Marine Resource Economics, Volume 14, pp. 333–341
0738-1360/99 $3.00 + .00

A Heteroskedastic Nested RUM of Freshwater Fishing

DOUGLAS J. MACNAIR
SEAN D. COX
Triangle Economic Research

Abstract *Choice set definition can be viewed as a nesting issue. Although the behavioral basis for grouping sites into nests is not well understood, one reason for grouping alternatives into nests may be the likelihood that they are "in" or "out" of an individual's choice set. However, a problem with using nests to evaluate choice set issues is that the researcher typically needs to impose the same nesting structure on all individuals and trips. Such an approach assumes that the degree of correlation among the alternatives does not vary across the sample. This paper develops and tests a more flexible nesting structure that allows the parameters on the inclusive values to vary systematically based on the sample demographics.*

Key words Angler behavior, nesting, random utility model, recreation demand.

Introduction

Nesting is a method for overcoming the restrictive statistical properties of the random utility model (RUM). Models without nests assume that the errors across all alternatives within the choice set are independently and identically distributed (i.i.d.), which is often an unrealistic assumption. Nested models are less restrictive because they construct groups of alternatives that share similar, but unobserved characteristics. Nested models impose the i.i.d. assumption on alternatives within the same group, but across groups the error distribution can vary.

The nesting structure imposed on the RUM usually has a behavioral basis. For example, fishing sites may be grouped into rivers and lakes. Anglers might be described as first choosing whether to fish at a lake or a river based on personal characteristics. Once that decision has been made, the angler then chooses among the sites within that nest.

Choice-set definition can be viewed as a nesting issue. Anglers might group sites as "relevant" or "irrelevant" to their site choice, then choose from among the "relevant" sites. In other words, either (*a*) the similar but unobserved characteristic the sites share is whether or not they are in the anglers' choice set or (*b*) the sites share similar but unobserved characteristics that cause them to be included or excluded from the choice set. From this standpoint, the parameter on the inclusive value for a nest can be an indicator of the probability that a group of sites is in the individual's choice set.

Using a nested model to directly test for choice-set relevance is not practical.

Douglas J. MacNair (corresponding author) is a senior economist with Triangle Economic Research, 2775 Meridian Parkway, Durham, NC 27713, e-mail: dmacnair@ter.com. Sean D. Cox is an associate economist with Triangle Economic Research.

The number of possible choice sets to consider as forming the "relevant" set is far too large. Therefore, to use a nested approach for choice-set definition, a researcher must be content to use a single nesting structure that mirrors the "relevant" vs. "irrelevant" distinction for the entire sample. Such an approach highlights a significant limitation of nested models. Using the same nests for the entire sample assumes that the degree to which sites share similar but unobserved characteristics is the same for all anglers. This may be unreasonably restrictive and can lead to biased parameter estimates for site characteristics if the assumption is not correct. With nested models, the best the researcher can do is evaluate several one-size-fits-all nesting structures and select the one with the best fit (Kling and Thomson 1996). Moreover, nested models only show the factors that affect the choice of a site within a nest, but not the factors that affect the extent to which sites are substitutes, which is important behavioral information.

This paper develops and tests a more flexible nested model. The heteroskedastic nested RUM (HNRUM) allows the degree to which sites within the same nest share similar but unobserved characteristics to vary by angler and by trip. While it does not eliminate the need to impose a nesting structure on the data, the HNRUM allows the researcher to go one step farther in assessing the relevance of the nests to each angler.

In this application, we do not directly test for choice-set relevance. However, the HNRUM is more amenable to assessing choice-set issues. The degree of substitution among lakes and rivers can vary by whether or not the angler owns a boat, which may better reflect the extent to which sites are relevant or irrelevant for boat vs. nonboat-owning anglers.

This paper has five sections. Section 1 derives the HNRUM and discusses its advantages. Section 2 describes the data used for this application. Section 3 describes the model results. Section 4 describes the welfare calculations for a simulated policy change. Section 5 discusses conclusions and future research.

