Economics Letters 72 (2001) 403–410

economics letters

www.elsevier.com/locate/econbase

Identifying non-consistent choice behavior in recreation demand models

Luis C. Nunes[a,*], Maria A. Cunha-e-Sá[a], Maria M. Ducla-Soares[a], Márcia A. Rosado[b], Brett H. Day[c]

[a]*Faculdade de Economia, Universidade Nova de Lisboa, Travessa Estevão Pinto, 1099-032 Lisboa, Portugal*
[b]*INOVA/FEUNL, Faculdade de Economia, Universidade Nova de Lisboa, 1099-032 Lisboa, Portugal*
[c]*CSERGE, University College of London, UK. Visiting INOVA/Faculdade de Economia, Universidade Nova de Lisboa, 1099-032 Lisboa, Portugal*

Received 22 March 2000; received in revised form 20 October 2000; accepted 7 November 2000

Abstract

In a RPL model, rather than imposing consistency with consumer theory by constraining the distribution of the price coefficient to have negative support, a more general procedure is developed. Non-consistent choice behavior is identified in a recreation demand model for game reserves in South Africa.

Keywords: Neoclassical theory of preferences; Recreation demand; Random parameter logit (RPL)

JEL classification: 022; 721; 211

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 cases decisions are taken sequentially, in two steps. While in the first step the decision is a discrete choice between alternatives, the second step decision, made conditional on the first, can be either discrete or continuous. Examples of purely qualitative choices, where the second step decision is also discrete, can be found in the transportation mode choice literature (Daly and Zachary, 1978), and in the demand for recreational activities (Ichimura and Thompson, 1998).

Nunes et al. (1998) derived the conditions under which an underlying rational preference exists in

*Tel.: +351-213-833-624; fax: +351-213-886-073.
E-mail address: lcnunes@fe.unl.pt (L.C. Nunes).

0165-1765/01/$ – see front matter
PII: S0165-1765(01)00449-9

the case of discrete/discrete choice data. Specifically, these authors have shown that, given the maximizing behavior of the consumers, if each of the second step conditional indirect utility functions is well behaved then the first step 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. In addition, the same authors have shown 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. If it can be shown that the consumers' conditional indirect utility functions are increasing in prices then it can be concluded that consumers are not behaving according to the neoclassical theory of preferences.

This paper applies the results derived in Nunes et al. (1998) in the context of a discrete choice model of recreational demand for game reserves in the KwaZulu-Natal province in South Africa. Regarding the estimated models, two key points should be stressed. First, rather than constraining the estimate of the price parameter to a single value, it varies in the population according to a given distribution, as in Train (1998). This more flexible specification captures eventual heterogeneity of attitudes across households. Second, rather than constraining the sign of the parameter of the price variable to conform with consumer theory, we consider a family of distributions that accommodates many different shapes. Conditional on the parametric distribution of the price parameter, the eventual identification for a fraction of the population of choice behavior non-consistent with economic theory is allowed.[1] This is in contrast with Train (1998), and the literature in general, where choice behavior consistency with neoclassical theory of preferences is commonly imposed a priori.

The remainder of the paper is organized as follows. Section 2 presents the econometric model. Section 3 describes the data. Section 4 discusses the results, and Section 5 summarizes the main conclusions of the paper.

2. The econometric model

When decisions involve discrete choices, researchers have frequently applied random utility models (RUMs) to model the choice between alternatives. To construct a RUM the researcher hypothesizes the conditional indirect utility function, $\bar{v}_{ij}$, for household i and alternative j. This function depends on the price of the alternative to the household, p_{ij}, and a vector of characteristics describing the qualities of the alternative as experienced by the household, $\boldsymbol{q}_{ij}$.

