AN APPLICATION OF THE NESTED MULTINOMIAL LOGIT MODEL TO ENROLLMENT CHOICE BEHAVIOR

William C. Weiler

Data bases containing information on the socioeconomic characteristics and post-high school activities of large numbers of young people are commonly used in enrollment demand studies. While researchers almost always use the multinomial logit approach to obtain estimates of the effects of the independent variables in this context, this technique has a potentially serious limitation. This study presents an alternative to multinomial logit and contrasts estimated results from the two approaches using data on 446 high-school graduates in a midwestern state. The results are different enough to suggest that researchers should explore alternatives to multinomial logit when using this type of data base.

Many researchers use data on individuals to analyze postsecondary attendance behavior. With these data the enrollment choices of the individuals are made over a limited number of "discrete" alternatives that constitute the exhaustive set of available education options. It is now well-known that using ordinary least squares (OLS) to analyze relationships in which the dependent variable is discrete or qualitative is not appropriate. If there are just two alternatives in the choice set, logit or probit analyses are often used to estimate the relationship between the option selected and the characteristics of the alternatives and of the individuals in the data sample. In analysis of enrollment choices these methods have been used to explain the choice between attendance and nonattendance or between attendance at a particular institution and not attending that institution.[1]

It is clearly of interest to extend the analysis to choice among several types of attendance options (e.g., four-year or two-year schools) or even to a limited number of individual institutions. When there are more than two alternatives in the choice set, the most widely used approach is multinomial or conditional logit

William C. Weiler, Management Information Division, 40 Wulling Hall, University of Minnesota, Minneapolis, MN 55455.

(MNL). This method is computationally simple with today's software and computers and also has another desirable property: It is easy to use the estimated results to forecast choices when a new alternative is introduced or when an existing one is eliminated so long as no parameters are added or deleted as a result. This is a useful property, for example, in estimating the effects of either opening or closing a postsecondary educational institution. However, MNL also has a distinct limitation, a property known as "independence of irrelevant alternatives" (IIA), which implies that the odds of choosing alternative i relative to alternative j are independent of the characteristics of or the availability of alternatives other than i and j. This is clearly a very restrictive property when the alternatives being studied have different degrees of "nearness" or similarity.[2]

The purpose of this paper is to present a generalization of MNL called nested multinomial logit (NMNL) that deals with this problem and is applicable to the study of attendance choice alternatives. The NMNL approach is not new; it has, for example, been used to study housing (McFadden, 1984) and commuter arrival time (Small and Brownstone, 1982) choices. However, to my knowledge it has not been applied to choice among postsecondary attendance options. Hence, it is useful to describe the NMNL technique in this context and to show its relevance when choice options available to individuals in the data set have different degrees of similarity.

A DESCRIPTION OF MNL AND NMNL

All models used to analyze individual choice among discrete alternatives begin with the assumption that each individual chooses the alternative that yields the most utility or satisfaction.[3] To fix ideas for analyzing postsecondary attendance choice, we consider a particular individual i who chooses among four attendance choice alternatives indexed by $j = 0, 1, 2,$ and 3 that constitute the choice set (nonattendance is among them). We also make the standard assumption that the utilities of the alternatives can be written as linear functions

$$U_{ij} = X_{ij}\alpha + \epsilon_{ij} \tag{1}$$

where X_{i0} is a vector of the values of the characteristics of option 0 that influence i's utility of option 0, X_{i1} is a vector of the values of the same characteristics for option 1, etc. The ϵ_{ij}'s are error terms accounting for unmeasured characteristics and random individual behavior, and α is an unknown parameter vector.

A particular statistical estimator is produced from this general formulation of decision making by assuming a specific distribution for the error terms. The MNL estimator results from the assumption that the ϵ_{ij}'s are independently and identically distributed with the type I extreme value distribution. Under this assumption the probability P_{ij} that individual i selects option j is

$$P_{ij}=\exp(X_{ij}\alpha)/\sum_{k=0}^{3}\exp(X_{ik}\alpha) \tag{2}$$

Three things are apparent from (2). First, components of the vectors X_{ij} that do not vary across the alternatives cancel out of (2). This is why socioeconomic attributes of individual i were not included among the variables in X_{ij}. We will discuss this issue in more detail in the next paragraph. Second,

$$\log(P_{ij}/P_{ik})=(X_{ij} - X_{ik})\alpha \tag{3}$$

for j and $k=0$, 1, 2, or 3 and $k\neq j$. This is the IIA property; it shows that the odds of selecting option j over k do not depend on characteristics of options other than j and k. Third, suppose option 4 is added to the choice set. The revised probabilities P'_{ij} of the five options now available are

