A Semi-Parametric Estimator for Revealed and Stated Preference Data

An Application to Recreational Beach Visitation

Craig E. Landry* and Haiyong Liu**

Department of Economics

East Carolina University

revised

May 2008

* Corresponding author: landryc@ecu.edu; 252-328-6383; 252-328-6743 (fax)

** Thanks are due to Dan Phanuef, Marty Smith, and two anonymous reviewers for helpful comments and to John Whitehead and the U.S. Army Corps of Engineers for providing the data.

A Semi-Parametric Estimator for Revealed and Stated Preference Data

An Application to Recreational Beach Visitation

Abstract

We present a semi-parametric approach for jointly estimating revealed and stated preference recreation demand models. The discrete factor method (DFM) allows for correlation across demand equations and incorporates unobserved heterogeneity. Our model is a generalized negative binomial with random effects; the random effect is composed of a discrete representation of unobserved heterogeneity and a factor loading that translates the heterogeneity measure into a demand effect. Our empirical application is to beach recreation demand in North Carolina. Statistical evidence supports our DFM specification, which imposes less restriction on model dispersion and incorporates unobserved heterogeneity in a flexible manner. Elasticity estimates are smaller than those derived from models with parametric specifications for unobserved heterogeneity, and welfare estimates are slightly larger (and less precise). While parametric models clearly dominate if the specification of unobserved heterogeneity is correct, the semi-parametric DFM provides a flexible alternative in cases where mis-specification is a potential problem.

JEL: C81, D12, Q51

Key words: beach recreation demand, revealed and stated preference, unobserved heterogeneity, semi-parametric

1. Introduction

Resource economists are increasingly gathering revealed preference (RP) data in conjunction with information on stated preference (SP). Adamowicz, Louviere, and Williams [2] note the potential for combining RP and SP data so that one can explore behavior associated with levels of environmental quality that are not observed. Aside from changes in environmental quality [22, 25, 30], others have used SP data to examine behavioral changes stemming from variations in travel cost [4, 10, 13], in access to resources [16, 43], and in management conditions [24]. Through combining information on revealed and stated behavior, the analyst can potentially learn more about underlying preferences and test for various restrictions. In addition, information on real behavior may be helpful in calibrating or validating stated preference data [22, 43].

In this paper we estimate parameters of revealed and stated recreation demand using a semi-parametric technique—the discrete factor method (DFM) (also known as discrete factor approximation). This method permits us to account for unobserved heterogeneity across agents, while at the same time allowing for correlation across RP and SP demand equations. We condition the joint distribution of revealed and stated demand on a factor that represents unobserved heterogeneity, which is approximated by a step function. The unconditional likelihood function is obtained by summing over the discrete distribution of the unobserved factor using empirically estimated probabilities for inclusion in the latent segments. Due to the use of a discrete distribution for unobserved heterogeneity the estimator falls in the class of finite mixture models. An attractive feature of this approximation approach is that a small number of supports for the discrete distribution (e.g., three or four) has been shown to compare favorably (on the basis of precision and unbiasedness) to alternative estimators that require more restrictive distributional assumptions [29].[1] Thus, the DFM is generally robust, computationally simple, and quite flexible. This approach has not, to our knowledge, been applied to recreation demand data.

Our econometric approach for modeling a quasi-panel of recreation demand makes use of a generalized negative binomial model. Conceptually, an individual-specific random effect enters each demand equation additively and is scaled by a factor loading that varies across equations. Individual-specific heterogeneity cannot be identified, but rather is approximated by a discrete distribution, the probability mass and supports of which are estimated empirically. Model parameters are estimated via quasi-maximum likelihood. Our data pertain to beach recreation demand in North Carolina.

Our DFM generalized negative binomial (DFM-GNB) specification performs quite well in comparison with competing models according to a number of statistical tests based on likelihood values. Using a likelihood ratio test (LRT), we reject the standard NB1 and NB2 specifications of the negative binomial model at conventional significance levels. Using LRT and information criteria, we find support for DFM-GNB over a generalized negative binomial without unobserved heterogeneity. Comparing DFM to parametric approaches for incorporating unobserved heterogeneity, information criteria and Vuong's [42] non-nested likelihood ratio test favor the DFM-GNB model over normality-based estimates.

Elasticities derived from the DFM model consistently exhibit less responsiveness than the parametric models, suggesting that more variation in recreation demand is attributed to unobserved heterogeneity as we move from a parametric specification to a less restrictive alternative. Average predicted recreation trips are very close to the sample means for the DFM-GNB, while predicted trips for normality-based estimates appear to exhibit more bias. This bias is extremely large for the multivariate Poisson log-normal (MPLN) model [13]. We construe the poor performance of the MPLN model as evidence of mis-specification of the distribution of unobserved heterogeneity. Nonetheless, MPLN estimates exhibit much tighter confidence intervals. On this basis, the normality-based estimates clearly dominate if the parametric specification of unobserved heterogeneity is correct. The semi-parametric DFM approach, however, provides a flexible alternative in cases where mis-specification is a potential problem. As mis-specification of unobservable components is difficult to assess, given its robustness and simplicity of estimation the DFM remains a viable alternative.

2. Combining Stated and Revealed Preference Data

We formulate a recreation demand model that incorporates revealed demand under current conditions and stated demand under both current conditions and hypothetical improvements in resource quality. Since we have multiple observations on each individual (quasi-panel data), we account for unobserved heterogeneity at the individual level in a way similar to the standard random effects model for panel data. Our econometric specification is a generalized negative binomial model, in which we use a semi-parametric technique—the discrete factors method (DFM)—to account for unobserved heterogeneity and permit correlation across RP and SP demand equations.

Our analysis of recreation demand is based to on the following negative binomial model [11]:

$$\Pr(y_{it} \mid x_{it}) = \frac{\Gamma(y_{it} + \alpha^{-1})\alpha^{y_{it}} \mu_{it}^{y_{it}} (1 + \alpha\mu_{it})^{-(y_{it}+\alpha^{-1})}}{\Gamma(y_{it} + 1)\Gamma(\alpha^{-1})}, \tag{1}$$

where i indexes individuals; t = 1, 2, …, T represents demand treatments (RP or SP demand under quality conditions q_t), y_{it} is recreation demand for individual i under treatment t, $\mu_{it} = \exp(x_{it}'\beta)$ is individual i's conditional recreation demand for treatment t, with x_{it} being a matrix of covariates and β denoting a vector of unknown parameters. The variance of trip demand under treatment t is given by Var[$y_{it} \mid x_{it}$] = $\mu_{it} + \alpha\mu_{it}^2$, and α is an additional parameter to be estimated. Cameron and Trivedi [11] refer to this specification as NB2. Dispersion (variance divided by the mean) in NB2 is proportional to the mean, $1 + \alpha\mu_{it}$.

A generalized version of the negative binomial model in equation (1) is:

$$\Pr(y_{it} \mid x_{it}) = \frac{\Gamma(y_{it} + \alpha_t^{-1}\mu_{it}^{2-p})\alpha_t^{y_{it}} \mu_{it}^{(py_{it}-2y_{it})} (1 + \alpha_t\mu_{it}^{p-1})^{-(y_{it}+\alpha_t^{-1}\mu_{it}^{2-p})}}{\Gamma(y_{it} + 1)\Gamma(\alpha_t^{-1}\mu_{it}^{2-p})}, \tag{2}$$

where p is an additional parameter to be estimated [12], and the scale parameter α_t is allowed to vary across treatments. The conditional variance for y_{it} is Var[$y_{it} \mid x_{it}$] = $\mu_{it} + \alpha_t\mu_{it}^p$. This specification provides a more flexible characterization of dispersion—$1 + \alpha_t\mu_{it}^{p-1}$—and nests specifications NB1 (p = 1) and NB2 (p = 2) [12]. We introduce a random effects component (ε_{it}) to conditional demand. Our random effect is decomposed into a scalar representation of unobserved heterogeneity (λ_i) that is common to all demand equations for the same individual and a coefficient, or factor loading (γ_t), that is common to a specific demand treatment across all individuals. Expected demand conditional on observable characteristics and unobserved heterogeneity type is:

$$\mu_{it}(\beta \mid x_{it}, \lambda_i) = \exp(x_{it}\beta + \varepsilon_{it}) = \exp(x_{it}\beta + \gamma_t\lambda_i) = \exp(x_{it}\beta)\exp(\gamma_t\lambda_i). \tag{3}$$

The common heterogeneity term allows for cross-equation correlation: $\text{cov}[\varepsilon_{ij}, \varepsilon_{im}] \neq 0$ for any two demand equations $j \neq m$. The semi-parametric DFM approximates λ_i with a step function that takes K discrete values; from hereon we subscript λ with k rather than i to denote this approximation. The factor loadings (γ_t) rescale unobserved heterogeneity into a recreation demand effect that can vary across demand equations.

