

Ronning, Gerd

Working Paper

Estimation of discrete choice models including socio-economic explanatory variables

Diskussionsbeiträge - Serie II, No. 42

Provided in Cooperation with:

Department of Economics, University of Konstanz

Suggested Citation: Ronning, Gerd (1987) : Estimation of discrete choice models including socio-economic explanatory variables, Diskussionsbeiträge - Serie II, No. 42, Universität Konstanz, Sonderforschungsbereich 178 - Internationalisierung der Wirtschaft, Konstanz

This Version is available at:

<https://hdl.handle.net/10419/101753>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Sonderforschungsbereich 178
„Internationalisierung der Wirtschaft“

Diskussionsbeiträge



Juristische
Fakultät

Fakultät für Wirtschafts-
wissenschaften und Statistik

Gerd Ronning

**Estimation of
Discrete Choice Models Including
Socio-Economic Explanatory Variables**

ESTIMATION OF DISCRETE CHOICE MODELS INCLUDING
SOCIO-ECONOMIC EXPLANATORY VARIABLES

Gerd Ronning

Serie II - Nr. 42

Ag 14 18 / 88
Weltwirtschaft
Kiel → *JK*

Oktober 1987

TABLE OF CONTENTS

1. INTRODUCTION	1
2. PURE DISCRETE CHOICE MODELS	3
2.1 Terminology	3
2.2 First and second derivatives of loglikelihood	4
2.3 The case of two or more explanatory variables	5
3. QUALITATIVE RESPONSE MODELS	7
3.1 Unidimensional stimulus	7
3.2 First and second derivatives of loglikelihood	8
3.3 Two or more explanatory variables	11
3.4 Additional remarks	12
4. MIXED SPECIFICATIONS	13
4.1 Introductory remarks	13
4.2 The simplest case	13
4.3 A compact formulation	18
4.4 The case $x_{ij} = x_j$ for all i ($m = 1$)	21
4.5 The general model for the mixed specification	23
4.6 The case $x_{ij} = x_j$, $z_{i1} \equiv 1$ ($m \geq 1$)	24
4.7 The case $x_{ij} = x_j$ and no constant term	28
REFERENCES	30

A B S T R A C T

Discrete Choice models and qualitative response models are intimately related. Usually both are based on a logistic specification. In many applications the two approaches have been combined leading to a 'discrete choice model including socio-economic explanatory variables'. The paper describes maximum likelihood estimation for all three approaches. Special attention is given to the case of missing variation (over individuals) of category-dependent explanatory variables.

ESTIMATION OF DISCRETE CHOICE MODELS INCLUDING
SOCIO-ECONOMIC EXPLANATORY VARIABLES

GERD RONNING , KONSTANZ *

1. INTRODUCTION

Econometric analysis of models with qualitative dependent variables is divided into two parts: discrete choice analysis and qualitative response models. The latter has been applied already for a long time by biometricians and psychometricians. The analysis of discrete (or qualitative) choice, however, has been mostly performed by economists and, in particular, by scientist involved in travel research. Broadly speaking, both approaches which are, of course, closely related, can be characterized as follows: Qualitative response models consider the effect of a certain stimulus (which may be multidimensional) on the probability of being thrown into a certain category of the qualitative dependent variable. The standard example from biometrics is the treatment of animals by different doses of a certain substance. The researcher then is interested in the probability that animals die, are greatly affected or are not affected at all. In economic context such analysis typically involves the use of socio-economic variables. Take the case of labour economics: The stimulus (family) income is analysed with respect to the employment state of married women (employed, non-employed, not on the labour market). Both example exhibit a causal chain of stimuli to a certain state of the dependent variables.

* Research is related to project A2, Sonderforschungsbereich 178 'Internationalisierung der Wirtschaft', Universität Konstanz. Some of the conclusions in this paper are based on computations done by Angelika Eymann and Martin Kukuk.

Discrete choice models, on the other hand, stress the fact that characteristics of the dependent variable may be responsible for the choice of a certain category. For example, in travel demand the choice of a certain travel mode naturally depends on the (relative) price of the different modes. So discrete choice models focus on category-dependent variables, whereas qualitative response models primarily consider individual characteristics and other stimuli.

In most empirical applications the researcher has to combine both explanatory components which here will be termed a discrete choice model including socio-economic explanatory variables. As an illustration, consider the case of a (European) tourist who wants to travel abroad. He has to choose a certain country. His decision will depend both on characteristics of the different countries and on his income, family size etc. Other examples can be found, e.g., in Domencich and McFadden (1975) and Lerman and Ben Akiva (1985).

Estimation of such models by the method of maximum likelihood is straightforward if the logistic specification is used. However, the usual formulation which transforms the socio-economic explanatory variables into variables of the discrete choice type (see, for example, Maddala 1983, p. 74) hides some problems in special situations. For example, if the purchase power of a certain country (or its exchange rate) is considered as an (category-dependent) explanatory variable in tourist's choice of country, a full parametrization is no longer possible since this variable does not vary across individuals which is usually assumed in discrete choice models.

In this research report we therefore first describe the estimation of (pure) discrete choice models and (pure) qualitative response models. We then give the formulae for the case of the mixed specification. Both in the pure discrete

choice model and in the mixed case we pay special attention to the case of category-dependent variables which do not vary across individuals.

The paper is organized as follows: Section 2 describes ML estimation of the pure discrete choice model, section 3 considers stimulus-response models and section 4 treats the mixed case.

2. PURE DISCRETE CHOICE MODELS

2.1 Terminology

Let p_{ij} be the probability that individual i chooses category j where $j = 1, \dots, r$. Note that we stick to the case of identical alternatives for all individuals.¹⁾ Let x_{ij} denote the values of the category-dependent explanatory variable with possible variation over individuals. The logistic specification²⁾ of the choice probability is then given by

$$p_{ij} = \frac{\exp(x_{ij} \beta)}{\sum_{k=1}^r \exp(x_{ik} \beta)}, \quad j = 1, \dots, r \quad (2.1)$$

for all i . All choice probabilities are characterized by the single parameter β which has to be estimated.