$$P'_{ij}=\exp(X_{ij}\alpha)/\sum_{k=0}^{4}\exp(X_{ik}\alpha) \tag{4}$$

Conversely, suppose option 3 is eliminated from the choice set. The revised probabilities (j=0, 1, or 2) are now

$$P'_{ij}=\exp(X_{ij}\alpha)/\sum_{k=0}^{2}\exp(X_{ik}\alpha) \tag{5}$$

If an option is either added to or removed from the choice set, it is clear from (4) or (5) that $\log(P'_{ij}/P'_{ik})=(X_{ij} - X_{ik})\alpha$ just as in (3). This means, for example, that if $P_{i0}=.5$, $P_{i1}=.25$, $P_{i2}=.1$, and $P_{i3}=.15$ before option 3 is eliminated, then $P_{i0}=.588$,$P_{i1}=.294$, and $P_{i2}=.118$ in the new situation whatever the characteristics of option 3. However, if options 0, 1, and 2 were nonattendance, attendance at four-year school F, and attendance at community college A, respectively, we would certainly expect different results if option 3 were community college B than if it were four-year school G. It is also clear that we can construct equally implausible examples if an option were added to the original choice set.[4]

It is possible to include the socioeconomic characteristics of individual i, represented by the vector Y_i, in the utilities (1) so long as their effects differ across the options. What this means is that we can change (1) to write the utilities as linear functions of X_{ij} and Y_i, or $U_{ij}=X_{ij}\alpha+Y_i\beta_j + \epsilon_{ij}$, so long as β_j differs for each option. Calculation of the probabilities P_{ij} proceeds from the obvious redefinition of (2). However, the IIA property is still retained since (3) now becomes

$$\log(P_{ij}/P_{ik}) = (X_{ij} - X_{ik})\alpha + (\beta_j - \beta_k)Y_i \tag{3'}$$

so that the odds of selecting j over k still do not depend on the characteristics of options other than j and k. We also lose the ability to produce a calculation like (4) because the new option introduces a new unknown parameter vector β_4 on the socioeconomic characteristics.

The reason MNL exhibits the IIA property is because the error terms in the utilities (1) are assumed to be independent. In a four-option example suppose the options are nonattendance, attendance at four-year school F, attendance at community college A, and attendance at community college B. In this situation it is probably not realistic to assume ϵ_{i2} and ϵ_{i3} are independent, although it may be reasonable to assume ϵ_{i0} and ϵ_{i1} are independent of each other and of the other two. This assumption yields a specific example of the NMNL approach. If we let c_i represent the option selected by individual i so that c_i assumes the values 0, 1, 2, or 3, then we can compute the probabilities

$$P(c_i=0) = P_{i0} = \exp(X_{i0}\alpha)/(\exp(X_{i0}\alpha) + \exp(X_{i1}\alpha) + D^{1-\sigma}) \tag{6}$$

$$P(c_i=1) = P_{i1} = \exp(X_{i1}\alpha)/(\exp(X_{i0}\alpha) + \exp(X_{i1}\alpha) + D^{1-\sigma}) \tag{7}$$

$$P(c_i=2/c_i \neq 0,1) = \exp(X_{i2}\alpha/(1-\sigma))/D \tag{8}$$

$$P(c_i=3/c_i \neq 0,1) = \exp(X_{i3}\alpha/(1-\sigma))/D \tag{9}$$

where $D = \exp(X_{i2}\alpha/(1-\sigma)) + \exp(X_{i3}\alpha/(1-\sigma))$. It can be shown that if the parameter $\sigma = 0$, the probabilities P_{ij} are calculated exactly as in (2). In fact, for this particular case, σ is approximately equal to the correlation between ϵ_{i2} and ϵ_{i3}, and in the general NMNL model, σ is a parameter that measures the similarity between options 2 and 3.

While formulas (6)–(9) look complicated, they have intuitively interesting forms. Expressions (8) and (9) suggest that the choice between the two community college attendance options, given that the individual attends a community college, is made on the basis of a simple bivariate logit model. Expressions (6) and (7) look very much like the MNL expression (2) except that options 2 and 3 are combined into a term $D^{1-\sigma}$ that looks something like a weighted average of the two similar alternatives. It is also the case that the IIA property does not hold for the similar alternatives in this model, although the arithmetic is rather complicated and will not be shown here. (Interested readers can consult Maddala, 1983, pp. 71–72.)

