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

February 2007

* Thanks are due to Dan Phanuef and Marty Smith for helpful comments and 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 revealed preference and stated preference and incorporates unobserved heterogeneity into the conditional expectation of recreation demand. Our model is a generalized negative binomial with random effects, in which the random effect is composed of a discrete representation of unobserved heterogeneity and a rescaling coefficient 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 more flexible manner. We find more predictive accuracy and smaller consumer surplus estimate with the DFM method in comparison with the standard random effects negative binomial. The structure of DFM allows for estimation of welfare by unobserved heterogeneity type.

JEL: C81, D12, Q51

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

Introduction

Resource economists are increasingly gathering revealed preference (RP) data in conjunction with information on stated preference (SP).[1] Adamowicz, Louviere, and Williams (1994) 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 (Niklitschek and León 1996; Huang, Haab, and Whitehead 1997; Loomis 1997; Whitehead et al. 2006), others have used SP data to examine behavioral changes stemming from variations in travel cost (Cameron 1992; Azevedo, Herriges, and Kling, 2003; Egan and Herriges 2006), in access to resources (Grijalva et al. 2002; Whitehead et al. 2006), and in management conditions (Layman, Boyce, and Criddle 1996). Through combining information on revealed and stated behavior, the analyst can 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 (Huang, Haab, and Whitehead 1997; Whitehead, Haab, and Huang 2000).

In this paper we estimate parameters of revealed and stated recreation demand using a semi-parametric technique—discrete factor method (DFM) (also known as discrete factor approximation)—to account for unobserved heterogeneity and allow for correlation across demand equations. This flexible and computationally simplistic approach has not, to our knowledge, been applied to recreation demand data. We condition the joint distribution of demand on a factor that represents unobserved heterogeneity, which is approximated by a step function, and integrate out over the distribution of the unobserved factor. Due to the use of a discrete distribution for

[1] For the travel cost model, one might also use the terms "observed" and "contingent behavior".

unobserved heterogeneity the estimator falls in the class of finite mixture distributions. One of the most attractive characteristics of this approximation is that a small number of discrete factors (e.g., three or four) has been shown to perform quite well under various distributional assumptions (Mroz 1999).[2]

Our econometric approach makes use of a generalized negative binomial model with individual heterogeneity modeled as discrete points, whose location and probability mass are estimated empirically. Model parameters are estimated via semi-parametric maximum likelihood. The data pertain to beach recreation demand in North Carolina. A random telephone survey provides estimates of current (revealed) demand for site trips. The survey also elicits stated demand under current conditions and hypothetical improvements in both access and beach quality.

Our DFM generalized negative binomial specification performs significantly better with the data than a standard random effects negative binomial model that is implemented in commercial statistical packages such as Stata and LIMDEP. While both models account for unobserved heterogeneity, the standard random effects negative binomial models imposed much more restrictive correlation structures on unobserved heterogeneity for a given individual. The DFM approach implemented in this paper relaxes the parametric form that the standard random effects negative binomial estimator imposes on the distribution of unobservables across individuals. For our North Carolina beach demand data, DFM estimates produce more accurate visitation predictions and smaller consumer surplus measures than the random effects negative binomial. We interpret the heterogeneity in our model as reflecting demand intensity, and illustrate how the structure of DFM allows for estimation of welfare by heterogeneity type.

[2] See Blau and Hagy (1998), Hu (1999), and Mocan and Tekin (2003) for applications of DFM.

Combining Stated and Revealed Preference Data

Cameron (1992) was the first, to our knowledge, to propose combining revealed and stated preference data to improve valuation estimates. She notes the potential for using RP and SP data to glean more information about the underlying structure of preferences. Combining these different kinds of data allows one to impose consistency across preference data, which could possibly attenuate hypothetical bias in stated preference data. In addition, the SP data can be used to "fill in" holes in the observed data by providing additional information about demand other than that exhibited under current market conditions. As resource economists are often interested in value stemming from alternate levels of environmental quality, a natural extension of the use of SP data is to market conditions associated with unobserved levels of public goods (Adamowicz, Louviere, and Williams 1994), and a number of researchers have made such an examination (Niklitschek and León 1996; Huang, Haab, and Whitehead 1997; Loomis 1997; Whitehead et al. 2006).

In this paper, 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. Collecting stated demand under current conditions is recommended in order to ensure that the underlying preference structure is comparable across revealed and stated behavior, and such data provides a heuristic in testing for hypothetical bias (Huang, Haab, and Whitehead 1997; Whitehead, Huang, and Haab 2000). Since we have multiple observations on each individual, we account for unobserved heterogeneity at the individual level in a way very 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 factor method (DFM)—to account for unobserved heterogeneity and permit correlation across RP and SP demand.

Our analysis of recreation demand is based to on the following negative binomial model (Cameron and Trivedi 1998):

$$\Pr(v_{it} \mid x_{it}) = \frac{\Gamma(v_{it} + \alpha_{it}^{-1})\alpha_{it}^{v_{it}} \lambda_{it}^{v_{it}} (1 + \alpha_{it}\lambda_{it})^{-(v_{it} + \alpha_{it}^{-1})}}{\Gamma(v_{it} + 1)\Gamma(\alpha_{it}^{-1})}, \qquad (1)$$

where i indexes recreators, t = 1, 2, …, T, represents revealed or stated demand under quality conditions q_t, v_{it} is recreation demand for recreator i under conditions t, $\lambda_{it} = \exp(x_{it}'\beta)$ is individual i's expected recreation demand for case t, with x_{it} being a matrix of covariates and β denoting a vector of unknown parameters. The variance for an individual's demand under conditions t is given by $\mathrm{Var}[v_{it} \mid x_{it}] = \lambda_{it} + \alpha_{it}\lambda_{it}^2$, where $\alpha_{it} = \alpha_0/\lambda_{it}$ and α_0 is a parameter to be estimated. With a panel specification, the effects of covariates x_{it} can be decomposed into within-group effects, which relate to different conditions of demand for the same individual, and between-group effects, which relate to a cross-section of different individuals. Note that the within-group dispersion (variance divided by the mean)[3] of v_{it} for individual i is $1+\alpha_0$ (i.e. constant within-group dispersion).

In the standard random effects negative binomial model the restriction on α_0 being constant is relaxed and it is allowed to vary randomly across groups; specifically, $1/(1+\alpha_i)$ is assumed to follow a parametric form. $1/(1+\alpha_i) \sim Beta(r, s)$ is one of the most commonly used distributions (Liang and Zeger 1986; Hausman and Griliches 1984; Stata

[3] This is also known as the "variance-mean ratio" (Winkelmann 1995; Mullahy 1997).

