Interdisziplinäres Institut für Raumordnung
Stadt- und Regionalentwicklung
Wirtschaftsuniversität Wien

Vorstand: o.Univ.Prof. Dr. Walter B. Stöhr
A-1090 Wien, Augasse 2-6, Tel. (0222) 34-05-25

1987

GUNTHER MAIER

DIE SCHÄTZUNG DISKRETER ENTSCHEIDUNGSMODELLE MIT HILFE DER SAS PROZEDUREN BPROBIT UND MNLOGIT

I I R - DISCUSSION 27/2.Auflage 1987

Überarbeitete und erweiterte Version von IIR-DISCUSSION 27, 1985

1. Einleitung

Diskrete Entscheidungsmodelle - oder präziser Modelle von Entscheidungen über diskrete Alternativen - haben sich in den letzten Jahren in mehreren ökonomischen Disziplinen zu Standardverfahren entwickelt. Ihre Fähigkeit, Entscheidungen über diskrete und qualitative Alternativen darzustellen macht sie zu wertvollen Instrumenten in Fachgebieten wie

* Verkehrsökonomie,
* Regionalökonomie,
* Arbeitsökonomie,

aber auch in betriebswirtschaftlichen Bereichen wie

* Marketing, und
* Public Relations.

In der Verkehrsökonomie werden sie verwendet zur Modellierung der Verkehrsmittelwahl, etwa der Entscheidung, ob eine bestimmte Strecke mit dem Auto oder mit öffentlichen Verkehrsmitteln zurückgelegt wird, und der Routenwahl. In der Regionalökonomie sind es Standortentscheidungen, Wanderungsentscheidungen, Pendelverhalten, Entscheidungen auf dem Wohnungsmarkt, etwa die Wahl zwischen einer Eigentumswohnung und einer Mietwohnung, in der Arbeitsökonomie Schulwahl, Berufswahl und die Entscheidung über die Teilnahme am Erwerbsprozeß (Partizipation) wo derartige Modelle zur Anwendung kommen. Neuere Anwendungen im Bereich des Marketing untersuchen etwa die Einflußfaktoren des Käuferverhaltens, die Marktchancen neuer Produkte, die Ursachen für Marktanteilsänderungen, und dergleichen.

Die vorliegende Arbeit verfolgt eine zweifache Zielsetzung. Einerseits soll sie einen kurzen Überblick über Theorie und Empirie diskreter Entscheidungsmodelle bieten (Abschnitt 2), andererseits die Anwendung der vom Autor erstellten SAS Prozeduren MNLOGIT und BPROBIT erläutern (Abschnitt 3). Diese Programme können wie Standard-Prozeduren des "statistischen Analysesystems" (SAS) aufgerufen werden und bieten neben der Schätzung eines multinomialen Logit-Modells (MNLOGIT) bzw. eines binären Probit-Modells (BPROBIT) alle üblichen Teststatistiken über die Signifikanz von Parametern und die Qualität des Gesamtmodells, die Möglichkeit von linearen Restriktionen und Konstantenrestriktionen sowie der Ausgabe der Schätzergebnisse, der Varianz-Kovarianz-Matrix und der Residuen. Die Verbindung mit SAS erlaubt die Verwendung der

Datenmanipulationsbefehle, der Matrixoperationen und anderer Statistikprozeduren dieses Programm-Pakets in Verbindung mit diskreten Entscheidungsmodellen.

2. Diskrete Entscheidungsmodelle

Dieser Abschnitt soll eine kurze Einleitung in Theorie und Empirie diskreter Entscheidungsmodelle bieten. Er beschränkt sich allerdings darauf, die für die Anwendung der SAS Prozeduren MNLOGIT und BPROBIT unbedingt notwendigen Informationen zu bieten und zielt nicht darauf ab, diese Modellfamilie grundlegend und ausführlich zu diskutieren. Ausführliche Darstellungen bieten etwa Domencich und McFadden, 1975, Hensher und Johnson 1981, Maddala 1983, Aldrich und Nelson 1984, Ben-Akiva and Lerman 1985.

Außerdem setzt diese Darstellung ein grundsätzliches Verständnis der mikroökonomischen Theorie des Konsumentenverhaltens (Green 1976, Gravelle und Rees 1981, Henderson und Quandt 1977) und des multivariaten linearen Regressionsmodells voraus (Johnston 1972, Maddala 1977, Wonnacott und Wonnacott 1979). Für den mathematischen Hintergrund sei auf Chiang (1974) verwiesen.

2.1. Grundlagen

Diskrete Entscheidungsmodelle unterstellen eine ähnliche Entscheidungssituation wie das Standard-Modell des Konsumentenverhaltens der Mikroökonomie. In beiden Fällen steht das Individuum (Konsument, Haushalt, etc.) vor dem Problem, aus einer Menge von möglichen Alternativen die beste auszuwählen. In der mikroökonomischen Konsumtheorie ist diese Menge an möglichen Alternativen durch die Einkommensrestriktion bestimmt, besteht aber theoretisch aus unendlich vielen Elementen. Jedes dieser Elemente stellt ein Güterbündel dar, ist also durch eine Kombination von Mengen verschiedener Güter beschrieben. Diskrete Entscheidungsmodelle hingegen gehen davon aus, daß nur eine endliche Zahl von Alternativen zur Verfügung steht. Im Falle einer Wohnungstypentscheidung etwa "Eigentumswohnung", "Mietwohnung", "eigenes Haus". Formal läßt sich diese Beziehung für die mikroökonomische Konsumtheorie darstellen als:

$$A_n = \{a_1, a_2, a_3, \ldots\} \qquad a_i = (q_i^1, q_i^2, q_i^3, \ldots) \tag{1}$$

A_n beschreibt die Menge der möglichen Alternativen des Individuums n, a ihre Elemente (Güterbündel) und q_{ij} die Menge von Gut j in Güterbündel i.

In der mikroökonomischen Konsumtheorie wird üblicherweise mit den Gütermengen (q) und nicht mit der Menge A und ihren Elementen gearbeitet. Für die vorliegende Fragestellung erweist sich allerdings diese Darstellung als geeigneter.

Die Bewertung von Alternativen durch das Individuum erfolgt in beiden Fällen auf Grund seiner Nutzenfunktion. Der Nutzen einer Alternative hängt in der Standardversion der mikroökonomischen Konsumtheorie von den sie beschreibenden Gütermengen ab. Es gilt also:

$$U_n(a_i) = U_n(q_i^1, q_i^2, q_i^3, \ldots) \tag{2}$$

Lancaster (1966) unterstellt, daß nicht die Gütermengen den Individuen Nutzen stiften, sondern die Charakteristika der Güter (siehe auch Green, 1976, Gravelle und Rees, 1981). Er definiert die Nutzenfunktion nicht über die Gütermengen sondern über Charakteristikamengen. Den Zusammenhang zwischen Gütermengen und Charakteristikamengen nennt Lancaster "Konsumtechnologie". Da dies für die vorliegende Fragestellung nicht relevant ist, soll auf den Zusammenhang zwischen der Menge der Alternativen, den Gütermengen und den Charakteristikamengen nicht näher eingegangen werden (siehe etwa Gravelle und Rees, 1981, Kapitel 5). Hier wird nur das Ergebnis von Lancaster's Analyse verwendet, nämlich daß die Alternativen durch Charakteristika, bezeichnet mit x, zu beschreiben sind und ihr Nutzen eine Funktion dieser Charakteristika ist.

$$U_n(a_i) = U_n(x_i^1, x_i^2, x_i^3, \ldots) \qquad a_i = (x_i^1, x_i^2, x_i^3, \ldots) \tag{3}$$

Betrachtet man diese Beziehung allein, so ist sie allgemeiner als Lancaster's Modell. Sie kann auch auf Alternativen angewendet werden, für die das Gütermengenkonzept nicht sinnvoll ist. Beispielsweise auf die verschiedenen Verkehrsmittel, mit denen eine Person ihren Arbeitsweg zurücklegen kann. In der weiteren Darstellung wird immer eine ordinale Nutzenfunktion der Form (3) unterstellt.

Die Entscheidungsmaxime beider Theorien, der neoklassischen Konsumtheorie und der Theorie diskreter Entscheidungen, lautet "wähle jene Alternative, deren Nutzen am größten ist." Um diese optimale Alternative zu bestimmen, verwendet die neoklassische Konsumtheorie die mathematische Methode der Optimierung unter Nebenbedingungen (siehe Chiang, 1974, Kapitel 12). Dies ist nur deshalb möglich, da sie von einer unendlich dichten (stetigen) Anordnung der Alternativen, also unendlich teilbaren Gütern ausgeht. Die Besonderheit einer Entscheidung über diskrete Alternativen besteht darin, daß nur eine endliche Anzahl deutlich unterscheidbarer Alternativen zur Wahl steht. Damit kann die Lösung des Optimierungsproblems nur folgendermaßen beschrieben werden:

$$a_{n,opt} = \{a_i \mid U_n(a_i) > U_n(a_j),\ a_j \varepsilon A_n,\ i \neq j\} \qquad (4)$$

Diese Lösung ist zwar allgemeiner als die Optimierungsbedingungen der neoklassischen Konsumtheorie - sie gilt auch für deren optimales Güterbündel - sie bietet aber keine direkte Möglichkeit, die Reaktion des Individuums auf Änderungen im System zu bestimmen. Während etwa Preisänderungen in der Konsumtheorie immer zu einer Veränderung in der Zusammensetzung des optimalen Güterbündels führen, führt Bedingung (4) nur dann zu einer Reaktion des Individuums, wenn a_i "überholt" wird. Im Fall von diskreten Alternativen kann daher ohne zusätzliche Annahmen aus dem beobachteten Verhalten von Indivduen nicht auf ihre Nachfragestruktur geschlossen werden.

2.2. Zufallsnutzentheorie

Um dieses Problem der Ableitung der Nachfragestruktur zu lösen, greifen diskrete Entscheidungsmodelle auf das in der Psychologie (siehe etwa Luce und Suppes, 1965) entstandene Konzept des Zufallsnutzens zurück. Die Zufallsnutzentheorie betrachtet den Nutzen einer Alternative nicht als deterministische Größe, sondern als Zufallsvariable. Auf der Grundlage dieser Interpretation kann (4) in eine Wahrscheinlichkeitsaussage umformuliert werden:

$$P(a_{n,opt}=a_i) = P(U_n(a_i) > U_n(a_j),\ a_j \in A_n,\ i \neq j) \qquad (5)$$

Sie besagt, daß die Wahrscheinlichkeit, daß für eine bestimmte Person Alternative i optimal ist - daß also diese Person Alternative i wählt - gleich der Wahrscheinlichkeit ist, daß der Nutzen von i größer als der aller anderen Alternativen in der Alternativenmenge A ist.

Für die Zufälligkeit des Nutzens gibt es mehrere Interpretationsmöglichkeiten. Die psychologische Interpretation vertritt die Ansicht, daß die Zufälligkeit des Zufallsnutzens sich aus dem Erkenntnis- und Bewertungsprozeß des Individuums ergibt. Die Zufälligkeit liegt damit in der Natur des entscheidenden Individuums, das weder in der Lage ist, die objektiven Charakteristika der Alternativen fehlerfrei zu erkennen, noch ihnen deterministisch Werte auf der Nutzenskala zuzuweisen. Die ökonomische Interpretation unterstellt weiterhin den deterministischen Nutzenmaximierer der Neoklassik, der auf der Grundlage seiner Nutzenfunktion die Alternativen bewertet und gemäß (3) die optimale auswählt. Der Analytiker allerdings hat nichtdie Fähigkeit "of 'peeping into the head' of each individual" (Hensher und Johnson, 1981, S.30), sodaß er:

- nicht alle relevanten Charakteristika der Alternativen kennt, es bleibt ein zufälliger Einfluß von unberücksichtigten Charakteristika.

- nicht alle sozioökonomischen Einflußfaktoren auf die Bewertung des Individuums kennt, es bleibt ein zufälliger Einfluß von unbeobachteten Geschmacksvariationen.

Außerdem kann der Analytiker

- die Werte von als relevant erkannten Charakteristika nicht exakt messen und

- einige Charakteristika überhaupt nur indirekt über "proxy"-Variable bestimmen.

Im Unterschied zur psychologischen Interpretation liegt die Zufälligkeit des Zufallsnutzens hier in der Unzulänglichkeit des analytischen Instrumentariums begründet. Welcher der beiden Interpretationen man folgt, ist für die theoretische Ableitung der Modellstruktur und die empirische Anwendung von diskreten Entscheidungsmodellen weitgehend irrelevant. In der weiteren Argumentation soll implizit immer die ökonomische Interpretation unterstellt werden.

Die Verwendung des Zufallsnutzenkonzepts in der theoretischen Ableitung führt zu weitgehender Konsistenz von Theorie und Empirie diskreter Entscheidungsmodelle. Da sie keine deterministischen Aussagen über das Verhalten der Individuen macht, kann die theoretisch ermittelte Modellstruktur in der empirischen Anwendung direkt verwendet werden. Die Begrenztheit des Wissens des Analytikers wird bereits in der Theorie berücksichtigt und nicht, wie sonst allgemein üblich, erst bei der empirischen Anwendung eingestanden.

Allgemein kann der Zufallsnutzen auch auf etwas andere Art angeschrieben werden, die für die weitere Darstellung in Abschnitt 2.4. von Vorteil ist.

$$U_{in} = V_{in} + \varepsilon_{in} \quad {}^{1)} \qquad (6)$$

Die Zufallsvariable U_i wird dabei unterteilt in einen deterministischen Teil, V_i, der als "mittlerer Nutzen" der Alternative interpretiert werden kann, und eine Zufallskomponente ε_i, die nun die Zufälligkeit des Zufallsnutzens als Abweichung vom "mittleren Nutzen" angibt.

Im Zusammenhang mit dem Ausdruck "mittlerer Nutzen" ist zu beachten, daß der Nutzen nicht kardinaler sondern ordinaler Natur ist. Solange wir keine Annahmen über die Verteilung von U - oder äquivalent dazu über die Verteilung von ε - treffen, ist das Modellergebnis invariant gegenüber einer monotonen Transformation von U. Unter den im Zusammenhang mit diskreten Entscheidungsmodellen üblichen Verteilungsannahmen ist der Zufallsnutzen zwar nicht mehr ordinal, das Modellergebnis ist jedoch invariant gegenüber Addition und gegenüber Multiplikation mit einer positiven Konstanten.

Die Unterteilung des Zufallsnutzen in Form von (6) dient dazu, die Zufallskomponente vom systematischen Einfluß der vom Analytiker meßbaren Charakteristika zu trennen. Die genaue Spezifikation der beiden Teile des Zufallsnutzens wird in Abschnitt 2.4 diskutiert werden.

Die Aussage der Zufallsnutzentheorie über das tatsächliche Verhalten der Individuen ist schwächer und damit allgemeiner als jene der deterministischen Theorie. Während (4) aussagt, daß a_i optimal ist und daher vom Individuum gewählt wird, behauptet die Zufallsnutzentheorie lediglich, daß Alternative i mit der in (5) beschriebenen Wahrscheinlichkeit gewählt wird. Die tatsächliche Entscheidung des Individumms stellt die Realisierung einer multinomial verteilten Zufallsvariablen dar, mit den Parametern n=1 und p, wobei p den Vektor der Wahrscheinlichkeiten für alle Alternativen darstellt, der durch (5) gegeben ist.

2.3. Die allgemeine Form diskreter Entscheidungsmodelle

Bedingung (5) trifft zwar eine Wahrscheinlichkeitsaussage über die Entscheidungen des Individuums, doch ist sie in dieser Form einer weiteren Behandlung nur schwer zugänglich. Als nächster Schritt soll daher (5) in eine geeignetere Form umgewandelt werden.

Da die Nutzen der einzelenen Alternativen, entsprechend dem Zufallsnutzenkonzept, Zufallsvariable sind, kann die Dichte ihre gemeinsame Verteilung als

$$f(U_{1n}, U_{2n}, U_{3n}, \dots) \qquad (7)$$

angeschrieben werden. Verschiedene Klassen von diskreten Entscheidungsmodellen unterscheiden sich durch unterschiedliche Annahmen über die Verteilung f, bzw. alternativ dazu, durch Annahmen über die Verteilung der Zufallskomponente ε. Hier treffen wir allerdings noch keinerlei Annahmen über die Verteilung der Zufallsnutzen, (7) definiert sie nur als multivariate Verteilung allgemeinster Form. [2)]

Unterstellt man für einen Moment, daß der Nutzen der Alternative i, deren Auswahlwahrscheinlichkeit bestimmt werden soll, fix und gleich U_i^* ist, so ist die Wahrscheinlichkeit, daß i gewählt wird, gleich der Wahrscheinlichkeit, daß die Nutzen aller anderen Alternativen kleiner als U_i^* sind. Bei insgesamt I Alternativen gilt also:

$$P(a_i | U_i = U_i^*) = \int_{-\infty}^{U_i^*} \int_{-\infty}^{U_i^*} .. \int_{-\infty}^{U_i^*} \int_{-\infty}^{U_i^*} .. \int_{-\infty}^{U_i^*} f(U_1, U_2, .., U_{i-1}, U_i^*, U_{i+1}, .., U_I)$$

$$dU_1 \, dU_2 .. dU_{i-1} \, dU_{i+1} .. dU_I \tag{8}$$

Dieser Integralausdruck bezeichnet nichts anderes als das vorangegangene verbale Statement, nämlich die Wahrscheinlichkeit, daß die Nutzen aller anderen Alternativen als i kleiner sind als U_i^*.

