017)
Travel Cost Substitute Site	0.067*** (0.022)	0.039** (0.016)
Water Sport Participation	0.166 (0.155)	0.437** (0.182)
Constant	3.529*** (0.184)	0.534* (0.280)
Alpha	0.051** (0.026)	0.722*** (0.176)
Class Probabilities	0.217*** (0.040)	0.783*** (0.040)
AIC	1605	
BIC	1679	
Log likelihood	-781	

Standard errors are in parentheses. *** indicates significance at the 1% level, ** indicates significance at the 5% level and * indicates significance at the 10% level. The income variable has been rescaled by dividing by 1000.

Table 4. Random Parameter Negative Binomial Contingent Behaviour Model accounting for Truncation and Endogenous Stratification

	Mean of coefficient	Standard Deviation of coefficient
Random Parameters in Utility Functions		
Age	0.19 (0.08)***	1.15 (0.02) ***
Income	0.005 (0.001)***	0.008 (0.001)***
Incidental Visit to Beach	-0.41 (0.10) ***	0.13 (0.06)**
Member of Recreation or Environmental Organisation	0.41 (0.085)***	0.71 (0.04)***
Travel Cost	-0.032 (0.005) ***	0.03 (0.001)***
Travel Cost Substitute Site	-0.028 (0.007) ***	0.03 (0.001)***
Water Sport Participation	0.69 (0.06) ***	0.06 (0.05)
Non-Random Parameters in Utility Functions		
Constant	1.37 (0.09)***	
Contingent Behaviour	0.47 (0.05)***	
Scale Parameter	0.029 (.01)***	
AIC	1557	
BIC	1617	
Log likelihood	-761	

Standard errors are in parentheses. *** indicates significance at the 1% level, ** indicates significance at the 5% level and * indicates significance at the 10% level. The income variable has been rescaled by dividing by 1000.

Table 5. Consumer Surplus (CS) and Change in Trips Taken Estimates from Alternative On-Site Model Specifications (all figures are per person).

Model Specification	Mean CS per Trip (€)	Change in number of trips taken a result of new walking trail	Change in Annual CS as a result of new walking trail (€)
Negative Binomial	30.54 (14.11, 46.96)	3.32	101.44
Latent Class NB Model			
LC Negative Binomial: Class 1	21.43 (4.20, 38.65)	6.04	129.39
LC Negative Binomial: Class 2	15.67 (7.36, 23.98)	6.04	94.61
Weighted LC Negative Binomial*	16.93 (6.66, 27.21)	6.04	102.26
Random Parameter NB Model			
25th Percentile	14.22 (6.01, 22.43)	6.28	89.30
50th Percentile	22.64 (13.18, 32.09)	6.28	142.18
75th Percentile	37.65 (29.39, 45.91)	6.28	236.44
Mean	22.30 (14.75, 29.86)	6.28	140.04

Ninety five percent confidence interval in brackets. * This is the weighted consumer surplus per trip estimate estimated by considering the class probabilities in the NB Latent Class Model.** Source: Calculated from model results reported in table 4 and based on 10,000 draws from the estimated population distribution.

semru

Socio-Economic Marine Research Unit