研究展望
Review

[PROC. OF JSCE No. 425/IV-14 (Infrastructure Planning and Management) January 1991]

Review

EFFICIENT ESTIMATION OF HIERARCHICAL LOGIT MODE CHOICE MODELS

By David A. HENSHER*

1. INTRODUCTION

Empirical mode choice studies are dominated by the estimation of a single level multinomial logit model on a choice set of mutually exclusive alternatives. In recognition of the limiting structural flexibility of the multinomial logit model, embedded in the condition of equal cross-alternative substitution, hierarchical structuring of single or multiple discrete decisions has been suggested as a means of accommodating a richer substitution pattern. The multinomial logit condition applies within a level of the hierarchy but is relaxed between levels. Despite the theoretical appeal of hierarchical modelling it has not been as widely used as was anticipated, especially in the decomposition of a single choice.

Studies which have applied the procedure have in the main been investigating multiple decisions, typically mode and destination choice in transport, and have sequentially estimated the structure as a linked set of multinomial logit models. Sequential estimation has been adopted in most studies because of the ability to use standard logit software. There are many practical and computational difficulties with sequential estimation that suggest the need for a serious effort to estimate hierarchical model systems simultaneously, regardless of the behavioural assumptions on the relationship between decisions. Full information maximum likelihood (FIML) estimation of any configured hierarchical structure for single or multiple decisions gives immediate asymptotically efficient estimates as well as saving the time and resources (and potential error) committed to the calculation of the linking indices and standard error corrections for a sequential estimator.

In this paper we apply efficient estimation to a hierarchical discrete choice model in the context of mode choice for urban commuter and intercity non-business travel and use the empirical studies to illustrate the potential benefits of efficient hierarchical logit model estimation. The paper is organised as follows. First we briefly outline the hierarchical logit model for mode choice and document some of the main advantages of efficient estimation compared to sequential estimation. Second, we justify the need for *yet another mode choice* model. Third we introduce the empirical setting including the specification of alternative hierarchical decision structures. Fourth, model estimation and interpretation of the results are presented, followed by a concluding section summarising the main contributions.

* Transport Research Centre and Graduate School of Management, Macquarie University, NSW 2109, Australia
Keywords: *mode choice, hierarchical discrete choice model, hierarchical logit, full information maximum efficient estimation*

2. AN HIERARCHICAL DISCRETE CHOICE MODEL

Consider a two-level mode choice decision involving two choice sets: the generic choice set (G) of private and public transport, and the elemental choice set (M) of car driver, car passenger, train and bus. The indirect utility function associated with an elemental mode m contained within the generic category g is given in equation (1).

$$U_{gm}=U_g+U_{m|g},\quad m\in M_g,\ g\in G \qquad (1)$$

The hierarchical structure is given in Fig. 1. The standard multinomial logit (MNL) model assumes that the ratio of the selection probabilities of any pair of modes is independent of any other mode(s) - that is, equality of all cross-elasticities, and so we can collapse the model structure into a single-level specification. The condition is associated with the assumption that the unobservable components of the indirect utility expression for each alternative are independently and identically distributed (IID). Hierarchical structuring is a mechanism for decomposing the single choice into *sub-choices* or clusters such that the MNL condition need not hold between clusters.

Let us write the U_{gm} expression in terms of component vectors for the observed (V) and unobserved $(\mu,\ \epsilon)$ influences on choice:

$$U_{gm}=V_g+V_{m|g}+\mu_g+\epsilon_{m|g},\quad m=1,\ \cdots,\ M_g,\ g=1,\ \cdots,\ G \qquad (2)$$

and define the variance-covariance matrix $\sum$ in terms of the associated elements

$$\textstyle\sum_{gm,g'm'}=E\,[(\mu_g+\epsilon_{m|g})*(\mu_{g'}+\epsilon_{m'|g})]. \qquad (3)$$

Invoking the independence assumption of the distributions in the two choice dimensions and random utility maximisation, numerous authors (e.g. McFadden 1981, Williams 1981) have shown that the joint probability of choosing alternative gm can be defined as equation (4), a hierarchical logit model:

$$P_{gm}=\frac{\exp[\lambda(V_g+V_{g*})]}{\sum_{g'\in G}\exp[\lambda(V_{g'}+V_{g'*})]}\ \frac{\exp[V_{m|g}]}{\sum_{m'\in M_g}\exp[V_{m'|g}]} \qquad (4)$$

where

$$V_{g*}=I_{g*}=(1/\lambda)\log\sum_{m'\in M}\exp[\lambda_{m'|g}]+\text{eulers's constant}. \qquad (5)$$

A global sufficiency condition for the hierarchical or nested choice model in equation (4) to be consistent with random utility maximisation is that the coefficient of inclusive value be in the 0-1 range (McFadden 1981). Under a local sufficiency condition proven by Boersch-Supan (1985), the parameter of inclusive value can be permitted to exceed unity. The parameter(s) of inclusive value(s) provide the basis for differences in cross-substitution elasticities as compared to the IID condition of MNL.