Beziehung (8) gilt nur für einen bestimmten fixen Wert von U_i. Die tatsächliche Wahrscheinlichkeit für die Wahl der Alternative i ist gleich der "Summe" von (8) über alle möglichen Werte von U_i. Da f in U_i stetig ist, wird diese "Summierung" durch Integration über alle möglichen Werte von U_i gebildet. Die für eine weitere Behandlung geeignetere Form von (5) ergibt sich daher als:

$$P_i = \int_{-\infty}^{U_i} \int_{-\infty}^{U_i} .. \int_{-\infty}^{U_i} \int_{-\infty}^{+\infty} \int_{-\infty}^{U_i} .. \int_{-\infty}^{U_i} f(U_1, U_2, .., U_I) \, dU_1 \, dU_2 \, .. \, dU_I \tag{9}$$

Da (7) nur die Zufallsnutzenannahme in formaler Form darstellt, sind (5) und (9) äquivalent. Beide beschreiben die Wahrscheinlichkeit, mit der ein nutzenmaximierendes Individuum eine Alternative wählt. Sie stellen eine allgemeine Form diskreter Entscheidungsmodelle dar.

Die deterministische Auswahlentscheidung, wie sie in der neoklassischen Theorie immer unterstellt wird, stellt einen Spezialfall von Bedingung (9) dar. Sie ist dann gegeben, wenn f für eine - die deterministisch gegebene - Kombination von Nutzenwerten den Wert 1, für alle übrigen den Wert Null annimmt. In diesem Fall wird die nutzenmaximale Alternative mit Sicherheit gewählt, alle anderen Alternativen haben Auswahlwahrscheinlichkeit Null.

2.4. Spezielle diskrete Entscheidungsmodelle

Wie bereits erwähnt unterscheiden sich die verschiedenen Klassen diskreter Entscheidungsmodelle durch unterschiedliche Annahmen über die Verteilung der Zufallskomponente des Zufallsnutzens und damit über die Verteilung von U. Bevor auf die beiden wichtigsten Modelltypen, das Probit- und das Logit-Modell, näher eingegangen wird, sollen noch einige kurze Anmerkungen zur Zahl der Alternativen und zur Struktur des deterministischen Nutzenteils V gemacht werden.

2.4.1. Binäre versus multinomiale Modelle

In den bisherigen Ausführen zu diskreten Entscheidungsmodellen wurden keinerlei Einschränkungen über die Zahl der Alternativen in der Alternativenmenge A getroffen. Implizit durch die Gegenüberstellung von Konsumtheorie und Theorie diskreter Entscheidungen wurde unterstellt, daß die Menge A aus einer endlichen Zahl diskreter Elemente besteht. Dies ist jedoch keine konstituierende Eigenschaft von Modellen, die auf der Zufallsnutzenannahme aufbauen. Stetige Varianten des Logit-Modells diskutieren etwa McFadden (1976), Ben-Akiva und Watanatada (1981), Ben-Akiva, Litinas und Tsunokawa (1985). Modelle mit einer endlichen aber nicht a-priori beschränkten Zahl an diskreten Alternativen werden üblicherweise als multinomiale Modelle bezeichnet. Beispiele hierfür sind etwa Standortentscheidungsmodelle und Modal-Split-Modelle mit den Alternativen "Auto", "öffentliche Verkehrsmittel", "Fahrrad", "zu Fuß gehen", etc.

Läßt ein Modell a-priori nur zwei Alternativen zu, etwa "ja" - "nein" oder "Auto" - "öffentliches Verkehrsmittel", so spricht man von einem binären Modell. Binäre Modelle sind ein Spezialfall der multinomialen Modelle. Die Unterscheidung ist deshalb von Bedeutung, da manche Verteilungsannahmen über die Zufallskomponente des Zufallsnutzen zwar im binären, nicht jedoch im multinomialen Fall zu mathematisch und rechentechnisch leicht handhabbaren Modellen führen. Beispiele sind etwa das - hier nicht diskutierte - lineare Wahrscheinlichkeitsmodell und das Probit-Modell.

Da im Probit-Modell die Integrale (9) nicht analytisch sondern nur numerisch berechnet werden könne, bringt jede zusätzliche Alternative eine Vervielfachung der Rechenzeit mit sich und die Anwendung stößt schon bei einer geringen Zahl von Alternativen an Grenzen.

2.4.2. Die Struktur des deterministischen Nutzens

Eines der Ziele bei der Anwendung diskreter Entscheidungsmodelle besteht darin, Erkenntnis über den Einfluß der Charakteristika und anderer Variabler auf die Entscheidung von Individuen zu gewinnen. Damit können Fragen nach der Reaktion von Straßenbahnpassagieren auf Fahrpreisänderungen, der Bewertung von Preis und Qualität eines Produktes durch die Konsumenten, oder der Auswirkung demographischer Änderungen auf die Siedlungsstruktur beantwortet werden. Da diese Einflußfaktoren für den Analytiker beobachtbar sind, wirken sie sich nicht in der Zufallskomponente sondern im deterministischen Teil des Zufallsnutzens aus. Wir bezeichen die Charakteristika der Alternative i, wie sie sich für das Individuum n darstellen, mit dem Vektor z_{in}. Da das Ziel der Analyse darin besteht, Aussagen für eine Gruppe von Personen zu machen und nicht für Einzelpersonen, muß auch der mögliche Einfluß von Besonderheiten der Individuen auf ihr Verhalten berücksichtigt werden. Wir postulieren, daß die Besonderheiten der Individuen annähernd durch ihre sozioökonomischen Eigenschafen - bezeichnet mit s_n - beschrieben werden können.

Damit kann der deterministische Teil des Nutzens, den ein Individuum n der Alternative i zumißt, angeschrieben werden als:

$$V_{in} = V(z_{in}, s_n) \tag{10}$$

Um zu allgemein verwendbaren Ergebnissen zu gelangen, definieren wir einen Vektor von erklärenden Variablen X_{in}, der sich folgendermaßen aus den Charakteristika der Alternativen und den Eigenschaften der Individuen ergibt

$$X_{in} = g(z_{in}, s_n) \tag{11}$$

X sind die in der Schätzung tatsächlich verwendeten Variablen, während die Funktion g jene Transformationen darstellt, die der Analytiker auf Grund theoretischer Überlegungen oder auf Grund von Besonderheiten der Daten mit den Rohdaten durchführt. Im einfachsten Fall bildet g ein Charakteristikum der Alternativen in ein Element von X ab, doch erlaubt diese Funktion auch das Quadrieren oder Logarithmieren von Rohdaten oder die Konstruktion von "Interaktionen" durch Multiplikation eines Charakteristikums mit einem anderen oder mit sozioökonomischen Variablen. In SAS sind dies Datenmanipulationen, die im DATA STEP durchgeführt werden. Auf Besonderheiten dieser Transformation im Zusammenhang mit diskreten Entscheidungsmodellen soll in Abschnitt 2.6 eingegangen werden.

Während die Funktion g vom Analytiker gewählt wird, ist der Zusammenhang zwischen dem deterministischen Nutzenteil und X zu schätzen. Dazu muß allerdings die funktionale Form dieser Beziehung näher spezifiziert werden. Die mit Abstand am häufigsten verwendete Spezifikation ist die einer "in den Parametern linearen" Funktion. Bezeichnet man die Parameter mit ß und ihre Anzahl mit K, so bedeutet dies

$$V_{in} = \sum_{k=1}^{K} \beta_k X_{ink} \quad \text{oder in Matrixnotation: } V_{in} = X_{in}\beta \qquad (12)$$

Die Parameter ß sind unbekannt und müssen aus dem beobachteten Verhalten geschätzt werden (Näheres siehe Abschnitt 2.5). Besonders wichtig in diesem Zusammenhang ist die Tatsache, daß (12) keinen linearen Zusammenhang zwischen dem deterministischen Nutzenteil und den Charakteristika bzw. sozioökonomischen Eigenschaften bedeutet. Über die entsprechende Wahl von g hat der Analytiker die Möglichkeit, Nichtlinearitäten und Interaktionen in diesen Zusammenhang einzubauen und durch Tests deren Bedeutung zu überprüfen.

Die Funktionen (11) und (12) bzw. (10), die die Charakteristika der Alternativen und sozioökonomischen Eigenschaften der Individuen in den deterministischen Nutzenteil abbilden, wurden sowohl über die Alternativen als auch über die Individuen als identisch angenommen. Dies ist einerseits eine statistische Notwendigkeit (Zahl der Freiheitsgrade), andererseits aber auch inhaltlich sinnvoll, da unser Interesse vor allem dem aggregierten Verhalten einer Gruppe von Individuen gilt. Da aber auch für die sozioökonomischen Eigenschaften nicht zu erwarten ist, daß der Analytiker ihren Einfluß vollkommen erfassen kann, beinhaltet die Zufallsgröße auch unbeobachtete Variationen über die Individuen. Zusammenfassend kann daher der Nutzen, den ein Individuum n einer Alternative i zumißt, angeschrieben werden als

$$U_{in} = X_{in}\beta + \varepsilon_{in} \tag{13}$$

wobei X durch (11) definiert ist. Die Auswahlwahrscheinlichkeit der Alternative i ergibt sich aus (5) bzw. (9).

2.4.3. Das binäre Probit-Model

Wegen der oben erwähnten Probleme im multinomialen Fall beschränkt sich diese Darstellung auf das binäre Probit-Modell. Außerdem ist auch die SAS Prozedur BPROBIT auf den binären Fall beschränkt.

Bezeichnet man die beiden Alternativen in der Alternativenmenge A mit i und j, so gelangt man durch Einsetzen von (13) in (5) zu folgender einfacher Form für die Auswahlwahrscheinlichkeit.

$$P_{in} = \text{Prob}\ [(X_{in} - X_{jn})\beta > (\varepsilon_{jn} - \varepsilon_{in})] \tag{14}$$

Gleichung (14) gilt für jedes binäre diskrete Entscheidungsmodell, nicht nur für das binäre Probit-Modell. Sie zeigt, daß nicht die absolute Höhe der erklärenden Variablen, sondern deren Differenz die Auswahlwahrscheinlichkeit beeinflußt. Aus (14) ist leicht zu ersehen, daß sowohl die Addition einer Konstanten zu beiden Nutzen, als auch die Multiplikation beider Nutzen mit einer positiven Konstanten die Ungleichung in den eckigen Klammern und damit die Auswahlwahrscheinlichkeiten unverändert läßt. Dieses Ergebnis ist an einem binären Modell leicht nachzuweisen, es gilt aber auch für jedes multinomiale Modell.

Die Verhaltensannahme, die zum binären Probit-Modell führt, ist die, daß ε_{in} und ε_{jn} bivariat normalverteilt sind mit Mittelwert Null, Varianz σ_i^2 bzw. σ_j^2 und Kovarianz σ_{ij}.

Aus den Eigenschaften der Normalverteilung, daß nämlich die Differenz normalverteilter Zufallsvariabler wieder normalverteilt ist, und aus (14) ergibt sich

$$P_{in} = \Phi\left(\frac{(X_{in} - X_{jn})\beta}{\sigma}\right)$$

$$\text{mit} \quad \sigma^2 = \sigma_i^2 + \sigma_j^2 - 2\sigma_{ij} \tag{15}$$

Φ ... Verteilungsfunktion der standardisierten Normalverteilung

Dadurch, daß sich die bivariate Verteilung der Zufallsterme durch die Differenzbildung auf eine univariate reduziert, kann die tabellierte Funktion verwendet werden. Bei multinomialen Versionen des Probit-Modells ist allerdings die numerische Berechnung des Integralausdrucks (9) notwendig.

Da sowohl der Parametervektor β unbekannt ist als auch das Skalar σ, können die einzelnen Elemente des Parametervektors erst geschätzt werden, wenn σ auf einen arbiträren positiven Wert fixiert wurde. Üblicherweise ist das der Wert Eins. Für die Interpretation bedeutet dies, daß die absolute Größe der geschätzten Parameter nicht interpretierbar ist, sondern nur ihre Relation zueinander. Dies ist ähnlich der aus der Konsumtheorie bekannten Beziehung zwischen den Grenznutzen und der Grenzrate der Substitution.

Auf Grund von Informationen über die von den Individuen tatsächlich gewählte Alternative und die erklärenden Variablen X kann der Parametervektor β geschätzt werden. Eines der möglichen Schätzverfahren, die Maximum Likelihood Schätzung unter Verwendung des Newton-Raphson-Algorithmus, wird in Abschnitt 2.5 diskutiert werden. Dieses Verfahren kommt auch in den SAS Prozeduren BPROBIT und MNLOGIT zur Anwendung.

2.4.4. Das multinomiale Logit-Modell

Im Gegensatz zu dem zuvor diskutierten binären Probit-Modell treffen wir beim multinomialen Logit-Modell keine Einschränkungen über die Zahl der in der Menge A enthaltenen Alternativen. Für verschiedene Personen kann diese Menge auch unterschiedlich groß sein.

Die zum Logit-Modell führende Verteilunsannahme ist die, daß die Zufallsterme des Zufallsnutzens unabhängig identisch Gumbel verteilt sind. Ihre Verteilungsfunktion lautet also

$$P(\varepsilon_{in} < \varepsilon) = \exp\{-\exp(-\mu(\varepsilon-\eta))\} \qquad (16)$$

Andere Bezeichnungen für diese Verteilung sind "Doppelexponentielle Verteilung" und "Extremwertverteilung erster Ordnung". Oft wird sie fälschlich auch als Weibull Verteilung bezeichnet.

Die Verteilung hat zwei Parameter, einen Lageparameter η, der proportional zum Mittelwert, und einen Streuungsparameter μ, der umgekehrt proportional zur Standardabweichung ist. (Für eine ausführlichere Diskussion siehe Ben-Akiva und Lerman, 1985 und Johnson und Kotz 1970). Da ε die Abweichung vom mittleren Nutzen repräsentiert, setzen wir den Lageparameter für alle Individuen und Alternativen gleich Null. Besteht auf Grund theoretischer Erkenntnisse oder auf Grund der Datenstruktur die Vermutung, daß zwischen den Alternativen Unterschiede in den Lageparametern bestehen, so können diese Unterschiede durch "alternativenspezifische Konstante" abgedeckt werden (Näheres siehe Abschnitt 2.6). Die Annahme identisch verteilter Zufallsvariabler besagt, daß der Streuungsparameter über alle Alternativen und Individuen gleich sein muß.

Diese Verteilungsannahmen führen unabhängig von der Zahl der Alternativen zu Auswahlwahrscheinlichkeiten von folgender analytisch geschlossener Form (für die Ableitung siehe etwa Domencich und McFadden, 1975, Hensher und Johnson, 1981, Anas, 1981)

$$P_{in} = \frac{\exp(\mu X_{in}\beta)}{\sum_{j\varepsilon A_n} \exp(\mu X_{jn}\beta)} \qquad (17)$$

Aus (17) ist leicht zu erkennen, daß - wegen der Exp-Funktion bzw. der Summe über alle Alternativen im Nenner - alle Wahrscheinlichkeiten zwischen Null und Eins liegen und sich über alle Alternativen auf Eins summieren. Außerdem hat der unbekannte Streuungsparameter μ eine ähnliche Funktion wie σ im Probit-Modell. Auch er muß auf einen arbiträren positiven Wert fixiert werden. Üblicherweise ist dies Eins, doch ist zu beachten, daß damit die Varianz des Zufallsterms auf ca. 1.645 fixiert wird. Dies ist für den Vergleich von Parametern binärer Probit- und Logit-Modelle von Bedeutung (Amemiya, 1981, Maddala, 1983).