This example illustrates the two general ideas inherent in the construction of the NMNL model. First, choices between the similar alternatives are made on the basis of a standard MNL model. Second, for an alternative distant from the similar alternatives, the similar alternatives are aggregated into a "composite" alternative for an MNL model comparison with the distant alternative. Note also that dividing (6) by (7) and taking the logs of both sides shows that the IIA property still holds for the dissimilar alternatives.

In principle, the NMNL approach can be extended to more than one set of similar options within all of the options in the choice set. In our example we might also want to assume that ϵ_{i1} is correlated with both ϵ_{i2} and ϵ_{i3} but that ϵ_{i0} is still uncorrelated with any of the other three. If we extended our example to include more individual institutions in the choice set, it would be natural to assume one similarity parameter for similarity between all four-year institutions, another similarity parameter measuring similarity between all community colleges, and possibly a third parameter for similarity between all four-year schools as a group and all community colleges as a group.

AN EMPIRICAL ILLUSTRATION

We apply the MNL and NMNL techniques to data on the enrollment choices of a random sample of 446 high school seniors who graduated in 1974 from schools in a particular region of a Midwest state. The postsecondary attendance choices of the individuals were aggregated into one of four options: nonattendance, attendance at a four-year college or university, attendance at a community college, or attendance at a technical institute. The latter are public postsecondary institutions specializing in vocational programs.

The only variable in the data set that measures a characteristic of the various options is the distance to the nearest institution categorized in each option (DISTANCE). Other characteristics of the options such as tuition, number of dorm rooms, or type of program either do vary across the individuals in the data sample or do not vary across the options. The other variables in the analysis measure characteristics of the individuals in the sample. These are dummy variables for sex (SEX, $1=$female) and parent's college background (PCOL, $1=$either parent attended college) and the student's high school rank percentile (HSRANK).

From (3′) we can see that adding the same constant to the coefficient vectors β_j and $\beta_{k'}$ leaves the left-hand side of this relation unchanged. Hence, we need to normalize one of these vectors, and the standard way to do so is to set one of them to zero; we therefore set each element of β_0 to zero. Since the "distance" to the nonattendance alternative is not defined, this means that the utilities for each option are

$$U_{i0} = 0$$
$$U_{i1} = \alpha\text{DISTANCE} + \beta_{11}\text{SEX} + \beta_{12}\text{PCOL} + \beta_{13}\text{HSRANK}$$
$$U_{i2} = \alpha\text{DISTANCE} + \beta_{21}\text{SEX} + \beta_{22}\text{PCOL} + \beta_{23}\text{HSRANK}$$
$$U_{i3} = \alpha\text{DISTANCE} + \beta_{31}\text{SEX} + \beta_{32}\text{PCOL} + \beta_{33}\text{HSRANK}$$

for a total of ten parameters (α and β_{11} through β_{33}) to be estimated for our MNL model. From these specifications of the utilities it is clear that the coefficients represent the difference between the effect of a variable on a particular choice of a type of school and its effect on nonattendance.

The first column of Table 1 reports the MNL estimates. The second column shows NMNL estimates from a model with the same utilities that also includes a parameter controlling for the similarity between the community college and technical institute choices.[5] There are some differences between the two sets of estimated coefficients, but all estimates that we would expect to have particular signs have those signs in both cases. For example, both sets of estimates indicate that students who rank higher in their high school classes are more likely to attend a college or university (β_{13}), while students who live farther from the attendance options are less likely to attend (α), an expected result since DISTANCE is a measure of attendance costs.[6]

There are two other results of interest. First, the similarity parameter estimate for the NMNL model is between 0 and 1 and is statistically significant. McFadden (1984) notes that the similarity parameter should fall in this interval for the results to be meaningful and also discusses a way of using this parameter estimate and its standard error to test the null hypothesis that the MNL model is the correct specification against the alternative that the NMNL model is the

TABLE 1. Estimated Results[a]

Parameter	MNL Model	NMNL Model
α	−.0256*	−.0208*
β_{11}	−.922*	−.920*
β_{21}	−.798	−.522
β_{31}	−.110	−.147
β_{12}	.952*	.926*
β_{22}	.669	.510
β_{32}	.0857	.245
β_{13}	.0108*	.00850*
β_{23}	−.0161*	−.00709
β_{33}	−.00726**	−.00674**
σ	—	.596*
Log of the likelihood function	−491.86	−485.90

[a] * and ** denote statistical significance at the 5% and 10% significance levels, respectively.