Manual 2005). The joint probability of observing a given set of trips demanded for individual i is given by:

$$L_{i|\alpha_i} = \int_{\alpha_i} \prod_{t=1}^{T} \Pr(v_{it} \mid x_{it}, \alpha_i) f(\alpha_i) d\alpha_i . \tag{2}$$

Note that this specification of the negative binomial model does not relax the restriction on the constant within-group dispersion. Moreover, given that the random effects only apply to the distribution of the dispersion parameter α_i, instead of expected demand λ_{it}, the within-group and between-group variations of λ_{it} are attributed only to observables, x_{it}. In other words, unobserved heterogeneity—such as idiosyncratic preferences or unobserved portions of budget constraints, each of which could affect demand intensity—has no impact on the central tendency of expected demand for recreation activities.

In this paper, we propose an alternative approach to introducing random effects into a negative binomial model for panel data. We use the generalized version of the negative binomial model in equation (1):

$$\Pr(v_{itk} \mid x_{it}) = \frac{\Gamma(v_{it} + \alpha_{it}^{-1}\lambda_{it}^{2-p})\alpha_{it}^{v_{it}}\lambda_{it}^{(pv_{it}-2v_{it})}(1+\alpha_{it}\lambda_{it}^{p-1})^{-(v_{it}+\alpha_{it}^{-1}\lambda_{it}^{2-p})}}{\Gamma(v_{it}+1)\Gamma(\alpha_{it}^{-1}\lambda_{it}^{2-p})}, \tag{3}$$

where p is an additional parameter to be estimated.[4] The conditional variance for v_{it} is now $\text{Var}[v_{it} \mid x_{it}] = \lambda_{it} + \alpha_{it}\lambda_{it}^{p}$ and the within-group dispersion in this generalized model becomes $\alpha_{it}\lambda_{it}^{p-1} = \alpha_0\lambda_{it}^{p-2}$, which in general varies across individuals i and conditions t as long as the maximum likelihood estimate of p is not (approximately) 2. Secondly, we allow an unobserved demand intensity component to enter expected recreation demand. We introduce a semi-parametric random effects component (μ_i) to each demand

[4] Note that given p=2, this model reduces to what many consider a "standard" negative binomial, which is referred to as NB2 by Cameron and Trivedi (1998).

equation, and utilize a discrete factors method (DFM) to approximate the distribution of unobserved heterogeneity (μ_i), which governs both RP and SP demand under conditions q_t. Conditional on an unobserved heterogeneity type,

$$\lambda_{it}(\beta \mid x_{it}, \mu_i) = \exp(x_{it}\beta + \rho_{it}) = \exp(x_{it}\beta + \gamma_t \mu_i) = \exp(x_{it}\beta)\exp(\gamma_t \mu_i), \tag{4}$$

where γ_t is a condition-specific loading factor, allowing for $\text{cov}[\rho_{ij}, \rho_{im}] \neq 0$ for any two demands $j \neq m$, and μ_i takes K discrete values. The loading factors rescale unobserved heterogeneity into a random recreation demand effect that can vary across demand types (revealed and stated preference and quality conditions q_t).

Discrete Factors Method and Likelihood Function

The discrete factors method (DFM) is proposed by Heckman and Singer (1984) 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 (μ_i), while maintaining the asymptotic efficiency of maximum likelihood estimators (Mroz 1999). The distribution of the heterogeneity type is approximated with a step function and integrated out through a weighted sum of step levels (Heckman and Singer 1984), where the weights are given by empirically estimated probabilities. Mroz (1999) 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 DFM performs better than normality based estimators when the underlying distribution is non-normal. Second, DFM is computationally simplistic. For instance, the MPLN model adopted by Egan and Herriges (2006) requires evaluating multivariate lognormal integrals based on the assumption that the random effects components across equations follow a multivariate lognormal distribution. Although simulation methods are typically adopted to make these evaluations feasible, the computations are somewhat cumbersome. Using DFM, the likelihood function conditional on unobserved heterogeneity (μ_i) can be constructed as follows.

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

$$\begin{aligned} L_i &= \sum_{k=1}^{K} \{\Pr(\mu_k)\Pr[v_{i1}, v_{i2}, v_{i3}, v_{i4} \mid x_{it}, \mu_k]\} \\ &= \sum_{k=1}^{K} \left\{ \Pr(\mu_k) \prod_t \Pr(v_{it} \mid x_{it}, \mu_k) \right\}, \end{aligned} \tag{5}$$

where $\Pr(\mu_k)$ is the probability of individual *i* having heterogeneity level *k*. Since cross-equation correlation is captured by the random effect term $\rho_{kt} = \gamma_t \mu_k$, the joint probability of vector v_i conditional on μ_k and observables x_i is equal to the product of their univariate densities conditional on μ_k and x_i, as indicated in equation (5).

Given the mixture of generalized negative binomial model (equation 3) and semi-parametric random effects (equation 4), our parameterization of the mean and dispersion parameter allows for systematic variation by heterogeneity type and across RP and SP demand equations. Note that the generalized negative binomial nests other specifications of the variance, including Cameron and Trivedi's NB1 and NB2 (1998). We further specify $\alpha_{itk} = \dfrac{\alpha_{0t}}{\lambda_{itk}}$, where α_{0t} is a condition-specific parameter, so that within-group dispersion could vary by quality condition q_t across individuals even in the case of NB2.

Using this specification for the distribution of demand, individual *i*'s contribution to the likelihood function, L_i , can be rewritten as:

$$\sum_{k=1}^{K}\left\{\Pr(\mu_k)\prod_t \frac{\Gamma(v_{itk}+\alpha_{itk}^{-1}\lambda_{itk}^{2-p})\alpha_{itk}^{v_{itk}}\lambda_{itk}^{(pv_{itk}-2v_{itk})}(1+\alpha_{itk}\lambda_{itk}^{p-1})^{-(v_{itk}+\alpha_{itk}^{-1}\lambda_{itk}^{2-p})}}{\Gamma(v_{itk}+1)\Gamma(\alpha_{itk}^{-1}\lambda_{itk}^{2-p})}\right\}. \qquad (6)$$

The sample likelihood function is derived as the product of (6) over all *N* individuals:

$$L=\prod_{i=1}^{N}\left(\sum_{k=1}^{K}\left\{\Pr(\mu_k)\prod_t \frac{\Gamma(v_{itk}+\alpha_{itk}^{-1}\lambda_{itk}^{2-p})\alpha_{itk}^{v_{itk}}\lambda_{itk}^{(pv_{itk}-2v_{itk})}(1+\alpha_{itk}\lambda_{itk}^{p-1})^{-(v_{itk}+\alpha_{itk}^{-1}\lambda_{itk}^{2-p})}}{\Gamma(v_{itk}+1)\Gamma(\alpha_{itk}^{-1}\lambda_{itk}^{2-p})}\right\}\right).$$

(7)

We refer to this model as the DFM generalized negative binomial (DFM-GNB) specification. The semi-parametric specification of the unobserved heterogeneity term μ allows unrestricted correlations across RP and SP demand equations for the same individual *i*. This approach should be advantageous in controlling for unobserved heterogeneity in data sets with information on revealed and stated preference and in relaxing the constant within-group dispersion restriction.