Daß auch im multinomialen Logit-Modell nur die Differenzen der erklärenden Variablen die Auswahlwahrscheinlichkeiten bestimmen, zeigt sich, wenn man Zähler und Nenner von (17) durch den Ausdruck im Zähler dividiert.

$$P_{in} = \frac{1}{\sum_{j \epsilon A_n} \exp[\mu (X_{jn} - X_{in}) \beta]} \qquad (18)$$

Damit ist aber auch das Logit-Modell invariant gegenüber der Addition einer Konstanten. Multipliziert man alle Nutzen mit einer positiven Konstanten, so verringert dies einerseits den Streuungsparameter um diesen Faktor, andererseits kann diese Konstante aus dem Ausdruck für die Differenz der deterministischen Nutzenteile herausgehoben und mit μ zusammengefaßt werden, womit (18) bzw. (17) auch gegenüber der Multiplikation mit einer positiven Konstanten invariant ist. Folglich gilt auch für das Logit-Modell, daß nur die relativen, nicht aber die absoluten Werte der geschätzten Parameter interpretiert werden können.

Die Annahme der Unabhängigkeit der Zufallsterme führt zu einer wichtigen Eigenschaft des Logit-Modells, zur "Unabhängigkeit von irrelevanten Alternativen" ("independence from irrelevant alternatives": IIA). Sie besagt, daß das Verhältnis der Auswahlwahrscheinlichkeit zweier Alternativen unabhängig von anderen Alternativen ist. Bildet man das Verhältnis zweier durch (17) definierter Auswahlwahrscheinlichkeiten, so ist diese Eigenschaft leicht nachzuweisen (μ=1)

$$\frac{P_{in}}{P_{kn}} = \frac{\exp(X_{in}\beta)}{\exp(X_{kn}\beta)} = \exp[(X_{in} - X_{kn})\beta] \qquad (19)$$

Die IIA-Eigenschaft ist auf der einen Seite eine Stärke des Logit-Modells, da sie die Aufnahme neuer Alternativen in bestehende Modelle erleichtert, auf der anderen Seite stellt sie eine starke a-priori Beschränkung für die Auswahlwahrscheinlichkeiten dar. Zur Illustration für die Restriktivität der IIA-Eigenschaft wird in der Literatur häufig auf das "red bus - blue bus"-Problem verwiesen (Mayberry 1970).

Dabei wird als Ausgangssituation unterstellt, für eine bestimmte Strecke bestehen zwei Transportmöglichkeiten, "Auto" und "Bus", jede mit Auswahlwahrscheinlichkeit von 0,5. Wird nun neben dem ersten, beispielsweise roten Bus ein zweiter eingeführt, der sich vom ersten in nichts als seiner blauen Farbe unterscheidet, so sollten sich, wird argumentiert, die Auswahlwahrscheinlichkeiten auf 0,5, 0,25, 0,25 ändern. Die IIA-Eigenschaft impliziert allerdings für alle drei Alternativen Auswahlwahrscheinlichkeiten von 1/3.

In diesem Zusammenhang sind allerdings zwei Dinge von besonderer Bedeutung: Erstens rührt eine Verletzung der IIA-Eigenschaft von Korrelation in den nicht berücksichtigten, also in den Zufallsterm abgeschobenen, Einflußfaktoren her. Diese Ursache kann durch Aufnahme dieser Einflüsse in den deterministischen Teil, also eine Verbesserung der Modellspezifikation, beseitigt werden. Zweitens gilt die IIA-Eigenschaft immer nur auf jenem Aggregationsniveau, auf dem die Modellschätzung durchgeführt wurde, nicht aber auf höherem Aggregationsniveau, das üblicherweise für die Prognose von Interesse ist. Dies soll an einem Beispiel illustriert werden (siehe auch Ben-Akiva und Lerman 1985, S. 109f).

Eine Bevölkerung bestehe aus nur 2 Personengruppen mit je 100 Personen, die die Wahl zwischen den beiden Transportmitteln "Auto" und "roter Bus" hat. In der Ausgangssituation sollen folgende Auswahlwahrscheinlichkeiten gelten:

	Gruppe 1	Gruppe 2	erwartete Nachfrage
Auto	o.92	o.o8	100
Bus rot	o.o8	o.92	100

Beide Verkehrsmittel haben damit eine erwartete Nachfrage von je 100 Personen. Führt man einen blauen Bus ein, so müssen sich die Auswahlwahrscheinlichkeiten der beiden Personen gemäß IIA-Eigenschaft ändern auf:

	Gruppe 1	Gruppe 2	erwartete Nachfrage
Auto	o.85	o.o4	89
Bus rot	o.o7	o.48	55
Bus blau	o.o7	o.48	55

Bildet man allerdings wiederum die erwartete Nachfragen, so zeigt sich, daß die Nachfrage nach dem roten Bus stärker zurückgegangen ist, als die nach dem Auto.

Dennoch stellt IIA eine restriktive Eigenschaft des Logit-Modells dar und empirisch geschätzte Modelle sollten immer auf Verletzung dieser Eigenschaft getestet werden (siehe Abschnitt 2.7). Weisen Tests auf eine Verletzung der IIA-Eigenschaft hin, so heißt dies nicht, daß das Logit-Modell an und für sich, sondern nur, daß das Logit-Modell in der verwendeten Spezifikation für das analysierte Problem inadäquat ist. Die Lösung des Problems sollte daher zuerst in einer besseren Beschreibung der Alternativen und entscheidenen Personen und einer Verbesserung der Spezifikation des deterministischen Nutzenteils gesucht werden.

Wegen seiner konzeptuellen Einfachheit ist das Logit-Modell die mit Abstand am häufigsten verwendete Version eines multinomialen diskreten Entscheidungsmodells und es wurde auch in vielfacher Hinsicht erweitert. Diese Erweiterungen reichen von der Behandlung aggregierter Alternativen über Modelle mit mehreren interdependenten Entscheidungsebenen ("nested logit") bis zu dynamischen Modellversionen. Näheres siehe Domencich und McFadden, 1975, Ameniya, 1981, Hensher und Johnson, 1981, Manski und McFadden, 1981, Maddala, 1983, Ben-Akiva und Lerman, 1985.

2.5. Schätzung

Methodisches Ziel der Darstellung in Abschnitt 2.4 war es, einen formalen Zusammenhang zwischen den erklärenden Variablen und den Auswahlwahrscheinlichkeiten abzuleiten. Für das binäre Probit-Modell ist dies Gleichung (15), für das multinomiale Logit-Modell (17). Wenn die getroffenen Annahmen zutreffen, so sind dies die Wahrscheinlichkeiten, die die Wahlentscheidung des Individuums steuern.

Wegen der diskreten Natur der Alternativen kann die Entscheidung des Individuums nur insofern beobachtet werden, welche Alternative es tatsächlich gewählt hat. Im statistischen Sinn ist die Entscheidung des Individuums eine multinomial verteilte Zufallsvariable mit den Parametern n = 1 und den sich aus (15) bzw. (17) ergebenden Wahrscheinlichkeiten. Die zu beobachtende tatsächliche Entscheidung ist eine Realisierung dieser Zufallsvariablen.

Wir wollen die tatsächliche Entscheidung des Individuums durch einen Vektor Y_n beschreiben. Dieser Vektor hat ebenso viele Elemente wie dem Individuum Alternativen zur Auswahl stehen und jenes Element, das der gewählten Alternative entspricht ist Eins, alle anderen sind Null. Die Elemente sollen mit y_{in} bezeichnet werden, Y_{in} sei jener Vektor, dessen i-tes Element Eins ist. Damit gilt

$$P(Y_{in}) = M_n \, P_{1n}^{y_{1n}} \, P_{2n}^{y_{2n}} \cdots P_{In}^{y_{In}} = P_{in}$$

$$\text{mit} \quad M_n = \frac{(\sum\limits_i y_{in})!}{y_{1n}! \, y_{2n}! \cdots y_{In}!} = 1 \tag{20}$$

Eine ähnliche Beziehung läßt sich auch dann aufstellen, wenn keine Individualinformationen, sondern nur Informationen über Gruppen von Individuen vorliegen. Allerdings muß dabei angenommen werden, daß die Gruppen intern homogen, also durch einen Vektor von sozioökonomischen Variablen zu beschreiben sind. Die Auswahlwahrscheinlichkeiten können durch eine Funktion beschrieben werden, die sich von (15) bzw. (17) nur dadurch unterscheidet, daß der Index n durch einen Index g, der die Gruppe bezeichnen soll zu ersetzen ist. Die Elemente des Vektors Y sind nicht mehr Null und Eins, sondern geben an, wie viele Personen der Gruppe die entsprechende Alternative gewählt haben. Die Wahrscheinlichkeit einen bestimmten Y-Vektor zu beobachten ist dann

$$P(Y_g) = M_g \, p_{1g}^{\,y_{1g}} \, p_{2g}^{\,y_{2g}} \cdots p_{Ig}^{\,y_{Ig}}$$

$$\text{mit} \quad M_g = \frac{(\sum_i y_{ig})!}{y_{1g}! \; y_{2g}! \cdots y_{Ig}!} \tag{21}$$

2.5.1. Das Prinzip der Maximum-Likelihood-Schätzung

Das Maximum-Likelihood-Verfahren stellt eine Möglichkeit dar, die unbekannten Parameter ß des Modells zu schätzen. Dabei wird versucht, jenen Parametervektor zu finden, für den die beobachteten Entscheidungen der Individuen am wahrscheinlichsten sind.

Geht man von einem beliebigen Vektor b aus, so können für das Probit-Modell aus (15), für das Logit-Modell aus (17) die diesem angenommenen Parametervektor entsprechenden Auswahlwahrscheinlichkeiten für alle Individuen berechnet werden. Mit diesen geschätzten Auswahlwahrscheinlichkeiten - sie sollen als $p_{in}(b)$ bezeichnet werden - kann die Wahrscheinlichkeit für das tatsächlich beobachtbare Verhalten der Individuen aus (20) berechnet werden. Da die Auswahlwahrscheinlichkeiten der Individuen annahmegemäß von einander unabhängig sind, ergibt sich die Wahrscheinlichkeit, mit der der angenommene Parametervektor b zu den beobachtbaren Entscheidungen führt, als Produkt dieser Wahrscheinlichkeiten für die Individuen.

$$L^*(b) = \prod_n M_n \, p_{1n}(b)^{y_{1n}} \, p_{2n}(b)^{y_{2n}} \cdots p_{In}(b)^{y_{In}} \tag{22}$$

Dies ist eine Funktion in b und sie wird als Likelihoodfunktion bezeichnet. Ziel einer Maximum-Likelihood-Schätzung ist es, jenen Vektor b zu finden, der die Funktion L^* maximiert. Aus rechentechnischen Gründen wird dazu üblicherweise der Logarithmus der Likelihoodfunktion verwendet. Da der Logarithmus eine monotone Transformation darstellt, maximiert jener Vektor, der die Log-Likelihoodfunktion maximiert, auch die Likelihoodfunktion. Die Log-Likelihoodfunktion zu (22) lautet

$$L(b) = \sum_n \ln M_n + \sum_n \sum_i y_{in} \ln p_{in}(b) \tag{23}$$

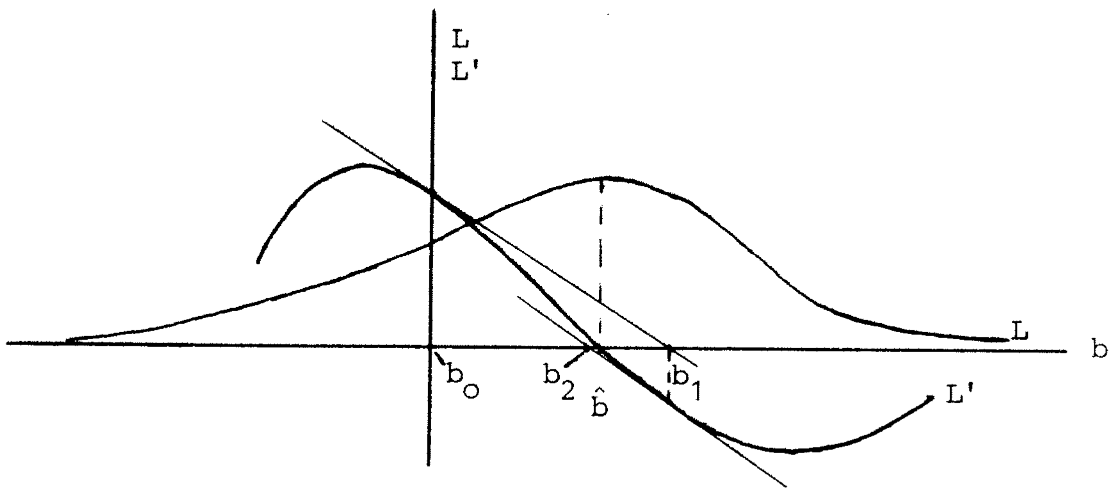
Notwendige Bedingung für das Maximum der Log-Likelihoodfunktion ist, daß ihre ersten Ableitungen nach den Elementen von b gleich Null sind. Da sowohl für das binäre Probit- als auch für das multinomiale Logit-Modell die Likelihoodfunktion unter sehr allgemeinen Bedingungen konkav ist (McFadden 1974), ist diese Bedingung üblicherweise auch hinreichend und liefert das globale Maximum der Likelihoodfunktion. Allerdings kann der Vektor b, der die Maximierungsbedingung erfüllt, wegen der nichtlinearen Beziehung zwischen b und L nur iterativ ermittelt werden. Dabei wird von einem Startvektor für b ausgegangen und in jeder Iteration versucht, einen besseren Vektor b zu ermitteln. Der optimale Vektor $\hat{b}$ und damit der Maximum-Likelihood-Schätzwert für ß ist dann erreicht, wenn die Maximierungsbedingung gut genug erfüllt ist. Die Maximum-Likelihood-Schätzer sind unter allgemeinen Bedingungen konsistent und asymptotisch normalverteilt (Amemiya, 1981, Rao, 1973), wobei die asymptotische Varianz-Kovarianz-Matrix über die für $\hat{b}$ berechnete negative Inverse der Matrix der zweiten Ableitung von L (Hesse Matrix) berechnet werden kann.

2.5.2. Der Newton-Raphson-Algorithmus

Ein wichtiges Problem im Zusammenhang mit Maximum-Likelihood-Schätzungen ist die Bestimmung des jeweils nächsten Wertevektors von b. Der Newton-Raphson-Algorithmus stellt eine von vielen Möglichkeiten dar, die jeweils verbesserte Schätzung zu ermitteln.

Das Prinzip des Alorithmus soll an Hand folgender vereinfachter Darstellung erläutert werden.

Darstellung 1



L stellt die zu maximierende Likelihoodfunktion dar. L' ihre erste Ableitung. Wir gehen vom Startwert b_0 aus und bestimmen die ersten beiden Ableitungen der Log-Likelihoodfunktion an diesem Punkt (L_0' und L_0''). Da die erste Ableitung ungleich Null ist, muß ein neuer Wert für b ermittelt werden. Dazu approximieren wir die Funktion L', von der wir im Moment nur Funktionswert und Steigung am Punkt b_0 kennen, durch eine Gerade:

$$y = L_0''(b-b_0) + L_0' \tag{24}$$

Die verbesserte Schätzung für b (b_1) finden wir dort, wo diese Gerade Null ist.

$$b_1 = b_0 - L_0' / L_0'' \tag{25}$$

Auch an diesem neuen Punkt können wir wieder erste und zweite Ableitung der Log-Likelihoodfunktion berechnen und gegebenenfalls eine verbesserte Schätzung für b.

Diese einfache Form gilt nur für den Fall eines einzigen zu schätzenden Parameters. Bezeichnen wir mit d den Vektor der ersten und mit H die Matrix der zweiten Ableitung von L, so lautet die allgemeine Version von (25)

$$b_1 = b_0 - H_0^{-1} d_0 \tag{26}$$

Der Newton-Raphson-Algorithmus führt im allgemeinen rasch zum optimalen Vektor $\hat{b}$. In Ausnahmefällen - in Darstellung 1 etwa bei einem Startwert bedeutend unter b_0 - kann er aber auch zu schlechteren Schätzwerten von b führen, wodurch üblicherweise Probleme bei der Berechnung der Ableitung auftreten. Ein Maximum-Likelihood-Schätzprogramm, das den Newton-Raphson-Algorithmus verwendet, sollte daher Möglichkeiten vorsehen, derartige Probleme zu umgehen. Wie dies in den SAS Prozeduren BPROBIT und MNLOGIT zu bewerkstelligen ist, wird in Kapitel 3 erläutert werden.

2.6. Datenstrukturen in diskreten Entscheidungsmodellen

Die oben diskutierten Eigenschaften diskreter Entscheidungsmodelle führen in Bezug auf die Struktur der erklärenden Variablen zu einigen Besonderheiten. Werden diese nicht beachtet, so führt dies üblicherweise zu Singularitäten und damit zum Abbruch des Schätzprogramms.

Am wichtigsten in diesem Zusammenhang ist die Tatsache, daß in die Berechnung der Auswahlwahrscheinlichkeiten von diskreten Entscheidungsmodellen nur die Differenzen der erklärenden Variablen eingehen. Für Variable, deren Werte über die Alternativen nicht variieren, sind die Differenzen immer Null und für sie können daher keine Parameter berechnet werden. Dies betrifft vor allem sozioökonomische Variable aber auch eine allgemeine Konstante, wie sie aus der Regressionsanalyse bekannt ist.

