UCI-ITS-WP-85-4

Efficient Estimation of Nested Logit Models

UCI-ITS-WP-85-4

David Brownstone
Kenneth A. Small

School of Social Sciences and
Institute of Transportation Studies
University of California, Irvine

June 1985

Institute of Transportation Studies
University of California, Irvine
Irvine, CA 92697-3600, U.S.A.
http://www.its.uci.edu

ABSTRACT

This paper examines the Sequential, Full Information Maximum Likelihood (FIML), and Linearized Maximum Likelihood (LML) estimators for a Nested Logit model of time-of-day choice for work trips. These estimators are compared using a Monte Carlo study based on specification and data from a previously published empirical study. The sequential estimator is found to be much less efficient than either LML or FIML; and its uncorrected second-stage standard-error estimates are strongly downward biased. LML is only slightly less efficient than FIML, but is often easier to compute. However there are cases where the sequential and LML estimators do not exist.

I. INTRODUCTION

The growing popularity and sophistication of discrete choice models has led to increasingly frequent use of the nested logit (NL) model. It is a natural generalization of multinomial logit (MNL), sharing some of its computational advantages but allowing freedom from the property of "independence from irrelevant alternatives" [McFadden (1981)]. Its applied uses have included transportation modal choice [Ben-Akiva (1974), Cosslett (1978)], consumer durable choice [Brownstone (1980)], household energy demand [Goett (1979), Cameron (1982)], and automobile demand [Train (1980, 1985), Hensher and Moncfield (1982), Hocherman et al (1983)].

One of the attractions of NL is the ability to use sequential estimation consisting of two or more MNL steps in succession. This sequential estimator has several well-known disadvantages: it is not asymptotically efficient, and it produces inconsistent standard error estimates [Amemiya (1978)]. Little work has been done assessing the severity of these problems in practice, or comparing the sequential estimator with alternative estimators. McFadden (1981) reports results by Cosslett using full information maximum likelihood (FIML) for two NL models; one is insignificantly different from MNL and the other yields parameter estimates outside the allowable range for NL to be a random utility model (see the next section). Brownstone (1980) and Hausman and McFadden (1984) report FIML but not sequential results. No other NL estimators have even been computed in any published literature known to us.[1]

This paper investigates three NL estimators, namely the sequential estimator and two asymptotically efficient estimators: full information maximum likelihood (FIML) and linearized maximum likelihood (LML). The latter consists of one step of the method-of-scoring algorithm starting from the sequential estimator. We use a Monte Carlo design based on data and specification from an applied study of work-trip timing. The example is one that has been thoroughly investigated using MNL, and for which NL is a plausible generalization. By examining the empirical distribution of parameter estimates obtained from repetitions of a single known stochastic model, we can compare the estimators' efficiencies and check the finite-sample accuracy of their asymptotic sampling distributions.

Our results support the desirability of using asymptotically efficient estimators wherever possible, and strongly caution against the use of uncorrected standard error estimates from the sequential estimator.

---

1 Some results using LML are reported in an earlier (unpublished) version of Hausman and McFadden (1984).

II. NOTATION

The NL model used in this paper is known as a two-level model and is represented diagrammatically by the "tree structure" shown in Figure 1. The twelve alternatives are divided into three groups, each represented by a "node" denoted a, b, or c. The probability of choosing alternative k attached to node s is:

$$P_k = P(k|s) \cdot P(s) \tag{1}$$

where

$$P(k|s) = \frac{\exp(V_k/\rho_s)}{\sum_{j\epsilon B_s} \exp(V_j/\rho_s)} \equiv \frac{\exp(V_k/\rho_s)}{\exp(I_s)} \tag{1a}$$