Take, for instance, our data set of recreational beach visitation, which includes both past visits and stated visits under current conditions as well as under hypothetical improvements in access and beach quality. *Ceteris paribus,* an individual who has frequented the beach in the past may be more likely to visit in the future, and may be more likely to visit conditional on improvements in access or beach quality, creating a positive correlation between RP and SP visitation. Our specification imposes a semi-parametric restriction that all heterogeneity and correlation among the individual demand equations enters the full model through the coefficients γ_t and the factor μ_i, which is approximated by a discrete distribution. In particular, $\Pr(\mu_k) > 0, \forall k = 1,...,K,$ where K is the total number of points support. Without loss of generality, μ is confined to the unit interval:

$$\begin{aligned}
\mu_1 &= 0 \\
\mu_2 &= \frac{1}{1+\exp(\overline{\mu_2})} \\
\mu_3 &= \frac{1}{1+\exp(\overline{\mu_3})} \\
&\cdots \\
\mu_{K-1} &= \frac{1}{1+\exp(\overline{\mu_{K-1}})} \\
\mu_K &= 1.
\end{aligned} \tag{8}$$

The transformed probability weights are given as follows:

$$\begin{aligned}
\Pr(\mu_k) &= \frac{\exp \overline{\theta_k}}{1+\sum_{k'=1}^{K-1} \exp \overline{\theta_{k'}}}, k = 1,...,K-1 \\
\Pr(\mu_k) &= \frac{1}{1+\sum_{k'=1}^{K-1} \exp \overline{\theta_{k'}}}, k = K.
\end{aligned} \tag{9}$$

Note that $\overline{\mu}$'s and $\overline{\theta}$'s are parameters to be estimated, along with α's, β's, and γ's. The support points, μ_k, and the transformed probabilities, $\Pr(\mu_k)$, can be calculated from the parameter estimates.

One practical complication in model specification is the comparison of the semi-parametric DFM-GNB model with that of the standard parametric random effects count data model characterized in equation (2). The two models are based on different formulations of the heterogeneity component and are therefore not nested. In order to examine the performance of alternative non-nested models with sample data, we adopt “rho-bar squared” ($\bar{\rho}^2$) (Horowitz 1980; Ben-Akiva and Lerman 1985) and Akaike’s Information Criterion (AIC) (Akaike 1973), which are two widely applied statistics for comparing mutually non-nested model specifications. Rho-bar squared is defined as $\bar{\rho}_s^2 = 1 - (L_s - \zeta_s)/L(0)$, where L_s is the log-likelihood of model specification s and ζ_s is the number of parameters to be estimated in specification s, and $L(0)$ is the log-likelihood function value of the null model, which for our application is a Poisson model containing a constant term only (Cameron and Trivedi 1998). The rho-bar squared test is analogous to the likelihood ratio test for non-nested models and the model with a higher $\bar{\rho}^2$ score implies a better fit to the data. To test statistical significance of the difference in $\bar{\rho}^2$ scores between two non-nested specifications s_1 and s_2, one can use the following property to calculate a p-value (Horowitz 1980; Ben-Akiva and Lerman 1985):

$$\Pr(\bar{\rho}_{s2}^2 - \bar{\rho}_{s1}^2 > \upsilon) \leq \Phi\left\{-\left[-2\upsilon L(0) + (\zeta_{s2} - \zeta_{s1})\right]^{1/2}\right\} \leq \text{, for any } \upsilon \geq 0, \quad (10)$$

where $\Phi(\cdot)$ indicates the normal cumulative distribution function.

Similarly, AIC can also be used to compare the performance of different models after accounting for the degree of freedom and sample size (N): $AIC = [L_s - \zeta_s]/N$. If the AIC value for a model is higher than that of others, then the former model specification is construed as more informative.

Beach Recreation Demand, Access, and Erosion

Empirical estimates suggest that recreation days at coastal beaches are of considerable economic value (Silberman and Klock 1988; Bell and Leeworthy 1990; Bin et al. 2005; Dumas et al. 2005; Landry and McConnell 2007). Rising incomes and employment have increased demand for coastal recreation and leisure, leading to widespread tourism and development on the coastal margin (Cordes and Yezer 1998). At the same time, natural forces—primarily coastal storms and sea level rise—threaten to erode beach sand, reducing the amount of space available for leisure and recreation activities. Beach replenishment—adding sand to the beach face—is a common tool to combat erosion; it is, nonetheless, expensive and controversial. In order for local communities to receive federal assistance for beach replenishment, public beach access must be maintained at specified levels (Dumas et al. 2006).

The survey data utilized herein, originally analyzed in Whitehead et al. (2006), record information on recreation demand that can be used to assess benefits of visitation for beach sites in southeastern North Carolina, and how benefits would change with improvements in access and beach width. In addition to revealed preference visitation behavior, the survey data include stated preference responses, describing intended visitation in the subsequent year assuming that conditions remain the same, as well as how beachgoers would change their visitation behavior in the subsequent year in response to hypothetical improvement in available parking spaces and beach width. The data are used to estimate recreation demand using the DFM-GNB model.

The study area encompasses seventeen beaches in five southeastern North Carolina counties. 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. The interested reader is referred to Whitehead et al. (2006) for more information on the survey.

Descriptive statistics are included in table 1. The average number of observed trips is 11, while planned trips under current conditions is slightly higher at 13. Respondents were offered a hypothetical scenario for improved parking. The scenario was as follows:

> Suppose that parking facilities and beach access at southeastern North Carolina oceanfront beaches were improved so that you would not have to spend time searching for a parking space or access area, the parking space and access area would be located within reasonable walking distance of the oceanfront beach, and parking was free or reasonably priced. Also suppose that the number of beach users at the oceanfront beaches does not change.