Spezifiziert man beispielsweise in einem binären Modell der Verkehrsmittelwahl für die beiden Alternativen "Auto" und "Bus" folgende Datenstruktur

Alternative	β_o	β_1	β_2	β_3
Auto	1	FZ_A	FK_A	Eink
Bus	1	FZ_B	FK_B	Eink

(27)

FZ ... Fahrzeit, FK ... Fahrtkosten, Eink ... Einkommen.

so sind die beiden Parameter β_0 und β_3 nicht bestimmbar und das Schätzprogramm wird keine Ergebnisse liefern. Diese Struktur ist gleichbedeutend mit der Hypothese, daß die Auswahlwahrscheinlichkeiten beider Alternativen auf Einkommensänderungen und auf Änderungen der Konstanten gleichartig reagieren. Da die Summe der beiden Auswahlwahrscheinlichkeiten immer Eins sein muß, ist diese Hypothese sinnlos. Hinter dieser Beschränkung der Datenstruktur stehen also nicht nur statistische sondern auch inhaltliche Gründe.

Eine sinnvolle und schätzbare Variante der Datenstruktur (27) wäre etwa die Folgende

Alternative	β_o	β_1	β_2	β_3
Auto	1	FZ_A	FK_A	Eink
Bus	Ø	FZ_B	FK_B	Ø

(28)

In diesem Fall stellt β_0 eine alternativenspezifische Konstante dar, die eine autonome Präferenz für oder (bei negativem β_0) gegen das Auto mißt. Der Parameter β_3 berücksichtigt die Tatsache, daß Personen höherer Einkommensschichten eher mit dem Auto als mit dem Bus fahren.

Alternativenspezifische Konstante können zwar die Anpassungsqualität des Modells oft erheblich verbessern, inhaltlich stellen sie aber eine Restgröße dar, in der sich nicht spezifizierte Einflüsse auswirken (etwa größere Bequemlichkeit des Autos). Um nicht nur die statistische sondern auch die inhaltliche Qualität des Modells zu erhöhen, sollten diese Einflüsse besser explizit in die Spezifikation des Modells aufgenommen werden. Im Zusammenhang mit alternativenspezifischen Konstanten und alternativenspezifischen sozioökonomischen Variablen ist zu beachten, daß sie nicht für alle I, sondern nur für I-1 Alternativen spezifiziert werden dürfen, da sie sonst gleichbedeutend mit allgemeinen Konstanten bzw. allgemeinen sozioökonomischen Variablen sind und zu den bereits diskutierten Problemen führen.

Die beiden Variablen Fahrzeit und Fahrtkosten in (27) und (28) stellen allgemeine ('generic') Variable dar, die für alle Alternativen gemessene Werte annehmen. Sie können nur deshalb so in die Schätzung aufgenommen werden, weil sie - zumindest für manche Individuen - über die Alternativen variieren. In der Spezifikation (27) bzw. (28) wird allerdings a-priori festgelegt, daß Bus-Fahrzeit und Auto-Fahrzeit bzw. Bus-Fahrtkosten und Auto-Fahrtkosten gleich bewertet werden. Diese a-priori Restriktion kann durch eine Datenstruktur von folgender Form vermieden werden.

Alternative	β_0	β_1	β_2	β_3	β_4	β_5	
Auto	1	FZ_A	Ø	FK_A	Ø	Eink	(29)
Bus	Ø	Ø	FZ_B	Ø	FK_B	Ø	

Sie erlaubt die Schätzung unterschiedlicher Parameterwerte für Auto- und Bus-Fahrzeit und -Kosten. Mit Hilfe statistischer Tests kann festgestellt werden, ob die in (28) enthaltene Hypothese zutrifft oder nicht.

Im Zusammenhang mit dem Logit-Modell wird oft zwischen dem in Abschnitt 2.4.4 dargestellten multinomialen Logit und einem als "conditional Logit" bezeichneten Modell unterschieden. Letzteres ist dadurch gekennzeichnet, daß die Parameter über die Alternativen variieren. Gleichung (12) wird damit zu

$$V_{in} = X_{in}\beta_i \tag{30}$$

Tatsächlich sind diese beiden Modelle identisch, sie können durch entsprechende Definition der Datenstruktur ineinander übergeführt und daher mit den selben Programmen geschätzt werden. Variieren die erklärenden Variablen über die Alternativen, so erfordert die Schätzung von (30) die in (29) für Fahrzeit und Fahrtkosten angesetzte Datenstruktur. Variieren die erklärenden Variablen nicht über die Alternativen, so ist zu beachten, daß wiederum nicht der volle Satz an Parametern identifiziert werden kann. In diesem Fall muß die Datenstruktur etwa folgendes Aussehen haben

Alt.	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8
1	X_1	X_2	X_3	X_4	Ø	Ø	Ø	Ø
2	Ø	Ø	Ø	Ø	X_1	X_2	X_3	X_4
3	Ø	Ø	Ø	Ø	Ø	Ø	Ø	Ø

(31)

Im binären Fall reduziert sich diese Datenstruktur auf

Alternative	β_1	β_2	β_3	β_4
1	X_1	X_2	X_3	X_4
2	Ø	Ø	Ø	Ø

(32)

Auf diese Datenstruktur werden wir bei der Beschreibung der SAS Prozedur BPROBIT zurückkommen.

Bei der Verwendung von SAS müssen diese Datenstrukturen in den Rohdaten noch nicht vollständig enthalten sein. Sie können durch Datenmanipulation im DATA STEP konstruiert werden. Für diesen Zweck ist es notwendig bei der Datenerfassung Alternativen und Personen durch Nummern zu kodieren. Außerdem erfordert jedes Schätzprogramm neben den erklärenden Variablen bzw. Rohdaten aus denen die erklärenden Variablen konstruiert werden können auch noch Informationen über den Vektor Y, also darüber, welche Alternative gewählt wurde und welche nicht.

2.7. Statisktiken und Tests

Bei den im Rahmen von diskreten Entscheidungsmodellen verwendbaren Statistiken und Tests können zwei Gruppen unterschieden werden: Einerseits solche, die für jede Maximum-Likelihood-Schätzung verwendet werden können, andererseits für diskrete Entscheidungsmodelle spezifische. Zur ersten Gruppe zählen etwa Signifikanztests und Statistiken über die Anpassungsgüte, zur zweiten beispielsweise der Test der Verletzung der IIA-Eigenschaft im Logit-Modell.

In dieser Arbeit können wir natürlich nur eine sehr oberflächliche Darstellung des statistischen Hypothesentestens geben. Ausführlichere Zusammenfassungen sind etwa in Blake (1979), Wonnacott und Wonnacott (1979), Judge, Griffiths, Hill und Lee (1980) zu finden. Im Zusammenhang mit diskreten Entscheidungsmodellen finden sich Darstellungen in Domencich und McFadden (1975), Hensher und Johnson (1981), Maddala (1983), Ben-Akiva und Lerman (1985).

2.7.1. Der asymptotische t-Test

Wie bereits erwähnt sind die Maximum-Likelihood-Schätzer asymptotisch normalverteilt. Die am Maximum der Likelihoodfunktion berechnete negative Inverse der Matrix der zweiten Ableitungen gibt einen Schätzwert für die Varianz-Kovarianz-Matrix dieser Schätzer.

In Analogie zur t-Statistik des linearen Regressionsmodells können auch für diskrete Entscheidungsmodelle t-Werte für die einzelnen Parameter berechnet werden. Diese Statistiken geben an, ob der für einen Parameter geschätzte Wert signifikant von einem vorgegebenen - meist Null - abweicht. Im Unterschied zum linearen Regressionsmodell gilt die Verteilung der Maximum-Likelihood-Schätzer nicht exakt sondern nur asymptotisch. Daher sind als kritische Werte für die Teststatistik Perzentile der standardisierten Normalverteilung zu verwenden.

Bezeichnet man mit $\hat{b}_k$ den geschätzten Parameter, mit s_k seinen Standardfehler und mit b_k den Wert gegen den er getestet werden soll, so errechnet sich die t-Statistik für diesen Parameter als

$$t_k = \frac{\hat{b}_k - b_k}{s_k} \quad . \qquad (33)$$

Ist dieser Wert größer als das Perzentil der standardisierten Normalverteilung bei der entsprechenden Irrtumswahrscheinlichkeit, so wird die Hypothese, daß b_k der wahre Wert von $\hat{b}_k$ ist, verworfen, sonst angenommen. Für eine Irrtumswahrscheinlichkeit von 5 % liegt die kritische Grenze für diese Statistik bei 1.96.

2.7.2. Der Likelihood-Ratio-Test

Ein sehr fexibel einsetzbares Instrument ist der Likelihood-Ratio-Test. Er kann sowohl als Signifikanztest für einen oder mehrere Parameter, als auch als Anpassungstest ("Goodness of fit"-Test) verwendet werden.

Ausgangspunkt für den Likelihood-Ratio-Test sind zwei Modelle, ein "vollständiges" und ein "restringiertes", wobei sich das restringierte Modell aus dem vollständigen durch die zu testende Hypothese ergibt. Eine mögliche Hypothese ist jene, die bereits im Zusammenhang mit dem asymptotischen t-Test verwendet wurde, nämlich, daß b_k der wahre Wert des Parameters $\hat{b}_k$ ist. Eine andere mögliche Hypothese wäre, daß alle Parameter eigentlich Null sind. Unter dieser Hypothese wird der Likelihood-Ratio-Test als Anpassungstest verwendet.

Bezeichnet man den Likelihood Wert des vollständigen (unrestringierten) Modells als $L(\hat{b}_V)$ und den des restringierten Modells als $L(\hat{b}_R)$, so ist der Likelihood-Ratio-Wert definiert als

$$LR = -2(L(\hat{b}_R) - L(\hat{b}_V)) \qquad (34)$$

Bei Zutreffen der Nullhypothese ist LR Chi-Quadrat verteilt mit R Freiheitsgraden, wobei R die Zahl der durch die Hypothese restringierten Parameter ist. Aus der Tabelle der Chi-Quadrat Verteilung kann wiederum der kritische Wert für die akzeptierte Irrtumswahrscheinlichkeit und R Freiheitsgrade ermittelt werden. Ist LR größer, so wird die Hypothese verworfen, ist er kleiner, wird sie angenommen.

Die beiden SAS Prodzeduren liefern automatisch einen LR-Wert, bei dem als restringiertes Modell die Anfangswerte der zu schätzenden Parameter verwendet werden. Gibt der Benutzer keine Anfangswerte vor (siehe Abschnitt 3.3, Option BSTART), so sind alle Anfangswerte Null und der LR-Wert gibt einen Indikator für die Anpassungsgüte. Für den Signifikanztest eines Parameters kann wahlweise der t-Test und der Likelihood-Ratio-Test verwendet werden. Beide liefern die selben Ergebnisse bezüglich Annahme oder Ablehnung der Hypothese. Da der Likelihood-Ratio-Test in dieser Anwendung eine zweite, restringierte Schätzung verlangt, die t-Werte aber in einem Schätzlauf ermittelt werden können, wird für Signifikanztests üblicherweise der t-Statistik der Vorrang gegeben.

Für den Test komplizierterer Hypthesen stellt der Likelihood-Ratio-Test ein flexibles Instrument dar. Beispielsweise um die Hypothese von alternativenspezifischen Unterschieden in den Parametern allgemeiner Variabler zu überprüfen (siehe Abschnitt 2.6). Um das Modell (28) gegen das allgemeinere (29) zu testen, muß einfach der maximale Log-Likelihood-Wert von Modell (28) für $L(\hat{b}_R)$, jener von Modell (29) für $L(\hat{b}V)$ in (34) eingesetzt werden und das Ergebnis mit dem Tabellenwert der Chi-Quadrat-Verteilung mit zwei Freiheitsgraden bei der gewählten Irrtumswahrscheinlichkeit verglichen werden. Wesentlich ist, daß das restringierte Modell eine durch die Hypothese beschränkte Version des vollständigen Modells sein muß.

2.7.3. Maße für die Anpassungsgüte

Im linearen Regressionsmodell ist das Bestimmtheitsmaß (R^2) eine aussagekräftige Maßzahl für die Anpassungsgüte der Schätzung. Für nichtlineare Modelle wie das Logit- und Probit-Modell läßt sich ein ähnlicher Indikator (Rho-Quadrat) konstruieren, der allerdings nicht wie das Bestimmtheitsmaß als "Prozent erklärt" interpretiert werden kann. Rho-Quadrat ist definiert als

$$\rho^2 = 1 - \frac{L(\hat{b})}{L(0)} \qquad (35)$$

wobei L(0) die Log-Likelihood bezeichnet, wo alle Parameter Null sind.

Rho-Quadrat berücksichtigt nicht den Verlust an Freiheitsgraden, der mit der Aufnahme weiterer erklärender Variabler in die Schätzung verbunden ist. Diesen Mangel korrigiert folgender Indikator

$$\bar{\rho}^2 = 1 - \frac{L(\hat{b}) - K}{L(0)} \qquad (36)$$

Er wird üblicherweide korrigiertes Rho-Quadrat genannt. Beide Maßzahlen werden von den SAS Prozeduren BPROBIT und MNLOGIT ausgegeben. Allerdings gibt es bei nichtlinearen Modellen keine statistischen Anhaltspunkt dafür, welche Werte von Rho-Quadrat als genügend groß für einen ausreichenden Erklärungsbeitrag des

Modells angesehen werden können. Für den relativen Vergleich verschiedener Spezifikationen eines Modells bietet das (korrigierte) Rho-Quadrat Maß jedoch einen wertvollen Anhaltspunkt.

2.7.4. Tests der IIA-Eigenschaft

Wie in Abschnitt 2.4 dargestellt wurde, impliziert die formale Struktur des Logit-Modells die IIA-Eigenschaft. Sie ist das Ergebnis der Annahme unabhängig identisch verteilter Zufallsterme im Zufallsnutzen. Ob diese Annahme zutrifft hängt einerseits von der Struktur des Entscheidungsproblems, andererseits aber auch von der Struktur des deterministischen Nutzenteils ab, da diese implizit bestimmt, welche Einflüsse sich im Zufallsterm auswirken. Tests der IIA-Eigenschaft sind damit auch immer Tests über die Struktur des deterministischen Nutzens und eine Verletzung der IIA-Eigenschaft weist auch immer auf Fehler in der Modellspezifikation hin.

Da die IIA-Eigenschaft bedeutet, daß das Verhältnis zweier Auswahlwahrscheinlichkeiten unabhängig vom Vorhandensein anderer Alternativen ist, können bei Gültigkeit von IIA Alternativen aus der Alternativenmenge eliminiert werden und die Schätzung liefert dennoch statistisch die gleichen Parameterwerte. Dieses Prinzip verwenden die IIA-Tests.

Wir bezeichnen mit A^* eine Teilmenge der Alternativenmenge A und schätzen das zu testende Logit-Modell mit beiden Alternativenmengen. Durch die Beschränkung auf A^* werden nicht nur für jedes Individuum eine oder mehrere Alternativen eliminiert, sondern auch jene Individuen, die eine der eliminierten Alternativen gewählt haben. Außerdem können mit der restringierten Alternativenmenge die Parameter für die alternativenspezifischen Variablen der eliminierten Alternativen nicht geschätzt werden. Die beiden Schätzungen liefern daher Parametervektoren von unterschiedlicher Länge.

Wir bezeichnen mit $\hat{b}_{A^*}$ die Ergebnisse der Schätzung mit der restringierten Alternativenmenge und mit $\hat{b}_A$ jenen Teil von $\hat{b}$, der auch in $\hat{b}_{A^*}$ enthalten ist. Die dazugehörenden Varianz-Kovarianz-Matrizen seien Σ_{A^*} und Σ_A, wobei letztere wiederum eine Teilmatrix der bei der Schätzung erhaltenen Matrix ist.

Hausman und McFadden (1985) zeigen, daß unter der Hypothese $b_A* = b_A$ die Statistik

$$(\hat{b}_A* - \hat{b}_A)' \ (\Sigma_A* - \Sigma_A)^{-1} \ (\hat{b}_A* - \hat{b}_A) \tag{37}$$

asymptotisch Chi-Quadrat verteilt ist mit K^* Freiheitsgraden. K^* ist die Anzahl der Elemente in $\hat{b}_A*$.