Typically, empirical studies have postulated a linear function for each conditional indirect utility function and an additive error term, ϵ_{ij}, according to:

$$\bar{v}_{ij} = \delta \boldsymbol{q}_{ij} + \beta p_{ij} + \epsilon_{ij} \tag{1}$$

where δ is the vector of parameters corresponding to the qualities of the alternative, $\boldsymbol{q}_{ij}$, and β is the parameter of the price variable p_{ij}. Finally, the researcher specifies a probability distribution for ϵ and, on the presumption that households will act rationally and choose the option amongst the available

[1] Alternatively, a non-parametric approach could be used.

alternatives that provides them with the greatest utility, generates a probabilistic choice model. When the error term, ϵ_{ij}, is i.i.d. extreme value, the probability that household i chooses j is given by:

$$\pi_{ij}(\beta, \delta) = e^{\bar{v}_{ij}} / \sum_j e^{\bar{v}_{ij}} \tag{2}$$

This is the familiar Conditional Logit (CL) model of McFadden (1973), usually estimated by maximum likelihood.

Specifications of the indirect utility function that constrain the estimates of the parameters to single values cannot reflect different tastes and attitudes across households. A more appropriate specification is to allow the value taken by the parameters to vary in the population, as in the Random Parameters Logit (RPL) model of Train (1998). In this formulation, the coefficient on price, β, may take different values for different households. Though these values are unobserved for each household, the distribution of these values in the population can be characterized by a probability density function, $f(\beta|\theta)$, where θ is the vector of the parameters of this distribution. Therefore, the probability that the researcher assigns to household i choosing j is the integral of $\pi_{ij}(\beta, \delta)$ over all possible values of β weighted by the density of β, that is,

$$Q_{ij}(\theta, \delta) = \int \pi_{ij}(\beta, \delta) f(\beta|\theta)\, d\beta \tag{3}$$

Estimation of the parameters in the likelihood function is not possible through exact maximum likelihood since the integral in (3) cannot be, in general, calculated analytically. Rather, Q_{ij} (.) is approximated through simulation, and the parameters estimated through maximization of the simulated likelihood function, as in Train (1998).

According to the results in Nunes et al. (1998), choice behavior not consistent with the neoclassical theory of preferences can be characterized as follows:

$$\frac{\partial \bar{v}_{ij}}{\partial p_{ij}} \geq 0 \Leftrightarrow \beta \geq 0 \tag{4}$$

Since the value of the price parameter now varies in the population, if part of the estimated distribution for β lies above zero, then, according to the condition in Eq. (4), the model suggests for a fraction of the population that choice behavior does not conform to the neoclassical theory of preferences.

In this paper, a RPL model is estimated for data relating to the recreational trips of households when choosing between different game reserves in the KwaZulu-Natal Province of South Africa. However, rational behavior is not imposed a priori as in Train (1998), where the choice of the distribution of the coefficients of the random parameters was made according to prior beliefs about their signs, that is, consistent with the neoclassical theory of preferences. When some coefficient was believed to be positive (negative) according to theory, a positive (negative) log-normal distribution was chosen. When there was no prior belief about the sign of some coefficient, the normal distribution was chosen.

Instead, in the example considered in this paper, we consider a family of distributions that includes, besides the normal, others with only positive or negative support, as follows:

$$\beta = \lambda_1 + \lambda_2\alpha + \lambda_3\alpha^2 \text{ where } \alpha \sim \mathrm{N}[0,1] \tag{5}$$

with λ_1, λ_2, λ_3, parameters to be estimated. In particular, while for $\lambda_3 = 0$ the coefficient β is normally distributed, for $\lambda_1 = \lambda_2 = 0$ and $\lambda_3 = 1$ it is distributed as a chi-square with one degree of freedom. Moreover, when $\lambda_2 = \lambda_3 = 0$ the coefficient β is not random, and the model is a standard CL.

If the estimated probability that β is positive is 'large', then one may claim that, conditional on the parametric distribution of the price parameter, there is a proportion of households whose choices are not consistent with preference theory. Therefore, by allowing the data to dictate the distribution of the coefficient, instead of constraining the parameter to follow a given distribution imposed a priori, the proposed procedure attempts to minimize the impact of eventual biases caused by misspecification of the underlying distribution function of the price coefficient.

