A Latent Class Analysis of Public Attitudes Towards Water Resources: Implications for Recreational Demand

Oren Ehrlich, Xiang Bi*, Tatiana Borisova, and Sherry Larkin

Department of Food and Resource Economics,

University Florida

*Corresponding author

P.O. Box 110240

University of Florida

Gainesville, FL 32611-0240

Phone: 352-294-7671

xiangbi@ufl.edu

Selected Paper prepared for presentation at the Southern Agricultural Economics Association's 2016 Annual Meeting, San Antonio, Texas, February, 6-9 2016

Abstract

The recent developments of nonmarket valuation have focused on identifying preference heterogeneity and examining the impact it has on consumer's willingness to pay. The objective of this study is to examine the extent to which heterogeneous environmental attitudes influence demand for freshwater recreational activities as well as the valuation of freshwater recreational benefits. We focus on the St. Johns River, the longest river in Florida, and use a telephone survey of Florida's residents to elicit information in regards to household outdoor recreational experiences on the river. Information regarding respondent attitudes and perceptions towards Florida's water resources and natural resource policies was gathered in the survey as well. We employed a latent class analysis to reveal two distinct classes of respondents based on their responses to questions regarding their environmental attitudes and perceptions. We then estimated a recreational demand model with respect to travel costs associated with getting to the river, household income, perceived water quality of the river, and respondents' environmental attitudes within each latent class. We found that class 1's individual recreational benefits are twice as large as class 2's. We contribute to the literature by emphasizing that environmental attitudes directly influence consumer recreational demand and valuation of the river, and should be taken into consideration for water resource management policies.

1. Introduction

Recent developments in nonmarket valuation have focused on identifying preference heterogeneity and examining its impact on willingness to pay (Boxall and Adamowicz, 2002). Recent literature further emphasizes the latent nature of preference heterogeneity and incorporates latent classes of stated attitudes in estimating recreational demand using discrete choice modeling (Morey, Thacher and Breffle, 2006). In contrast, revealed preference methods such as the travel cost method (TCM) incorporate individual or household demographic characteristics to approximate the impact of preference heterogeneity.

Demographic variables are either used as covariates in single-site models or to identify latent consumer classes (Scarpa, Thiene, and Tempesta, 2007; Bujosa, Riera, and Hicks, 2010). Although demographic characteristics may be correlated with unobserved heterogeneous preferences, using demographic characteristics to approximate preference heterogeneity is less likely to reveal direct relationships between consumption frequencies and consumers' stated perception/attitudes. This is because demographics do not explain how said heterogeneous preferences influence consumption patterns of a particular nonmarket good (Boxall and Adamowicz, 2002). Furthermore, identifying latent classes using demographic covariates may overlook individual preference heterogeneity, i.e. the specific choices of the individual consumer's demographic characteristics and the sufficient variations within selected demographic characteristics.

As opposed to being explained by their demographic characteristics, consumer preferences are more likely to be revealed through their stated perceptions and attitudes about the particular product under examination (Boxall and Adamowicz, 2002). Morey et al. (2006) further note that expressions of attitudes provide insight and information regarding underlying latent preferences in addition to revealed-preference data. Specifically, they suggest that individuals have well-defined preferences, those preferences are latent, and said latent preferences determine the respondents' choices (Morey, Thacher, and Breffle, 2006).

In order to account for these latent preferences, the Latent Class Analysis (LCA) can be used where heterogeneity is examined using a discrete distribution (Aitkin and Rubin, 1985). In this type of analysis a population is split up into groups based on an estimated probability of a particular person belonging to a particular group. The estimated probability is determined by survey responses to questions regarding attitudes and perceptions. LCA is often estimated jointly with discrete site-choice model, following Boxall and Adamowicz (2002). They combined a LCA (that was estimated using only attitudinal data) with a latent-class site choice model (that was estimated using only choice data). Under the single-site TCM framework, Scarpa et al. (2007) combined the LCA with a general count model to examine consumer preference heterogeneity. Though a joint estimation like in Scarpa et al. (2007) is more efficient, the formation of latent classes (i.e., the number of latent classes identified) and the estimates from the TCM may hinge upon the choices of covariates in determining the latent class membership (Vermunt, 2010).

The objective of this paper is to examine the extent to which heterogeneous environmental attitudes influence demand for freshwater recreational activities and valuation of freshwater recreational benefits. This study differs from the previous literature in that we incorporate the LCA with a single-site TCM by estimating the TCM conditional on the latent class formation. Though the two-step approach may not be as efficient as a joint estimation, we demonstrate that this approach is more efficient than estimating a TCM that ignores consumers' distinct preferences. Additionally, consumers' conditional class membership probabilities can be used for market segmentation.

We focus on the longest river in Florida, the St. Johns River, and use a telephone survey of a random sample of 500 residents to elicit information on household outdoor recreational experiences on the river. Additional information about respondent attitudes and perceptions about Florida's water resources and natural resource policy are also collected.

We use a latent class analysis to reveal two distinct classes based on their environmental attitudes and perceptions. Although these two classes share similar demographic characteristics, one class demands greater environmental protection (i.e., the environmentally concerned class) and represents 52% of the sample. Households in this same class also recreate outdoors more frequently thus are more likely to benefit from improved environmental quality. We then estimate a recreational demand model with respect to travel costs to the river, household income, perceived water quality of the river, and respondent's environmental attitudes for each latent class.

We find that the two classes have statistically different willingness to pay estimates for improved water quality. Particularly, the average willingness to pay from the more-pro-environment class is $83 per household per trip and the willingness to pay from the other class is $40 per household per trip.

We contribute to the literature in the following ways. First, our results emphasize that environmental attitudes directly influence consumers' recreational demand and valuation of the river thus should be taken into consideration for managing water resources. Second, environmental attitudes can be used to examine the potential distributional impacts of proposed environmental policy by estimating the shares of distinct classes in the population.

2. Empirical Model

2.1 The Latent Class Model

A latent class analysis (LCA) is "a statistical method used to identify a set of discrete, mutually exclusive latent classes of individuals based on their responses to a set of observed categorical variables" (Lanza, Collins, Lemmon, & Schafer, 2007). This type of analysis is used to reveal underlying (or latent) classes based on multiple variables that are characterized by a pattern of conditional probabilities. In this study attitudinal questions from the telephone survey were used as the explanatory variables to define the latent classes. Specifically, responses to questions regarding home water quality and quantity, perceptions of the laws and regulations regarding the quality of Florida waterways, and perceptions of the amount of government spending in different areas were used in the analysis. Based on the probabilities of responses to these questions, respondents were grouped to best characterize the classes determined by their responses.