Ein auf dem Likelihood-Ratio-Konzept beruhender Test der IIA-Eigenschaft geht zurück auf McFadden, Tye und Train (1977) und Small und Hsiao (1982) (siehe auch Ben-Akiva und Lerman, 1985). Diese Statistik verwendet die über die restringierte Alternativenmenge berechneten Log-Likelihood-Werte bei $\hat{b}_A*$ und $\hat{b}_A$ $(L_A*(\hat{b}_A*),\ L_A*(\hat{b}_A))$. Sie lautet

$$-2\ C\ (L_A*(\hat{b}_A) - L_A*(\hat{b}_A*)) \qquad \text{mit} \quad C = \frac{1}{1 - N_A*/\alpha N_A} \tag{38}$$

N_A und $N_A{}^*$ bezeichnen die Anzahl der Individuen bzw. Gruppen in den beiden Datensätzen. α ergibt sich aus dem Vergleich der beiden Varianz-Kovarianz-Matrizen, Small und Hsiao schlagen daür einen Wert von Eins vor. Die Statistik (38) ist wiederum asymptotisch Chi-Quadrat verteilt mit K^* Freiheitsgraden. Für ein Anwendungsbespiel dieser Teststatistik siehe Ben-Akiva und Lerman, 1985, S. 186 ff. Da die SAS Prozedur MNLOGIT eine sehr einfache Möglichkeit bietet, die geschätzten Parameter und Varianz-Kovarianz-Matrizen auf SAS Dateien auszugeben, und SAS mit der Prozedur MATRIX ein Instrument für Matrixoperationen zur Verfügung stellt, erscheint die Statistik (37) im Zusammenhang mit MNLOGIT besser geeignet zu sein, die IIA-Eigenschaft zu testen.

3. Die Verwendung der SAS Prodzeduren BPROBIT und MNLOGIT

Die beiden SAS Prodzeduren BPROBIT und MNLOGIT dienen zur Schätzung binärer Probit- und multinomialer Logit-Modelle. Der Vorteil von SAS Prozeduren gegenüber Stand-Alone-Programmen liegt darin, daß zur Vorbereitung der Schätzung die mächtigen Datenmanipulationsmöglichkeiten und Prozeduren von SAS zur Verfügung stehen. BPROBIT und MNLOGIT wurden möglichst allgemein und einfach ausgelegt. Modifikationen der Datenstruktur (siehe Abschnitt 2.6), die genaue Formulierung von Restriktionen etc. muß der Benutzer durch Datenmanipulationen im DATA STEP von SAS ausführen. Damit soll einerseits erreicht werden, daß der Benutzer den Schätzvorgang versteht und die Prozeduren nicht als "Black Box" verwendet, andererseits, daß er durch Ausformulieren und Programmieren von Hypothesen und nicht durch Probieren von verschiedenen Modellversionen zu Ergebnissen gelangt. Dies bedeutet aber nicht, daß der Benutzer für die erste Anwendung von BPROBIT und MNLOGIT irgendwelche umfangreiche Vorbereitungen treffen muß. Er sollte nur sicherstellen, daß die in Abschnitt 2.6 diskutierten Anforderungen an die Datenstruktur zutreffen.

Da die beiden Prozeduren BPROBIT und MNLOGIT für den Benutzer sehr ähnlich sind, wird ihre Anwendung im Folgenden gemeinsam beschrieben. Die Beschreibung gliedert sich in die Abschnitte "Aufruf und Variablenübergabe" (3.1), "Ausgabe der Ergebnisse" (3.2), "Optionen" (3.3) und "Parameter" (3.4).

Dabei wird die folgende Syntax verwendet:

Großschreibung unterstrichen ... Diese Teile müssen exakt in der vorgegebenen Art und Weise eingegeben werden.

Großschreibung nicht unterstrichen ... Ist ein Teil des Wortes unterstrichen, ein Teil nicht, so kann der gesamte Aufruf durch den unterstrichenen Teil abgekürzt werden (Beispiel: VARIABLES). Ist das gesamte Wort nicht unterstrichen, so bezeichnet es Teile der Ausgabe der Prozeduren (Beispiel: RHO-QUADRAT).

Groß-Kleinschreibung ... Diese Teile müssen vom Benutzer spezifiziert werden (Beispiel: Optionen Parameter). Unterstrichene Teile sind wiederum unbedingt notwendig.

Strichpunkt (;) ... Dies ist das Endezeichen für SAS Befehle. Es ist unbedingt notwendig.

Genauere Informationen über die Syntax von SAS Befehlen, insbesonders über die Datenmanipulationsbefehle des DATA STEP sind den SAS Publikationen "SAS Introductory Guide" und "SAS User's Guide" zu entnehmen.

3.1. Aufruf und Variablenübergabe

Die beiden Prozeduren BPROBIT und MNLOGIT werden jeweils durch PROC und die Angabe ihres Namens aufgerufen. Auf den Namen der Prozedur folgt die optionale Angabe von Dateinamen, Optionen und Parametern.

Für BPROBIT lautet der Aufruf

PROC BPROBIT Dateien Optionen Parameter;

für MNLOGIT

PROC MNLOGIT Dateien Optionen Parameter;

Als Dateien können sowohl die Eingabedatei als auch Ausgabedateien für die Parameterschätzungen, die Varianz-Kovarianz-Matrix und die geschätzten Wahrscheinlichkeiten und bei MNLOGIT die Inklusivwerte angegeben werden. Für alle diese Dateien werden, wenn nicht explizit spezifiziert, Standardnamen verwendet. Standardeingabedatei etwa ist die zuletzt angelegte SAS Datei. Soll eine andere SAS Datei verwendet werden, so kann diese durch den Befehl

DATA = Name der Eingabedatei

festgelegt werden. Die Spezifikation der Ausgabedateien wird bei den dazugehörenden Optionen erläutert werden.

Ein zweiter Befehl dient der Übergabe der Variablen. Er beginnt mit der Listenbezeichnung VARIABLES. Da die beiden Prozeduren keine "MODEL" Statement kennen, verlangen sie die Übergabe der Variablen in einer vorgegebenen Reihenfolge. Für jede Alternative in der Alternativenmenge des Individuums (einer Gruppe von Individuen) muß ein Datenrecord übergeben werden. Außerdem müssen jene Datenrecords, die die Alternativen eines Individuums beschreiben, direkt aufeinander folgen. Die Prozeduren erwarten also folgende Datenorganisation:

Individuum 1 : Alternative 1
Individuum 1 : Alternative 2
:
:
Individuum 1 : Alternative I
Individuum 2 : Alternative 1
:
:

Sind die Rohdaten nicht in dieser Reihenfolge, so müssen sie durch die SAS Prozedur SORT umsortiert werden. Für die Prozedur BPROBIT ist I definitionsgemäß gleich 2. Eine mögliche Abweichung von dieser Regel im Zusammenhang mit BPROBIT und Datenstruktur (32) wird in Abschnitt 3.3 beschrieben werden. Für die Funktionsweise von MNLOGIT hat die Zahl und Reihenfolge der Alternativen eines Individuums keine Bedeutung. "Alternative 1" darf also für Indivduum 1 etwas anderes beschreiben als für Individuum 2.

Das VARIABLES Statement lautet für BPROBIT

VARIABLES Choice Var1 Var2 Var3 ... ;

für MNLOGIT

VARIABLES Indiv Choice Var1 Var2 Var3, ... ;

Var1, Var2, Var3 usw. bezeichnen die in der Schätzung zu verwendenden erklärenden Variablen. Für jede dieser Variablen wird ein Parameter mittels Maximum-Likelihood geschätzt. Die Zahl der erklärenden Variablen ist von der Programmstruktur her nicht beschränkt, ebensowenig wie die Zahl der Individuen und die Zahl der Alternativen. Lediglich der Umfang der zur Verfügung stehen-den Informationen (Zahl der Freiheitsgrade) und die Speicherkapazität der verwendeten Anlage beschränken die Schätzung. Selbstverständlich ist mindestens eine erklärende Variable anzuführen.

Für Choice ist jene Variable einzusetzten, die angibt, ob diese Alternative gewählt wurde oder nicht. Im ersten Fall ist Choice Eins, im zweiten Null. Bei gruppierten Daten sagt Choice den Prozeduren, wie oft diese Alternaitve gewählt wurde.

MNLOGIT verlangt außerdem noch die Angabe einer Variablen Indiv, die angibt, zu welchem Individuum (bzw. Gruppe von Individuen bei aggregierten Daten) diese Alternative gehört. Bei BPROBIT ist diese Information nicht notwendig da pro Individuum genau zwei Alternativen vorliegen müssen. MNLOGIT verwendet die Variable Indiv nur um den Wechsel zu einem anderen Indivduum (gekennzeichnet durch einen anderen Wert von Indiv) festzustellen. Folgen die Werte von Indiv nicht in aufsteigender Reihenfolge, so gibt MNLOGIT eine Warnung aus. Ebenso, wenn für irgendein Individuum nur eine für die Schätzung verwendbare Alternative gefunden wird. Der Programmablauf wird jedoch fortgesetzt. Derartige Warnungen entstehen auch, wenn Indiv vergessen oder ein falscher Variablenname angegeben wird.

Fehlende Werte in der Eingabedatei (gekennzeichnet mit '.') werden standardmäßig so behandelt, daß das entsprechende Datenrecord für die Schätzung nicht verwendet wird. Im binären Fall ist dies gleichbedeutend mit der Elimination des Indivduums. Die Option MGES erlaubt auch im multinomialen Fall bei einem fehlenden Wert die Elimination der gesamten Information für dieses Individuum.

Ein dritter, wahlweise verwendbarer Befehl, das ID Statement, dient der Übergabe von Variablen in die Ausgabedatei für die geschätzten Wahrscheinlichkeiten (siehe auch Option SPROB, unten). Er lautet

ID IdVar1 IdVar2 IdVar3 ... ;

Damit können aus der Eingabedatei Variable, die die Datenrecords identifizieren, in diese Ausgabedatei übertragen werden. Dies ist dann von Bedeutung, wenn die Datenrecords der Ausgabedatei zur weiteren Bearbeitung in eine bestimmte, von jener der Eingabedatei abweichende Reihenfolge gebracht werden müssen.

3.2. Ausgabe der Ergebnisse

Die Ausgabe von Ergebnissen erfolgt über die LISTING Datei, Warnungen und Fehlermeldungen finden sich in der SASLOG Datei.

Als erste Information nach eventuellen Warnungen geben die Prozeduren die **ANFANGSWERTE FUER BETA** aus. Wenn vom Benutzer nicht explizit - über die Option BSTART - oder implizit - über eine Restriktion - anders bestimmt, starten die Prozeduren immer bei Anfangswerten von Null. Bei diesen Werten sind für jedes Individuum alle Alternativen gleich wahrscheinlich. Die ANFANGSWERTE FUER BETA sind deshalb wichtig zu wissen, da für diese Werte die LOG LIKELIHOOD-ANFANG und damit auch der Likelihood-Ratio-Wert und die RHO-QUADRAT-Statistiken berechnet werden.

Für jede Iteration geben die Prozeduren als Zwischenergebnisse **DELTA BETA** und den Wert der **LOG-LIKELIHOOD** an. DELTA BETA gibt die durchschnittliche Veränderung der Schätzwerte zwischen zwei Iterationen an und es ist jener Wert, der von den Prozeduren mit dem Konvergenzkriterium verglichen wird um festzustellen, ob das Maximum der Log-Likelihoodfunktion erreicht wurde. Das Konvergenzkriterium kann über den Parameter LIMIT eingestellt werden. Erreicht das Schätzprogramm einen Wert von DELTA BETA kleiner als das Konvergenzkriterium, so rechnet es noch eine Iteration, um auch die ersten und zweiten Ableitungen für diesen Parametervektor zu errechnen.

Bei einem Programmabbruch mit FORTRAN Fehlermeldung geben die Zwischenergebnisse für DELTA BETA und die LOG-LIKELIHOOD einen Hinweis auf die Ursache und helfen die notwendigen Korrekturmaßnahmen zu finden. Ein Beispiel für die Ausgabe der Zwischenergebnisse einer Schätzung mit MNLOGIT zeigt Darstellung 2 im Anhang.

Beendet das Schätzprogramm die Iterationen, so beginnt es eine neue Seite und gibt an, ob und nach wieviel Iterationen es **KONVERGENZ** erreicht hat (DELTA BETA unterschreitet das Konvergenzkriterium), oder ob es **KEINE KONVERGENZ** erreicht hat (Maximale Zahl der Iterationen erreicht).

Anschließend beginnt die Ausgabe der endgültigen Ergebnisse. Als erstes listen beide Prozeduren den Namen der abhängigen Variablen, MNLOGIT auch den Namen der Gruppierungsvariablen. Für jeden geschätzten Parameter gibt das Schätzprogramm den Namen der Variablen, **BETA**, **T-WERT**, **STEIGUNG** der Log-Likelihoodfunktion und

Standardfehler (**STD.FEHLER**) aus. Für Berechnung und Interpretation von Standardfehler und t-Wert siehe Abschnitt 2.7.

Im Anschluß daran geben beide Prozeduren Informationen über:

ANZAHL GRUPPEN, dies ist die Zahl der Individuen bzw. Gruppen, die unterschieden werden konnten,

ANZAHL BEOB., dies ist die Zahl der gelesenen Datenrecords,

ANZAHL FAELLE, dies ist die Zahl der Freiheitsgrade des Datensatzes. Da sich die Auswahlwahrscheinlichkeiten immer auf Eins summieren müssen, ist dies die Anzahl der Beobachtungen minus der Anzahl der Gruppen.

FEHLENDE WERTE, dies ist die Zahl der auf Grund fehlender Werte eliminierten Datenrecords.

MNLOGIT gibt außerdem noch die maximale Anzahl an Alternativen (**MAX.ANZ.ALT.**) eines Indivduums bzw. einer Gruppe an.

Diese Informationen geben einen Überblick darüber, ob die Daten in der gewünschten Art verarbeitet wurden. Fehler, bei MNLOGIT etwa in der Gruppierungsvariablen, sollten sich in diesen Werten auswirken. Allerdings muß davor gewarnt werden, diese Werte als einzige Grundlage für die Überprüfung zu verwenden. Eine detailierte Überprüfung des Datensatzes kann nur durch die Ausgabe der SAS Datei mit PROC PRINT erfolgen.

Nächste Information, die die Prozeduren ausgeben, sind die Werte der Log-Likelihoodfunktion am **ANFANG** der Schätzung, also für die Anfangswerte von Beta, und für die Beta-Werte am **ENDE** der Schätzung. Unter **LR-TEST** (Likelihood-Ratio-Test) wird der Wert der Statisktik (34) ausgegeben. Außerdem wird angegeben, wie viele nichtrestringierte erklärende Variable in der Schätzung verwendet wurden. Erreicht das Schätzprogramm allerdings keine Konvergenz, so ist der Log-Likelihood-Wert am Ende der Schätzung nicht der maximale und die Likelihood-Ratio-Statistik liefert keinen sehr sinnvollen Wert.

Als weiteren Indikator für die Qualität der Schätzung liefern die Prozeduren die Werte **RHO-QUADRAT** (35) und **RHO-QU.KORR.**, das um die Zahl der erklärenden Variablen korrigierte Rho-Quadrat (36). Im Falle von Nichtkonvergenz gilt für diese beiden Indikatoren ähnliches wie für die Likelihood-Ratio-Statistik. Bei gruppierten Daten ergeben sich üblicherweise Probleme mit den Rho-Quadrat-

Statistiken. Da die Prozeduren BPROBIT und MNLOGIT die Log-Likelihoodwerte vollständig und nicht verkürzt berechnen, geben sie sowohl bei Individualdaten als auch bei aggregierten Daten korrekte Rho-Quadrat-Statistiken aus. Darstellung 3 im Anhang zeigt ein Beispiel für die Ausgabe der Schätzergebnisse von MNLOGIT.

3.3. Optionen

Beide Prozeduren verfügen über die Optionen PROB, COVAR, BSTART, EST, SPROB und SCOV. BPROBIT kennt außerdem noch die Option EINZEIL, MNLOGIT die Option MGES. Sie sollen im Folgenden beschrieben werden.

PROB: Diese Option führt zur Ausgabe der beobachteten und der geschätzten Häufigkeiten und ihrer Differenz, den Residuen in die Listing Datei. Erstere sind die Werte der Variablen Choice, zweitere die geschätzten Auswahlwahrscheinlichkeiten multipliziert mit der Anzahl der Individuen in der Gruppe (Eins im Fall von Individualdaten). In einem einfachen Plot-Diagramm werden die Residuen dargestellt. Die Skala dieses Diagramms wird durch die maximale Anzahl von Individuen in einer Gruppe bestimmt.

Bei der Prozedur MNLOGIT gibt die Option PROB auch noch Inklusivwerte aus. Dies ist der Logarithmus des Nenners von (17) und er dient zur Verbindung der einzelnen Schätzungen in "Nested Logit"-Modellen (Näheres siehe Hensher und Johnson, 1981, Maier und Fischer, 1985, Ben-Akiva und Lerman, 1985). Darstellung 4 im Anhang gibt eine Beispiel für die Ausgabe der Option PROB in MNLOGIT.