correct specification.[7] The test statistic, which has the chi-square distribution with one degree of freedom, is $(1-\sigma)^2/\text{SE}$ where σ is the estimate and SE is its estimated standard error. This statistic has a value of 8.55 in our case which means we reject the null hypothesis at the 0.005 significance level. Second, another statistic for testing the same null and alternative hypotheses is calculated by multiplying the difference of the values of the log likelihood function by -2 (again see McFadden, 1984). This statistic also has the chi-square distribution with one degree of freedom, so that the value of 11.92 for our case leads us to reject the null hypothesis at the 0.001 significance level.[8]

While there are differences in the coefficient estimates in Table 1, it is impossible to see the impacts of clustering the community college and technical institute alternatives from those results. A way to show these differences is to compare the elasticities of the choice probabilities with respect to each independent variable for the two approaches. These elasticities are reported in Table 2. We first note that the signs of the elasticities of each attendance option with respect to SEX, PCOL, and HSRANK are the same for both sets of estimates. Women are more likely not to attend or to attend a technical institute than are men. Individuals whose parents attended college are more likely to attend a community college or a four-year school and less likely to attend a technical school or not attend, although the sizes of these elasticities are not large. High-school rank is positively associated with four-year school attendance and negatively related to the other choices. It is somewhat surprising that the

TABLE 2. Elasticities of Attendance Probabilities with Respect to Each Independent Variable[a]

	Alternative			
Independent Variable	Non-Attendance	Attend 4-Year	Attend Comm. Col.	Attend Technical
SEX	0.15	−0.30	−0.24	0.10
	(0.15)	(−0.30)	(−0.31)	(0.14)
PCOL	−0.093	0.19	0.11	−0.067
	(−0.099)	(0.18)	(0.14)	(−0.052)
HSRANK	−0.030	0.53	−0.87	−0.41
	(−0.041)	(0.40)	(−0.43)	(−0.39)
DISTANCE to	0.27	−0.73	0.27	0.27
nearest 4-yr.	(0.23)	(−0.58)	(0.23)	(0.23)
DISTANCE to	0.064	0.064	−0.98	0.064
nearest c. c.	(0.045)	(0.045)	(−1.78)	(0.33)
DISTANCE to	0.11	0.11	0.11	−0.47
nearest tech.	(0.086)	(0.086)	(0.63)	(−0.54)

[a] All elasticities are calculated at the mean values of the independent variables. Two values are shown for each independent variable and alternative; the one in parentheses is calculated from the NMNL estimates, and the other is calculated from the MNL estimates.

negative elasticity of nonattendance with respect to this variable is much smaller in absolute value than the elasticities for community college and technical school attendance. A possible explanation is that potential students' high-school ranks have little effect on the choice of attendance versus nonattendance, but a larger effect on the choice of the selected option given attendance.

The elasticities of attendance with respect to DISTANCE, the only variable that measures an attribute of the attendance alternatives, are substantially different for the two sets of estimates. The MNL elasticities of the probability of choosing nonattendance, attendance at a four year school, and attendance at a technical institute with respect to distance to the nearest community college are equal; similar results hold for the distances to the nearest four-year school and technical institute. These results are forced by the IIA property. However, with NMNL we do not have the IIA property for the similar alternatives. As a result we see much larger elasticities of the community college attendance probability with respect to distance to the nearest technical institute and of the technical institute attendance probability with respect to distance to the nearest community college. These are exactly the types of changes we would expect if these options are truly close substitutes for each other because they imply that students are more responsive to changes in the costs of options that are close rather than distant substitutes.

Two additional points regarding the elasticities are of interest. First, the pattern of equal elasticities for distance to the nearest four-year college or university is the same for both the MNL and NMNL estimates. In other words, because we assume no similarity between attendance at a college or university and any other option in constructing our NMNL estimates, the IIA property is still present for this "dissimilar" alternative. Second, the NMNL estimates of the elasticities of the probabilities of nonattendance and attendance at a four-year institution with respect to the distances to the two similar options are equal for the same reason: We assumed similarity between the community college and technical school options, but that the error terms in the nonattendance and attendance at four-year school utilities were independent of each other and independent of the error terms in each of the similar options.

CONCLUSIONS

Data bases containing information on the socioeconomic characteristics and postsecondary attendance choices of large numbers of individuals are commonly used in analyses of enrollment demand. However, the appropriate statistical technique to use in this context is not so clear. The MNL approach has several advantages including widespread familiarity among researchers and computational ease, but one major disadvantage, the IIA property. In this paper we have presented NMNL, an alternative, but very similar approach that does not have this disadvantage. The empirical results from the two procedures are different enough to suggest that researchers should investigate using the NMNL approach, or possibly some other alternative to MNL that does not incorporate the IIA property, when working with this type of data base.