Assume the sample population is composed of a number of different preference groups denoted *C,* and an individual's preference group is latent, or unobserved. What is observed is the individual i's responses to attitudinal questions and the observed characteristics of the individual z_i as a set (x_i, z_i). The latent class model has been extensively used in economic literature, and the explanation of the model that is most relevant to this study is presented in Morey, Thacher, and Breffle (2006). Morey, Thacher, and Breffle (2006) build on foundational studies such as those by McCutheon (1987) and pioneer environmental economics studies such as Boxall and Adamowicz (2002) and Provencher et al. (2002). Following Morey, Thacher, and Breffle (2006), the model includes the following four probabilities:

$$Pr(c: z_i),\ \ Pr(c: z_i|x_i),\ \ \pi_{qs|c},\ \ Pr(x_i: z_i) \quad (1)$$

$Pr(c: z_i)$ is the unconditional probability that individual *i* belongs to group *C* based on the observable characteristics *z*. This probability is unconditional because it does not rely on the specific answer to the attitudinal questions. Respondents with the same observable characteristics *z* belong to group *c* because of the unconditional membership probabilities.

$Pr(c: z_i|x_i)$ is the conditional membership probability that individual *i* belongs to group *c* based on the observable characteristics z and is conditional on the individual's answers to attitudinal questions. This allows for a more accurate prediction of the respondent's group membership.

$\pi_{qs|c}$ is the probability that an individual in group *c* answers level *s* to attitudinal question q. This is a function of an individual's preferences.

$Pr(x_i: z_i)$ is the probability that an individual with characteristics z_i has the response pattern x_i. These are functions of the $\pi_{qs|c}$ response probability.

If x_{iqs} represents an individual *i*'s answer to attitudinal question q at level *s*, then x_{iqs}=1 otherwise x_{iqs}=0. The unobservable characteristics of which the latent groups are formed is the basis of why individual response patterns from the same group are more correlated to each other as opposed to individuals from the other membership group, basically showing that those who share commonalities are more likely to answer the same questions similarly.

The latent class model assumes that once group membership is accounted for, the attitudinal responses are independent. Keeping this in mind, the probability that an individual with given characteristics has a specific response pattern is explained as follows:

$$\Pr(x_i : z_i) = \sum_{C=1}^{C} \Pr\,(\mathrm{c}: z_i) \Pr(x_i | c) = \sum_{c=1}^{C} \Pr\,(\mathrm{c}: z_i) \prod_{q=1}^{Q} \prod_{s=1}^{S} (\pi_{qs|c})^{x_{iqs}} \tag{2}$$

Note that $\Pr(x_i|c) = \prod_{q=1}^{Q} \prod_{s=1}^{S} (\pi_{qs|c})^{x_{iqs}}$ is the probability that the individual response pattern x_i is conditional on belonging to group c, which ultimately results in the probability of observing an individual's response pattern.

The main goal of this type of estimation is to find the parameter values that can describe the response patterns most effectively. This is achieved by finding the probabilities that will maximize the log likelihood function using $Pr(c: z_i|x_i)$ and $\pi_{qs|c}$, which are both functions of the conditional probability $Pr(c: z_i|x_i)$;

$$\ln L = \sum_{i}^{N} \ln\,[\Pr(x_i : z_i)] = \sum_{i}^{N} \ln\,[\Pr(c: z_i) \prod_{q=1}^{Q} \prod_{s=1}^{S} (\pi_{qs|c})^{x_{iqs}}] \tag{3}$$

subject to $\sum_{s=1}^{S} \pi_{qs|c} = 1$ and $\sum_{c=1}^{C} \Pr\,(\mathrm{c}: z_i) = 1$.

The function $\pi_{qs|c}$ that maximizes the log likelihood function (3) is:

$$\pi_{qs|c} = \frac{\sum_{i=1}^{N} \Pr\,(c: z_i|x_i)\, x_{iqs}}{\sum_{i=1}^{N} \Pr\,(c: z_i|x_i)} \tag{4}$$

In equation (4), the numerator results in the number of times a respondent *i* gives a particular answer *s* to a question q, weighted by the conditional probability that the respondent is in group *c*. The denominator results in the number of individuals in group *c*. Thus, equation (4) is the proportion of the number of times respondent *i* in group *c* gives a particular answer *s* to question *q*.

Before looking at the unconditional probability that maximizes equation (3) it is worth discussing element z_i as it can either vary continuously or have a finite number of discrete values. Since z_i does in fact have elements that vary continuously, $\Pr(c{:}\,z_i)$ is specified as a function of some vector class-specific parameters β_c such that $0 \leq \Pr(c{:}\,z_i) \leq 1$ and $\sum_c \Pr(c{:}\,z_i)=1$.

Considering that z_i will have continuous values a logit specification is used:

$$\Pr(c{:}\,z_i) = \frac{e^{\beta_c z_i}}{\sum_{k=1}^{C} e^{\beta_k z_i}}, c = 1, \ldots, C/, \tag{5}$$

The elements of β are estimated, but since the closed-form solutions for β do not exist, a numerical optimization routine must be embedded in an expectation-maximization algorithm.

Considering the parameters in the equations (4) and (5) are unknown, it is not possible to obtain the maximum likelihood estimates of the functions $Pr(c{:}\,z_i|x_i)$ and $\pi_{qs|c}$. The remedy to this situation is to use the expectation-maximization (E-M) algorithm, which can be used to perform a maximum likelihood estimation when there is incomplete information (Bartholomew and Knott 1999; Dempster et al. 1977; Morey, Thacher, and Breffle 2006). The E-M algorithm estimates the maximum likelihood in two steps; an expectation step and a maximization step. The expectation step determines the expected value of the latent information, then the maximization step estimates the maximum likelihood while treating the latent information's true value the same as the latent information's expected value. Upon reviewing the results the expected value of the latent information is compared to the true value and this process is reiterated until the log-likelihood function converges. Convergence occurs when the percentage change in L_β approaches a small, pre-specified number. In the LCM the conditional membership probability is $Pr(c: z_i|x_i)$, which means that the E-M algorithm is determining the values of $Pr(c: z_i|x_i)$ and $\pi_{qs|c}$. What makes this an expected likelihood function is the treatment of the expected values of the conditional membership probability being the same as that of the true values.

In short, the model is estimated by using a guessed number of *N* values of $Pr(c: z_i|x_i)$ at first. Equations (4) and (5) are then used to calculate $\Pr(c: z_i)$ and $\pi_{qs|c}$. The resulting equation is then used to recalculate the new $Pr(c: z_i|x_i)$ and the estimation is repeated using the new probability. Each iteration should use equation (3) to calculate the lnL.

