A nonparametric multiple choice method within the random utility framework

Ju-Chin Huang\textsuperscript{a,}\textsuperscript{*}, Douglas W. Nychka\textsuperscript{b,}\textsuperscript{c}

\textsuperscript{a}Department of Economics, University of New Hampshire, McConnell Hall, Durham 03824, USA
\textsuperscript{b}National Center for Atmospheric Research, Climate and Global Dynamics Division, Boulder, CO 80307, USA
\textsuperscript{c}Department of Statistics, North Carolina State University, Raleigh, NC 27695, USA

Received 1 August 1997; received in revised form 28 July 1999; accepted 1 August 1999

Abstract

Many researchers use categorical data analysis to recover individual consumption preferences, but the standard discrete choice models require restrictive assumptions. To improve the flexibility of discrete choice data analysis, we propose a nonparametric multiple choice model that applies the penalized likelihood method within the random utility framework. We show that the deterministic component of the random utility function in the model is a cubic smoothing spline function. The method subsumes the conventional conditional logit model (McFadden, 1973, in: Zarembka, P., (Ed.), Frontiers in Econometrics) as a special case. In this paper, we present the model, describe the estimator, provide the computational algorithm of the model, and demonstrate the model by applying it to nonmarket valuation of recreation sites. © 2000 Elsevier Science S.A. All rights reserved.

JEL classification: C14

Keywords: Polychotomous choices; Cubic smoothing splines; Random utility model; Welfare measurement; Nonmarket valuation

*Corresponding author. Tel: (603) 862-3279; fax: (603) 862-3383.
E-mail addresses: jehuang@cisunix.unh.edu (J.-C. Huang), nychka@ucar.edu (D.W. Nychka).
1. Introduction

Following the development of the random utility model (RUM, McFadden, 1973) and given the increased availability of detailed micro-data that describe people's consumption choices, categorical data analysis has become an important tool for recovering individual preferences. In particular, multiple choice analysis based on RUM is now used to analyze market data for the study of individual choice decisions in finance, marketing research, transportation research, occupational choices, and many other fields. In the past twenty years, multiple choice methods have also been applied to nonmarket valuation to examine individuals' consumption choices of nonmarket goods. For example, researchers and policy makers may want to estimate the recreational value of improving quality of a sportfishing site or to derive the welfare loss of closing a recreation site for commercial uses (e.g., Morey, 1981; Bockstael et al., 1991; Morey et al., 1991; Kaoru et al., 1995; Kling and Thomson, 1996). Under certain assumptions, these benefits or losses can be estimated through individuals' recreational choice decisions.

The basic assumption of RUM is that the utility function is unknown but it can be partially recovered by relating an individual's consumption choices to the individual's socio-economic characteristics, the relative prices that he or she faces, and the characteristics of the goods available for consumption. A part of the utility function, however, cannot be recovered by the researcher because of unknown factors about individuals. Hence, the utility function in RUM contains two components: a deterministic component that can be explained by known factors and a random error component that indicates the unknown variation across individuals. Conventional discrete choice models cast within the random utility framework assume a parametric specification for the deterministic component and a distributional assumption for the error component of the random utility function. Since the true preference structure cannot be actually observed, any model entails some degree of misspecification. Because the conventional models impose these parametric assumptions, they tend to increase the degree of misspecification, which frequently leads to biased estimates of marginal values (Hanemann, 1984; Ozuna et al., 1993).

A number of more flexible discrete choice methods within the random utility framework has been developed by applying nonparametric estimation techniques. These variations can be divided into two groups. The first group adopts distribution free estimation methods while keeping a parametric specification for the deterministic component of the random utility function. This method has

---

1 Through December, 1998, the seminal paper by McFadden (1973) has generated over 1000 citations according to the Social Science Citation Index and the Science Citation Index. The total number of citations on the sequence of papers by McFadden on this subject is over 1600.
been applied to binary choice, multiple choice, sequential choice, and contingent ranking models (e.g., Manski, 1975, 1985; Cosslett, 1983; Gallant and Nychka, 1987; Han, 1987; Ichimura and Lee, 1991; Horowitz, 1992; Klein and Spady, 1993; Lee, 1995). Relaxing the distributional assumption on the random component of RUM helps provide a better connection between choices and underlying preference structure, although misspecification of the deterministic component of the utility function can still be present.

The second group of estimation methods generalizes the discrete choice models by relaxing the parametric specification assumptions of the deterministic component of RUM. Most of the studies in this group focus on the binary choice data analysis (Matzkin, 1992; Manski, 1991; Ahn and Manski, 1993; Huang et al., 1999). The two multiple choice applications in this group are Matzkin (1991) and Abe (1999). Ahn (1995) combines the approaches applied in these two groups and proposes a nonparametric binary choice model that uses both a distribution free error structure and a nonparametric deterministic utility function.

In general, the theory of nonparametric binary choice methods is studied more extensively. Although there are a few existing applications, little theory on nonparametric multiple choice methods has been explicitly addressed. In this paper, we examine a nonparametric multiple choice model that assumes a smooth nonparametric deterministic component of the utility function to allow more flexibility in describing the impact of a particular variable on utility. The proposed model fits into the second group of nonparametric generalization of the discrete choice methods. The objective function of this model is a penalized likelihood function that contains a likelihood function and a roughness penalty function to control the smoothness of the nonparametric utility function. When the penalized likelihood function is maximized over the second-order Sobolev function space, we show that the deterministic component of the random utility function is a cubic smoothing spline function. As the penalty of roughness increases, the cubic smoothing spline function approaches linearity. Hence, the proposed method subsumes the conventional conditional logit model as a special case and allows for more flexible fitting of the utility function in analyzing multiple choice data. The method extends the use of the penalized likelihood method in binary choice data analysis by O'Sullivan et al. (1986) and Cox and O'Sullivan (1990). This paper provides the basic theory behind the applications of spline fitting to multiple choice data. In addition, we apply the method to nonmarket valuation of recreation site choices for its need for more flexible estimation of the marginal utility of income.

