

CHAPTER 6

**COMPARING DISCRETE CHOICE MODELS: SOME HOUSING MARKET
EXAMPLES**

Chapter 6 in D Pitfield (ed), London Papers in Regional Science, 1984,
Vol 14, 163-180 Pion, London

Comparing Discrete Choice Models: Some Housing Market Examples

P A LONGLEY
University of Reading

1 Introduction

Since the mid nineteen seventies there has been strong interest within various branches of social science in the adaptation of the discrete choice modeling methodology towards a wide range of research problems. This has required recognition of a wide variety of alternative decision-contexts (Landau et al. 1982) and behaviour-patterns (Lerman, 1979), and has also raised general issues concerning the variable extent to which individual or subgroup choices may be restricted by spatial and temporal constraints. Further interest has been expressed about the spatial and temporal transferability of alternative discrete choice models (Atherton and Ben-Akiva, 1976; Galbraith and Hensher, 1982). This substantive diversification has been accompanied by a variety of technical and methodological refinements of the multinomial logit (MNL) and multinomial probit (MNP) models, ranging from new estimation procedures (Hausman and Wise, 1978) to the development of less-restrictive, computationally tractable discrete choice model forms (for example, Williams, 1977; Daly and Zachary, 1978). Faced with both a wider selection of methodological tools and a broader spectrum of substantive enquiry, there exists a clear need for formal comparison procedures which the analyst can call upon to evaluate a given model specification or framework.

In this paper, I attempt to review briefly some trends amongst recent housing choice studies which employ discrete choice modeling methods. A new procedure is then presented (Hubert and Golledge, 1981; Halperin et al. 1984) which may be used to compare discrete choice models specified and/or structured in accordance with different a priori hypotheses. It is argued that this method fills a gap between existing discrete choice model comparison-procedures which are inapplicable to 'nonnested' model specifications, that is, to competing discrete choice models which comprise totally different variable specifications and that such procedures can usefully aid selection of the discrete choice model most appropriate to any given decision context.

2 The development of housing applications

From an historical perspective, it would appear that the initial impetus behind the development of discrete choice models of the housing market largely articulated growing dissatisfaction with the established space-distance models in the Alonso-Muth tradition, particularly with respect to their oversimplistic conception of the housing commodity and the notion

of instantaneous equilibrium of supply and demand (Maclennan, 1982). It has also been claimed that discrete choice models have advantages over the other aggregate spatial interaction approaches, because of their improved predictive and explanatory capabilities.

Although some early studies were concerned only with physical dwelling attributes and conventional sociodemographic characteristics, subsequent applications have developed choice contexts which embrace work location (Friedman, 1981), qualitative aspects of previous dwellings (Boehm, 1982), the learning process (Cronin, 1979), public service provision (Lerman, 1977), and a wide variety of other neighbourhood externality factors (Segal, 1979). Work in the United States of America (for example, Anas, 1982) is currently building upon these developments to integrate long-run and short-run considerations of housing supply and demand, although some recent British and continental European research suggests that these developments might be rather premature in their present form (O'Brien, 1982; Dieleman, 1983). These reservations express the widely held view that pervasive local and central government intervention in European housing markets conspires to sustain short-run and long-run disequilibria within each of a number of distinctive tenure-based submarkets. Moreover, it is recognised that the compounded effect of different interventionist policies within these various submarkets is to make utility functions reliant upon less conventional individual characteristics, alternative attributes, and interaction effects, such as wealth accumulation, tax relief, rate rebate eligibility, etc. Although the relative magnitude of choice-based and constraint-based factors itself forms the focus on an ongoing debate (Saunders, 1981), there is some consensus that the net effect of public policy upon dwelling selection is often to create distinct housing submarkets⁽¹⁾, which are defined along tenure lines in the first instance (Murie et al, 1976).

From a research standpoint, owner-occupation in Britain has been seen as an attractive tenure to which most sections of the population aspire. Access to this sector is initially dependent upon negotiation of loan finance from a building society (or other lending institution) which requires the borrower to fulfil a variety of formal criteria. These requirements are geared towards ability to repay the loan, and hence hinge upon the likelihood that the borrower will maintain an adequate and steady income over the entire repayment period. Seen from a constraint-based perspective, therefore, present income and socioeconomic status (whereby professional and white-collar workers are viewed as the lowest-risk borrowers) are therefore likely to be important determinants of the ability to gain access to the owner-occupied sector. If access to the owner-occupied sector is not feasible, choice becomes restricted to the rented sector.

(1) Although US research recognises the importance of tenure divisions, there is no need to recognise the existence of sustained disequilibria.

Within the rented sector, overt public policy intervention in dwelling choice is much more direct, since the bulk of the rented stock is owned and administered by local authorities. Allocation of local authority dwellings is not performed in accordance with any uniform national scheme, although preference is given to prospective tenants with large families and high housing needs, tenants who have been forced to move from their previous dwelling (which may also have been local authority owned), or tenants whose age or health necessitate purpose-built accommodation. To adopt a crude managerialist and constraint-based perspective, therefore, is to suggest an overall group-preference structure in which the initial split between owner-occupation and renting is made in accordance with the ability to get mortgage finance: subtenure selection for many of those who remain in the rented sector then proceeds in accordance with household needs and the ability to fulfil local authority eligibility criteria. Those who do not gain access to owner-occupied or local authority rented housing become restricted to choice within the private rented sector. This implies the sequential decision structure depicted in figure 1.