1) Domencich and McFadden (1975, p. 51) call this the case of "ranked" alternatives.

2) This specification can be derived from utility maximization of a utility function with extreme value distributed random component. See, for example, Amemiya (1985).

2.2 First and Second Derivatives

Under multinomial sampling scheme the loglikelihood function is given by

$$\begin{aligned}
 L &= \sum_i \sum_j \delta_{ij} \log p_{ij} \\
 &= \sum_i \sum_j \delta_{ij} (\beta x_{ij} - \log (\sum_k \exp(\beta x_{ik})))
 \end{aligned} \tag{2.2}$$

where $\delta_{ij} = 1$ if individual i chooses category j and zero otherwise. The first order condition is given by

$$\begin{aligned}
 \frac{\delta L}{\delta \beta} &= \sum_i \sum_j \delta_{ij} \left\{ x_{ij} - \frac{\sum_k x_{ik} \exp(\beta x_{ik})}{\sum_{k'} \exp(\beta x_{ik'})} \right\} \\
 &= \sum_i \sum_j \delta_{ij} x_{ij} - \sum_i \left(\sum_j \delta_{ij} \right) \sum_k p_{ik} x_{ik} \\
 &= \sum_i \sum_j (\delta_{ij} - p_{ij}) x_{ij} \stackrel{!}{=} 0
 \end{aligned} \tag{2.3}$$

since $\sum_j \delta_{ij} = 1$.

The second derivative is given by

$$\begin{aligned}
 \frac{\delta^2 L}{(\delta \beta)^2} &= - \sum_i \frac{\sum_j x_{ij}^2 \exp(\beta x_{ij}) \sum_k \exp(\beta x_{ik}) - \sum_k x_{ik} \exp(\beta x_{ik}) \sum_j x_{ij} \exp(\beta x_{ij})}{\left(\sum_k \exp(\beta x_{ik}) \right)^2} \\
 &= - \sum_i \left(\sum_j p_{ij} x_{ij}^2 - \left(\sum_k p_{ik} x_{ik} \right)^2 \right) \\
 &< 0
 \end{aligned} \tag{2.4}$$

because of the Cauchy inequality.

Therefore this loglikelihood function is globally strictly concave.

Consider now the special case $x_{ij} = x_j$ for all i . Note that this implies $p_{ij} = p_j$. Then the first order condition (2.3) reduces to

$$\sum_j x_j \left(\sum_i \delta_{ij} - \sum_i p_{ij} \right) = \sum_j n_j x_j (1 - p_j) = 0 \quad (2.5)$$

where n_j is the number of individuals choosing category j .

Since there exists no β for which all p_j equal one simultaneously, a non-degenerate solution of β still exists.

2.3 The case of two or more explanatory variables

If we have more than one explanatory variable x , we employ the following notation:

$$\beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_m \end{pmatrix} \quad x_{ij} = (x_{ij1}, \dots, x_{ijm})$$

where m is the number of explanatory variables. Then the choice probabilities are formally identical to (2.1). Maximization of the loglikelihood function with respect to the vector β gives

$$\frac{\delta}{\delta \beta} L = \sum_i \sum_j (\delta_{ij} - p_{ij}) x'_{ij} \stackrel{!}{=} 0 \quad (2.6)$$

which is formally equivalent to (2.3). However, (2.6) is a set of m equations. The matrix of second derivatives (Hessian matrix) is given by

$$\frac{\delta}{\delta \beta \delta \beta^T} L = - \sum_i \sum_j x'_{ij} \frac{\delta}{\delta \beta^T} p_{ij} \quad (2.7)$$

Since

$$\begin{aligned}
 \frac{\delta}{\delta \beta^T} \frac{\exp(x_{ij} \beta)}{\sum_k \exp(x_{ik} \beta)} &= \\
 &= \frac{\exp(x_{ij} \beta) \left(\sum_k \exp(x_{ik} \beta) \right) x_{ij} - \exp(x_{ij} \beta) \sum_k \exp(x_{ik} \beta) x_{ik}}{\left(\sum_k \exp(x_{ik} \beta) \right)^2} \\
 &= p_{ij} x_{ij} - p_{ij} \sum_k p_{ik} x_{ik} \quad (2.8)
 \end{aligned}$$

we obtain for the Hessian matrix

$$\begin{aligned}
 \frac{\delta^2}{\delta \beta \delta \beta^T} L &= - \sum_i \left(\sum_j p_{ij} x'_{ij} x_{ij} - \left(\sum_k p_{ik} x'_{ik} \right) \left(\sum_l p_{il} x_{il} \right) \right) \\
 &= - \sum_i \left(\sum_j p_{ij} x'_{ij} x_{ij} - \bar{x}'_i \bar{x}_i \right) \quad (2.9)
 \end{aligned}$$

where $\bar{x}_i = \sum_k p_{ik} x_{ik}$. Negative definiteness of this matrix has been proved by McFadden (1973). Note that a (trivial) necessary condition is $n > m$ where n is the number of observations.

The case $x_{ij} = x_j$ for all i has the same comments as in section 2.2.

3. QUALITATIVE RESPONSE MODELS

3.1 Unidimensional stimulus

Let z_i be the stimulus observed by individual i . Then the simplest version of a qualitative response model specifies the probabilities p_{ij} as

$$p_{ij} = \frac{\exp(z_i \alpha_j)}{\sum_k \exp(z_i \alpha_k)} \quad , j = 1, \dots, r \quad (3.1)$$

Note that contrary to (2.1) there are r different parameters α_j now.

Note further that

$$z_i \equiv 1 \quad \text{for all } i \quad (3.2a)$$

is included as a special case. We then write

$$p_{ij} = p_j = \frac{\exp(\alpha_j)}{\sum_k \exp(\alpha_k)} \quad (3.2b)$$

However, the parametrization is unique only if some constraint is placed on to the α_j . Henceforth we shall use the normalization

$$\alpha_r = 0 \quad (3.3)$$

so that (3.1) becomes

$$p_{ij} = \begin{cases} \frac{\exp(z_i \alpha_j)}{1 + \sum_{k \neq r} \exp(z_i \alpha_k)} & , j \neq r \\ \frac{1}{1 + \sum_{k \neq r} \exp(z_i \alpha_k)} & , j = r \end{cases} \quad (3.4)$$

Therefore only $r-1$ Parameters α_j , $j=1, \dots, r-1$ have to be estimated.