In this study, the number of latent classes was determined to be the best fit of the model based on the Bayesian Information Criterion (BIC) likelihood based criteria. The BIC is a statistic criteria based on the likelihood function that measure the quality of the model while introducing penalty terms in order to reduce overfitting of the model (Kass and Wasserman, 1995). The difference between the AIC and BIC is the BIC has a larger penalty term and thus was determined to be more suitable for the purpose of this study. When assessing the best fit of the model, the lowest BIC will likely result in the best fit.

The basic latent class model (that uses respondents' attitudes to identify unmeasured class membership) can be extended to Multiple Groups LCA and LCA with covariates. When existing subgroups exist in the data that represent different populations, a multiple group LCA can be used to compare certain aspects of the latent class model across these groups (Clogg & Goodman, 1985; Collins & Lanza, 2010; Hagenaars, 2003). In turn, a LCA with covariates includes additional variables (not necessarily attitudinal) that may also affect which class each individual may be categorized into. While both methods (multiple Groups LCA and LCA with covariates) have advantages, this study utilized a basic LCA. Introducing covariates into a LCA can be very useful; however, it is primarily used for the studies with large samples of responses (more than 500). It was found that the logistic regression coefficients for the covariates showed relatively high biases when the sample size was relatively small (Wurpts & Geiser, 2014). For these reasons it was decided not to include covariates in the estimation of the latent classes and instead focus on the survey responses pertaining to consumer attitudes and perceptions.

2.2 Travel Cost Model

The travel cost method (TCM) is a commonly used revealed preference approach that can be used to estimate the total economic value a consumer derives from travelling to a site for recreation by accounting for the costs incurred in taking the trip. These costs include transportation, access fees, lodging, the opportunity cost of time, etc. Utilizing data regarding the number of trips taken by the sample population and the costs that have been incurred, a function for the recreational demand of a consumer can be estimated. Once the demand function is estimated it can be used to assess consumer surplus which is the difference between the total economic value derived and the total cost incurred from taking a recreational trip by the sample of consumers. Consumer surplus can be visualized on a demand curve by observing the area under the demand function and above the travel cost level (Hanley et al., 2007)

Estimating a single-site TCM with a substitute site is modeled as follows:

$$x_1 = f(p_1, p_2, q_1, y) \tag{6}$$

The model describes the number of trips x_1 as a function of the travel cost to the visited site p_1 , the travel cost to an alternate site p_2, the quality at the visited site q_1 , and the visitor demographics y (such as household income or education) that are included in the model that might influence the number of trips an individual would take.

Employing the travel cost method requires some initial calculations of the incurred travel costs to the visited site p_1 and the alternate site p_2 from the individual respondents' starting point. These costs are estimated using the monetary cost of travel and the opportunity cost of travel time. In this study the costs are estimated from the mid-point of the provided zip code of each respondent to the visited site p_1 by using the following relationship:

$$p_{i1} = cd_{i1} + \gamma w_i(\frac{d_{i1}}{mph}) \tag{7}$$

where c represents the cost per mile traveled, d_{i1} is the round trip distance from home to site, and mph is the travelling speed in miles per hour. The implicit wage rate is calculated using the respondent household income provided in the survey as a fraction 0< γ<1 of the hourly wage rate w_i. The fraction γ of the wage rate is assumed to be 1/3 based on previous literatures of Anderson (2010) and Parsons (2003), as well as the initial economic study conducted for the study region (Bi et al., 2015). Individual respondent households are represented by $i = 1, \dots, N$. The cost per mile c is \$0.575 per mile based on the standard mileage rate determined by the Internal Revenue Service (IRS, 2015).

Considering that the study area spans a large area of the state of Florida, and that the respondents were residing in three regions (north, central, south Florida), an alternative site for freshwater recreation was decided upon for each region based on popularity, proximity to respondents' residences (based on respondent home zip codes), and the availability of comparable opportunities for recreational activities.

While the Poisson model is a commonly used regression model for the TCM, it has the drawback of assuming that the conditional mean is equal to the variance (equi-dispersion): $E(x_i|z_i\beta) = V(x_i|z_i\beta) = \lambda_i$. However, this assumption is violated in our dataset since there is a larger than expected variability in the data, otherwise known as overdispersion. In order to account for overdispersion a common version of the negative binomial model was estimated as shown in (8):

$$\log\big(E(x_i)\big) = z_i\beta + \theta_i \tag{8}$$

where $z_i\beta$ is the conditional mean of the Poisson model and θ_i is the error term (unobserved heterogeneity). This model accounts for systematic and random variations in the means for every respondent.

Next the distribution of trips conditional on θ_i is estimated by substituting the right hand side of (8) into the probability for a Poisson random variable:

$$\Pr(x_i|\theta) = \frac{\exp(-\exp(z_i\beta + \theta_i))\exp(z_i\beta + \theta_i)\,n}{x_i!} \tag{9}$$

If $\exp(\theta_i) = v_i$ has a normalized gamma distribution where $E(v_i) = 1$ then the gamma distribution is given as:

$$h(v) = \frac{\alpha^\alpha}{\Gamma(\alpha)}\exp\,(-\alpha v)v^{\alpha-1} \tag{10}$$

In order to find the probability function for the number of trips, x_i, and arrive at the negative binomial probability function the error υ must be factored out resulting in:

$$\Pr(x_i) = \frac{\Gamma\left(x_i + \frac{1}{\alpha}\right)}{\Gamma(x_i + 1)\Gamma\left(\frac{1}{\alpha}\right)} \left(\frac{\frac{1}{\alpha}}{\frac{1}{\alpha} + \lambda_i}\right)^{\frac{1}{\alpha}} \left(\frac{\lambda_i}{\frac{1}{\alpha} + \lambda_i}\right)^{x_i} \tag{11}$$

where $\lambda_i = \exp{(z_i\beta)}$.

The mean of the negative binomial is now $E(x_i) = \lambda_i = \exp{(z_i\beta)}$ and the variance of the dependent variable is $V(x_i) = \lambda_i(1 + \alpha\lambda_i)$. The parameter α represents overdispersion so if $\alpha = 0$ then the negative binomial reverts back to a Poisson model.