We begin in Section 2 with a review of RUM in multiple choice applications. Section 3 discusses the theory of splines, develops the nonparametric multiple choice model within the random utility framework, presents the estimator, and describes the computational algorithm. In Section 4, we apply the proposed method to non-market valuation, specifically, to derive the willingness to pay.
(WTP) for maintaining the access to a recreation site. We discuss the WTP estimators implied by both the conventional conditional logit model and the proposed nonparametric multiple choice method. Section 5 provides a case study that estimates the WTP for maintaining access to a recreation site in the Monongahela River Basin in Pennsylvania. Section 6 gives concluding remarks and suggests future research.

2. Review of multiple choice random utility model

Consider a utility maximization for the individual $i$ subject to his budget constraint.

$$\max_{y_{ij}} U(y_{ij}, \ldots, y_{iJ}, q_{i1}, \ldots, q_{iJ}, z_i, a_i) + \sum_{j=1}^{J} y_{ij}e_{ij}$$

subject to

$$I_i = \sum_{j=1}^{J} p_{ij}y_{ij} + z_i$$

where $U(\cdot)$ is the utility function. $y_{ij}$’s indicate the consumption of a group of $J$ goods that $y_{ij} = 1$ if the $j$th good is consumed and $y_{ij} = 0$, otherwise. The consumption of these $J$ goods is assumed to be mutually exclusive; that is, $y_{il}y_{ik} = 0$ for all $l \neq k$. $q_{ij}$ is the quality attribute corresponding to $y_{ij}$. $z_i$ is the numeraire good and $a_i$ is a vector of exogenous variables. $e_{ij}$’s are random variables. The randomness of the utility function represents the unknown characteristics of an individual. $p_{ij}$ is the price or cost of consuming $y_{ij}$.

It is assumed that the partial derivative of the direct utility function with respect to $q_{ij}$ is zero if the $j$th good is not consumed. This is the weak complementarity assumption introduced by Mäler (1974). Under this assumption, an individual does not care about the quality attributes of a good if it is not consumed. Hence, the indirect utility function conditional on the consumption of the $j$th good is

$$U_{ij} = V(q_{ij}, I, p_{ij}, a_i) + e_{ij} = V_{ij} + e_{ij}.$$ 

For the $j$th good to be chosen for consumption, the utility derived from consuming the $j$th good must be the highest. Consequently the unconditional indirect utility function for the individual $i$ is the maximum of the $J$ conditional indirect utility functions, $U_i = \max\{V_{i1} + e_{i1}, \ldots, V_{ij} + e_{ij}\}$.\(^2\) The probability

\(^2\)This expression is in fact a conditional indirect utility function given that one of the $J$ goods is consumed. It is, however, unconditional comparing to the expression of the indirect utility function in (2). In this paper, we assume that one and only one of the $J$ goods is consumed. For an analysis that incorporates no consumption of all $J$ goods, see Morey et al. (1991).
(π_{ij}) that the jth good is consumed by the individual i, given that one of the J goods is consumed, is equivalent to the probability that \( \max\{V_{ik} + \varepsilon_{ik}, \text{for } k = 1, \ldots, J\} = V_{ij} + \varepsilon_{ij} = U_{ij} \). Hence, \( π_{ij} \) is the probability that the Jth order statistic of \( (U_{i1}, U_{i2}, \ldots, U_{iJ}) \) is \( U_{ij} \), for any value of \( U_{ij} \). Assume that \( \varepsilon_{ij} \)'s are independent and identically distributed (i.i.d.) and each follows a type I extreme value distribution with location and scale parameters being zero and one, respectively. It can be shown that probability of consuming good j by the individual i is

\[
π_{ij} = \Pr(U_{(J)} = U_{ij}) = e^{V_j} / \sum_{k=1}^{J} e^{V_k}
\]  

(3)

where \( U_{(J)} \) is the Jth order statistic of \( (U_{i1}, U_{i2}, \ldots, U_{iJ}) \). Once the functional form of the utility function is specified, we can apply the maximum likelihood estimation method to estimate the probabilities and the parameters in the utility function.

\[
\max L = \sum_{i=1}^{n} (y_{i1} \log(π_{i1}) + y_{i2} \log(π_{i2}) + \cdots + y_{iJ} \log(π_{iJ}))
\]

(4)

where \( L \) is the log likelihood function. Notice that the expression of \( π_{ij} \) in (3) is invariant to an additive constant of all \( V \)'s. Hence, the conditional indirect utility function can only be uniquely estimated up to an additive constant. Nevertheless, the welfare measures derived from this model are uniquely defined because the Hicksian measures of welfare changes are invariant to monotonic transformations of the utility function (Mishan, 1977).

3. Nonparametric multiple choice method

The spline functions are a class of piecewise polynomial functions that satisfy continuity properties and therefore are a natural generalization of polynomial functions. They have been found to exhibit desirable characteristics for approximating and interpolating functions (Green and Silverman, 1994). We examine the nonparametric multiple choice model that uses the penalized likelihood

---

3The i.i.d. extreme value distributional assumption of errors implies the hypothesis of independence of irrelevant alternatives (IIA). Under IIA, the ratio of two probabilities (e.g., \( π_{i1}/π_{i2} \)) is independent of the other goods available. When two or more of the J alternatives are close substitutes, the IIA assumption is not plausible. The IIA hypothesis can be lifted by assuming a generalized extreme value distribution for the errors, which allows interdependence across choices. The relaxation of the i.i.d. and IIA assumptions will be considered in future research.
method to incorporate the theory of splines into the conditional logit model. In this section we show the characterization and representation of the estimator, discuss the determination of smoothness of the utility function, and provide the computational algorithm.