Nesting Structures

Nonnested RUMs assume that the errors follow a generalized extreme value (GEV) function with a constant variance across sites and individuals. The nested RUMs also assume a GEV, but nested RUMs can have a different variance for different groups of sites. Following Morey's (1999) notation the error distribution is:

$$\sum_{m=1}^{M}\left[\sum_{j=1}^{J_m}\exp(\varepsilon_{mj}s_m)\right]^{1/S_m} \tag{1}$$

where j denotes one of N sites, there are M nests containing J_m sites and S_m is the substitution parameter for nest m. S_m is equal to $1/(1-\sigma_m)$, where σ_m is approximately equal to the correlation in the errors for the nest (Maddala 1983). When σ_m equals 1, the model reduces to a nonnested model. In order to be globally consistent with utility maximization, σ_m must be greater than or equal to 1 (McFadden 1978). This error distribution means the probability of site j being selected on trip i from nest m is:

$$P_{ij} = \frac{\exp(X_{ij}\beta)\left[\sum_{k=1}^{N_m}\exp(X_{ik}\beta S_m)\right]^{(1/S_m)-1}}{\sum_{m=1}^{M}\left[\sum_{k=1}^{N_m}\exp(X_{ik}\beta S_m)\right]^{(1/S_m)}} \tag{2}$$

S_m has a significant effect on the estimated probabilities and the log-sums (or inclusive values) used in welfare calculations. For a given $X_{ij}\beta$, a site in the nest with the higher S_m will have lower selection probability than if it were in a nest with a lower S_m. The probability is lower because sites within that nest have a higher correlation (σ_m) and are less distinct from each other. Therefore, the probability of selecting any one of the sites in the high S_m nest is lower. A higher S_m also implies a lower probability of selecting a site from the nest. The sites are closer substitutes for each other, and the probability of selecting a site from that group is lower than the sum of the exp($X_{ij}\beta$) would indicate. As a result of the lower selection probabilities, the log-sum for the nest with a higher S_m is also lower than it would otherwise be. Therefore, the welfare effects of changes in site characteristics within the nest will be diminished.

Because the log-sums for the nests are the basis for the welfare calculations, it is important to accurately measure the S_m for policy analysis. The error distribution in equation (1) only varies by site; the characteristics of the individual do not affect the degree of substitutability among the sites. One way to reduce this weakness would be to allow the error distribution to vary by trip, *i*, such as:

$$\sum_{i=1}^{I}\sum_{m=1}^{M}\sum_{j=1}^{N_m}\exp(\varepsilon_{mj}s_{mi})^{1/S_{mi}} \tag{3}$$

There are insufficient degrees of freedom to estimate a substitution parameter for each trip for each nest. However, the substitution parameter can be estimated as a function:

$$S_{mi} = \alpha_m + Z_i\lambda_m \tag{4}$$

where (Z_i) is a vector of individual and trip-specific characteristics such as age and income and angling avidity. Because this model parameterizes the error distribution and allows for a nonconstant variance across individuals, we refer to the model as a heteroskedastic nested RUM (HNRUM). The traditional nested model is a restricted (or nested) version of the HNRUM, in that the traditional model assumes $\lambda_m = 0$.

The HNRUM does not reduce the independence from irrelevant alternatives (IIA) property for a given trip compared to a nested model. As with the nested model, the selection probability ratio for two sites in different nests will not impose IIA. The odds of selecting one site relative to another within the same nest is still independent of the characteristics of all other sites in the model. However, the HNRUM allows these odds to vary across individuals and trips.

Data

This study uses data from a 14-month panel survey of Montana anglers from July 1992 to August 1993. The respondents were recruited using random-digit dialing and 75% of anglers agreed to participate. Once recruited, the respondents received a trip diary every two months where they recorded details of their fishing trips. Then the respondents were called and asked to read the information from their trip summaries to the interviewer. The response rates for each of the seven panels range from 61% to 78%. In total, 2,919 trips were reported. After removing trips that lack key information and trips lasting for more than 1 day, 1,473 trips remain for use in this analysis. Table 1 provides demographic information of survey respondents and the locations of their trips.