$$P(s) = \frac{\exp(\rho_s I_s)}{\sum_r \exp(\rho_r I_r)} \tag{1b}$$

$$V_j = \beta z_j \ , \ j=1,\ldots,12 \tag{1c}$$

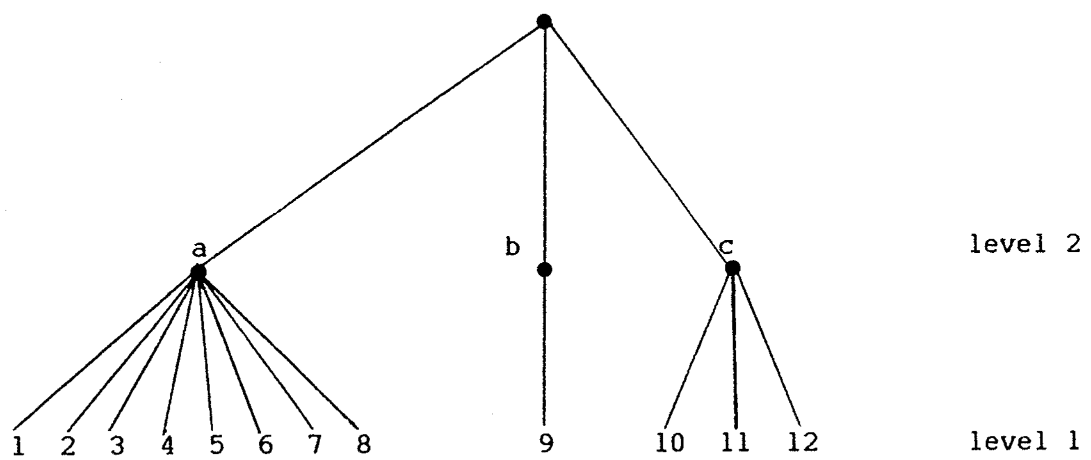
and

$$I_r = \log \sum_{j\epsilon B_r} \exp(V_j/\rho_r) \qquad r = a,b,c \ . \tag{1d}$$

McFadden (1981) shows that if

Figure 1

Nested Logit Tree Structure



$$0 < \rho_r \leq 1, \; r = a,b,c \; ,$$

these probabilities can be derived from a random utility model in which V_j is the "average" or "strict" utility of alternative j. This utility is specified in (1c) to be a linear function of observable characteristics z_j describing both the alternative and the individual making the choice. The quantity I_r is called the inclusive value of node r. A node such as b with only one attached alternative is said to be degenerate, and it is easy to see that $I_b = V_9/\rho_b$, $P(9|b)=1$, $P(b)$ is independent of ρ_b, and hence the parameter ρ_b drops out of equation (1). In this paper, furthermore, we constrain

$$\rho_a = \rho_c \equiv \rho \; . \tag{2}$$

Denote a member of the sample (also called a "case") by superscript n; denote the alternative chosen by that member by k^n and the corresponding node by s^n. The log-likelihood function is then

$$\begin{aligned} L &= \sum_n \log P^n_{k^n} \\ &= \sum_n \log P^n(k^n|s^n) + \sum_n \log P^n(s^n) \\ &\equiv L_1 + L_2 \; . \end{aligned} \tag{3}$$

Note that β appears in both L_1 and L_2, but that the scalar ρ appears in L_1 only through the quotients β/ρ. The sequential estimator takes advantage of this fact: it first estimates β/ρ by maximizing L_1 (first stage);[2] uses these estimates to compute I_r; then estimates ρ by maximizing L_2 (second stage). The advantage of this procedure is that both L_1 and L_2 have the form of MNL log-likelihoods in parameters β/ρ and ρ, respectively, hence can be maximized using MNL computer programs.[3] The MNL algorithm at the

---

2 One or more components of β may not be identified at the first stage because the corresponding variables do not vary over alternatives within groups B_r. An example is a dummy variable equal to 1 for $j \epsilon Ba$ and 0 otherwise. Such variables can simply be omitted in calculating inclusive value, and entered separately as additional variables at higher stages. To see this, partition $\beta = (\beta^1, \beta^2)$ and $z_j = (z_j^1, z_j^2)$, where $z_j^2 \equiv z_r^2$ for all $j \epsilon B_r$ so that β^2 is not identified at the first stage. Then from equation (1d),

$$I_r = \log \left[\exp(\beta^2 z_r^2) \cdot \sum_{j \epsilon \beta_r} \exp(\beta^1 z_j^1) \right] = \beta^2 z_r^2 + I_r^1,$$

where I_r^1 is the inclusive value computed from (1d) omitting variables z^2.

3 If ρ_a and ρ_c are not constrained equal, distinct first-stage estimates of $\beta/\rho r$ are obtained on each subsample whose choices are in one of the nondegenerate groups Br. If these are constrained to be proportional so as to yield identical estimates of β, the ability to use "off-the-shelf" MNL algorithms is lost; if they are not so constrained, the best estimate of β, given the second-stage estimates of ρr, is unclear. Imposing constraint (2) solves this problem by allowing a considerably more efficient technique at the first stage: β/ρ is estimated on the union of the subsamples just mentioned. To our knowledge this procedure has not been explicitly described, though it appears to have been used in practice [e.g. Train (1980)]. Note that in our example, since node b is degenerate, the first stage still does not use the entire sample: β/ρ is estimated ignoring the information from those sample members choosing alternative 9.

second stage generates what we call the "uncorrected" standard-error estimate of $\hat{\rho}$, using the Hessian of L_2 and hence implicitly assuming that the values β/ρ used in calculating L_2 are nonstochastic. The correct (i.e., consistent)standard-error estimate of $\hat{\rho}$ is computed as explained in the Appendix.