The average number of contingent trips under the scenario of improved parking was 17. Respondents were also offered a scenario on improved beach width. The scenario was as follows:

> Suppose a beach nourishment policy is implemented for all southeastern North Carolina oceanfront beaches. Beach nourishment would be performed in each county periodically, at least once every 3 to 5 years, for the 50-year life of the project. Periodic nourishment is done to maintain an increased beach width to provide shore protection and recreation benefit. The goal would be to make the average beach width increase by 100 feet.

The average number of contingent trips under the scenario of improved beach width was 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 southeaster 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:

$$v_{it} = \exp(\beta_0 + \beta_1 op_i + \beta_2 cp_i + \beta_3 m_i + \beta Z_i + \rho_{it}) \quad (11)$$

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, North Carolina—a substitute beach recreation site for individual i, m_i is income, Z is a vector of indicators for stated preference responses and access and beach width scenario indicators and relevant interactions, and ρ_{it} is a random effect. The random effect can be decomposed: $\rho_{it} = \gamma_t \mu_i$, where γ_t are loading factors, with $\text{cov}[\gamma_j, \gamma_m] \neq 0$ for demands $j \neq m$, and μ_i represents unobserved heterogeneity common to both RP and SP demand. Both γ_t and μ_i are unobserved; we cannot identify them individually. A step function with K partitions is used to approximate the distribution of μ_i, while γ_t are parameters to be estimated.[5]

Results

We turn now to parameter estimates for our DFM-GNB model (7-9). The parameters are estimated using a quasi-maximum likelihood method with discrete factors to account for unobserved agent heterogeneity. The random effects estimator with discrete factor approximation imposes a semi-parametric restriction that all heterogeneity and correlation among the random effects from individual equations enters the full model through the discrete factor μ.[6] The unconditional sample likelihood function (equation 7)

---

[5] Our dataset includes both visitors that made day trips to the beach and visitors that stayed overnight. Many empirical applications of recreation demand focus on a single user type or estimate separate equations across user types due to possible differences in onsite costs incurred or differences in preferences. Whitehead et al. (2006) find no evidence of bias with these data in pooling different user types. Thus, we ignore differences in user type.

[6] 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 (Mroz 1999; Picone et al. 2003; Mocan and Tekin 2003; Holmes 2005). The results presented in this paper are based on four points of support. A specification with

is programmed into FORTRAN and all parameters are obtained by maximizing the likelihood function using Davidson-Fletcher-Powell (DFP) optimization algorithm.

The parameter estimates for two specifications of the DFM-GNB model are presented in table 2. The first specification includes own-price (*op*), cross-price (*cp*), income (*m*), an indicator for stated preference (*sp*) (t = 2, 3, or 4), an indicator for the access treatment (t = 3), and an indicator for the beach width improvement treatment (t = 4). The second specification includes the same covariates as the first, but also interactions for *sp* with prices and income. We conduct a series of likelihood ratio tests to explore specifications. We find joint significance for each model at conventional levels ($p < 0.0001$). The generalized negative binomial is preferred over the NB2 specification (in which the p term in equations (3, 6, and 7) is restricted to be 2) ($p < 0.0001$). We explored using four separate parameters for the γ term in equation (4)—essentially allowing for the rescaling coefficient to vary across all four scenarios—but the unrestricted model did not significantly improve the log-likelihood. Our final model includes only two γ terms, corresponding with RP and SP demand. In comparing the displayed DFM-GNB specifications in table 1, the likelihood ratio test supports the restricted model in column 1 (p=0.1519).

Most of the demand parameter estimates of the DFM-GNB model are statistically significant at the 5% level, with the exception of the stated preference (*sp*) and *width* treatment variables. All of the parameters associated with the distribution of unobserved heterogeneity are statistically significant at the 5% level. The estimates of $\overline{\mu}$ represent parameters of the mass points of our heterogeneity distribution in equations (8), while

five points of support did not improve the likelihood function value significantly or affect our main findings.

estimates of $\bar{\theta}$ correspond with parameters of the probability distribution for heterogeneity types in equations (9). The support points, μ_k, and the transformed probabilities, $\Pr(\mu_k)$, are calculated from the parameter estimates.

Figure 1 displays the distribution of unobserved heterogeneity for the preferred DFM-GNB specification (column 1 of table 2). Unobserved heterogeneity enters the exponential of the demand equation as an additive term with a rescaling coefficient γ_t, where $t = 1, 2, 3, 4$, but $\gamma_2 = \gamma_3 = \gamma_4$ in our restricted model. Thus, γ_t varies only across RP and SP. The rescaling coefficients translate unobserved heterogeneity mass points (which are confined to the unit interval) into a recreation demand effect. Each rescaling coefficient is positive and statistically significant at conventional levels. Elasticities are calculated in the conventional manner and are displayed in the first column of table 3.

For comparison purposes, we estimate a generalized negative binomial model that does not account for unobserved heterogeneity ("GNB (no hetero)") as well as a standard random effects negative binomial ("RE-NB") with *Beta* specification. For the GNB model without heterogeneity, we restrict all DFM parameters to be zero, essentially assuming that there is only one type of unobserved heterogeneity. Parameter estimates for the generalized negative binomial model without heterogeneity are presented in the column 3 of Table 2. All coefficients are statistically significant at the 5% level, with the exception of the binary treatment indicators—*sp*, *access*, and *width*. The log-likelihood value for the GNB model with a single type of heterogeneity is much smaller than the log-likelihood associated with the DFM-GNB model. A likelihood ratio test supports the DFM-GNB specification at $p < 0.0001$ significance level. We present elasticity estimates for the generalized negative binomial model without heterogeneity in the second column of table 3.

Parameter estimates for the RE-NB model are presented in the column 4 of Table 2. All coefficients are statistically significant at the 5% level, with the exception of *cross price*. According to the AIC and $\overline{\rho}^2$ scores, the DFM-GNB specification outperforms the standard RE-NB model. In particular, we note from Table 2 that among the three random effects specifications, the DFM-GNB in column 1 has the largest $\overline{\rho}^2$ value (=0.7564) and the difference between it and the $\overline{\rho}^2$ value for the standard RE-NB is 0.0037. The remaining issue is whether this difference is statistically significant. Using the property in equation (10), we obtain

$$\Pr(\overline{\rho}_{s2}{}^2 - \overline{\rho}_{s1}{}^2 > 0.0037) \leq \Phi(-15.438).$$

The probability that this difference could have occurred by chance is less than 0.0001. Hence, we conclude that, taking the number of parameters into account, the DFM-GNB specification provides the best fit to the data. Elasticity estimates for the RE-NB model are presented in the last column of table 3.