Table 1
Demographic and Trip Information

Variable	Mean	Standard Deviation
AGE	48	15.61
INCOME ($ 1992)	25,893	16,027.45
OWNBOAT	0.47	
TARGET SALMON	0.06	
MAJOR SITE	0.63	
RIVER SITE	0.51	
MAJOR RIVER	0.37	
MAJOR LAKE	0.26	
NONMAJOR RIVER	0.14	
NONMAJOR LAKE	0.23	

"Major" refers to sites that are defined by the *Angler's Guide to Montana Fishing* (Sample 1984) as major fishing areas.

The choice set for this model comprises 182 river sites and 71 lake sites. In most cases, the lake sites are defined around a single lake. River site definitions are based on Montana River Information System river reaches, the smallest segments of each river. The fishing sites are characterized using the variables listed in table 2.

Results

Our preliminary analysis of alternative nesting structures shows that the data are best modeled with four nests: major river, nonmajor river, major lake, and nonmajor lake. Major refers to sites that are defined by the "Angler's Guide to Montana Fishing" (Sample 1984) as major fishing areas.[1]

Table 3 presents two contrasting models using these four nests. NRUM is a traditional nested RUM with a constant S_m for each nest (i.e., $\lambda_m = 0$). HNRUM includes a variety of demographic characteristics as determinants of the substitution parameters. For both models, the site characteristic parameters are constrained to be the same between the major and nonmajor river nests and also between major and nonmajor lake nests. The travel cost parameter is constrained to be equal across all four nests.

The results show that the HNRUM provides a significant improvement in explanatory power over the traditional nested RUM. Because the HNRUM nests the NRUM, we can use a likelihood ratio test to determine whether restrictions imposed by the NRUM are appropriate. The difference in log-likelihood between the two models is 152.0, which for 11 degrees of freedom has a p-value of 0.0000.

The substitution parameter functions contain the same variables for both lake nests. The results show that younger anglers view major lake sites as closer substitutes, as do lower income anglers. Conversely, older anglers view nonmajor lakes as closer substitutes than younger anglers do. We had no priors on the signs of these variables; they just reflect differences in preferences.

[1] The fact that the major versus nonmajor nesting classifications work well supports a behavioral interpretation of nests. Nests may not reflect characteristics that are endemic to the sites, but the information available to anglers about the sites. The results suggest that more research should be focused on the role of information in developing the nesting structure for a RUM.

Table 2
Site Characteristics

Variable	Description	Mean	Standard Deviation
Lake Sites Only			
Abundant Fish	Dummy variable for lakes with "abundant" fish	0.54	0.50
Campgrounds	Number of campgrounds relative to circumference of lake	0.14	0.24
Surface Area	Log of the surface area of the lake	5.59	2.11
River Sites Only			
Biomass	Biomass rating measure of pounds per 1,000 feet of river	82.94	154.85
Catch Restrictions	Number of species with catch restrictions	0.03	0.07
Reach	Log the length of reach in miles	2.59	0.71
Aesthetics	Aesthetics rating for rivers	0.20	0.40
Lake and River Sites			
Major	Dummy variable for site defined as major fishing sites	0.35	0.48
River	Dummy variable for river sites	0.72	0.45
Trip Cost	Costs of trip calculated as trips costs plus maintenance costs plus oil costs	19.83	17.14

The river substitution parameters are affected by several demographic characteristics. *Income* is negative and significant, while *Age* is positive and significant. *Target Salmon* is negative and significant for trips to major rivers, which means these sites are more likely to be chosen for trips where salmon is the target species. There were no trips to nonmajor rivers where salmon was the target species; therefore, the variable *Target Salmon* could not be included in the substitution parameter function for the nonmajor river nest. The *Own Boat* indicator is positive which means boat owners view river sites as less desirable substitutes than do nonboat owners. Because anglers cannot use boats on many rivers, most boat fishing trips occur on lakes. Therefore, it is not surprising that boat owners view lakes as less desirable substitutes. The results do not suggest that rivers should be excluded from boat owners' choice set. In fact, boat owners take 36% of their fishing trips to rivers, which clearly indicates that rivers are important components of their choice set. The results show only that rivers have correlated errors, are closer substitutes for each other, and have lower selection probabilities for boat owners. This important aspect of boat owners' behavior would not be evident if a traditional nested model were used or if rivers were excluded from boat owners' choice set.