3.1. Smoothing splines in analyzing multiple choice data

The penalized likelihood method is initially proposed by Good and Gaskins (1971). Anderson and Blair (1982) apply the estimation method to logistic regression that examines impact of individual characteristics on binary choices. In this paper, we examine a nonparametric multiple choice model within the random utility framework that considers the effect of a choice specific variable on choice decisions. The following objective function is used to derive the optimal nonstochastic component of the utility function in RUM based on multiple choice data.

\[
\max_{V \in W^2_2[a,b]} Q(V) \equiv \max_{V \in W^2_2[a,b]} L(V, x_{11}, \ldots, x_{ij}, \ldots, x_{nJ}) - \lambda \int_a^b (V''(x))^2 \, dx
\]

\[
= \max_{V \in W^2_2[a,b]} \left\{ \sum_{i=1}^n (y_{i1} \log(\pi_{i1}) + y_{i2} \log(\pi_{i2}) + \ldots + y_{ij} \log(\pi_{ij})) - \lambda \int_a^b (V''(x))^2 \, dx \right\}
\]

\[
= \max_{V \in W^2_2[a,b]} \left\{ \sum_{i=1}^n \sum_{j=1}^J \frac{e^{V(x_{ij})}}{\sum_{k=1}^J e^{V(x_{ik})}} \log \left( \sum_{k=1}^J e^{V(x_{ik})} \right) - \lambda \int_a^b (V''(x))^2 \, dx \right\}
\]

\[
= \max_{V \in W^2_2[a,b]} \left\{ \sum_{i=1}^n \sum_{j=1}^J y_{ij} \left[ V(x_{ij}) - \log \left( \sum_{k=1}^J e^{V(x_{ik})} \right) \right] - \lambda \int_a^b (V''(x))^2 \, dx \right\}
\]

Assume that \( V \), the deterministic component of the random utility function, is a function of a choice specific variable, \( x \). The double subscript of \( x \) emphasizes that the value of \( x \) varies with both individuals \( i \) and choices \( j \), and \( a \leq x_{ij} \leq b \). It is assumed that the values of \( x_{ij} \)'s are distinct. The penalized likelihood function \( Q(V) \) consists of two terms. The first term, \( L(\cdot) \), is the log likelihood function for a discrete response model. The second term, \( \int_a^b (V''(x))^2 \, dx \), is the integral of the squared second derivative of \( V(x) \). This is commonly called the roughness penalty function that measures the roughness of the fitted \( V(x) \) curve. The measure increases with the roughness of \( V \).

Instead of assuming a parametric function, \( V \) is assumed to belong to the function space \( W^2_2[a,b] \), where \( W^2_2[a,b] \) is the second-order Sobolev space such
that any function in the space has absolutely continuous first derivative and square integrable second derivative on the interval \([a, b]\). \(\lambda\) is called the smoothing parameter for its ability to determine the degree of smoothness. The value of \(\lambda\) controls the relative weights of the two terms in the penalized likelihood function. If \(\lambda\) is set very large, it imposes a high penalty to roughness of the fitted function. As \(\lambda\) goes to infinity, to maximize the penalized likelihood function the second term must be set to zero, which implies a zero second derivative and a linear function for \(V\). Thus, the proposed nonparametric multiple choice model subsumes the linearity of \(V\) as a special case.

There are various presentations of the solution to (5). A well-known representation employs reproducing kernel Hilbert space theory. Suppose that \(W^2_2\) has a decomposition \(W^2_2 = W_0 \oplus W^2_{2,BC}\), where \(W_0\) is a two dimensional space spanned by the vectors of 1’s and \(x\)’s, \(\{1, x\}\).\(^4\) \(W^2_{2,BC}\) is a subspace of \(W^2_2\) with boundary conditions (BC) such that for any function \(g \in W^2_{2,BC}\), (i) \(g'\) is absolutely continuous; (ii) \(g''\) is squared integrable; (iii) \(g(a) = g(b) = 0\), where \(a\) is the left end point of the range. With the inner product being defined as

\[
\langle g_1, g_2 \rangle = \int g_1'' g_2'' \, dx, W^2_{2,BC} \text{ is a Hilbert space. If } V \in W^2_2, \text{ then there is a unique set of } d_0, d_1, \text{ and } g \in W^2_{2,BC} \text{ such that } d_0 + d_1(x) + g(x) = V(x). \text{ Further, } \max_{v \in W^2_2} Q(V) \text{ is equivalent to } \max_{d_0, d_1, g \in W^2_{2,BC}} Q(d_0 + d_1 x + g).
\]

The basic idea is to first fix \(d_0\) and \(d_1\), then maximize over \(g\). The solution, if a maximum exists, will always have the form \(g(x) = \sum_{i=1}^n \sum_{j=1}^J c_{ij} \rho_{x_i}(x)\), where the functions \(\rho_{x_i}(\cdot)\) are derived from the inner products for \(W^2_{2,BC}\):

\[
\langle \rho_{x_i}, \rho_{x_n} \rangle = \rho_{x_i}(x_{hi}).
\]

It should be noted that while reproducing kernels are associated with a function space, the verification of the formula in (6) can be established simply by integration by parts twice. In particular, the inner products of representors on \(W^2_{2,BC}\) associated with the \(x\)’s are piecewise cubic polynomial functions of the \(x_{ij}\)’s.

\[
\langle \rho_{x_i}, \rho_{x_n} \rangle_{W^2_{2,BC}} = \int \rho_{x_i}'' \rho_{x_n}'' \, dx \quad \text{where } i, h = 1, 2, \ldots, n \text{ and } j, l = 1, \ldots, J
\]

\[
= \begin{cases} 
\frac{x_{ij} x_{hl}}{2} - \frac{x_{ij}^3}{6} & \text{if } x_{hl} > x_{ij}, \\
\frac{\rho_{x_i} x_{hl}^2}{2} - \frac{x_{ij}^3}{6} & \text{if } x_{hl} \leq x_{ij}
\end{cases}
\]

\(^4\) \(W_0 \oplus W^2_{2,BC}\) is the direct sum of \(W_0\) and \(W^2_{2,BC}\). It denotes the set of all functions of the form \(t + g\), where \(t \in W_0\) and \(g \in W^2_{2,BC}\) and \(W_0 \cap W^2_{2,BC} = \{0\}\); that is, \(W_0 \oplus W^2_{2,BC}\). \(W^2_{2,BC} = W_0 \oplus W^2_{2,BC}\) denotes that each \(f\) in \(W^2_2\) can be written in the form of \(f = t + g\), with \(t \in W_0\) and \(g \in W^2_{2,BC}\) in one, and only one, way.
With this form for \( g \), one can then maximize (5) over the linear portion involving \( d_0 \) and \( d_1 \). The following theorem suggests a representation of the solution to (5).

