Munich Personal RePEc Archive

An Information-Theoretic Approach to Estimating Willingness To Pay for River Recreation Site Attributes

Henry, Miguel and Mittelhammer, Ron and Loomis, John

Greylock McKinnon Associates, Washington State University, Colorado State University

January 2018

Online at https://mpra.ub.uni-muenchen.de/89842/
MPRA Paper No. 89842, posted 18 Nov 2018 04:28 UTC

AN INFORMATION-THEORETIC APPROACH TO ESTIMATING WILLINGNESS TO PAY FOR RIVER RECREATION SITE ATTRIBUTES

MIGUEL HENRY[a*], RON MITTELHAMMER[a] and JOHN LOOMIS[b]

[b]*School of Economic Sciences, Washington State University, Pullman, WA*

[b]*Dept. of Agricultural and Resource Economics, Colorado State University, Fort Collins, CO*

This version: August 2016

Abstract

This study applies an information theoretic econometric approach in the form of a new maximum likelihood-minimum power divergence (ML-MPD) semi-parametric binary response estimator to analyze dichotomous contingent valuation data. The ML-MPD method estimates the underlying behavioral decision process leading to a person's willingness to pay for river recreation site attributes. Empirical choice probabilities, willingness to pay measures for recreation site attributes, and marginal effects of changes in some explanatory variables are estimated. For comparison purposes, a Logit model is also implemented. A Wald test of the symmetric logistic distribution underlying the Logit model is rejected at the 0.01 level in favor of the ML-MPD distribution model. Moreover, based on several goodness-of-fit measures we find that the ML-MPD is superior to the Logit model. Our results also demonstrate the potential for substantially overstating the precision of the estimates and associated inferences when the imposition of unknown structural information is not accounted explicitly for in the model. The ML-MPD model provides more intuitively reasonable and defensible results regarding the valuation of river recreation than the Logit model.

Keywords: Minimum power divergence, contingent valuation, binary response models, information theoretic econometrics, river recreation

JEL Codes: C14, C5, Q5

The data set used in this study was collected with the support of the National Science Foundation under Grant No. 0308414. We also gratefully acknowledge helpful comments and suggestions by J. Scott Shonkwiler, J.M. Gonzalez-Sepulveda, Jonathan Yoder, Philip R. Wandschneider, and comments received from seminar attendees at Virginia Commonwealth University, University of Connecticut, Washington State University, the Economic Research Service of USDA, and at the AAEA/WAEA/CAES Conferences held in Portland, OR and Seattle, WA. The views expressed are strictly those of the authors.

[*]Corresponding Author. Present Address: Greylock McKinnon Associates, 75 Park Plaza, 4th Floor, Boston, MA 02116, USA. Email: mhenry@gma-us.com

1 INTRODUCTION

Economists have used diverse valuation techniques to estimate the economic values of non-market water resources, including revealed preference (aka recreation travel demand models and hedonic property models) and stated preference techniques (aka contingent valuation (CV) methods and choice experiments).

In practice, resource economists seek to know marginal effects of changes in different attributes of water resources (rivers or lakes) that water resource specialists are capable of managing. In principle, marginal values of changes in river recreation attributes can be obtained using choice experiments or pooling dichotomous choice CV responses derived from either hypothetical scenarios of river quality or from a natural experiment of pooling data obtained at river sites that have varying site quality attributes. When the goal is to estimate marginal values of these management relevant attributes using dichotomous choice CV responses, strong functional form assumptions of the conditional expectation (i.e., fully parametric functional specifications) are frequently imposed for this purpose. Without invoking such restrictions or when using appropriate non-parametric econometric approaches (e.g., the Kriström (1990) fully non-parametric estimator) identification of the marginal effects of changes in management relevant explanatory variables will not usually be possible. Thus, most CV studies found in the water resource valuation literature rely on fully parametric discrete choice models (e.g., binary response models [BRMs]).These BRMs allow managers to determine how the value of a water-based recreation area varies with possible changes in water management (e.g., flow regime, reservoir levels).

The large majority of empirical analysts who have utilized discrete choice models have chosen parametric statistical procedures on the basis of precedent and readily available software. These methods require a full parametric functional specification of the relationship between regressors and the response variable, and more importantly, a full specification of a parametric distribution of the disturbances (e.g., the probit (normal) or Logit cumulative distribution functions [CDFs]). Although some distributional assumptions can be benign, especially if the parameterization is flexible enough to describe behavior adequately (McFadden, 1994, McFadden and Train, 2000), the implementation of an incorrect parametric functional form can lead to spurious statistical inferences due to biased and inconsistent estimates. Moreover, underlying economic theory provides little guidance for using these functional specifications, so there is insufficient information regarding the appropriate distribution to adopt in practice (Mittelhammer *et al.,* 2000, Crooker and Herriges, 2004). Thus, any parametric functional specification for either the stochastic error or the utility differences used in these methods is in general uncertain and questionable (Creel and Loomis, 1997).

This study applies a new BRM to dichotomous CV data, which is semi-parametric by nature and, unlike many non-parametric approaches, provides both density structure and flexibility, allowing identification of marginal effects of changes in the explanatory variables on WTP. Its density structure is supported by a large and varied new family of CDFs, which has the flexibility to fit a very wide variety of distributional shapes to the choices made by survey respondents. It also nests the behavior of some prominent Binary Response estimators (e.g., the Logit model), and can approximate a vast array of left- and right-skewed probability density functions due to its highly flexible class of density shapes (Mittelhammer and Judge, 2011).

This new approach, known as the maximum likelihood-minimum power divergence (ML-MPD) binary response estimator, avoids using a priori model specification information about the functional form of the conditional expectation of the response variable. This feature of the ML-MPD estimator reduces the potential for model specification errors that arise when relying on tenuous parametric distributional assumptions underlying binary responses. Moreover, the ML-MPD maintains the full set of familiar ML sampling properties (consistency, asymptotic normality, and asymptotic efficiency) and, unlike most non-parametric methods, does not employ the usual kernel density estimation methodology with the attendant required implementation choices relating to bandwidth, kernel functions, and other tuning parameter issues. That is, the ML-MPD estimator is free of user-specified tuning parameters.

The ML-MPD estimator begins in a *non-parametric* information theoretic context regarding model specification. Then, *semi-parametric* orthogonality relationships, in the form of empirical sample moments, are introduced that, within the non-parametric information theoretic framework, ultimately lead to a (new) *parametric* family of probability distributions (CDFs) for BRMs (see Judge and Mittelhammer [2012]; Ch. 9), as well as a conditional expectation function for BRMs and estimators for the unknowns in the model. Thus, the ML-MPD approach combines attractive features of parametric, semi-parametric, and non-parametric estimation methodologies.

Several distribution-free estimators for estimating BRMs have already been proposed in the literature to overcome model misspecification issues (e.g., Manski, 1975, Turnbull, 1976, Cosslett, 1983, Kriström, 1990, Horowitz, 1992, Matzkin, 1992, Klein and Spady, 1993, Li, 1996, Chen and Randall, 1997, Creel and Loomis, 1997, Araña and León, 2005, Huang *et al.*, 2008).

However, the majority of these estimators have not found widespread application in the empirical discrete choice literature for a number of reasons that may include: 1) users' lack of understanding regarding the estimation and inference properties of the approaches in empirical applications; 2) difficulties in providing economic interpretations of the results of the analysis; 3) nonidentification of model parameters (e.g., Kriström, 1990, and the Klein and Spady (1993) estimator[1] [KS]) and marginal effects; and 4) ambiguity and/or uncertainty regarding the appropriate choices for tuning parameters and other estimator implementation-computational issues.