The majority of mode choice models estimated as a nested logit model use sequential estimation. This involves firstly the separate estimation of each choice situation as the lowest level in the hierarchy—CD vs CP and TN vs BS in Fig. 1, then the calculation of inclusive values I_{g*}, followed by estimation at the upper level(s). Because separate models are estimated on each branch at the lower level, there is potential for a considerable loss of information. This affects both the sample size and the distribution of the levels of the exogenous variables. Clearly one can only include in the lower level estimation those observations which

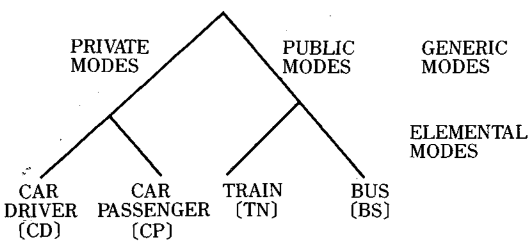


Fig. 1 An Hierarchical Modal Structure.

have a chosen mode and at least one alternative in that branch. In reality, mode choice observations typically have varying choice sets containing the chosen mode in one branch and the non-chosen mode(s) in another branch (Hensher 1986 a). Sequential estimation is inefficient in that the parameter estimates at all levels above the lowest level do not have minimum variance parameter estimates. This is the consequence of using *estimates of estimates* in the calculation of the contribution of aggregate utility or inclusive value indices of the subsets of alternatives below the relevant node. Inclusive values are not relevant for a degenerate node—node with a choice set of one. Correction procedures to obtain asymtotically efficient estimates from sequential estimation are extremely complex (McFadden 1981), especially for more than two levels. It is also inefficient because the task of calculating inclusive values and relating them to the upper level is quite demanding.

The literature gives the misleading impression that sequential nested logit modelling requires less data than simultaneous MNL. For example with 6 destinations and 4 modes there are 24 mode by destination combinations of data required to estimate an MNL model whereas a nested logit model of mode given destination requires 4 data points for each individual and destination choice requires up to 6 data points. However the calculation of inclusive value requires data points for all modes to all destinations in the choice set regardless of which destination is chosen. The ultimate saving in data is zero. The gains in nesting are linked to the variation in cross-alternative substitution.

The presence of degenerate nodes at the lower level also requires that data on such alternatives be included in the upper levels but not via inclusive values. This creates problems in interpretation, especially where a variable can appear in more than one level (Hensher 1986 a). Furthermore, if theory argues for certain conditions on the parameters of particular variables, such as the equality of cost coefficients throughout the structure (Truong and Hensher 1985), this cannot be accommodated in total—it can be handled *within a level* by stacking such that each branch is treated as a separate obervation. This however forces an *unranked* and hence generic specification on all explanatory variables. Alternative (type)-specific constants can however still be calculated for *types* of alternatives contained in each alternative in the unranked set.

Full information maximum likelihood (FIML) estimation avoids the problems of sequential estimation, but does require somewhat complex software. Two software packages are currently available for FIML nested logit modelling (ALOGIT by Daly (1985) and HLOGIT by Boersch-Supan, Smith and Milthorpe (Smith and Milthorpe 1989)). FIML also introduces some new issues. First the hierarchical (HL) model as a set of independently estimated MNL models has a unique optimal solution and no local suboptimal maxima for the likelihood function. FIML-HL however must have a unique optimal solution, unless there is a lack of identification, but there may be local suboptimal maxima. Great care must be exercised in the selection of starting values for all of the parameter estimates. Second, the optimisation procedure will accept analytical first derivatives, and analytical second derivatives but the latter are not available in simple form. Thus the practice has been to accept either approximate analytical second derivatives (Small and Brownstone 1981, Hensher 1986 a, Daly 1985) or numerical second derivatives (Matzoros 1982, Boersch-Supan 1984). Given the highly non-linear behaviour of the parameters of inclusive value, use of the first derivatives only is not recommended.

The full log-likelihood expression for a FIML hierarchical model extended to three levels is of the form:

$$\log L=\sum_{q=1}^{Q}\sum_{m\subseteq(1,\cdots,M)} S_{mq}\log[P(m\,|\,x_m,\ \beta,\ \theta,\ \lambda)] \quad (6)$$

where $q=1, \cdots, Q$ observations, $m=1, \cdots, M$ modes, θ and λ are the parameter estimates of inclusive value for the top and middle levels, S_{mq} equals 1 if mode m is chosen and zero otherwise, x_m are the set of other explanatory variables, and β their parameter estimates. For a three-level logit model of the form:

$$P_{agm}=P(a)*P(g\,|\,a)*P(m\,|\,a,\ g), \quad (7)$$

the joint cumulative distribution function of the errors is (McFadden 1979, 1981) :

$$F(\epsilon_1, \epsilon_2, \cdots, \epsilon_z)=\exp\{-G[\exp(-\epsilon_1), \cdots, \exp(-\epsilon_z)]\} \quad (8)$$

with

$$G(y_1, \cdots, y_z)=(\sum_{a}\sum_{g'\in a'}\sum_{m'\in g'} y_m'^{(1/\lambda)\lambda/\theta})^{\theta} \quad (9)$$

where we sum over the highest level containing alternative a', the subsets of alternative g' contained in each branch of the top level, and then over the elemental alternative m' in each of the middle level branches. Model form (4) is applied in the empirical studies for (i) the choice of mode for the urban journey to work and (ii) the choice of mode for intercity non-business travel.

3. EMPIRICAL APPLICATIONS

An individual's choice of transport mode has been extensively studied both as an isolated choice and as one of a set of interdependent travel decisions. The *popularity* of this most-studied trip is a consequence of three main influences. Firstly the belief that the physical capacity of the transport system is determined by the peak period traffic which in urban contexts is primarily commuter oriented. Secondly by the opportunity to observe individuals with at least two available alternatives with relatively well-defined trade-offs in terms of time and money, enabling calculation of empirical values of travel time savings—a major user benefit of transport improvements. Thirdly the relative ease of collecting data on the small set of universal alternatives has made it an attractive application of discrete-choice models. As the empirical *workhorse* of individual discrete-choice modelling, the modal choice problem has provided a rich experience of the strengths and weaknesses of these methods, especially the MNL specification which is itself the *workhorse* of discrete choice modelling.