3.2 First and second derivatives of loglikelihood

Under multinomial sampling scheme the loglikelihood function is again given by (2.2) where p_{ij} is now defined in (3.1). From the normalization (3.3) and from $p_{ir} = 1 - \sum_{k \neq r} p_{ik}$ we then obtain for the likelihood function

$$\begin{aligned}
 L &= \prod_i \left[\prod_{j \neq r} \left(\frac{1}{1 + \sum_{k \neq r} \exp(z_i \alpha_k)} \right)^{\delta_{ij}} \prod_{j \neq r} \exp(z_i \alpha_j)^{\delta_{ij}} \right] \\
 &= \prod_i \left[\frac{1}{1 + \sum_{k \neq r} \exp(z_i \alpha_k)} \prod_{j \neq r} \exp(z_i \alpha_j)^{\delta_{ij}} \right] \quad (3.5)
 \end{aligned}$$

The loglikelihood function therefore is given by

$$L = \sum_i \left[\sum_{j \neq r} \delta_{ij} z_i \alpha_j - \log \left(1 + \sum_{k \neq r} \exp(z_i \alpha_k) \right) \right] \quad (3.6)$$

The first order conditions are

$$\begin{aligned}
 \frac{\delta L}{\delta \alpha_j} &= \sum_i \left[\delta_{ij} z_i - \frac{z_i \exp(z_i \alpha_j)}{1 + \sum_{k \neq r} \exp(z_i \alpha_k)} \right] \\
 &= \sum_i (\delta_{ij} - p_{ij}) z_i \stackrel{!}{=} 0, \quad j=1, \dots, r-1, \quad (3.7)
 \end{aligned}$$

and for the matrix of second derivatives (Hessian matrix) we obtain

$$\begin{aligned}
\frac{\delta^2 L}{(\delta \alpha_j)^2} &= - \sum_i \frac{z_i^2 \exp(z_i \alpha_j) (1 + \sum_{k \neq r} \exp(z_i \alpha_k)) - z_i^2 (\exp(z_i \alpha_j))^2}{(1 + \sum_{k \neq r} \exp(z_i \alpha_k))^2} \\
&= - \sum_i \left[(p_i - p_i^2) z_i^2 \right] \tag{3.8a} \\
&= - \sum_i p_i (1 - p_i) z_i^2 < 0
\end{aligned}$$

and

$$\begin{aligned}
\frac{\delta L}{\delta \alpha_j \delta \alpha_k} &= + \frac{z_i^2 \exp(z_i \alpha_j) \exp(z_i \alpha_k)}{(1 + \sum_{k \neq r} \exp(z_i \alpha_k))^2} \\
&= p_{ij} p_{ik} z_i^2 \quad j \neq k \tag{3.8b}
\end{aligned}$$

The Hessian matrix H therefore has the following structure:³⁾

$$\frac{\delta L}{\delta \alpha \delta \alpha'} = - \sum_i \begin{bmatrix} p_{1i}(1-p_{1i}) & -p_{1i} p_{2i} & -p_{r-1,i} p_{1i} \\ -p_{2i} p_{1i} & p_{2i}(1-p_{2i}) & \\ \vdots & \vdots & \ddots \\ -p_{r-1,i} p_{1i} & -p_{r-1,i} p_{2i} & p_{r-1,i}(1-p_{r-1,i}) \end{bmatrix} z_i^2 \tag{3.9}$$

The matrix has multinomial covariance structure and therefore is positive definite for each i (see, e.g., Ronning 1983). This seems to imply that we could obtain a unique ML estimate for all α_j even from a single observation. However, inspection of the first order conditions (3.7) shows that no meaningful estimates can exist if $n < r-1$ where n is the number of observations.

3) We define the $(r-1)$ dimensional vector α by $\alpha = (\alpha_1, \dots, \alpha_{r-1})'$.
See also page 12.

Take the example of $n = 2$ observations and assume that both category 1 and 2 are taken once, i.e. $\delta_{11} = 1, \delta_{22} = 1, \delta_{12} = 0, \delta_{21} = 0$. Furthermore assume $r - 1 = 3$. Then from (3.7) we obtain

$$\begin{aligned} \text{for } j = 1 & \quad (1 - p_{11}) z_1 - p_{21} z_2 = 0 \\ \text{for } j = 2 & \quad - p_{12} z_1 + (1 - p_{22}) z_2 = 0 \\ \text{for } j = 3 & \quad - p_{13} z_1 - p_{23} z_2 = 0 \end{aligned}$$

Solution for z_1/z_2 for all three equations results in

$$\frac{p_{21}}{1 - p_{11}} = \frac{1 - p_{22}}{p_{12}} = -\frac{p_{23}}{p_{13}}$$

which shows that no meaningful solution exists.

For the case of $z_1 \equiv 1$, all i (see (3.2)) the first order condition (3.7) reduces to

$$n_j - np_j = 0 \tag{3.10}$$

from which we get $\hat{p}_j = n_j/n$, where n_j is the number of observations from category j . Note that for constant z_i estimation is already possible for $n < r - 1$; on the other hand, even for large n an estimated probability might be zero.

3.3 Two or more explanatory variables

Let z_i be an q -dimensional (row) vector of explanatory variables and define α_j by and z_i by

$$\alpha_j = \begin{pmatrix} \alpha_{1j} \\ \vdots \\ \alpha_{qj} \end{pmatrix}, \quad z_i = (z_{i1}, z_{i2}, \dots, z_{iq})$$

Then the probabilities are formally identical to those in (3.1). Under normalization

$$\alpha_r = 0$$

we obtain the formulae (3.4). Now $(r-1)q$ parameters from the vectors $\alpha_1, \dots, \alpha_{r-1}$ have to be estimated.