Creel and Loomis (1997) underscore that the required scale and local normalizations for the identification of KS parameter estimates are questionable because they go beyond restrictions implied by demand theory. Moreover, it has been found that other suggested semi-parametric methods do not achieve root-*n* consistency (e.g., the Manski [1975] and Horowitz [1992] estimators), and their finite sample behavior is in question (e.g., the Cosslett [1983], KS and Ichimura [1993] estimators).[2] While fully non-parametric estimation techniques tend to be more robust to incorrect functional specifications of conditional expectation functions as well as probability distributions, most of them involve various choices of tuning parameters, kernels, and other implementation choices. Sampling behavior in smaller-sized samples is also problematic. Crooker and Herriges (2004) state that the gains and losses from using non-parametric and semi-parametric estimators to recover WTP measures relative to the standard parametric approaches

[1] Despite the infrequency of empirical applications, the KS estimator is considered in the econometrics literature to be one of the statistically "best" semi-parametric estimators in the sense that its asymptotic covariance matrix achieves the semi-parametric efficiency bound for the single index BRM.

[2] See Chen and Khan (2003) for more details.

are still unknown. There remains a continuing need to seek robust and efficient methods for analyzing binary decision making behavior.

This empirical application of the ML-MPD estimator relates to a stated-preference CV on-site dataset collected along two rivers at the Caribbean National Forest (CNF) in Puerto Rico. The ML-MPD estimator is used to analyze the underlying behavioral economic decision process leading to a person's WTP for river recreation site attributes. In order to compare the new ML-MPD approach to a leading empirical method for analyzing BRMs, we also implement the commonly used fully parametric Logit estimator.[3]

In the next section, we discuss the ML-MPD estimator and its implementation. Section 3 describes the dataset utilized in this study and the nature of the decision making context. In section 4, we present the results from our analyses. We conclude in section 5 with a summary of our findings and its corresponding implications.

[3] Probit ML results are available from the authors upon request.

2 BRM Model and ML-MPD Estimation Framework

We present an overview of the ML-MPD estimation procedure in this section. The Logit model is well documented in the literature and is not reviewed here. The statistical approaches used in this study allow one to model the underlying consumers' binary choices as the outcomes of a latent utility-maximizing process, where respondents are assumed to choose the alternative that result in the greatest indirect utility.

2.1 Minimum Power Divergence Distributions for the BRM

To motivate the ML-MPD estimator, note that the n-dimensional vector of unknown Bernoulli probabilities corresponding to a BRM,

$$P(y_i = 1 | x_i) = p_i = 1 - F(-\mathbf{x}_{i.}\boldsymbol{\beta}) = F_*(\mathbf{x}_{i.}\boldsymbol{\beta}),\ i = 1,...,n, \tag{1}$$

is associated with an unknown *link or transformation function* $F_*(\cdot)$ of factors affecting the decision environment and that in practice is expressed in terms of an *index function*, assumed to be linear[4]. More generally, one can always characterize the Bernoulli random variables $[Y_1, Y_2, ..., Y_n]' \in \mathbb{R}^n$ as being defined by $Y_i = p_i + \varepsilon_i$, $\forall i$, with zero-mean error, $E(\varepsilon_i) = 0$, $\forall i$. Without knowledge of the particular parametric distributional specification of the link function, the traditional ML approach is not available. One might then consider a Quasi-ML approach, but this method does not assure the full set of attractive ML sampling properties (Mittelhammer *et al.*, 2000), and moreover, it can be challenging to characterize its sampling properties in any given

[4] This index is usually a function of the covariates $\mathbf{x}$ and a vector of $\boldsymbol{\beta}$ unknown parameters, which is estimated along with the link function. Although non-linear specifications and the linear Box-Cox utility function are also possible, the commonly used linear index representation $\mathbf{x}_{i.}\boldsymbol{\beta}$, with $\boldsymbol{\beta}$ being a vector of parameters, is considered in this study. The ML-MPD approach can be generalized to utilize nonlinear index representations.

application. Alternatively, one might consider the two-stage Generalized Method of Moment estimator; however, the approach is not appealing for the current application due to the ill-posed, underdetermined nature of the estimating equations (see equation (2) ahead).

In this study, we adopt an information theoretic estimator for $\boldsymbol{\beta}$. Unlike classical estimation procedures, these estimators rely on Kullback's (1959) information theoretic minimum discrimination information principle[5] and on data-moment constraints (sample moments), as defined in Mittelhammer and Judge (2011) and Judge and Mittelhammer (2012). The information-theoretic estimation principle allows joint estimation of the unknown parameters of the model along with the empirical sampling distributions that exhibit minimum discrepancy relative to a reference distribution. The MPD approach is robust in terms of the uncountably infinite number of highly varied candidate distributions that are members of the distribution class, and it maintains the full set of usual ML estimation and inference sampling behavior properties under familiar regularity conditions. As noted by Judge and Mittelhammer (2012), the MPD has been shown through Monte Carlo simulations to exhibit mean square error (MSE) superiority relative to probit and Logit estimators. Moreover, the ML-MPD approach has been shown to be MSE superior to the best semi-parametric estimator (i.e., the KS estimator) under certain sampling conditions and has the flexibility to fit a wide variety of distributional shapes to the latent variables underlying the Bernoulli process.

[5] An alternative to this principle is the maximum entropy principle, also known as the Shannon's (1948) entropy measure or the generalized maximum entropy approach. Although there are some recent theoretical and empirical contributions in the econometric literature using the latter approach (e.g., Golan *et al.*, 1996, Crooker and Herriges, 2004, Marsh and Mittelhammer, 2004) a user of the method is also confronted with a notable number of "tuning parameter" type of decisions to make, for which the performance consequences are not well known currently.

The application of the MPD procedure can be conceptualized in two stages, although implementation of the estimation methodology can actually be performed in one computational step. One begins with an ill-posed inverse problem consisting of the non-parametric linear model $Y_i = p_i + \varepsilon_i, \ \forall i$ noted above, along with generally applicable semi-parametric orthogonality conditions between explanatory variables and model noise of the general form $E\left[\mathbf{g}(\mathbf{x})^{'}(\mathbf{Y}-\mathbf{p})\right]=\mathbf{0}$. Then, an MPD solution for the probabilities is found that identifies a complete set of (new) parametric probability distributions for the BRM. In a second stage, based on the MPD solution of probability distributions, ML estimation is used to estimate the unknowns that occur in the class of probability distributions. The results produce estimates of the effects of explanatory variables on the conditional Bernoulli probabilities, and also concurrently identify a link function for those probabilities. In effect, the method estimates the functional form of the probability model along with estimates of the unknown probabilities in the model.