Recall that the conditional mean for the DFM specification is:

$$E(v_{it} \mid x_{it}, \mu_i) = \exp(x_{it}\beta + \gamma_t \mu_i) = \exp(x_{it}\beta)\exp(\gamma_t \mu_i)$$

We assume that x and μ are independent, and we can thus express the conditional mean as product of expectations:

$$E(v_{it} \mid x_{it}) = E[\exp(\gamma_t \mu_i)]E[\exp(x_{it}\beta)]$$

Substituting our discrete representation for unobserved heterogeneity and integrating, we have:

$$E(v_{it} \mid x_{it}) = \sum_{k=1}^{K} [\Pr(\mu_k) \exp(\gamma_t \mu_k)] \exp(x_{it}\beta), \tag{12}$$

where the first term represents the expected value of μ, which is independent of x. Annual consumer surplus for beach access under conditions t is the integral of equation (12) over travel cost (own-price—*op*) from the average price to infinity:

$$\begin{aligned} CS_t &= \int_{op^0}^{\infty} \sum_{k=1}^{K} [\Pr(\mu_k) \exp(\gamma_t \mu_k)] \exp(\tilde{\beta}_t + \beta_1 C) dC \\ &= \sum_{k=1}^{K} [\Pr(\mu_k) \exp(\gamma_t \mu_k)] \int_{op^0}^{\infty} \exp(\tilde{\beta}_t + \beta_1 C) dC \\ &= \sum_{k=1}^{K} [\Pr(\mu_k) \exp(\gamma_t \mu_k)] \frac{\exp(\tilde{\beta}_t + \beta_1 op^0)}{-\beta_1} \end{aligned} \tag{13}$$

where $\tilde{\beta}_t$ is the linear combination of coefficients and variable means—other than travel cost—in equation (11), and $\bar{\lambda}_t = \exp(\tilde{\beta}_t + \beta_1 op^0)$ is the conditional mean of beach trips under conditions t in the absence of unobserved heterogeneity. The summation term in the last line of equation (13) is the expected demand effect associated with unobserved heterogeneity. For the RE-NB model, consumer surplus is calculated in the conventional manner:

$$CS_t = -\frac{\bar{\lambda}_t}{\beta_1}, \tag{14}$$

which is equivalent to the last term on the last line of equation (13). The conditional mean of beach trips and average annual consumer surplus measures for each model under current conditions RP and SP and improved conditions SP are presented in table 4. To aid in interpretation, we divide CS_t by the actual number of trips to produce per-trip measures. We use the Krinsky-Robb (1985) procedure to produce confidence intervals for consumer surplus estimates.

Discussion

Statistical evidence supports the DFM-GNB specification over the parametric random effects negative binomial (RE-NB). The $\bar{\rho}^2$ test statistic that compares non-nested specifications for our data indicates that the DFM-GNB performs better than the RE-NB model, and the difference is statistically significant at $p < 0.0001$. Parameters of the DFM-GNB model are more precisely estimated. The demand parameters for the DFM-GNB specification represent *weighted* average[7] effects across heterogeneity types, in which the weights are endogenously determined by the data as shown in equations (9) and unobserved heterogeneity enters the exponential of conditional expectation of demand as an additive random term with non-zero mean as in equation (4).

The standard RE-NB specification places equal weight on each observation in estimating parameters and incorporates unobserved heterogeneity as an additive term in the exponential of expected demand with zero mean. As should be expected, the estimation procedures produce different elasticity estimates, with DFM-GNB indicating significantly more own-price responsiveness (ε_{op} = -0.50 versus ε_{op} = -0.25) and less responsiveness to income (ε_m = 0.08 versus ε_m = 0.29). The cross-price elasticity from the RE-NB model is not different from zero. The coefficient on *sp* is not different from zero in any of the DFM models. *Prima facie*, our results do not appear to support the existence of hypothetical bias in the data (Huang, Haab, and Whitehead 1997; Whitehead, Huang, and Haab 2000) after conditioning on observable and unobservable characteristics.

[7] The "weighted" nature of the parameters is due to the quasi-maximum likelihood estimation procedure.

A likelihood ratio test rejects the degenerate specification of the DFM model with only one type of unobserved heterogeneity ($p < 0.0001$). All of the parameters of the DFM-GNB model associated with the distribution of unobserved heterogeneity are statistically significant at conventional levels. Identification of these parameters derives from two sources. The first is restrictions on the covariance between the disturbance terms in RP and SP equations. Specifically, we only use one discrete factor (μ) for all four demand equations. We also use functional form restrictions on the distributions of the discrete factor as another identification source; both the location and probability weight for each value of μ are given a logistic form, which is well-behaved, facilitating optimization procedures.

Figure 1 displays the empirical estimates of unobserved heterogeneity and the associated probabilities. Equations (8) and (9) provide the basis for these estimates. The lowest heterogeneity type ($\mu_1 = 0$) has the highest probability, and the larger values of μ—$\mu_2 = 0.3272$, $\mu_3 = 0.6073$, $\mu_4 = 1.000$—exhibit monotonically decreasing probabilities. Unobserved heterogeneity affects recreation demand in the DFM-GNB model as an additive term with a rescaling coefficient in the exponential of equation (4). The rescaling coefficients differ across RP and SP, with $\gamma_{rp} = 2.897$ and $\gamma_{sp} = 3.199$. While we test for differences in γ_t across all four demand equations, statistical evidence supports the specification with only two rescaling coefficients.

The DFM-GNB model produces predicted trip levels that are closer to the observed levels. For example, 8.57 trips per year for the RP baseline versus 6.42 trips per year from the RE-NB model—the observed level is 11.08. This pattern holds for all conditions of demand ($t = 1 - 4$).