COVAR: Diese Option führt zur Ausgabe der Varianz-Kovarianz-Matrix in die Listing Datei. Sie gibt Aufschluß über Interdependenzen zwischen Parametern und wird für manche Tests benötigt. Darstellung 5 im Anhang gibt ein Beispiel für MNLOGIT.

BSTART: Diese Option sagt den Prozeduren, daß nicht Null als Startwert für die Schätzung verwendet werden soll, sondern Startwerte von einer Datei eingelesen werden sollen. Diese Datei muß dem SAS Programm mittels des CMS Statements (siehe Arbeitsbuch für das System IBM 4331 unter VM/SP-CMS)

CMS FILEDEF 19 DISK fn ft fm;

zur Verfügung gestellt worden sein. Wird die Option BSTART verwendet ohne daß eine Datei zur Verfügung steht, so führt dies zum Programmabbruch.

Darstellung 6 im Anhang zeigt ein Beispielprogramm und die verwendete Datei.

EINZEIL: Diese Option steht nur für BPROBIT zur Verfügung. Sie wird dann verwendet, wenn ein Modell mit der Datenstruktur (32) geschätzt werden soll, im Datensatz aber nur Informationen über jene Alternative enthalten sind, deren erklärende Variable ungleich Null sind. In diesem Fall wird nur ein Datenrecord pro Individuum erwartet, das zweite Datenrecord wird von BPROBIT intern generiert. Im Zusammenhang mit gruppierten Daten ist diese Option nicht sinnvoll, da in diesem Fall die zweite Zeile von (32) zusätzliche Informationen enthält.

MGES: Diese Option steht nur für MNLOGIT zur Verfügung. Sie führt zum Ausschluß aller Alternativen des Individuums (der Gruppe), wenn in einem der Datenrecords ein fehlender Wert entdeckt wird.

EST: Diese Option führt zur Ausgabe der geschätzten Parameterwerte in eine SAS Datei. Diese stehen damit direkt für die weitere Bearbeitung mit SAS zur Verfügung. Die Option EST gibt zwei Variable aus, nämlich 'NAME', den Variablennamen des geschätzten Parameters, und 'EST', den geschätzten Parameterwert. Die Ausgabedatei hat standardmäßig den Namen 'EST', er kann jedoch durch Angabe von

EST = Dateiname

abgeändert werden. Eine PROC PRINT Ausgabe einer EST Datei zeigt Darstellung 7 im Anhang.

SPROB: Diese Option führt zur Ausgabe der beobachteten und geschätzten Häufigkeiten und bei MNLOGIT der Inklusivwerte in eine SAS Datei. Diese Variablen haben die Namen 'PROGN' (geschätzte Häufigkeit) und 'INCL' (Inklusivwert). Für die beobachtete Häufigkeit und eventuell mit einem ID Statement übergebene Variable wird ihr Name in der Eingabedatei verwendet. Die Ausgabedatei hat standardmäßig den Namen 'RESID', er kann durch Angabe von

RESID = Dateiname

abgeändert werden. Eine PROC PRINT Ausgabe einer RESID Datei zeigt Darstellung 8 im Anhang.

SCOV: Diese Option führt zur Ausgabe der Varianz-Kovarianz-Matrix in eine SAS Datei. Diese Datei hat eine Variable 'NAME', die den Namen des entsprechenden Parameters angibt, und Variable mit den Namen der entsprechenden Parameter. Standardname der Ausgabedatei ist 'COV', er kann durch Angabe von

COV = Dateiname

abgeändert werden. Eine PROC PRINT Ausgabe einer COV Datei zeigt Darstellung 9 im Anhang.

3.4. Parameter

In den Parametern bestehen keine Unterschiede zwischen BPROBIT und MNLOGIT. Beide Prozeduren kennen die Parameter MAXIT, LIMIT, INVERT, STEP und RESTRICT. Alle diese Parameter verlangen numerische Parameterwerte und werden in folgender Form verwendet.

Parametername = Parameterwert

MAXIT: Dieser Parameter dient dazu, die maximale Anzahl von Iterationen zu fixieren. Sein Standardwert ist 20.

Beispiel: MAXIT = 10

LIMIT: Dieser Parameter setzt das Konvergenzkriterium mit dem DELTA BETA verglichen wird. Standardwert dieses Parameters ist 0.0001.

Beispiel: LIMIT = 0.00001

INVERT: Dieser Parameter setzt eine Grenze, die im Zuge der Invertierung der Matrix der zweiten Ableitung überprüft wird. Unterschreitet der Absolutbetrag des Nenners einer bei der Invertierung durchzuführenden Division diese Grenze, so bricht das Schätzprogramm mit der Fehlermeldung **INVERTIERUNGSGRENZE UNTERSCHRITTEN** ab. Diese Prüfung verhindert einen unkontrollierten Programmabbruch mit FORTRAN Fehlermeldung. Das Unterschreiten der Invertierungsgrenze deutet auf einen Fehler in der Datenstruktur hin (siehe Abschnitt 2.6), etwa auf nicht alternativenspezifisch verwendete sozioökonomische Variable. Standardwert der Invertierungsgrenze ist 0.0001.

Beispiel: INVERT = 0.002

STEP:Dieser Parameter setzt eine maximale Schrittweite für den Newton-Raphson-Algorithmus. Überschreitet DELTA BETA die maximale Schrittweite, so reduziert das Programm die Parameteränderungen derart, daß DELTA BETA gleich der maximalen Schrittweite wird. Die berechnete Richtung der Parameteränderung wird beibehalten.

Der STEP Parameter ist für jene Fälle interessant, in denen der Newton-Raphson-Algorithmus über das Ziel hinaus schießt und zu schlechteren Schätzwerten und rechentechnischen Problemen führt (siehe Abschnitt 2.5.2). Durch Angabe des STEP Parameters kann dies verhindert werden. Standardmäßig ist die maximale Schrittweite auf den sehr großen Wert 7.E+75 gesetzt.

Beispiel: STEP = 5.5

RESTRICT: Mit Hilfe dieses Parameters kann der Koeffizient einer erklärenden Variablen auf den Wert Eins restringiert werden. Der Parameterwert gibt die Nummer der zu restringierenden erklärenden Variablen im VARIABLES Statement an. Soll beispielsweise in MNLOGIT die Variable VAR2 restringiert werden und das VARIABLES Statement lautet

VARIABLES INDIV CHOICE VARO VAR1 VAR2;

so ist **RESTRICT** = 3 anzugeben, da VAR2 die dritte erklärende Variable im VARIABLES Statement ist.

Die Restriktion auf dem Wert Eins ist direkt für Modelle mit aggregierten Alternativen verwendbar (siehe Hensher und Johnson, 1981). Bei entsprechenden Datenmanipulationen kann der Parameter RESTRICT jedoch für jede lineare Restriktion und Konstantenrestriktion verwendet werden. Soll beispielsweise das Modell

$$V_{in} = X_{1in}\beta_1 + X_{2in}\beta_2 + X_{3in}\beta_3 + X_{4in}\beta_4 \tag{39}$$

mit der Restriktion

$$a_1\beta_1 + a_2\beta_2 = R \tag{40}$$

geschätzt werden, so kann dies folgendermaßen erreicht werden: Löst man (40) nach - beispielsweise - β_1 und setzt in (39) ein, so ergibt sich

$$V_{in} = \frac{R}{a_1}X_{1in} - \frac{a_2}{a_1}X_{1in}\beta_2 + X_{2in}\beta_2 + X_{3in}\beta_3 + X_{4in}\beta_4 \tag{41}$$

Konstruiert man im DATA STEP die beiden Variablen

$$X_{5in} = \frac{R}{a_1}X_{1in} \qquad \text{und} \qquad X_{6in} = X_{2in} - \frac{a_2}{a_1}X_{1in}\,, \tag{42}$$

so wird (41) zu

$$V_{in} = X_{3in}\beta_3 + X_{4in}\beta_4 + X_{5in}\beta_5 + X_{6in}\beta_2 \quad \text{mit } \beta_5 = 1 \qquad (43)$$

Dieses Modell kann mit Hilfe des Parameters RESTRICT geschätzt werden. Sind zwei oder mehr Parameter auf Eins zu restringieren, so bildet man die Summe der entsprechenden Variablen, verwendet diese neue Variable an Stelle der ursprünglichen und restringiert ihren Parameter auf Eins.

Darstellung 10 im Anhang gibt ein Beispiel für Aufrufe von MNLOGIT und BPROBIT, in denen alle Dateibezeichnungen, Optionen und Parameter spezifiziert werden.

Fußnoten

1) Zur Vereinfachung der Schreibweise wird hier und in der weiteren Darstellung der Nutzen der Alternative i als U_{in} statt $U_n(a_i)$, und die Wahrscheinlichkeit für die Wahl der Alternative i als P_{in} statt $P(a_{n,opt} = a_i)$ bezeichnet.

2) Aus Darstellungsgründen wird für den Rest des Abschnitts 2.3 auf den Index für das Individuum (n) verzichtet.

Literaturverzeichnis

Aldrich J.H., Nelson F.D., 1984. Linear Probability, Logit, and Probit Models. Series: Quantitative Applications in the Social Sciences, Nr. 45. Beverly Hills.

Amemiya T., 1981. Qualitative response models: a survey. In: Journal of Economic Literature, Vol.XIX, S.1483-1536.

Anas A., 1982. Residential Location Markets and Urban Transportation, Economic Theory, Econometrics, and Policy Analysis with Discrete Choice Models. New York.

Ben-Akiva M., Lerman St.R., 1985. Discrete Choice Analysis: Theory and Application to Predict Travel Demand. Cambridge.

Ben-Akiva M., Litinas N., Tsunokawa K., 1985. Continuous Spatial Choice: The Continuous Logit Model and Distreibutions of Trips and Urban Densities. In: Transportation Research, Vol.19a, S.119-154.

Ben-Akiva M., Watanatada T., 1981. Application of a Continuous Choice Logit Model. In: Manski C., McFadden D., (Eds.) Structural Analysis of Discrete Data with Economic Applications. Cambridge.

Blake I.F., 1979. An Introduction to Applied Probability. New York.

Chiang A.C., 1974. Fundamental Methods of Mathematical Economics. Tokyo.

Domencich T.A., McFadden D., 1975. Urban Travel Demand, a Behavioral Analysis. Amsterdam.

Gravelle H., Rees R., 1981. Microeconomics. London.

Green H.A.J., 1976. Consumer Theory. London.

Hausman J., McFadden D., 1985. Specification Tests for the Multinomial Logit Model. In: Econometrica, Vol.52, S. 1219-1240.

Henderson J.M., Quandt R.E., 1977. Mikroökonomische Theorie, eine mathematische Darstellung. München.

Hensher D., Johnson L., 1981. Applied Discrete Choice Modelling. London, New York.

Johnson N., Kotz S., 1970. Distributions in Statistics - Continuous Univariate Distributions. New York.

Johnston J., 1972. Econometric Methods. New York.

Judge G.G., Griffiths W.E., Hill R.C., Lee T., 1980. The Theory and Practice of Econometrics. New York.

Lancaster K.J., 1966. A new approach to consumer theory. In: Journal of Political Economy, Vol.74, S.132-157.

Luce R., Suppes, P., 1965. Preference, Utility and Subjective Probability. In: Luce R., Bush R., Galanter E., (Eds.) Handbook of Mathematical Psychology, Vol. 3. New York.

Maddala G.S., 1977 Econometrics. New York.

Maddala G.S., 1983. Limited-Dependent and Qualitative Variables in Econometrics. Cambridge.

Maier G., Fischer M.M., 1985. Random utility modelling and labour supply mobility analysis. In: Papers and Proceedings of the Regional Science Association, im Erscheinen.

Manski C.F., McFadden D. (Eds.), 1981. Structural Analysis of Discrete Data with Econometric Applications. Cambridge.

Mayberry J.P., 1970. Structural requirements for abstract-mode models of passenger transportation. In: Quandt R.E., The Demand for Travel: Theory and Measurement. Lexington.

McFadden D., 1974. Conditional logit analysis of qualitative choice behavior. In: Zarembka P. (Ed.), Frontiers in Econometrics, S.105-142. New York.

McFadden D., 1976. The Mathematical Theory of Demand Models. In: Stopher P., Meyburg, A., (Eds.) Behavioral Travel Demand Models. Lexington.

McFadden D., Tye W., Train K., 1977. An Application of Diagnostic Tests for the Irrelevant Alternatives Property of the Multinomial Logit Model. In: Transportation Research Record, 637, S.39-46.

Rao C.R., 1973. Linear Statistical Inference and its Applications. New York.

Small K., Hsiao C., 1982. Multinomial Logit Specification Tests. Working Paper, Department of Economics, Princeton University, Princeton, N.J.

Wonnacott R.J., Wonnacott T.H., 1979. Econometrics, Second Edition. New York.

Anhang

Darst. 2: Parameterschätzung, Zwischenergebnisse

```
                                                              SAS
MULTINOMIALES LOGIT - COPYRIGHT: GUNTHER MAIER, IIR, WU - WIEN
==============================================================

 ANFANGSWERTE FUER BETA:

   0.0000000D+00  0.0000000D+00  0.0000000D+00


  NR.IT.       DELTA BETA       LOG-LIKELIHOOD
    1          0.74716371         -109.86
    2          0.54895866          -57.04
    3          0.52750719          -43.97
    4          0.38493299          -39.50
    5          0.14880276          -38.49
    6          0.01628404          -38.41
    7          0.00016456          -38.41
    8          0.00000188          -38.41
    9          0.00000180          -38.41
```

Darst. 3: Parameterschätzung, Endergebnisse

```
                                                              SAS
MULTINOMIALES LOGIT - COPYRIGHT: GUNTHER MAIER, IIR, WU - WIEN
==============================================================

 KONVERGENZ NACH     9 ITERATIONEN

 ABHAENGIGE VARIABLE     : CHOICE
 GRUPPIERUNGSVARIABLE    : IND

 VARIABLE          BETA          T-WERT        STEIGUNG
                STD.FEHLER

      X1         2.268721       5.380110    -0.1366430D-03
                 0.4216868

      X2        -3.352666      -5.262329    -0.8494731D-04
                 0.6371067

      X3         0.5665798      0.7475463     0.3296115D-04
                 0.7579194

 ANZAHL GRUPPEN :    100
 ANZAHL BEOB.   :    300
 ANZAHL FAELLE  :    200
 MAX. ANZ. ALT. :      3
 FEHLENDE WERTE :      0

 LOG-LIKELIHOOD
 ANFANG :    -109.8612
 ENDE   :     -38.40541
 LR-TEST:     142.9115      BEI      3 VARIABLEN

 RHO-QUADRAT :    0.6504187
 RHO-QU.KORR.:    0.6231115
```

