ELSEVIER

Economics Letters 60 (1998) 255–261

economics letters

Testing for rationality: the case of discrete choice data

Luis C. Nunes*, Maria Cunha-e-Sá, Maria Ducla-Soares

Faculdade de Economia, Universidade Nova de Lisboa, Travessa Estevão Pinto, 1070 Lisboa, Portugal

Received 29 December 1997; accepted 7 May 1998

Abstract

When working with micro data on consumer demand, there are many different situations where decisions involve only discrete choices. In this context, conditions under which an underlying rational preference structure exists are derived. Moreover, by introducing flexibility into the model, it is possible to identify nonrational behavior in the sample.

Keywords: Discrete choice data; Conditional indirect utility function; Rationality

JEL classification: 022; 721

1. Introduction

When working with micro data on consumer demand, there are many different situations where decisions involve both continuous and discrete choices. In particular, in many different situations decisions are taken sequentially, in two steps. While in the first step the decision is discrete, in the second step it can be either discrete or continuous. When the consumer decides which brand to select of a particular good, his choice is discrete, as the consumer prefers to select only one brand at any time. However, after the brand is selected, the choice regards the number of units he will choose to buy. Therefore, the choice is now continuous. In contrast, there are cases in which the second choice decision is discrete as well. For instance, when the choice is over different brands of an indivisible durable good, the second choice is also discrete. Other examples of purely qualitative choices can be found on transportation mode choice literature, as well as on the demand for environmental quality literature. Discrete/Discrete choice models are particularly well suited for dealing with those issues.

Given the nature of the problem, once the first step choice is made, utility only depends on the price of the chosen alternative, and does not depend on the prices of the remaining ones. This fact determines a very particular structure for the unconditional indirect utility function, that is, piecewise differentiable. The conditional indirect utility functions are the crucial building blocks for the unconditional indirect one.

It is well established that a utility theoretic framework can be used to develop statistical models

*Corresponding author. Tel.: +351-1-3833624; Fax: +351-1-3886073; E-mail: lcnunes@fe.unl.pt

0165-1765/98/$19.00
PII: S0165-1765(98)00135-9

suitable for the analysis of discrete choices. The purpose of this paper is twofold. First, testable conditions for the existence of an underlying rational preference structure are derived, in this particular context. Second, it is shown how a flexible enough specification allows the researcher to identify nonrational individual behavior in the sample.

It is shown in Proposition 1 that if each of the conditional indirect utility functions is well behaved, given the maximizing behavior of the consumers, the unconditional function is well behaved too. Therefore, in order to test for the existence of an underlying rational preference structure, one only needs to focus on each conditional indirect utility function. It also shown in Proposition 2 that the conditions to be tested are greatly simplified in this context, since the conditional indirect utility functions are defined in the real line. In particular, monotonicity with respect to prices is a necessary and sufficient condition for the existence of a well-behaved unconditional indirect utility function.

When the error terms enter additively, the existence of a well-behaved indirect utility function is either globally rejected or not rejected. In contrast to the additive error specification's results,[1] by allowing flexibility, in particular, considering the Random Coefficients Model, it is shown that it can be the case that for some observations, that is, for some individuals, a well-behaved indirect utility function exists, while for others it does not. Instead of assuming a priori rational individual choice behavior, a way to identify nonrational behavior in the sample is provided.

2. The discrete/discrete random utility model

Following Hanemann (1982), (1984), in the Budget-Constrained Random Utility Qualitative Choice Model the consumer is assumed to maximize utility $u(\cdot) = u(x, q, z, \epsilon)$ defined over $x = (x_1, x_2, \ldots, x_n)$, and z, where x_j represents the quantity of good j, for $j = 1, \ldots, n$, and z is the numeraire. Moreover, the utility function depends also on attributes of the x's, denoted by $q = (q_1, q_2, \ldots, q_n)$. The utility function is assumed to satisfy weak complementarity, $x_j = 0 \rightarrow (\partial u / \partial q_j) = 0$, for $j = 1, \ldots, n$, that is, the attributes of a good do not matter unless that good is actually consumed. For the individual consumer $\epsilon = (\epsilon_1, \ldots, \epsilon_n)$ is a set of fixed constants (or functions), but for the researcher it is a set of random variables with some joint c.d.f. $F_\epsilon(\epsilon_1, \ldots, \epsilon_n)$ which induces a distribution on $u(.)$.