**Theorem.** If there exists a maximizer to (5), then it must have the form

\[
s(x) = d_0 + d_1 x + \sum_{i=1}^{n} \sum_{j=1}^{J} c_{ij} \rho_{x_{ij}}(x) \tag{8}
\]

where \( d_0, d_1 \), and the \( c_{ij} \) are \( nJ + 2 \) unknown constants.\(^5\)

\( s(x) \) is a natural cubic smoothing spline function. That is, \( s(x) \) is a piecewise cubic polynomial with continuous derivatives up to order 2 at \( x \) (the knot) within the data range and is linear beyond the data range. The theorem says that the estimator of \( V(x) \), for any value of \( x \), has the form described in (8). The proof of the Theorem is in the Appendix. Hence, for a fixed \( \lambda \), the maximizer of the penalized likelihood function in (5) is a natural cubic smoothing spline function. The basic approach of continuous spline smoothing is expounded by Wahba (1990). Our model extends her work to discrete choice data analysis.\(^6\)

Let \( s(x_{hl}) \) be the estimator of \( V \) evaluated at \( x_{hl} \). Applying the formula in (6), we have: \( s(x_{hl}) = d_0 + d_1 + \sum_{i=1}^{n} \sum_{j=1}^{J} c_{ij} \rho_{x_{ij}}(x_{hl}) = d_0 + d_1 + \sum_{i=1}^{n} \sum_{j=1}^{J} c_{ij} \rho_{x_{ij}}(x) \). Let \( S(x) \) be the \( nJ \times 1 \) vector of \( \{s(x_{ij})\}, \ i = 1, \ldots, n \) and \( j = 1, \ldots, J \). It can be shown that \( S(x) = T d + Kc \), where \( K \) is an \( nJ \times nJ \) matrix with a typical element \( k_{(ij)(kl)} = \rho_{x_{ij}}(x_{kl}) = \langle \rho_{x_{ij}}(x), \rho_{x_{kl}}(x) \rangle = \int \rho_{ij}'' \rho_{kl}'' \ dx \), which is the inner product of representors defined in (7); \( d \) is the \( 2 \times 1 \) vector \( \{d_0, d_1\} \); \( c \) is the \( nJ \times 1 \) vector \( \{c_{ij}\} \); and \( T \) is the \( nJ \times 2 \) matrix with the first column of ones and the second column of \( x_{ij} \)'s stacked first by \( j \). The \( nJ + 2 \) unknowns \( (d_0, d_1, c_{ij}) \) are just identified by the \( nJ \) observations plus the boundary conditions that restrict the fitted function to be linear outside the data range. The boundary conditions \( s'(a) = s''(a) = s'(b) = s''(b) = 0 \) can be represented by \( T^T c = 0 \) (\( T \) ‘transpose’ \( c = 0 \)), which adds two constraints to the estimation: \( (T_{nJ \times 2})^T c_{nJ \times 1} = 0_{2 \times 1} \).

---

\(^5\)The theorem proceeds under the assumption that a maximizer exists. Due to the highly nonlinear forms in the likelihood, the conditions to ensure a maximizer are not obvious. Also, in conditional logit models, certain restriction must be imposed in order for \( V \) to be identifiable. The identification condition could be set \( d_0 \equiv 0 \). The restriction imposed in the estimation is described in \(^6\).

\(^6\)In defining the set of \( \rho_{x_{ij}} \), it is necessary that all \( \{x_{ij}\} \) be distinct. For defining the basis, repeated values should be dropped. For example, let \( \{x_{\alpha}\} \) denote the unique set of values in the covariates \( \{x_{ij}\} \), then \( s(x) = d_0 + d_1 + \sum_{\alpha} \rho_{x_{\alpha}}(x) \).

\(^7\)Cox and O’Sullivan (1990) provide an error analysis of the binary choice case for the penalized likelihood estimation. Under suitable conditions on \( \lambda \) and sample size they establish existence of the estimator and its consistency. The extension of their results to the multiple choice case should be straightforward but beyond the scope of this paper.
This can be verified by taking the second and the third derivatives of \( s(x) \) with respect to \( x \) and setting them equal to zero. Consequently, the optimization problem in (5) is equivalent to the maximization over a finite collection of basis functions \([T, K]\) to derive \([d, c]\) in \( S(x) \).

3.2. Computation and smoothness of the penalized likelihood estimate

It is assumed that there is one preference structure to characterize individuals’ choice decisions. Hence, there is only one nonparametric function to be estimated in this multiple choice setting. In estimation all \( x_{ij} \)’s are stacked first by choices \( (j) \) then by individuals \( (i) \) to create the value vector of the explanatory variable. The estimation of \( V \) is based on \( nJ \) observed values of \( V(x_{ij}) \) with choice probabilities summed to one for each individual. \( x_{ij} \) can be a characteristic or the price of the good \( j \), or any economic variable that varies across choices. For simplicity, we will use \( V_{ij} \) for \( V(x_{ij}) \) in the rest of this paper.

