On discrete choice under uncertainty

Rouwendal, J.

1990

SERIE RESEARCH MEMORANDA

ON DISCRETE CHOICE UNDER UNCERTAINTY :

A GENERALIZATION OF THE LOGIT MODEL AND ITS APPLICATION

J. Rouwendal

Researchmemorandum 1990-28 juni 1990

VRIJE UNIVERSITEIT
FACULTEIT DER ECONOMISCHE WETENSCHAPPEN
EN ECONOMETRIE
AMSTERDAM

ON DISCRETE CHOICE UNDER UNCERTAINTY :

A GENERALIZATION OF THE LOGIT MODEL AND ITS APPLICATION

Jan Rouwendal*
Dpt. of Economics
Free University
P.O. Box 7161
1007 MC Amsterdam
The Netherlands

* The author gratefully acknowledges financial support of the Dutch Association for Scientific Research (NWO) for part of the research reported here.

Abstract

In this paper we consider the consequences of realization uncertainty, i.e. uncertainty about the possibility of realization of a chosen alternative, for the formulation of discrete choice models and for empirically observable choice behaviour in the Dutch housing market. They turn out to be significant in both respects.

It is shown that the introduction of realization uncertainty introduces correlation between the random terms of the utilities and that the usual derivation of GEV-models breaks down in the presence of such correlation. As an alternative we propose an axiomatic approach which gives rise to a reasonable model, which is a generalisation of the logit model.

This model is applied to choice behaviour in the Dutch housing market. It is shown that the realization uncertainty has a significant influence on choice behaviour in this market, in the sense that people in general tend to avoid the most heavily rationed types of dwellings and that this effect is stronger when their original sitiation is worse.

1 Introduction

Discrete choice models have been used widely in recent years for analysis of choice behaviour of actors who are faced with a finite number of alternatives. Such situations occur e.g. in traffic mode choice (see e.g. Domencich and McFadden [1974]) and housing choice (see e.g. Anas [1982]). The models that are used in this analysis were often used earlier in other disciplines, but can sometimes be related to economic theory by interpreting them as the result of choice behaviour based on utility maximization (see McFadden [1981]). In particular, this is the case for the probit and generalized extreme value (GEV) models.

A tacit assumption made in these models is that the choice for a particular alternative will always lead to the realization of that alternative. Although this may be appropriate in many situations (e.g. in traffic mode choice), it may be less so in other circumstances. Examples can readily be given. A household that is searching for a particular type of dwelling may not be able immediately to find vacancies of the relevant type. A worker who wants to apply for a job may in the short run be faced with a lack of vacant positions which are attractive to him.

One may conjecture that uncertainty about the realization possibilities of particular choice alternatives may influence behaviour, and will, therefore, have consequences for the formulation of discrete choice models. In this paper we will start an investigation of this question by considering a situation in which an actor faces (known) realization probabilities for each of the alternatives with which he is confronted.

The paper is built up as follows. In the next section we sketch the general framework of discrete choice models that can be derived on the basis of utility maximization and discuss the consequences of introducing the uncertainty about realisation. In section 3 we derive an alternative model.

2 Utility Maximization under Uncertainty

In conventional discrete choice models it is assumed that the utility u_n attached to a particular alternative n is a random variable which can be written as the sum of its mean v_n and a stochast ϵ_n, which reflects the variation around it :

$$u_n = v_n + \epsilon_n, \qquad n=1,\dots,N. \tag{1}$$

In this equation N denotes the number of alternatives.

The probability π_n that alternative n will be chosen can, on the basis of utility maximizing behaviour, be determined as :

$$\pi_n = \text{Prob}(u_n > u_{n'} \ ; \ n'=1,\dots,N,\ n' \neq n), \qquad n=1,\dots,N. \tag{2}$$

Denoting the simultaneous distribution function of the ϵ_n's as F, and its first-order partial derivative with respect to ϵ_n as F_n, we can - on the basis of (2) - formulate the following expression for the choice probabilities :

$$\pi_n = \int_{-\infty}^{\infty} F_n(v_n+\epsilon_n-v_1, \ \dots \ , v_n+\epsilon_n-v_N).d\epsilon_n, \qquad n=1,\dots,N. \tag{3}$$

When F is the multivariate normal distribution, the resulting discrete choice model is known as the probit model, when it is a generalized extreme value distribution, the resulting model is called a GEV-model. The best-known member of this family is the logit model, which results when all ϵ_n's are independently and identically Weibull distributed. For a general discussion of these models we refer to McFadden [1984].

The question we seek to investigate in the present paper is : what changes in these models when the realization of the various alternatives is no longer guaranteed. This means that, in addition to the stochastic element in the model that is embodied in the ϵ_n's, a different form of uncertainty is introduced, which refers to the realization possibility of alternatives, once they are chosen. We will sometimes refer to this new stochastic element as 'realization uncertainty'.

The analysis that follows will be based on two assumptions. In the first place the individual actor knows the probability q_n that alternative n will be realized when he chooses it for all $n=1,\dots,N$. In the second place, the non-realization of alternative n will be identified with the realization of another alternative, indicated by an index 0. It is assumed that the utility v_0 attached to this alternative obeys the same formulation as that of the other alternatives (see equation (1)).

Some remarks concerning the interpretation of the model seem to be in order at this point.

1 The realization probabilities q_n may be individual-specific. This may be of importance in situations in which the model refers to search on the labour or housing market and employers or landlords have preferences over the various actors applying to their vacancy.

2 The 0-alternative may be identical to one of the alternatives 1,...,N, say n'. This may be of relevance e.g. in housing market search where one choice alternative may be to stay (voluntarily) in the dwelling presently occupied, while the 0-alternative means that one stays involuntarily (i.e. only because a preferred alternative was not availble) in that dwelling. The consequences of such a state of affairs may be explored by setting v_0 equal to $v_{n'}$ and assuming perfect correlation between the ϵ_0 and $\epsilon_{n'}$. It should also be remarked that identification of the 0-alternative with one of the other alternatives implies that this alternative can always be realized, and that $q_{n'}$ should therefore be equal to 1.