The number of trips a household takes is represented by the λ_i parameter, otherwise known as latent demand. The demand function is represented in a log-linear form to ensure nonnegative probabilities and is written as:

$$\ln{(\lambda_{i1})} = \beta_0 + \beta_1 p_1 + \beta_2 p_2 + \beta_3 q_1 + \beta_4 y_i \tag{12}$$

Which can then be transformed to:

$$\hat{\lambda}_{i1} = exp(\hat{\beta}_0 + \hat{\beta}_1 p_1 + \hat{\beta}_2 p_2 + \hat{\beta}_3 q_1 + \hat{\beta}_4 y_i) \tag{13}$$

In order to obtain the number of trips her household to the study site (referred to as site 1) in a year.

The consumer surplus of a trip to site 1 can then be assessed using the results from the estimation using:

$$CS/household/trip = -\frac{1}{\hat{\beta}_1} \tag{14}$$

Then the value consumer surplus can be found by multiplying (12) with (13):

$$CS/household/year = -\frac{\hat{\lambda}_{i1}}{\hat{\beta}_1} \tag{15}$$

The consumer surplus each household receives per trip in a year (14) can then be multiplied with the number of households using data gathered from the most recent Census (US Census Bureau, 2014).

3. Telephone Survey

A previous economic study of recreational demand in the St. Johns River employed two surveys aimed at identifying two groups of recreationists (Bi, et al., 2015). An online survey was used to target those who would frequently recreate in the river, while a telephone survey was developed to target a sample population that would be representative of the state of Florida's general population. The estimations in this study employed the data gathered from the telephone survey that consisted of 500 completed responses evenly distributed across the northern, central, and southern regions of Florida.

Designed with the estimation of the TCM in mind the survey consisted of questions pertaining to recreation habits, frequencies, and preferred activities as well as site-specific water and site quality perceptions. The questions concerning the respondents' stated attitudes gathered information regarding their perceptions of the water quality and quantity in their home counties, donations towards environmental causes, and their level of satisfaction with governing laws and regulations regarding Florida waterways. Also included were questions about the respondents' perceptions regarding the amount of government spending towards education, the environment, economic development, and current infrastructure. The last section of the survey gathered socio-demographic information about the respondents.

The attitudinal questions aimed at gauging respondent perceptions and satisfaction utilized Likert scale responses in order to assess a certain level of approval, disapproval, or neutrality. A Likert scale employs a numerical scale, most commonly on a five to seven point scale. The data was not changed in order to conduct the LCA, however after the LCA was estimated the attitudinal survey responses were later recoded and used in the TCM. The attitudinal questions and their Likert scale responses are presented in Table 1 as they were used in the LCA.

4. Results

4.1 Latent Class Analysis Segments

Upon estimation of the LCA, two latent groups were revealed based on their individual preference heterogeneity. Class 1 resulted in a total of 261 respondents and class 2 was comprised of 238 respondents out of the 496 respondents. Using a rank-sum test indicated that not many demographics were statistically different among classes. The largest variations were seen in the number of respondents that traveled more than ten miles to participate in an inland outdoor activity such as hiking, biking, wildlife viewing, horseback riding, or freshwater-related activities like swimming, boating, or swimming in the past 12 months. While 44% of class 1 reportedly traveled more than ten miles to recreate in the past 12 months, class 2 seemed to have fewer visits with reportedly 34% (Table 2).

The two classes were set apart by their major differences with regard to their responses to the attitudinal questions. Conducting a rank-sum test showed that the attitudinal variables were statistically different between the two classes. The distribution of attitudinal variables among classes resulting from the LCA can be seen in Table 3.

When asked about their home county water quality, 70% of class 1 expressed concern as opposed to 50% of class 2 as shown in Figure 1. Similarly, 55% of class 1 felt that a water shortage in their home county was likely to occur compared to 46% of class 2. Fifty-two percent of class 1 reported contributing time or money to environmental causes compared with 32% of class 2.

There was significant split between the classes when it came to their attitudes and perceptions regarding the reach of laws and regulations protecting Florida's freshwater quality and government spending. Specifically, 77% of class 1 felt that existing laws do not provide enough protection of Florida's freshwater where as 60% of class 2 felt that the protection struck the right balanced and should not be intervened with as represented by Figure 2.

With regard to respondent satisfaction with FL's amount of environmental spending, 90% of class 1 felt that there is not enough money being allocated to this area contradictory to the results for class 2, where the majority (59%) felt that the amount of spending is at the proper level (Figure 3).

Continuing with the respondents' perceptions regarding the allocation of Florida's budget for economic development, the responses between the two classes varied. Specifically, 38% of class 1 felt the amount spent on economic development is just right where as 41% felt that there is not enough and 15% felt there was too much money allocated towards economic development. A similar split was seen in class 2 of which 41% felt the amount of spending on economic development is just right and 37% felt there is not enough being spent while 15% were not sure, as shown in Figure 4. Similar trends were seen in regards to infrastructure spending, with 54% of class 1 felt that there is not enough money being spent on FL infrastructure as opposed to 33% of class 2.

Continuing with the trend, 48% of class 2 is satisfied with the amount of infrastructure spending compared to 29% of class 1 who share the same opinion.

Education was the only area presented in the survey that the majority of both groups felt should have a larger budget, of which 87% of class 1 and 47% of class 2 both shared the same opinion of wanting more money for education.

In sum, Class 1 seemed to be more concerned with the quality of the environment in both their home counties and their reported recreation site. Their concern with environmental quality is apparent by the larger amount of environmental contributions class 1 makes compared to class 2. The importance that class 1 places on the environment is further illustrated by nearly the entire group's desire to see more government spending towards the environment.

For class 2, the results from the LCA suggest that they are generally content with the level of protection of water resources, the quality of water in their home counties, and the amount of spending toward environmental protection.

4.2 Travel Costs

A standard negative binomial regression (NBR) was preferred because of the overdispersion in the explanatory variables. Table 4 presents the results from the main NBR estimation that accounts for the LCA results. Table 5 presents the marginal results from the main NBR estimation.

Results from the main NBR shown in Table 4 indicated a negative inverse relationship between the travel cost and number of trips taken to have statistically significant coefficients at 10% significance in class 1 and 1% significance in class 2. Holding all other variables constant, the marginal changes presented in Table 5 shows a one unit increase in the travel cost would result in 0.04 fewer trips taken for class1 at 10% significance and 0.03 fewer trips taken for class 2 at 5% significance. In regards to the effect of the cost to travel to an alternate site on the number of trips taken, with all other variables held constant, a one unit increase in the travel cost to an alternate site would result in increases of 0.31 trips for class 1 and 0.11 trips for class 2 at 5% significance confirming that the alternate sites in the SJR Basin are substitutes, ceteris paribus.