Darst. 4: Ausgabe der Option PROB (Listing Datei)

```
MULTINOMIALES LOGIT - COPYRIGHT: GUNTHER MAIER, IIR, WU - WIEN   SAS           17:09 WEDNESDAY, OCTOBER 9, 1985
==================================================================

                 RESIDUEN PLOT

NR.BEOB. NR.ALT.  INCL.WERT  CHOICE  PROGNOSE      RESIDUEN        -                   0                   +
                                                                   1                                       1
   1        1    -14.18488     1     0.9971442      0.2855837D-02  .                   *                   .
   2        2    -14.18488     0     0.9631023D-04 -0.9631023D-04  .                   *                   .
   3        3    -14.18488     0     0.2759526D-02 -0.2759526D-02  .                   *                   .
   4        1    -15.94385     0     0.1070124D-01 -0.1070124D-01  .                   *                   .
   5        2    -15.94385     1     0.9892779      0.1072212D-01  .                   *                   .
   6        3    -15.94385     0     0.2087500D-04 -0.2087500D-04  .                   *                   .
   7        1    -14.44348     0     0.1546187D-01 -0.1546187D-01  .                   *                   .
   8        2    -14.44348     1     0.9818595      0.1814051D-01  .                   *                   .
   9        3    -14.44348     0     0.2678641D-02 -0.2678641D-02  .                   *                   .
  10        1    -19.01606     1     0.9872551      0.1274493D-01  .                   *                   .
  11        2    -19.01606     0     0.1076201D-01 -0.1076201D-01  .                   *                   .
  12        3    -19.01606     0     0.1982920D-02 -0.1982920D-02  .                   *                   .
  13        1    -18.48442     1     0.5977274      0.4022726      .                   I       *           .
  14        2    -18.48442     0     0.4007690     -0.4007690      .           *       I                   .
  15        3    -18.48442     0     0.1503573D-02 -0.1503573D-02  .                   *                   .
  16        1    -20.02959     0     0.4941576D-01 -0.4941576D-01  .                  *I                   .
  17        2    -20.02959     1     0.9274973      0.7250271D-01  .                   I*                  .
  18        3    -20.02959     0     0.2308696D-01 -0.2308696D-01  .                   *                   .
  19        1    -18.78379     0     0.1152096     -0.1152096      .                 * I                   .
  20        2    -18.78379     0     0.1300649D-01 -0.1300649D-01  .                   *                   .
  21        3    -18.78379     1     0.8717839      0.1282161      .                   I *                 .
  22        1    -16.90745     0     0.7309619     -0.7309619      .    *              I                   .
  23        2    -16.90745     1     0.2485379      0.7514621      .                   I              *    .
  24        3    -16.90745     0     0.2050016D-01 -0.2050016D-01  .                   *                   .
  25        1    -15.48914     0     0.6970281     -0.6970281      .     *             I                   .
  26        2    -15.48914     1     0.3029631      0.6970369      .                   I             *     .
  27        3    -15.48914     0     0.8790651D-05 -0.8790651D-05  .                   *                   .
  28        1    -15.48367     0     0.5594833D-02 -0.5594833D-02  .                   *                   .
  29        2    -15.48367     1     0.9912791      0.8720872D-02  .                   *                   .
  30        3    -15.48367     0     0.3126038D-02 -0.3126038D-02  .                   *                   .
  31        1    -15.31833     1     0.9675476      0.3245238D-01  .                   I*                  .
  32        2    -15.31833     0     0.2832021D-01 -0.2832021D-01  .                   *                   .
  33        3    -15.31833     0     0.4132165D-02 -0.4132165D-02  .                   *                   .
  34        1    -16.93394     1     0.4585730      0.5414270      .                   I          *        .
  35        2    -16.93394     0     0.5396801     -0.5396801      .         *         I                   .
  36        3    -16.93394     0     0.1746891D-02 -0.1746891D-02  .                   *                   .
  37        1    -18.01453     1     0.9768510      0.2314901D-01  .                   *                   .
  38        2    -18.01453     0     0.2278656D-01 -0.2278656D-01  .                   *                   .
  39        3    -18.01453     0     0.3624509D-03 -0.3624509D-03  .                   *                   .
  40        1    -15.40667     0     0.6425517D-02 -0.6425517D-02  .                   *                   .
  41        2    -15.40667     1     0.9935327      0.6467251D-02  .                   *                   .
  42        3    -15.40667     0     0.4173415D-04 -0.4173415D-04  .                   *                   .
  43        1    -16.05630     0     0.2531834D-02 -0.2531834D-02  .                   *                   .
  44        2    -16.05630     1     0.9771724      0.2282756D-01  .                   *                   .
  45        3    -16.05630     0     0.2029573D-01 -0.2029573D-01  .                   *                   .
  46        1    -19.81463     0     0.7449468     -0.7449468      .    *              I                   .
  47        2    -19.81463     1     0.2534825      0.7465175      .                   I              *    .
  48        3    -19.81463     0     0.1570728D-02 -0.1570728D-02  .                   *                   .
```

Darst. 5: Ausgabe der Option COVAR (Listing Datei)

```
                                                                   SAS
MULTINOMIALES LOGIT - COPYRIGHT: GUNTHER MAIER, IIR, WU - WIEN
==============================================================


 VARIANZ - KOVARIANZ - MATRIX


            X1               X2               X3
 X1          0.1778198       -0.2429439       0.7153466D-01
 X2         -0.2429440        0.4059050       0.7830167D-01
 X3          0.7153457D-01    0.7830178D-01   0.5744419
```

Darst. 6: Beispielsprogramm und -datei zur Option BSTART

```
FILE: LOGTEST  SAS      A1

CMS FILEDEF TEST DISK TEST DATEN C;
CMS FILEDEF 9 DISK BETA START A;
DATA TEST;
INFILE TEST;
INPUT NR IND ALT CHOICE X1 X2 X3;
PROC MNLOGIT BSTART;
VARIABLES IND CHOICE X1 X2 X3;



FILE: BETA     START    A1

2.27 -3.35 0.57
```

Darst. 7: Ausgabe der Option EST (SAS Datei 'EST')

SAS

OBS	NAME	EST
1	X1	2.2687
2	X2	-3.3527
3	X3	0.5666

Darst. 8: Ausgabe der Option SPROB (SAS Datei 'RESID')

SAS

OBS	NR	IND	ALT	INCL	CHOICE	PROGN
1	1	1	1	-14.185	1	0.997144
2	2	1	2	-14.185	0	0.000096
3	3	1	3	-14.185	0	0.002760
4	4	2	1	-15.944	0	0.010701
5	5	2	2	-15.944	1	0.989278
6	6	2	3	-15.944	0	0.000021
7	7	3	1	-14.443	0	0.015462
8	8	3	2	-14.443	1	0.981859
9	9	3	3	-14.443	0	0.002679
10	10	4	1	-19.016	1	0.987255
11	11	4	2	-19.016	0	0.010762
12	12	4	3	-19.016	0	0.001983
13	13	5	1	-18.484	1	0.597727
14	14	5	2	-18.484	0	0.400769
15	15	5	3	-18.484	0	0.001504
16	16	6	1	-20.030	0	0.049416
17	17	6	2	-20.030	1	0.927497
18	18	6	3	-20.030	0	0.023087
19	19	7	1	-18.784	0	0.115210
20	20	7	2	-18.784	0	0.013006
21	21	7	3	-18.784	1	0.871784
22	22	8	1	-16.907	0	0.730962
23	23	8	2	-16.907	1	0.248538
24	24	8	3	-16.907	0	0.020500
25	25	9	1	-15.489	0	0.697028
26	26	9	2	-15.489	1	0.302963
27	27	9	3	-15.489	0	0.000009
28	28	10	1	-15.484	0	0.005595
29	29	10	2	-15.484	1	0.991279
30	30	10	3	-15.484	0	0.003126
31	31	11	1	-15.318	1	0.967548

Darst. 9: Ausgabe der Option SCOV (SAS Datei 'COV')

SAS

OBS	NAME	X1	X2	X3
1	X1	0.17782	-0.24294	0.071535
2	X2	-0.24294	0.40590	0.078302
3	X3	0.07153	0.07830	0.574442

Darst. 10: Programmbeispiel mit allen Aufrufen

```
FILE: LOGTEST  SAS      A1


CMS FILEDEF TEST DISK TEST DATEN C;
CMS FILEDEF 9 DISK BETA START A;
DATA TEST;
INFILE TEST;
INPUT NR IND ALT CHOICE X1 X2 X3;
PROC MNLOGIT PROB COVAR BSTART EST SPROB SCOV MGES
        MAXIT=10 LIMIT=0.00001 INVERT=0.002 STEP=5.5 RESTRICT=1
        DATA=TEST EST=OUTEST RESID=OUTRES COV=OUTCOV;
VARIABLES IND CHOICE X1 X2 X3;
ID NR IND ALT;
```

ADDENDUM: Programmerweiterungen und fortgeschrittene Analysemethoden

1. Einleitung

Die nachfolgende Beschreibung baut auf den vorangegangenen Kapiteln auf und umfaßt zwei Punkte:

1.) Erweiterungen der in den vorangegangenen Kapiteln beschriebenen Möglichkeiten.

2.) Die Verwendung der in SAS verfügbaren Möglichkeiten für die Schätzung diskreter Entscheidungsmodelle. Dieser Punkt enthält vor allem schätztechnische und diagnostische "Tips und Tricks", beschreibt aber auch die erweiterten Möglichkeiten der Programme.

2. Erweiterungen

Die Erweiterungen, die seit der ersten Auflage von IIR-Disc. 27 in den Programmen vorgenommen wurden, ergaben sich aus zwei Quellen: Einerseits aus dem Ziel, in den Prozeduren BPROBIT und MNLOGIT möglichst alle jene Informationen auszugeben, die ein Benutzer von anderen Schätzungen gewöhnt ist, andererseits aus den Erfahrungen, die an der Wirtschaftsuniversität Wien mit der Anwendung der Prozeduren gemacht wurden und den dabei aufgetretenen Wünschen und Anregungen.

Wie in Schätzprogrammen üblich gibt die neue Version zusätzlich zu allen Statistiken auch die Wahrscheinlichkeit an, die diesen bei Zutreffen der Null-Hypothese entspricht. Zu jedem T-WERT gibt es bei der Ausgabe der Ergebnisse einen Wert PROB, der zeigt, mit welcher Wahrscheinlichkeit hinter dem errechneten Schätzwert des Parameters ein wahrer Wert von Null steht. Sowohl T-WERT als auch PROB beziehen sich immer auf diese Null-Hypothese. Für eine andere Null-Hypothese muß die entsprechende Statistik aus den Schätzwerten für Koeffizienten und Standardfehler vom Anwender selbst berechnet werden (siehe IIR-Disc. 27, S. 26).

Eine zweite Statistik, die von den Prozeduren automatisch berechnet wird ist LR-TEST. Sie vergleicht die Anfangs- und Endwerte der Schätzung. Der neben der Statistik LR-TEST ausgegebene PROB-Wert gibt daher die Wahrscheinlichkeit an, mit der die Anfangswerte der Schätzung die wahren Werte sind. Auch hier müssen Teststatistiken, die sich auf eine andere Null-Hpothese beziehen vom Benutzer berechnet werden. Er kann aber auch die Werte der Null-Hypothese als Startwerte der Schätzung spezifizieren und so die Berechnung von Teststatistik und Wahrscheinlichkeit dem Programm überlassen.

Eine wichtige Erweiterung betrifft die Prozedur BPROBIT. Bei Aufruf der Option SPROB gibt sie einen mit INV MILL bezeichneten Wertvektor aus. Es ist dies die Inverse der Mills-Ratio, die zur Korrektur eines "truncation bias" benötigt wird (genaueres siehe Maddala, 1981). BPROBIT kann damit auch zur Schätzung von "truncation bias"- und "censoring"-Modellen eingesetzt werden.

Neben den acht in IIR-Disc. 27 beschriebenen Optionen umfaßt die neue Version von BPROBIT und MNLOGIT vier weitere. Es sind dies: BSAS, ITER, SCREEN, und ERRORS. Sie werden im folgenden beschrieben.

BSAS: Diese Option hat dieselbe Aufgabe wie die Option BSTART, nämlich Startwerte für die Schätzung vorzugeben. Während BSTART eine CMS-Datei erwartet, verwendet BSAS eine SAS-Datei. Der Name der SAS-Datei wird durch

ESTIN = Dateiname

bestimmt. Die Vorgabe für ESTIN ist eine leere Datei, sodaß bei Angabe der Option BSAS ohne ESTIN die Standardwerte (Null) verwendet werden.

ITER: Die Option ITER führt zur Ausgabe der Schätzwerte der Koeffizienten nach jeder Iteration. Ohne Spezifikation von ITER werden sie erst nach der letzten Iteration aufgelistet. Die Ausgabe der Zwischenergebnisse ist dann hilfreich, wenn der Schätzalgorithmus nicht konvergiert oder zu einer singulären Matrix der zweiten Ableitungen führt. Probleme in der Datenstruktur können damit leichter identifiziert werden.

SCREEN: Mit dieser Option können die Prozeduren BPROBIT und MNLOGIT angewiesen werden, die Zwischenergebnisse nach den einzelnen Iterationen (DELTA BETA, LOG-LIKELIHOOD) auf eine CMS-Datei (FORTRAN-Dateinummer 16) auszugeben. Mit dem CMS Statement

CMS FILEDEF 16 TERMINAL;

wird diese Ausgabe auf den Bildschirm umgeleitet, sodaß die Zwischenergebnisse der Iterationen vom Benutzer verfolgt werden können. Dies ist vor allem dann ratsam, wenn die Gefähr besteht, daß die Schätzung nicht konvergiert. Bei den entsprechenden Zwischenergebnissen kann sie vom Benutzer abgebrochen werden. Auch bei sehr großen Datensätzen und entsprechend langen Rechenzeiten ist eine interaktive Kontrolle hilfreich. Wird die Option SCREEN ohne CMS-Statement angegeben, so werden die Zwischenergebnisse auf eine CMS-Datei auf der Magnetplatte geschrieben.

ERRORS: Der häufigste rechentechnische Fehler, der im Zusammenhang mit BPROBIT und MNLOGIT auftritt, ist ein Overflow oder Underflow bei Berechnung der EXP-Funktion. Da diese Funktion sehr häufig verwendet wird, würde der Versuch, diesen FORTRAN-Fehler zu vermeiden, sich sehr negativ auf die Rechenzeit auswirken. Bei Auftreten dieses Fehlers bricht FORTRAN allerdings die Bearbeitung nicht ab, sondern setzt das Resultat der EXP-Funktion auf einen sehr großen Wert (Overflow) bzw. auf Null (Underflow) und gibt eine Menge von Informationen aus. Da die Standard-Korrektur von FORTRAN genau der gewünschten entspricht, unterdrücken BPROBIT und MNLOGIT standardmäßig die von den Fehlerroutinen ausgegebenen Informationen (die die Lesbarkeit der Programmausgabe sehr beeinträchtigen).

Die Option ERRORS erlaubt es dem Benutzer, diese Unterdrückung rückgängig zu machen und die volle Information der Fehlerroutinen von EXP-Overflow und EXP-Underflow anzufordern. Dies ist dann sinnvoll, wenn andere Fehler auftreten, die möglicherweise eine Folge von EXP-Fehlern sind. Mit Ausgabe der Option ERRORS kann erreicht werden, daß die zeitliche Abfolge der einzelnen Fehler dargestellt wird.

Die Option ERRORS sollte für die normalen Anwendungen von BPROBIT und MNLOGIT eigentlich nicht benötigt werden. Auch FORTRAN-Fehler sollten in den üblichen Anwendungen nicht auftreten. Die Option wurde nur inkludiert, um dem Anwender den Zugang zur vollen Information der FORTRAN-Fehlerroutinen offen zu halten, und um eine Fehlinterpretation von Folgefehlern zu vermeiden.

3.) "Tips und Tricks"

Dieser Punkt beschreibt einige schätztechnische und diagnostische Hilfsmittel, die sich bei der Anwendung diskreter Entscheidungsmodelle als hilfreich erwiesen haben.

Die folgenden "Tips und Tricks" werden behandelt:

- Die Erstellung von Residuenplots.
- Die Schätzung von "nested logit" - Modellen
- Die Berechnung von marginalen Auswahlwahrscheinlichkeiten und Elastizitäten.
- Der Test der IIA-Eigenschaft.

Die Punkte zwei und vier sind nur im Zusammenhang mit MNLOGIT sinnvoll. Auch die anderen beiden sind für das Logit-Modell einfacher darstellbar als für das Probit-Modell. Die nachfolgende Diskussion bezieht sich daher prinzipiell auf das Logit-Modell. Für die entsprechenden Darstellungen für das Probit-Modell siehe Daganzo (1979).

Die vier Punkte gehen über die direkte Anwendung von BPROBIT und MNLOGIT hinaus und beschreiben auch das Zusammenspiel dieser Prozeduren mit den übrigen Möglichkeiten von SAS. Da diese Verfahren zum Teil von der konkreten Anwendung abhängen, werden sie hier nur schematisch dargestellt. Ihre exakte Implementierung bleibt dem Anwender überlassen.

3.1. **Die Erstellung von Residuenplots.**

Bei der Suche nach einem adäquaten Schätzmodell sind Residuenplots ein wertvolles Hilfsmittel. Dabei werden Streudiagramme zwischen den Residuen eines Logit- oder Probit-Modells und möglichen erklärenden Variablen erstellt. Sie geben Hinweise auf etwaige Spezifikationsfehler, fehlende Variable oder Heteroskedastizität

SAS verfügt mit der Prozedur PLOT über ein universelles Werkzeug zur Erstellung von Streudiagrammen, das auch für die Konstruktion von Residuenplots diskreter Entscheidungensmodelle verwendet werden kann. Die Übergabe der entsprechenden Variablen an PLOT besorgt die Option SPROB. Sie führt zur Ausgabe der beobachteten und prognostizierten Entscheidungen auf eine SAS-Datei. Das ID-Statement erlaubt die Übertragung jener Variabler in die Ausgabedatei, gegen die die Residuen im Streudiagramm dargestellt werden sollen.