3 When actors make a choice from an initial position that can be identified with one of the choice alternatives, the choice probabilities can be interpreted as transition probabilities. In fact we are dealing with decision-making in a dynamic context. However, in the present paper we will confine ourselves to an analysis of the behaviour of one actor in one period. For this reason we do not have to deal explicitly with the dynamic effects.

In order to analyze the consequences of the realization uncertainty we will make the assumption that the actors behave on the basis of maximization of the expected utilities $\hat{u}_n$, which are defined as :

$$\hat{u}_n = q_n . u_n + (1-q_n) . u_0, \qquad n=1,\ldots,N. \tag{4}$$

Since the u_n's are random variables, this expression seems to be a little bit different from expected utility used in the normal sense of the term. However, it should be realized that for every actor the values of the ϵ_n's, and therefore of the u_n's are fixed. This implies that the situation in which the decision-making occurs is not so different from the conventional (von Neumann-Morgenstern) setting as may seem to be the case at first sight[1].

Substitution of (1) in (4) gives rise to the following expression :

$$\hat{u}_n = \hat{v}_n + \hat{\epsilon}_n, \qquad n=1,\ldots,N, \tag{5}$$

where $\hat{v}_n = q_n.v_n+(1-q_n).v_0$ and $\hat{\epsilon}_n = q_n.v_n+(1-q_n).v_0$. The significance of equation (5) is that it brings out clearly that the formal structure of the discrete choice model in the present context is similar to the conventional one (see equation (1)).

Notwithstanding this, discrete choice in the present situation is much more difficult to analyze because of the influence of the uncertainty about the realization of the alternatives on the variances and covariances on the stochastic components of the utilities. Denoting the covariance of ϵ_n and $\epsilon_{n'}$ as $\sigma_{nn'}$ and the variance of ϵ_n as σ_n^2, and using carets to indicate the situation of uncertainty, the following formula for the covariance $\hat{\sigma}_{nn'}$ that is relevant in the presence of realization uncertainty can be derived (see appendix A1) :

$$\begin{aligned}\hat{\sigma}_{nn'} = q_n.q_{n'}.\sigma_{nn'} &+ \\ &+ q_n.(1-q_{n'}).\sigma_{n0} + q_{n'}.(1-q_n).\sigma_{n'0} + \\ &+ (1-q_n).(1-q_{n'}).\sigma_0^2, \\ &n,n'=1,\ldots,N,\ n\neq n'.\end{aligned} \tag{6}$$

The corresponding formula for the variances is analogous (see appendix A1). Equation (6) brings out clearly that the covariances depend on the realization probabilities. Some remarks are appropriate.

1 The random terms $\hat{\epsilon}_n$ will in general be correlated, even when the random terms ϵ_n are not. This is caused by the fact that the expected utilities of all alternatives n for which $q_n<1$ depend in part on ϵ_0.

2 Only when ϵ_0 is identically zero (i.e. u_0 is not a random variable) and all ϵ_n's are independently distributed will the $\hat{\epsilon}_n$'s be independent of each other.

3 When the ϵ_n's are uncorrelated and have equal variance σ^2 the

covariances $\hat{\sigma}_{nn'}$ will be equal to $[(1-q_n).(1-q_{n'}) + q_n.q_{n'}].\sigma^2$ and the variances $\hat{\sigma}_n^2$ will be equal to $[(1-q_n)^2+q_n^2].\sigma^2$. This implies that, as a consequence of the realization uncertainty, the covariances $\hat{\sigma}_{nn'}$ will be of the same order of magnitude as the variances $\hat{\sigma}_n^2$.

4 When alternative 0 can be identified with one of the alternatives 1,...,N, say n', the resulting formula's for the variances and covariances can be found by substitution of n' for 0 and setting $q_{n'}=1$.

It will be clear from the discussion given above that the correlations between the various alternatives in the model may be radically changed as a consequence of the uncertainty, even though the formal structure of the model was shown to be maintained.

Let us now examine the consequences for the best-known examples of discrete choice models more closely.

1 The Probit model. With respect to this model the introduction of uncertainty implies that the structure of the variance-covariance matrix should be adapted to the formula's given in (6) for the covariance, and in the appendix for the variance. This seems to be possible without causing great problems. The consequences of the introduction of uncertainty for the probit model therefore seem to be limited.

2 GEV-models. These models are based on a distribution function F of the form :

$$F(\epsilon) = \exp\{-G(e^{-\epsilon_1}, \ldots, e^{-\epsilon_N})\}, \qquad (7)$$

where G should satisfy a number of restrictions (see e.g. McFadden [1977] for the complete list). For our purposes one of the requirements is of particular importance, viz. that it should be homogeneous of degree 1 in the variables $\exp(-\epsilon_n)$. This property is of crucial importance for the derivation of the choice probabilities π_n as :

$$\pi_n = \frac{e^{v_n}.G_n(e^{v_1}, \ldots, e^{v_N})}{G(e^{v_1}, \ldots, e^{v_N})}, \qquad (8)$$

where G_n denotes the first-order partial derivative of G with respect to $\exp(v_n)$. The function G reflects the mutual correlations of the ϵ_n's.

It is shown in appendix A2 that the introduction of realization uncertainty with respect to the realization of the various alternatives destroys the homogeneity property of the fuction G and that equation (8) is no longer valid. Moreover, it seems impossible to arrive at an alternative closed form.

One may hope to be able to reach more useful results when looking at a specific member of the family of GEV-models, viz. :

3 The logit model. This model results from the specification of G as $\sum_n \exp(-\epsilon_n)$ and gives rise to a distribution function of the form :

$$F(\epsilon) = \exp(-\sum_{n=1}^{N} e^{-\epsilon_n}). \qquad (9)$$

However, it turns out that even in this simple case one arrives at an integral for which no analytical solution is known. Again, we refer to appendix A2 for the relevant derivations.