Regarding the first stage of the method, the information theoretic Cressie-Read (CR)[6] power-divergence family of statistics (see Read and Cressie, 1988, Imbens *et al.*, 1998) measures the discrepancy between the probabilities to be estimated and a reference distribution for those probabilities. Including sample moment constraints based on zero-mean population conditions, the minimum power divergence extremum problem is specified as:

$$\underset{\mathbf{p}}{\text{Min}}\left\{CR(\mathbf{p},\mathbf{q},\gamma) \text{ s.t. } n^{-1}\left(\mathbf{g}(\mathbf{x})'(\mathbf{y}-\mathbf{p})\right)=\mathbf{0} \text{ and } 0\le p_i \le 1, \ \forall i\right\} \tag{2}$$

[6] This goodness-of-fit measure encompasses the empirical likelihood objective as a special case when $\gamma=-1$ and the maximum entropy or Kullback-Leibler objective when $\gamma=0$, among others. As γ ranges from $-\infty$ to ∞, the CR divergence measure leads to different information theoretic estimators (see Mittelhammer *et al.*, 2000 (Chapter 13.4), Lee *et al.*, 2010, and Judge and Mittelhammer, 2012).

where $CR(\mathbf{p},\mathbf{q},\gamma)$ is a member of the CR family, $\mathbf{q}=[q_1,q_2,...,q_n]' \in \times_{i=1}^{n}[0,1]$ is an n-dimensional vector of reference Bernoulli probabilities, and $\gamma \in (-\infty,\infty)$ is the scalar power parameter of the divergence measure. The sample moment constraint vector equation $n^{-1}(\mathbf{g}(\mathbf{x})'(\mathbf{y}-\mathbf{p}))=\mathbf{0}$ is of dimension $m \times 1$, where $\mathbf{g}:\mathbb{R}^k \to \mathbb{R}^m$ is a real-valued measurable function. The inequality (non-negativity and upper bound) constraints on the solved probability values, $\mathbf{p}=[p_1,p_2,...,p_n]' \in \times_{i=1}^{n}[0,1]$, are logically necessary for **p** to represent an n-dimensional vector of conditional-on-**x** Bernoulli probabilities underlying the binary decisions. In the current application, the specification $g(\mathbf{x})=\mathbf{x}$ is utilized, where **x** is a vector of explanatory variables affecting the binary decision outcome.

The estimation objective function in (2) is represented in terms of the information theoretic CR power-divergence criterion, which in this binary application takes the following form (Mittelhammer and Judge, 2011):

$$\mathrm{CR}(\mathbf{p},\mathbf{q},\gamma)=\frac{\sum_{i=1}^{n}\left\{p_i\left(\left(\frac{p_i}{q_i}\right)^{\gamma}-1\right)+(1-p_i)\left(\left(\frac{1-p_i}{1-q_i}\right)^{\gamma}-1\right)\right\}}{\gamma(\gamma+1)} \tag{3}$$

This discrepancy measure is always positive valued unless $p_i = q_i$, no matter the choice of γ. The value of the discrepancy measure becomes larger the more divergent p_i and q_i are, and the measure is convex in the p_i's and second order continuously differentiable. The MPD family of CDFs (i.e., a minimum power divergence solution for the probabilities) that solve the extremum problem represented in equation (2), using equation (3) for the CR power-divergence criterion, is given by:

$$
\begin{aligned}
p(w_i;q_i,\gamma) &\Rightarrow \arg_{p_i}\left\{\left[\left(\frac{p_i}{q_i}\right)^{\gamma}-\left(\frac{1-p_i}{1-q_i}\right)^{\gamma}\right]-w_i\gamma=\mathbf{0}\right\} \text{ when } \gamma<0 \\
&\Rightarrow \arg_{p_i}\left\{Ln\left(\frac{p_i}{q_i}\right)-Ln\left(\frac{1-p_i}{1-q_i}\right)-w_i=\mathbf{0}\right\} \text{ when } \gamma=0 \qquad (4) \\
&\Rightarrow \arg_{p_i}\left\{\begin{matrix} 1 \\ \left(\frac{p_i}{q_i}\right)^{\gamma}-\left(\frac{1-p_i}{1-q_i}\right)^{\gamma}-w_i\gamma=\mathbf{0} \\ 0 \end{matrix}\right\} \text{when } \gamma>0 \text{ and } w_i\left\{\begin{matrix} \geq \gamma^{-1}q_i^{-\gamma} \\ \in\left(-\gamma^{-1}(1-q_i)^{-\gamma},\gamma^{-1}q_i^{-\gamma}\right) \\ \leq -\gamma^{-1}(1-q_i)^{-\gamma} \end{matrix}\right\}
\end{aligned}
$$

where $w_i = \mathbf{x}_{i.}\boldsymbol{\beta}$, $\boldsymbol{\beta}$ represents the $m \quad 1$ vector of Lagrange multipliers of the moment constraints when (2) is expressed in Lagrange form, and $\mathbf{x}_{i.}$ denotes the i^{th} $1\times k$ row vector contained within the $n\times k$ matrix of covariates, $\mathbf{x}$. A complete analytic derivation of the MPD family of CDFs solution is provided in Appendix A.

The definition in (4) characterizes an uncountably infinite number of CDFs, with argument w_i, indexed by the values of γ and q_i. For example, when $\gamma=0$ and $q_i=0.5$, the standard Logit model is subsumed by the family of distributions. It is clear that the inverse CDF of the MPD family always exists in closed form, but except for a measure zero set of possibilities for γ and q_i, the probabilities themselves must be solved for numerically. Fortunately, strict monotonicity properties of the terms involving the p_i's in (4) make for a relatively straightforward and efficient numerical solution procedure that is guaranteed to provide an appropriate solution for p_i for any admissible argument, w_i, of the CDF. In the application ahead, the empirical probabilities were solved for numerically using the interval bisection method (root-finding algorithm). Additional detail on both computational methodology and characteristics of the MPD family of distributions, including their myriad different shapes and properties, can be found in Mittelhammer and Judge (2011) and Judge and Mittelhammer (2012).

2.2 The ML-MPD estimator

The family of parametric probability distributions in (4) is used as a basis for specifying the likelihood function associated with the data outcomes, leading to a log-likelihood function of the general form. In the implementation of the MPD distribution family, we specify $q_i = q \ \forall i$, which is tantamount to assuming that the same basic functional form for the probability distribution, $p(\mathbf{x}_{i.}\boldsymbol{\beta};q,\gamma)$, is used across the observations for representing the conditional Bernoulli probabilities. In this context, it is the $\mathbf{x}_{i.}$'s, and thus the arguments of the distributions, that cause the probabilities to vary across survey respondents.

The negative of the log-likelihood function is maximized using the non-gradient based Nelder-Mead simplex minimization algorithm proposed by Nelder and Mead (1965).[7] This optimization method belongs to the general class of “direct search methods” and has become one of the most widely used techniques for non-linear unconstrained optimization. The Nelder-Mead algorithm does not rely on gradients or Hessians, so that it typically generates iterations more quickly than search methods that depend on derivatives of the objective function (e.g., Newton-Raphson). The Nelder-Mead approach also mitigates numerical problems caused by highly nonlinear

[7] A detailed explanation of this algorithm and its implementation can be found in Nelder and Mead (1965) and Jacoby *et al.* (1972).

problems, extreme flatness of log-likelihood functions, and associated unstable gradient and/or Hessian calculations from iteration to iteration.

Preliminary graphical investigations with our data regarding the behavior of the log-likelihood function suggested that the log-likelihood was notably flat in the γ direction over a range of values in which the maximum of the likelihood function was likely to reside. In order to facilitate stability and accuracy in the search for the ML optimum over a concave but relatively flat likelihood function, while mitigating the possibility of convergence to local optima, we implement a grid search in the γ direction. Specifically, we begin the search using the standard logistic distribution $(\gamma=0,\ q=0.5)$ as the starting probability distribution, and solve (2) for $(q,\boldsymbol{\beta})$ given $\gamma=0$. Then, a grid above and below $\gamma=0$ in ± 0.2 increments, is implemented, using a recursively updated set of starting values based on immediately preceding solutions, identifying a γ-grid indexed series of solutions for (2) that bracketed the optimum within an interval of length 0.2. A second grid search based on 0.01 increments between the bracket values is then utilized to determine the optimum with a 0.01 tolerance level. While this grid search method, with recursively updated starting values, does not guarantee a global optimum, it reduces the possibility of not searching in the neighborhood of the global optimum.