FIML, in contrast, maximizes L simultaneously with respect to β and ρ, using a nonlinear maximization algorithm that typically requires computing L and some of its derivatives at each iteration. We used the quadratic hill-climbing algorithm of Goldfeld and Quandt (1973, pp. 5-8) except that in place of the Hessian we substituted the expected cross-product of the gradient, both for step direction and for standard-error estimates; hence our procedure is a modified method of scoring (Rao, 1973, p. 370).

LML maximizes a quadratic approximation to L based on the values of L and its first two derivatives at the sequential estimate $\hat{\theta}_{SEQ} \equiv (\hat{\beta}'_{SEQ}, \hat{\rho}'_{SEQ})'$. The second derivative is estimated by the expected value of the cross-product of the gradient. Thus the LML estimate is:

$$\hat{\theta}_{LML} = \hat{\theta}_{SEQ} + H^{-1}g \tag{4}$$

where

$$g = \sum_n \frac{\partial \log P^n_{k^n}}{\partial \theta} \tag{4a}$$

$$H = \sum_n \sum_j P^n_j \left(\frac{\partial \log P^n_j}{\partial \theta} \right) \left(\frac{\partial \log P^n_j}{\partial \theta'} \right) \tag{4b}$$

with (4a) and (4b) both evaluated at $\hat{\theta}_{SEQ}$. Note that $\hat{\theta}_{LML}$ is obtained from one step of the method-of-scoring algorithm; hence it is consistent and efficient (Theil, 1971, p. 527). It requires essentially the same computations as FIML except they need not be iterated to convergence.

III. EMPIRICAL EXAMPLE

The empirical example is the choice of time-of-day for work trips, previously modelled by McFadden et. al. (1977), Small (1982), and Abkowitz (1981). Because of analytical difficulties with treating the choice as continuous,[4] plus a tendency of respondents to round off replies to the nearest five minutes, all of these authors estimated an MNL model of choice among 12 discrete alternatives, each representing arrival at work within a particular 5-minute interval. The choice set consists of intervals centered from 40 minutes before to 15 minutes after the official work start time for the individual. Data were collected on the actual arrival times, the official work start times, and other characteristics of 527 individuals who commuted by auto to a major city in the San Francisco Bay area in 1972 (see McFadden et. al., 1977). These were supplemented with engineering calculations of the travel times each would have faced at each of the 12 alternative arrival times.

The model estimated here is a simple extension of the previously published MNL results of Small (1982). The most well-behaved specification found in that paper is reestimated in the first column of Table 1, with variables defined as follows:[5]

---

4 For example, the congestion curves (giving travel time versus time of day) are neither monotonic, concave, nor convex. Hence, stating the condition for utility maximization requires dividing the possible range of time of day into several regions, and corner solutions are frequent.

5 The sample used here is larger than that in Small (1982) because of reconstruction of some previously missing carpool data. The MNL coefficient estimates are nearly identical, except that SDE , SDL , and SDLX have been divided by 10 in the current paper for computational reasons.

SD = Schedule Delay: actual arrival time minus official work start time in minutes, for a given alternative. Thus its value for alternative j is $SD_j = 5(j-9)$, j = 1,...,12.

R15 $= \begin{cases} 1 \text{ if SD} = -30, -15, 0, 15 \\ 0 \text{ otherwise.} \end{cases}$

R10 $= \begin{cases} 1 \text{ if SD} = -40, -30, -20, -10, 0, 10 \\ 0 \text{ otherwise.} \end{cases}$

TIM = Travel time in minutes.

SDE = Max {-SD/10 , 0}.

SDL = Max {SD/10 , 0}.

FLEX = Answer to question: "How many minutes late can you arrive at work without it mattering very much?"

SDLX = Max {(SD-FLEX)/10 , 0}

D2L $= \begin{cases} 1 \text{ if SD} \geq \text{FLEX} \\ 0 \text{ otherwise.} \end{cases}$

CP = Dummy for carpool.

This model captures the trade-off between the desire to avoid congestion on the one hand, and the desire to avoid arriving too early or late on the other. The estimated marginal rates of substitution imply that the average noncarpooler would incur .53 minute of travel time to avoid arriving an extra minute early; 1.24 minute to avoid arriving an extra minute late; and an additional 1.53 minute to avoid arriving an extra minute beyond the reported employer's flexibility range.[6]

---

6 These are the ratios of the coefficient on SDE , SDL , and SDLX , respectively, to the coefficient on TIM ; divided by 10 to get in the right units.