Respondent perceptions' regarding freshwater quality at the recreation site last visited resulted in the largest marginal impact and is the most influential variable among class 1's demand curve. In the main model a one-unit increase in the perception of freshwater quality resulted in 4.36 additional trips in class 1 (not statistically significant) and 2.23 more trips taken among class 2 (statistically significant), ceteris paribus. This shows the effects of accounting for individual preference heterogeneity among consumers. Even though the objective quality of the water in the SJR Basin is classified as impaired, per unit increase in their perception of the freshwater quality improving class 1 would take four additional trips and class 2 would take two. Considering class 1 tends to recreate more frequently this is consistent with the notion that the better they perceive the quality of the water, the more often they are likely to recreate in it.

4.3 Robustness of Models

Table 6 reports the results from the NBR accounting for the LCA results and the attitudinal variables, which is used for comparison purposes. Table 7 reports the marginal effects of the attitudinal variables from the models in Table 6. We find that including altitudinal variables did not significantly improve the model fit, as compared to the main results in Table 4; however, it further provides evidence on the source of consumer preference heterogeneity. Specifically, the two classes differ in their altitudes towards the current laws and regulations protecting freshwater in Florida. As a result, the number of recreational trips to the SJR was significantly affected. Comparing to the consumers who were satisfied with the current Florida freshwater protection laws, those who believe the current laws were not sufficient in Class 1 recreated more often in the SJR. In contrast, this dummy variable is not statistically significant in Class 2 (Table 7).

To compare with the results in Table 4, we further estimate a joint TCM model in Table 8, which contains all of the observations pooled together. The estimates are consistent with the main results in Table 4. We further use the log likelihood ratio test for the hypothesis that a pooled model should be estimated (as in Table 8) vs. the two separate models should be estimated (as in Table 4), and we cannot reject the null hypothesis.

4.4 Post-estimation Results

Information from the NBR was used to estimate consumer surplus and predicted number of future trips in order to estimate the benefits realized by the respondents. The main model predicted the number of trips per year for classes 1 and 2 to be 2.66 and 2.10 with a standard deviation of 5.45 and 4.86 respectively. Using the coefficients from Table 4 this translated to a consumer surplus per household per trip of $83.33 for class 1 and $40 for class 2. The annual consumer surplus per household for classes 1 and 2 was $221.67 and $84 respectively. The total annual benefits for all Florida households estimated to belong to class 1 was $142,130,800, and it was $28,770,123 for class 2. While both classes predicted a relatively similar number of trips, class 1 received a higher annual consumer surplus per household per trip resulting in a significantly higher level of total annual benefits received.

Since the results from the LCA suggest that class 1 is more environmentally concerned than class 2 and they recreate more frequently, it is consistent with the findings that the benefits they receive from recreation are much higher than that of class 2.

Post-estimation results were also calculated for comparison using the coefficients from the simple model that did not include the LCA (Table 8). The simple model predicted the number of trips per year to be 2.27 with a standard deviation of 4.86. Using the coefficients from Table 8 this resulted in a consumer surplus per household per trip of $76.92 and an annual consumer surplus per household of $174.62. The resulting total annual benefit received is $171,639,210. Consumer surplus estimations are reported in Table 9.

5. Conclusions

The addition of consumer perceptions and attitudes regarding freshwater resources and government policies did in fact have a significant impact on the estimated demand for freshwater recreation. Conducting a LCA in order to account for the individual preference heterogeneity of consumers resulted in a more efficient model of demand for freshwater recreation. The lower BIC (a popular measure for examining the overall fit of maximum likelihood models) indicated that the inclusion of latent groups resulted in a better model fit. The likelihood ratio test indicated that we cannot reject the hypothesis that accounting for individual preference heterogeneity in the model improves the model estimation.

The results showed that the class 1 visitors (more frequent users) gain higher total annual benefits from recreating in the SJR Basin compared to class 2 (the less frequent users). The more frequent users of class 1 are also likely to take more recreational trips than class 2 as well as more likely to visit the SJR Basin specifically.

Restricting the TCM by the results of the LCA enhanced the model by accounting for consumer perceptions and their individual preference heterogeneity. Future studies can benefit from utilizing the methods presented in this study in order to account for the differences in underlying groups of consumers that might not be captured in standard valuations.

Even more so, the straightforward manner in which this estimation was conducted is preferential for policy analysts as it allows for similar studies to be easily replicated. By recognizing different types of consumer groups based on their individual preference heterogeneity policy makers can also better consider the trade-offs of implementing certain policies by accounting for what groups may benefit more and which groups have more to lose. A tangible example of current policy that could benefit from this type of analysis would be the Surface Water Improvement and Management (SWIM) plan and Basin Management Action Plan (BMAP) developed to restore and manage water resources in SJR Basin. The results from this study can be used in the plans and other water quality policy initiatives by accurately analyzing the benefits provided by water quality improvement, and the distribution of benefits between (latent) groups of Florida residents.

Implications for future management strategies may also be realized by potential improvements in marketing and management strategies resulting from a more accurate representation of the groups that recreate in the Basin. While these groups share many similar demographic characteristics it is their personal values, attitudes, and preferences that lead to a more accurate grouping and targeting of recreationists.

Note that forming latent groups based on the recreationists' perceptions and attitudes improved the overall model fit. Future studies can benefit from the inclusion of more stated preference questions in their survey design in order to better capture individual preference heterogeneity and improve the overall model estimation.

References

Aitkin, Murray, and Rubin, D.B. (1985). Estimation and Hypothesis Testing in Finite Mixture Models. Journal of the Royal Statistical Society 47(1), 67- 75. Print.

Anderson, D.M. (2010). Estimating the economic value of ice climbing in Hyalite Canyon: an application of travel cost count data models that account for excess zeros. Journal of Environmental Management. 91, 1012-1020.

Bartholomew, D., and M. Knott. (1999) Latent Class Models and Factor Analysis.2Oxford University Press Inc.

Bi, X., Borisova, T., Larkin, S., and Longanecker, J. (2015) St. Johns River Economic Study. St. Johns River Water Management District, 226-262. Retrieved from http://www.fred.ifas.ufl.edu/faculty/xiangbi/.

Boxall, P. C., and Adamowicz, W.L. (2002). Understanding Heterogeneous Preferences in Random Utility Models: A Latent Class Approach. Environmental and Resource Economics 23, 421-46.

Bujosa, A., Riera, A., and Hicks, R.L. (2010). Combining Discrete and Continuous Representations of Preference Heterogeneity: A Latent Class Approach. Environmental and Resource Economics, 47(4), 477-93.

Clogg, C.C., and L.A. Goodman (1985). Simultaneous latent structure analysis in several groups. Sociological Methodology, 81-110.