The discussion above forces us to conclude that the introduction of realization uncertainty within the framework of discrete choice models causes complications which are hard to overcome for models of the GEV-type. For probit models the consequences seem far less severe. One may therefore be tempted to conclude that the probit model is the appropriate choice for the analysis of discrete choice behaviour whenever realization uncertainty is present. The problem with this conclusion is that the practical usefulness of the probit model is limited to cases where the number of alternatives (N in the present context) is small. Although the computational methods used are continually improving Maddala's [1984] conjecture that 'it is doubtful that the multinomial probit model is worth all the computational trouble when the number of choices is greater than four' [p. 64] is still relevant. It may be concluded therefore that there is a need for models that can deal with at least some of the consequences of realization uncertainty in order to see whether this phenomenon is of empirical relevance. This will be the subject of the next section.

3 An Alternative Model

The multinomial logit model is by far the most widely used discrete choice model. The distribution function of its error terms has already been given above (see (9)), the choice probabilities that correspond with it are :

$$\pi_n = e^{v_n} / \sum_{n'=1}^{N} e^{v_{n'}} \qquad (10)$$

$$n=1,\ldots,N.$$

The most important property of this model is the so-called independence of irrelevant alternatives (IIA) which is directly related to the simplicity of its functional form and can be formulated as saying that the ratio of two choice probabilities, say n and n', is independent of the inclusion of other alternatives in the choice set. It is indeed easy to see from (10) that :

$$\pi_n/\pi_{n'} = e^{v_n - v_{n'}}, \qquad (11)$$

$$n,n'=1,\ldots,N.$$

Although the theoretical implications of the IIA property are awkward (cf. the celebrated red bus/blue bus example), the computational tractability and empirical robustness of the logit model have made it by far the most popular discrete choice model.

In the previous section it has been shown that the introduction of realization uncertainty seems to make the analytical derivation of the model impossible. Since there is nevertheless some reason to look for tractable models that capture some of the consequences of the presence of such uncertainty, we will in the present section explore an alternative route for formulating such models, viz. the investigation of the consequences of some plausible conditions that can be imposed on the choice probabilities.

Since we want to arrive at a model that can be applied as easy as the conventional logit model we will use as our first condition a modified version of the IIA-property. As has been discussed in section 2 (see equation (5)), the relevant utilities in the presence of realization uncertainty are the $\hat{u}_n$'s, which depend on v_n,v_0 and q_n (and the random terms ϵ_n and ϵ_0) and not the u_n's which depend only on v_n (and ϵ_n).

Analogous to (13) we could therefore formulate the condition $\pi_n/\pi_{n'}=\exp[\hat{v}_n-\hat{v}_{n'}]$. However, it turns out that this condition is too strict for our purposes, and we will therefore use a more general one, viz. :

Condition 1

$$\pi_n/\pi_{n'} = f_n(v_n,v_0,q_n)/f_{n'}(v_{n'},v_n,q_{n'}), \qquad n,n'=1,\ldots,N. \tag{12}$$

When $f_n=\exp[q_n.v_n+(1-q_n).v_0]$ this gives the narrower condition mentioned above. It is easy to see (by summation over n) that condition 1 implies choice probabilities of the form :

$$\pi_n = f_n(v_n,v_0,q_n)/\sum_{n'=1}^{N} f_{n'}(v_{n'},v_0,q_{n'}), \qquad n=1,\ldots,N, \tag{13}$$

which is indeed close to the conventional logit model.

In order to find a meaningful second condition on the choice probabilities we return to the definition of the choice probabilities (2) and observe that it implies for the case of uncertainty :

$$\pi_n = \mathrm{Prob}(\hat{u}_n>\hat{u}_{n'};\ n'=1,\ldots,N,\ n'\neq n), \qquad n=1,\ldots,N. \tag{14}$$

The inequality $\hat{u}_n>\hat{u}_{n'}$ can be written out as :

$$q_n.u_n + (1-q_n).u_0 > q_{n'}.u_{n'} + (1-q_{n'}),u_0. \tag{15}$$

Now consider the special case in which q_n equals $q_{n'}$. Inequality (15) then changes into :

$$u_n > u_{n'}, \tag{16}$$

which is the inequality that is valid in the conventional case (i.e. without realization uncertainty). This motivates :

Condition 2 When all realization probabilities q_n are equal, the choice probabilities are the same as in the absence of realization uncertainty.

Taking the logit model as our point of reference, it follows that equation (12) should, in the special case considered in condition 2 be equal to (10). This implies that (12) can be written as follows :

$$\pi_n/\pi_{n'} = [g(q_n,v_0).e^{v_n}]/[g(q_{n'},v_0).e^{v_{n'}}] \qquad (17)$$

and that the discrete choice model can be written as :

$$\pi_n = e^{v_n+\log[g(q_n,v_0)]} / \sum_{n'=1}^{N} e^{v_{n'}+\log[g(q_{n'},v_0)]} . \qquad (18)$$

Equation (18) is a generalisation of the logit model which is easy to interpret. According to this equation the presence of realization uncertainty influences choice behaviour in the same way as changes in the utilities v_n do. When all q_n's are equal to 1, the conventional logit model results. When the realization probability $q_{n'}$ becomes smaller than 1 the resulting change in choice behaviour is the same as when the systematic utility $v_{n'}$ had been decreased by an amount $g(1,v_0)-g(q_{n'},v_0)$.

One may wonder whether any other useful restriction can be placed on the form of (18). One that suggests itself immediately from the discussion above is that the choice probabilities π_n should be increasing in q_n, n=1,...,N. Although this condition is intuitively appealing, it cannot be motivated as a consequence of the maximization of expected utility. In appendix A3 it is shown that - somewhat surprisingly - the sign of the partial derivative $\partial\pi_n/\partial q_n$ in general cannot be determined on the basis of expected utility maximization.

A final condition concerns the role of v_0. In appendix A3 it is shown that the following statement is valid under rather general conditions :

Condition 3 The probability that the alternative with the highest realization probability will be chosen is a decreasing function of v_0 ; the probability that the alternative with the lowest realization probability will be chosen is an increasing function of v_0.

This condition is also intuitively appealing. When non-realization of the chosen alternative is less harmful, the uncertainty will be of smaller influence on choice behaviour.