Table 1: Estimation Results on Actual Data

	MNL MODEL	NESTED LOGIT MODEL		
	FIML	Sequential	LML	FIML
ß/ρ (S.E.)				
R15	1.106	1.133	1.157	1.134
	(.101)	(.129)	(.115)	(.110)
R10	.398	.416	.439	.419
	(.102)	(.128)	(.114)	(.108)
TIM	-.141	-.195	-.179	-.163
	(.053)	(.072)	(.064)	.060
TIM•CP	.105	.163	.139	.129
	(.076)	(.104)	(.090)	(.084)
SDE	-.75	-.69	-.76	-.75
	(.06)	(.09)	(.07)	(.07)
SDE•CP	.23	.04	.24	.23
	(.09)	(.13)	(.11)	(.10)
SDL	-1.75	-1.92	-2.33	-2.07
	(.29)	(.81)	(.52)	(.50)
SDLX	-2.16	-3.10	-3.49	-2.81
	(.81)	(2.09)	(1.47)	(1.28)
D2L	-1.057	-1.015	-1.467	-1.314
	(.170)	(1.197)	(.395)	(.362)
$\hat{\rho}$		.882	.648	.807
(S.E.)		(.419)	(.134)	(.178)
[uncorrected S.E.]		[.075]		
Log Likelihood	-994.90	-998.10	-995.48	-994.43

Notes:
Dependent variable is choice among 12 time-of-day alternatives, each a 5-minute arrival interval. Alternative 9 is on-time arrival.
No. cases = 527.
Asymptotic standard errors are in parentheses.

Clearly the assumption of independence from irrelevant alternatives, implicit in MNL, is dubious here. At least two correlation patterns other than independence might plausibly be postulated for the unobserved preferences for these 12 alternatives. One, explored by Small (1981), is induced by the ordering of the alternatives and involves a closer correlation among "nearby" alternatives. The other, explored here, occurs if commuters have unmeasured preferences for arriving early (alternatives 1-8), on-time (alternative 9), or late (alternatives 10-12), thereby inducing correlation within the corresponding groups $B_a=\{1,...,8\}$, $B_b=\{9\}$, and $B_c=\{10,11,12\}$. This is precisely the kind of situation for which NL is designed. We also computed estimates using several other groupings (i.e. tree structures), including some three-level trees [Small and Brownstone (1982)]. The groupings used here gave among the most plausible results of any tried.

The sequential estimation results are given in the second column of Table 1. The inefficiency of the sequential estimator is manifested in the large standard error estimates. This arises because the 187 individuals choosing alternative 9 (on-time arrival) are dropped from the first-stage estimation since node b is degenerate (see footnote 3). Further evidence of the sequential estimator's inefficiency is that the likelihood attained is <u>lower</u> than that attained by the less general MNL model -- an unsettling result which we found for other tree structures as well.

Table 1 also shows the uncorrected standard error estimates for $\hat{\rho}$ as computed by the MNL algorithm in the second stage. This is a gross underestimate relative to the correct asymptotic standard error estimates, corroborating Cosslett's (1978) findings. Use of uncorrected standard errors might lead an investigator to unwarranted rejection of the hypothesis that the true model is MNL (i.e., $\rho = 1$).

Table 1 also presents results for LML and FIML, which yield considerably lower asymptotic standard error estimates than the sequential estimator. This apparent efficiency gain is particularly important for ρ and for those coefficients (of SDLX and D2L) that are only identified from the 22 individuals choosing late alternatives.

Computationally, each of these estimators has its advantages and disadvantages.[7] The sequential estimator is the simplest, but the need to correct its standard error estimates greatly reduces this advantage: the correction requires programming effort and computer time comparable to that needed to compute LML. We see little reason to use the sequential estimator if standard errors for $\hat{\rho}$ are needed. Comparing FIML and LML is more difficult. On the one hand, FIML does not require the sequential estimator as starting values (indeed, we found it often converges more easily starting from $\rho = 1$ and from coefficients set either to zero or to the MNL estimates). On the other, FIML requires a nonlinear maximization algorithm, uses more computer time, and sometimes encounters convergence problems or multiple maxima.

---

7 We used the program QUAIL [Berkman, et al (1979)] for MNL and for sequential NL estimates. The quadratic hill-climbing algorithm "GRADX" in GQOPT, distributed by Richard Quandt of Princeton University, was used for FIML estimation of NL . LML used a separate FORTRAN program with matrix inversion routine "LINV2P" from International Mathematical and Statistical Libraries.