The first order conditions are obtained from the likelihood functions (3.6) with z_i and α_j as vectors. Inspection of result (3.7) shows that we obtain again

$$\frac{\delta L}{\delta \alpha_j} = \sum_i (\delta_{ij} - p_{ij}) z_i \stackrel{!}{=} 0, \quad j = 1, \dots, r-1 \quad (3.11)$$

The matrix of second derivatives (Hessian matrix) is given by

$$\frac{\delta^2 L}{\delta \alpha_j \delta \alpha_j} = - \sum_i p_{ij} (1 - p_{ij}) z_i^! z_i$$

and

$$\frac{\delta^2 L}{\delta \alpha_j \delta \alpha_k} = \sum_i p_{ij} p_{ik} z_i^! z_i \quad j \neq k \quad (3.12)$$

Therefore the Hessian matrix has the structure

$$\frac{\delta^2 L}{\delta \alpha \delta \alpha'} = - \sum_i \left(\begin{array}{cccc} p_{i1}(1-p_{i1}) & -p_{i1} p_{i2} & \cdots & -p_{i1} p_{i,r-1} \\ -p_{i2} p_{i1} & p_{i2}(1-p_{i2}) & \cdots & -p_{i2} p_{i,r-1} \\ \vdots & \vdots & \ddots & \vdots \\ -p_{i,r-1} p_{i1} & -p_{i,r-1} p_{i2} & \cdots & p_{i,r-1}(1-p_{i,r-1}) \end{array} \right) \otimes z_i' z_i \quad (3.13)$$

where \otimes denotes the Kronecker product and $\alpha' = (\alpha_1', \alpha_2', \dots, \alpha_{r-1}')$ is the vector of all $q(r-1)$ parameters. This matrix is negative definite whenever at least q vectors z_i are linearly independent. See, e.g. Dhrymes (1978).

3.4 Additional remarks

We have already discussed the case $z_i \equiv 1$ in (3.2) and (3.20). For the case of a q -dimensional vector z_i the missing variation over individuals would cause trouble. Let

$$z_i = a \quad \text{for all } i$$

where a is some arbitrary chosen vector. Then the diagonal blocks in the Hessian matrix (3.13) are no longer of full rank (see (3.12)) and therefore (3.13) cannot obtain full rank. One could also note that $p_{ij} = p_j$, all i , so that (3.13) has at most rank $r-1$.

4. MIXED SPECIFICATIONS

4.1 Introductory remarks

In this section we consider models which combine category-dependent variables x_{ij} with variables z_i of the stimulus type. Again we start with the simplest specification and then proceed to more general models. In particular we discuss problems which arise from missing variation of the explanatory variables. For convenience we will abbreviate variables in the vector x_{ij} as CD (= category dependent) variables and those from the vector z_i as ST (= stimulus) variables.

4.2 The simplest case

The simplest case of a combination of CD variables and ST variables is as follows:

$$p_{ij} = \frac{\exp(x_{ij}\beta + z_i \alpha_j)}{\sum_{k=1}^r \exp(x_{ik}\beta + z_i \alpha_k)} \quad j = 1, \dots, r \quad (4.1)$$

where both x_{ij} and z_i are scalars. Again we need (see section 3) a normalization for the α_j . Using (3.3) we obtain

$$p_{ij} = \frac{\exp(x_{ij}\beta + z_i \alpha_j)}{\exp(x_{ir}\beta) + \sum_{k \neq r} \exp(x_{ik}\beta + z_i \alpha_k)} \quad , j \neq r$$

$$p_{ir} = \frac{\exp(x_{ir}\beta)}{\exp(x_{ir}\beta) + \sum_{k \neq r} \exp(x_{ik}\beta + z_i \alpha_k)} \quad (4.2)$$

This model has $r - 1 + 1 = r$ parameters which have to be estimated.

The first order conditions are obtained from the loglikelihood function under multinomial sampling scheme (see (2.2)) with p_{ij} from (4.2). This gives (compare (3.6))

$$L = \sum_i \left[\sum_{j \neq r} \delta_{ij} (x_{ij}^\beta + z_i \alpha_j) + \delta_{ir} x_{ir}^\beta - \log \left(\exp(x_{ir}^\beta) + \sum_{k \neq r} \exp(x_{ik}^\beta + z_i \alpha_k) \right) \right] \quad (4.3)$$

Differentiating with respect to β and α_j we obtain the first order conditions:

$$\begin{aligned} \frac{\delta L}{\delta \beta} &= \sum_i \left[\sum_{j \neq r} \delta_{ij} x_{ij} + \delta_{ir} x_{ir} - \frac{\exp(x_{ir}^\beta) x_{ir} + \sum_{k \neq r} \exp(x_{ik}^\beta + z_i \alpha_k) x_{ik}}{\exp(x_{ir}^\beta) + \sum_{k \neq r} \exp(x_{ik}^\beta + z_i \alpha_k)} \right] \\ &= \sum_i \sum_j (\delta_{ij} - p_{ij}) x_{ij} \end{aligned} \quad (4.4a)$$

$$\begin{aligned} \frac{\delta L}{\delta \alpha_j} &= \sum_i \left[\delta_{ij} z_i - \frac{\exp(x_{ij}^\beta + z_i \alpha_j) z_i}{\exp(x_{ir}^\beta) + \sum_{k \neq r} \exp(x_{ik}^\beta + z_i \alpha_k)} \right] \\ &= \sum_i (\delta_{ij} - p_{ij}) z_i, \quad j = 1, \dots, r-1. \end{aligned} \quad (4.4b)$$

Note that these conditions are equivalent to the conditions obtained from separate specifications in sections 2 and 3, respectively.