The last line of Equations (13) provides the formula for consumer surplus for the DFM-GNB specification, while welfare estimation for the RE-NB specification uses the conventional formulation in equation (14). Consider first the DFM-GNB estimates. Annual consumer surplus for RP demand is $1521.35 (95% C.I. $1499 - $1560). Consumer surplus measures for SP demand are: $2076.84 (95% C.I. $1966 - $2199) for current conditions; $2738.03 (95% C.I. $2258 - $2896) for improved beach access; and $2279.83 (95% C.I. $2157 - $2412) for increased beach width. Since neither the *sp* nor *width* treatment variables are statistically significant in the DFM-GNB estimates, we also calculate welfare measures that exclude the influence of these parameters in the estimation of expected demand (numerator of the last term in equation (13)). Under these conditions, we find annual SP consumer surplus measures of $1871.17 (95% C.I. $1770 - $1979) for current conditions and increased beach width (since the width effect is not statistically significant) and $2466.88 (95% C.I. $2337 - $2612) for improved beach access. Lastly, as it is conceivable that hypothetical bias could be expressed through the SP rescaling parameter (γ_{sp} - which is common to all SP demand equations) (Huang, Haab, and Whitehead 1997; Whitehead, Huang, and Haab 2000) we calculate SP welfare using the rescaling coefficient associate with RP demand. Under these conditions, we find no difference in welfare associated with RP demand, the SP baseline, or increased beach width. Annual consumer surplus for improved access is $2005.69 (95% C.I. $1902 - $2125).

DFM-GNB estimates suggest that SP consumer surplus under current conditions exceeds RP by 36% if the *sp* treatment variable is included in estimation of conditional demand and by 23% if it is not. If we do not include the *sp* treatment variable and use the

RP rescaling coefficient (γ_{rp}) in place of the SP rescaling coefficient we obtain the same consumer surplus estimates under current beach conditions for RP and SP demand. Consumer surplus for improved access (increased beach width) exceeds RP by 80% (50%) if insignificant treatment variables (*sp* and *width* where appropriate) are included in estimation of conditional demand, and by 62% (23%) if they are not. The most conservative estimates are produced by excluding insignificant treatment variables and using the RP rescaling coefficient in estimation of SP demand. In this case, consumer surplus for improved access increases 32% over current conditions and increased beach width has no effect on welfare.

Due to the smaller absolute value on parameter estimate of own-price, the RE-NB welfare estimates are considerably larger than those derived from the DFM-GNB model. However, one should bear in mind that the consumer surplus estimates from DFM-GNB represent a weighted average[8] across unobserved heterogeneity types, as indicated in equations (13). Moreover, the RE-NB model imposes constant within-group dispersion, which is fairly restrictive. This is a major practical difference between the specifications. Thus, one might expect differences in welfare estimates. The enhanced efficiency of the DFM-GNB estimator produces tighter confidence intervals on parameter transformations. This is true for elasticity and consumer surplus estimates.

The distribution of unobserved heterogeneity and the rescaling coefficients have an intuitive interpretation that can assist in understanding the differences between consumer surplus estimates. Since the rescaling coefficients 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 $\mu_1 = 0$

[8] Again, the weighted nature is due to quasi-maximum likelihood estimation.

representing low intensity demand, type $\mu_4 = 1.000$ representing high intensity type, and the remaining types (μ_2 and μ_3) being intermediate. Moreover, given the structure of equations (11) we can produce welfare estimates for each class of heterogeneity and demand treatment by:

$$CS_{tk} = -\exp(\gamma_t \mu_k)\exp(\bar{x}_t \beta) / \beta_1 \quad (15)$$

These estimates are displayed in table 5. Insignificant treatment effects (*sp* and *width*) are not used in estimating conditional demand for the welfare measures in table 5. The welfare estimates by heterogeneity type exhibit systematic differences in the conditional mean of trips with lower values of μ associated with lower recreational demand. Likewise, consumer surplus estimates are smaller for lower values of μ. Given the probability estimates in figure 1, these welfare measures can be used to assess the impact of environmental quality changes on different user segments.

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 (1992) was the first to propose combining observed and contingent behavior 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.

In this paper we estimate RP and SP beach recreation demand simultaneously using a semi-parametric technique—discrete factor approximation (also known as discrete factor method (DFM))—to permit correlation across equations and to account for unobserved heterogeneity. This flexible and computationally simplistic approach has not, to our knowledge, been applied to recreation demand data. DFM uses a discrete distribution to take account of unobserved heterogeneity, with distribution parameters and probabilities estimated empirically from the data. We show how the DFM likelihood function is derived and how the specification is used to produce welfare measures. We compare our results against those derived from a generalized negative binomial model without heterogeneity control and from a standard random effects negative binomial model.

The degenerate DFM model that does not incorporate unobserved heterogeneity is outperformed by the DFM-GNB model (likelihood ratio test statistic, $p < 0.0001$). The DFM-GNB parameter estimates differ from those attained by a standard random effects negative binomial model that assumes a parametric form for the dispersion term. We interpret these differences as reflecting the more flexible form for incorporation of heterogeneity among and within individual recreators and the less restrictive specification of dispersion in the DFM model. 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. This flexibility allows policymakers and analysts to explore welfare effects within the user population.

References

Adamowicz, W., J. Louviere, and M. Williams. 1994. "Combining Revealed and Stated Preference Methods for Valuing Environmental Amenities" *Journal of Environmental Economics and Management* 26: 271-92.

Akaike, H. 1973. "Information Theory and the Extension of the Maximum Likelihood Principle" Petrov, N., and Csaki (eds.), 2nd International Symposium on Information Theory, Akademiai Kaido Budapest, 267-281.

Azevedo, C.D., J.A. Herriges, and C.L. Kling. 2003. "Combining Revealed and Stated Preferences: Consistency Tests and Their Interpretations" *American Journal of Agricultural Economics* 85(3): 525-37.

Bell, F.W. and V.R. Leeworthy. 1990. "Recreational Demand by Tourists for Saltwater Beach Days" *Journal of Environmental Economics and Management* 18(3): 189-205.

Ben-Akiva, M. and Lerman, S. 1985. *Discrete Choice Analysis* Cambridge, MIT University Press.

Bin, O., C.E. Landry, C. Ellis, and H. Vogelsong. 2005. "Some Consumer Surplus Estimates for North Carolina Beaches" *Marine Resource Economics* 20(2): 145-61.

Blau, D.M. and D.B. Gilleskie. 2001. "Retiree Health Insurance and the Labor Force Behavior of Older Men in the 1990s" *Review of Economics and Statistics* 83(1):64-80.

Blau, D.M. and A.P. Hagy. 1998. "The Demand for Quality in Child Care" *Journal of Political Economy* 106: 104–46.

Cameron, T.A. 1992. "Combining Contingent Valuation and Travel Cost Data for the Valuation of Nonmarket Goods" *Land Economics* 68(3): 302-17.

Cameron, A.C. and P.K. Trivedi. 1998. *Regression Analysis of Count Data* Cambridge University Press: Cambridge, UK.