3. The data

The province of KwaZulu-Natal lies in the north-eastern corner of the Republic of South Africa. It is a region of extreme natural diversity, boasting lagoons, coral reefs, mountains and some of Africa's oldest game reserves. The four game reserves that are the focus of this empirical application (Hluhluwe, Umfolozi, Mkuzi and Itala) are administered by the KwaZulu-Natal Parks Board (KNPB), an organization that is responsible both for the reserves' protection and enhancement, and also for providing facilities and accommodation for visitors.

The four reserves are relatively different in the game-viewing experience they afford visitors. Households visit the reserves for a number of reasons; Umfolozi is the largest and possibly wildest of the four reserves, Hluhluwe is the only park in which there is a reasonable chance of seeing large herds of elephant, whilst Mkuzi boasts the greatest diversity of bird species.

A sample of 1000 trips made by residents of KwaZulu-Natal province between August 1994 and August 1995 was collected from data stored on the KNPB reservations database. The database indicated each household's choice of reserve and accommodation. The database was also used to determine which other reserve-accommodation combinations would have been available to a household when making the choice (some accommodation types may already have been booked by other parties or have been unavailable due to maintenance activities). It was also possible to determine exactly how many units of these alternative reserve-accommodation options would be needed to house the party and to calculate the cost they would have paid if they had chosen that option.

A more detailed description of the data and the construction of the data set are provided in Day (1998).

4. Results and discussion

Using the econometric models discussed in Section 2, the choice between the four game reserves is modeled as a function of the accommodation costs of staying in each reserve. Accommodation costs for non-chosen game reserves were obtained by calculating costs averaged over the accommodation types available in that particular reserve. Variability of the prices across individuals within a game reserve is assured, according to the description of the data in Section 3. Three dummy variables (Hluhluwe, Itala and Umfolozi) were also included. These alternative specific constants capture the influence of all characteristics of each game reserve that were not considered as regressors in the model.

Specifically, a CL model, a RPL model where the random cost coefficient was assumed to follow a normal distribution, and a more general distribution, as in Eq. (5), are estimated. The results of the estimated models are presented in Table 1.

For the CL model, the highest coefficient is for Hluhluwe, indicating the preference of households for this reserve over the other reserves. Because all estimated choice specific indicators coefficients are positive, it seems that if accommodation costs were the same in all parks, the least preferred park would be Mkuzi. The sign of the coefficient on accommodation costs is negative as expected. All the coefficients are significant at the 1% significance level.

To allow for heterogeneity in the accommodation cost coefficient, RPL models were estimated assuming a normal distribution, and a more general case as in Eq. (5). The results obtained for the reserve specific dummy variables are similar to those obtained for the CL model. At the estimated values of the parameters it follows from Eq. (5) that $\beta = -18.9914 + 16.5646\alpha + 10.3177\alpha^2$, where $\alpha \sim N[0,1]$. This estimated distribution is illustrated in Fig. 1.

The results clearly suggest that the assumption of a constant β across the population ($\lambda_2 = \lambda_3 = 0$) should be rejected. The assumption of normality ($\lambda_3 = 0$) is also rejected in favor of the more general distribution. Although the (negative) log-normal is not a particular case of the general distribution in Eq. (5), it was also considered, following Train (1998). In this case, the value of the log-likelihood function is -1083.0837. This value is smaller than that obtained in the normal case (-1077.58), indicating that the negative log-normal is not adequate.

Table 1
Model estimation results

Parameter	CL	RPL (Normal)	RPL (General)
Accommodation cost			
λ_1	-8.83^a	-12.06^a	-18.99^a
	(0.49)	(1.01)	(1.81)
λ_2	0	11.11^a	16.56^a
		(1.62)	(2.52)
λ_3	0	0	10.32^a
			(1.87)
Hluhluwe dummy	1.64^a	1.91^a	1.96^a
	(0.12)	(0.14)	(0.15)
Itala dummy	1.23^a	1.44^a	1.49^a
	(0.14)	(0.16)	(0.17)
Umfolozi dummy	0.78^a	0.95^a	1.03^a
	(0.12)	(0.14)	(0.15)
Log likelihood	-1085.46	-1077.58	-1069.60

Note: The numbers in parentheses below the estimated coefficients are the estimated asymptotic standard errors.
[a] 1% significance level.