Note that in implementing the DFM, we decompose the errors across multiple equations so that one random component is common to individuals and another is unique to each demand equation.[2] In other words, with the common discrete factor, we allow all heterogeneity components to be correlated. The magnitude of correlations, however, is determined by the factor loadings (γ_t). This approach of introducing correlation is different from that of parametric random effects as there is no need to explicitly specify the structure of the variance-covariance matrix. Nevertheless, this approach can still suffer from the curse of dimensionality if multiple random effects are specified. For instance, one may assume there is unobserved heterogeneity in the own-price, cross-price, and income effects. In this case, estimation would become much more difficult: a model with 3 independent random effects with 4 points of support each would imply 64 conditional likelihood functions to enter the unconditional likelihood function.[3]

3. Discrete Factors Method and Likelihood Function

DFM was proposed by Heckman and Singer [19] as an approach for modeling unobserved heterogeneity. This method has two distinct advantages in the class of mixture distribution estimators. First, DFM does not impose *a priori* arbitrary distributional forms for unobserved heterogeneity, while maintaining the asymptotic efficiency of maximum likelihood estimators [29]. The distribution of heterogeneity type is approximated with a step function and integrated out through a weighted sum of step levels [19], where the weights are given by empirically estimated probabilities. Mroz [29] demonstrates that when the true correlation of the error terms is multivariate normal, DFM performs well in comparison with estimators which assume multivariate normality; and when the underlying distribution is non-normal, DFM dominates other normality based estimators in terms of unbiasedness and precision. Second, DFM is computationally simple. For instance, the MPLN model adopted by Egan and Herriges [13] requires evaluating multidimensional integrals based on the assumption that the random effects across equations follow a multivariate lognormal distribution. Although simulation methods are adopted to make these evaluations feasible, the computations are somewhat cumbersome. Using DFM, the likelihood function conditional on unobserved heterogeneity can be constructed as follows.

A priori we do not know to which class of heterogeneity each individual belongs. The likelihood function for individual i is thus:

$$\begin{aligned} L_i &= \sum_{k=1}^{K} \{\Pr(\lambda_k) \Pr[y_{i1},\ldots,y_{iT} \mid x_{it},\lambda_k]\} \\ &= \sum_{k=1}^{K} \left\{\Pr(\lambda_k) \prod_t \Pr(y_{it} \mid x_{it},\lambda_k)\right\}, \end{aligned} \tag{4}$$

where $\Pr(\lambda_k)$ is the probability of individual i having heterogeneity level k, and there are K levels of heterogeneity. Combining the generalized negative binomial specification (equation 2) with DFM (equations 4), individual i's contribution to the likelihood function, L_i, can be rewritten as:

$$\sum_{k=1}^{K}\left\{\Pr(\lambda_k)\prod_t \frac{\Gamma(y_{it}+\alpha_t^{-1}\mu_{itk}^{2-p})\alpha_t^{y_{it}}\mu_{itk}^{(py_{it}-2y_{it})}(1+\alpha_t\mu_{itk}^{p-1})^{-(y_{it}+\alpha_t^{-1}\mu_{itk}^{2-p})}}{\Gamma(y_{it}+1)\Gamma(\alpha_t^{-1}\mu_{itk}^{2-p})}\right\}. \tag{5}$$

Note the conditional mean μ_{itk} reflects individual-level observables, demand treatment, and heterogeneity type. The sample likelihood function is derived as the product of (5) over all N individuals:

$$L=\prod_{i=1}^{N}\left(\sum_{k=1}^{K}\left\{\Pr(\lambda_k)\prod_t \frac{\Gamma(y_{it}+\alpha_t^{-1}\mu_{itk}^{2-p})\alpha_t^{y_{it}}\mu_{itk}^{(py_{it}-2y_{it})}(1+\alpha_t\mu_{itk}^{p-1})^{-(y_{it}+\alpha_t^{-1}\mu_{itk}^{2-p})}}{\Gamma(y_{it}+1)\Gamma(\alpha_t^{-1}\mu_{itk}^{2-p})}\right\}\right). \tag{6}$$

We refer to this model as the discrete factors method-generalized negative binomial (DFM-GNB) specification. The semi-parametric specification of unobserved heterogeneity allows unrestricted correlations across RP and SP demand equations for the same individual *i*.

Our specification imposes a restriction that all unobserved heterogeneity and correlation among the individual demand equations enters the full model through the factor loadings γ_t and the factor λ_k, with $\Pr(\lambda_k)>0, \forall k=1,\dots,K,$ where K is the total number of points support for the discrete distribution. Without loss of generality, λ is confined to the unit interval:

$$\begin{aligned}
\lambda_1 &= 0\\
\lambda_2 &= \frac{1}{1+\exp(\overline{\lambda_2})}\\
\lambda_3 &= \frac{1}{1+\exp(\overline{\lambda_3})}\\
&\dots\\
\lambda_{K-1} &= \frac{1}{1+\exp(\overline{\lambda_{K-1}})}\\
\lambda_K &= 1.
\end{aligned} \tag{7}$$

The transformed probability weights are given as follows:

$$\Pr(\lambda_k) = \frac{\exp \overline{\theta_k}}{1 + \sum_{k'=1}^{K-1} \exp \overline{\theta_{k'}}}, k = 1, ..., K-1$$
$$\Pr(\lambda_k) = \frac{1}{1 + \sum_{k'=1}^{K-1} \exp \overline{\theta_{k'}}}, k = K. \qquad (8)$$

The logit kernals in (7) and (8) are strictly concave, facilitating optimization procedures. Note that the $K - 2$ vector $\overline{\lambda}$ and the $K - 1$ vector $\overline{\theta}$ are parameters to be estimated, along with T vectors α and γ, and the standard vector of demand parameters β. The support points, λ_k, and the transformed probabilities, $\Pr(\lambda_k)$, can be calculated from the parameter estimates via equations (7) and (8).

The incorporation of latent segments with endogenous probabilities makes the DFM similar to the latent class model,[4] recently applied to recreation choice data [8, 32, 33, 35, 38] and choice experiments [17, 34]. The discrete factor method and latent class model differ in two important ways. First, the latent class model produces separate covariate parameter estimates for each latent segment, while DFM does not unless the discrete factor is interacted with model covariates. In the standard setup the discrete factor is only interacted with a constant term. Second, the latent class model often imposes *a priori* assumptions about determinants of classification by including covariates in the class probabilities, while DFM does not. For instance, Boxall and Adamowicz [8] employ a factor analysis of 20 "motivational indicators" so that probability weights of the latent classes can be constructed. More commonly, individual characteristics are included as covariates for the class probabilities [17, 32, 34, 38].[5] As indicated by Greene [15, pp. 440], if the class probabilities are fixed parameters and only a constant term is included as a covariate in the mean of the random parameter distribution, then the latent class model is equivalent to a DFM approach in modeling the distribution of unobserved heterogeneity.

While parametric mixing approaches such as the mixed logit (e.g. [39]) are available for random utility models of recreation choice, the latent class approach offers a potential advantage in that it makes use of more information in the dataset by allowing for inclusion of individual characteristics as covariates in class probabilities in the conditional logit model. Scarpa, Thiene, and Tempesta [36] is one of the only studies that apply a finite mixture model to introduce preference heterogeneity in modeling total demand for recreation (i.e., not a site choice model). They apply the latent class model to a system of hiking demand equations for the Italian Eastern Alps.

We contend that the DFM is more appropriate for this type of recreation demand data for a number of reasons. First, the latent class model is demanding in terms of the number of parameters to be estimated. As estimated parameters increase drastically with rising number of classes, Scarpa, Thiene, and Tempesta [36] can only include three or four covariates in most specifications even when the information criteria favor a model with only two classes of unobserved heterogeneity. DFM, on the other hand, utilizes a relatively parsimonious specification. Second, in the zero-inflated count models employed by Scarpa, Thiene, and Tempesta, all parameters are constrained to be equal across 18 destination sites (the panel dimension). DFM allows for some variability across the panel dimension (on our case, RP and SP demand), with the degree of variability depending upon the specification.[6] Third, the basic structure of DFM allows for correlation in the panel dimension (across sites in Scarpa, Thiene, and Tempesta or across demand equations in our study), while the latent class model would require modification to accommodate this type of correlation. Fourth, identification can be problematic in latent class specifications, as individual covariates can appear in both the class probabilities and the demand equations. Most often we do not have exclusion conditions – information on variables that belong to class probability but not to the main demand equations; in many cases, identification is based purely on functional form.