A further important problem facing housing market analysts concerns the representation of spatial choice alternatives and spatial choice sets in dwelling choice models. Representation of any extensive spatial choice set will generate very considerable computational difficulties, which cannot always be satisfactorily redressed by aggregation procedures (Tardiff, 1980), in which spatial alternatives are assigned to broad categories and are represented by a corresponding set of alternative-specific constant terms. Moreover, inclusion of the global choice set will be unrealistic if any subset of individuals have a restricted search space and do not consider all of the available alternatives. Although manipulation of the utility function offers a technical solution to this second problem (Anas, 1982), computational problems effectively dictate that the global choice set be segmented into feasible and nonfeasible choice subsets for the purpose of empirical investigation. Such choice-set segmentation presents the housing market analyst with problems akin to those of modeling tenure-based housing submarkets, discussed above.

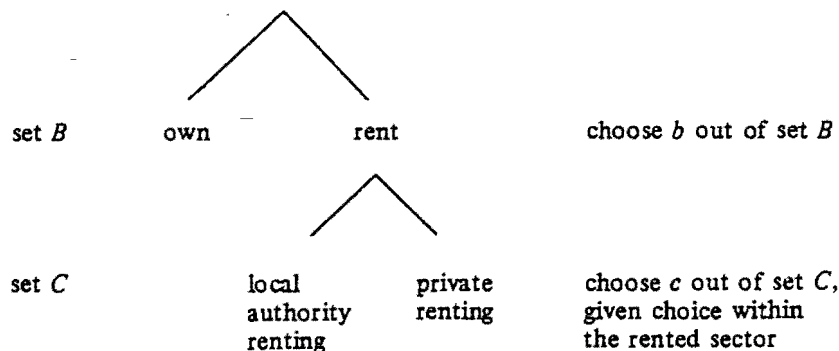


Figure 1. Nested formulation of the tenure-choice problem.

From the preceding discussion, it may be suggested that residential location modeling could gainfully be divided into a multistage choice sequence. In this way, feasible spatial choice sets may be restrained to manageable proportions, and constrained choices within tenure submarkets can be couched within an appropriate decision context. A range of broad methodological tools are now available to tackle these sorts of problems. In the period since the mid 1970s we have witnessed the development of a wide variety of alternative discrete choice model forms to obviate the conceptual difficulties posed by the independence from irrelevant alternatives (IIA) assumption of the multinomial logit model, on the one hand, and to circumvent the computational difficulties associated with multinomial probit, on the other (for a review, see Wrigley, 1984). In particular, it is recognised that the nested logit model may be used to structure a choice process into a series of sequential decisions, and transportation scientists have also exploited the computational convenience inherent in this model (Sobel, 1980).

The simple tenure-choice situation depicted in figure 1 illustrates a two-level decision structure, in which the choice whether to rent or to own a dwelling logically precedes tenants' choices between the local authority and private rented sectors. Adapting the notation developed by Wrigley (1984), the model for any individual can be seen to comprise two linked equations: the first of these gives the conditional probabilities of private or public sector renting for any individual whose choice is made within the rented sector; the second gives the marginal probabilities of the choice between owning and renting. Thus:

$$p_{c/rent} = \frac{\exp(V_c + V_{bc})}{\sum_{c^*=1}^C \exp(V_{c^*} + V_{bc^*})}, \quad (1)$$

where c stands for private or public renting,

$$p_b = \frac{\exp(V_b + \delta \tilde{U}_b)}{\sum_{b^*=1}^B \exp(V_{b^*} + \delta \tilde{U}_{b^*})}, \quad (2)$$

where b stands for owning or renting. V_c and V_b denote representative components of utility, which are specific to either the lower-level or the higher-level decision in figure 1, V_{bc} denotes components which relate to both nests in the hierarchy, and V_b also includes any attributes whose parameters cannot be identified within the lower nest. The \tilde{U}_b term is the inclusive value component which can be defined as

$$\tilde{U}_b = \ln \sum_{c^*=1}^C \exp(V_{c^*} + V_{bc^*}). \quad (3)$$

This term acts as the link between equations (1) and (2) and allows the analyst to assume that individuals taking decisions at the higher level of

the hierarchy will take account of the 'expected maximum utility' of the lower nest (for a full discussion see Williams, 1977; Daly and Zachary, 1978; Wrigley, 1984). This type of structure might be used as the starting point for the decomposition of the residential location decision into a number of component conditional decisions each of which represents a relevant decision context⁽²⁾.

Figure 2 represents just one such plausible choice hierarchy, in which constrained choices can be modeled alongside computationally manageable numbers of feasible spatial choice alternatives. Here, it is suggested that the initial decision represents tenure choice-allocation, and that subsequent housing decisions might be generalised along various tenure-specific sequences. For example, within the owner-occupied sector, it is reasonable to postulate a further two-tier structure comprising choice of dwelling type (for example, house, flat, or other) and subsequent residential location choice conditional upon this decision (for example, area A, B, or C). By contrast, there is evidence (Murie et al, 1976) that the more restricted margins of choice that are common within the local authority renting are centred primarily about spatial preferences, with some lesser subsequent provision for dwelling type preferences conditional upon the area in which the client is allocated a dwelling. Sequential decisions within the private rented sector are likely to reflect an amalgam of dwelling type preferences within the furnished and unfurnished sectors, with spatial choices dependent upon local availability.

However, adoption of a nested model based upon a specialised decision structure does imply some reliance upon an a priori hypothesis of the precise way in which choices are sequenced. Indeed, some recent housing location studies (Quigley, 1976; Lerman, 1979; van Lierop, 1981; Onaka and Clark, 1983) adopt rather different structures in modeling residential location and related decisions. To some extent, this problem can be resolved by careful reflection about the choice sequence used, and by judicious use of the empirical findings of existing studies: Onaka and Clark, for example, cite the empirical results obtained by Barrett (1973) to justify their sequencing of dwelling type and locational choice. However, given the claims that discrete choice models should offer improved predictive and explanatory power, it is clearly desirable that formal procedures should be available to facilitate comparison both of alternative model-structures. This issue is of importance if discrete choice models are to move successfully beyond idiographic study of uncoordinated decision situations; indeed, one of the early but as yet unrealised goals of disaggregate choice models was that well-specified models would be transferable in space and time.