In einem zwischen den Prozeduren MNLOGIT und PLOT eingeschobenen DATA-STEP müssen die Residuen als Differenz zwischen beobachteter und prognostizierter Entscheidung errechnet werden:

```
RESID = CHOICE-PROGN;                                    (1)
```

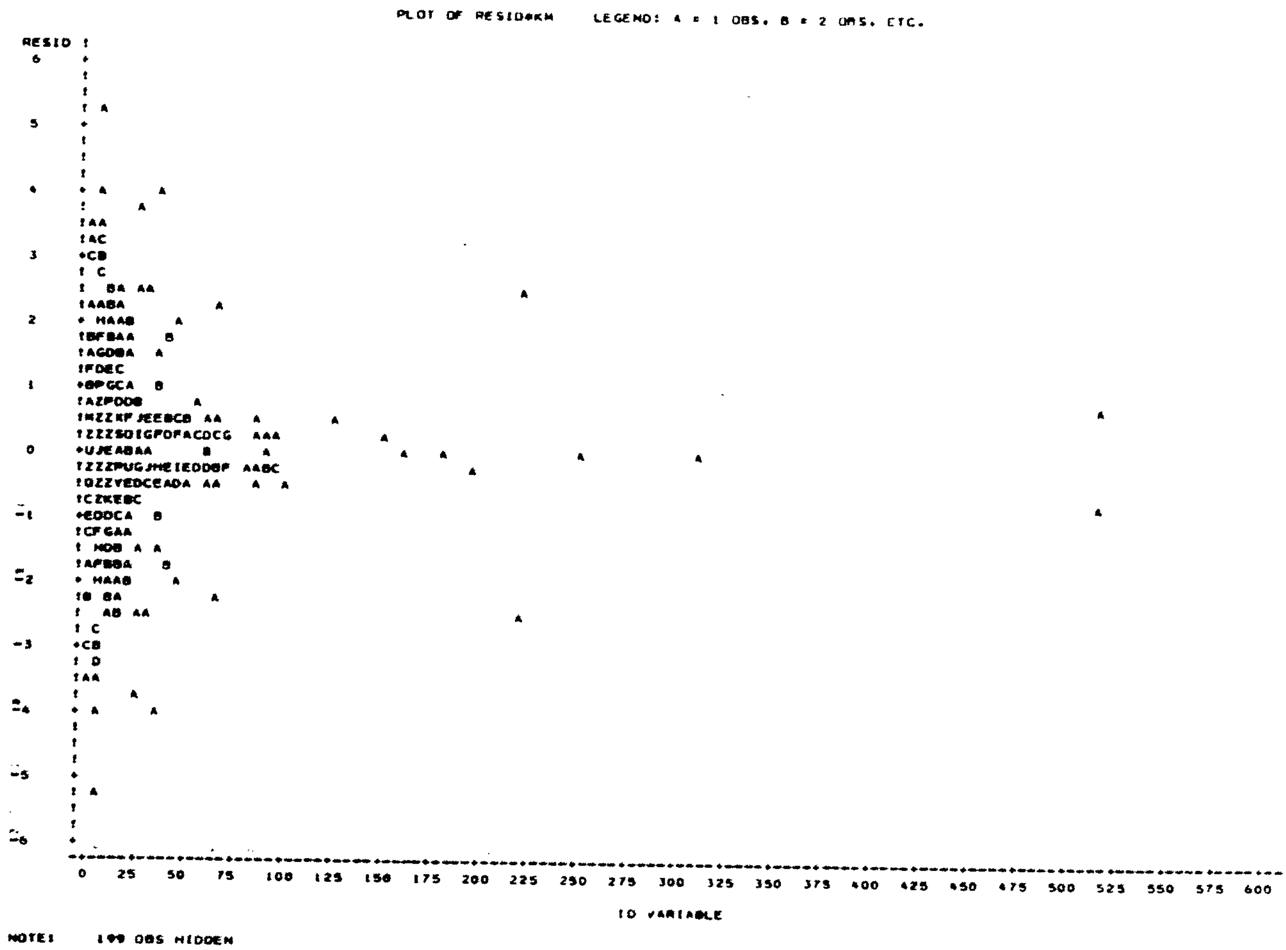
Alternativ dazu können auch standardisierte Residuen berechnet werden als (Wrigley, 1985):

$$\text{RESID} = (\text{CHOICE}-\text{PROGN})/\text{SQRT}(\text{PROGN}(1 - \text{PROGN})); \qquad (2)$$

Diese haben den Vorteil, daß die extremen Abweichungen stärker hervortreten.

Die Variable RESID kann nun gegen mögliche Einflußvariable dargestellt werden. Dabei ist zu beachten, daß die Residuen der gewählten Alternative immer positiv, jene der nicht-gewählten immer negativ sind. Außerdem führt die Summenrestriktion des Logit-Modells im binären Fall dazu, daß einem positiven Residuum immer ein dem Absolutbetrag nach gleich großes negatives Residuum gegenübersteht. Abbildung 1 zeigt ein Beispiel eines Residuenplots.

Abbildung 1. Residuenplot

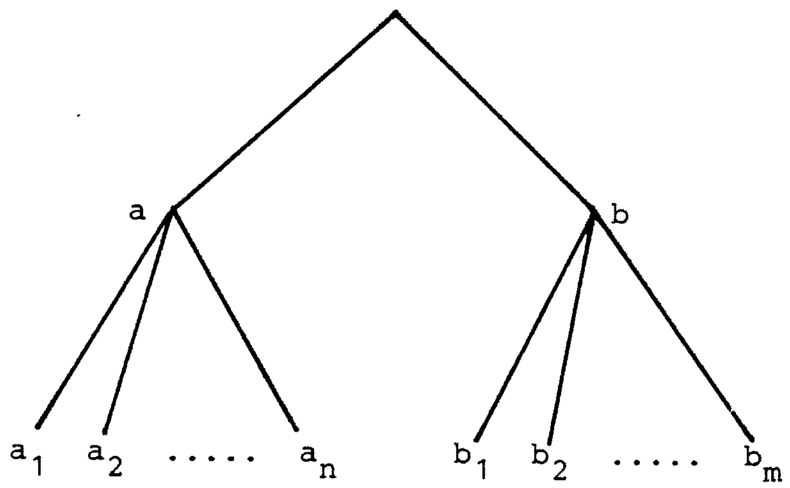


3.2. Die Schätzung von "nested logit"-Modellen.

Das "nested logit"-Modell stellt eine Verallgemeinerung des multinomialen Logit Modells dar. Es erlaubt teilweise Korrelation zwischen den Nutzen der Alternativen und auch unterschiedliche Streuungsparameter. Damit ist das "nested logit"-Modell nicht durch die IIA-Eigenschaft beschränkt. (Für Darstellungen des "nested logit"-Modells siehe unter anderem Ben Akiva & Lerman, 1985, Wrigley, 1985, Maier & Fischer, 1986).

Das Grundprinzip des "nested logit" - Modells liegt darin, daß es in einer hierarchischen Struktur dargestellt werden kann (siehe Abbildung 2).

Abbildung 2. Die Hierarchien des "nested logit"-Modells



Auf der ersten Ebene werden die Entscheidungen zwischen a und b modelliert, auf der zweiten jene zwischen a1 und an bzw. zwischen b1 und bm. Die Auswahlwahrscheinlichkeiten auf der Ebene zwei sind bedingt darauf, daß auf der ersten Ebene der entsprechende Ast gewählt wird. Das Konzept der Nutzenmaximierung erfordert andererseits, daß bei der Entscheidung auf der ersten Ebene die Möglichkeiten, die Ebene 2 bietet, berücksichtigt werden. Die Ebenen des "nested logit"-Modells sind also untereinander eng verbunden.

Trotz seiner allgemeinen Struktur kann das "nested logit"-Modell mit einem normalen Logit-Programm wie MNLOGIT geschätzt werden. Bei diesem "sequentielle Maximum Likelihood Schätzung" genannten Verfahren werden zuerst die beiden Modelle auf der Ebene 2 geschätzt. Ebene 1 dient dazu, den Datensatz entsprechend zu zerlegen. Um die Entscheidung auf Ebene 1 zu modellieren, muß der erwartete maximale Nutzen der beiden Modelle auf Ebene 2 errechnet werden, der - gemäß dem Konzept der Nutzenmaximierung - die übergeordnete Entscheidung beeinflußt. Im Logit-Modell errechnet sich dieser Wert (Inklusivwert) als

$$I_k = 1/\mu_k \ln \sum_{i \varepsilon K} \exp(\mu_k X_{in} \beta) \qquad (3)$$

Dabei beschreibt n das Individuum, k die Entscheidungen auf Ebene 1 und K die entsprechende Alternativenmenge auf Ebene 2. μ, x und ß haben die übliche Bedeutung (siehe IIR-Disc. 27). In dem in Abbildung 2 dargestellten Beispiel kann k also die Werte a und b annehmen, bei k = a besteht K aus den Elementen {a1, a2, ..., an}, bei k = b aus {b1, b2, ..., bm}.

Im Programm MNLOGIT werden diese Inklusivwerte automatisch berechnet und bei Spezifikation der Option SPROB auf eine SAS-Datei ausgegeben. Sonstige benötigte Variable können mittels ID-Statement in die Ausgabedatei übertragen werden. Zu beachten ist, daß nur die beiden auf der Ebene 2 generierten Dateien gemeinsam die für die Schätzung auf Ebene 1 benötigten Informationen enthalten. Sie müssen also mit den SAS-Statements SET bzw. MERGE miteinander und mit eventuell benötigten anderen Daten kombiniert werden. Außerdem enthalten die Ausgabedateien der Ebene 2 pro Element der entsprechenden Alternativenmenge und Individuum ein Record. Für die Schätzung der Ebene 1 wird aber nur ein Record pro Alternativenmenge und Individuum benötigt. Die Dateien müssen also auch entsprechend verkürzt werden. Etwa dadurch, daß in einem zwischengeschalteten DATA-STEP alle bis auf das erste Record jedes Individuums entfernt werden. Dazu muß die das Individuum identifizierende Variable jedes Records in einer Hilfsvariablen aufgehoben und anschließend mit der entsprechenden Variablen des nächsten Records verglichen werden. Die Hilfsvariable muß mit dem SAS-Statement RETAIN davor bewahrt werden, daß sie von SAS automatisch mit der Kennung für einen fehlenden Wert überschrieben wird. Die vorangegangenen Ausführungen bezogen sich immer auf die zwei in Abbildung 2 dargestellten Ebenen und Alternativen auf Ebene 1. Die Methode ist jedoch keineswegs darauf beschränkt und kann ebenso bei mehreren Ebenen und Alternativen verwendet werden.

3.3. Die Berechnung von marginalen Auswahlwahrscheinlichkeiten und Elastizitäten.

Marginale Auswahlwahrscheinlichkeiten und Elastizitäten sind zwei Maßzahlen für die Reaktionen auf Änderungen einer erklärenden Variablen. In einem "modal split"-Modell geben sie beispielsweise Aufschluß darüber, wie sich die Fahrgastzahlen öffentlicher Verkehrsmittel ändern, wenn die Tarife erhöht werden oder die Benzinpreise steigen. Die beiden Maßzahlen unterscheiden sich dadurch, daß die marginalen Auswahlwahrscheinlichkeiten die Reaktion auf eine Veränderung der erklärenden Variablen um eine Einheit, die Elastizitäten die prozentuelle Veränderung bei einer einprozentigen Veränderung der erklärenden Variablen angeben. Die beiden Maßzehlen stehen damit auch formal in engem Zusammenhang:

$$E_{in}^{jk} = M_{in}^{jk} \frac{X_{jk}}{P_{in}} \tag{4}$$

Dabei bezeichnet E die Elastizität, M die marginale Auswahlwahrscheinlichkeit. Gleichung (4) bezeichnet die Elastizität für Individuum n und Alternative i bei einer Änderung der Variablen k für Alternative j. Ist j gleich i, so spricht man von einem direkten Effekt, sonst von einem Kreuzeffekt.

Im allgemeinen sind nicht die Reaktionen eines Individuums sondern die der gesamten Bevölkerung von Interesse. Die entsprechenden Werte für die Gesamtbevölkerung errechnen sich als:

$$E_{i}^{jk} = \frac{\sum_{n} P_{in} E_{in}^{jk}}{\sum_{n} P_{in}} \qquad M_{i}^{jk} = \frac{\sum_{n} P_{in} M_{in}^{jk}}{\sum_{n} P_{in}} \tag{5}$$

Die individuellen marginalen Auswahlwahrscheinlichkeiten in (4) bezeichnen die Veränderung der Auswahlwahrscheinlichkeit bei einer marginalen Änderung der erklärenden Variablen:

$$M_{in}^{jk} = \partial P_{in}/\partial V_{jk} \; \partial V_{jk}/\partial X_{jk} \tag{6}$$

Die Aufspaltung in die Reaktion der Wahrscheinlichkeit auf eine Änderung des deterministischen Nutzens und die Reaktion des Nutzens auf eine Änderung der erklärenden Variablen ist vorteilhaft, da der zweite Teil von der verwendeten Schätzfunktion abhängt, der erste Teil jedoch vom verwendeten Modelltyp.

Für das Logit-Modell ergibt sich:

$$M_{in}^{jk} = \begin{cases} P_{in}(1-P_{in})\ \partial V_{in}/\partial X_{in} & \text{für } i=j \\ -P_{in}P_{jn}\ \partial V_{jn}/\partial X_{jk} & \text{für } i\neq j \end{cases} \qquad (7)$$

Die Ableitung von V nach X variiert von Schätzung zu Schätzung. Im Falle einer in den Variablen linearen Nutzenfunktionen ist sie gleich dem Parameter.

Für das Logit-Modell können marginale Auswahlwahrscheinlichkeiten und Elastizitäten damit folgendermaßen berechnet werde:

1.) Ausgabe der prognostizierten und beobachteten Häufigkeiten, der erklärenden Variablen und der Kennung der Individuen und Alternativen auf eine SAS-Datei mit der Option SPROB.

2.) Einlesen dieser Datei in einen DATA-STEP und Auswahl der entsprechenden Alternativen (i und j), berechnen der Ableitung von V nach X, berechnen von (7) und von (6).

3.) Zur Berechnung von Elastizitäten einsetzen in (4)

4.) Im nächsten Schritt müssen Zähler und Nenner der entsprechenden Gleichung in (5) berechnet und ausgegeben werden. Die Division in (5) erfolgt am besten außerhalb von SAS.

Da bei der Berechnung von Kreuzeffekten i und j nicht identisch sind, gestaltet sich die Auswahl der entsprechenden Alternativen etwas komplizierter als im Fall von direkten Effekten. Besonders ist dabei zu beachten, daß Hilfsvariable mit dem SAS-Statement RETAIN vor dem Überschreiben mit der Kennung für einen fehlenden Wert geschützt werden müssen.

3.4. Test der IIA - Eigenschaft

In IIR-Disc. 27 (Seite 30, Gleichung 37) wurde eine Teststatistik dargestellt, mit der überprüft werden kann, ob die impliziten Annahmen des Logit-Modells für eine bestimmte Modellversion zutreffen oder nicht. Mit Hilfe der Möglichkeiten von PROC MATRIX ist dieser Test in SAS sehr leicht durchzuführen. Wir benötigen dafür jeweils den Vektor der Koeffizienten und die Varianz-Kovarianz-Matrix für zwei Schätzungen; eine mit der vollen und eine mit einer reduzierten Alternativmenge. Bei Angabe der Optionen EST und SCOV werden diese von MNLOGIT auf SAS-Dateien ausgegeben. Diese Dateien enthalten allerdings jeweils eine Variable NAME, die vor der Verwendung in PROC MATRIX entfernt werden muß. Dies erfolgt am besten in einen kurzen DATA-STEP mit dem SAS-statement DROP . Die aus der Schätzung mit der vollen Alternativenmenge gewonnenen Dateien enthalten normalerweise auch Schätzwerte, die bei der restringierten Alternativenmenge nicht identifizierbar sind. Auch diese müssen aus den Dateien entfernt werden um die Vektoren und Matrizen auf identische Dimension zu bringen. Aus der Varianz-Kovarianz-Matrix müssen dabei entsprechende Spalten (Variable) und Zeilen (Beobachtungen) entfernt werden. Letzteres erfolgt mit dem SAS-Statement DELETE.

Sind die Dateien derart vorbereitet, so können sie von PROC MATRIX übernommen werden. Die Berechnung der Teststatistik in PROC MATRIX erfordert nur wenige Operationen. Sie können direkt der Definition der Teststatistik in IIR-Disc. 27, Gleichung 37 entnommen werden.

Literaturverzeichnis

Ben-Akiva M., Lerman St.R., 1985. Discrete Choice Analysis: Theory and Application to Predict Travel Demand. Cambridge.

Daganzo C.F., 1979. Multinomial Probit: The Theory and its Application to Demand Forecasting. New York.

Maier G., 1985. Die Schätzung diskreter Entscheidungsmodelle mit Hilfe der SAS Prozeduren BPROBIT und MNLOGIT, IIR-Discussion 27, Interdisziplinäres Institut für Raumordnung, Stadt- und Regionalentwicklung, Wirtschaftsuniversität Wien.

Maier G., Fischer M.M., 1986. Random utility modelling and labour supply mobility analysis. In: Papers of the Regional Science Association, Vol. 56.

Wrigley N., 1985. Categorial Data Analysis for Geographers and Environmental Scientists. New York.