Collins, L. M., and S. T. Lanza, (2010). Latent Class and Latent Transition Analysis With Application in the Social, Behavioral, and Health Sciences. Pennsylvania State University, Chapter 5, 113.

Dempster, A., N. Laird, and D. Rubin. (1977). Maximum Likelihood from Incomplete Observations. Journal of the Royal Statistical Society, B39,1- 38.

Hagenaars, J. A., and A. L. McCutcheon (2003). Applied Latent Class Analysis.

Hanley, N., Shogren, J. F., & White, B. (2007). Environmental Economics in Theory and Practice. New York: Oxford University Press.

Internal Revenue Service (IRS). (2015). Standard Mileage Rate (2015).

Kass, R. E., and L. Wasserman (1995). A Reference Bayesian Test for Nested Hypotheses and Its Relationship to the Schwarz Criterion. Journal of the American Statistical Association, 90(431), 928-34

Lanza, S. T., L. M. Collins, D. R. Lemmon, and J. L. Schafer (2007). PROC LCA: A SAS Procedure for Latent Class Analysis. Struct Equ Modeling, 14(4): 671-694.

McCutheon, A. L. (1987). Latent Class Analysis. Quantitative Applications in the Social Sciences, (64).

Morey, E., J. Thacher, and W. Breffle (2006). Using Angler Characteristics and Attitudinal Data to Identify Environmental Preference Classes: A Latent- Class Model. Environmental and Resource Economics, 34(1), 91-115.

Parsons, G. R. (2003). The travel cost model. Chapter 9 in A Primer on Nonmarket Valuation, edited by P. A. Champ, K. J. Boyle, and T. C. Brown, London: Kluwer Academic Publishing.

Provencher, B., K. A. Baerenklau, and R. C. Bishop (2002). "A Finite Mixture Logit Model of Recreational Angling with Serially Correlated Random Utility." American Journal of Agricultural Economics 84(4), 1066-1075.

Scarpa, R., M. Thiene, and T. Tempesta (2007). Latent Class Count Models of Total Visitation Demand: Days out Hiking in the Eastern Alps. Environmental and Resource Economics, 38(4), 447-60.

U.S. Census Bureau. (2014). Florida. Retrieved from http://quickfacts.census.gov.

Vermunt, J. K. (2010). Latent Class Modeling with Covariates: Two Improved Three-Step Approaches, Political Analysis, 18: 450-469.

Wurpts, I. C., and C. Geiser (2014). Is Adding More Indicators to a Latent Class Analysis Beneficial or Detrimental? Results of a Monte-Carlo Study. Frontiers in Psychology, 5.

Table 1. Attitudinal Survey Questions Used in LCA Estimation

Survey Question	Likert Scale Response
In your home county, how much of a problem is the quality of water in the lakes, streams, rivers, and springs?	1 = No problem at all 2 = A small problem 3 = A moderate problem 4 = A very big problem 8 = Not sure 9 = Prefer not to answer
How likely do you think it is that your home county will experience severe shortages of fresh water in the next 10 years?	1 = Not at all likely 2 = Slightly likely 3 = Somewhat likely 4 = Moderately likely 5 = Very likely 8 = Not sure 9 = Prefer not to answer
At the present time, do you think laws and regulations protecting water quality in Florida's rivers, lakes, and springs have gone too far, struck the right balance, or have not gone far enough?	1 = Not enough 2 = Right balance 3 = Too far 8 = Not sure 9 = Prefer not to answer
Florida's budget includes a wide array of government expenses. Some areas will be presented, and for each one, please express whether you think we're spending too much money on it, about the right amount of money, or too little money.	
The Environment	1 = Too little 2 = Right balance 3 = Too much 8 = Not sure 9 = Prefer not to answer
Economic Development	1 = Too little 2 = Right balance 3 = Too much 8 = Not sure 9 = Prefer not to answer
Education	1 = Too little 2 = Right balance 3 = Too much 8 = Not sure 9 = Prefer not to answer
Infrastructure	1 = Too little 2 = Right balance 3 = Too much 8 = Not sure 9 = Prefer not to answer

Table 2. Demographics Divided by Latent Class

	Class 1	Class 2
Male	0.36	0.42
Average Age	63.77	62.42
Full Time FL Resident	0.96*	0.99
Property:		
Owner	0.84	0.87
Renter	0.15	0.12
Home Type:		
Apartment / Condo	0.14	0.13
House	0.78	0.77
Mobile Home	0.07	0.08
Household Income		
Income <$50k	0.34	0.34
Income >$50k	0.38	0.43
Income undisclosed	0.28	0.22
Education:		
High School or less	0.21	0.20
Some College / Tech School	0.29	0.28
College or higher	0.49	0.51
Recreated in past 12 months	0.44*	0.34
Total	261	238

*Statistically different from Class 2 at 5% significance using a Rank-sum test

Table 3. Attitudinal Data Divided by Latent Class

	Class 1	Class 2
Home County Water Quality Concern	.70**	.51
Home County Water Shortage Likeliness	.55*	.46
Environmental Contributions	.52**	.32
Satisfaction with Current Florida Freshwater Protection Laws		
Too Far	.01**	.10
Balanced	.14**	.60
Not Enough	.77**	.13
Not Sure	.08**	.17
Satisfaction with Current Level of Educational Spending		
Too Far	.02**	.13
Just Right	.09**	.27
Not Enough	.87**	.47
Not Sure	.02**	.13
Satisfied with Current Level of Environmental Spending		
Too Far	.00**	.16
Just Right	.10**	.59
Not Enough	.90**	.07
Not Sure	.00**	.18
Satisfied with Current Level of Infrastructure Spending		
Too Far	.15**	.08
Just Right	.29**	.48
Not Enough	.54**	.33
Not Sure	.02**	.12
Satisfied with Current Level of Economic Development Spending		
Too Far	.15**	.07
Just Right	.38**	.41
Not Enough	.41**	.37
Not Sure	.06**	.15

*, **Statistically different from Class 2 at 5% and 1% significance using a Rank-sum test

Table 4. Main NBR Accounting for LCA

Number of Trips		Class 1	Class 2
β_1	Travel Cost	-0.012	-0.025
		(0.007)*	(0.008)***
β_2	Travel Cost to Alternate Site	0.090	0.098
		(0.007)***	(0.009)***
q_1	Perceived Site Freshwater Quality	1.268	2.010
		(0.515)**	(0.664)***
y_i	Household Income	0.305	-0.419
		(0.263)	(0.368)
	Environmental Contributions	0.920	2.443
		(0.368)**	(0.496)***
	Constant	-10.476	-6.105
		(3.599)**	(3.948)
α	Alpha (Overdispersion)	1.933	1.903
		(0.168)***	(0.249)***
	Log Pseudolikelihood	-237.95	-123.37
	BIC	514.78	284.99
N	Observations	258	236
	(Percentage of Class Membership)	(0.52)	(0.48)