(2) The presentation of the model reflects a 'bottom-up' estimation procedure contrary to the logical hierarchy of decisions.

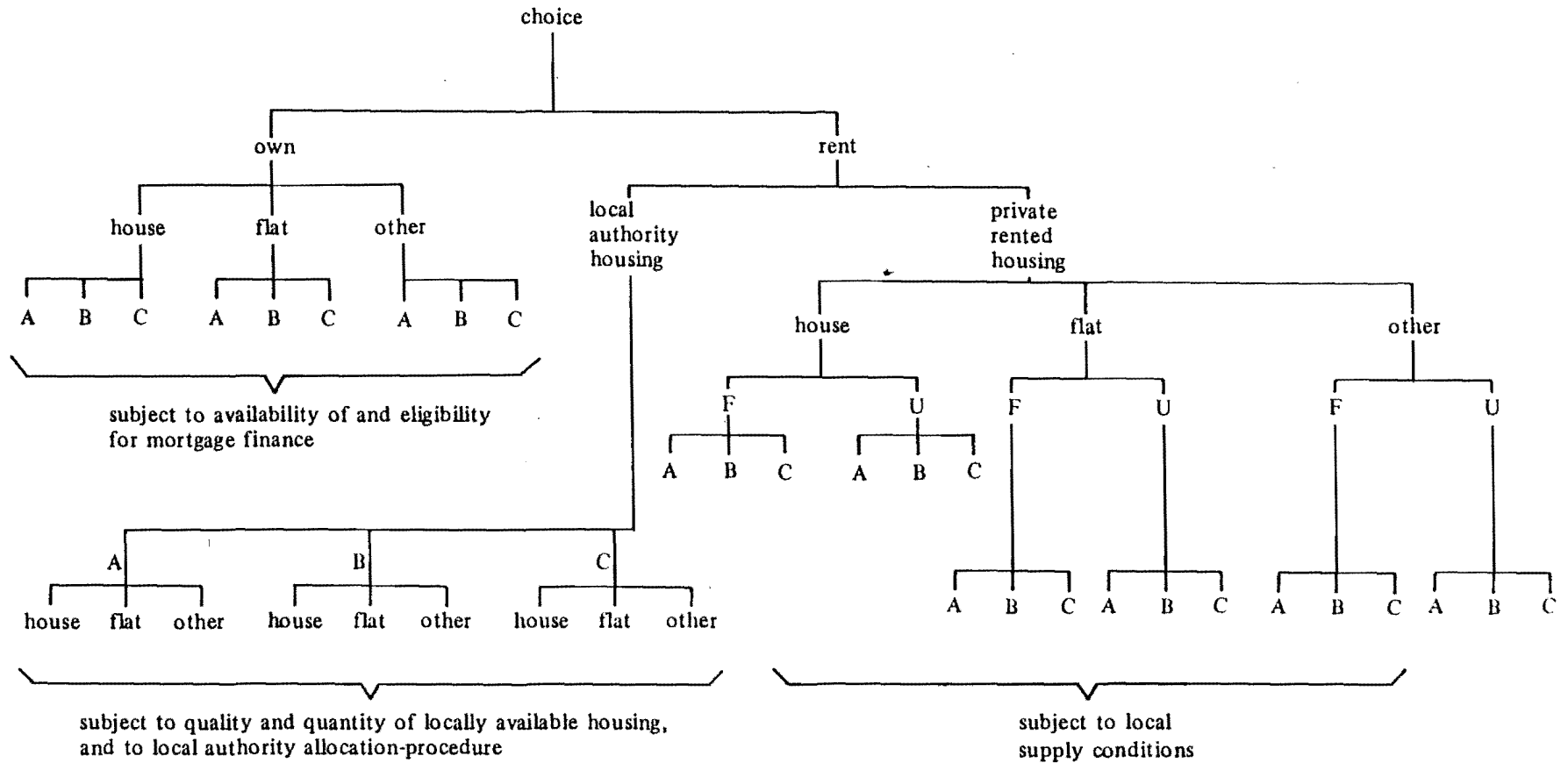


Figure 2. One hypothetical decision hierarchy for dwelling choice within a constrained housing market (A, B, and C are areas; U and F are unfurnished and furnished, respectively).

3 A heuristic matrix-comparison strategy

It should be stated at the outset that formal statistical tests can only fulfil a complementary role to sensitive selection of the most appropriate discrete choice model form (Wrigley, 1982), and awareness of a variety of contextual issues such as sampling method, variable representation, and choice of the optimal form of the utility function (Wrigley and Longley, 1984). However, the desire for complementary formal indices of model adequacy has stimulated the development of a number of procedures, ranging from relatively simple pseudo- R^2 goodness-of-fit measures (Domencich and McFadden, 1975), prediction success tables (McFadden, 1979), and residual plots (Wrigley, 1984) to tests of the IIA property of the multinomial logit model (McFadden et al, 1977), partitioned likelihood ratio statistics for nested models (Sobel, 1980), and a procedure for comparing logit and probit models (Horowitz, 1980). Relatively little attention has been devoted towards routine comparison of different discrete choice model specifications and structures. Indeed, the general point may be made that, although progress has previously been made towards comparison of competing 'nonnested' models (Cox, 1962), operational difficulties have resulted in a dearth of empirical applications of these methods within the wider social science literature.

An interesting general comparison strategy has been developed by Hubert and Golledge (1981), which extends the technique developed by Wolfe (1976) for testing the difference between two dependent correlations. This method has also been adapted and applied in a geographical context by Halperin et al (1984). Essentially, the method attempts to ascertain which of two models best represents a third matrix of observed choices. The two models may differ in the type of variables included, in the overall model-structure, or both. Thus a number of alternative simultaneous models might be compared, or a nested model formulated in accordance with an a priori hypothesis could be compared with a variety of alternative model-structures.