IV. MONTE CARLO RESULTS

The conclusions in the previous section are based on a single model of unknown validity. Hence they may be affected by misspecification, and the asymptotic approximations underlying the standard error estimates may be inaccurate. These problems are overcome in this section through a Monte Carlo study based on the empirical model. Starting with "true" parameter values chosen in advance (for realism they are taken to be the simple average of the sequential and FIML estimates from Table 1), individual choices are generated randomly from the probabilities in equations (1) - (1d); these choices are then used to compute the various estimators. Furthermore, the process is repeated many times in order to obtain an empirical sampling distribution of the parameter estimates; as the number of repetitions becomes large, this empirical distribution should approach the true small-sample distribution. All the results reported here are based on 100 Monte Carlo repetitions; there was little change in the empirical distributions as the number of repetitions was increased beyond 60.

This method of estimating small-sample moments is identical to Efron's (1982) "bootstrap" estimator, a very general technique whose practical application in econometrics has been limited to linear models [Freedman and Peters (1984)] or to unrealistically simple binomial logit models [Brownstone (1984), Davidson and MacKinnon (1984)]. By contrast, the study described here is based on a realistic model and data set.

One immediate problem with our design is that the sequential and LML estimators turn out to exist for only 18 percent of the Monte Carlo repetitions. In the others, the first-stage log-likelihood L_1 increases monotonically as the coefficient of SDLX and/or D2L goes to $-\infty$. The reason is that SDLX and D2L are identically zero on alternatives 1-9, so only for those relatively few sample members choosing to arrive late do they vary across the alternatives included in L_1. Hence only through such cases do these variables' coefficients affect L_1. If it happens that in every such case the alternative chosen minimizes SDLX, then L_1 is maximized by setting the coefficient of SDLX to $-\infty$. A similar argument holds for D2L.

This problem hardly ever occurs with FIML because all 12 alternatives are considered simultaneously for each case. Since SDLX and D2L often vary between alternatives 9 and 10, a given Monte Carlo repition usually generates several "people" who choose alternative 10 even though it has a higher value of SDLX and/or D2L than alternative 9. In only one repetition did FIML fail to produce finite coefficient estimates.

Our example should not be viewed as pathological because of this problem. On the contrary, our design shows that estimator nonexistence can easily arise in data generated by a perfectly reasonable true model. The problem illustrates dramatically the general point that important information is lost by ignoring certain alternatives at the first stage of the sequential estimation process. In practice, an investigator using sequential estimation on data generated by the true model considered here might well be led to reject the correct specification because of "unreasonable" results, and to use some incorrect specification instead.

Selected results of the Monte Carlo experiment are shown in Table 2. In order not to bias the comparisons of the three estimators, their empirical sampling distributions are computed on the same subset of the total set of 545 Monte Carlo repetitions: namely, those 100 for which all three estimators exist. Unsurprisingly, the coefficients on SDLX and D2L are strongly upward biased in this subset, since it excludes many of the repetitions with choice vectors favoring highly negative estimates of these coefficients.

The last column describes the FIML empirical distribution on a more complete set of repetitions: 100 of the first 101 (excluding the one for which FIML failed to exist). The upward bias disappears, though the coefficient of SDLX now shows a sizable downward bias. It appears that FIML gives results about as good (i.e. coefficient estimates about as tightly distributed around the true parameters) whether or not the sequential estimator exists.

Since ρ enters the log-likelihoods in a highly nonlinear fashion, we repeated the entire experiment changing the true value of ρ to 0.20. This results in Table 3. The major difference is that the sequential estimator now exists for 77 percent of the repetitions; the smaller value of ρ causes the late alternatives to be chosen more frequently, which in turn increases the number of cases for which SDLX and D2L vary across alternatives included in L_1.

Tables 2 and 3 corroborate the conclusions of the previous section: LML and FIML are more efficient than sequential estimation, and the uncorrected standard error estimate for $\hat{\rho}$ is likely to be a gross underestimate. The efficiency gain is greatest in the case ρ=.844 ,

Table 2: Monte Carlo Results for ρ = .844

Coefficient (True Value)	Sample and Estimator			
	18% for which SEQ exists			99% for which FIML exists
	SEQ	LML	FIML	FIML
TIM (-.179)				
Bias	.011	.008	.019	.002
RMSE	.077	.062	.063	.061
$\widehat{SD}$	.076	.059	.056	.058
SDE (-.720)				
Bias	.000	.023	.006	.001
RMSE	.093	.079	.066	.072
$\widehat{SD}$	.089	.068	.063	.066
SDLX (-2.960)				
Bias	1.212	.600	.604	-.432
RMSE	1.761	1.231	1.154	1.698
$\widehat{SD}$	1.938	.970	1.100	1.618
D2L (-1.164)				
Bias	1.035	-.238	.119	-.017
RMSE	1.372	.752	.344	.378
$\widehat{SD}$	1.218	.392	.313	.345
ρ (.844)				
Bias	.011	-.060	.181	.041
RMSE	.368	.217	.380	.249
$\widehat{SD}$	.406	.176	.271	.217
$\widehat{SD}$ (uncorrected)	.075			
Skewness	1.90	.047	1.80	1.12
Log L	-1038.27	-998.15	-990.12	-985.77