The HNRUM only has a modest effect on the site characteristic parameters compared to the NRUM. Table 3 reports the marginal values of the site characteristics. There are five site attribute parameters for the river nests. All the marginal values decline with the HNRUM, mostly because of the 12% increase in the travel cost parameter. *Biomass* is positive and highly significant for both models, but *Biomass* has a reduced marginal value with the HNRUM, \$1.76 vs. \$1.46, a 15% decline. *Major* and *Catch Restrictions* are significant in both models and decline about 7% in the HNRUM. The marginal values of the *Aesthetics Dummy* declined about 4%. *Reach* is insignificant in both models. There are four significant site-attribute parameters for the lake nests. As with the river nests, the HNRUM does not have a significant effect on the marginal values of these characteristics. However, within the lake nest the marginal value of *Campgrounds* increases by 7%.

Table 3
RUM Results Parameters

	Variable	Nested RUM	Heteroskedastic Nested RUM
River	Biomass	1.76	1.46
		(6.69)	(5.75)
	Reach	0.94	0.52
		(1.54)	(0.87)
	Aesthetics dummy	7.90	7.58
		(6.01)	(6.02)
	Catch restrictions	–4.27	–3.97
		(–5.81)	(–5.58)
	Major dummy	18.33	17.14
		(8.14)	(7.47)
Lake	Abundant fish	5.60	5.30
		(5.28)	(5.42)
	Surface area	3.58	3.23
		(10.71)	(10.18)
	Campgrounds	4.64	4.99
		(1.68)	(1.88)
	Major dummy	8.16	7.11
		(4.50)	(4.23)
Common to All Nests	Travel cost	–0.056	–0.063
		(–14.21)	(–14.46)
Substitution Parameters	Major river		
	Constant	1.78	1.54
		(13.61)	(8.50)
	Income/10,000		–0.02
			(–8.40)
	Target Salmon		–0.47
			(–2.50)
	Boat		0.18
			(2.45)
	Age/100		1.28
			(3.75)
	Nonmajor river		
	Constant	1.51	1.27
		(11.76)	(8.21)
	Income/10,000		–0.01
			(–4.12)
	Boat		0.23
			(2.54)
	Age/100		0.59
			(2.31)
	Major lake		
	Constant	2.01	2.75
		(12.38)	(9.95)
	Income/10,000		–0.02
			(–7.83)
	Age/100		–0.79
			(–1.90)
	Nonmajor lake		
	Constant	1.73	1.45
		(12.32)	(6.88)
	Income/10,000		–0.01
			(–5.42)
	Age/100		1.17
			(2.47)
	Log-likelihood	–4,657	–4,581

Notes: All site characteristic parameters have been normalized by travel cost parameters. T-statistics are in parentheses.

Welfare Estimates

Although the marginal value of site characteristics are not significantly affected by using an HNRUM, policy analysts and researchers are often interested in the value of specific sites and the distribution of welfare changes. We use an OLS regression to evaluate the significance and cause of differences in welfare estimates between the two models for the most popular river and lake sites. Table 4 shows the results.

The compensating variation measures for each trip for each individual (i) are:

$$CV_i^N = \frac{\ln(D_i^N) - \ln(\tilde{D}_i^N)}{\beta_{TC}^N} \tag{5}$$

$$CV_i^H = \frac{\ln(D_i^H) - \ln(\tilde{D}_i^H)}{\beta_{TC}^H} \tag{6}$$

where the N and H superscripts refer to the nested and heteroskedastic nested RUMs, respectively; β_{TC} is the parameter on travel cost; D_i is the log-sum or inclusive value (denominator of equation 2) for all sites; and $\tilde{D}_i$ is the log-sum with either the most popular lake or river eliminated from the choice set.

The dependent variables for the regressions are the difference between the absolute values of CV under the HNRUM model and the NRUM model: $(CV_i^N) - (CV_i^H)$. Therefore, a positive parameter on the independent variable means higher losses under HNRUM. For both models, all but one of the demographic characteristics are significant, the R^2 are high, and the F-test p-values are 0.000. This confirms that the models yield significantly different CV values for different individuals. This result is not caused by the difference in the travel cost parameter, which drives many of the differences in site characteristic marginal values. Travel cost only acts as a scalar on the difference in expected utility, the D_i. Therefore, the differences in the D_i must be the cause of the results.