In brief, Wolfe demonstrated that, if three random variables, x_1 , x_2 , and x_3 follow some joint distribution for which $\text{var}(x_2) = \text{var}(x_3)$, then the correlation between the first and second variables ($\rho_{x_1x_2}$) is equal to that between the first and third variables ($\rho_{x_1x_3}$), only if x_1 and $(x_2 - x_3)$ are uncorrelated. Assessment of any difference between $\rho_{x_1x_2}$ and $\rho_{x_1x_3}$ consequently reduces to a test of zero correlation for $\rho_{x_1, (x_2 - x_3)}$. Thus

$$\rho_{x_1, (x_2 - x_3)} = \frac{\rho_{x_1x_2} - \rho_{x_1x_3}}{[2(1 - \rho_{x_2x_3})]^{1/2}}, \quad \rho_{x_2x_3} \neq 1. \quad (4)$$

Given three $n \times n$ proximity matrices, Hubert and Golledge have demonstrated that the matrix equivalent in empirical studies⁽³⁾ is given

⁽³⁾ In empirical studies, the assumption that $\text{var}(B) = \text{var}(C)$ is satisfied by standardising the entries of these matrices (matrix A is also standardised as a consequence, to maintain matrix compatibility).

by the expression

$$r_{A, (B-C)} = \frac{r_{AB} - r_{AC}}{[2(1 - r_{BC})]^{1/2}}, \quad r_{BC} \neq 1. \quad (5)$$

The applicability of this strategy to the comparison of alternative discrete choice models becomes clear if we postulate that **A** is a proximity matrix derived from the sparse matrix of observed choices. Elements a_{ik} of the proximity matrix **A** are given by

$$a_{ik} = \frac{1}{2} \sum_{j=1}^J |Y_{ji} - Y_{jk}|, \quad (6)$$

where Y_{ji} and Y_{jk} are the observed choices of alternative j by individuals i and k . **B** and **C** are proximity matrices generated from the predicted probabilities obtained from two alternative discrete choice models (b) and (c). Elements b_{ik} and c_{ik} of proximity matrices **B** and **C** are given by

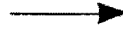
$$b_{ik} \text{ (or } c_{ik}) = \left[\sum_{j=1}^J (\hat{P}_{j|i} - \hat{P}_{j|k})^2 \right]^{1/2}, \quad (7)$$

where $\hat{P}_{j|i}$ denotes the predicted probability of individual i selecting alternative j . A simple worked example is presented in figure 3. Figure 3(a) might be the revealed choices of six individuals between two alternatives, such as owner-occupation and renting. Figure 3(b) might show the predicted probabilities obtained using a simultaneous logit model, and figure 3(c) might show the predicted probabilities obtained from an alternative nested logit structure. Matrix **A** is derived using expression (6), and matrices **B** and **C** are derived using expression (7): matrix **(B - C)** [figure 3(d)] is simply the difference between these two standardised matrices.

The matrix-comparison method seeks to compare the correlation $r_{A, (B-C)}$ with a reference distribution to test the null hypothesis that neither **B** nor **C** provide a better representation of matrix **A**. Applied to discrete choice models, this would suggest that there is no 'significant' difference between the predicted probabilities generated by two different models. The reference distribution provides a nonparametric method of assessing the possibility that no correlation exists between the standardised matrix of observed choices (**A**) and the difference between the standardised predicted probabilities produced by the two alternative discrete choice models (**B - C**). If this indeed proves to be the case, then it is reasonable to assert that neither model (b) nor model (c) presents a more adequate representation of the observed choices in model (a). The reference distribution for a sample comprising n observations is obtained by using a randomisation model in which the rows and columns of **(B - C)** are permuted and a total of $n!$ sample correlations calculated using all possible permutations (Mantel, 1967). In practice, this procedure becomes

Raw matrix of observed choices

		<i>j</i>	
		1	2
<i>i (= k)</i>	1	1	0
	2	1	0
	3	1	0
	4	0	1
	5	1	0
	6	0	1



Proximity matrix A

		<i>k</i>					
		1	2	3	4	5	6
<i>i</i>	1	0	0	0	1	0	1
	2	0	0	0	1	0	1
	3	0	0	0	1	0	1
	4	1	1	1	0	1	0
	5	0	0	0	1	0	1
	6	1	1	1	0	1	0

(a)

Predicted probabilities
obtained from model (b)

		<i>j</i>	
		1	2
<i>i (= k)</i>	1	0.4	0.6
	2	0.7	0.3
	3	0.5	0.5
	4	0.2	0.8
	5	0.9	0.1
	6	0.6	0.4



Proximity matrix B

		<i>k</i>					
		1	2	3	4	5	6
<i>i</i>	1	0	$\sqrt{0.18}$	$\sqrt{0.02}$	$\sqrt{0.08}$	$\sqrt{0.50}$	$\sqrt{0.08}$
	2	$\sqrt{0.18}$	0	$\sqrt{0.08}$	$\sqrt{0.50}$	$\sqrt{0.08}$	$\sqrt{0.02}$
	3	$\sqrt{0.02}$	$\sqrt{0.08}$	0	$\sqrt{0.18}$	$\sqrt{0.32}$	$\sqrt{0.02}$
	4	$\sqrt{0.50}$	$\sqrt{0.50}$	$\sqrt{0.18}$	0	$\sqrt{0.98}$	$\sqrt{0.32}$
	5	$\sqrt{0.08}$	$\sqrt{0.08}$	$\sqrt{0.32}$	$\sqrt{0.98}$	0	$\sqrt{0.18}$
	6	$\sqrt{0.08}$	$\sqrt{0.02}$	$\sqrt{0.02}$	$\sqrt{0.32}$	$\sqrt{0.18}$	0

(b)

Predicted probabilities
obtained from model (c)