On the other hand, if one does has *a priori* knowledge about covariates that affect class membership, the standard DFM model will not take account of this information. Moreover, the manner in which the unobserved heterogeneity component is introduced in DFM is rather restrictive compared to a latent class model. In particular, the heterogeneity term is additive in the main demand equation unless covariate interaction terms are introduced.

If we allow the discrete factor to interact with not only a constant term, but also model covariates, we obtain a discrete analog to the random parameters model [1]:

$$\mu_{it}(\beta \mid x_{it}, \lambda_i) = \exp[x_{it}(\beta + \delta\lambda_i) + \gamma_t \lambda_i] = \exp[x_{it}(\beta + \delta\lambda_i)]\exp(\gamma_t \lambda_i), \qquad (3')$$

where δ is a vector of parameters for the interaction terms. Employing specification (3') will produce class-specific parameter estimates, but in a more parsimonious way than the latent class model for rank[x] > k > 2. The number of parameter estimates increases by rank[x] in moving from specification (3) to (3') regardless of the number of classes, while for the latent class model the number of parameter estimates increases by rank[x]×(k-1) because separate parameters are estimated for each class. The downside of specification (3') is the fairly restrictive way that class-specific estimates are produced; they are rescalings of the common heterogeneity component and may be difficult to interpret.

For comparison purposes, we also estimate a GNB model with additive unobserved heterogeneity that follows a standard normal distribution and the multivariate Poisson-lognormal (MPLN) model of Egan and Herriges [13]. For the MPLN model,

$$\Pr(y_{it} \mid x_{it}) = \frac{\exp(-\mu_{it})(\mu_{it})^{y_{it}}}{y_{it}!},$$

which is the standard Poisson density, and $\mu_{it} = \exp(x_{it}'\beta + \varepsilon_{it})$, where $\varepsilon_i \equiv (\varepsilon_{i1},\ldots,\varepsilon_{iT})$ is assumed to follow a multivariate normal distribution, i.e. $\varepsilon_i \sim N(0,\Omega)$. Following Egan and Herriges, we impose the same restrictions on the structure of the variance-covariance matrix in the MPLN model. In particular, the unobserved error component for stated trips (t = 2, 3, 4) have the same correlations with each other (though different variances) and the same correlations with observed trips (t = 1). The variance-covariance matrix is therefore assumed to take the following form:

$$\Omega = \begin{bmatrix} \sigma_1^2 & \rho_{rp}\sigma_1\sigma_2 & \rho_{rp}\sigma_1\sigma_3 & \rho_{rp}\sigma_1\sigma_4 \\ & \sigma_2^2 & \rho_{sp}\sigma_2\sigma_3 & \rho_{sp}\sigma_2\sigma_4 \\ & & \sigma_3^2 & \rho_{sp}\sigma_3\sigma_4 \\ & & & \sigma_4^2 \end{bmatrix}.$$

4. Data

The survey data, originally analyzed in Whitehead et al. [44], contain information on recreation demand for seventeen beaches in five southeastern North Carolina counties, including both revealed and stated visitation. The stated preference responses describe intended visitation in the subsequent year under current conditions, as well as how beachgoers would change visitation in the subsequent year in response to hypothetical improvements in parking spaces/access points and beach width. The data were gathered via random telephone survey of North Carolina counties within 120 miles of the study site in spring of 2004. The response rate was 52 percent, with a final sample size of 664.[7]

For purposes of our analysis, demand is aggregated over the seventeen beach sites. Descriptive statistics are included in table 1. The average number of observed trips is 11, while planned trips in the subsequent year under current conditions is slightly higher at 13. Improvements in parking facilities that would obviate beachgoers' need to search for a parking space or beach access point (holding parking fees and beach congestion constant) increase the average number of stated trips to 17 in the subsequent year. Implementation of a beach replenishment policy to improve beach width by an average of 100 feet increases the average number of stated trips in the subsequent year to 14. Travel distance is measured as distance between population centers at the home ZIP code and the nearest beach county ZIP code. Travel costs are measured as the sum of pecuniary ($0.37/mile) and time (33% of wage) costs, assuming an average speed of 50 miles per hour. Average trip cost to southeastern NC beaches is $89, while average cost to the Outer Banks (a substitute site) is $202. Average income is $59 thousand per year.

We thus focus on the demand for recreational beach trips:

$$\begin{aligned} y_{itk} &= \exp(\beta_0 + \beta_{op} op_i + \beta_{cp} cp_i + \beta_m m_i + \beta_X X_i + \beta_Z Z_{it} + \varepsilon_{kt}) \\ &= \exp(\beta_0 + \beta_{op} op_i + \beta_{cp} cp_i + \beta_m m_i + \beta_X X_i + \beta_Z Z_{it} + \gamma_t \lambda_k) \end{aligned} \quad (9)$$

where op_i is the own price travel cost to southeastern North Carolina beaches for individual i, cp_i represents cross price travel cost to the Outer Banks of North Carolina—a substitute beach recreation site for individual i, m_i is income, X_i includes individual characteristics and interactions, Z_{it} is a vector of indicators for various stated preference responses (including access and beach width scenarios), and ε_{kt} is a random effect. Both γ_t and λ_k are unobserved; we cannot separately identify them. It is only meaningful that they be multiplied in equation (3) and interpreted jointly in analyzing their effect on recreation demand.[8] For the MPLN model, recreation demand is specified according to the first line of equation (9), with the T-vector $\varepsilon_i \sim N(0,\Omega)$ replacing ε_{kt}, and the form of Ω given above.

For specification (3'), the conditional expectation of demand is given by:

$$\begin{aligned} y_{itk} &= \exp(\beta_0 + \beta_{op}op_i + \beta_{cp}cp_i + \beta_m m_i + \beta_X X_i + \beta_Z Z_{it}) \times \\ &\exp(\delta_{op}op_i\lambda_k + \delta_{cp}cp_i\lambda_k + \delta_m m_i\lambda_k + \delta_X X_i\lambda_k + \delta_Z Z_{it}\lambda_k + \gamma_t\lambda_k) \end{aligned} \quad (9')$$

This model is a discrete analog of a random parameters model, with the first line of equation (9') representing the conditional expectation of demand for type k = 1 (because λ_1=0 as indicated in (7) above) and the complete equation applying to types k = 1 - K. Thus, this form of the DFM model allows for different coefficient estimates by class through the interaction of the discrete factor.

5. Results and Discussion

We turn now to parameter estimates for our DFM-GNB model (equations 6-8). The unconditional sample likelihood function is programmed into FORTRAN and all parameters are obtained by the Davidson-Fletcher-Powell (DFP) optimization algorithm. Given a finite sample size, econometric theory does not provide the optimal number of points of support in DFM. In general, researchers add points of support until the likelihood function value fails to improve significantly, based on a likelihood ratio test [20, 27, 29, 31].

The likelihood ratio testing procedure, however, has shortcomings in this application. Specifically, under the null hypothesis that fewer support points represents the true model, the Hessian matrix for the alternative model (i.e., more points of support) is singular. As a result, the likelihood ratio test (LRT) statistic does not follow an asymptotic Chi-square distribution under the null hypothesis. In a Monte Carlo study, however, Mroz [29] demonstrates that this test performs fairly well when deciding between small numbers of support points (up to four). We first estimate the DFM-GNB models with three points of support and incrementally increase the number of supports to five, at which the likelihood function fails to improve significantly based on Akaike's Information Criterion (AIC) [3] and the Bayesian Information Criterion (BIC) [37]. A likelihood ratio test also favors the selection of four points of support. The results presented in this paper are based on four points of support for the discrete distribution.

Our econometric specification (9) includes age, marital status, interactions of age and marital status with own-price, cross-price, and income in the X vector, and indicators for stated preference treatments (*sp*) ($t = 2$, 3, or 4), the access treatment ($t = 3$), and the beach width improvement treatment ($t = 4$) in the Z vector. Note, the revealed preference treatment ($t = 1$) is the excluded category, and access and beach quality conditions (q_t) for treatments 1 and 2 are identical. Parameter estimates and robust standard errors for the DFM-GNB model are presented in column 1 of table 2.[9,10,11] Using the LRT we find joint significance for the model at conventional levels ($p < 0.0001$), and most of the parameter estimates are statistically significant at the 1% level, except for the *sp* and beach width treatment indicators, and the interactions of marital status with own-price and income. The LRT rejects the restriction that the p parameter in equation (6) is equal to 1 or 2 for a p-value less than 0.0001.[12] All of the parameters of the DFM-GNB model associated with the distribution of unobserved heterogeneity are statistically significant at conventional levels, except for $\overline{\theta}_3$. The $K-2$ estimates of $\overline{\lambda}$ represent parameters of the mass points of our heterogeneity distribution in equations (7), while the $K-1$ estimates of $\overline{\theta}$ correspond with parameters of the probability distribution for heterogeneity types in equations (8). The support points, λ_k, and probabilities, $\Pr(\lambda_k)$, are calculated from the parameter estimates.