Table 4
Results of OLS Regressions

	Most Popular River Regression (a)		Most Popular Lake Regression (b)	
	Parameter	T-Statistic	Parameter	T-Statistic
Constant	0.02	5.96	–0.344	–21.00
Income/10,000	0.013	28.41	0.066	22.12
Age/100	–0.10	–20.96	0.40	12.82
Boat	–0.013	–8.97	0.32	3.24
Target Salmon	0.05	15.81	0.001	0.04
N	1,473		1,473	
R-squared	0.47		0.35	
F-statistic p-value	0.0000		0.0000	

Notes:
(a) dependent variable is $\text{CV(HNRUM)}_{\text{most popular river}} - \text{CV(NRUM)}_{\text{most popular river}}$
(b) dependent variable is $\text{CV(HNRUM)}_{\text{most popular lake}} - \text{CV(NRUM)}_{\text{most popular lake}}$

The results are consistent with expectations based on the RUM results. For example, eliminating the most popular river for boat owners results in lower CV for boat owners with the HNRUM compared to the NRUM. Conversely, boat ownership results in higher losses for eliminating the most popular lake. Because boat owners view rivers as less desirable substitutes under HNRUM, the losses for the most popular river are lower, while the losses for the most popular lake are higher. Similarly, anglers who target salmon have higher losses for rivers because they are more likely to choose a river site than the NRUM would suggest.

Conclusions

The HNRUM model specifies a substitution parameter that is a function of individual characteristics, which lets the degree of substitution between nests vary by angler and by trip. The traditional nested RUM model, which estimates a constant substitution parameter for each nest, is a constrained version of the HNRUM model because it imposes the restriction that all the substitution parameters, except the constant, are zero. The results reported here suggest the substitution variables are important considerations in estimating nested models. For example, in our preliminary models we first tested a two-nest model with river vs. lake and with *Major* a variable in the substitution parameter function. The results were strong enough to suggest testing a four-nest model. Our results also show that including substitution variables at the very least can change the distribution of gains and losses.

The purpose of properly determining an individual's choice set is to ensure that sites with very low (or zero) actual selection probabilities also have low (or very close to zero) predicted probabilities. The task is complex: it is difficult for a researcher to know whether a site has low selection probability because it is not in the choice set or because the site characteristics make the site undesirable. Nesting is one approach for dealing with choice set relevance, because it only requires determining the sites that share common characteristics. It does not require that the characteristics be specified. Ultimately the goal is to develop more flexible RUMS that estimate more accurately the site selection probabilities; therefore, it becomes less important to specifically determine whether the alternative is in or out of the choice site. The HNRUM may be one tool for accomplishing that goal. While the changes in probabilities are significant but modest in this application, other data may show more significant results, especially for the value of site characteristics and site values.

The results show that incorporating demographic information in the nesting structure can improve the explanatory power of nested models. The HNRUM also shows the factors that affect the degree of substitutability among sites by different anglers, which can help in selecting the overall nesting structure. Moreover, to the extent that site choice is a nesting issue, the HNRUM may be a useful tool in evaluating site relevance while maintaining the site characteristics in the model.

References

Kling, C.L., and C.J. Thomson. 1996. The Implications of Model Specification for Welfare Estimation in Nested Logit Models. *American Journal of Agricultural Economics* 78(1).

Maddala, G.S. 1983. *Limited Dependent and Qualitative Variables in Econometrics*. Cambridge University Press.

McFadden, D. 1978. Modeling the Choice of Residential Location, Chapter 3 *Spatial Interaction Theory and Planning Models,* A. Karlqvist, L. Lundqvist, F. Snickars, and J. Weibull, eds., Amsterdam: North Holland.

Morey, E. 1999. Two RUMs Uncloaked: Nested-Logit Models of Site Choice and Nested-Logit Models of Participation and Site Choice, *Valuing Recreation and the Environment,* C.L. Kling and J.A. Herriges, eds., Edward Elgar Publishing Ltd.

Sample, M.S. 1984. *The Angler's Guide to Montana*. Billings, MT: Falcon Press Publishing Co., Inc.