The matrix of second derivatives now has the following elements:

$$\frac{\delta^2 L}{\delta \beta^2} = - \sum_i \left[\frac{(\exp(x_{ir}^\beta) x_{ir}^2 + \sum_{j \neq r} \exp(x_{ij}^\beta + z_i \alpha_j) x_{ij}^2) (\exp(x_{ir}^\beta) + \sum_{k \neq r} \exp(x_{ik}^\beta + z_i \alpha_k))}{[\exp(x_{ir}^\beta) + \sum_{k \neq r} \exp(x_{ik}^\beta + z_i \alpha_k)]^2} \right]$$

$$= - \frac{(\exp(x_{ir}\beta)x_{ir} + \sum_{k \neq r} \exp(x_{ik}\beta + z_i\alpha_k) x_{ik})^2}{(\exp(x_{ir}\beta) + \sum_{k \neq r} \exp(x_{ik}\beta + z_i\alpha_k))^2} \Bigg]$$

$$= - \sum_i \left[\sum_j p_{ij} x_{ij}^2 - \sum_j p_{ij} x_{ij} \right]^2 \quad (4.5a)$$

$$\frac{\delta^2}{\delta\beta\delta\alpha_j} = - \sum_i \left[\frac{\exp(x_{ij}\beta + z_i\alpha_j) z_i x_{ij} [\exp(x_{ir}\beta) + \sum_{k \neq r} \exp(x_{ik}\beta + z_i\alpha_k)]}{[\exp(x_{ir}\beta) + \sum_{k \neq r} \exp(x_{ik}\beta + z_i\alpha_k)]^2} \right.$$

$$\left. - \frac{[\exp(x_{ir}\beta)x_{ir} + \sum_{k \neq r} \exp(x_{ik}\beta + z_i\alpha_k)x_{ik}] \exp(x_{ij}\beta + z_i\alpha_j) z_i}{[\exp(x_{ir}\beta) + \sum_{k \neq r} \exp(x_{ik}\beta + z_i\alpha_k)]^2} \right]$$

$$= - \sum_i \left[p_{ij} x_{ij} z_i - \left(\sum_k p_{ik} x_{ik} \right) p_{ij} z_i \right]$$

$$= - \sum_i \left[x_{ij} - \left(\sum_k p_{ik} x_{ik} \right) \right] p_{ij} z_i \quad (4.5b)$$

$$\frac{\delta^2}{(\delta\alpha_j)^2} = - \sum_i \left[\frac{\exp(x_{ij}\beta + z_i\alpha_j) z_i^2 [\exp(x_{ir}\beta) + \sum_{k \neq r} \exp(x_{ik}\beta + z_i\alpha_k)]}{[\exp(x_{ir}\beta) + \sum_{k \neq r} \exp(x_{ik}\beta + z_i\alpha_k)]^2} \right.$$

$$\left. - \frac{[\exp(x_{ij}\beta + z_i\alpha_j) z_i]^2}{[\exp(x_{ir}\beta) + \sum_{k \neq r} \exp(x_{ik}\beta + z_i\alpha_k)]^2} \right]$$

$$= - \sum_i \left[p_{ij} z_i^2 - (p_{ij} z_i)^2 \right]$$

$$= - \sum_i p_{ij} (1-p_{ij}) z_i^2 \quad (4.5c)$$

$$\frac{\delta^2 L}{\delta \alpha_j \delta \alpha_k} = + \sum_i \frac{\exp(x_{ij}\beta + z_i \alpha_j) \exp(x_{ik}\beta + z_i \alpha_k) z_i^2}{[\exp(x_{ir}\beta) + \sum_{k \neq r} \exp(x_{ik}\beta + z_i \alpha_k)]^2}$$

$$= + \sum_i p_{ij} p_{ik} z_i^2, \quad j \neq k \quad (4.5d)$$

Setting $\theta' = (\beta, \alpha_1, \alpha_2, \dots, \alpha_{r-1})$ we obtain the following structure of the Hessian matrix:

$$\frac{\delta^2 L}{\delta \theta \delta \theta'} = - \sum_i$$

$$\left[\begin{array}{c} \sum_i p_{ij} x_{ij}^2 - \left(\sum_j p_{ij} x_{ij} \right)^2 \\ \left(x_{ij} - \sum_k p_{ik} x_{ik} \right) p_{i1} z_i \\ \left(x_{i2} - \sum_k p_{ik} x_{ik} \right) p_{i2} z_i \\ \vdots \\ \left(x_{i,r-1} - \sum_k p_{ik} x_{ik} \right) p_{i,r-1} z_i \end{array} \quad \left[\begin{array}{c} \left(x_{i1} - \sum_k p_{ik} x_{ik} \right) p_{i1} z_i \\ \left(x_{i2} - \sum_k p_{ik} x_{ik} \right) p_{i2} z_i \\ \vdots \\ \left(x_{i,r-1} - \sum_k p_{ik} x_{ik} \right) p_{i,r-1} z_i \end{array} \right] \left[\begin{array}{c} \left(x_{i,r-1} - \sum_k p_{ik} x_{ik} \right) p_{i,r-1} z_i \\ \left(x_{i,r-1} - \sum_k p_{ik} x_{ik} \right) p_{i,r-1} z_i \\ \left(x_{i,r-1} - \sum_k p_{ik} x_{ik} \right) p_{i,r-1} z_i \\ \vdots \\ \left(x_{i,r-1} - \sum_k p_{ik} x_{ik} \right) p_{i,r-1} z_i \end{array} \right]$$

(4.6)

4.3 A compact formulation

Inspection of the blockdiagonal matrices shows that these are identical with those from sections 2 and 3, respectively. A compact formulation (which will be particularly useful in the case of vectors z_i and x_{ij}) starts from the following definitions:

$$W_i = \begin{matrix} (r \times r) \\ \left[\begin{array}{cccccc} x_{i1} & z_i & 0 & & & 0 \\ x_{i2} & 0 & z_i & & & 0 \\ x_{i3} & 0 & 0 & \dots & & 0 \\ & \vdots & \vdots & \dots & \dots & \vdots \\ x_{i,r-1} & 0 & 0 & & & z_i \\ x_{ir} & 0 & 0 & \dots & & 0 \end{array} \right] \end{matrix} \quad (4.7)$$

$$\theta = \begin{matrix} (r \times 1) \\ \left(\begin{array}{c} \beta \\ \alpha_1 \\ \vdots \\ \alpha_{r-1} \end{array} \right) \end{matrix}, \quad w_{ij} = (x_{ij}, 0, \dots, z_i, \dots, 0) \quad (4.8)$$