Cordes, J.J. and A.M.J. Yezer. 1998. "In Harm's Way: Does Federal Spending on Beach Enhancement and Protection Induce Excessive Development in Coastal Areas?" *Land Economics* 74(1): 128-45.

Dumas, C.F., J.C. Whitehead, J.H. Herstine, R.B. Buerger, and J.M. Hill. 2005. "Estimating Peak Demand for Beach Parking Spaces under Capacity Constraints" working paper University of North Carolina-Wilmington.

Egan, K. and J. Herriges. 2006. "Multivariate Count Data Regression Models with Individual Panel Data from an On-site Sample" *Journal of Environmental Economics and Management* 52(2): 567-81.

Grijalva, T.C., R.P. Berrens, A.K. Boharra, and W.D. Shaw. 2002. "Testing the Validity of Contingent Behavior Trip Responses" *American Journal of Agricultural Economics* 84(2): 401-14.

Hausman, J., B.H. Hall, and Z. Griliches. 1984. "Econometric Models for Count Data with an Application to the Patents-R & D Relationship" *Econometrica* 52: 909-938.

Heckman, J.J. and B. Singer. 1984. "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data" *Econometrica* 52: 271-320.

Holmes, G.M. 2005. "Increasing Physician Supply in Medically Underserved Areas" *Labour Economics* 12: 697-725.

Horowitz, J. 1980. "The Accuracy of the multinomial Logit Model as an Approximation to the Multinomial Probit Model of Travel Deman" *Transportation Research* 14B: 331-341.

Hu, W.Y. 1999. "Child Support, Welfare Dependency and Women's Labor Supply" *Journal of Human Resources* 34: 71–103.

Huang, J.C., T.C. Haab, and J.C. Whitehead. 1997. "Willingness to Pay for Quality Improvements: Should Revealed and Stated Preference Data Be Combined?" *Journal of Environmental Economics and Management* 34: 240-55.

Landry, C.E. and K.E. McConnell. 2007. "Hedonic Onsite Cost Model of Recreation Demand" forthcoming *Land Economics* 83(2).

Layman, R.C., J.R. Boyce, and K.R. Criddle. 1996. "Economic Valuation of the Chinook Salmon Sport Fishery of the Gulkana River, Alaska, under Current and Alternative Management Plans" *Land Economics* 72(1): 113-28.

Liang, K.Y. and S.L. Zeger. 1986. "Longitudinal data analysis using generalized linear models" *Biometrika* 73: 13-22.

Mocan, H.N. and E. Tekin. 2003. "Nonprofit Sector and Part-Time Work: An analysis of Employer-Employee Matched Data on Child Care Workers" *Review of Economics and Statistics* 85: 38-50.

Mroz, T.A. 1999. "Discrete Factor Approximation in Simultaneous Equation Models: Estimating the Impact of a Dummy Endogenous Variable on a Continuous Outcome" *Journal of Econometrics* 92: 233–74.

Mullahy, J. 1997. "Heterogeneity, Excess Zeros, and the Structure of Count Data Models" *Journal of Applied Econometrics* 12: 337-50.

Niklitschek, M. and J. León. 1996. "Combining Intended Demand and Yes / No Responses in the Estimation of Contingent Valuation Models" *Journal of Environmental Economics and Management* 31: 387-402.

Picone, A.G., F. Sloan, S. Chou, and D.H. Taylor, Jr. 2003. "Does Higher Hospital Cost Imply Higher Quality of Care" *Review of Economics and Statistics*, 85(1): 51-62.

Silberman, J., and M. Klock. 1988. "The Recreational Benefits of Beach Renourishment" *Oceanand Shoreline Management* 11: 73-90.

Stata Longitudinal/Panel Data Reference Manual. 2005. Stata Press.

Whitehead, J.C., T.C. Haab, and J.C. Huang. 2000. "Measuring Recreation Benefits of Quality Improvements with Revealed and Stated Behavior Data" *Resource and Energy Economics* 22: 339-54.

Whitehead, J.C., C.F. Dumas, J. Herstine, J. Hill, and B. Buerger. 2006. "Valuing Beach Access and Width with Revealed and Stated Preference Data" working paper Appalachian State University: Boone, NC.

Winkelmann, R. 1995. "Duration Dependence and Dispersion in Count-Data Models" *Journal of Business and Economic Statistics* 13(4): 467-74.

Table 1: Descriptive Statistics

Variable	Mean	Standard Deviation
own-price (*op*) – travel cost to southeastern NC beaches	89.489	61.094
cross-price (*cp*) – travel cost to Outer Banks, NC beaches	202.051	57.046
income (*m*) – household income in thousands of US dollars	58.75	26.897
trips for t=1 (v_1) – revealed preference trips	11.082	23.004
trips for t=2 (v_2) – stated preference trips under current conditions	13.049	24.838
trips for t=3 (v_3) – stated preference trips under improved access	17.099	30.504
trips for t=4 (v_4) – stated preference trips under improved beach width	14.173	26.396

Table 2: Beach Recreation Demand Models

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

Variable	DFM-GNB		GNB (no hetero)	RE-NB
own-price	-0.00564 (0.00019)	-0.00536 (0.00062)	-0.01096 (0.00023)	-0.00274 (0.00069)
cross-price	0.00129 (0.00020)	0.00093* (0.00064)	0.00467 (0.00017)	-0.00048* (0.00083)
income	0.00137 (0.00046)	0.00335 (0.00119)	0.00495 (0.00039)	0.00486 (0.00173)
sp	0.10428* (0.06803)	0.18676* (0.13514)	0.20865* (0.16293)	0.23755 (0.03024)
access	0.27639 (0.04790)	0.27723 (0.04724)	0.24057* (0.14936)	0.28098 (0.02652)
width	0.09325* (0.06156)	0.09351* (0.06101)	0.07866* (0.28206)	0.09512 (0.02751)
$op \times sp$	-	-0.00049* (0.00070)	-	-
$cp \times sp$	-	0.00046* (0.00072)	-	-
$income \times sp$	-	-0.00231 (0.00128)	-	-
constant	0.93733 (0.04943)	0.86979 (0.10913)	1.93820 (0.07847)	1.91666 (0.14553)
α_{0rp}	0.01962 (0.00355)	0.01789 (0.00338)	1.24385 (0.14374)	-
α_{0sp}	0.00255 (0.00072)	0.00226 (0.00066)	0.94258 (0.08797)	-
p	4.16307 (0.06431)	4.19674 (0.06620)	3.07842 (0.03489)	-
γ_{rp}	2.89731 (0.13117)	2.87113 (0.13468)	-	-
γ_{sp}	3.19901 (0.03336)	3.19488 (0.03407)	-	-
$\bar{\mu}_2$	-0.43612 (0.02993)	0.72957 (0.02914)	-	-
$\bar{\mu}_3$	0.72091 (0.02848)	-0.41942 (0.02980)	-	-
$\bar{\theta}_1$	1.51159 (0.15093)	1.48100 0.14926	-	-
$\bar{\theta}_2$	0.47509 (0.18435)	0.99329 (0.16134)	-	-
$\bar{\theta}_3$	1.02043 (0.16300)	0.46072 (0.18230)	-	-
r	-	-	-	2.25156 (0.13958)
s	-	-	-	1.75035 (0.11847)
Log Likelihood	-7567.658	-7565.014	-9196.440	-7690.427
LRT (df)	47116.66(16)	47121.95(19)	43859.09(9)	46871.12 (8)
$\bar{\rho}^2$	0.7564	0.7563	0.7043	0.7527
AIC	-10.373	-11.422	-13.863	-11.594
# unique persons	664	664	664	664
* - not statistically significant at 5% level; Standard errors are provided in parentheses.				