Focusing on the purely qualitative utility maximizing choice case, that is, assuming that the quantities of the x_j's are fixed, the corresponding unconditional consumer's problem can be stated as follows:

$$\max_{\{x,z\}} u(x, q, z, \epsilon)$$

s.t.

$$\sum_{j=1}^{n} p_j x_j + z = y$$

$$x_i x_j = 0, \forall i \neq j$$

$$x_j = \bar{x}_j \text{ } or \text{ } 0, \ j = 1, \ldots n$$

$$x_j \geq 0, z \geq 0.$$

[1]See Hanemann (Hanemann, 1982, 1984).

Assuming that the consumer has selected good j, his utility conditional on this decision will be denoted by $\bar{u}_j(.)$. It follows that

$$\bar{u}_j = u(0,\ldots,0,\bar{x}_j,\ldots,0,q_1,\ldots,q_n,y-p_j\bar{x}_j,\epsilon) = u_j(\bar{x}_j,\bar{q}_j,z,\epsilon), \text{for } j=1,\ldots,n.$$

The resulting conditional ordinary demand functions are given by $x_j(.)=\bar{x}_j$ and $z=z(p_j, q_j, y, \epsilon)=y-p_j\bar{x}_j$. The corresponding conditional indirect utility function is

$$\bar{v}_j(q_j, y-p_j\bar{x}_j, \epsilon) \equiv u_j(\bar{x}_j, q_j z(p_j, q_j, y, \epsilon), \epsilon),$$

where $(\bar{v}_1(.),\ldots,\bar{v}_n(.))$ represent the conditional indirect utility functions. Each $\bar{v}_j(.)$, for $j=1,\ldots,n$, is increasing in $(y-p_j\bar{x}_j)$.

The discrete choice can be represented by a set of binary valued indices $\delta_1,\ldots\delta_n$, where $\delta_j=1$ if $x_j>0$, and $\delta_j=0$ if $x_j=0$. The choice can be expressed in terms of the conditional indirect utility functions as follows:

$$\delta_j(p,q,y,\epsilon) = \begin{cases} 1 \text{ if} & \bar{v}_j(q_j, y-p_j\bar{x}_j, \epsilon) \geq \bar{v}_i(q_i, y-p_i\bar{x}_i, \epsilon) \quad \forall i \\ 0 & \text{otherwise} \end{cases}$$

is a Bernoulli random variable with mean $E\{\partial_j\} \equiv \pi_j$, given by

$$\pi_j = Pr\{\bar{v}_j(q_j, y-p_j\bar{x}_j, \epsilon) \geq \bar{v}_i(q_i, y-p_i\bar{x}_i, \epsilon), \forall i \neq j\}$$

which can be expressed in terms of the joint c.d.f. of the $(n-1)$ differences of the random terms.

By establishing the relationship between the unconditional utility maximization problem and the conditional one, the unconditional indirect utility function is obtained as follows:

$$\bar{v}(.) = \bar{v}(p,q,y,\epsilon) = \text{Max}\,\{\bar{v}_1(q_1, y-p_1\bar{x}_1, \epsilon),\ldots,\bar{v}_n(q_n, y-p_n\bar{x}_n, \epsilon)\},$$

where $\bar{v}_j(.)$, $j=1,\ldots,n$, represents the utility derived conditional on the choice of alternative j. Therefore, $\bar{v}(p,q,y,\epsilon)$ is the utility attained by the individual maximizing consumer when confronted with the choice set (p,q,y). This is a known number for the consumer but for the researcher it is a random variable with a c.d.f. obtainable from the assumed distribution for the random terms, F_ϵ.