For computational purpose, we want to rewrite the maximization problem into a minimization problem by examining the first-order conditions of the likelihood function, \( L \). The partial derivatives of the log-likelihood function with respect to \( V_{ij} \) (i.e., \( V \) evaluated at \( x_{ij} \)) are a set of nonlinear functions of \( V_{ij} \).

\[
\frac{\partial L}{\partial V_{ij}} = y_{ij} - \sum_{k=1}^{J} e^{V_{ij}} \left( \sum_{j=1}^{J} y_{ij} \right)
\]

\[
= y_{ij} - \sum_{k=1}^{J} e^{V_{ij}} \quad \text{since} \quad \sum_{j=1}^{J} y_{ij} = 1
\]

\[
= y_{ij} - \pi_{ij}
\]

where \( \pi_{ij} \) is the probability that the individual \( i \) selects the \( j \)th choice. By taking the first order Taylor series expansion at some initial value \( V_{ij}^0 \), expression (9) can be linearized as follows.

\[
\left( y_{ij} - \frac{e^{V_{ij}^0}}{\sum_{k=1}^{J} e^{V_{ij}^0}} \right) - \frac{e^{V_{ij}^0}(e^{V_{ij}^0} + \sum_{k \neq j} e^{V_{ij}^0}) - (e^{V_{ij}^0})^2}{(\sum_{k=1}^{J} e^{V_{ij}^0})^2}(V_{ij} - V_{ij}^0)
\]

\[
= (y_{ij} - \pi_{ij}^0) - \pi_{ij}^0(1 - \pi_{ij}^0)(V_{ij} - V_{ij}^0)
\]

\[
= \left[ \frac{y_{ij} - \pi_{ij}^0}{\pi_{ij}^0(1 - \pi_{ij}^0)} + V_{ij}^0 \right] - V_{ij} \left[ \pi_{ij}^0(1 - \pi_{ij}^0) \right].
\]
Let $z_{ij}$ be the ‘standardized’ $y_{ij}$ plus the initial value $V_{ij}^0$, and $w_{ij}$ be one half times $\pi_{ij}^0(1 - \pi_{ij}^0)$.

$$z_{ij} = \frac{y_{ij} - \pi_{ij}^0}{\pi_{ij}^0(1 - \pi_{ij}^0)} + V_{ij}^0,$$

$$w_{ij} = \frac{1}{2}\pi_{ij}^0(1 - \pi_{ij}^0).$$

The expression in (10) is equivalent to the first derivative of $-\sum_i \sum_j (z_{ij} - V_{ij})^2 w_{ij}$ that can be written in a matrix form: $-(Z - S(x))^TW(Z - S(x))$, where $Z$ is the vector of $\{z_{ij}\}_{nJ \times 1}$, $W$ is the $nJ \times nJ$ matrix with the diagonal elements $\{w_{ij}\}_{nJ \times 1}$ and 0 elsewhere, and $S(x)$ is the estimator of the vector of $\{V_{ij}\}_{nJ \times 1}$. Further, substituting the estimator of $\{V_{ij}\}_{nJ \times 1}$, $S(x) = Td + Kc$, into the roughness penalty function, it can be easily verified that $\left[\int S''(x)\right]^2 dx = \left[\int c_j \rho'' \right]^2 = c^T K c$ (Green and Silverman, 1994, pp. 24–25). Recall that $K$ is an $nJ \times nJ$ matrix of inner products of representors associated with the $x$’s and $c$ is the $nJ \times 1$ vector of $\{c_{ij}\}$. Substituting $[Td + Kc]$ for $\{V_{ij}\}_{nJ \times 1}$ and $c^T K c$ for the roughness penalty function, the maximizer of the optimization problem in (5) can be approximated by the minimizer of the following optimization problem (expressed in a matrix form).

$$\min_{c,d} \sum_{i=1}^n \sum_{j=1}^J (z_{ij} - V_{ij})^2 w_{ij} + \lambda \int V''(x) \, dx$$

$$\equiv \min_{c,d} (Z - Td - Kc)^TW(Z - Td - Kc) + \lambda c^T K c. \quad (12)$$

The expression in (12) can be viewed as a generalization of the ridge regression model. Hence, the smoothing spline estimator is a generalized ridge estimator that belongs to the family of shrinkage estimators. After simple manipulation, the normal equations for minimizing (12) can be presented as follows.

$$Td + (\lambda W^{-1} + K)c = Z,$$

$$T^T c = 0. \quad (13)$$

Notice that for a fixed value of $\lambda$, there are $nJ + 2$ equations to solve $nJ + 2$ unknowns. As discussed in Section 3.1, equations $T^T c = 0$ imply the boundary conditions.

The method can be extended to account for repeated $x$ values, it is computationally more complex due to an additional step in updating $z$ and $w$ in the estimation. Nonetheless, the basic model is the same.

The smoothing parameter, $\lambda$, controls the roughness of the fitted curve. Various methods to determine the optimal value of $\lambda$ have been proposed in the literature (e.g., Hastie and Tibshirani, 1990, pp. 42–52). For its simplicity and
popularity, we choose to use the generalized cross-validation (GCV) as the criterion to determine the optimal $\lambda$ – a data based method proposed by Craven and Wahba (1979). Let $\hat{F}$ be the vector of estimates $\hat{F} = T\hat{d} + K\hat{z} = H(\lambda)\hat{Z}$ for a given $\lambda$. $H(\lambda)$ is the weight matrix commonly called the hat matrix. The GCV function for selecting $\lambda$ in the proposed multiple choice model can be formulated as follows.$^8$

$$GCV(\lambda) = \frac{\sum_{i=1}^{n} \sum_{j=1}^{J} (y_{ij} - \pi_{ij}^\lambda)^2 / w_{ij}^\lambda}{n - \text{Tr}(H(\lambda))},$$ (14)

where $\pi_{ij}^\lambda$ is the estimated probability; $w_{ij}^\lambda$ is the estimated variance; and $\text{Tr}(H(\lambda))$ is the trace of the hat matrix. The optimal $\lambda$ is determined by minimizing the GCV function with respect to $\lambda$.

To compute the estimates for $V(x)$, we suggest a computational algorithm that is equivalent to a combination of the iteratively reweighted ridge regression procedure (O’Sullivan et al., 1986) plus an embedded probability constraint to incorporate the restriction that choice probabilities for each individual must sum to 1. We define $\text{const}_{ij} = \sum_{k \neq j} e^{V_k} \Rightarrow \pi_{ij} = e^{V_j} / (e^{V_j} + \text{const}_{ij})$. The variable $\text{const}_{ij}$ is used to impose the constraint in the estimation/iterations that $\text{const}_{ij} + e^{V_j}$ is a constant across $j$ for the individual $i$. The computational algorithm to estimate $d$ and $c$ is summarized as follows.