		<i>j</i>	
		1	2
1	0.5	0.5	

Proximity matrix C

		<i>k</i>					
		1	2	3	4	5	6
1	0	$\sqrt{0.02}$	$\sqrt{0.08}$	$\sqrt{0.32}$	$\sqrt{0.18}$	$\sqrt{0.02}$	

computationally expensive given large numbers of observations, so the reference distribution may be constructed about a smaller prespecified number (for example 100) of random permutations, which is nevertheless large enough to assess critically the observed value of $r_{A,(B-C)}$. The null hypothesis that $r_{A,(B-C)} = 0$ is rejected if the observed value lies at one or other extreme of the reference distribution: an observed value in the upper tail suggests that B is the best representation of A; a value in the lower tail suggests that C provides the better fit. The level of significance attached to any extreme values produced by the randomisation procedure depends on the number (if any) of more extreme values produced by the randomisation procedure, relative to the number of permutations performed.

The fact that this strategy does not depend upon distributional assumptions can be seen as a strength, but is also a weakness insofar as it is not possible to correct for degrees of freedom. As a consequence, if a complicated model is shown to represent the original data better than a simple one, this technique cannot assess whether the improvement is significant in any absolute sense (of course, if a complicated model does not represent the original data better, then the simple model is the most parsimonious fit). This shortcoming restricts the scope of discrete choice modeling applications to model comparisons involving similar numbers of parameters. Thus, it could not be used to compare logit and probit models, but could be used to compare alternative multinomial logit structures or alternative nested schemas. In the following examples, this method is used to compare the efficiency of a number of alternative model specifications of a tenure-choice problem.

4 Some empirical examples

The discussion in section 2 suggested that state intervention in British housing markets can mould choices between renting and owner-occupation in accordance with a variety of policy criteria. It follows that choices between owner-occupation and renting and between renting within the private sector or within local authority sectors should each be related to different sets of individual characteristics and alternative attributes. Moreover, if some overall group preference exists for owner-occupation, it would prove intuitively appealing as well as analytically expedient to couch tenure choice within a nested logit framework as shown in figure 1. In this section, data obtained from 309 interviews as part of the 1976 English House Condition Survey (DoE, 1978; 1979) are analysed using the BLOGIT computer program (Crittle and Johnson, 1980) to present some preliminary evidence relevant to such a policy-manipulation hypothesis: the matrix-comparison procedure described in section 3 is then invoked to indicate the appropriateness of some alternative model-specifications and model-structures.

As a first stage in this empirical analysis, two logit models were used to assess the relationship between revealed tenure preference and selected

individual characteristics and housing tenure attributes. The models were of forms:

$$\ln\left(\frac{p_i^{LA}}{p_i^{PR}}\right) = \hat{\beta}_0 + \hat{\beta}_1 N_i + \hat{\beta}_2 L_i + \hat{\beta}_3 R_i + \hat{\beta}_4 F_i, \quad (8)$$

$\hat{\beta}_0 = 3.905^{**}$	$\hat{\beta}_1 = 1.635^{**}$	$\hat{\beta}_2 = 0.882$	$\hat{\beta}_3 = 0.530^*$	$\hat{\beta}_4 = 1.132^*$
(1.58)	(0.59)	(0.62)	(0.28)	(0.58)
{2.47}	{2.77}	{1.43}	{1.86}	{1.94}

where

standard errors are given in parentheses;

t-statistics are given in curly brackets;

** means that $t > 2.0$;

* means that $1.6 < t < 2.0$;

$\ln(p_i^{LA}/p_i^{PR})$ is the logarithmic probability that the revealed preference of individual *i* is to rent a dwelling in the local authority sector, rather than in the private sector;

N is the number of bedrooms required by the household according to English House Condition Survey standards (DoE, 1978; 1979);

L is a dummy variable recording whether (1) or not (0) the household is of a late stage in the life cycle (that is, about or above retirement age);

R is the household's assessment of the state of repair of the dwelling, recorded on a five-point scale (local authority responsibility for redecoration and maintenance of their dwellings inferring an a priori expectation that they will be in better condition than private rented sector equivalents);

F is a dummy variable recording whether (1) or not (0) the previous move was forced;

and

$$\ln\left(\frac{p_i^R}{p_i^{OO}}\right) = \hat{\beta}_0 + \hat{\beta}_1 C_i^D + \hat{\beta}_2 S_i + \hat{\beta}_3 I_i, \quad (9)$$

$\hat{\beta}_0 = 1.421^{**}$	$\hat{\beta}_1 = 0.118^{**}$	$\hat{\beta}_2 = -0.626$	$\hat{\beta}_3 = -0.116^{**}$
(0.48)	(0.02)	(0.54)	(0.02)
{2.94}	{8.11}	{-1.16}	{-6.72}

where

$\ln(p_i^R/p_i^{OO})$ is the logarithmic probability that the revealed preference of individual *i* is to rent rather than to own a dwelling;

C^D is the generalised annual cost of the dwelling (ignoring asset-accumulation attributes of owner-occupied dwellings);

S is a dummy variable recording whether (1) or not (0) the head of the household is a member of a professional or white-collar socioeconomic group;

I is the annual income of the head of the household.

Negative coefficients imply that a variable exhibits a positive relationship with local authority renting in equation (8), or a positive relationship with the aggregated renting alternative in equation (9).

The parameter estimates shown beneath equation (8) suggest that the revealed preference of local authority vis-à-vis private renting is more likely to be attained by large households, retired households (although the parameter is not statistically significant), and households who have been forced to move from their previous dwellings. Local authority landlord responsibility for redecoration and maintenance seems to enhance perceptions of the condition of public sector dwellings. These results correspond with a priori expectations about the type of need of clients for whom local authority dwellings were designed. The parameter estimates associated with equation (9) illustrate that the annualised costs of renting are above those of owner-occupation and that the revealed preference of owner-occupation is associated with high incomes and professional or white-collar socioeconomic status⁽⁴⁾. This suggests that magnitude and probable stability of income is associated with the revealed preference of owner-occupation and complies with an expectation that these variables influence the ability to acquire the necessary loan finance to enter this sector. These results thus lend some support to the notion that tenure choice between renting and owner-occupation should be logically separated from the choice between the private and local authority divisions within the rented sector.