3. Testing for rationality

In the previous section, the general structure of a random utility pure discrete choice model was described. As it is shown below, the conditional indirect utility functions $\bar{v}_j(.)$, for $j=1,\ldots,n$, are the crucial building blocks for the unconditional indirect utility function.

3.1. Basic results

Proposition 1. *Given that each $\bar{v}_j(q_j, y-p_j\bar{x}_j, \epsilon)$, for all $j=1,\ldots,n$, is a conditional indirect utility function, and that*

$$\bar{v}(.) = \bar{v}(p,q,y,\epsilon) = \text{Max}\{\bar{v}_1(q_1, y-p_1\bar{x}_1, \epsilon),\ldots,\bar{v}_n(q_n, y-p_n\bar{x}_n, \epsilon)\},$$

then, $\bar{v}(.)$ is also an indirect utility function.

Proof. If $\bar{v}(.)$ is an indirect utility function, then it is (i) continuous at all $p>>0$, $y>0$, (ii) nonincreasing in p and nondecreasing in y, (iii) quasi-convex in p, that is,

$$\bar{v}(\lambda p+(1-\lambda)p', q, y, \epsilon)\leq \text{Max}\{\bar{v}(q, y-p\bar{x}, \epsilon), v(\bar{q}, y-p'\bar{x}, \epsilon)\}\text{ for all }j=1,\ldots,n,$$

and $0<\lambda<1$, and (iv) homogeneous of degree zero in (p, y).

(i) Since each of the $\bar{v}_j(q_j, y-p_j\bar{x}_j\epsilon)$, for all $j=1,\ldots,n$, is continuous so it is $\bar{v}(.)$. This follows from the continuity of the composite of continuous functions.

(ii) By definition,

$$v(p, q, y, \epsilon)=\text{Max}\{\bar{v}_1(q_1, y-p_1\bar{x}_1, \epsilon),\ldots,\bar{v}_n(q_n, y-p_n\bar{x}_n, \epsilon)\}.$$

Given (i), as $\bar{v}_j(q_j, y-p_j\bar{x}_j, \epsilon)$ is nonincreasing in p_j and nondecreasing in y, for all $j=1,\ldots,n$, the properties of the Max operator allow us to extend these results to the unconditional indirect utility function $\bar{v}(p, q, y, \epsilon)$.

(iii) Since each $\bar{v}_j(q_j, y-p_j\bar{x}_j\epsilon)$, for all $j=1,\ldots,n$, is quasi-convex in p, so it is $\bar{v}(p, q, y, \epsilon)$ (see Appendix 1).

(iv) As $\bar{v}_j(q_j, y-p_j\bar{x}_j, \epsilon)$ is homogeneous of degree zero in (p_j, y), for all $j=1,\ldots,n$, by inspection the result follows, that is, $\bar{v}(p, q, y, \epsilon)$ is homogeneous of degree zero in (p, y). □

By Proposition 1 checking on the properties of each conditional indirect utility function, that is, on each $\bar{v}_j(q_j, y-p_j\bar{x}_j, \epsilon)$, for all $j=1,\ldots,n$, is enough. However, for this particular case, the results are greatly simplified, as stated in Proposition 2.

Proposition 2. *As the conditional indirect utility functions are defined in the real line, and given Proposition 1, a necessary and sufficient condition for $\bar{v}(p, q, y, \epsilon)$ to be an indirect utility function is that*:

$$\frac{\partial\bar{v}_j(q_j, y-p_j\bar{x}_j, \epsilon)}{\partial p_j}\leq 0,\text{ for all }j=1,\ldots,n.$$

Proof. As the conditional indirect utility functions are defined in the real line, if they are monotonic in the respective price, then, they are quasi-convex (and quasi-concave). As, by Proposition 1 they are homogeneous of degree zero in price and income, then monotonicity in the respective prices is a sufficient condition for each conditional function to be an indirect utility function. On the other hand, given Proposition 1, if each conditional indirect utility function is well-behaved, then it is nonincreasing in prices. □