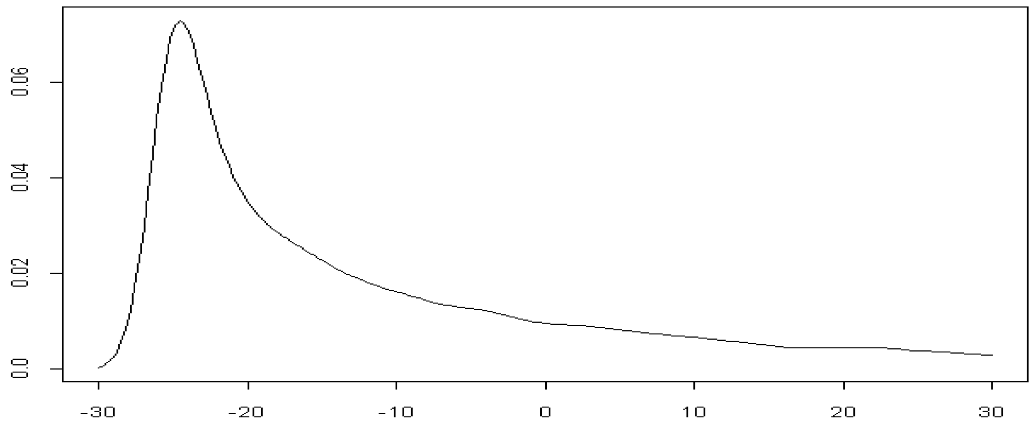

Fig. 1. Estimated general distribution for β.

Since the distribution of β in the general RPL does not conform to any standard form, statistical tables cannot be referenced to determine the percentage of households showing a non-consistent behavior in the sample. However, by Monte-Carlo integration, simple statistics can be easily derived by drawing a large number of times from the estimated distribution. Using this technique, the mean and standard deviation of β are calculated to be -8.62 and 22.05, respectively, which are quite different from those in the normal distribution case (-12.06, 11.11).

The share of households behaving non-consistently with the neoclassical theory of preferences can be computed in a similar manner by drawing a large number of times from the estimated distribution and computing the percentage of draws that gives a positive β. The resulting estimate gives a value of 23%, which is larger than in the normal case, 14%. Thus, with this data set, the normality assumption underestimates non-consistent choice behavior.

So far, the reported estimated probability followed directly from the point estimates of the parameters. A confidence interval for the proportion of households that behave non-consistently can be constructed by an approximation to the distribution of the estimator of the unknown probability using a bootstrap procedure. Each bootstrap sample is generated by drawing with replacement from the whole sample of households. For each bootstrap sample, the model is re-estimated and a new probability is computed following the procedure described above. Given a large enough number of bootstraps, an approximation to the distribution of the estimator of the probability that β is positive can be obtained. Fig. 2 presents the histogram of these bootstrap probabilities, assuming the more general distribution for the accommodation cost coefficient.

The histogram clearly suggests that the probability of choice behavior non-consistent with theory is large. In fact, the distribution of the proportion of households showing non-consistent choice behavior is concentrated around 0.23, and well away from zero.

Finally, a comparison of the likelihood ratio indices (LRIs) for the various models suggests that the general RPL model (LRI = 0.1735) achieves greater explanatory power than the RPL (normal) model (LRI = 0.1673), which in turn has greater explanatory power than both the CL model (LRI = 0.0058) and the (negative) log-normal.

Obviously, the general RPL model is the preferred one for this data set. Using the proposed procedure, this model gives a clear indication that non-consistent choice behavior exists in the population.