The estimated covariance matrix of the ML-MPD estimator is based on the optimized log-likelihood function and the standard "outer-product-of-gradients" (OPG) approach for defining the asymptotic covariance matrix of the MLE. In particular, by letting $\boldsymbol{\theta}\equiv\left[\boldsymbol{\beta}^{'}\,|\,q\,|\,\gamma\right]^{'}$, the covariance matrix was estimated via $\left.\left[\left(\frac{\partial \mathbf{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^{\prime}\left(\frac{\partial \mathbf{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)\right]^{-1}\right|_{\hat{\boldsymbol{\theta}}}$, where $\hat{\boldsymbol{\theta}}$ is the ML-MPD estimate of the parameters and $\frac{\partial \mathbf{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ is the $n \times (k+2)$ matrix of derivatives of the log of the likelihood function contributions, $L_i(\boldsymbol{\theta})$, $i=1,...,n$ with respect to $\boldsymbol{\theta}$.[8] Because there is generally no closed form solution for the probabilities in (4), one needs to apply implicit differentiation in order to define the derivatives of the log-likelihood with respect to $\boldsymbol{\beta}$, q and γ. Expressions for the derivatives are somewhat involved, but are straightforward to compute using generalized Hadamard multiplication and division operators (see Appendix B).

For comparison purposes, we also calculate and report the conditional-on-$(\hat{q},\hat{\gamma})$ covariance matrix of $\hat{\boldsymbol{\beta}}$ based on the OPG method. Note that this comparison is informative in the sense that conditioning on a particular q and γ (e.g., standard symmetric logistic distribution if $\gamma=0$ and $q=0.5$) is analogous to the common empirical practice of choosing a specific functional form for the distribution underlying binary responses, like specifying the probit or Logit model in the parametric ML estimation method. Since the covariance matrix is generally understated when conditioning on unknown parameters that were estimated, the empirical comparison of the ML-MPD and the Logit Model suggests the potential for overstating the precision of estimates and associated inferences when lack of information regarding the appropriate functional form of the probability distribution is not explicitly accounted for in the model.

Computational implementation of all of the preceding procedures relating to the ML-MPD estimator and Logit MLE is based on Aptech Systems' GAUSS™.

[8] An alternative approach could follow the result of Patefield (1985), based on the profile likelihood function concentrated on only the $\boldsymbol{\beta}$ parameter, if one were not interested in inferences concerning the MPD distribution parameters.

3 Data and Problem Setting

The dataset contains information on 718 in-person interviews with visitors at ten different recreation sites along the Mameyes and Espiritu Santo rivers at the Caribbean National Forest (CNF) in Puerto Rico during the summers of 2004 and 2005. One of the key recreation features of this collection of sites is the many “recreation pools” of water that are available to visitors for swimming, wading, and sitting , where some of the pools have the attraction of providing a “hydromassage” experience similar to that of a Jacuzzi.

The data was collected through dichotomous-response CV surveys administered at the recreation pool sites, employing the single-bounded[9] dichotomous choice approach as the elicitation protocol, which is also referred to in the literature as the “*closed-ended*” CV approach or the “*take-it-or-leave-it”* approach. Additional details of the survey and its design are given in Gonzalez-Sepulveda (2008).

The survey asked each randomly sampled recreation user the following CV question: “*As you know the price of gasoline often goes up. Taking into consideration that there are other rivers as well as beaches nearby where you could go visit, if the cost of this visit to this river was $____ more than what you have already spent, would you still have come today? Yes or No?*.” The hypothetical additional cost of the visit was randomly drawn from a set of 18 bid thresholds[10] for each respondent, and ranged from $1 to $200. Thus, increase in travel cost was the payment vehicle. A reminder of the availability of substitute recreation sites was explicitly mentioned as

---

[9] This approach has the potential to be less efficient than the double-bounded protocol. However, McFadden (1994) and Cooper *et al.* (2001) have documented that the single-bounded CV question eliminates the response inconsistency and its associated bias.

[10] The chosen bid vector of 18 prices was $\{1,5,10,15,20,30,40,50,60,80,100,120,130,140,150,160,180,200\}$.

well. Additional information used in the models include on site attributes (road quality, volume and speed of water in the pools, and size of rocks or sand surrounding the pools), a threshold indicator of the level of a recreation user's income, and travel time information as Cameron and James (1987) propose. Tables 1 and 2 summarize the variables included in the estimated models, along with descriptive statistics.

Table 1. Variables Used in the Analyses

Variable Name	Description
Choice	= 1 if willing to pay the visit price; 0 otherwise
Bid	Offered U.S. dollar amount (threshold) that households were asked to pay
Road	= 1 if non-paved road; 0 otherwise
Discharge	Mean annual speed of water in the pool (in cubic feet)
Size	Median sand size around the pools (in millimeters)
Volume	Volume of the pool (in cubic feet)
Income	= 1 if the family annual income (in U.S. dollars) is greater than $20,000 = 0 otherwise
Travel Time	= 1 if the travel time exceeds 30 minutes; 0 otherwise

Note: The variables volume and size were scaled by 10 and 100, respectively, in estimation to support numerical stability and accuracy in calculations, and allow similar orders of magnitude for parameter estimates.

Table 2. Descriptive Statistics for Variables Used in the Analyses (n = 718)

Variable Name				
Qualitative Variables	**Proportion (%)**			
▪ Choice	67.97			
▪ Road	39.00			
▪ Income	55.43			
▪ Travel time	61.56			
Quantitative Variables	**Mean**	**Std.Dev.**	**Minimum**	**Maximum**
▪ Bid	63.53	58.27	1	200
▪ Discharge	0.83	0.57	0.11	1.67
▪ Size	5.09	6.28	1.02	23.37
▪ Volume	44.67	41.40	4.2	186.84

4 RESULTS

Results presented in this section are based on the explanatory variables defined in Table 1. The WTP measures are interpreted on an aggregate per group basis[11], and are calculated for the Logit ML and ML-MPD models using Hanemann's (1984, 1989) approach. Following Hanemann, the difference in utilities of visitors from accepting or rejecting the hypothetical additional cost (i.e., the bid A) of the visit to the recreational pools of water at the CNF is denoted by

$$\Delta U(A_i, \mathbf{Z}) = U(1, I - A_i, \mathbf{Z}) - U(0, I, \mathbf{Z}) \tag{5}$$

where $U(\cdot)$ is the utility function, I denotes annual income, $\mathbf{Z}$ is a vector of visitor and recreational site characteristics, and A_i is the offered bid to visitor i. The bid value that makes the visitor indifferent between acceptance/rejection defines the WTP and the calculation of its mean is obtained by dividing the negative grand constant term by the absolute value of the estimated coefficient associated with A. The grand constant term is calculated as a linear combination of the means of the explanatory variables times their associated coefficients, with the effect of A suppressed.