(i) Read in $(x_{ij}, y_{ij})$ and stack the data by choices ($j$) then by observations ($i$).
(ii) Create a grid of $\lambda$ values.
(iii) Create pseudo data $z_{ij} = (y_{ij} - \pi_{ij})((\pi_{ij}(1 - \pi_{ij}))/V_{ij} + w_{ij} = \pi_{ij}(1 - \pi_{ij})/2$.
(iv) Estimate $d$ and $c$ for a given value of $\lambda$.
(v) Calculate predicted $\text{const}_{ij}$ and $\pi_{ij}$; calculate predicted $z_{ij}$ and $w_{ij}$.
(vi) Repeat steps (iv) and (v) with $z_{ij}$ and $w_{ij}$ updated each time until $\pi_{ij}$’s converge; calculate $\text{Tr}(H(\lambda))$ and GCV.
(vii) Change $\lambda$ value; repeat steps (iv), (v), and (vi).
(viii) Determine the optimal $\lambda$ according to the minimum GCV.
(ix) Repeat steps (iv), (v), and (vi) with $\lambda = \text{optimal } \lambda$; calculate predicted $V(x_{ij})$ and $\pi_{ij}$.

$^8$The theoretical background for using GCV function to select the optimal $\lambda$ value in the binary choice model can be found in O’Sullivan et al. (1986).

$^9$Notice that $V$ cannot be uniquely identified in the conditional logit models. In estimation a constraint must be imposed to hold the values of $V$ within a range. We restrict the mean of the estimated $V_{ij}$ over $j$ to be zero in each iteration. The constraint dictates the values of the fitted utility function. However, the scaling does not affect the probability and welfare estimation since they are invariant to monotonic transformations of the utility function. Hastie and Tibshirani (1990) suggest a similar restriction in estimating the generalized additive models to resolve the identification issue they called concurvity.
4. Willingness to pay for the accessibility to a good

Multiple choice models have been widely applied in describing the relationship between recreation sites and the site characteristics. Individual choices of recreation sites can be used to recover his value for a particular site. Consider an individual who wants to go to a beach with \( J \) possible sites from which he can choose. Once he decides on a site to visit, he cannot visit the other sites simultaneously and it is reasonable to assume that he does not care about the quality attributes of the other sites. If the random errors are assumed to follow the i.i.d. extreme value distributions, it can be shown that the expected utility of the individual \( i \) who faces \( J \) site choices is

\[
E(\max(V_{i1} + \varepsilon_{i1}, V_{i2} + \varepsilon_{i2}, \ldots, V_{iJ} + \varepsilon_{iJ})) = \ln(\sum_{j=1}^{J} e^{V_{ij}}) + k,
\]

where \( k \) is the Eulers constant \( \approx 0.57 \) (Bockstael et al., 1991). Based on the conditional utility functions for all site choices, we can derive the willingness to pay (WTP) for reserving the access (or willingness to accept (WTA) for the loss of access) to a recreation site by equating the expected utilities from two states: one state with access to the site and one without. If the conditional utility function is assumed quasi-linear in income, the welfare measure of the availability of site \( t \) has a convenient analytical solution as follows (Bockstael et al., 1991).

\[
C^\gamma_t = \frac{\ln(\sum_{j=1}^{J} e^{V_{ij}}) - \ln(\sum_{j=1, j \neq t}^{J} e^{V_{ij}})}{\gamma}
\]

where \( \gamma \) is the fixed coefficient of income in the conditional utility function; i.e., the constant marginal utility of income. If the price of consuming a site choice is treated as a reduction of income, then the coefficient of the price is \( -\gamma \). The welfare measure \( C^\gamma_t \) is derived under the assumption of constant marginal utility of income – a rather restrictive assumption. If the utility function is allowed to be nonlinear in income, the estimation of the utility function is more flexible that might lead to better welfare estimate. However, the calculation of \( C^\gamma_t \) from a nonlinear (or nonparametric) utility function is more difficult – a situation discussed in Bockstael et al. (1991)\(^{10}\). Certain approximation is needed. With

---

\(^{10}\)If the conditional indirect utility function is not quasi-linear in income, the loss (WTA or compensating variation) of eliminating an alternative and the gain (WTP or equivalent variation) of maintaining an alternative will differ. The WTP for maintaining alternative \( t \) is derived by solving

\[
\sum_{j=1}^{J} e^{V(p, q, J - \text{WTP})} = \sum_{j=1, j \neq t}^{J} e^{V(p, q, J)}.
\]

and the WTA for losing alternative \( t \) is solved from

\[
\sum_{j=1, j \neq t}^{J} e^{V(p, q, J + \text{WTA})} = \sum_{j=1}^{J} e^{V(p, q, J)}.
\]
The income information is not available in the data set for the application. The assumption we make is that the derivative of \( V \) with respect to price at a given price indicates the negative marginal utility of income at that particular price since the price of a trip could be thought as a reduction of income.

Alternatively the marginal utility of income at site \( t \) for the individual \( i \) can be estimated by the weighted average of \( \frac{\partial V}{\partial x_{it}} \) over all \( J \) sites. Two possible measures of average marginal utility of income are considered in Huang (1994). The first one is the negative sum of \( \frac{\partial V}{\partial x_{ij}} \) weighted by the probabilities, \( \frac{n_{ij}}{n_{it}} \). Both \( n_{ij} \) and \( \frac{\partial V}{\partial x_{ij}} \) can be estimated from the nonparametric multiple choice model. This measure is also suggested by one of the reviewers. Since survey respondents took multiple trips to the same site within the given time period, the second measure of the average marginal utility of income is to use the actual proportions of trips taken to various sites as weights:

\[
C_{it} = \frac{\ln(\sum_{j=1}^{J} e^{V_{ij}}) - \ln(\sum_{j=1, j \neq t}^{J} e^{V_{ij}})}{-V'_{it}}.
\]

It must be emphasized that since the welfare measure in (16) is an approximation, not an exact measure, the error of approximation is embedded in the nonparametric welfare estimates and cannot be separated from the estimation errors.