Figure 1 displays the distribution of unobserved heterogeneity for the DFM-GNB specification. The lowest heterogeneity type ($\lambda_1 = 0$) has the highest probability (41 percent), and larger values of λ — $\lambda_2 = 0.325$, $\lambda_3 = 0.617$, $\lambda_4 = 1.00$—exhibit monotonically decreasing probabilities of 30 percent, 16 percent, and 13 percent, respectively. In the standard DFM model, the discrete factors enter the exponential of the demand equation as an additive term with a factor loading γ_t, where $t = 1, 2, 3, 4$ represents the demand treatment for our quasi-panel data. The factor loadings translate unobserved heterogeneity mass points (which are confined to the unit interval) into a recreation demand effect (that is unrestricted). We explored using four distinct parameters for the γ_t term in equation (4)—essentially allowing for the factor loadings to vary across all four scenarios—but the unrestricted model did not significantly improve the log-likelihood value.[13] Our final model in column 1 of table 2 includes only two γ terms, both of which are positive suggesting a positive correlation across RP and SP demand.

The distribution of unobserved heterogeneity and the factor loadings have a plausible and intuitive interpretation in this application that can provide potential insight into recreation demand panel data. Since the factor loadings are positive, conditional recreation demand is increasing in unobserved heterogeneity. As such, the classes of unobserved heterogeneity can be thought of as demand-intensity types, with type $\lambda_1 = 0$ representing low intensity demand, type $\lambda_4 = 1.00$ representing high intensity type, and the remaining types (λ_2 and λ_3) being intermediate.

Column 2 of table 2 displays results for a generalized negative binomial model without controls for unobserved heterogeneity (NUH-GNB). The log-likelihood value for this model, which restricts all parameters of the DFM model to zero, is much smaller than the DFM-GNB log-likelihood. The LRT does not support the restriction imposed by NUH-GNB (p-value < 0.0001).[14] The DFM-GNB is preferred according to information criteria (smallest AIC and BIC values).

We turn next to comparisons of the DFM-GNB model with parametric counterparts, specifically the multivariate Poisson-lognormal (MPLN) model [13] and a GNB specification with unobserved heterogeneity following a standard normal distribution (norm-GNB). The MPLN model is estimated using a maximum simulated likelihood procedure with 1000 Halton draws,[15] and the norm-GNB is estimated via quasi-maximum likelihood with unobserved heterogeneity integrated out by using Gaussian-Hermite quadrature. Parameter estimates for the MPLN model are presented in column three of table 2. All coefficients associated with this model are statistically significant at the 1% level for a type I error. The norm-GNB estimates are presented in the last column of table 2, and all coefficients associated with this model are statistically significant at the 1% level, with the exception of the beach width scenario indicator.

A practical complication in comparing DFM with parametric models is that since the models are based on different formulations of unobserved heterogeneity, they are not nested. In order to examine the performance of alternative non-nested models with sample data, we adopt AIC, BIC, and Vuong's non-nested test [42].[16] According to the AIC and BIC scores, the DFM-GNB specification outperforms the MPLN model.[17] The AIC and BIC are lower for the DFM-GNB. The first step of Vuong's test indicates rejection of the null hypothesis that the MPLN and DFM-GNB models cannot be distinguished at a significance level of 0.01 using a likelihood ratio variance test. The directional test in the second step indicates that the DFM-GNB model is preferred, but the test statistic is not statistically significant.[18] Hence, we conclude that, taking the number of parameters into account, the DFM-GNB specification provides a better fit to the data than the MPLN model, while the advantage of DFM-GNB is not statistically significant at conventional levels. In comparison with the norm-GNB model, the DFM-GNB outperforms in terms of information criteria and the Vuong test.[19] Thus, comparative evidence in support of our semi-parametric model is favorable.

Demand model parameters for each specification represent average effects across heterogeneity types, while the NUH-GNB model ignores unobserved heterogeneity. Elasticity estimates in table 3, exhibit a fairly wide range across specifications. The average DFM-GNB effect suggests price inelastic demand (ε_{op} = -0.54), while the models with standard normal heterogeneity exhibit more price responsiveness—inelastic for norm-GNB (ε_{op} = -0.82) and elastic for MPLN (ε_{op} = -1.65). The model that ignores heterogeneity indicates price elasticity close to unitary (ε_{op} = -1.01). A similar pattern is found in the cross-price elasticities, with relatively low responsiveness found in the DFM-GNB (ε_{cp} = 0.40) and more responsiveness in the other models (ε_{cp} = 0.74 for the NUH-GNB, ε_{cp} = 1.58 for the MPLN, and ε_{cp} = 1.03 for the norm-GNB). All estimations imply that beach visitation is a normal good when evaluated at the means of the data (ε_m

= 0.29 for DFM-GNB, ε_m = 0.60 for NUH-GNB, ε_m = 0.53 for MPLN, and ε_m = 0.36 for norm-GNB). Overall, the pattern of results shows less responsiveness in DFM parameter estimates, suggesting that more variation in recreation demand is attributed to unobserved heterogeneity as we move from a parametric specification to a less restrictive alternative.

The coefficient on the *sp* dummy variable is not different from zero in the DFM-GNB or NUH-GNB models. *Prima facie*, these results do not appear to support the existence of hypothetical bias in the data after conditioning on observable and unobservable characteristics. The coefficient on the *sp* dummy is positive and statistically significant in MPLN and norm-GNB models, consistent with the hypothetical bias heuristic of Huang, Haab, and Whitehead [22, 43]. The DFM and norm-GNB specifications also include a factor loading (γ) that varies across RP and SP demand. Since the factor loadings rescale unobserved heterogeneity, the γ_{sp} coefficient can be interpreted as an interaction term for SP treatments and unobserved heterogeneity. The RP and SP factor loadings are significantly different from one another, with $\gamma_{sp} > \gamma_{rp}$. Thus, some degree of hypothetical bias (that associated with higher intensity visitors) could manifest through this parameter. The *access* scenario (t = 3) indicator is statistically significant and positive in all models (with the exception of the NUH-GNB model), indicating that the average visitor would make more beach trips if parking and access conditions were improved. The *width* scenario (t = 4) indicator is not statistically significant in any of the models except MPLN, in which it is positive.

The parameter estimates of the discrete analog of a random parameters model given by specifications (3') and (9') are presented in table 4. Using the same criteria as above, four classes of heterogeneity is the preferred specification. The BIC favors the standard DFM model over the random parameters specification. The AIC, however, favors random parameters and a likelihood ratio test rejects the hypothesis that the interactions terms are jointly zero.[20]

The random parameters DFM specification can provide additional insight into the latent classifications revealed in the data. Distinct parameter estimates for classes (i.e. demand intensity types) are calculated as:

$$\beta_k = \beta + \delta\lambda_k, \text{ for } k = 1 - K, \qquad (10)$$

where β are the covariate parameters in the first column of table 4 and δ are the covariate interaction parameters in the second column of table 4 vectors. Since $\lambda_1 = 0$ for the low intensity demand type, the first column of table 4 applies. For this type, the own-price coefficient is about half the magnitude of the standard DFM-GNB specification, but the elasticity evaluated at the means and taking interactions into account is about the same (ε_{op} = -0.53). The cross-price elasticity for this type, on the other hand, is negative (ε_{cp} = -0.23) though not statistically significant (standard error = 0.59). The *sp* treatment effect is statistically insignificant, while the access treatment (t = 3) effect is positive and significant.

Only five of the fourteen interactions terms in the second column of table 4 are statistically significant, suggesting that only the effects of cross-price, income, the *sp* treatment, and interactions of income with marital status and age vary across heterogeneity types. The cross-price elasticity, for example, varies from ε_{cp} = -0.002 for λ_2 to ε_{cp} = 0.0007 for λ_4, though none of the estimates are statistically significant at conventional levels. While parsimonious in comparison with latent class models, the interaction of discrete factors with model covariates is a fairly restrictive way to produce class-specific parameter estimates because heterogeneity in covariate effects is reduced to a rescaling of the common discrete factor type. Despite the fact that the AIC and likelihood ratio test favor the random parameters specification, the lack of statistical significance in the majority of covariate parameter estimates and restrictive form of heterogeneity lead us to focus on the standard DFM model. Nonetheless, the random parameters DFM model could be worth pursuing in this and other applications, as alternate specifications (i.e. dropping some primary interaction terms or limiting the number of DFM interaction terms) could prove to fit the data well and provide for more insight into class-specific preferences.