↑
(element j+1)

Please note that $w_{ij}\theta = x_{ij}\beta + z_i\alpha_j$. We furthermore define

$$V_i = D(p_i) - p_i p_i^i \quad (4.9a)$$

where $p_i^i = (p_{i1}, p_{i2}, \dots, p_{i,r})$ and

$$D(p_i) = \left(\begin{array}{cccc} p_{i1} & & & \\ & p_{i2} & & \\ & & \dots & \\ & & & p_{i,r} \end{array} \right) \quad (4.9b)$$

Finally we need the vector

$$y_i = \begin{pmatrix} \delta_{i1} \\ \vdots \\ \delta_{ir} \end{pmatrix} \quad (4.10)$$

which contains exactly one element equal to one and all others equal to zero.

The choice probabilities of the mixed specification may now be written as follows:

$$p_{ij} = \frac{\exp(w_{ij}\theta)}{\sum_k \exp(w_{ik}\theta)} \quad , \quad j = 1, \dots, r \quad (4.11)$$

Note that in this formulation the parameters regarding the ST variables do not need a normalization.⁴⁾ We now make use of our results in section 2.3 where we considered the model

$$p_{ij} = \frac{\exp(x_{ij}\beta)}{\sum_k \exp(x_{ik}\beta)}$$

with x_{ij} and β as vectors. Application of (2.6) gives the first order conditions

$$\sum_i \sum_j (\delta_{ij} - p_{ij}) w_{ij} \stackrel{!}{=} 0 \quad (4.12a)$$

or

$$\sum_i W_i^! (y_i - p_i) = 0 \quad (4.12b)$$

If we write (4.12a) in detail, we obtain the results already derived in (4.4).

In order to derive the Hessian matrix for the mixed specification, we first write (2.9) in the following form:

4) In fact, normalization is accomplished by the structure of the matrix W_i . See (4.7) and (4.8).

$$\begin{aligned} & \sum_i \left(\sum_j p_{ij} x'_{ij} x_{ij} - \left(\sum_k p_{ik} x'_{ik} \right) \left(\sum_k p_{ik} x_{ik} \right) \right) \\ & = \sum_i X'_i V_i X_i \end{aligned} \quad (4.13)$$

where

$$X_i = \begin{pmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ir} \end{pmatrix} \quad (4.14)$$

(r × m)

and each x_{ij} is an m -dimensional row vector (m denotes the number of explanatory CD variables).

In order to see the equivalence of (4.13) and (2.9) we consider first

$$\begin{aligned} X'_i D(p_i) X_i &= (x'_{i1} \ x'_{i2} \ \dots \ x'_{ir}) \begin{pmatrix} p_{i1} & & \\ & \ddots & \\ & & p_{ir} \end{pmatrix} X_i \\ &= (p_{i1} x'_{i1} \ , \ p_{i2} x'_{i2} \ \dots \ p_{ir} x'_{ir}) \begin{pmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ir} \end{pmatrix} \\ &= \sum_j p_{ij} x'_{ij} x_{ij} \end{aligned}$$

and

$$\begin{aligned} X'_i p_i &= (x_{i1} \ x_{i2} \ \dots \ x_{ir}) \begin{pmatrix} p_{i1} \\ p_{i2} \\ \vdots \\ p_{ir} \end{pmatrix} \\ &= \sum_j p_{ij} x_{ij} \\ &= \bar{x}_i \end{aligned}$$

This shows the equivalence. We now use W_i from (4.7). Note that in general W_i will have $m + r - 1$ columns and r rows. In our simplest example $m = 1$ so that W_i is quadratic. Application of (4.13) then gives the compact form of (4.6):

$$\frac{\delta^2 L}{\delta \theta \delta \theta'} = - \sum_i W_i' V_i W_i \quad (4.15)$$

4.4 The case $x_{ij} = x_j$ for all i ($m = 1$)

In this subsection we consider again the case that the CD variables do not vary over individuals (x_{ij} the exchange rate faced by tourist i who chooses country j). In particular we want to show that estimation (identification) is impossible if at the same time the special ST variable

$$z_i \equiv 1 \quad \text{all } i$$

(see (3.2a)) is employed. We use the explicit results obtained in subsection 4.2. The choice probabilities are now

$$p_{ij} = \frac{\exp(x_j \beta + z_i \alpha_j)}{\sum_k \exp(x_k \beta + z_i \alpha_k)} \quad (4.16)$$

and the first order conditions (4.4a) reduce to

$$\sum_i \sum_j (\delta_{ij} - p_{ij}) x_j = 0 \quad (4.17)$$

whereas (4.4b) remains the same. However, if at the same time we additionally assume that $z_i \equiv 1$ then (4.4b) reduces to

$$\sum_i (\delta_{ij} - p_{ij}) = 0 \quad j = 1, \dots, r-1. \quad (4.18)$$

which is equivalent to (3.10) and already determines all probabilities

$p_{ij} = p_i = \exp(x_j\beta + \alpha_j) / \sum_k \exp(x_k\beta + \alpha_k)$. The defect can also be seen from inserting (4.18) into (4.17):

$$\sum_j x_j \underbrace{(\sum_i (\delta_{ij} - p_{ij}))}_{= 0} = 0 \quad (4.19)$$

irrespective of the value of x_j . (To put it differently, the equations of the first order conditions are not independent.)