5. An example of nonparametric nonmarket valuation

This section applies the nonparametric multiple choice estimator to valuation of recreation sites. The application employs the well known data collected by Smith and Desvousges (1986). The survey, conducted in five counties of the

---

11 The income information is not available in the data set for the application. The assumption we make is that the derivative of \( V \) with respect to price at a given price indicates the negative marginal utility of income at that particular price since the price of a trip could be thought as a reduction of income.

12 Alternatively the marginal utility of income at site \( t \) for the individual \( i \) can be estimated by the weighted average of \( -V'_{it} \) over all \( J \) sites. Two possible measures of average marginal utility of income are considered in Huang (1994). The first one is the negative sum of \( V'_{ij} \) weighted by the probabilities, \( -\sum_{j=1}^{J} \pi_{ij} V'_{ij} \). Both \( \pi_{ij} \) and \( V'_{ij} \) can be estimated from the nonparametric multiple choice model. This measure is also suggested by one of the reviewers. Since survey respondents took multiple trips to the same site within the given time period, the second measure of the average marginal utility of income is to use the actual proportions of trips taken to various sites as weights:

\[
-\left(\frac{\sum_{j=1}^{J} n_{ij} V'(x_{ij})}{\sum_{j=1}^{J} n_{ij}}\right) / \left(\frac{\sum_{j=1}^{J} n_{ij}}{n_{it}}\right),
\]

where \( n_{ij} \) is the number of trips taken to site \( j \). A comparison of the nonparametric welfare measures based on alternative estimates of marginal utility of income can be found in Huang (1994).

13 A Monte-Carlo simulation is conducted to compare the two welfare measures implied by the nonparametric and parametric multiple choice models, respectively. Under three assumed true utility functions – two simple random utility models (taken from Morey et al., 1993) and one share model (based on Morey, 1981) – the simulation results show no significant statistical difference between the two welfare estimators. The details of sampling experiments and response surface analysis of mean squared errors of the welfare estimates are available upon request from the first author.
It is common, although not always correct, to assume that the decision of each of the multiple trips taken by the same individual/household is independent of other trips. In this application, we maintain the assumption of independent decisions of multiple trips by the same household.

Monongahela River Basin in Pennsylvania in 1981, concerns households' recreational decisions in the Basin. A total of 397 households was randomly selected for in-person interviews. The number of fully completed interviews is 303. A subset of 112 households that are users of seven selected recreation sites is used in this application. The seven sites are: Kittanning, Keystone Dam, Lake Arthur in Moraine State Park, Ohiopyle State Park, North Park Lake, Youghiogheny River Lake Reservoir, and Pittsburgh (The Point, Smithfield Bridge). Multiple trips to the recreation sites by the same household in the season will be stacked and treated as independent visits. The total number of trips to all sites used in estimation is 945. The data set provides information on costs of visiting various sites, fish catch rates at the sites, number of trips taken by individuals, and socio-economic characteristics of the interviewed individuals.\textsuperscript{14}

The regressor employed in the empirical models is the cost (price) of visiting a recreation site. The per trip WTP for maintaining the accessibility of the site near Smithfield Bridge is calculated from both parametric and nonparametric multiple choice models. Table 1 summarizes the estimated per trip benefits of the recreation site (averaged across the sample of individuals) from both parametric and nonparametric models. In addition to the nonparametric WTP measure based on the optimal $\lambda$ value ($\log\lambda = -11$), the estimates under three fixed $\lambda$ values are also presented for comparison. It appears that in general WTP estimates increase with $\lambda$. Fitting parametric logit model is equivalent to setting the smoothing parameter to infinity in the proposed nonparametric model. The optimal $\lambda$ value is much smaller than infinity. Based on the bootstrapped confidence intervals, the welfare measure derived from the parametric conditional logit model is significantly higher than the estimate from the nonparametric multiple choice model when the optimal value of $\lambda$ is used. Notice that the resampling distribution for the parametric benefit estimate appears to be symmetric with small standard deviation, while the resampling distribution of the estimate from the nonparametric model is skewed to the right with larger standard deviation. The larger variance of estimates from the nonparametric model is induced by allowing the variation of smoothing parameter values in the repetitions. The statistically significant difference of WTP estimates derived from the parametric and nonparametric models suggests that under the simple single-variable model, the (linear) conditional logit model might not be the best choice for approximating the underlying preference structure in this application. It must be emphasized that the estimation error induced by the approximation in the formula (16) for calculating the nonparametric WTP may also contribute

\textsuperscript{14} It is common, although not always correct, to assume that the decision of each of the multiple trips taken by the same individual/household is independent of other trips. In this application, we maintain the assumption of independent decisions of multiple trips by the same household.
Table 1
Point estimates and bootstrapped results of the average per trip WTP

<table>
<thead>
<tr>
<th></th>
<th>Logit</th>
<th>Spline</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>opt. log(λ)*</td>
<td>log(λ) = 5</td>
</tr>
<tr>
<td>Point estimate</td>
<td>0.3722</td>
<td>0.3721</td>
</tr>
<tr>
<td>Mean*</td>
<td>0.3725</td>
<td>0.3725</td>
</tr>
<tr>
<td>STD</td>
<td>0.0083</td>
<td>0.0085</td>
</tr>
<tr>
<td>Median</td>
<td>0.3722</td>
<td>0.3724</td>
</tr>
<tr>
<td>Lower bound b</td>
<td>0.3874</td>
<td>0.3865</td>
</tr>
<tr>
<td>Upper bound</td>
<td>0.3600</td>
<td>0.3584</td>
</tr>
</tbody>
</table>

*It is the mean of the 500 bootstrapped average benefit estimates.