Although there is a substantial literature on the wisdom of different discrete-choice approaches (Williams 1977, 1981, Hensher and Johnson 1981, Amemiya 1981, McFadden 1981, Ortuzar 1982, Maddala 1983, Langdon 1984) and of the over-emphasis on modal choice in the contexts of other travel decisions (e. g. destination, frequency, route choice—Tye *et al.* 1982) and broader travel-related activities (e. g. residential location choice, timing of activities choice—Jones 1979, Small and Brownstone 1982, Landau *et al.* 1982), there is still a need for better explanatory and predictive modal choice models. The interest in improved model specification in the current paper stems from the recognition that the independence of irrelevant alternatives (IIA) axiom of the MNL model includes mixes of public and private alternatives. The hierarchical logit model offers an empirically tractable alternative way of minimising violation of the IIA property. This concern is widely echoed in the literature. However very few empirical studies have structured the mode-choice decision in a logit context to accommodate this concern ; where this has occurred estimation has been sequential (Wildermuth 1982, Sobel 1980, Hensher 1983, McFadden 1981, and Ortuzar 1983) with the consequent problems as outlined above.

The previous sequentially estimated models have not corrected the asymptotic standard errors of parameters in the upper level of the nested structure although software is now available to do this (Greene 1990) whcih also corrects for an error in the formulae given in McFadden (1981). Cameron (1982, 1985), Small and Brownstone (1982) and McFadden (1981) discuss correction procedures for the covariance matrix of sequential model. McFadden (1981, 267-8) reports some results from Cosslett (1978) which estimate a two-level mode choice nested model using FIML ; however the coefficient of inclusive value does not satisfy the global sufficiency condition. Boersch-Supans's test for local sufficiency was not known at the time.

To illustrate the potential empirical benefit of FIML estimation of nested logit models we draw on two data sets collected in Australia. The first data set relates to a sample of commuters resident in the Sydney Metropolitan Area, who are faced with a choice of mode from a universal choice set of car as driver, car as passenger, bus and train. Since not all modes are physically available to all of the sample, we have a differential choice set. The second data set was collected as part of an intercity mode choice study and is specialised to the subsample of non-business travellers travelling between Sydney, Canberra and Melbourne. All travellers choose from among four modes—plane, car, coach and train. A number of nested logit models are estimated using FIML, and compared to the MNL specification. In an earlier paper (Hensher 1986) we compared sequential and FIML nested logit models using the urban commuter data set.

4. THE CHOICE OF MODE FOR THE URBAN JOURNEY TO WORK

An effective sample of 1 434 households were obtained from home interviews undertaken in the Sydney Metropolitan Area between September 1981 and May 1982. Details of the exogenously stratified sampling strategy are given in Hensher (1986). A distinction was made between trips to a fixed workplace on a regular basis, trips to fixed but different locations such as the casual employee or inspector, and travel as part of work. The modelling is limited to individuals who by the nature of their work and its location have available at least two alternative main means of transport. The captive mode users such as travelling salespersons and self-employed tradesmen are excluded in the formal modelling but must be identified as a representative percentage of the workforce in any projection on modal split.

The sample yielded 1 700 commuters who have a choice set of two or more modes of transport. The *universal* choice set for each individual varies according to where one lives and works. Watercraft modes (hydrofoil, ferry) for example would only be feasible for a small group. Full details have been complied on the *universal* set of alternatives available to each individual, which is a maximum of four modes out of the Sydney-wide offerings of ten modes (car driver, car passenger, train, bus, bicycle, walk, ferry, hydrofoil, motorbike and taxi). Unlike the majority of mode-choice studies, the current study has identified the *full set* of feasible alternatives *for each individual*, avoiding the problem of exogenising the decision on *choice of choice sets* which arises when the individual's *reported* set is used. Only when the reported set is a random subset of the full choice set is the problem removed. Reported choice sets are typically non-random.

The final data is reduced for sample size reasons to individuals whose chosen mode was one of car driver, car passenger, train and bus with the provision that one of the other three modes was an alternative. This deleted 245 commuters, leaving a total modelling sample of 1 455. This represents 70% of all work-related trips for the household sector.

The modal composition of the data is summarised in Table 1 and APPENDIX A, together with the variables used in estimation. Each mode is well represented in the choice sets (1 251-car driver, 792-car passenger, 779-train and 994-bus); however car driver accounts for 65.5% of chosen modal trips. The public transport modes are 29.1% of chosen trips. The standard level of service variables and socioeconomic descriptors are postulated as the main influences on mode choice; we have included two additional variables to pick up the influence that the incidence of business-registered vehicles in the household (NVHHS, PCSTT) has on the use of the car as a driver. We have identified in Hensher (1985) that business-registered vehicles have higher annual kilometers than privately registered vehicles (11 900

Table 1 Summary Statistics for Each Model Structure (standard deviation in parenthesis).

	Invehicle cost (cents)	Invehicle time (mins)	Walk time (mins)	Wait time (mins)	No. of choosers
Car Driver	65.46 (66.9)	28.65 (20.6)	0.76 (0.9)	0.15 (1.1)	953
Car Passenger	4.37 (26.5)	28.32 (22.4)	0.71 (2.9)	2.89 (3.1)	78
Train	98.23 (64.4)	43.84 (26.4)	10.50 (7.7)	8.37 (5.6)	279
Bus	81.61 (55.1)	38.15 (23.3)	7.47 (6.1)	7.11 (4.4)	145
Car Driver Only	Parking cost = 26.59 cents, Number of household business-registered vehicles = 0.186, Per cent of travel costs paid for = 17.58				

Note : These figures are for the subset of observations that have the alternative in the choice set.