We also show that the Hessian matrix cannot be regular. This time we use the compact formulation from subsection 4.3. For $x_{ij} = x_i$ and $z_i = 1$ we get

$$W_i = \begin{pmatrix} x_1 & 1 & 0 & 0 \\ x_2 & 0 & 1 & \\ & & \ddots & \\ x_{r-1} & 0 & 0 & 1 \\ x_r & 0 & 0 & 0 \end{pmatrix} = W \quad (4.20)$$

and $V_i = V$ since $p_{ij} = p_j$. Therefore

$$\frac{\delta^2 L}{\delta\theta\delta\theta'} = n W' V W \quad (4.21)$$

where n is the number of observations. Note that W and V are quadratic. Obviously W has rank m . However, the matrix V has only rank $r-1$ (see Ronning 1983). Therefore the Hessian matrix which is a product of matrices W and V , can have only rank $r-1$ (or less). Therefore the Hessian matrix is singular.

4.5 The general model for the mixed specification

In the most general formulation we have an m -dimensional vector x_{ij} of CD variables and an q -dimensional vector z_i of ST variables. The choice probabilities are given by

$$p_{ij} = \frac{\exp(x_{ij}\beta + z_i\alpha_j)}{\sum_k \exp(x_{ik}\beta + z_i\alpha_k)} \quad , j = 1, \dots, r \quad (4.22)$$

which is formally equivalent to (4.1). (Vectors x_{ij} , z_i , β and α_j have been defined in sections 2.3 and 3.2.) If we employ the compact formulation of section 4.3, (4.22) can also be written

$$p_{ij} = \frac{\exp(w_{ij}\theta)}{\sum_k \exp(w_{ik}\theta)}$$

(see (4.11)). First order conditions are then

$$\sum_i \sum_j (\delta_{ij} - p_{ij}) w_{ij} = 0$$

or

$$\sum_i W_i^! (y_i - p_i) = 0$$

(see (4.12)) and the Hessian matrix is given by

$$\frac{\delta^2 L}{\delta\theta\delta\theta^T} = - \sum_i W_i^! V_i W_i$$

(see (4.15)).

4.6 The case $x_{ij} = x_j$, $z_{i1} \equiv 1$ ($m \geq 1$)

We now consider again the case $x_{ij} = x_j$ where x_{ij} may be an m -dimensional vector. Additionally, we assume that a "constant term" has been specified together with other ST variables. This leads to

$$p_{ij} = \frac{\exp(x_j \beta + \alpha_{j1} + z_i \bar{\alpha}_j)}{\sum_k \exp(x_k \beta + \alpha_{k1} + z_i \bar{\alpha}_k)} \quad (4.23)$$

where α_{j1} is the first element from the vector α_j , z_i is an $q-1$ dimensional vector and $\bar{\alpha}_j$ contains the remaining $q-1$ elements of α_j . From (4.11a) we get the following first order conditions ($w_{ij} = x_{ij} \dots, 0 \dots (1, z_i) \dots 0$):

$$\sum_i \sum_j (\delta_{ij} - p_{ij}) x_j = 0$$

or

$$\sum_j x_j' \left(\sum_i (\delta_{ij} - p_{ij}) \right) = 0 \quad (4.24)$$

Note that if only some elements of x_{ij} do not depend on i (for all i), then these components could be written in this way.

The remaining equations are given by

$$\sum_i (\delta_{ij} - p_{ij}) = 0 \quad , j = 1, \dots, r-1 \quad (4.25)$$

which we already obtained for the simplest case in (4.18) and, for the components of z_i ,

$$\sum_i (\delta_{ij} - p_{ij}) z_i = 0 \quad , j = 1, \dots, r-1 \quad (4.26)$$

We see that the inclusion of other variables z_i does not alter the statement from section 4.4: The equations (4.25) imply that (4.24) always holds, that is these $r+m$ equations are not independent.

We now want to show that this leads to singularity of the Hessian matrix, but only at parameter values which satisfy the first order conditions. (Note that no longer we can write $p_{ij} = p_j$ since we have included a vector z_i .)

It seems useful to use a slightly different compact notation. Set

$$\begin{aligned} \bar{W}_i &= \left(\bar{W}_{i1} \quad , \quad \bar{W}_{i2} \right) \\ &= \left(\begin{array}{cccc|cc} x_1 & 1 & & & z_i & \\ x_2 & & 1 & & & z_i \\ & & & \ddots & & \ddots \\ x_{r-1} & & & & 1 & z_i \\ x_r & & & & 0 & 0 \end{array} \right) \end{aligned} \quad (4.27)$$

Note that $\bar{W}_{i1} = W_1$ is independent of i ! In fact we have just permuted the columns of the matrix W_i . The new parameter vector $\bar{\theta}$ is given by

$$\bar{\theta}' = (\beta', \alpha_1^{(1)} \quad \dots \quad \alpha_{r-1}^{(1)}, \alpha_1^{(2)} \quad \dots, \alpha_{r-1}^{(2)})$$

where $\alpha_j^{(1)}$ are scalars and $\alpha_j^{(2)}$ denotes $(q-1)$ dimensional vectors. From completeness we also indicate the dimensions of the matrices \bar{W}_1 and \bar{W}_{2i} :

$$\begin{array}{cc} \bar{W}_1 & \bar{W}_{2i} \\ [r \times (m + r - 1)] & [r \times (q - 1)(r - 1)] \end{array}$$

The first order condition (4.12b) now can be written as follows:

$$\sum_i (\bar{W}_1 \quad \bar{W}_{21})' (y_i - p_i) = 0$$

or

$$W_1' \sum_i (y_i - p_i) = 0 \tag{4.28}$$

$$\sum_i \bar{W}'_{2i} (y_i - p_i) = 0$$

The first line of (4.28) shows that the r rows of W_1 are not linearly independent if the first order conditions are satisfied by the parameter estimates of the vector θ . Since W_1 has more (not less) columns than rows (if $m \geq 1$), the maximal rank of W_1 is r . However, under (4.28) there is one linear dependence between rows so that

$$\text{rg}(W_1) = r - 1 \quad \text{if (4.28) holds} \tag{4.29}$$

Now consider the Hessian matrix (4.15). Using the permuted partitioned matrix \bar{W} , we obtain

$$\frac{\delta^2 L}{\delta \theta \delta \theta'} = - \sum_i \bar{W}'_i V_i \bar{W}_i$$

$$= - \left(\begin{array}{c|c} \bar{W}'_i (\sum V_i) \bar{W}_i & \bar{W}'_i \sum V_i \bar{W}_{2i} \\ \hline (\sum_i \bar{W}'_{2i} V_i) \bar{W}_1 & \sum_i \bar{W}'_{2i} V_i \bar{W}_{2i} \end{array} \right) \tag{4.30}$$