To see what this condition implies for our model, we should return to equation (18). Condition 3 points to a combined influence of q_n and v_0. When the function g(.) is continuous in its two arguments condition 3 implies that the change in g(.) caused by an increase in v_0 should be a decreasing function of q_n. Assuming differentiability, this gives :

$$\frac{\partial^2 g(q_n, v_0)}{\partial q_n \partial v_0} < 0. \qquad (19)$$

One should note that condition 3 excludes the possibility that g is additive (possible after a transformation) in its two arguments. If it would be additive, there would be no influence of v_0 on choice behaviour.

It should be remarked that a model satisfying conditions 1-3 seems to give a plausible description of decision making in the presence of uncertainty : (i) in the absence of such uncertainty the model becomes equal to the multinomial logit model which is known to give an empirically useful framework for the analyses of discrete choices; (ii) changes in the realization probabilities have effects that are comparable to changes in the systematic utilities v_n; (iii) lower values of a realization probability imply a smaller choice probability for the associated alternative; (iv) a lower utility of the 0-alternative implies that people become more inclined to choose an alternative with a high realization probability.

On the other hand, it should be realized that the generalized logit model that has been presented above is not in all respects an ideal one. Consistency with utility maximization is a desirable characteristic for economic models of choice behaviour. However, it can be shown that the model of equation (18) does not have this property (see appendix A.4).

This result is a little bit surprising in view of the fact that two of our conditions were derived explicitly on the basis of utility maximizing behaviour. The implication seems to be that the one condition that has not been derived in this way, the weak form of independence of irrelevant alternatives (condition 1) is incompatible with utility maximazation in the presence of realization uncertainty. It seems useless, therefore, to search for models that can be as easily used as the logit model and are at the same time able to deal with expected utility maximization in the presence of realization uncertainty. This suggests that the present model is close to the best we can get if we want to deal in a relatively easy way with realization uncertainty.

In conclusion, it seems fair to state that a model based on the conditions 1-3, although not entirely satisfactory, may be useful in studying some of the effects of realization uncertainty on choice behaviour. In the next section it will be applied for this purpose.

4 Empirical Application

In the present section we will apply the model formulated above to the analysis of housing choice behaviour on the heavily regulated rented segment of the Dutch housing market. Rent control on this part of the market has resulted in a situation of permanent excess demand. The demand is allocated over the dwellings that become vacant by local authorities, that use a number of different rules which are usually not completely transparant. Also the number of dwellings that become vacant in a certain period differs. The resulting situation for the searching household is probably best modelled as one of uncertainty. Since excess demand for some types of dwellings is much more substantive than for others it is well known that the probability that a move to a desired type of dwelling can be made within a reasonable period of time is dependent on the choice of the dwelling type.

Earlier attempts to model choice behaviour on regulated housing markets by means of incorporating uncertainty into discrete choice models have been made in Anas and Cho [1987] and Rouwendal [1988]. In the former an *ad hoc* formulation of the logit model is adopted, which can, in the present notation, be written as :

$$\pi_n = e^{q_n.v_n+(1-q_n).v_0+\alpha.\log(q_n)} \Big/ \sum_{n'=1}^{N} e^{q_{n'}.v_{n'}+(1-q_{n'}).v_0+\alpha.\log(q_{n'})} \qquad (20)$$

$$n=1,\ldots,N.$$

This formula differs from the conventional logit model by introducing the sum of the expected value of the utility that will be reached when a certain alternative is chosen and a term $\alpha.\log(q_n)$ which is incorporated in order to deal with risk-avoidance. This formulation looks intuitively as appealing as the model (18) proposed above. However, it does not satisfy condition 2. Anas and Cho give no results of estimation of their model, although it seems to have been used (presumably with 'guesstimated' coefficients) in their tentative simulation exercises.

Rouwendal [1988] uses a model which is essentially similar to the one proposed here. The analysis of the present paper should be viewed as an extension of this work. The main difference between the earlier results and the ones presented here are (i) the explicit introduction of u_0 in the function to be estimated and (ii) the use of data for various regions instead of the Dutch Rimcity alone.

The utility function has been specified as follows :

$$u_n = \alpha_{0n} + \alpha_1.\log(m/k) + \alpha_2.\log(y-p), \qquad n=1,\dots,N, \tag{20}$$

where m denotes the number of persons in the household, k the number of rooms in the dwelling, y after-tax household income and p the rent to be paid for the dwelling. The coefficients α_1 and α_2 are expected to be positive.

For the function $\log[g(.)]$ the following specification has been chosen :

$$\log[g(.)] = \beta_1.f(q_n) + \beta_2.f(q_n).u_0, \tag{21}$$

with f an increasing function. Both coefficients are expected to have a positive value.

A complication arises because of the queuing effects that occur as a consequence of the disequilibrium situation. As long as the searching households persist in their choice for a particular type of dwelling, even though realization is not immediately possible, it must be expected that households that have chosen for a dwelling with a low realization probability will be overrepresented in the sample.

The effect will be that the coefficient β_1 gets a negative bias. In extreme situations this may cause β_1 to become negative.[2] The effect of queuing thus counteracts that of risk avoidance. One possible solution to this problem is to select only those households that have been searching for a short period only. However, this reduces the size of the sample drastically. We have therefore chosen to estimate the model on the complete sample and to introduce dummy variables for households that have been searching for a long time. More specifically, we did not simply estimate one coefficient β_1, but :

$$\beta_{11} + \beta_{12}.d_1 + \beta_{13}.d_2 + \beta_{14}.d_3, \qquad (22)$$

where d_1 is a dummy for households that have been searching at least half a year, d_2 for those that have been searching at least one year and d_3 for those that have been searching for more than a year. We expect the coefficient β_{11} to be positive and the others to be negative. As a second modification we introduced the possibility that behaviour in the third region, the so-called Rimcity, where housing market problems are concentrated, is somewhat different from that in the rest of the country. For this purpose we used a fourth dummy, d_4, referring to inhabitants of region 3. We add a term $\beta_{15}.d_4$ to the expression (22) and estimate :

$$\beta_{21} + \beta_{22}.d_4 \qquad (23)$$

instead of β_2.