4.1 Parametric Model Results

To solve for the MLE of the Logit model (see Table 3) we utilized the Berndt-Hall-Hall-Hausman optimization approach (Berndt *et al.*, 1974). The Logit model results indicate that the bid variable is highly significant and its sign is consistent with economic theory, indicating that the higher the price of a visit to the "recreation pools" of water, the less willing users of the pools are to pay for a visit. A continuous measure of the dollar amount of respondent's income has

[11] The average group in the ten different river recreation sites under study at the CNF consisted of 3.5 visitors.

often been dropped in dichotomous choice CV studies. Sometimes it is dropped based on theoretical grounds (see e.g. Hanemann, 1984, Haab and McConnell, 2002), but it can also dropped due mainly to lack of statistical significance (Gonzalez *et al.*, 2008). However, when expressed as a dichotomous income threshold indicator its coefficient in this application was found to be significantly different from zero and positively related to the probability of paying the bid amount. The variables bid, size (size of the rocks or sand surrounding the water pools) and road (non-paved roads) contribute to the explanation of the dependent variable at the 0.01 level of type I error. Volume of water in the pools of the two rivers is also positively related to the probability of paying the bid amount, whereas the variables bid, discharge, road, size and travel time are negatively associated with the probability of paying the bid amount.

Table 3. Estimation Results for the Logit and ML-MPD Models based on Conditional(c) and Unconditional(u) Covariance Matrix Results

Variable	LOGIT-MLE	ML-MPD(c)	ML-MPD(u)
BID	-0.01596***	-0.0527***	-0.0527*
	(0.00152)	(0.0132)	(0.0344)
DISCHARGE	-0.97192**	-1.7767*	-1.7767
	(0.4861)	(1.1390)	(1.5180)
SIZE	-0.05080***	-0.1083***	-0.1083*
	(0.0184)	(0.0439)	(0.0779)
VOLUME	0.00624*	0.0137	0.0137
	(0.0043)	(0.0111)	(0.0134)
INCOME	0.41435**	-0.0921	-0.0921
	(0.1832)	(0.3230)	(0.3360)
ROAD	-1.08619***	-2.1023**	-2.1023*
	(0.3981)	(1.0050)	(1.6180)
TRAVEL TIME	-0.33628**	-0.2580	-0.2580
	(0.1925)	(0.3320)	(0.3560)
INTERCEPT	3.09335***	3.8895***	3.8895*
	(0.5786)	(1.5470)	(3.0650)
q		0.8835	0.8835***
			(0.0309)
γ		-4.23	-4.23***
			(1.5598)
McFadden R^2	0.1105	0.1936	
AIC	846.9398	742.2263	
BIC	883.5515	778.8381	
Mean WTP ($)	114.74	27.46	
Lower CI ($)	101.28	13.01	
Upper CI ($)	128.19	41.91	
Log Likelihood	**-415.47**	**-363.11**	

Notes: ML-MPD(c) and ML-MPD(u) are the ML-MPD models based on conditional and unconditional variance-covariance matrices of $\boldsymbol{\beta}$, respectively. Asymptotic standard errors are shown in parentheses. AIC and BIC are the Akaike information criterion and Schwarz's information criterion, respectively. Z-Lower and Z-Upper Confidence Interval Levels for the mean WTP, shown by Lower CI and Upper CI for 95% confidence levels, are calculated using the delta method. For the computational implementation of the Logit model, an iterative algorithm with analytical gradients and analytical Hessian was implemented in GAUSS™ 12. ***, **, and * denote statistical significance at the 99%, 95%, and 90% confidence levels, respectively, using one-sided critical values $|z| > 2.33, 1.65, \text{and } 1.28$, except a two-sided test that is used for the intercept. The initial grid search process described in section 2.2 identified that the optimal solution for γ resided in the interval $[-4.4, -4.2]$. The second grid search within $\gamma \in [-4.4, -4.2]$ at a 0.01 level of accuracy led to an optimum at the values $\hat{\gamma}^* = -4.23$ and associated $\hat{q}^* = 0.8835$.

Table 3 also reports the estimated mean WTP from the Logit model, as well as confidence intervals (CIs) for the compensated WTP measure. Employing Hanneman's approach for estimating the mean WTP, as described above, and the delta method to derive the asymptotic standard errors, the mean Logit WTP measure is $114.74 and the 95% CI based on z-scores ranges from $101.28 to $128.19. We note that the estimated mean WTP obtained from the fully parametric Logit model appears to be relatively high given the typical levels of income of visitors surveyed at the CNF (60% of visitors' annual income is less than $15,000) and the values in the literature for swimming and other non-boating, non-fishing water based recreation (Loomis, 2005), as well as the type of recreation experience obtained by visiting the recreation pools.

4.2 ML-MPD Model Results

Parameter estimates , $\hat{\boldsymbol{\beta}}_{ML-MPD}$, of the $\boldsymbol{\beta}$ vector corresponding to $\hat{\gamma}^* = -4.23$ and $\hat{q}^* = 0.8853$ are presented in Table 3, along with associated asymptotic standard errors derived from the conditional and unconditional variance-covariance matrices for $\hat{\boldsymbol{\beta}}_{ML-MPD}$. Goodness-of-fit measures (McFadden R^2, AIC and BIC) and the mean ML-MPD WTP estimate and its corresponding CIs are also reported in this table.

Based on the goodness-of-fit measures, the ML-MPD model performs better than the Logit model. Regarding parameter estimates, the coefficients for the ML-MPD and Logit models have identical signs, except for the income variable. As mentioned previously, the dichotomous income indicator is positively related to the probability of paying the bid amount in the Logit model, but when using the ML-MPD approach, a negative effect of income arises, suggesting that a visit to the river pools is an inferior good, although the income effect is not statistically significant at any conventional level as in other studies of this type.

The ML-MPD approach does not result in uniformly smaller estimated standard errors relative to Logit, especially in reference to the standard errors derived from the full unconditional covariance matrix (see Table 3). As indicated in section 4.1, the variables that are statistically significant at the 0.01 level of type 1 error using the Logit model are bid, size, and road. The ML-MPD results, based on the *conditional* covariance matrix, for the bid and size regressors are also significant at the 0.01 level, with the road regressor being significant at the 0.05 level. However, these same three variables achieve significance at only the 0.10 level when the *unconditional* covariance matrix is used for the ML-MPD results, which recognizes the uncertainty involved in choosing the appropriate functional form of the probability distribution underlying binary responses. The discharge variable is significant at the 0.05 and 0.10 levels for the Logit and conditional MPD-ML cases, respectively and statistically insignificant at conventional levels in the unconditional ML-MPD case. However, its probability value of 0.12 suggests that its effect should not necessarily be ignored. The travel time indicator variable is significant at the 0.05 level under the Logit estimator, but insignificant under the ML-MPD estimator, regardless of whether conditional or unconditional covariance matrices are utilized.

Regarding the ML-MPD results for q and γ, we conducted a Wald test of the symmetric logistic distribution hypothesis $H_o : q = 0.5$ and $\gamma = 0$, and the null hypothesis is rejected at the 0.01 level. This suggests that a logistic approach is inappropriate for this application. We also tested separately the individual hypotheses of symmetric distributions, $H_o : q = 0.5$, and a general (i.e., possibly skewed) logistic distribution, $H_o : \gamma = 0$. Both hypotheses are also rejected at the 0.01 level.

Finally, Table 3 shows that the mean ML-MPD WTP estimate of $27.46 for a visit to the river recreation sites is substantially *lower* in value than the result estimated with the Logit ($114.74) approach. Moreover, it is apparent from the fact that the 95% CI for the ML-MPD approach ($13.01, $41.91) does not overlap the CI of the Logit estimate that the mean WTP estimated via ML-MPD is statistically significantly smaller than the Logit estimate (at any conventional level of confidence). Given the characteristics of the visitors to the water pools, in particular their low incomes, and the values in the literature for similar recreation experiences, the estimated mean WTP derived from the ML-MPD approach appears to be considerably more plausible.