Further and more direct support for this framework can be demonstrated if the variable specification used in equation (8) is shown to be inappropriate to resolving the choice situation shown in equation (9). This would suggest that the decision between renting from the local authority and renting within the private sector is based upon variables and criteria which are less appropriate to the decision whether to own or rent. With this consideration in mind, those variables which were most significant in equation (8) were used in a model which is directly comparable to equation (9), thus:

$$\ln\left(\frac{P_i^R}{P_i^{OO}}\right) = \beta_0 + \beta_1 N_i + \beta_2 R_i + \beta_3 F_i \quad (10)$$

$\hat{\beta}_0 = -0.873$	$\hat{\beta}_1 = -0.613^{**}$	$\hat{\beta}_2 = -0.288^*$	$\hat{\beta}_3 = 1.626^{**}$
(0.65)	(0.16)	(0.15)	(0.29)
{-1.34}	{-3.79}	{-1.95}	{5.56}

Predicted probabilities of owner-occupation versus renting were estimated for 154 individuals using the parameter estimates from equation (9) and equation (10) [see also table 1(a)]. These predicted probabilities were used to develop proximity matrix **B**, comprising proximity measures from equation (9), and proximity matrix **C** comprising measures from equation (10). The sorted distribution of 100 index values obtained by random permutation of the resulting (**B** - **C**) matrix is reproduced in table 2.

⁽⁴⁾ Exploratory analysis did not suggest that any strong correlation existed between income of head of household and the particular socioeconomic status classification used in this analysis: nevertheless, some multicollinearity may account for the failure of the socioeconomic status variable to breach statistical significance levels in analyses employing 300 observations.

Table 1. Parameter estimates and matrix comparison indices for the estimated models.

Mod ^a	Parameter estimate									No ^b	Statistics ^c				SL ^d		
	β_0	C_i^D	S_i	I_i	N_i	L_i	R_i	F_i	\tilde{U}_i		L(B)	ρ^2	OI	EV	90	95	99
(a)																	
B(9)	1.421*	0.118*	-0.626	-0.116*						4	-81.06	0.5749	5192	3925 (11.52)	☆	☆	☆
C(10)	-0.873				-0.613*		-0.288*	1.626*		4	-166.65	0.1260			-	-	-
(b)																	
B(11)	1.028	0.117*	-0.931	-0.103*	-0.572*	0.698	-0.097	1.393*		8	-71.75	0.6237	1923	328.1 (15.30)	☆	☆	☆
C(9)	1.421*	0.118*	-0.626	-0.116*						4	-81.06	0.5749			-	-	-
(c)																	
B(9)	1.421*	0.118*	-0.626	-0.116*						4	-81.06	0.5749	2426	2851 (-2.307)	-	-	-
C(12)	-1.516	0.120*	-0.711	-0.114*				-0.676*		5	-79.48	0.5832			☆	☆	-
(d)																	
B(12)	-1.516	0.120*	-0.711	-0.114*					-0.676*	5	-79.48	0.5832	-1675	-2651 (4.076)	-	-	-
C(13)	1.766*	0.114*	-0.675	-0.114*			1.244*			5	-76.94	0.5965			-	-	-
(e)																	
B(11)	1.028	0.117*	-0.931	-0.103*	-0.572*	0.698	-0.097	1.393*		8	-71.75	0.6237	4349	3179 (7.280)	☆	☆	☆
C(12)	-1.516	0.120*	-0.711	-0.114*				-0.676*		5	-79.48	0.5832			-	-	-

* denotes significant *t*-statistics.

^a Mod model; the numbers refer to the equations defining the models in the text.

^b No number of parameters.

^c L(B) log-likelihood; L(0) = 190.67; OI observed index; EV expected value; Z-scores are given in brackets; β -parameters were estimated for a sample of 309 results; matrix-comparison results use a stratified subsample of 154.

^d SL records the statistical significance level at which the starred model outperforms the competing specification.

The observed value of 5192 [table 1(a)] falls well beyond the upper tail of this distribution, suggesting that the model generated by equation (9) is a significant improvement over equation (10). This suggests that the choice between homeownership and renting is more adequately explained by the variables pertaining to building society allocation-criteria than the need-based criteria originally used to model subtenure choice within the rented sector. In short, homeownership is more clearly associated with the ability to pay than with household needs.

This result can be interpreted as offering some qualified support for an urban managerialist argument, insofar as the variables used appropriately reflect the role of local authorities in influencing choice. A related problem concerns verification of the sequential decision structure shown in figure 1. As a first step, the heuristic matrix-comparison strategy was used to compare equation (9) with a fully simultaneous model:

$$\ln\left(\frac{p_i^R}{p_i^{OO}}\right) = \beta_0 + \beta_1 C_i^D + \beta_2 S_i + \beta_3 I_i + \beta_4 N_i + \beta_5 L_i + \beta_6 R_i + \beta_7 F_i, \quad (11)$$

which included all of the variables in equations (9) and (10). The observed value of 1923 [table 1(b)] falls beyond the upper tail of the reference distribution, demonstrating that the inclusion of four additional variables causes a significant improvement in the model representation of the observed choices. However, given that the matrix-comparison procedure is incapable of correcting for degrees of freedom, it is not possible to assert whether equation (11) or equation (9) represents the most parsimonious

Table 2. Sorted distribution of index values obtained by permutation of the (B-C) matrix.