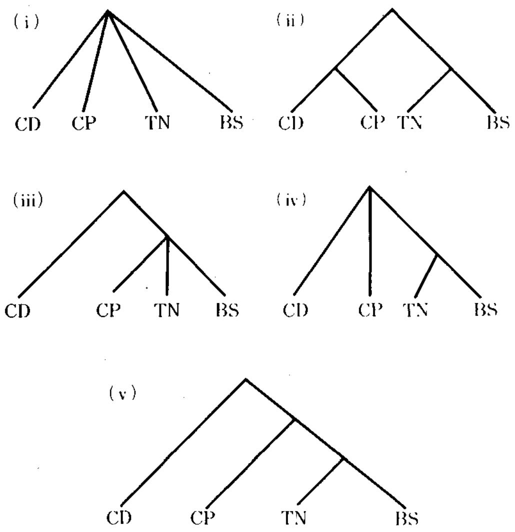


Fig. 2 Alternative Estimation Strategies.

private, 19 400 household business and 17 300 other-business), and that the percentage of such use attributed to work travel is also higher (36% private, 73% household business, and 76% other-business registered).

Five logit structures (Fig. 2) are postulated for comparative analysis. Structure (i) is nested within (ii)-(v), with structures (iii) and (iv) nested within (v). All structures are estimated simultaneously. Structures (i)-(iii) were also considered in Hensher (1986 a). Starting values for the FIML-HL are obtained from both the MNL model (i) using a preliminary search of the parameter space and intelligent guesswork (Cosslett 1978, 219). Given the lack of a unique optimum for FIML-HL, it is essential that the FIML-HL estimates are confirmed by more than one source of starting values. This strategy is also in accord with a number of other authors (e. g. Small and Brownstone 1982, Ortuzar 1983 and Ben-Akiva 1977), who have recommended a preliminary screening of models using the relatively simpler MNL or sequential nested-logit estimators.

Good starting values, in addition to the real goal of consistent estimates are the key to cost effectiveness in FIML-HL estimation. Evidence by Brownstone and Small (1985) using a Monte Carlo design suggests that if FIML-HL is used with arbitrary starting values care must be taken that the global optimum is reached since otherwise consistency is not guaranteed. We would not recommend the use of starting values from a sequential estimator unless each observation faces the full choice set (Hensher 1986 a).

Nested structure (ii) is apriori the most plausible. Train and bus are likely to display similar correlation patterns on the unobserved influences due to the nature of the unobservables. This same argument can be made for car driver and car passenger, although perhaps less strongly. Car passenger is an awkward mode - it has elements of both private and public transport in that it relies on the provider for service while having door-to-door convenience and a perceived low or zero marginal cost. It is fully under the control of a single household in the majority of cases. Hierarchical structures (iii), (iv) and (v) are included to assess the implications of both grouping car passenger with the public modes (iii) and treating it separately (as degenerate choices in two (iv) and three (v) level models). Dandy and Neil (1981) provide some evidence that *passenger* modes, as distinct from *driver* modes, have the common feature of inflexibility which is not shared with the driver mode, although the empirical results are not conclusive. Since the choice sets do vary across the sample, we apply Cosslett's exclusive deletion of nodes such that for an individual, a node is deleted if and only if all nodes connected to it from below have been deleted (Cosslett 1978).

The MNL model was estimated using BLOGIT (Hensher and Johnson 1981). The FIML-HL program is driven by Quandt's GQOPT 2 program, using the quadratic hill-climbing method GRADX. It takes that step at each iteration which maximises a quadratic approximation to a function $f(x)$ on a sphere of suitable radius (Goldfeld and Quandt 1972). First and second analytical derivatives of the log-likelihood function are calculated, and second derivatives using an approximation to the 2nd derivatives equal to $-E([\partial \log L/\partial\theta]*[\partial \log L/\partial\theta'])$. The convergence criterion for FIML-HL is that the norm of the gradient is less than the accuracy required by the user. All software is programmed to check for the eligible elements of the choice set.

5. ANALYSIS AND RESULTS

The parameter estimates and measures of goodness-of-fit for all five structures are given in Table 2. The set of significant influences on mode choice were identified by the MNL model and applied in all FIML-NL specifications. Socio-demographic variables were found to have little influence on the choice of mode, hence we confine analysis to the cost and level of travel service variables. NVHHS is a household level variable which is a constraint on the feasible choice set, and together with PCSTT is a proxy for the burden of travel cost on the household rather than on non-household sources. The expected positive sign for both variables which are alternative specific to car driver supports the hypothesis of a bias towards car as a driver.

The likelihood ratio (LR) test is an appropriate test for an exogenous sample. If we had selected a choice-based sample (which is not uncommon in commuter mode choice given the dominance of the car) then the LR test is not valid since the LR test statistic does not have a chi square distribution. Rather it is distributed as a weighted average of chi square variates with one degree of freedom. The lagrange multiplier test would then be preferred since it works with both choice-based and exogenous samples. It is

Table 2 Multinomial and Hierarchical Logit Models of Mode Choice (Full Information Maximum Likelihood, 1 455 Observations)