Lastly, we tested whether there is a difference in WTP *distributions* based on the Logit and ML-MPD approaches (i.e. H_o: $WTP_{Logit} = WTP_{ML\text{-}MPD}$), using the nonparametric *complete combinatorial convolution* approach of Poe *et al.* (2005). The resulting two-tailed p-value of zero (to five decimal places) rejected the null hypothesis very convincingly, indicating that the two empirical WTP distributions are statistically significantly different.

4.3 Comparisons of Marginal Values of River Recreation Attributes

Based on the fact that the Logit model is rejected, marginal effects of changes in the explanatory variables on WTP are calculated here for the ML-MPD model only. In pursuit of this task, marginal values of changes in river recreation attributes are computed by forming the ratio of the value of the coefficient of the explanatory variable associated with the site attribute and the absolute value of the coefficient of the bid variable.

Using the ML-MPD approach, the marginal values of changes in site attributes indicate that visitors would be willing to pay \$33.70 and \$39.90 for decreasing high in-stream flows and to have paved roads, respectively. The change in WTP relating to decreasing the size of rocks or sand becomes \$2.10, whereas the marginal effect associated with the increased volume of water in the pools is \$0.26 per cubic foot of water.

5 CONCLUSIONS AND IMPLICATIONS

In this paper, we present an information theoretic econometric approach to analyzing the willingness to pay for river recreation site attributes, using a new ML-MPD binary response estimator for dichotomous CV data. In contrast to many alternative estimators, the ML-MPD estimator mitigates the need for subjective choices of distributional functional form and tuning parameters. The ML-MPD has the flexibility to fit a wide range of varied distributional shapes to conform to the choice process observed while utilizing only very general (non-parametric) moment assumptions relating to the data. Moreover, the ML-MPD method allows for statistical testing of the symmetry-of-distribution assumption, $q = 0.5$, as well as a test of the validity of the logistic family of distributions, $\gamma = 0$ (either symmetric with $q = 0.5$, or non-symmetric for general q). As such, the ML-MPD approach to estimation appears to represent a useful innovative estimation framework for providing general and testable representations of data-generating processes underlying economic decision-making behavior in binary choice situations.

In addition, in our empirical application, goodness-of-fit measures (McFadden R^2, AIC and BIC) indicate that the ML-MPD model performs better than the Logit model. The ML-MPD model rejects the special case of a Logit model restriction at the 0.01 level of significance. In addition, another empirical finding of this study is that the ML-MPD approach yields a substantially *lower* estimate of the mean WTP ($27.46) for a visit to the "river recreation pools" of water (recreation sites) at the CNF compared to the mean WTP obtained from the fully parametric Logit ($114.74) approach. Based on the nature of the recreational experience and the demographics of typical visitors to the water pools, the *lower* mean WTP value generated by the ML-MPD method appears to be a more reasonable and defensible estimate of the mean WTP for visiting the "water pool" recreation sites. If policymakers were to contemplate entry fees to these recreation sites in the Caribbean National Forest, the *lower* mean WTP estimate of the ML-MPD approach would suggest a substantially more narrow range of fee possibilities than those implied by the Logit model.

Except for income, the effects of all other decision maker or site attributes on WTP were the same in sign using either the Logit or ML-MPD approach. While the two methods were in agreement that the size of the bid, size of rocks or sand, and road quality all significantly impacted the acceptance of the bid, they differed in the magnitude of the impacts. Moreover, the methods differed on the significance of other attributes. The results beg the question of whether the parametric Logit approach is understating uncertainty about model components and overstating significance of estimation results by assuming that the functional form of the probability distribution underlying decision outcomes is fully known and certain. The ML-MPD method has an inherent feature of mitigating unwarranted indications of over-precision by explicitly introducing uncertainty with respect to the functional form of the distribution underlying the binary choice process.

APPENDIX A: Analytic Derivation of the MPD family of CDFs solution

$$\underset{p_i \in \Xi}{\text{Min}} \frac{\sum_{i=1}^{n}\left\{ p_i \cdot \left(\left(\frac{p_i}{q_i} \right)^{\gamma} - 1 \right) + (1-p_i) \cdot \left(\left(\frac{1-p_i}{1-q_i} \right)^{\gamma} - 1 \right) \right\}}{\gamma(\gamma+1)}$$

s.t. $n^{-1}\left(\mathbf{x}'(\mathbf{y}-\mathbf{p})\right) = \mathbf{0}$

$0 \le p_i \le 1$

On the basis of the previous set-up, the Lagrange function associated with the non-linear constrained minimization problem is given by:

$$\Im = \frac{\sum_{i=1}^{n}\left\{ p_i \cdot \left(\left(\frac{p_i}{q_i} \right)^{\gamma} - 1 \right) + (1-p_i) \cdot \left(\left(\frac{1-p_i}{1-q_i} \right)^{\gamma} - 1 \right) \right\}}{\gamma(\gamma+1)} + \boldsymbol{\lambda}'\left[\mathbf{x}'(\mathbf{y}-\mathbf{p})\right] + \varphi(p_i) + \eta(1-p_i)$$

For simplicity we will start ignoring the inequality constraints on the probability value. Solving the first order conditions (FOCs) for p_i for an interior optimum yields:

$$\frac{d\Im}{dp_i} = \frac{1}{\gamma(\gamma+1)} \sum_{i=1}^{n}\left\{ (1+\gamma)\left(\frac{p_i}{q_i} \right)^{\gamma} - 1 - (1+\gamma)\left(\frac{1-p_i}{1-q_i} \right)^{\gamma} + 1 \right\} - \mathbf{x} \cdot \boldsymbol{\lambda} = \mathbf{0}$$

$$= \frac{1}{\gamma(\gamma+1)} \cdot (1+\gamma) \sum_{i=1}^{n}\left\{ \left(\frac{p_i}{q_i} \right)^{\gamma} - \left(\frac{1-p_i}{1-q_i} \right)^{\gamma} \right\} - \mathbf{x} \cdot \boldsymbol{\lambda} = \mathbf{0}$$

$$= \sum_{i=1}^{n}\left\{ \left(\frac{p_i}{q_i} \right)^{\gamma} - \left(\frac{1-p_i}{1-q_i} \right)^{\gamma} \right\} - \mathbf{x} \cdot \boldsymbol{\lambda} \cdot \gamma = \mathbf{0}$$

$$= \left[\left(\frac{p_i}{q_i} \right)^{\gamma} - \left(\frac{1-p_i}{1-q_i} \right)^{\gamma} \right] - \mathbf{x} \cdot \boldsymbol{\lambda} \cdot \gamma = \mathbf{0} \text{ for a particular } p_i \text{ and } \gamma \neq 0$$

For $\gamma = 0$ the CR objective function becomes undefined. Because of that, we apply L'Hôpital's rule to $\Im$. Rewriting the Lagrange function we have:

$$\Im = \frac{p_i \cdot \left(\left(\frac{p_i}{q_i} \right)^{\gamma} - 1 \right) + (1-p_i) \cdot \left(\left(\frac{1-p_i}{1-q_i} \right)^{\gamma} - 1 \right)}{\gamma(\gamma+1)} + \frac{\boldsymbol{\lambda}' \left[\mathbf{x}' (\mathbf{y} - \mathbf{p}) \right] \cdot (\gamma^2 + \gamma)}{\gamma(\gamma+1)}$$