Note: Each entry represents 100 repetitions excluding those for which the relevant estimator fails to exist. SEQ denotes the sequential estimator.

Table 3: Monte Carlo Results for ρ = .20

Coefficient (True Value)	Sample of Estimator			
	77% for which SEQ exists			100% for which FIML exists
	SEQ	LML	FIML	FIML
TIM (-.179)				
Bias	.003	.003	.004	-.002
RMSE	.076	.077	.077	.079
SD	.089	.084	.084	.084
SDE (-.720)				
Bias	-.024	-.018	-.020	-.023
RMSE	.106	.100	.100	.102
SD	.105	.099	.099	.099
SDLX (-2.960)				
Bias	-.088	-.147	-.093	-.096
RMSE	.744	.687	.676	.755
SD	.800	.742	.740	.758
D2L (-1.164)				
Bias	.082	-.038	-.039	-.142
RMSE	.631	.540	.539	.628
SD	.870	.586	.581	.612
ρ (.200)				
Bias	-.003	-.004	.002	-.003
RMSE	.037	.036	.038	.037
SD	.042	.039	.040	.039
SD (uncorrected)	.029			
Skewness	0.42	0.25	0.21	0.28
Log L	-1027.28	-1026.32	-1026.28	-1024.62

Note: Each entry represents 100 repetitions excluding those for which the relevant estimator fails to exist. SEQ denotes the sequential estimator.

and for the two coefficients (SDLX and D2L) that, from the earlier discussion, are rather tenuously estimated.

One striking result is the excellent performance of LML relative to FIML. Generally these estimators are equally efficient for this model. LML may therefore be preferred in many situations except where it fails to exist.

The tables also show that the small sample distributions are mostly quite close to their asymptotic limits. The best way to see this is by comparing the estimated standard deviations $\widehat{SD}$ with the root-mean-squared error RMSE, since most practical applications assume zero bias and use the asymptotic standard errors to generate confidence intervals. Comparison of the rows $\widehat{SD}$ and RMSE in the tables shows that this procedure works quite well for most coefficients and estimators. The exceptions are the coefficient of D2L and the parameter $\hat{\rho}$, especially in the case of ρ=.844. This may be due to the large amount of skewness found in their empirical distributions. (The empirical skewness of $\hat{\rho}$ is given in the next to last row of each table; skewness of the coefficient estimates, not shown, is small except occasionally for SDLX and D2L.) It is possible that reparameterizing the NL model to reduce the nonlinearity in ρ would improve the accuracy of the asymptotic approximations for $\hat{\rho}$, but that is well beyond the scope of this paper.

V. CONCLUSION

The theoretical advantages of efficient estimation seem to be realized for this particular model and data set. The sequential estimator may be useful for initial screening of models, but it cannot be relied upon for standard errors of the ρ parameters without a difficult correction. LML performs substantially better than the sequential estimator and is no harder to compute, if that correction is performed. FIML does not require computation of the sequential estimator and seems not to be adversely affected by the configurations of variables that cause the sequential estimator (and hence LML) not to exist. However, if FIML is used with arbitrary starting values, care must be taken that the global optimum is reached since otherwise consistency is not guaranteed.

In this study the asymptotic approximations of the small-sample distributions of most parameter estimates are adequate. This result is at odds with previous work on the simpler binomial logit model. Much more work is needed to discover just what conditions cause the asymptotic approximations to fail.

Until then, the bootstrap technique used here appears a useful though expensive method for estimating standard errors. As computing costs fall, investigators should consider substituting this for the traditionally reported asymptotic standard errors, which our results show to be sometimes misleading for hard-to-estimate parameters.

ACKNOWLEDGEMENT

This work received financial support from the National Science Foundation through grant No. SES-8007010, and from the University of California Institute of Transportation Studies. The results and views expressed are the authors' responsibility and should not be attributed to the National Science Foundation. We are grateful to Kenneth Boese, Douglas Holtz-Eakin, Avner Bar-Ilan, and David Bowman for creative and efficient research assistance; and to Jerry A. Hausman and an anonymous referee for comments on an earlier draft.