Appendix B gives some information about the data material that has been used. For maximum-likelihood estimation we used the program GRMAX.[3] It turned out that a logarithmic transformation of the realization probabilities gave the best results. For reasons of brevity we present only estimates for the final specification that has been chosen. Some remarks are appropriate.

1 The coefficient β_{11} is positive and significantly different from zero. This implies that avoidance of the heavily rationed alternatives is existing.

2 The coefficients β_{12}, β_{13} and β_{14} are all negative as they should be. This indicates that we would have found a smaller value for β_{11} if we restrict the sample to contain only those households that have been searching for a long time. As has been explained above, the negative sign of these coefficients should be associated with queueing. The fact that there is no increase in absolute value of these coefficient may be interpreted as the effect of a reconsideration of the initial choice which leads to a change in preferences towards a type of dwelling that is more

Table 1 Results of Maximum Likelihood Estimation

α_{0n}	constant		α_{19}	7.51 (12.8)
α_{12}		2.51 (5.9)	α_{110}	10.41 (17.3)
α_{13}		2.81 (5.4)	α_{111}	-4.53 (11.9)
α_{14}		-14.05 (18.7)	α_{112}	-2.14 (5.5)
α_{15}		-10.64 (14.4)	α_{113}	-2.22 (4.5)
α_{16}		-10.38 (13.3)	α_{114}	-14.64 (19.0)
α_{17}		-19.46 (17.9)	α_{115}	-12.06 (16.0)
α_{18}		-18.33 (16.4)	α_{116}	-11.84 (14.7)
α_2	log(#rooms/#persons)			32.84 (19.9)
α_3	log(income - rent)			13.13 (8.3)
β_1	log(realisation prob.)			
	β_{11}			1.15 (3.4)
	β_{12}			-0.25 (1.3)
	β_{13}			-0.25 (1.3)
	β_{14}			-0.12 (0.8)
	β_{15}			-0.92 (2.4)
β_2	log(realisation prob.)*u_0			
	β_{21}			-0.021 (3.2)
	β_{22}			0.036 (4.2)

number of observations : 942

\- log(likelihood) : 1965.46

readily available.

3 The coefficient β_{21} is positive as we would expect on the basis of the discussion in section 3. It indicates that households that are in a relatively bad housing situation are more inclined to choose for a dwelling type that is relatively easy to find than households whose housing situation is good.

4 The values of the coefficients β_{15} and β_{22} indicate that the effects of

realization uncertainty on choice behaviour are almost non-existant for region 3, the Rimcity. Since the excess demands are largest in this region this may perhaps be viewed as an indication that people regard is as quite normal that they have to wait for months or years before they will be able to find a dwelling of their preferred type. This confirms our earlier results for the Dutch Rimcity, that suggested that only effects of queing were present in this part of the Netherlands.

5 Conclusion

The findings of the present paper can be summarized as follows :

1 The introduction of realization uncertainty complicates the derivation of discrete choice models by means of additive random utility maximization, since it introduces correlation between the random parts of the utility functions.

2 The consequences for the probit model are modest, but this model is difficult to use when the number of choice alternatives is large.

3 The analytical derivation of the GEV-models, including the logit model, which are easier to use, breaks down.

4 For this reason we tried to derive an alternative model on the basis of three 'plausible' conditions. The first one is a version of the independence of irrelevant alternatives, which is a crucial property of the popular logit model. The other two are explicitly derived as consequences of expected utility maximization.

5 The resulting model turned out to be a generalization of the logit model and is intuitively appealing as a description of choice behaviour under realization uncertainty.

6 The model could be shown to be incompatible with utility maximization, the implication being that independence of irrelevant alternatives is inconsistent with utility maximization.

7 However, the derivation of an alternative, easy-to-use model seems to be a very hard job. The generalized logit model seems to be the only one that is able to give an indication of the possible effects of realization uncertainty on choice behaviour.

8 The application of the model to choice behaviour in the heavily regulated Dutch housing market indicated that the disequilibrium in this market has significant effects on choice behaviour. These effects were much less significant for the Dutch Rimcity, where housing market problems are

concentrated. This may be interpreted as a consequence of adaptation of the people to a permanent situation of disequilibrium.

In short : the paper has shown that realization uncertainty has important consequences, both for empirically observable choice behaviour and for the formulation of discrete choice models.

Notes

1 The usual loss of the ordinal character of the utility as a consequence of the expected utility formulation holds (of course) also here.

2 On the basis of a fixed number b of households that start searching in each period and fixed realization probabilities the total number of households searching for a prticular type of dwelling will be equal to $\sum_{o}^{\infty}(1-q_n)^t.\pi_n.b = \pi_n.b/q_n$. When the choice probability π_n equals $\exp(w_n)/\sum\exp(w_{n'})$, the proportion of households searching for a dwelling of type n in the total population of searching households will not be equal to π_n, but to $[\exp(w_n)/q_n]/\sum[\exp(w_n)/q_n]$.

3 GRMAX is a FORTRAN-based general maximum likelihood program. It should be noted that the packages for logit models that are available cannot be used in the present context because of the appearance of the multiplicative term $\log(q_n)*u_0$.

Appendix

A Derivations

A1 The variance-covariance matrix under realization uncertainty

Let σ_n^2 be the variance of ϵ_n and $\sigma_{nn'}$ the covariance between ϵ_n and $\epsilon_{n'}$. The variance of $\hat{\epsilon}_n$ will be denoted as $\hat{\sigma}_n^2$ and can be derived as follows :

$$\begin{aligned}
\hat{\sigma}_n^2 &= E[(\hat{\epsilon}_n - E(\hat{\epsilon}_n))^2] \\
&= E[(\hat{\epsilon}_n)^2] \\
&= E[(q_n.\epsilon_n + (1-q_n).\epsilon_0)^2] \\
&= q_n^2.E(\epsilon_n^2) + 2.q_n.(1-q_n).E(\epsilon_n.\epsilon_0) + (1-q_n)^2.E(\epsilon_0^2) \\
&= q_n^2.\sigma_n^2 + 2.q_n.(1-q_n).\sigma_{n0} + (1-q_n)^2.\sigma_0^2.
\end{aligned}$$

The symbol E denotes the mathematical expectation. For the first three equalities we have made use of the fact that $E(\hat{\epsilon}_n)=0$, of the definition of $\hat{\epsilon}_n$.