3.2. Flexible functional forms

Typically, the majority of the empirical studies have postulated a linear function for each conditional indirect utility function $\bar{v}_j(.)$,[2] and an additive error term ϵ_j, as follows:

[2]See Braden and Kolstad (1991).

$$\bar{v}_j(q_j, y - p_j\bar{x}_j, \epsilon_j) = h_j(q_j) - \alpha_j \log\left(\frac{p_j\bar{x}_j}{y}\right) + \epsilon_j.$$

In this case, the simplest and most common assumptions are that either the errors are Weibull distributed, resulting in a logit model, or normal distributed, resulting in a probit model.

For $\bar{x}_j = 1$, it follows from Proposition 2 that the corresponding condition to be tested in this case is:

$$\frac{\partial \bar{v}_j(q_j, y - p_j\bar{x}_j, \epsilon_j)}{\partial p_j} \leq 0 \Leftrightarrow \frac{-\alpha_j}{p_j} \leq 0 \Leftrightarrow \frac{\alpha_j}{p_j} \geq 0,\ j = 1, \ldots, n.$$

Thus, the conditions required are limited to the sign of α_j, which has to be nonnegative. If this is the case, the data supports the existence of a well-behaved indirect utility function for the whole sample.

More flexible random utility models can be generated, amongst others, by (i) allowing the random elements to enter the conditional indirect utility functions in different ways, and (ii) allowing variability in the parameters (Random Coefficients Model).

Focusing only on (ii),[3] consider the case in which there is parameter variation among individuals, that is, a Random Coefficients Model specification. One possible source of parameter heterogeneity has been viewed in the literature as due to stochastic variation. In this case, the parameter α_j, for $j = 1, \ldots, n$, is not constant across individuals.

Letting α_{ij} represent the parameter for individual i and site j, the corresponding individual conditional indirect utility function is as follows:

$$\bar{v}_{ij}(q_j, y - p_j\bar{x}_j, \epsilon_{ij}) = h_j(q_j) - \alpha_{ij} \log\left(\frac{p_j\bar{x}_j}{y}\right) + \epsilon_{ij}.$$

For $\bar{x}_j = 1$, it follows from Proposition 2 that the condition to be tested in this more flexible case is:

$$\frac{\partial \bar{v}_{ij}(q_j, y - p_j\bar{x}_j, \epsilon_{ij})}{\partial p_j} \leq 0 \Leftrightarrow \frac{-\alpha_{ij}}{p_j} \leq 0 \Leftrightarrow \frac{\alpha_{ij}}{p_j} \geq 0,\ j = 1, \ldots, n,\ i = 1, \ldots, m.$$

Therefore, with a more flexible form for the conditional indirect utility function it can be the case that for some observations, that is, for some individuals, a well-behaved function exists, while for others it does not, depending on the sign of α_{ij}. Thus, nonrational behavior can be identified in the sample.

4. Conclusions

One of the goals of applied demand analysis is to make welfare evaluations for policy purposes. When making these welfare evaluations, based on the data available, the researcher relies on the neoclassical theory of preferences. In this context, the question to be addressed is how to legitimate the use of the data in order to have welfare significance. Focusing on the case of discrete choice data,

[3] It is easy to derive similar conditions for case (i).

the conditions under which an underlying rational preference exists are derived. These conditions are shown to be greatly simplified in this case. Moreover, by introducing flexibility into the model, a way to identify nonrational behavior in the sample is provided. This is in contrast with previous work in the literature, as, typically, rational behavior is assumed a priori.

Acknowledgements

Fellowship from Faculdade de Economia, Universidade Nova de Lisboa is gratefully acknowledged.