Figure 1: Unobserved Heterogeneity

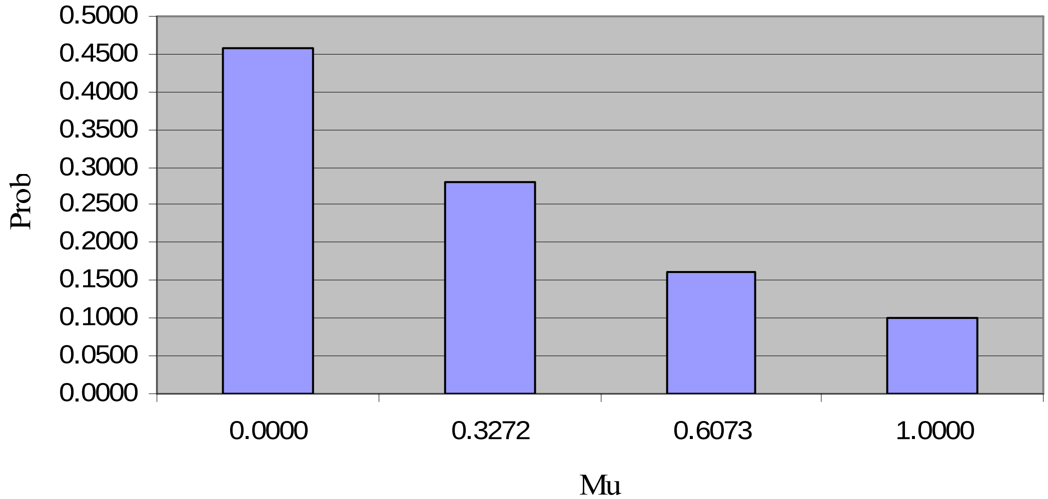

Table 3: Elasticity Estimates

Elasticity	DFM-GNB *	GNB (no hetero)	RE-NB
own-price	-0.5043 (-0.4776 – -0.5320)	-0.9807 (-0.9471 – -1.0154)	-0.2452 (-0.1435 – -0.3447)
cross-price	0.2615 (0.3218 – 0.4545)	0.9436 (0.8069 – 1.0014)	0.0969 (-0.1821 – 0.3768)
income	0.0807 (0.0364 – 0.1265)	0.2908 (0.0797 – 0.3146)	0.2855 (0.1239 – 0.4503)

* - We present elasticity estimates for the restricted DFM-GNB model (column one in table 2), since this is the preferred specification.
95% Confidence Intervals are indicated in parentheses (Krinsky and Robb 1986).

Table 4: Predicted Trips and Consumer Surplus

	DFM-GNB			GNB (no hetero)			RE-NB		
t	**Predicted Trips**	**CS**	**CS/trip**	**Predicted Trips**	**CS**	**CS/trip**	**Predicted Trips**	**CS**	**CS/trip**
RP	8.57	$1521.35 ($1499 - $1560)	$137.30	8.95	$816.72 ($788 - $845)	$73.71	6.42	$2344.42 ($1656 - $3899)	$211.59
SP – baseline	11.70	$2076.84 ($1966 - $2199)	$159.15	11.02	$1006.20 ($972 - $1042)	$77.11	8.14	$2973.05 ($2100 - $5095)	$227.83
SP – access	15.43	$2738.03 ($2589 - $2896)	$160.12	14.02	$1279.87 ($1234 - $1325)	$74.85	10.78	$3937.60 ($2783 - $6623)	$230.28
SP - width	12.84	$2279.83 ($2157 - $2412)	$160.85	11.93	$1088.55 ($1050 - $1127)	$76.80	8.95	$3269.74 ($2313 - $5544)	$230.70
SP – baseline/ width*	10.54	$1871.17 ($1770 - $1979)	$143.39	-	-	-	-	-	-
SP – access*	13.90	$2466.88 ($2337 - $2612)	$144.27	-	-	-	-	-	-
SP – access**	11.30	$2005.69 ($1902 - $2125)	$117.29	-	-	-	-	-	-
SP – width**	8.57	$1521.35 ($1499 - $1560)	$107.34	-	-	-	-	-	-
	95% Confidence Intervals are provided in parentheses * SP excluding insignificant parameters in calculation of conditional demand; ** SP excluding insignificant parameters and using RP rescaling coefficient								

Table 5: Consumer Surplus Measures by Heterogeneity Type

k	μ_k	**Treatment***	**Trips**	**Consumer Surplus**
1	0.000	RP	2.17	$385.22
1	0.000	SP	2.86	$507.87
1	0.000	SP - access	2.17	$385.22
2	0.3272	RP	5.60	$994.07
2	0.3272	SP	6.18	$1,097.21
2	0.3272	SP - access	8.15	$1,446.52
3	0.6073	RP	12.61	$2,238.29
3	0.6073	SP	15.15	$2,688.42
3	0.6073	SP - access	19.97	$3,544.30
4	1.000	RP	39.35	$6,982.36
4	1.000	SP	53.20	$9,441.32
4	1.000	SP - access	70.14	$12,447.06

* SP – width treatment does not differ from SP;
RP estimates are calculated using the RP rescaling coefficient (γ_{rp});
SP estimates are calculated using the SP rescaling coefficient (γ_{sp})