We measure the economic value of recreation trips as the area under the compensated demand function that intersects the uncompensated demand function at the observed price and quantity—a measure of compensating variation [18]. Bockstael, Hanneman, and Strand [7] show this formula to be:

$$CV_{itk} = \frac{1}{\tilde{\beta}_m} \ln\left(1 + \frac{\mu_{itk}\tilde{\beta}_m}{\tilde{\beta}_{op}}\right), \tag{11}$$

where $\mu_{itk} = \exp(x_{it}\beta + \gamma_t \lambda_k)$ is the conditional mean for the standard DFM specification for heterogeneity type k, $\tilde{\beta}_m = (\beta_m + \beta_{m,age} \times age + \beta_{m,mar} \times married) \times (1000)^{-1}$ is the income coefficient (taking account of interactions in equation (3)) rescaled to be in the same units as price, and $\tilde{\beta}_{op} = (\beta_{op} + \beta_{op,age} \times age + \beta_{op,mar} \times married)$ is the own-price coefficient (accounting for interactions). Thus by equation (11), we employ duality to evaluate economic welfare for every individual in the sample under conditions t and conditional on heterogeneity type k. Measures of average conditional trip demand and (absolute value of) average annual compensating variation by heterogeneity type for relevant demand conditions are presented in table 5. Asymptotic standard errors for compensating variation are calculated using the delta method.

Estimation results by heterogeneity type exhibit systematic differences in the conditional mean of trips with lower values of λ associated with lower recreation demand. Likewise, estimates of compensating variation are smaller for lower values of λ as indicated by equation (11). For example, for the RP data predicted demand is 2.2, 5.5, 12.8, and 38.2 days per year for heterogeneity types 1 through 4 respectively, with compensating variation measures of $360, $914, $2109, and $6380. While these annual welfare estimates can be divided by number of trips to produce unit measures, the form of (11) produces roughly comparable per trip estimates across heterogeneity types, ranging from $164 to $167 per trip for the low-intensity demand and high-intensity demand, respectively. These estimates are similar to recent approximations (consumer surplus) of the value of beach recreation days—ranging from $38 to $274 per trip for an array of North Carolina beach sites [5] and $137 to $178 per trip for two Georgia beach sites [23]—and somewhat larger than what Whitehead et al. [44] found—roughly $90 per trip—using the same data. Trip estimates associated with SP data are slightly higher, ranging from 2.4 to 58 days per year, with corresponding welfare measures of $400 and $9759.

Turning next to the hypothetical improvement scenario, we find significantly higher demand when parking and access points are improved so that beachgoers would not have to spend an inordinate amount of time searching for a parking space and would not have to walk a great distance to access the beach.[21] Under this scenario conditional beach trips increase 46 to 99 percent, with the degree of amplified demand increasing monotonically with demand-intensity type. Compensating variation increases by roughly similar proportions, suggesting that improvements in access and parking could induce considerable economic benefits. It is conceivable, however, that hypothetical bias could be expressed through the SP factor loading (γ_{sp} - which is common to all SP demand equations). To explore this possibility we calculate conditional trip demand and economic welfare excluding the statistically insignificant *sp* parameter estimate and employing the factor loading associate with RP demand (γ_{rp}).[22] Under these conditions, we find that beach recreation demand increases a more moderate 31 to 32 percent, with similar increases in economic welfare. Thus, evidence does suggest that the SP data may exhibit hypothetical bias, but this bias can be excised using procedures similar to Whitehead et al. [43].

Our DFM heterogeneity types can be thought of as latent sub-populations, and welfare for the full population [41] can be summarized by employing our empirically estimated probability weights:

$$CV_{it} = \sum_{k=1}^{K} [\Pr(\lambda_k) CV_{itk}] \tag{12}$$

producing an average welfare measure analogous to the standard compensating variation measure. For our other specifications, individual welfare is calculated as:

$$CV_{it} = \frac{1}{\widetilde{\beta}_m} \ln\left(1 + \frac{\mu_{it}\widetilde{\beta}_m}{\widetilde{\beta}_{op}}\right), \tag{13}$$

where $\mu_{it} = \exp(x_{it}\beta + \varepsilon_i)$, $\varepsilon_i \sim MVN(0, \Omega)$ for the MPLN model; $\mu_{it} = \exp(x_{it}\beta + \gamma_t \varepsilon_i)$, $\varepsilon_i \sim N(0,1)$ for the norm-GNB model; and $\mu_{it} = \exp(x_{it}\beta)$ for the NUH-GNB model. Average compensating variation is the mean of (12) or (13) over the entire sample.

Conditional trip means and summary welfare measures for each of the four models under relevant demand conditions are presented in table 6.

Most models under-predict the number of trips, except for the MPLN which greatly over-predicts trip levels. For example, the DFM-GNB predicts 9.49 trips per year for the RP baseline compared to 10.93 for the NUH-GNB model, 62.04 for the MPLN model, and 7.90 for the norm-GNB specification—the observed level is 11.08.

The DFM-GNB model produces a summary compensating variation estimate of $1574 per visiting household per year, or about $166 per trip, for economic cost attributable to loss of access to beach sites in southeastern North Carolina. Compensating variation under SP demand is somewhat higher at $2202 per visiting household per year. Controlling for potential hypothetical bias (SP-access* estimates), improvements in access increase compensating variation by 32 percent, to $2079 per year. The standard errors for DFM-GNB are rather large in comparison with the other models.

The GNB specification without controls for unobserved heterogeneity produces mean conditional trip estimate very close to the sample mean. Welfare measures are lower than those produce by DFM-GNB. Compensating variation for loss of access for the RP data is $975 per visiting household per year, or about $89 per trip. This baseline is close to that found in Whitehead et al. [44]. Compensating variation for SP data is slightly higher at $1176 per visiting household per year, while economic welfare for improved access (excluding the *sp* coefficient) is $1275 per year. The $300 difference in welfare (31% increase) between RP and the hypothetical-bias adjusted access scenario is roughly equivalent to that found by Whitehead et al. [44].

The MPLN model produces extremely high trip estimates, ranging from 62.04 for RP to 99.15 for SP-access scenario. These estimates are approximately five times the size of the sample statistics, and much greater than estimates from other specifications. Compensating variation estimates are also somewhat large, ranging from $2407 for RP to $4015 for SP-access. We construe this as evidence that the MPLN model is mis-specified. The parameter estimates do not fit the data very well.[23] Standard errors, however, are relatively small (e.g. $45.50 for RP). Lastly, the norm-GNB results produce conditional trip estimates that are somewhat lower than the other models and welfare estimates that are similar to the NUH-GNB model.

Empirical evidence tends to favor the DFM-GNB specification over a specification that ignores unobserved heterogeneity (NUH-GNB) and models that impose multivariate normality on unobserved heterogeneity (MPLN and norm-GNB). Thus, results suggest the flexibility inherent in the semi-parametric characterization of unobserved heterogeneity has advantages in model-fitting. The asymptotic standard errors of variable transformations (elasticities and measures of compensating variation), however, tend to be rather large. The MPLN model produces the most precise parameter estimates, but these estimates will be biased if the normal distribution for unobserved heterogeneity is mis-specified or if the restrictions on the covariance matrix are invalid. Our results are consistent with mis-specification of the MPLN model in our application, though removal of outliers improves the fit of MPLN.

6. Conclusions

Information on revealed (RP) and stated preference (SP) is often gathered by those interested in valuing hypothetical changes in environmental quality or resource management regimes. Cameron [10] proposed combining these sources of data to improve model estimation and welfare calculations. Joint estimation allows one to glean more information about the underlying structure of preferences by imposing cross-equation restrictions on model parameters and assessing behavior associated with unobserved levels of exogenous factors in a way that was consistent with the observed levels.

Our results provide support to the DFM-GNB specification for modeling panel (and quasi-panel) count demand data. The method provides for a flexible characterization of count dispersion and incorporates unobserved heterogeneity while allowing for cross-equation correlations. Parameters associated with unobserved heterogeneity tend to exhibit statistical significance, and the model that ignores unobserved heterogeneity (NUH-GNB) is rejected at conventional significance levels. Each specification includes age and marital status (dummy variable for married) of the survey respondents, as well as interactions of age and marital status with own-price, cross-price, and income. Results suggest that both observable and unobservable heterogeneity are important aspects of recreation demand.