The lower and upper bounds are created based on Efron’s (1982) percentile method. They form a 90% confidence interval for each of the benefit estimates.

The “log” here is the natural logarithm. The optimal log(λ) selected by the GCV function for the point estimate is -11. λ varies according to the GCV criterium in the bootstrap when deriving the confidence interval. However, the λ value is fixed in resamplings in the next three columns.

to the difference between the parametric and nonparametric welfare measures. Further investigation is needed. Nonetheless, this simple exercise demonstrates that the proposed nonparametric multiple choice model allows more flexibility in estimating WTP and provides a test for linear approximation of the utility function.

6. Summary and remarks

In this paper we have demonstrated the use of the nonparametric multiple choice method within the random utility framework by employing the penalized likelihood estimation method. We have applied the method to nonmarket valuation to value recreation sites. The estimation results indicate that the benefit estimate derived from the nonparametric multiple choice method is significantly different from the estimate from the conventional conditional logit model; that is, the optimal smoothing parameter is finite and does not approach infinity. In this simple application, we show that a linear utility function might not be a good approximation for the true preference structure. Given that we do not know the true underlying utility function and specification errors are present, the nonparametric multiple choice model allows more flexible estimation. The advantage of the proposed nonparametric multiple choice method is its flexibility. However, for policy makers or those who prefer analyzing marginal effects parametrically, it is less intuitive how to derive ‘parameter’ estimates of interest from a nonparametric model. Since the proposed nonparametric model subsumes the conventional conditional logit model as its
special case so it can be used to re-examine and select the most plausible parametric model to derive the important parameter estimates that influence policy decisions.

The model can be extended to include multiple covariates. For example, with additional assumptions on the relationship of covariates, methods such as generalized additive models (Hastie and Tibshirani, 1987, 1990) and semiparametric generalized linear models (Green and Silverman, 1994) can be incorporated to handle multivariate situations. Based on our results, the theory and derivation of the estimators for these models should be straightforward.

The i.i.d. assumption of errors and the IIA hypothesis can be relaxed by exploring more generalized statistical distributions such as a generalized extreme value distribution that allows nonconstant variances and interdependence across errors. There are other extensions. We may, for example, examine through sampling experiments the performance of the nonparametric discrete choice method under highly nonlinear behavioral models. We can also extend the method to multivariate cases as well as explore the asymptotic and small sample properties of the nonparametric welfare estimators. Nonparametric functions can also be used to analyze more complex choice settings such as nested and ranked data that are now commonly elicited in surveys. Indeed, the potential of nonparametric multiple choice methods is unlimited. The properties of the estimators and economic implications of these models must be further explored so the flexibility of nonparametric function estimation can be fully appreciated.

Appendix A

Proof of Theorem: As seen in Section 3.1, by the properties of reproducing kernels, \( l_{ij}(V) = \langle \rho_{x_i}, V \rangle = f(V_{ij}) \). The optimization in (5) can be presented in the following simple form.

\[
\max_{V\in W^2_1} Q(V) = \max_{V\in W^2_1} \left( L(l_{11}(V), \ldots, l_{ij}(V), \ldots, l_{nj}(V)) - \lambda \int_a^b (V''(x))^2 \, dx \right) \quad (A.1)
\]

The Hilbert space projection theorem (Taylor and Lay, 1986) starts with the assumption that \( M \) is a closed subspace of the Hilbert space. If \( M \in H \) (in our case, \( H \equiv W^2_2 \)), define the space perpendicular to \( M \) by \( M^\perp = \{ w \in H, \langle w, z \rangle = 0 \ \forall z \in M \} \). For any \( x \in H \), can be decomposed uniquely as \( x = z + w \) such that \( z \in M, w \in M^\perp \). Based on the projection theorem and the decomposition of \( W^2_2 \) to \( W_0 ^2 \) and \( W^2_{2,BC} \), let the maximizer of (5) be of the form \( \hat{V} = s + v = d_0 + d_1 x + \sum_{i=1}^n \sum_{j=1}^J c_{ij} \rho_{x_{ij}}(x) + v \), where \( s \) is as defined in (8) and \( v \) is some element in \( W^2_{2,BC} \) that is perpendicular to \( 1, x, \rho_{x_1}, \ldots, \rho_{x_n} \); that
is, \( < v, \rho_{x_i} > = 0 \) and \( < v, d_0 + d_1 x > \gtrless \int_0^b d_2 (d_0 + d_1 x) dx^2 v' dt = 0 \). Substituting \( \hat{V} \) into (A1), we have

\[
Q(\hat{V}) = L(l_{11}(\hat{V}), \ldots, l_{nJ}(\hat{V})) - \int_a^b (\hat{V}''')^2 dx
\]

\[
= L(l_{11}(\hat{V}), \ldots, l_{nJ}(\hat{V})) - \lambda < \hat{V}, \hat{V} >
\]

\[
= L(l_{11}(s + v), \ldots, l_{nJ}(s + v) - \lambda < s + v, s + v >
\]

\[
= L(<\rho_{11}, s + v>, \ldots, <\rho_{nJ}, s + v >) - \lambda [ < s, s > + 2 < s, v > + < v, v > ]
\]

\[
= L(<\rho_{11}, s > + <\rho_{11}, v >, \ldots, <\rho_{nJ}, s > + <\rho_{nJ}, v >)
\]

\[
- \lambda < s, s > - 2\lambda < s, v > - \lambda < v, v >
\]

\[
= L(<\rho_{11}, s >, \ldots, <\rho_{nJ}, s >) - \lambda < s, s > - \lambda < v, v >
\]

\[
= L(l_{11}(s), \ldots, l_{nJ}(s)) - \lambda \int (s''')^2 dx - \lambda < v, v >
\]

\[
= Q(s) - \lambda < v, v > \leq Q(s) \quad \cdot \quad < v, v > \geq 0
\]

\[
\Rightarrow Q(\hat{V}) \leq Q(s).
\]

References