For the covariance $\hat{\sigma}_{nn'}$ we derive analogously :

$$\begin{aligned}
\hat{\sigma}_{nn'} &= E[(\hat{\epsilon}_n - E(\hat{\epsilon}_n)).(\hat{\epsilon}_{n'} - E(\hat{\epsilon}_{n'}))] \\
&= E[\hat{\epsilon}_n.\hat{\epsilon}_{n'}] \\
&= E[(q_n.\epsilon_n + (1-q_n).\epsilon_0).(q_{n'}.\epsilon_{n'} + (1-q_{n'}).\epsilon_0)] \\
&= q_n.q_{n'}.E(\epsilon_n.\epsilon_{n'}) + q_n.(1-q_{n'}).E(\epsilon_n.\epsilon_0) + \\
&\qquad + q_{n'}.(1-q_n).E(\epsilon_{n'}.\epsilon_0) + (1-q_n).(1-q_{n'}).E(\epsilon_0^2) \\
&= q_n.q_{n'}\sigma_{nn'} + q_n.(1-q_{n'}).\sigma_{n0} + q_{n'}.(1-q_n).\sigma_{n'0} + \\
&\qquad + (1-q_n).(1-q_{n'}).\sigma_0^2
\end{aligned}$$

The last expression is the one given in the text.

A2 GEV-models under realization uncertainty

On the basis of the definition of $\hat{\epsilon}_n$, which has been given just below equation (5) we can derive the following relationship :

$$\epsilon_n = [\hat{\epsilon}_n - (1-q_n).\epsilon_0]/q_n$$

Substtution of this equation in (7) allows us to find the density function of the $\hat{\epsilon}_n$'s for a given value of ϵ_0 :

$$F(\hat{\epsilon}|\epsilon_0) = \exp(-G(e^{-[\hat{\epsilon}_1-(1-q_1)\epsilon_0]/q_1}, \ldots, e^{-[\hat{\epsilon}_N-(1-q_N)\epsilon_0]/q_N})$$

Making use of the homogeneity of G, we can rewrite this as :

It should be observed that G is not homogeneous in the variables $\exp(-\hat{\epsilon}_n)$. In order to be able to use (3) for the derivation of the choice probabilities we have to compute the partial derivative of F with respect to $\hat{\epsilon}_n$:

$$F_n(\hat{\epsilon}|\epsilon_0) = e^{-\epsilon_0}.e^{-(\hat{\epsilon}_n-\epsilon_0)/q_n}.G_n(\ \ldots\ ).\exp(-e^{-\epsilon_0}.G(\ \ldots\ )),$$

where G_n denotes the first-order partial dervative of G_n with respect to $\exp[-(\hat{\epsilon}_n-\epsilon_0)/q_n$. In order to derive the choice probability π_n one should insert for the variables $\hat{\epsilon}_{n'}$ the values $\hat{v}_n+\hat{\epsilon}_n-\hat{v}_{n'}$, and integrate over (3). In the case without realization uncertainty this integration is possible because of the homogeneity of the function G in the variables $\exp[-\epsilon_n]$ (see McFadden [1977]). In the present situation an analogous procedure would require homogeneity of G in the variables $\exp[-\hat{\epsilon}_n]$. Since this homogeneity does (except in very special cases) not exist, we have to conclude that an expression like the one given in (8) can not be arrived at.

In order to investigate the possibility that for certain formulations of G one may, nevertheless, arrive at an analytic expression for the choice probabilities, we consider the case of the multinomial logit model. The relevant distribution function can be determined as :

$$F(\hat{\epsilon}|\epsilon_0) = \exp(-e^{-\epsilon_0}.\sum_{n'=1}^{N} e^{-(\hat{\epsilon}_{n'}-\epsilon_0)/q_{n'}}),$$

and its partial derivative with respect to $\hat{\epsilon}_n$ as :

$$F_n(\hat{\epsilon}|\epsilon_0) = \exp(-e^{-\epsilon_0}.\sum_{n'=1}^{N} e^{-(\hat{\epsilon}_{n'}-\epsilon_0)/q_{n'}}).e^{-\epsilon_0}.e^{-(\hat{\epsilon}_n-\epsilon_0)/q_n}.(1/q_n).$$

Inserting $\hat{\epsilon}_{n'} = \hat{v}_n+\hat{\epsilon}_n-\hat{v}_{n'}$ one arrives at an equation of the form :

$$F_n(.|\epsilon_0) = \exp(-\sum_{n'=1}^{N} c_{n'}.e^{-\hat{\epsilon}_n/q_{n'}}).c_n.e^{-\hat{\epsilon}_n/q_n}.(1/q_n)$$

with $c_{n'}=\exp[-\epsilon_0-(\hat{v}_n-\hat{v}_{n'}-\epsilon_0)/q_{n'}]$. One has to find a primitive function for this relationship in order to be able to find a closed-form expression of the choice probability π_n. However, such a primitive function is not known and we have to conclude that we are unable to find a useful formula for the choice probability π_n that corresponds to the logit model in the presence of realization uncertainty.

A3 The Derivation of the Generalized Logit Model

After summation over n, condition 1 becomes :

$$1/\pi_{n'} = \sum_{n=1}^{N} f_n(v_n,v_0,q_n)/f_{n'}(v_{n'},v_0,q_{n'}),$$

since $\sum_n \pi_n=1$. After inversion one arrives at equation (13).