Information criteria and Vuong's non-nested likelihood ratio test support the DFM-GNB over a generalized negative binomial model in which unobserved heterogeneity follows a standard normal distribution (norm-GNB). The same criteria lend statistical support to the DFM-GNB over the multivariate Poisson log-normal (MPLN) model, but the second step of the Vuong test is not statistically significant. Each of the parametric heterogeneity models exhibits more responsiveness in terms of covariate elasticities (own-price, cross-price, and income), suggesting that more variation in recreation demand is attributed to unobserved heterogeneity as we move from a parametric specification to a less restrictive alternative. The MPLN model, however, provides poor fit to the data, with estimated trips roughly five times the sample statistics.

Welfare estimates derived from the DFM model are roughly 50% larger than those derived from the generalized negative binomial models that ignore unobserved heterogeneity (NUH-GNB) and impose standard normality (norm-GNB). Moreover, the confidence intervals associated with DFM are significantly larger. Welfare estimates for the MPLN model are much larger than any of the competing models, largely reflecting the apparently large bias in predicted trips. In particular, the MPLN results appear sensitive to outliers. The confidence intervals for MPLN, however, are relatively tight (an order of magnitude less than DFM). With joint normal unobserved heterogeneity, the MPLN estimator will be asymptotically efficient. If, however, normality does not hold then the MPLN estimates is biased and inconsistent. The DFM specification provides a flexible specification that should produce reliable, albeit potentially less efficient, estimates under general conditions. We recommend that both models be explored in empirical work.

Including latent segments with endogenous probabilities makes the DFM approach similar to the latent class model, but the standard form of DFM posits the class probabilities as fixed (rather than functions of covariates) and does not produce separate parameter estimates by class. We argue that, in comparison with the latent class model, the DFM framework is appropriate for count demand panel data, since: (i) the introduction of discrete factors explicitly incorporates cross-equation correlations; (ii) the specification is more straightforward as covariates do not appear in both the demand equation and class probabilities; and (iii) the specification is parsimonious, while allowing for variation across agents and across the panel dimension. These attributes of the model, however, may not always be advantageous given the data at hand, as *a priori* knowledge of information that affect class membership is not utilized, and the way in which unobserved heterogeneity is introduced may be viewed as restrictive.

Given our empirical estimates of beach recreation under current and improved conditions, we are able to interpret heterogeneity types as reflecting demand intensity. Since heterogeneity is approximated by a discrete distribution, we can use the support points to produce welfare estimates that vary by heterogeneity type, not unlike the results from latent class RUMs. This flexibility can allow policymakers and analysts to explore welfare effects within the user population.

References

[1] M. Alfò, G. Trovato, Semiparametric misture models for multivariate count data, with application, Econometrics J. 7 (2004) 426-454.

[2] W. Adamowicz, J. Louviere, M. Williams, Combining revealed and stated preference methods for valuing environmental amenities, J. Environ. Econ. Manage. 26 (1994) 271-292.

[3] H. Akaike, Information theory and the extension of the maximum likelihood principle, in: N. Petrov, Csaki (Eds.), 2nd International Symposium on Information Theory, Akademiai Kaido: Budapest (1973) 267-281.

[4] C.D. Azevedo, J.A. Herriges, C.L. Kling. Combining revealed and stated preferences: consistency tests and their interpretations, Amer. J. Ag. Econ. 85 (2003) 525-537.

[5] O.Bin, C.E. Landry, C. Ellis, H. Vogelsong, Some consumer surplus estimates for North Carolina beaches, Marine Res. Econ. 20 (2005) 145-161.

[6] D.M. Blau, A.P. Hagy, The demand for quality in child care, J. Pol. Econ. 106 (1998) 104–146.

[7] N.E. Bockstael, W.M. Hanemann, I.E. Strand, Measuring the benefits of water quality improvements using recreation demand models, Environ. Prot. Agency, Dept. of Policy, Planning and Evaluation (EPA-230-10-89-069), 1989.
[8] P.C.Boxall, W.L. Adamowicz, Understanding heterogeneous preferences in random utility models: A latent class approach, Environ. Res. Econ. 23 (2002) 421-446.
[9] T.T. Brown, R.M. Scheffler, S. Seo, M. Reed, The empirical relationship between community social capital and the demand for cigarettes, Health Econ. 15 (2006) 1159-1172.
[10] T.A. Cameron, Combining contingent valuation and travel cost data for the valuation of nonmarket goods, Land Econ. 68 (1992) 302-317.
[11] A.C. Cameron, P.K. Trivedi, Econometric models based on count data: comparisons and applications of some estimators and tests, J. App. Econometrics 1 (1986) 29-53.
[12] A.C. Cameron, P.K. Trivedi, Regression analysis of count data, Cambridge University Press: Cambridge, UK. 1998.
[13] K. Egan, J. Herriges, Multivariate count data regression models with individual panel data from an on-site sample, J. Environ. Econ. Manage. 52 (2006) 567-581.
[14] J. Englin, D. Lambert, Measuring angling quality in count data models of recreational fishing, Environ. Res. Econ. 6 (1995) 389-399.
[15] W. Greene, Econometric analysis, fifth ed. Prentice Hall, Upper Saddle River, New Jersey, 2003.
[16] T.C. Grijalva, R.P. Berrens, A.K. Boharra, W.D. Shaw, Testing the validity of contingent behavior trip responses, Amer. J. Ag. Econ. 84 (2002) 401-414.
[17] M.K. Haener, P.C. Boxall, W.L. Adamowicz, Modeling recreation site choice: do hypothetical choices reflect actual behavior?, Amer. J. Ag. Econ. 83 (2001) 629-642.
[18] J.A. Hausman, Exact consumer's surplus and deadweight loss, Amer. Econ. Rev. 71 (1981) 662-676.
[19] J.J. Heckman, B. Singer, A method for minimizing the impact of distributional assumptions in econometric models for duration data, Econometrica 52 (1984) 271-320.
[20] G.M. Holmes, Increasing physician supply in medically underserved areas, Labour Econ. 12 (2005) 697-725.
[21] W.Y. Hu, Child support, welfare dependency and women's labor supply, J. Human Resources 34 (1999) 71–103.
[22] J.C. Huang, T.C. Haab, J.C. Whitehead, Willingness to pay for quality improvements: should revealed and stated preference data be combined? J. Environ. Econ. Manage. 34 (1997) 240-255.
[23] C.E. Landry, K.E. McConnell, Hedonic onsite cost model of recreation demand, Land Econ. 83 (2007) 253-267.
[24] R.C. Layman, J.R. Boyce, K.R. Criddle, Economic valuation of the chinook salmon sport fishery of the Gulkana River, Alaska, under current and alternative management plans, Land Econ. 72 (1996) 113-128.
[25] J.B. Loomis, Panel estimators to combine revealed and stated preference dichotomous choice data, J. Ag. Res. Econ. 22 (1997) 233-245.
[26] B. Mcwilliams, Y. Tsur, E. Hochman, D. Zilberman, Count-data regression models of the time to adopt new technologies, Applied Economics Letters 5(1998) 369-373.

[27] H.N. Mocan, E. Tekin, Nonprofit sector and part-time work: An analysis of employer-employee matched data on child care workers” Rev. Econ. Stat. 85 (2003) 38-50.
[28] E. Morey, J. Thacher, W. Breffle, Using angler characteristics and attitudinal data to identify environmental preference classes: a latent-class model, Environ. Res. Econ. 34 (2006) 91–115.
[29] T.A. Mroz, Discrete factor approximation in simultaneous equation models: Estimating the impact of a dummy endogenous variable on a continuous outcome, J. Econometrics 92 (1999) 233–274.
[30] M. Niklitschek, J. León, Combining intended demand and yes/no responses in the estimation of contingent valuation models, J. Environ. Econ. Manage. 31 (1996) 387-402.
[31] A.G. Picone, F. Sloan, S. Chou, D.H. Taylor, Jr., Does higher hospital cost imply higher quality of care, Rev. Econ. Stat. 85 (2003) 51-62.
[32] B. Provencher, K.A. Baerenklau, R.C. Bishop, A finite mixture logit model of recreational angling with serially correlated random utility, Amer. J. Ag. Econ. 84 (2002) 1066-1075.
[33] B. Provencher, R. Bishop, Does accounting for preference heterogeneity improve the forecasting of a random utility model? A case study, J. Environ. Econ. Manage. 48 (2004) 793–810.
[34] R. Scarpa, A.G. Drucker, S. Anderson, N. Ferraes-Ehuan, V. Gómez, C.R. Risopatrón, O. Rubio-Leonel ,Valuing genetic resources in peasant economies: the case of ‘hairless’ creole pigs in Yucatan, Ecol. Econ. 45 (2003) 427- 443.
[35] R. Scarpa, M. Thiene, Destination choice models for rock climbing in the Northeastern Alps: a latent-class approach based on intensity of preference. Land Econ. 85 (2005) 426-444.
[36] R. Scarpa, M. Thiene, T. Tempesta, Latent class count models of total visitation demand: Days out hiking in the Eastern Alps, Environ. Res. Econ. 38 (2007) 447-460.
[37] G. Schwarz, Estimating the dimension of a model, Ann. Stat. 6 (1978) 461-464.
[38] J.S. Shonkwiler, W.D. Shaw, A finite mixture approach to analyzing income effects in random utility models: reservoir recreation along the Columbia river, in: N. Hanley, W. Shaw, R. Wright (Eds.) The new economics of outdoor recreation. Edward Elgar: UK, 2003.
[39] K. Train, Recreation demand models with taste differences over people. Land Econ. 74 (1998) 185–194.
[40] K. Train, Halton Sequences for Mixed Logit, working paper, University of California, Berkeley, 1999.
[41] R.H. von Haefen, D.J. Phaneuf, Estimating preferences for outdoor recreation: a comparison of continuous and count demand system frameworks, J. Environ. Econ. Manage. 45 (2003) 612-630.
[42] Q.H. Vuong, Likelihood ratio tests for model selection and non-nested hypotheses, Econometrica 57 (1989) 307-333.
[43] J.C. Whitehead, T.C. Haab, J.C. Huang, Measuring recreation benefits of quality improvements with revealed and stated behavior data, Resource Energy Econ. 22 (2000) 339-354.