EXPLANATORY VARIABLE	ACRONYM	(i) [CD, CP, TN, BS]		(ii) [CD, CP], [TN, BS]		(iii) [CD], [CD, TN, BS]		(iv) [CD], [CP], [TN, BS]		(v) [CD], [CP (TN, BS)], [TN, BS]	
		estimate	t-value	estimate	t-value	estimate	t-value	estimate	t-value	estimate	t-value
Car driver dummy	ASC-CD	0.8937	4.86	0.9397	4.48	0.8699	4.15	0.9809	4.20	0.8461	4.42
Car passenger dummy	ASC-CP	−2.2154	−10.36	−2.4589	−8.79	−2.1817	−8.85	−2.6209	−5.56	−2.4361	−7.38
Train dummy	ASC-TN	1.3286	9.10	1.4040	8.41	1.3127	8.27	1.4278	7.34	1.2956	7.05
Invehicle time (mins)	INVT	−0.0227	−4.70	−0.0255	−4.48	−0.0224	−4.51	−0.0258	−4.11	−0.0217	−4.33
Wait time (mins)	WT	−0.1336	−6.68	−0.1489	−6.21	−0.1313	−6.27	−0.1504	−5.22	−0.1365	−6.25
Walk time (mins)	WLKT	−0.0672	−5.44	−0.0769	−5.17	−0.0658	−5.03	−0.0799	−4.61	−0.0695	−5.33
In-vehicle cost (cents)	INVC	−0.0063	−5.03	−0.0071	−4.72	−0.0062	−4.54	−0.0072	−4.18	−0.0065	−4.97
Parking cost for car driver (¢)	CDPC	−0.0086	−5.04	−0.0098	−4.71	−0.0084	−4.34	−0.0101	−4.09	−0.0086	−5.08
Number of household business regd vehicles : CD specific	NVHHS	0.4524	1.83	0.5289	1.84	0.4396	1.80	0.5230	1.77	0.4199	1.72
Percentage of travel cost covered by nonhousehold source : CD specific	PCSTT	0.0119	3.71	0.0131	3.45	0.0165	3.46	0.0138	3.28	0.0120	3.71
Inclusive Value (level 2)	INCLV 2	–	–	0.860	10.80	1.030	8.25	0.857	6.76	0.8915	6.65
Inclusive Value (level 3)	INCLV 3	–	–	–	–	–	–	–	–	1.1929	5.85
Log-likelihood at :											
zero : equal shares		−2 017.1		−2 017.1		−2 017.1		−2 017.1		−2 017.1	
zero : observed shares		−1 271.1		−1 271.1		−1 271.1		−1 271.1		−1 271.1	
convergence		−598.9		−596.9		−598.3		−597.8		−598.7	
Number of iterations		6		8		7		10		14	
Values of time savings (¢/person minute)											
invehicle time		3.59		3.56		3.61		3.55		3.34	
wait time		21.14		20.80		21.18		20.69		21.00	
walk time		10.63		10.74		10.61		10.99		10.69	

Note : (i)=1−Level, (ii)−(iv)=2−Level, (v)=3−Level also computationally much simpler since it does not require maximising the likelihood function of the generalised model, operating using only the maximum likelihood or WESML parameter estimates for the FIML-HL model.

The log-likelihoods at convergence of all models are virtually indistinguishable, as are the marginal rates of substitution between invehicle cost and each component of travel time. The likelihood-ratio test comparing all pairs of models $2[L_2^* - L_1^*] \sim \chi^2$ with T degrees of freedom fails to reject the null hypothesis of no statistically significant differences between each pair of models: the highest LR is 4.0 for models (i) and (ii).

Model Pair	LR	Model Pair	LR
i, ii	4.0	ii, iv	1.8
i, iii	1.2	ii, v	3.6
i, iv	2.2	iii, iv	1.0
i, v	0.4	iii, v	1.0
ii, iii	2.8	iv, v	1.8

The coefficients of inclusive value satisfy the global sufficiency condition for structures (ii) and (iv) but exceed unity for the 3-level model (1.19) and structure (iii), the latter equal to 1.03. From a statistical perspective the only hierarchical specification which is significantly different from an equivalent MNL form (where inclusive value=1.0) and is consistent globally with random utility maximisation is model (ii). Using a one-tailed test of the coefficient of inclusive value equal to 1.00 against it being less than one based on the asymptotic t-ratio with fully efficient parameters, we obtain 1.8 for model (ii) and 1.12 for model (iv). Although the equivalent value for models (iii) and (v) are not significantly different from 1.00, the inability to satisfy global sufficiency is tempting evidence to reject these specifications.

The application of Boersch-Supan's local sufficiency test found some evidence to suggest that models with inclusive values greater than 1.0 were consistent with random utility maximisation, but the results are not fully conclusive. Specifications (iii) and (v) which are a priori plausible are difficult to justify. A striking result is that mixing car passenger with public modes (train, bus) in (iii), and conditionally in (v) appears to be a major contributing influence to the inclusive value result, especially when we note that the only fully consistent models are those where car passenger is either paired with car driver (ii) or treated as a degenerate alternative in a two-level model (iv). There is little evidence from this data set in favour of the nested logit form compared to the simple MNL model.

A generalised definition of the elasticity of mode choice with respect to the k-th exogenous variable, g^* obtained from the hierarchical nested logit model, $P_s \cdot P_{b|s} \cdot P_{m|sb}$, is:

$$\partial \log P_{aht} / \partial \log g^*_{aht} = \partial P_{aht} / \partial g^*_{aht} \cdot g^*_{aht} / P_{aht} \quad (10)$$

$$= g^*_{sbm,k} \, x_k [\delta - P_{sbm}] \quad (11)$$

where

$$\delta = \Delta_{sv} \cdot \Delta_{bw} \cdot \Delta_{mx} \cdot 1/\lambda_{sb} + \Delta_{bw} \cdot \Delta_{mx} \cdot [1/\lambda_{sb} - 1/\tau_s] \cdot P_{m|sb} + \Delta_{mx} \cdot [\tau_s - 1] \cdot \tau_s \cdot P_{b|s} \cdot P_{m|sb} \quad (12)$$

and $\Delta_{sv}=1$ if $s=1$, otherwise$=0$, s, b and m define the three levels of an hierarchical-logit tree.