Applying L'Hôpital's rule to $\Im$ and using exponential rules:

$$\frac{d\Im_{\text{numerator}}}{d\gamma} = p_i \left(\frac{p_i}{q_i} \right)^{\gamma} Ln\left(\frac{p_i}{q_i} \right) + (1-p_i) \left(\frac{1-p_i}{1-q_i} \right)^{\gamma} Ln\left(\frac{1-p_i}{1-q_i} \right) + \boldsymbol{\lambda}\left(\mathbf{x}(\mathbf{y} - \mathbf{p}) \right)(2\gamma + 1)$$

$$\frac{d\Im_{\text{denominator}}}{d\gamma} = 2\gamma + 1$$

It follows,

$$\upsilon = \lim_{\gamma \to 0} \frac{p_i \left(\frac{p_i}{q_i} \right)^{\gamma} Ln\left(\frac{p_i}{q_i} \right) + (1-p_i) \left(\frac{1-p_i}{1-q_i} \right)^{\gamma} Ln\left(\frac{1-p_i}{1-q_i} \right) + \boldsymbol{\lambda}\left(\mathbf{x}(\mathbf{y} - \mathbf{p}) \right)(2\gamma + 1)}{2\gamma + 1}$$

$$= p_i Ln\left(\frac{p_i}{q_i} \right) + (1-p_i) Ln\left(\frac{1-p_i}{1-q_i} \right) + \boldsymbol{\lambda}\left(\mathbf{x}(\mathbf{y} - \mathbf{p}) \right)$$

By taking the FOCs on this last result and using the product rule with the chain rule we have:

$$\frac{d\upsilon}{dp_i} = \left[Ln\left(\frac{p_i}{q_i} \right) + p_i \cdot \frac{\partial Ln(p_i / q_i)}{\partial p_i} \right] + \left[-1 \cdot Ln\left(\frac{1-p_i}{1-q_i} \right) + (1-p_i) \cdot \frac{\partial Ln\left((1-p_i)/(1-q_i) \right)}{\partial p_i} \right] + (-\boldsymbol{\lambda}\mathbf{x}) = \mathbf{0}$$

$$= Ln\left(\frac{p_i}{q_i} \right) - Ln\left(\frac{1-p_i}{1-q_i} \right) - \boldsymbol{\lambda}\mathbf{x} = \mathbf{0} \text{ for a particular } p_i \text{ and } \gamma = 0$$

Solving for p_i when $\gamma = 0$ it yields to the following closed-form solution for p_i :

$$Ln\left[\frac{(p_i / q_i)}{\left((1-p_i)/(1-q_i) \right)} \right] = \mathbf{x}\boldsymbol{\lambda} \Rightarrow \frac{(p_i / q_i)}{\left((1-p_i)/(1-q_i) \right)} = e^{\mathbf{x}\boldsymbol{\lambda}} \Rightarrow p_i = \frac{q_i e^{\mathbf{x}\boldsymbol{\lambda}}}{(1-q_i) + q_i e^{\mathbf{x}\boldsymbol{\lambda}}}$$

Now, when including the non-negativity inequality constraints on the probability value, Karush-Kuhn-Tucker conditions as well as their corresponding complementary slackness conditions must be specified. For any feasible solution to this problem we have the following three possibilities for p_i :

i) $p_i = 0$;

ii) $p_i > 0$; and

iii) $p_i = 1$

For $p_i > 0$ we get an interior solution, i.e $p(w_i; q_i, \gamma) = \left[\left(\frac{p_i}{q_i} \right)^{\gamma} - \left(\frac{1-p_i}{1-q_i} \right)^{\gamma} \right] - w_i \cdot \gamma = \mathbf{0}$ for $\gamma > 0$.

Whenever $p_i = 1$ the largest value of the previous bracket expression is $\left(\frac{1}{q_i} \right)^{\gamma}$ so that $\left(\frac{1}{q_i} \right)^{\gamma} \le w_i \gamma \Rightarrow w_i \ge \gamma^{-1} q_i^{-1}$. Conversely, when $p_i = 0$ the smallest value of the bracket expression is equal to - $\left(\frac{1}{1-q_i} \right)^{\gamma} \ge w_i \gamma \Rightarrow w_i \le -\gamma^{-1} (1-q_i)^{-\gamma}$.

APPENDIX B: Gradients of the MPD Distribution for Use in ML Covariance Matrix Calculations

In the derivative expressions below, $\odot$ and 0 denote the *generalized* Hadamard (elementwise) product and division operators, respectively. In addition, a column vector raised to the -1 power indicates an elementwise vector reciprocal operator, and $\mathbf{1}_n$ denotes an $n \times 1$ vector of unit values.

B.1. Derivatives of MPD Distributions

$$\frac{\partial \mathbf{p}(\mathbf{w};q,\gamma)}{\partial \boldsymbol{\beta}} = \mathbf{x} \, 0 \, \mathcal{D}$$

$$\frac{\partial \mathbf{p}(\mathbf{w};q,\gamma)}{\partial q} = \left[\left((\mathbf{1}_n - \mathbf{q})^{-\gamma-1} \odot (\mathbf{1}_n - \mathbf{p})^{\gamma} \right) + \left(\mathbf{q}^{-\gamma-1} \odot \mathbf{p}^{\gamma} \right) \right] \, 0 \, \mathcal{D}$$

$$\frac{\partial \mathbf{p}(\mathbf{w};q,\gamma)}{\partial \gamma} = \left[((\mathbf{1}_n - \mathbf{p}) 0 (\mathbf{1}_n - \mathbf{q}))^{\gamma} \odot Ln((\mathbf{1}_n - \mathbf{p}) 0 (\mathbf{1}_n - \mathbf{q})) - (\mathbf{p} 0 \mathbf{q})^{\gamma} \odot Ln(\mathbf{p} 0 \mathbf{q}) + \mathbf{x}\boldsymbol{\beta} \right] \, 0 \, (\gamma \mathcal{D})$$

where $\mathcal{D} \equiv \left((\mathbf{1}_n - \mathbf{q})^{-\gamma} \odot (\mathbf{1}_n - \mathbf{p})^{\gamma-1} \right) + \left(\mathbf{q}^{-\gamma} \odot \mathbf{p}^{\gamma-1} \right)$ and $\mathbf{w} = \mathbf{x}\boldsymbol{\beta}$.

B.2. ML-MPD Asymptotic Covariance Matrix Definition

As mentioned in section 2, the estimate of the asymptotic covariance matrix of the ML-MPD estimator was based on the OPG approach using

$$Asy\,\mathrm{cov}(\hat{\boldsymbol{\theta}}) = \left[\left(\frac{\partial \mathbf{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)' \left(\frac{\partial \mathbf{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) \right]^{-1} \Bigg|_{\hat{\boldsymbol{\theta}}} ,$$

where $\boldsymbol{\theta} \equiv [\boldsymbol{\beta}' \mid q \mid \gamma]'$, $\hat{\boldsymbol{\theta}}$ is the ML-MPD estimate of the unknown parameters, and

$$\frac{\partial \mathbf{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \left(\mathbf{y} 0\, \mathbf{p} - (\mathbf{1}_n - \mathbf{y}) 0\, (\mathbf{1}_n - \mathbf{p})\right) \odot \left[\frac{\partial \mathbf{p}(\mathbf{w}; q, \gamma)}{\partial \boldsymbol{\beta}} \;\middle|\; \frac{\partial \mathbf{p}(\mathbf{w}; q, \gamma)}{\partial q} \;\middle|\; \frac{\partial \mathbf{p}(\mathbf{w}; q, \gamma)}{\partial \gamma} \right].$$

The derivatives in the above expression are defined in B.1.