* p<0.1; ** p<0.05; *** p<0.01

Table 5. Marginal Effects of Main Model (From Table 4)

Number of Trips		Class 1	Class 2
β_1	Travel Cost	-0.043	-0.028
		(0.023)*	(0.012)**
β_2	Travel Cost to Alternate Site	0.308	0.108
		(0.136)**	(0.049)**
q_1	Perceived Site Freshwater Quality	4.363	2.233
		(2.300)	(0.940)**
y_i	Household Income	1.050	-0.466
		(0.988)	(0.481)
	Environmental Contributions	3.166	2.713
		(1.934)	(1.417)
N	Observations	258	236

* p<0.1; ** p<0.05; *** p<0.01

Table 6. NBR Accounting for LCA and Explaining Attitudinal Variables

Number of Trips		Class 1	Class 2
β_1	Travel Cost	-0.017 (0.006)**	-0.040 (0.010)**
β_2	Travel Cost to Alternate Site	0.090 (0.008)**	0.109 (0.013)**
q_1	Perceived Site Freshwater Quality	1.386 (0.474)**	3.307 (0.681)**
y_i	Household Income	0.074 (0.270)	-0.302 (0.290)
	Environmental Contributions	0.608 (0.388)	2.991 (0.692)**
	Home County Water Quality Concern	-0.148 (0.462)	0.607 (0.507)
	Not Enough Laws Currently Protecting Freshwater in FL	1.185 (0.513)*	-0.817 (0.993)
	Too Many Laws Currently Protecting Freshwater in FL	-15.944 (1.020)**	0.381 (0.592)
	Too Little Environmental Spending	-0.055 (0.551)	0.036 (0.799)
	Too Much Environmental Spending	0.000 (0.000)	0.037 (0.770)
	Too Little Economic Development Spending	-0.696 (0.483)	1.472 (0.497)**
	Too Much Economic Development Spending	-0.917 (0.534)	1.583 (1.090)
	Constant	-8.474 (3.361)*	-12.389 (3.402)**
α	Alpha (Overdispersion)	1.863 (0.174)**	1.761 (0.255)**
	Log Pseudolikelihood	-235.65	-119.71
	BIC	543.49	310.45
N	Observations	258	236

* $p<0.05$; ** $p<0.01$

Table 7. Marginal Effects of Attitudinal Variables (From Table 6)

Number of Trips		Class 1	Class 2
	Home County Water Quality Concern	-0.450	0.924
		(1.481)	(1.016)
	Not Enough Current Laws Protecting Freshwater in FL	2.340	-0.889
		(1.175)*	(1.115)
	Too Many Current Laws Protecting Freshwater in FL	-1.031	0.738
		(0.507)*	(1.134)
	Too Little Environmental Spending	-0.172	0.055
		(1.733)	(1.243)
	Too Much Environmental Spending	0.000	0.058
		(0.000)	(1.207)
	Too Little Economic Development Spending	-2.218	2.177
		(1.462)	(1.706)
	Too Much Economic Development Spending	-2.655	2.510
		(1.664)	(3.395)
N	Observations	258	236

* *p*<0.05; ** *p*<0.01

Table 8. Simple NBR Results without LCA

Number of Trips		
β_1	Travel Cost	-0.013 (0.006)*
β_2	Travel Cost to Alternate Site	0.091 (0.006)**
q_1	Perceived Site Freshwater Quality	1.337 (0.455)**
y_i	Household Income	0.154 (0.224)
	Environmental Contributions	1.254 (0.300)**
	Constant	-9.496 (2.966)**
α	Alpha (Overdispersion)	1.980 (0.142)**
	Log Pseudolikelihood	-364.89
	BIC	773.21
N	Observations	494

* $p<0.05$; ** $p<0.01$

Table 9. Consumer Surplus Estimations

		Class 1	Class 2	Simple Model
	Consumer Surplus/ Household/ Trip	$83.33	$40.00	$76.92
λ	Predicted Number of Trips/ Year	2.66	2.10	2.27
	(Standard Deviation)	(5.45)	(4.86)	(4.86)
	Annual Consumer Surplus per Household	$221.67	$84.00	$174.62
	Total Surveyed FL Households	6,237,279	6,237,279	6,237,279
	Probability of Visiting SJR	19.77%	11.44%	15.76%
	Total Annual Benefits	$142,130,800	$28,770,123	$171,639,210
	Observations	258	236	494
	(Percentage of Class Membership)	(0.52)	(0.48)	(1.00)

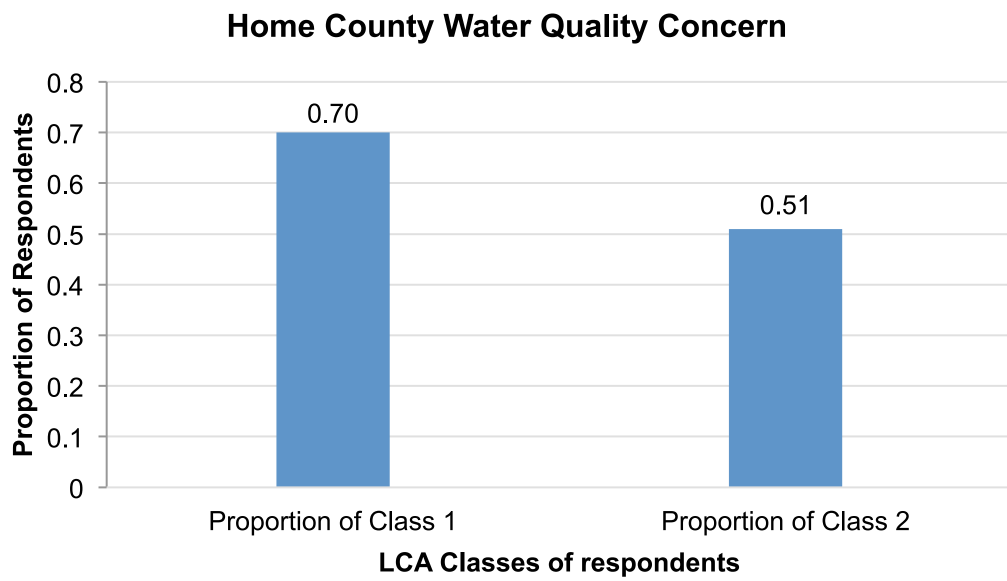

Figure 1. Quality of water in Home County: combined responses for answer choices "Moderate problem" and "Very big problem", by LCA Class