By the application of equation (10) we are able to derive direct and cross elasticities within and between levels of the hierarchical modal choice system. Elasticities can be obtained for each individual using sample enumeration, with weighted aggregate elasticities calculated as probability weighted individual-specific elasticities. The elasticities reported in Table 3 are derived using equation (11).

The cross-elasticities throw light on the consequences of adopting an MNL specification, and also highlight the sensitivity of such elasticities to different hierarchical structures. Elasticities at sample means are, in this study, quite different, highlighting the error due to naive aggregation of a non-linear function. The latter are given in parenthesis in Table 3. We limit the comparison to three models and two variables, although the implications transfer to all models and variables. The cross elasticities for the MNL model (i) are independent of the mode associated with the variable under consideration. The most notable effect of the hierarchical specification is the significant difference between the mean cross elasticities within the (TN, BS) branch and those calculated for an MNL model. The mean elasticities in

Table 3 Selective Direct and Cross Elasticities for Alternative Model Structures (i), (ii) and (iv)
(Naive Elasticities in parenthesis.)
(Results for all models and variables are available from the author.)

A. INVEHICLE TIME	WITH RESPECT TO ATTRIBUTE LEVEL OF											
ELASTICITY ASSOCIATED WITH	CD			CP			TN			BS		
	(i)	(ii)	(iv)	(i)	(ii)	(iv)	(i)	(ii)	(iv)	(i)	(ii)	(iv)
CD	−0.117	−0.121	−0.119	0.409	0.564	0.477	0.409	0.470	0.477	0.409	0.470	0.470
	(−0.181)	(−0.210)	(−0.206)	(0.344)	(0.474)	(0.390)	(0.344)	(0.386)	(0.390)	(0.344)	(0.386)	(0.390)
CP	0.034	0.059	0.033	−0.304	−0.382	−0.350	0.034	0.033	0.033	0.034	0.033	0.032
	(0.018)	(0.025)	(0.021)	(−0.320)	(−0.416)	(−0.363)	(0.018)	(0.020)	(0.021)	(0.018)	(0.020)	(0.020)
TN	0.170	0.187	0.189	0.170	0.187	0.189	−0.360	−0.416	−0.421	0.170	0.274	0.270
	(0.102)	(0.114)	(0.115)	(0.102)	(0.114)	(0.115)	(−0.428)	(−0.513)	(−0.520)	(0.102)	(0.177)	(0.181)
BS	0.072	0.072	0.073	0.072	0.072	0.073	0.072	0.135	0.138	−0.493	−0.602	−0.610
	(0.056)	(0.063)	(0.064)	(0.056)	(0.063)	(0.064)	(0.056)	(0.098)	(0.100)	(−0.509)	(−0.638)	(−0.641)
B. INVEHICLE COST												
CD	−0.077	−0.080	−0.080	0.253	0.351	0.301	0.253	0.292	0.301	0.253	0.292	0.300
	(−0.114)	(−0.132)	(−0.131)	(0.216)	(0.298)	(0.249)	(0.216)	(0.242)	(0.249)	(0.216)	(0.242)	(0.240)
CP	0.002	0.004	0.002	−0.013	−0.015	−0.015	0.002	0.002	0.002	0.002	0.002	0.001
	(0.001)	(0.001)	(0.001)	(−0.014)	(−0.018)	(−0.016)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
TN	0.098	0.108	0.111	0.098	0.108	0.111	−0.231	−0.267	−0.275	0.098	0.162	0.160
	(0.063)	(0.071)	(0.073)	(0.063)	(0.071)	(0.073)	(−0.266)	(−0.319)	(−0.328)	(0.063)	(0.110)	(0.110)
BS	0.042	0.043	0.044	0.042	0.043	0.044	0.042	0.082	0.085	−0.292	−0.356	−0.360
	(0.033)	(0.038)	(0.039)	(0.033)	(0.038)	(0.038)	(0.033)	(0.058)	(0.061)	(−0.302)	(−0.379)	(−0.391)

the hierarchical model are 60–70% higher for the train choice with respect to bus attributes, and 20% higher for the bus choice with respect to train attributes. The inclusion of CP and CD or its treatment as a singleton does not affect the cross or direct elasticities in any significant way within the (TN, BS) branch, but it does outside the public branch between CD and CP. These results support the argument that the cross-elasticities are not independent of mode and that the MNL model consistently underestimates the mean point elasticities, sometimes marginally but often quite significantly. This empirical study suggests that although mean generic values of travel time savings are insensitive to model specification, the direct and cross elasticities are very sensitive. The error in forecasting modal split using elasticities from an MNL model is clearly of concern.

6. INTERCITY MODE CHOICE FOR NON-BUSINESS TRAVEL

The second empirical study focusses on non-business travel between three major cities (Sydney, Canberra, Melbourne) in South-Eastern Australia. 118 individuals were surveyed in 1986 as part of a larger study into the feasibility of introducing a new high speed rail service. The choice set contained four modes: plane, coach, train and car. All four alternatives were included in each person's chocie set. The level of service data were obtained from a face to face interview. Two models were estimated—an MNL model and a three level hierarchical model. The results are summarised in Table 4. Other FIML-HL models were also estimated but are not reported herein. The hierarchical structure is of the form {plane vs slow (public [train vs coach] vs private [car])}. Theta (θ) is the coefficient of inclusive value which links the train vs bus choice to the public branch, and Tau (τ) is the inclusive value linking the public vs private choice to the slow branch. The air mode is degenerate at the lowest and middle levels, as is the car mode at the lowest level in the tree. Intuitively, individuals choose between the fast and slow modes, then within the slow modes they choose between public and private transport, and then within the slow public modes they choose between train and coach.