Equation (13) implies that we should have :

$$\frac{\pi_n}{\pi_{n'}} = \frac{f_n(v_n,v_0,q_n)}{f_{n'}(v_{n'},v_0,q_{n'})}, \qquad n,n'=1,\ldots,N,$$

while, according to condition 2, we should have :

$$\frac{\pi_n}{\pi_{n'}} = \frac{e^{v_n}}{e^{v_{n'}}}, \qquad n,n'=1,\ldots,N.$$

These equations are compatible with each other if and only if (see Green [1964], ch.4) :

$$f_n(v_n, v_0, q_n) = g(v_0, q_n).e^{v_n}.$$

This gives rise to equation (17).

In order to analyze the effect of changes in the realization probabilities and the utility in the initial situation on the choice probabilities we use the partial derivative of the distribution function of the logit model (see the last equation of A2).

The choice probability π_n is defined in equation (3). After substitution of the first equation of A2 and, subsequently, of the relevant bounds $\hat{v}_n+\hat{\epsilon}_n-\hat{v}_{n'}$, we arrive at an equation of the form :

$$\pi_n = \int_{-\infty}^{\infty} F_n([\hat{v}_n+\hat{\epsilon}_n-\hat{v}_1-(1-q_1)\epsilon_0]/q_1, \ldots$$
$$\ldots [\hat{v}_n+\hat{\epsilon}_n-\hat{v}_N-(1-q_N)\epsilon_0]/q_N.|\epsilon_0).d\hat{\epsilon}_0$$

This equation is conditional upon the realization of ϵ_0.

In order to analyze the effect of a change in q_n we have to compute the partial derivative $\partial\pi_n/\partial q_n$:

$$\frac{\partial \pi_n}{\partial q_n} = \int_{-\infty}^{\infty} \frac{\partial F_n}{\partial q_n}.d\hat{\epsilon}_n.$$

If we can show that the sign of $\partial F_n/\partial q_n$ is unambiguously determined (for all values of ϵ_n and ϵ_0), we have also found the sign of $\partial\pi_n/\partial q_n$. We find :

$$\frac{\partial F_n}{\partial q_n} = \sum_{\substack{n'=1 \\ n'\neq n}}^{N} F_{nn'}.(v_n-v_0).q_n/q_1 - F_{nn}.(\epsilon_n+\epsilon_0)/q_n^2,$$

where $F_{nn'}$ denotes the first-order partial derivative of F_n with respect to its n'-th argument ($n' \neq n$). These partial derivatives may all be assumed positive. The problem is that the sign of the partial derivative $\partial F_n/\partial q_n$ is not determined, but depends on ϵ_n. This makes it impossible to impose general restictions on the sign of the effect of changes in q_n on π_n. It can be verified that this remains true when the special case of the logit-formulation is considered.

Now consider the effects of a change in v_0. We find in an analogous way :

$$\frac{\partial F_n}{\partial v_0} = \sum_{n'=1}^{N} F_{nn'} \cdot (q_{n'} - q_n)/q_{n'} .$$

Although in general the sign of this partial derivative is also undetermined, the result for the alternatives with the highest and lowest realization probabilities is unambiguous. It may be concluded therefore that an increase in v_0 decreases the probability that the alternative with the highest realization probability will be chosen and increases the probability that the one with the lowest realization probability will be chosen.

A4 Inconsistency with Expected Utility Maximization

In this section we will show that the generalized logit of equation (18) violates a necessary consequence of expected utility maximization and is therefore inconsistent with such behaviour. The necessary consequence of utility maximization we have in mind is known as the symmetry condition (see Smith [1984] for a discussion). This condition states :

$$\frac{\partial \pi_n}{\partial v_{n'}} = \frac{\partial \pi_{n'}}{\partial v_n}, \qquad n,n'=1,\ldots,N.$$

It can easily be derived from equation (3) when it is realized that (under very general circumstances) the second-order cross-partial derivatives of a real function are symmetric.

In the present context we are concerned with a sitiuation of uncertainty and the relevant utilities are therefore the expected utilities $\hat{v}_n$.

In order to check whether or not this condition is satisfied we rewrite the generalized logit model of equation (17) in terms of the expected utilities $\hat{v}_n$ on which the household is assumed to base its decisions. Inserting $v_n = [(1-q_n).v_0 - \hat{v}_n]/q_n$ (n=2,...,N), we find :

$$\pi_n = \frac{g(v_0,q_n).e^{(v_0-\hat{v}_n)/q_n}}{\sum_{n'=1}^{N} g(v_0,q_{n'}).e^{(v_0-\hat{v}_{n'})/q_{n'}}}, \qquad n=1,\ldots,N.$$

We compute the partial derivatives with respect to $v_{n''}$ $(n''\neq n)$ and find :

$$\frac{\partial \pi_n}{\partial v_{n''}} = \pi_n.\pi_{n''}/q_{n''} \qquad n,n''=1,\ldots,N.$$

This is not symmetric because of the occurence of the probability $q_{n''}$.

It follows that the model of equation (18) is inconsistent with maximization of expected utilities. This is somewhat surprising since conditions 2 and 3 have been derived explicitly as consequences of such behaviour. We must therefore conclude that condition 1 (the only one left), a variant of the independence of irrelevant alternatives assumption, is incompatible with expected utility maximization. This implies that a model which is as easy to use as the logit model will not be available for situations of utility maximization under uncertainty.

It also sheds some light on the difficulties we experienced when we tried to derive a model explicitly on the basis of such behaviour.

B The Data

The data that have been used are those of the Dutch Housing Needs Survey (WoningBehoefte Onderzoek, WBO) of 1981. This is a 1% sample of the total Dutch population consisting of more than 60,000 respondents (households). We selected those that had moved to their present house during the four years before 1981 and those that indicated to be willing to move to another type of dwelling within a year. We selected the households that were occupying a rented dwelling and were willing to move to a rented dwelling. Since the rented part of the Dutch housing market is regulated this seems to be the appropriate strategy for the detection of effects of realization uncertainty. Moreover, we selected only those households that intended to remain within the boundaries of the same region, since it may be expected that migrants behave differently (e.g. by moving to a cheap and readily dwelling type that give them a good opportunity to look for a better dwelling in their new environment).