Appendix: Standard Error Formulas for the Sequential Nested Logit Estimator.

This appendix specializes the general formulas given in McFadden (1981, pp. 252-260) to the 2-level Nested Logit model with a single ρ parameter.

Equation (3) gives the log-likelihood function as:

(A1) $$L = \sum_n \log P^n(K^n|s^n) + \sum_n \log P^n(s^n) \equiv L_1 + L_2$$

Partition β and z into two groups $\beta = (\beta^1, \beta^2)$ and $z = (z^1, z^2)$ such that β^2 is not identified at the first stage (i.e., via maximization of L_1). Let Z^1 and Z^2 be the quantities actually entered as variables at stages 1 and 2, respectively; and γ^1 and γ^2 their coefficients. Thus $Z^1=z^1$, $\gamma^1 = \beta^1/\rho$, $Z^2 = (z^2, I^1)$, and $\gamma^2 = (\beta^2, \rho)$, where I^1 is the inclusive value from the first stage. Z^1 takes on distinct values $Z^1_{k,s,n}$ for alternative k attached to node s for individual n ; whereas Z^2 , which by construction does not vary among alternatives attached to a given node, takes on values $Z^2_{s,n}$.

Let E_i be the expectation operator appropriate for L_i ; i.e., E_1 uses probabilities $P(k|s)$ and E_2 uses $P(s)$. Define

(A2) $$M_{ij} = E_i \frac{\partial L_i}{\partial \gamma^i} \frac{\partial L_i}{\partial \gamma^j}'$$

McFadden (1981) shows that the asymptotic covariance matrix of the sequential estimator of γ is consistently estimated by

$$\text{(A3)}\qquad V = \begin{pmatrix} M_{11}^{-1} & -M_{11}^{-1}M_{21}'M_{22}^{-1} \\ -M_{22}^{-1}M_{21}M_{11}^{-1} & M_{22}^{-1}+M_{22}^{-1}M_{21}M_{11}^{-1}M_{21}'M_{22}^{-1} \end{pmatrix} .$$

M_{ii} is just the Information Matrix for the MNL likelihood function at the i:th stage, so the MNL computer packages produce standard-error estimates asymptotically equal to M_{ii}^{-1}. It is clear from formula (A3) that these "uncorrected" estimates are correct only for γ^1 and are downward biased for γ^2 .

Define the random variables $S_{k,s,n}$ to equal 1 if individual n chooses alternative k attached to node s. Let $S_{s,n} \equiv \sum_k S_{k,s,n}$. Then

$E_1 S_{k,s,n} = p^n(k|s)$ and $E_2 S_{s,n} = p^n(s)$.

Using this notation we have

$$\text{(A4)}\qquad L_1 = \sum_n \sum_s \sum_k S_{k,s,n} \log P^n(k|s)$$

and

$$\text{(A5)}\qquad L_2 = \sum_n \sum_s S_{s,n} \log P^n(s)$$

Differentiating (A4) and (A5) yields

(A6) $$\frac{\partial L_1}{\partial \gamma^1} = \sum_n \sum_s \sum_k S_{k,s,n}(Z^1_{k,s,n} - \bar{Z}^1_{s,n})$$

(A7) $$\frac{\partial L_2}{\partial \gamma^1} = \sum_n \sum_s \rho\, S_{s,n}\left(\bar{Z}^1_{s,n} - \sum_t P^n(t)\bar{Z}^1_{t,n}\right)$$

(A8) $$\frac{\partial L_2}{\partial \gamma^2} = \sum_n \sum_s S_{s,n}(Z^2_{s,n} - \bar{Z}^2_n)$$

where $\bar{Z}^1_{s,n} = \sum_k P^n(k|s)Z^1_{k,s,n}$ and $\bar{Z}^2_n = \sum_s P^n(s)Z^2_{s,n}$

Note that $\frac{\partial L_1}{\partial \gamma^2} = 0$, and therefore $M_{12} = 0$. Taking expectations of (A6-A8) we have:

(A9) $$M_{11} = \sum_n \sum_s \sum_k (Z^1_{k,s,n} - \bar{Z}^1_{s,n}) P^n(k|s)\ (Z^1_{k,s,n} - \bar{Z}^1_{s,n})'$$

(A10) $$M_{21} = \sum_n \sum_s (Z^2_{s,n} - \bar{Z}^2_n)\ \rho\ P^n(s)\left[\bar{Z}^1_{s,n} - \sum_t \bar{Z}^1_{t,n}P^n(t)\right]'$$

(A11) $$M_{22} = \sum_n \sum_s (Z^2_{s,n} - \bar{Z}^2_n)\ P^n(s)\ (Z^2_{s,n} - \bar{Z}^2_n)'$$

Formulas (A9)-(A11), with probabilities evaluated at the sequential estimates of γ, were used in formula (A3) to produce the correct standard errors in Table 1. The "uncorrected standard errors" are simply the square roots of the diagonal elements of M^{-1}_{ii}.