The exogenous variables include invehicle time (in hours), time spent waiting at public transport terminals (minutes), frequency of services (per hour in an 18 hour travel day, with car set to 1), the actual invehicle time (hours), household income (dollars per annum), number of persons in the travelling party, and the mode-specific constants for J-1 alternatives.

The FIML-HL model uses the MNL parameter estimates as starting values. A comparison of the two

Table 4 Intercity Non-Business Mode Choice Models (Alternatives : 1= plane, 2=train, 3=bus, 4=car. The Mode column indicates which alternative each variable is associated with).

		MNL MODEL		HIERARCHICAL MODEL	
Exogenous Variable	Mode(s)	Parameter Estimate	T-Value	Parameter Estimate	T-Value
Invehicle Cost	1-4	−0.006017	−0.85	−0.013705	−0.98
Terminal Time	1-4	−0.021389	−1.97	−0.011379	−0.66
Hhld Income	1	0.00232	1.97	0.004142	2.97
Size of Group	1	−0.495204	−2.10	−0.881091	−2.97
Service Freq.	1-4	−0.228895	−2.77	−0.109775	−0.70
Travel Time	1-4	−0.191950	−2.51	−0.116918	−0.81
MSC-Plane	1	−0.138229	−0.09	0.056551	0.030
MSC-Train	2	2.269314	2.88	1.907287	0.75
MSC-Bus	3	0.145447	0.25	0.972361	0.71
θ Public				0.097236	6.39
τ Slow				0.296311	1.86
Log likelihood		−125.5502		−96.9558	
Log likelihood at zero		−163.5827		−163.5827	
Pseudo-R^2		0.23		0.41	
Number of iterations		5		5	
No. of evaluations		42		43	
Mean square gradient		27.77		3.884	

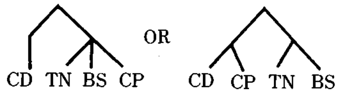
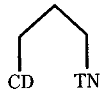
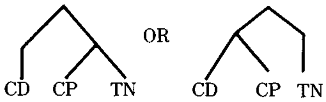
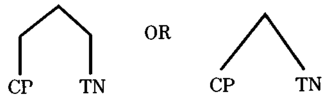
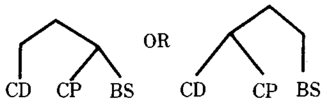
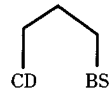
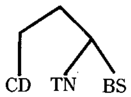
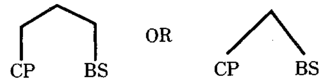
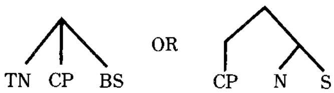
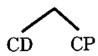
models in Table 4 suggests a substantial improvement in the overall goodness of fit of the model when FIML-HL replaces the MNL specification. On the likelihood ratio test at any generally acceptable level of confidence we can confidently reject the null hypothesis of no significant difference between the two models. We might have anticipated this in the intercity context (in contrast to an urban commuter context) given the greater variation in levels of service and possibly more binding financial constraint on the travelling family and hence the benefit of a conditional structure. The pseudo-R^2 increases from 0.23 to 0.41. The inclusive values θ (public) and τ (slow) are much closer to zero than unity, suggesting that the tree structure is justified relative to the MNL specification.

7. CONCLUSION

This paper has highlighted the econometric and empirical benefits of adopting full-information maximum likelihood estimation procedures for nested or hierarchical logit discrete choice modelling. Although the empirical illustrations have concentrated on two modal choice contexts, the arguments are applicable to all discrete-choice modelling in transportation.

Computer software is readily available to enable efficient estimation (and data preparation) of FIML hierarchical logit models so that researchers and practitioners are not *forced* to opt for the MNL form because of the *lack of suitable software*. Despite recent advances in the estimation of multinomial probit models, the nested-logit model is likely to be the most-practical method for minimising the unattractive properties of MNL models for quite some time. The opportunity to efficiently estimate hierarchical logit models is now available with ease, making sequential estimation unneccessary.

APPENDIX—Empirical Configurations Indicated by the Data

TREE STRUCTURES	NO. OF CASES THAT CHOOSE EACH MODE	TREE STRUCTURES	NO. OF CASES THAT CHOOSE EACH MODE
 CD TN BS CP OR CD CP TN BS	CD=6 CP=1 BS=0 TN=1	 CD TN	CD=132 TN=27
 CD CP TN OR CD CP TN	CD=108 CP=5 TN=7	 CP TN OR CP TN	CP=14 TN=49
 CD CP BS OR CD CP BS	CD=131 CP=8 BS=13	 CD BS	CD=269 BS=22
 CD TN BS	CD=161 TN=55 BS=13	 CP BS OR CP BS	CP=34 BS=67
 TN CP BS OR CP N S	CP=14 TN=13 BS=65	 CD CP	CD=146 CP=2
		 TN BS	TN=34 BS=28