[44] J.C. Whitehead, C.F. Dumas, J. Herstine, J. Hill, B. Buerger, Valuing beach access and width with revealed and stated preference data, working paper, Appalachian State University, 2006.
[45] A. Young, The gift of the dying: the tragedy of AIDS and the welfare of future African generations, NBER Working Paper No. 10991, 2004.

---

[1] See Blau and Hagy [6], Hu [21], and Mocan and Tekin [27] for applications of DFM.

[2] We show the details of this mechanism using a stylized model with bivariate random errors in the online Appendix. Note that this single random effect approach with different factor loadings for each equation could not capture the full generality of the underlying multivariate normal distribution, as this specification places restrictions on the structure of the variance-covariance matrix.

[3] The latent class approach resolves this problem by assuming that there are a finite number of person types. With several parameters that vary across type and several types, the curse of dimensionality arises again.

[4] Latent class modeling can be viewed as a finite mixture approach to random parameter models [e.g. 15].

[5] Morey, Thatcher, and Breffle [28] explore the use of Likert-scale attitudinal responses in estimating a latent class attitudinal model and how this approach could be used in conjunction with choice models.

[6] We note that our data are aggregated over sites, so our results also restrict coefficients across sites. This is due, however, to the nature of the data, not the DFM approach.

[7] The interested reader is referred to Whitehead et al. [44] for more information on the survey.

[8] Our dataset includes both visitors that made day trips to the beach and visitors that stayed overnight. Whitehead et al. [44] find no evidence of bias with these data in pooling different user types.

[9] Considering that the log-likelihood function might not be globally concave, we start the estimation with a variety of initial parameter values, which are generated by a grid search. Our estimation programs consistently converge to virtually the same parameter estimates except for *married*, *op*×*married* and *inc*×*married*. The parameter estimates presented in table 2 correspond with the highest log-likelihood value.

[10] We explored interactions of treatment variables (specifically *sp* (t = 2, 3, or 4), *access* (t = 3), and *width* (t = 4) with own-price, cross-price, and income to allow for structure change between treatments [44], but the parameter estimates were not statistically significant; a LRT supports restricting the parameters to zero ($\chi^2_{df=9} = 12.423$).

[11] The LRT supports pooling data across equations and estimating one set of coefficients for the primary demand parameters, rather than different estimates for each of the four equations ($\chi^2_{df=33} = 10.041$).

[12] The online appendix contains parameter estimates for DFM models with NB1 and NB2 variance specifications. For the restriction implied by NB1, $\chi^2_{df=1} = 1099.555$, and for the restriction implied by NB2, $\chi^2_{df=1} = 191.718$.

[13] For the restriction implied by the model with only two factor loadings, $\chi^2_{df=2} = 1.263$.

[14] For the restriction implied by NUH-GNB, $\chi^2_{df=7} = 3023.761$.

[15] Following Train [40], Halton sequences are used to numerically compute a multidimensional integral in the MPLN model. Thanks are due to Kevin Egan and Joe Herriges for providing the GAUSS code for this estimation procedure.

[16] Details of these steps can be found in the Appendix of Englin and Lambert [14].

[17] AIC $= 2\zeta_s - 2L_s$ and BIC = $-2L_s + \zeta_s ln(N)$, where L_s is the log-likelihood of model specification *s*, *N* is the sample size, and ζ_s is the number of parameters in specification *s*.

[18] The directional test sample-weighted t-statistic is 1.12.

[19] The first step null hypothesis for the Vuong test is rejected at a significance level of 0.01, and the second step favors DFM-GNB over norm-GNB with a sample-weighted t-statistic of 3.67.

[20] For the restriction implied by standard DFM-GNB, $\chi^2_{df=14} = 83.328$.

[21] The *width* ($t = 4$) scenario did not appear to shift recreation demand, thus trips and welfare measures for this scenario are not reported.

[22] If, on the other hand, the magnitude of SP parameters also reflects an expectation of higher income [43], our hypothetical bias-adjusted estimates will represent a lower bound.

[23] Note that the MPLN model in Egan and Herriges [13] over-predicts the mean visits by nearly 30% for a type of visitation in their data. Our results suggest that fit of MPLN parameters is hampered in the presence of extreme values in dependent variables. Exploring a subset of our data with outliers removed, the goodness-of-fit for MPLN estimates is considerably improved.

Table 1: Descriptive Statistics (N = 664)

Variable	Mean	Standard Deviation
Own-price (*op*) – travel cost to southeastern NC beaches	89.49	61.09
Cross-price (*cp*) – travel cost to Outer Banks, NC beaches	202.05	57.04
Income (*m*) – household income in thousands of US dollars	58.75	26.89
Age – age of respondent	43.82	15.14
Married – dummy variable for married respondent	0.68	0.47
Trips for $t = 1$ (y_1) – revealed preference trips	11.08	23.00
Trips for $t = 2$ (y_2) – stated preference trips under current conditions	13.05	24.84
Trips for $t = 3$ (y_3) – stated preference trips under improved access	17.10	30.50
Trips for $t = 4$ (y_4) – stated preference trips under improved beach width	14.17	26.40

Table 2: Beach Recreation Demand Models

Dependent variable: annual beach trips (y_{it})

	DFM-GNB		NUH-GNB		MPLN		norm-GNB	
Variable	parm	std err	parm	std err	parm	std err	parm	std err
own-price	-0.007	0.001	-0.016	0.001	-0.012	0.001	-0.002	0.001
cross-price	0.006	0.001	-0.005	0.001	-0.011	0.001	0.012	0.001
income	-0.006	0.001	0.007	0.001	0.005	0.002	-0.018	0.001
married	-0.841	0.076	-1.229	0.069	-2.110	0.140	-0.179	0.078
age	0.006	0.002	-0.038	0.002	-0.052	0.004	0.011	0.002
sp	0.104*	0.070	0.185*	0.154	0.363	0.035	0.180	0.060
access	0.274	0.050	0.266*	0.139	0.334	0.026	0.280	0.051
width	0.091*	0.065	0.097*	0.242	0.093	0.032	0.086*	0.064
op×married	0.0004*	0.0004	0.004	0.001	0.007	0.001	0.002	0.001
op×age	0.0001	0.0001	0.0001	0.0001	-0.0001	0.0001	-0.0002	0.0001
cp×married	0.003	0.001	0.003	0.001	0.008	0.001	-0.002	0.001
cp×age	-0.0001	0.0001	0.0001	0.0001	0.0003	0.0001	-0.0001	0.0001
inc×married	-0.001*	0.001	-0.004	0.001	-0.013	0.001	0.003	0.001
inc×age	0.0003	0.0001	0.0001	0.0001	0.0003	0.0001	0.0005	0.0001
constant	0.809	0.123	4.265	0.156	5.377	0.247	0.648	0.113
α_{rp}	0.016	0.003	0.878	0.098	-	-	0.033	0.005
α_{sp}	0.002	0.0007	0.632	0.056	-	-	0.005	0.001
p	3.211	0.069	2.204	0.033	-	-	3.002	0.055
γ_{rp}	2.856	0.135	-	-	-	-	-0.753	0.031
γ_{sp}	3.167	0.036	-	-	-	-	-0.816	0.010
$\bar{\lambda}_2$	0.730	0.029	-	-	-	-	-	-
$\bar{\lambda}_3$	-0.476	0.034	-	-	-	-	-	-
$\bar{\theta}_1$	1.166	0.149	-	-	-	-	-	-
$\bar{\theta}_2$	0.870	0.158	-	-	-	-	-	-
$\bar{\theta}_3$	0.213*	0.180	-	-	-	-	-	-
σ_1	-	-	-	-	1.365	0.025	-	-
σ_2	-	-	-	-	1.327	0.020	-	-
σ_3	-	-	-	-	1.289	0.017	-	-
σ_4	-	-	-	-	1.293	0.027	-	-
ρ_{rp}	-	-	-	-	0.815	0.005	-	-
ρ_{sp}	-	-	-	-	0.985	0.001	-	-
lnL	-7552.09		-9063.97		-7646.99		-7653.55	
LRT (df) [p-value]	39,578.19 (24) [<0.0001]		36,554.42 (17) [<0.0001]		62,420.8 (20) [<0.0001]		39,375.27 (19) [<0.0001]	
AIC	15154.19		18163.95		15335.99		15347.11	
BIC	15266.65		18244.92		15430.46		15437.07	

* - not statistically significant at 5% level for type I error. For each model, the number of unique individuals included in estimation is 664. Standard errors are robust.