REFERENCES

Abkowitz, Mark D., 1981, An Analysis of the Commuter Departure Time Decision, Transportation, 10, no.3, Sept., 283-297.

Amemiya, Takeshi, 1978, On a Two-Step Estimate of a Multivariate Logit Model, Journal of Econometrics, 8, 13-21.

Ben-Akiva, Moshe, E., 1974, Structure of Passenger Travel Demand Models, Transportation Research Record, 256, 26-41.

Berkman, J., D. Brownstone, et al, 1979, QUAIL 4.0 User's Manual, Computer Center, University of California, Berkeley.

Brownstone, David, 1980, An Economic Model of Consumer Durable Choice and Utilization Rate, Ph.D. Dissertation, University of California, Berkeley.

Brownstone, D., 1984, Information Criterion and Estimation of Misspecified Qualitative Choice Models, Industrial Institute for Economic and Social Research, Stockholm, Working Paper, August.

Cameron, Trudy, A., 1982, Qualitative Choice Modeling of Energy Conservation Decisions, unpublished Ph.D. Dissertation, Princeton University.

Cosslett, S., 1978, Efficient Estimation of Discrete-Choice Models from CHoice-based Samples, Ph.D. Dissertation, University of California, Berkeley.

Davidson, R. and J.G. MacKinnon, 1984, Convenient Specification Tests for Logit and Probit Models, Journal of Econometrics, 25, 241-262

Efron, B., 1982, The Jacknife, Bootstrap, and other Resampling Plans, CBMS-NSF Regional Conference Series in Applied Mathematics, Monograph 38, Society for Industrial and Applied Mathematics, Philadelphia.

Freedman, D.A. and S.C. Peters, 1984, Bootstrapping a Regression Equation: Some Empirical Results, Journal of the American Statistical Association, 79, no.385, , 97-107.

Goett, A., 1979, A Structured Logit Model of Appliance Investment and Fuel Choice, mimeo, Cambridge Systematics, Inc., West Berkeley, California.

Goldfeld, Stephen M. and Richard E. Quandt, 1972, Nonlinear Methods in Econometrics (North-Holland, Amsterdam).

Hausman, J.A. and D. McFadden, 1984, Specification Tests for the Multinomial Logit Model, Econometrica, 52, Sept., 1219-1240.

Hensher, D.A. and T. Moncfield, 1982, A Structured Logit Model of Automobile Acquisition and Type Choice, paper presented to Australian Transport research Forum, Hobart, Tasmania.

Hocherman, I., J. Prashkar, and M. Ben-Akiva, 1983, Estimation and Use of Dynamic Transaction Models of Automobile Ownership, Transportation Research Record, 944, 134-141.

McFadden, Daniel, 1981, Econometric Models of Probabilistic Choice, in: C.F. Manski and D. McFadden, eds., Structural Analysis of Discrete Data with Econometric Applications (MIT Press, Cambridge, Massachusetts).

McFadden, Daniel, Antti P. Talvitie, and Associates, 1977, Demand Model Estimation and Validation, Vol. V of The Urban Travel Demand Forecasting Project Phase I Final Report Series, Institute of Transportation Studies, University of California, Berkeley.

Rao, C.R., 1973, Linear Statistical Inference and its Applications, 2nd ed., Wiley, New York.

Small, Kenneth A., 1981, Ordered Logit: A Discrete Choice Model with Proximate Covariance Among Alternatives, Econometric Research Program Research Memorandum No. 292, Princeton University, December.

Small, Kenneth A., 1982, The Scheduling of Consumer Activities: Work Trips, American Economic Review, 72, June, 467-479.

Small, Kenneth A., and David Brownstone, 1982, Efficient Estimation of Nested Logit Models: An Application to Trip Timing, Econometric Research Program Research Memorandum No. 296, Princeton University (revised), December.

Theil, Henri, 1971, Principles of Econometrics (Wiley: New York).

Train, K., 1980, A Structured Logit Model of Auto Ownership and Mode Choice, Review of Economic Studies, 47, 357-370.

Train, K., 1985, Qualitative Choice Analysis: Theory, Econometrics, and an Application to Automobile Demand (M.I.T. Press, Cambridge, Massachusetts), forthcoming.