REFERENCES

1) Amemiya, T. : Qualitative response models : a survey, Journal of Economic Literature, XIX, pp. 1483-1536, December 1981.
2) Ben-Akiva, M. E. : Passenger travel demand forecasting : applications of disaggregate models and directions for research, World Conference on Transport Research, Rotterdam, pp. 183-193, April 1977.
3) Boersch-Supan, A. : The Demand for Housing in the United States and West Germany : A Discrete Choice Analysis, Unpublished PhD thesis, Department of Economics, MIT, June 1984.
4) Boersch-Supan, A. : Hierarchical choice models and efficient sampling : with applications on the demand for housing, Methods of Operations Research, 50, pp. 175-186, 1985.
5) Brownstone, D. and Small, K. A. : Efficient estimation of nested logit models, School of Social Sciences, University of California at Irvine, June 1985.
6) Cameron, T. A. : Qualitative Choice Modelling of Energy Conservation Decision : A Microeconomic Analysis of the Determinants of Residential Space-Heating, unpublished PhD thesis, Department of Economics, Princeton University, 1982.
7) Cameron, T. : A nested logit model of energy conservation activity by owners of existing single family dwellings, Review of Economics and Statistics, LXVII (2), pp. 205-211, May 1985.
8) Cosslett, S. : Efficient Estimation of Discrete Choice Models from Choice-Based Samples, unpublished PhD thesis, Department of Economics, University of California, Berkeley, 1978.
9) Daly, A. J. : Estimating 'tree' logit models, Cambridge Systematics Europe BV, Den Haag, 1985.
10) Dandy, G. and Neil, R. : Alternative mathematical structures for modelling mode choice, Report No. R 30, Department of Civil Engineering, University of Adelaide, 1981.
11) Goldfeld, S. M. and Quandt, P. E. : Nonlinear Methods in Econometrics, North-Holland Publishing Company, Amsterdam, 1972.
12) Greene, W. H. : LIMDEP 5. 1, Econometric Software Inc., New York.
13) Hensher, D. A. : A sequential attribute dominance model of probabilistic choice, Transportation Research, 17 A (3), pp. 215-218, 1983.
14) Hensher, D. A. : An econometric model of vehicle use in the household sector, Transportation Research, 19 B (4), pp. 303-313, 1985.
15) Hensher, D. A. : Longitudinal surveys in transport : an assessment, in Ampt, E., Brog, W. and Richardson, A. J. (eds.) New Survey Methods in Transport, VNU Science Press, Utrecht, pp. 77-97, 1986.

16) Hensher, D. A. : Sequential and full information maximum likelihood estimation of a nested logit model, Review of Economics and Statistics, LXVIII (4), pp. 657-667, November 1986 a.

17) Hensher, D. A. and Johnson, L. W. : Applied Discrete-Choice Modelling, Croom Helm, London (Wiley, New York), 1981.

18) Jones, P. M. : New approaches to understanding travel behaviour : the human activity approach, in Hensher, D. A. and Stopher, P. R. (eds.) Behavioural Travel Modelling, Croom Helm, London, pp. 55-80, 1979.

19) Langdon, M. G. : Methods of determining choice probability in utility maximising multiple alternative models, Transportation Research, 18 B (3), pp. 209-234, June 1984.

20) Landau, U., Prashker, J. N. and Alpern, B. : Evaluation of activity constrained choice sets to shopping destination choice modelling, Transportation Research, 16 A (3), pp. 199-207, 1982.

21) Maddala, G. : Limited-Dependent and Qualitative Variables in Economics, Cambridge University Press, Cambridge, 1983.

22) Matzoros, A. : The Estimation of Nested Logit Choice Models, Unpublished MSc Thesis, Institute for Transport Studies, The University of Leeds, 1982.

23) McFadden, D. : Quantitative methods for analyzing travel behaviour of individuals : some recent developments, in Hensher, D. A. and Stopher, P. R., (eds.) Behavioural Travel Modelling, Croom Helm, London, pp. 279-318, 1979.

24) McFadden, D. : Econometric models of probabilistic choice, in Manski, C. F. and McFadden, D. (eds.) Structural Analysis of Discrete Data with Econometric Applications, M. I. T. Press, Cambridge, pp. 198-272, 1981.

25) McFadden, D. : Econometric analysis of qualitative response models, in Griliches, A. and Intriligator, M. D. (eds.) Handbook of Econometrics, Elsevier Science Publisher BV, Amsterdam, pp. 1395-1457, 1984.

26) Ortuzar, J. de Dios : Nested logit models for mixed-mode travel in urban corridors, Transportation Research, 17 A (4), pp. 282-299, 1983.

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

28) Smith, N. C. and Milthorpe, F. W. : Hierarchical logit models : a new software package, Proceedings of the 14 th Australian Transport Research Forum, Volume 1, Perth, Western Australia, pp. 439-457, September 1989.

29) Sobel, K. L. : Travel demand forecasting by using the nested multinomial logit model, Transportation Research Record 775, 48-55.

30) Truong, T. P. and Hensher, D. A. : Measurement of travel time values and opportunity cost from a discrete-choice model, Economic Journal, 95 (378), pp. 438-451, June 1985.

31) Tye, W., Sherman, L., Kinnucan, M., Nelson, D. and Tardiff, T. : Application of disaggregate travel demand models, National Cooperative Highway Research Program Report 253, Transportation Research Board, Washington D. C., 1982.

32) Wildermuth, H. K. : The Perth (transport) 2000 forecasting model-sensible hybrid, 7 th Australian Transport Research Forum, Hobart, pp. 477-504, 1982.

33) Williams, H. C. W. L. : On the formation of travel demand models and economic evaluation measures of user benefit, Environment and Planning A, 9 (3), pp. 285-344, 1977.

34) Williams, H. C. W. L. : Random theory and probabilistic choice models, in A. G. Wilson, J. D. Coelho, S. M. Macgill and H. C. W. L. Williams, Optimization in Locational and Transport Analysis, John Wiley and Sons, Chichester, pp. 46-84, 1981.

(Received December 19 1990)