1699	1801	1825	1834	1850
1854	1859	1861	1872	1880
1884	1903	1907	1908	1920
1933	1934	1953	1959	1979
1979	1980	1981	1983	1996
2011	2021	2021	2033	2034
2036	2037	2040	2048	2049
2065	2068	2069	2071	2073
2073	2079	2080	2086	2086
2094	2097	2099	2101	2106
2107	2122	2130	2135	2141
2142	2143	2143	2144	2162
2164	2167	2168	2171	2177
2186	2194	2219	2223	2234
2235	2236	2239	2247	2253
2267	2290	2291	2308	2318
2323	2330	2341	2352	2358
2380	2385	2401	2419	2420
2431	2442	2449	2467	2481
2503	2559	2562	2563	2607

fit [conversely, however, had there been no significant difference between equation (11) and equation (9), then equation (9) would have been the more parsimonious model].

A more interesting comparison would seem to exist between the simple simultaneous model of equation (9) and a sequential nested logit structure:

$$\ln\left(\frac{p_i^R}{p_i^{OO}}\right) = \beta_0 + \beta_1 C_i^D + \beta_2 S_i + \beta_3 I_i + \beta_4 \tilde{U}_i, \quad (12)$$

[where \tilde{U}_i is an inclusive value term representing the 'expected maximum utility' of equation (8), derived in accordance with equation (3)], conforming to the choice situation depicted in figure 1. Equation (12) differs from equation (9) by the addition of an inclusive value term which uses the method developed in section 2 to capture the 'expected maximum utility' of the choice between local authority and private renting modeled in equation (8). Results obtained using the matrix-comparison method [table 1(c)] suggest that the addition of this inclusive value term produces an improvement in the fit of the model to the observed data at the 95% level. However, once again this improvement accrues at the expense of increased model complexity, and so definitive identification of the most parsimonious model is not possible.

Any more direct and interesting comparison must therefore equalise the number of parameters, and hence the degrees of freedom, used in the models. This is achieved [equation (13)] by comparing the nested model [equation (12)] with a variant of the basic simultaneous model [equation (9)] which includes the most statistically significant parameter from equation (10):

$$\ln\left(\frac{p_i^R}{p_i^{OO}}\right) = \beta_0 + \beta_1 C_i^D + \beta_2 S_i + \beta_3 I_i + \beta_4 F_i. \quad (13)$$

Comparison of equation (12) with equation (13) produces an observed index which falls towards the centre of the reference distribution of index values [table 1(d)]. It is not, therefore, possible to assert that inclusion of the composite 'expected maximum utility' variables improves the statistical performance of the basic simultaneous model any more than an equivalently parameterised model to which the F_i variable has been added. For completeness [table 1(e)], the nested model [equation (12)] was compared with the enlarged simultaneous model [equation (11)], yet the observed index of 4349 shows that the efficiency of the nested model was insufficient to counterbalance the improved predictive capability of a model with three additional parameters.

Taken together, these results offer some empirical support for the notion that choices between owner-occupation and renting are formulated on a different basis to choice between local authority and private dwellings within the rented sector. Table 1(a) clearly suggests that a simultaneous model tuned towards local authority allocation criteria is less appropriate

than equation (9) for modeling the split between owner-occupation and renting. The model-comparison procedure further demonstrates that the nested logit model [equation (12)] attains approximately the same level of efficiency as a model selected because of its anticipated level of statistical fit rather than its correspondence with any a priori hypothesis. In view of this, it is clear that the nested logit framework offers no obvious advantages of statistical efficiency in this example. Nevertheless, the inconclusive result recorded in table 1(d) has favourable implications for sequential disaggregation of the dwelling choice process into a number of tractable submodels, since the nested framework accrues no costs in terms of statistical inefficiencies. This suggests, for example, that the utility of spatial choices within one niveau of a model structure might also be effectively represented in a higher-level choice (for example, tenure choice) using the inclusive value term. In this manner, nested logit models may be used both to clarify a priori decision sequences and to ameliorate some of the considerable difficulties encountered in modeling spatial choice alternatives. The matrix-comparison strategy produces results which closely correspond to the conventional log-likelihood and ρ^2 indices (table 1), and yet allows important additional insights with regard to formal comparison of models with different variable specifications.

5 Discussion

This paper has briefly reviewed the development of discrete choice modeling approaches to housing applications, and has used a new method for assessing the statistical efficiency of alternative operational models of the housing market. The discussion in section 2 and preliminary results reported in section 4 suggest that the nested logit model has an important role to play, both as a way of rationalising hitherto intractable spatial choice situations, and of structuring choice sets in a theoretically appealing way. However, it is important to remember that a uniform decision structure across an entire sample was used in this analysis. This is clearly unrealistic in many circumstances, and it is important that the flexibility inherent in disaggregate choice models be used explicitly to allow variation in subgroup behaviour and differential subgroup decision situations. Moreover, this in itself poses further challenges (Hanson and Hanson, 1981), which are ultimately of a qualitative nature and cannot be entirely reduced to a mechanistic procedure. Nevertheless, formal model comparison procedures such as that described above can usefully supplement qualitative judgments and aid critical evaluation of any given model framework.

Acknowledgements. I am grateful to Dr Richard Dunn for assistance with the main computing routine used in this paper, which was devised by Mike Costanzo. Data derive from the 1976 English House Condition Survey: the Department of the Environment bears no responsibility for the views and analysis set out in this paper, which are wholly the responsibility of the author.