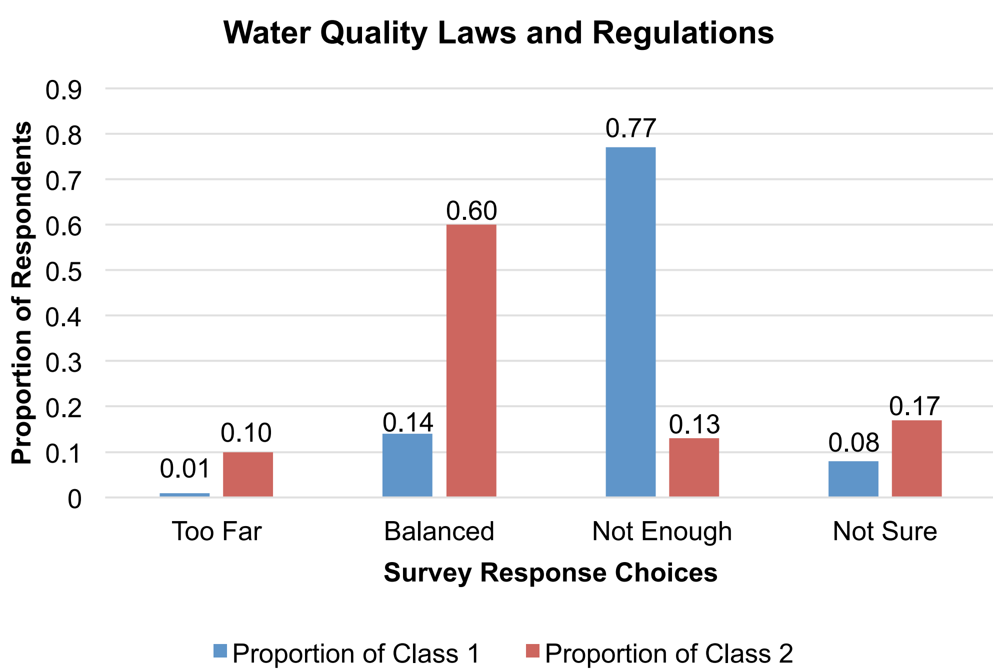

Figure 2. LCA Distribution of Balance of Water Quality Laws: Proportion of Each Response, by LCA Class

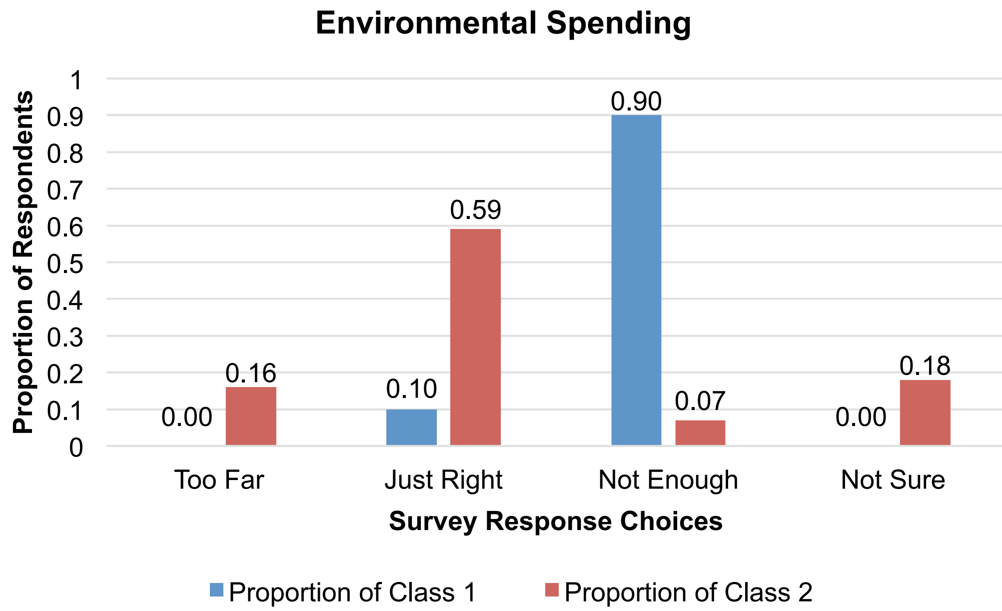

Figure 3. LCA Distribution of Environmental Spending in FL: Proportion of Each Response, by LCA Class

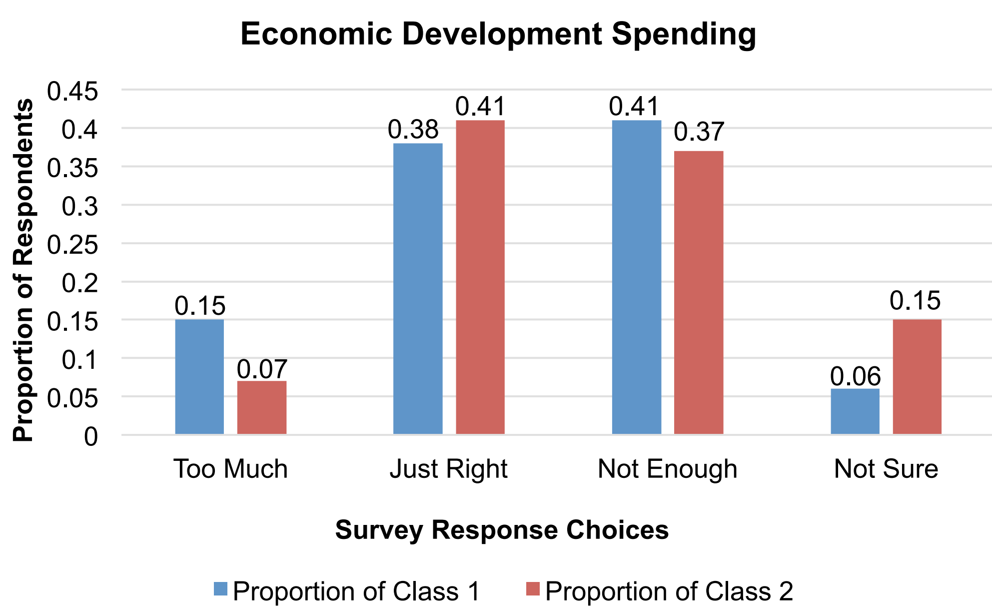

Figure 4. LCA Distribution of Economic Development Spending in FL: Proportion of Each Response, by LCA Class