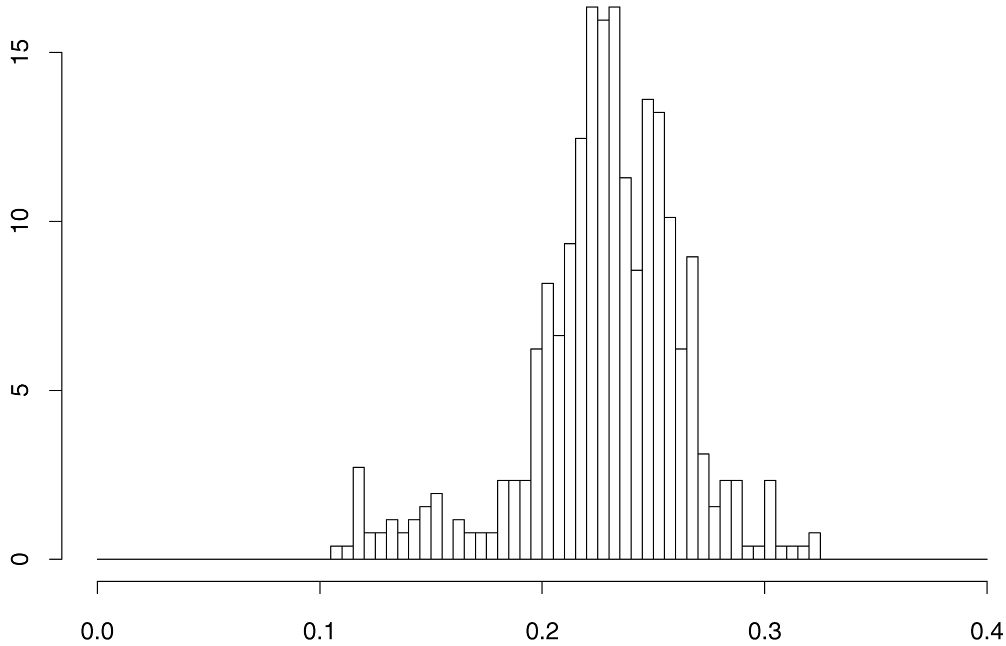

Fig. 2. Histogram of bootstrap probabilities for general RPL model.

5. Conclusions

This paper develops a procedure that attempts to minimize the impact of eventual biases caused by misspecification of the underlying distribution function that describes the heterogeneity of the price coefficient. The results are obtained in the context of a recreational demand model for trips to four game reserves in the KwaZulu-Natal province of South Africa. By considering a family of parametric distributions for the price parameter that accommodates many different shapes, the results suggest the presence of choice behavior not consistent with the neoclassical theory of preferences. This is in contrast to previous work in the literature, where rational behavior has been typically assumed a priori. This procedure could also be applied to many other different choice structures.

Empirically, this can be of great relevance. In fact, for welfare estimates to be meaningful and appropriate for policy purposes, it should be examined whether household choice behavior actually conforms to neoclassical theory of preferences.

Acknowledgements

We thank an anonymous referee for helpful comments.

References

Daly, A., Zachary, J.S., 1978. Improved multiple choice models. In: Hensher, D.A., Dalvi, D.A.Q. (Eds.), Determinants of Travel Choice. Saxon House, Teakfield Limited, Westmead, Farnborough, Hants, England, pp. 335–357.

Day, B.H., 1998. Day trips and away breaks: modelling recreational demand for the game reserves of the KwaZulu-Natal province of South Africa. CSERGE-London, unpublished manuscript.

Ichimura, H., Thompson, S.T., 1998. Maximum likelihood estimation of a binary choice model with random coefficients of unknown distribution. Journal of Econometrics 86, 269–295.

McFadden, D., 1973. Conditional logit analysis of qualitative choice behavior. In: Zarembka, P. (Ed.), Frontiers in Econometrics. Academic Press, New York, pp. 105–144.

Nunes, L.C., Cunha-e-Sá, M., Ducla-Soares, M., 1998. Testing for rationality: the case of discrete choice data. Economics Letters 60, 255–261.

Train, K.E., 1998. Recreation demand models with taste differences over people. Land Economics 74, 230–239.