The households that were willing to move to another dwelling were asked to indicate the preferred number of rooms of the new dwelling, whether it should be a single-family house or an apartment and what price they would be willing to pay for such a dwelling. These characteristics were also known from the dwellings that were presently occupied. They were used to construct a classification of the dwelling types that referred to both the dwellings presently occupied and the dwellings to which the households wished to move. 16 types of dwellings have been distinguished.

The disequilibrium situation was investigated by a comparison of the number of intended moves to the various dwelling types that had been distinguished by the averaged number of yearly moves towards such dwellings in the preceding for years. On the basis of the assumption that the Dutch housing market is more or less in a stationary state we can use the ratio between these two variables as an indication of the realization probability.

The additional information needed consists of the income of the household and the number of persons it contains. Information about these variables was also contained in the Housing Needs Survey.

References

Green, H.A.J. [1964] Aggregation in Economic Analysis, Princeton University Press, Princeton(N.J.).

Maddala, G.S. [1982] Limited Dependent and Qualitative Variables in Econometrics, Cambridge University Press, Cambridge.

McFadden, D. [1978] Modelling the choice of the residential location, in : A.Karlqvist et al.(eds) Spatial Interaction Theory and Planning Models, Amsterdam, North Holland, 76-96.

McFadden, D. [1981] Econometric models of probabilisatic choice, ch. 5 in : C.F.Manski and D. McFadden, Structural Analysis of Discrete Data with Econometric Applications, MIT-Press, Cambridge (Ma).

Rouwendal, J. [1988] Discrete Choice Models and Housing Market Analysis, Ph.D. thesis, Free University, Amsterdam.

Ridder, G. [198] GRMAX, A general maximum-likelihood program, user manual.

Smith, T.E. [1984] A choice probability characterization of generalized extreme value models, Applied Mathematics and Computation, 14, 35-62.

1988-1	H. Visser	Austrian thinking on international economics
1988-2	A.H.Q.M. Merkies T. van der Meer	Theoretical foundations for the 3-C model
1988-3	H.J. Bierens J. Hartog	Nonlinear regression with discrete explanatory variables, with an application to the earnings function
1988-4	N.M. van Dijk	On Jackson's product form with 'jump-over' blocking
1988-5	N.M. van Dijk M. Rumsewicz	Networks of queues with service anticipating routing
1988-6	H. Linneman C.P. van Beers	Commodity Composition of Trade in Manufactures and South-South Trade Potential
1988-7	N.M. van Dijk	A LCFS finite buffer model with batch input and non-exponential sevices
1988-8	J.C.W. van Ommeren	Simple approximations for the batch-arrival $M^X/G/1$ queue
1988-9	H.C. Tijms	Algorithms and approximations for bath-arrival queues
1988-10	J.P. de Groot H. Clemens	Export Agriculture and Labour Market in Nicaragua
1988-11	H. Verbruggen J. Wuijts	Patterns of South-South trade in manufactures
1988-12	H.C. Tijms J.C.W. van Ommeren	Asymptotic analysis for buffer behaviour in communication systems
1988-13	N.M. van Dijk E. Smeitink	A non-exponential queueing system with batch servicing
1988-14	J. Rouwendal	Existence and uniqueness of stochastic price equilibria in heterogeneous markets
1988-15	H. Verbruggen	GSTP, the structure of protection and South-South trade in manufactures
1988-16	Mevr. H. Weijland Mevr. R. Herweijer J. de Groot	Female participation in agriculture in the Dominican Republic
1988-17	N.M. van Dijk	Product Forms for Random Access Schemes
1988-18	A.H.Q.M. Merkies I.J. Steyn	Adaptive Forecasting with Hyperfilters
1988-19	J. Rouwendal	Specification and Estimation of a Logit Model for Housing Choice in the Netherlands
1988-20	J.C.W. van Ommeren R.D. Nobel	An elementary proof of a basic result for the GI/G/1 queue
1988-21	H. Kool	A Note on Consistent Estimation of Heterostedastic and Autocorrelated Covariance Matrices
1988-22	C.P.J. Burger	Risk Aversion and the Family Farm
1988-23	N. van Dijk I.F. Akyildiz	Networks with mixed processor sharing parallel queues and common pools
1988-24	D.J.F. Kamann P. Nijkamp	Technogenesis: Incubation and Diffusion
1988-25	P. Nijkamp L. van Wissen A. Rima	A Household Life Cycle Model For the Housing Market
1988-26	P. Nijkamp M. Sonis	Qualitative Impact Analysis For Dynamic Spatial Systems
1988-27	R. Janssen P. Nijkamp	Interactive Multicriteria Decision Support For Environmental Management
1988-28	J. Rouwendal	Stochastic Market Equilibria With Rationing and Limited Price Flexibility
1988-29	P. Nijkamp A. Reggiani	Theory of Chaos in a Space-Time Perspective
1988-30	P. Nijkamp J. Poot J. Rouwendal	R & D Policy in Space and Time
1988-31	P. Nijkamp F. Soeteman	Dynamics in Land Use Patterns Socio-Economic and Environmental Aspects of the Second Agricultural Land Use Revolution
1988-32	J. Rouwendal P. Nijkamp	Endogenous Production of R & D and Stable Economic Development
1988-33	J.A. Hartog E. Hinloopen P. Nijkamp	Multicriteria Methoden: Een gevoeligheidsanalyse aan de hand van de vestigingsplaatsproblematiek van kerncentrales
1988-34	R. van der Mark P. Nijkamp	The Development Potential of High Tech Forms in Backward Areas - A Case study for the Northern Part of The Netherlands
1988-35	E.R.K. Spoor J.W.B. Vermeulen	Principes en gebruik van Envisage
1988-36	C. Gorter P. Nijkamp P. Rietveld	The Duration of Unemployment: Stocks and Flows on Regional Labour Markets in the Netherlands
1988-37	M. Hofkes	Parametrization of simplicial algorithms with an application to an empirical general equilibrium model
1988-38	J. van Daal A.H.Q.M. Merkies	A Note on the Quadratic Expenditure Model