For the two blocks at the left side of the matrix we can write $A \bar{W}_1$ where

$$A = \begin{bmatrix} \bar{W}_1' \cdot \left(\sum_i V_i \right) \\ \left(\sum_i \bar{W}_{2i}' V_i \right) \bar{W}_1 \end{bmatrix} \quad (4.31)$$

$[m+q(r-1) \times r]$

$A \bar{W}_1$ has $(m+r-1)$ columns which should be independent. In general $\text{rg}(A \bar{W}_1) \leq \min \{ \text{rg}(A), \text{rg}(\bar{W}_1) \}$. Since for $m \geq 1$ and $q > 1$ (as assumed here) we have $m + q(r-1) \geq r$, the maximal rank of $A \bar{W}_1$ is r . However, under first order conditions (4.28) the rank condition (4.29) restricts the rank of $A \bar{W}_1$ to $r - 1$ linear independent columns.

In practice, this results in the following:

- (a) If only one CD Variable which does not depend on i is specified ($m = 1$) together with a constant term, then the Hessian matrix will become singular whenever the first order conditions are satisfied. In iterative procedures (such as Newton-Raphson), the minimal eigenvalue of the Hessian matrix will tend towards zero if the gradient tends towards zero.
- (b) If two or more CD variables which do not depend on i are specified ($m > 1$) (and combined with a constant term), then our analysis indicates that always
- $$m + q(r-1) - r = m + (q-1)r - q$$
- eigenvalues of the Hessian matrix should be equal to zero. Under first order conditions one additional eigenvalue will tend towards zero. Therefore already at first iteration a singular Hessian matrix (if used) will stop the procedure.

4.7 The case $x_{ij} = x_j$ and no constant term

We finally discuss the case that all CD variables are constant over individuals. However, all ST variables vary over individuals. The analysis is similar to that of section 4.7. However, it will turn out, that the default condition for a singular Hessian matrix is complexer than in the previous situation (as considered in subsection 4.6). We henceforth assume $q \geq 1$, $m \geq 1$, $r \geq 2$.

We start again from the matrix W_i in (4.7). However, this time we partition it as follows:

$$W_i = (W_1 \quad \vdots \quad W_{21}) \quad (4.32)$$

where

$$W_i = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_r \end{bmatrix} \quad , \quad W_{2i} = \begin{bmatrix} z_i & & & \\ & z_i & & \\ & & \ddots & \\ & & & z_i \\ & & & & 0 \end{bmatrix}$$

$[r \times m]$ $[r \times q(r-1)]$

The first order conditions (4.12b) are as follows:

$$\sum_i W_i^1 (y_i - p_i) = 0$$

or

$$W_1^1 \sum_i (y_i - p_i) = 0 \quad (4.33)$$

$$\sum_i W_{2i}^1 (y_i - p_i) = 0$$

The first line of (4.33) shows that the r rows of W_1 are not independent.

Therefore

$$\text{rg}(W_1) = \min \{ m, r-1 \} \quad (4.34)$$

if (4.33) holds. This result differs significantly from the result (4.28)

in the previous subsection. The Hessian matrix (4.25) can be written as

$$- \begin{pmatrix} W_1' \left(\sum_i v_i \right) W_1 & W_1' \sum_i V_i W_{2i} \\ \text{---} & \text{---} \\ \sum_i W_{2i}' V_i W_1 & \sum_i W_{2i}' V_i W_{2i} \end{pmatrix} \quad (4.35)$$

The left two blocks of this matrix can be written as AW_1 (ignoring the minus sign) where

$$A = \begin{bmatrix} W_1' \sum_i V_i \\ \sum_i W_{2i}' V_i \end{bmatrix} \quad (4.36)$$

$[m+q(r-1) \times r]$

Since we assume in this subsection $q \geq 1$, $m \geq 1$, $r \geq 2$, we always have $m - q(r - 1) \geq r$ so that

$$\text{rg}(A) \leq r \quad (4.37)$$

The maximal rank of W_1 depends on the number of its rows and columns:

$$\text{rg}(W_1) \leq \min \{ r, m \} \quad (4.38)$$

Therefore $\text{rg}(AW_1) \leq \min \{r, m\}$. However, under first order conditions (4.34) holds. In practice this leads to the following three situations:

a) $m < r$

Since AW_1 has m columns, these will be independent even if (4.33) holds.

b) $m = r$

In general AW_1 has m independent columns. However, under (4.33) no longer all columns remain independent, that is the Hessian matrix will become singular when the gradient approaches zero.

c) $m > r$

In this case the columns of AW_1 are never independent. The Hessian matrix is always singular.

REFERENCES

- Amemiya, T. (1985), *Advanced Econometrics*, Basil Blackwell, Oxford.
- Ben-Akiva, M., and S. Lerman (1985), *Discrete Choice Analysis*. MIT Press, Cambridge.
- Domencich, T.A., and D. McFadden (1975), *Urban Travel Demand*, North Holland, Amsterdam.
- Dhrymes, P. (1978), *Introductory Econometrics*, Springer Verlag, New York.
- Maddala, G.S. (1983), *Limited-Dependent and Qualitative Variables in Econometrics*, Cambridge University Press, Cambridge.
- McFadden, D. (1974), "Conditional Logit Analysis of Qualitative Choice Behavior", in: Frontiers of Econometrics, edited by P. Zarembka, Academic Press: New York, 105-142.
- Ronning, G. (1982), "Characteristic Values and Triangular Factorisation of the Covariance Matrix for Multinomial, Dirichlet and Multivariate Hypergeometric Distributions and Some Related Results", Statistische Hefte 23, 152-176.