Appendix 1

Proof of (iii) in Proposition 1

(iii) We want to show that, for $p=(p_1,\ldots,p_j,\ldots,p_n)$ and $p'=(p'_1,\ldots,p'_j,\ldots,p'_n)$,

$$\bar{v}(\lambda p+(1-\lambda)p', q, y, \epsilon)\leq \mathrm{Max}\{\bar{v}(p, q, y, \epsilon), \bar{v}(p' q, y, \epsilon)\} \text{ for } 0<\lambda<1.$$

As by definition,

$$\bar{v}(p, q, y, \epsilon)=\mathrm{Max}\{\bar{v}_1(q_1, y-p_1\bar{x}_1, \epsilon),\ldots,\bar{v}_n(q_n, y-p_n\bar{x}_n, \epsilon)\}$$

we have

$$\begin{aligned}
&\mathrm{Max}\{\bar{v}(p, q, y, \epsilon)\}, \{\bar{v}(p', q, y, \epsilon)\}=\mathrm{Max}\{\mathrm{Max}\{\bar{v}_1(q_1, y-p_1\bar{x}_1, \epsilon),\ldots,\\
&\bar{v}_n(q_n, y-p_n\bar{x}_n\epsilon)\}, \mathrm{Max}\{\bar{v}_1(q_1, y-p'_1\bar{x}_1, \epsilon),\ldots,\bar{v}_n(q_n, y-p'_n\bar{x}_n, \epsilon)\}\}=\\
&=\mathrm{Max}\{\mathrm{Max}\{\bar{v}_1(q_1, y-p_1\bar{x}_1, \epsilon), \bar{v}_2(q_2, y-p_2\bar{x}_2, \epsilon), \bar{v}_n(q_n, y-p_n\bar{x}_n, \epsilon),\ldots,\\
&\bar{v}_1(q_1, y-p'_1\bar{x}_1, \epsilon), \bar{v}_2(q_2, y-p'_2\bar{x}_2, \epsilon),\ldots,\bar{v}_n(q_2, y-p'_n\bar{x}_n, \epsilon)\}\}=\\
&\geq \mathrm{Max}\{\bar{v}_1(\lambda p_1+(1-\lambda)p'_1, q_1, y, \epsilon), \mathrm{Max}\{\bar{v}_2(q_2, y-p_2\bar{x}_2, \epsilon), \bar{v}_2(q_2, y-p'_2\bar{x}_2, \epsilon)\},\ldots,\\
&\mathrm{Max}\{\bar{v}_n(q_n, y-p_n\bar{x}_n, \epsilon), \bar{v}_n(q_n, y-p'_n\bar{x}_n, \epsilon)\}\}\geq
\end{aligned}$$

by quasi-convexity of $\bar{v}_1(.)$, and because the Max operator is nondecreasing,

$$\begin{aligned}
&\geq \mathrm{Max}\{\bar{v}_1(\lambda p_1+(1-\lambda)p'_1, q_1 y, \epsilon), \bar{v}_2(\lambda p_2+(1-\lambda)p'_2, q_2, y, \epsilon),\ldots,\\
&\mathrm{Max}\{\bar{v}_n(q_n, y-p_n\bar{x}_n, \epsilon), \bar{v}_n(q_n, y-p'_n\bar{x}_n, \epsilon)\}\}\geq\\
&\geq \bar{v}(\lambda p+(1-\lambda)p', q, y, \epsilon)\}
\end{aligned}$$

by quasi-convexity of $\bar{v}_2(.)$ and of all the $(j-2)$ remaining elements of the left-hand side set. Therefore, the unconditional indirect utility function $\bar{v}(p, q, y, \epsilon)$ is quasi-convex in p.

References

Braden, J., Kolstad, C., 1991. Measuring the Demand for Environmental Quality. North-Holland, Amsterdam.
Hanemann, W.M., 1982. Applied welfare analysis with qualitative response models. Working-Paper nº 241, Giannini Foundation of Agricultural Economics, University of California, Berkeley, October 1982.
Hanemann, W.M., 1984. Discrete/continuous models of consumer demand. Econometrica 52, 541–561.