REFERENCES

Araña, J. E., and C.J. León. 2005. Flexible mixture distribución modelan of dichotomous choice contingent valuation with heterogenity. *Journal of Environmental Economics and Management*. 50(1): 170 - 188.

Berndt, E.K., Hall, B.H., Hall, R.E. and J.A. Hausman. 1974. Estimation and inference in nonlinear structural models. *Annals of Economic and Social Measurement*. **3**(4): 653 – 665.

Cameron, T.A. and M.D. James. 1987. Estimating Willingness to Pay from Survey Data: An Alternative Pre - Test – Market Evaluation Procedure. *Journal of Marketing Research*. **24**(4): 389 – 395.

Chen, H.Z. and A. Randall. 1997. Semi-nonparametric estimation of binary response models with an application to natural resource valuation. *Journal of Econometrics*. **76**(1-2): 323 – 340.

Chen, S. and S. Khan. 2003. Rates of convergence for estimating regression coefficients in heteroskedastic discrete response models. *Journal of Econometrics*. **117**: 245 – 278.

Cooper, J.C., Hanemann, W.M. and G. Signorello. 2001. One-and-one-half-bound dichotomous choice contingent valuation. CUDARE working paper No. 921. University of California, Berkeley.

Cosslett, S. 1983. Distribution-Free Maximum Likelihood Estimation of the Binary Choice Model. *Econometrica*. **51**: 765 – 782.

Creel, M. and J. Loomis. 1997. Semi-nonparametric Distribution-Free Dichotomous Choice Contingent Valuation. *Journal of Environmental Economics and Management*. **32**: 341 – 358.

Crooker, J.R. and J.A. Herriges. 2004. Parametric and Semi-Nonparametric Estimation of Willingness-to-Pay in the Dichotomous Choice Contingent Valuation Framework. *Environmental and Resource Economics*. **27**: 451 – 480.

Golan, A., Judge, G. and J. Perloff. 1996. A Maximum Entropy Approach to Recovering Information from Multinomial Response Data. *Journal of the American Statistical Association*. **91**: 841 – 853.

Gonzalez-Sepulveda, J.M. 2008. *Challenges and Solutions in Combining RP and SP Data to Value Recreation*. Doctoral Dissertation. Colorado State University. Department of Agricultural and Resource Economics.

Gonzalez, J.M., Loomis, J.B. and A. Gonzalez-Caban. 2008. A Joint Estimation Method to Combine Dichotomous Choice CVM Models with Count Data TCM Models Corrected for Truncation and Endogenous Stratification. *Journal of Agricultural and Applied Economics*. **40**(2): 681 – 695.

Haab, T. and K. McConnell. 2002. *Valuing environmental and natural resources: the econometrics of non-market valuation*. Northampton, MA: Edward Elgar.

Hanemann, W.M. 1984. Welfare evaluations in Contingent valuation Experiments with Discrete Responses. *American Journal of Agricultural Economics*. **66**(3): 332 – 341.

______________. 1989. Welfare evaluations in Contingent valuation Experiments with Discrete Response Data: Reply. *American Journal of Agricultural Economics*. **71**(4): 1057 – 1061.

Horowitz, J.L. 1992. A Smoothed Maximum Score Estimator for the BRM. *Econometrica*. **60**: 505 – 531.

Huang, J., Nychka, D.W. and V.K. Smith. 2008. Semi-parametric discrete choice measures of willingness to pay. *Economics Letters*. **101**: 91 – 94.

Ichimura, H. 1993. Semiparametric least squares (SLS) and weighted SLS estimation of single-index-models. *Journal of Econometrics*. **58**: 71 – 120.

Imbens, G.W., Spady,R.H. and P.Johnson. 1998. Information Theoretic Approaches to Inference in Moment Condition Models. *Econometrica*. **66**(2): 333 – 357.

Jacoby, S.L.S., Kowalik, J.S. and J.T. Pizzo. 1972. *Iterative Methods for Nonlinear Optimization Problems*. Englewood Cliffs, NJ: Prentice Hall.

Judge, G. and R.C. Mittelhammer. 2012. *An Information Theoretic Approach to Econometrics*. Cambridge University Press, Cambridge.

Klein, R.W. and R.H. Spady. 1993. An Efficient Semiparametric Estimator for Binary Response Models. *Econometrica*. **61**(2): 387 – 421.

Kriström, B. 1990. A Non-Parametric Approach to the Estimation of Welfare Measures in Discrete Response valuation Studies. *Land Economics*. **66**(2): 135 – 139.

Kullback, S. 1959. *Information Theory and Statistics*. John Wiley and Sons. NY.

Lee, J., Tan Chao, W.K. and G.G. Judge. 2010. Stigler's approach to recovering the distribution of first significant digits in natural data sets. *Statistics and Probability Letters*. **80**(2): 82 – 88.

Li, C. 1996. Semiparametric Estimation of the Binary Choice Model for Contingent Valuation. *Land Economics*. **72**(4): 462 – 473.

Loomis, J. 2005. Updated Outdoor Recreation Use Value on National Forests and Other Public Lands. General Technical PNW-GTR-658. Pacific Northwest Research Station, USDA Forest Service, Portland, OR.

Manski, C.F. 1975. The Maximum Score Estimation of the Stochastic Utility Model of Choice. *Journal of Econometrics*. **3**: 205 – 228.

Marsh, T.L. and R.C. Mittelhammer. 2004. Generalized Maximum Entropy Estimation of a First Order Spatial Autoregressive Model. Spatial and Spatiotemporal Econometrics. Advances in Econometrics. **18**: 199–234.

Matzkin, R. 1992. Non-Parametric and Distribution-Free Estimation of the Binary Threshold Crossing and the Binary Choice Models. *Econometrica*. **60**: 239 – 270.

McFadden, D. 1994. Contingent Valuation and Social Choice. *American Journal of Agricultural Economics*. **76**: 689 – 708.

McFadden, D. and K. Train. 2000. Mixed MNL Models for Discrete Response. *Journal of Applied Econometrics*. **15**: 447 – 470.

Mittelhammer, R.C., Judge, G.G. and D.J. Miller. 2000. *Econometric Foundations*. Cambridge.

Mittelhammer, R.C. and Judge, G.G. 2011. A Family of Empirical Likelihood Functions and Estimators for the Binary Response Model. *Journal of Econometrics*. **164**(2): 207 - 217.

Nelder, J.A. and R. Mead. 1965. A simplex method for function minimization. *The Computer Journal*. **7**(4): 308 – 313.

Patefield, W.M. 1985. Information From the Maximized Likelihood Function. *Biometrika*. 72(3): 664 – 668.

Poe, G.L., Giraud, K.L. and J.B. Loomis. 2005. Computational Methods For Measuring The Difference of empirical Distributions. *American Journal Agricultural Economics*. **87**(2): 353 - 365.

Read, T.R. and N.A. Cressie. 1988. *Goodness of Fit Statistics for Discrete Multivariate Data*. Springer Verlag. NY.

Shannon, C.E. 1948. A Mathematical Theory of Communication. *Bell System Technical Journal*. **27**: 379 – 423.

Turnbull, B. 1976. The empirical distribution function with arbitrarily grouped, censored and truncated data. *Journal of the Royal Statistical Society*. Series B. **38**: 290 – 295.