NOTES

1. Leslie and Brinkman (1987) include a table that lists, among other things, the type of data sample and estimation method for a large selection of student demand studies.
2. Several researchers (Radner and Miller, 1975; Kohn, Manski, and Mundel, 1976; Chapman, 1979) have used MNL in enrollment choice analyses. The first two studies include a discussion of the effects of the IIA property.
3. Several other approaches besides MNL and NMNL have been suggested to analyze individuals' selections among discrete alternatives. Two that do not have the IIA property and have been used to analyze enrollment choices are sequential logit (Elliott and Hollenhorst, 1981; Ghali, Miklius, and Wada, 1977; Weiler, 1986) and universal logit (Weiler and Wilson, 1984). These techniques, along with MNL and NMNL, are described in surveys of estimation methods developed for use with discrete dependent variables (Amemiya, 1981; Maddala, 1983; McFadden, 1981, 1984). See these references for more detailed discussions of the theories of behavior that generate these estimators and of the maximum likelihood methods that are used to estimate unknown parameters.
4. Consider, for example, building a new community college five miles away from an existing one.
5. There are several statistical software packages such as TSP that have MNL routines, but none to my knowledge that provide estimates for the NMNL model. However, since estimating the parameters of the NMNL model is just a matter of maximizing a function of several variables (the likelihood function), mathematical software libraries that contain routines to perform this task can be used. The parameters reported in Table 1 were estimated using the routine ZXMIN from the IMSL mathematical software library.
6. Other specifications including separate dummy variables indicating each parent's college attendance and dummy variables indicating that the distance to an option was within commuting distance yielded similar results. In particular, the test statistics discussed below indicated that the NMNL estimates were theoretically preferred.
7. If σ does not fall in the unit interval, it is evidence that the NMNL model is misspecified in terms of the assumed similarity grouping of the alternatives. Had this occurred with our NMNL results, we would have reestimated the model with similarity between the community college and four year options.
8. There is another statistic (see Maddala, 1983, section 3.11) that tests the null hypothesis that the IIA property is correct against the alternative that it is not. The rationale behind the statistic is that if the IIA property holds, estimates of the MNL model coefficients from the full data sample should not differ much from the estimates obtained using that part of the data set constructed by eliminating all individuals choosing one or more of the alternatives in the choice set. A difficulty with this test is that it can produce conflicting results depending on which observations are dropped from the data sample.

REFERENCES

Amemiya, T. (1981). Qualitative response models: A survey. *Journal of Economic Literature* 19(4): 1483–1536.

Chapman, R. G. (1979). Pricing policy and the college choice process. *Research in Higher Education* 10(1): 37–57.

Elliott, D., and Hollenhorst, J. (1981). Sequential unordered logit applied to college selection with imperfect information. *Behavioral Science* 26(4): 366–378.

Ghali, M., Miklius, W., and Wada, R. (1977). The demand for higher education facing an individual institution. *Higher Education* 6(4): 477–487.

Kohn, M. G., Manski, C. F., and Mundel, D. S. (1976). An empirical investigation of the factors influencing college going behavior. *Annals of Economic and Social Measurement* 5(4): 391–419.

Leslie, L. L., and Brinkman, P. T. (1987). Student price response in higher education: The student demand studies. *Journal of Higher Education* 58(2): 181–204.

Maddala, G. S. (1983). *Limited-Dependent and Qualitative Variables in Econometrics.* New York: Cambridge University Press.

McFadden, D. (1981). Econometric models of probabilistic choice. In C. F. Manski and D. McFadden (eds.), *Structural Analysis of Discrete Data with Econometric Applications,* pp. 198–272. Cambridge, MA: MIT Press.

McFadden, D. L. (1984). Econometric analysis of qualitative response models. In Z. Griliches and M. D. Intriligator (eds.), *Handbook of Econometrics, Volume II,* pp. 1395–1447. Amsterdam: Elsevier.

Radner, R., and Miller, L. S. (1975). *Demand and Supply in U. S. Higher Education.* New York: McGraw-Hill.

Small, K. A., and Brownstone, D. (1982). Efficient estimation of nested logit models: An application to trip timing. Paper No. 296, Econometric Research Program, Princeton University.

Weiler, W. C. (1986). A sequential logit model of the access effects of higher education institutions. *Economics of Education Review* 5(1): 49–55.

Weiler, W. C., and Wilson, F. S. (1984). Prediction of the enrollment effects of institutional closure. *Research in Higher Education* 20(1): 23–33.

Received November 11, 1987