Figure 1: Unobserved Heterogeneity

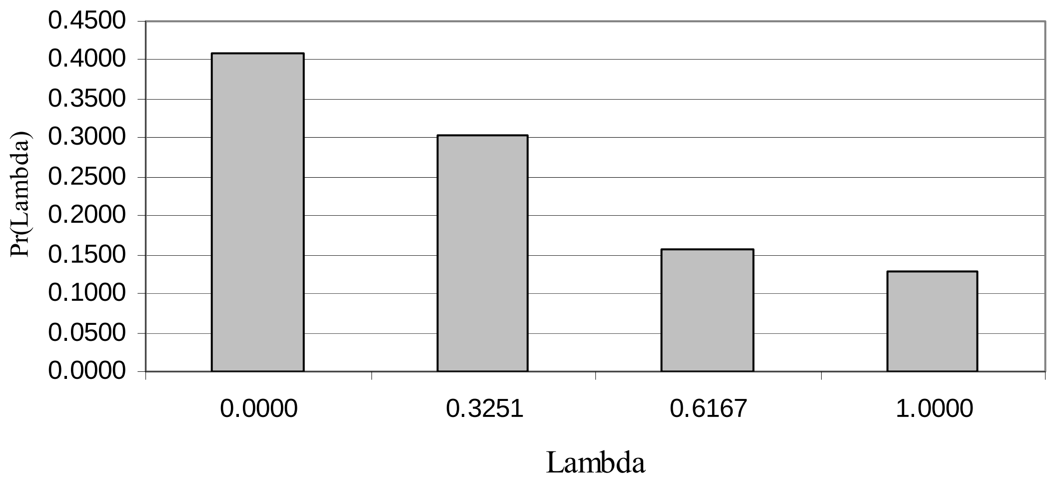

Table 3: Elasticity Estimates

Elasticity	DFM-GNB	NUH-GNB	MPLN	norm-GNB
own-price	-0.5443 (0.0183)	-1.0106 (0.0294)	-1.6523 (0.0244)	-0.8173 (0.3579)
cross-price	0.4027 (0.1905)	0.7388 (0.0414)	1.5825 (0.0054)	1.0312 (0.4337)
income	0.2938 (0.0357)	0.5973 (0.0294)	0.5292 (0.0001)	0.3591 (0.0322)

Asymptotic standard errors are calculated via the delta method and are indicated in parentheses.

Table 4: Random Parameters DFM-GNB Model

Dependent variable: annual beach trips (y_{it})

	Primary and DFM parameters ($\beta, \alpha, \rho, \gamma, \bar{\lambda}, \bar{\theta}$)		**Heterogeneity interaction** (δ)	
Variable	parm	std err	parm	std err
own-price	-0.0037	0.0010	0.0037*	0.0020
cross-price	-0.0041	0.0011	0.0049	0.0025
income	-0.0019*	0.0023	-0.0097	0.0048
married	-0.6586	0.1192	-0.0163*	0.2351
age	-0.0168	0.0032	0.0079*	0.0074
sp	0.0717*	0.0985	1.3838	0.1982
access	0.3244	0.0914	-0.1295*	0.2001
width	0.1224*	0.1252	-0.0729*	0.2681
op×married	0.0013*	0.0008	-0.0028*	0.0015
op×age	-0.0001	0.0000	-0.0001*	0.0000
cp×married	0.0024	0.0008	0.0027*	0.0015
cp×age	0.0001*	0.0001	0.0001*	0.0001
inc×married	0.0016*	0.0018	-0.0095	0.0036
inc×age	0.0002	0.0000	0.0002	0.0001
constant	2.2433	0.1739	-	-
α_{rp}	0.0187	0.0036	-	-
α_{sp}	0.0023	0.0007	-	-
p	3.1583	0.0661	-	-
γ_{rp}	2.2447	0.3821	-	-
γ_{sp}	1.2148	0.1983	-	-
$\bar{\lambda}_2$	0.7518	0.0316	-	-
$\bar{\lambda}_3$	-0.4389	0.0381	-	-
$\bar{\theta}_1$	1.5760	0.1646	-	-
$\bar{\theta}_2$	1.2220	0.1720	-	-
$\bar{\theta}_3$	0.6630	0.1913	-	-
lnL	-7510.43			
LRT (df) [p-value]	39,661.51 (37) [<0.0001]			
AIC	15098.86			
BIC	15274.30			
* - not statistically significant at 5% level for type I error; standard errors are robust.				

Table 5: Predicted Trips and Welfare by Treatment and Heterogeneity Type for Standard Discrete Factors Method-Generalized Negative Binomial

Treatment	λ_k	k	**Predicted Trips**	**Compensating Variation#**
RP	0.000	1	2.19	\$360.57 (126.55)
	0.325	2	5.55	\$913.94 (320.76)
	0.617	3	12.77	\$2109.00 (740.04)
	1.000	4	38.19	\$6379.64 (2236.60)
SP	0.000	1	2.43	\$400.34 (140.51)
	0.325	2	6.82	\$1123.15 (394.18)
	0.617	3	17.18	\$2841.79 (997.06)
	1.000	4	57.84	\$9759.12 (3418.00)
SP-access No correction for hypothetical bias	0.000	1	3.21	\$526.90 (166.55)
	0.325	2	8.98	\$1479.17 (467.40)
	0.617	3	22.61	\$3748.46 (1183.30)
	1.000	4	76.10	\$12,962.23 (4070.60)
SP-access* Corrected for hypothetical bias	0.000	1	2.89	\$474.55 (166.55)
	0.325	2	7.31	\$1203.42 (422.34)
	0.617	3	16.81	\$2779.99 (975.38)
	1.000	4	50.24	\$8443.79 (2958.60)

- Absolute values of compensating variation are presented. RP estimates employ the γ_{rp} factor loading; SP estimates employ the γ_{sp} factor loading, except for SP-access*, which are calculated using the γ_{rp} factor loading and excluding the *sp* coefficient. Asymptotic standard errors (in parentheses) are calculated using the delta method.

Table 6: Summary Predicted Trips and Compensating Variation for Demand Models

	DFM-GNB		NUH-GNB		MPLN		norm-GNB	
Treatment	**Predicted Trips**	**Compensating Variation#**	**Predicted Trips**	**Compensating Variation#**	**Predicted Trips**	**Compensating Variation#**	**Predicted Trips**	**Compensating Variation#**
RP	9.49	\$1574.19 (548.67)	10.93	\$975.08 (129.95)	62.04	\$2407.05 (45.50)	7.90	\$915.65 (172.69)
SP	13.18	\$2201.62 (763.39)	13.51	\$1175.86 (156.58)	77.48	\$2966.61 (65.50)	9.93	\$1151.86 (206.90)
SP-access	17.34	\$2914.28 (905.66)	17.16	\$1538.94 (204.60)	99.15	\$4014.95 (91.61)	13.14	\$1527.29 (274.04)
SP-access*	12.48	\$2078.73 (722.81)	14.26	\$1275.47 (169.77)	68.96	\$3053.41 (63.60)	10.45	\$1213.49 (228.70)

- Absolute values of compensating variation are presented. RP estimates employ the RP factor loading (γ_{rp}), while SP estimates employ the SP factor loading (γ_{sp}). The exception is SP-access* estimates, which are calculated using the RP factor loading (γ_{rp}) and excluding the *sp* coefficient in demand estimation to correct for potential hypothetical bias in the data. Asymptotic standard errors for compensating variation are calculated using the delta method and are displayed in parentheses.