References

- Anas A, 1982 *Residential Location Markets and Urban Transportation: Economic Theory, Econometrics, and Policy Analysis with Discrete Choice Models* (Academic Press, New York)
- Atherton T J, Ben-Akiva M E, 1976, "Transferability and updating of disaggregate travel demand models" *Transportation Research Record* number 610, 12-18
- Barrett F, 1973 *Residential Search Behaviour* research monograph, York University, Downsview, Ontario
- Boehm T P, 1982, "A hierarchical model of housing choice" *Urban Studies* 19 17-31
- Cox D R, 1962, "Further results on tests of separate families of hypotheses" *Journal of the Royal Statistical Society Series B (Methodological)* 24 406-424
- Crittelle F J, Johnson L W, 1980 *Basic Logit (BLOGIT)—Technical Manual* ATM number 9, Australian Road Research Board, Nunawading, Victoria 3131
- Cronin J, 1979 *Low-income Households' Search for Housing: Preliminary Findings on Racial Differences* The Urban Institute, Washington, DC
- Daly A J, Zachary S, 1978, "Improved multiple choice models" in *Determinants of Travel Choice* Eds D A Hensher, M Q Dalvi (Teakfield, Aldershot, Hants) pp335-357
- Dieleman F M, 1983, "Tenure and allocation policy in the Tilburg housing market" *Tijdschrift voor Economische en Sociale Geografie* 74 162-174
- DoE, 1978 *English House Condition Survey 1976, Part 1: Report of the Physical Condition Survey* Housing Survey Report number 10, Department of the Environment (HMSO, London)
- DoE, 1979 *English House Condition Survey 1976, Part 2: Report of the Social Survey* Housing Survey Report number 11, Department of the Environment (HMSO, London)
- Domencich T, McFadden D, 1975 *Urban Travel Demand: A Behavioural Analysis* (North-Holland, Amsterdam)
- Friedman J, 1981, "A conditional logit model of the role of local public services in residential choice" *Urban Studies* 18 347-358
- Galbraith R A, Hensher D A, 1982, "Intra-metropolitan transferability of model choice models" *Journal of Transport Economics and Policy* 16 7-29
- Halperin W C, Richardson G D, Gale N D, Costanzo C M, 1984, "A generalised procedure for comparing models of spatial choice" *Environment and Planning A* 16 (forthcoming)
- Hanson S, Hanson P, 1981, "The travel-activity patterns of urban residents: dimensions and relationships to sociodemographic characteristics" *Economic Geography* 57 332-347
- Hausman J, Wise D A, 1978, "A conditional probit model for qualitative choice: discrete decisions recognising interdependence and heterogeneous preferences" *Econometrica* 46 403-426
- Horowitz J, 1980, "The accuracy of the multinomial logit model as an approximation to the multinomial probit model of travel demand" *Transportation Research B* 14 331-341
- Hubert L J, Golledge R G, 1981, "A heuristic method for the comparison of related structures" *Journal of Mathematical Psychology* 23 214-226
- Landau U, Prashker J N, Alpern B, 1982, "Evaluation of activity constrained choice sets to shopping destination choice modelling" *Transportation Research A* 16 199-207
- Lerman S R, 1977, "Location, housing, automobile ownership, and mode to work: a joint choice model" *Transportation Research Record* number 610, 6-11
- Lerman S R, 1979, "The use of disaggregate choice models in semi-Markov process models of trip-chaining behaviour" *Transportation Science* 13 273-291

- McFadden D, 1979, "Quantitative methods for analysing travel behaviour of individuals: some recent developments" in *Behavioural Travel Modelling* Eds D A Hensher, P R Stopher (Croom Helm, Beckenham, Kent) pp 279-318
- McFadden D, Train D, Tye W, 1977, "An application of diagnostic tests for the independence from irrelevant alternatives property of the multinomial logit model" *Transportation Research Record* number 637, 39-45
- Maclennan D, 1982 *Housing Economics* (Longman, Harlow, Essex)
- Mantel N, 1967, "The detection of disease clustering and a generalised regression approach" *Cancer Research* 27 209-220
- Murie A, Niner P, Watson C, 1976 *Housing Policy and the Housing System* (George Allen and Unwin, Hemel Hempstead, Herts)
- O'Brien L G, 1982 *Categorical Data Analysis for Geographical Research: With Applications to Public Sector Residential Mobility* unpublished PhD dissertation, Department of Geography, University of Bristol, Bristol
- Onaka J, Clark W A V, 1983, "A disaggregate model of residential mobility and housing choice" School of Architecture and Urban Planning, University of California, Los Angeles, CA
- Quigley J M, 1976, "Housing demand in the short run—an analysis of polytomous choice" *Explorations in Economic Research* 3 76-102
- Saunders P, 1981 *Social Theory and the Urban Question* (Hutchinson, London)
- Segal D, 1979, "A quasi-log linear model of neighbourhood choice" in *The Economics of Neighbourhood* Ed. D Segal (Academic Press, New York) pp 57-82
- Sobel K, 1980, "Travel demand forecasting with the nested multinomial logit model" *Transportation Research Record* number 775, 48-55
- Tardiff T J, 1980, "Definition of alternatives and representation of dynamic behaviour in spatial choice models" *Transportation Research Record* number 723, 25-30
- van Lierop W F J, 1981, "Towards a new disaggregate model for the housing market" in *Locational Developments in Urban Planning* Eds W F J van Lierop, P Nijkamp (Sijthoff and Noordhoff, Alphen aan den Rijn) pp 221-243
- Williams H C W L, 1977, "On the formation of travel demand models and economic evaluation of user benefit" *Environment and Planning A* 9 285-344
- Wolfe D A, 1976, "On testing equality of related correlation coefficients" *Biometrika* 63 214-215
- Wrigley N, 1982, "Quantitative methods: developments in discrete choice modelling" *Progress in Human Geography* 6 547-562
- Wrigley N, 1984 *Categorical Data Analysis for Geographers and Environmental Scientists* (Longman, Harlow, Essex)
- Wrigley N, Longley P A, 1984, "Discrete choice modelling in urban analysis" in *Geography and the Urban Environment: Progress in Research and Applications, Volume 6* Eds D T Herbert, R J Johnston (John Wiley, Chichester, Sussex) chapter 2, pp 45-94