

Regensburger
DISKUSSIONSBEITRÄGE
zur Wirtschaftswissenschaft

Conjoint-Analyse und Marktsegmentierung

Winfried J. Steiner und Bernhard Baumgartner

Lehrstuhl für Betriebswirtschaftslehre, insbesondere Marketing

Universität Regensburg, Universitätsstraße 31, D-93053 Regensburg

Nr. 382

März 2003

Conjoint-Analyse und Marktsegmentierung

A. Einführung

Die Conjoint-Analyse kann als eine Klasse von dekompositionellen Verfahren gekennzeichnet werden, die auf Basis empirisch erhobener Gesamtnutzenurteile für eine Menge von Stimuli (z.B. alternative Neuproduktkonzepte) die Parameter eines Präferenzmodells schätzen und auf diese Weise relative Nutzenbeiträge einzelner Stimulusmerkmale (z.B. Produkteigenschaften) zum Gesamtnutzen ermitteln. Sie kann heute als diejenige Form der Präferenzanalyse bezeichnet werden, die die höchste praktische Relevanz gefunden hat¹. Neben Neuproduktplanung und Preisgestaltung hat sich die Marktsegmentierung zu einem wesentlichen Einsatzgebiet der Conjoint-Analyse entwickelt². Die Marktsegmentierung liefert dabei Informationen über die Struktur des Marktes auf der Abnehmerseite, während sich aus der Conjoint-Analyse Empfehlungen für die Marketingpolitik – und insbesondere die Produktpolitik – in den einzelnen Segmenten ableiten lassen. Durch die Verknüpfung von Conjoint-Analyse und Marktsegmentierung lassen sich damit Lösungen für eine differenzierte Marktbearbeitung gewinnen.

Vor dem Hintergrund "traditioneller" zweistufiger Vorgehensweisen, die entweder (1) von einer nicht-präferenzbasierten a priori-Segmentierung des Marktes ausgehen und separate Conjoint-Analysen für jedes Segment vorsehen oder (2) die Segmentierung erst nach der Erhebung von Präferenzdaten bzw. erst nach Schätzung der Präferenzstrukturen einer für den Gesamtmarkt repräsentativen Stichprobe von Auskunftspersonen vornehmen, sind in jüngerer Zeit neue Methoden entwickelt worden, die eine simultane Segmentierung und Präferenzschätzung erlauben. Der vorliegende Beitrag gibt einen Überblick über die Möglichkeiten, Conjoint-Analyse und Marktsegmentierung zu verbinden, und beleuchtet insbesondere die neueren Simultanansätze, die gegenüber den traditionellen zweistufigen Methoden deutliche Vorzüge aufweisen.

In Abschnitt B werden zunächst die wesentlichen Grundzüge der rating-, ranking- und der wahlbasierten Conjoint-Analyse erörtert sowie auf im Zusammenhang mit der Modell-schätzung relevante Aspekte der Datenaggregation eingegangen. Darauf basierend werden in Abschnitt C nach einem kurzen Abriss der traditionellen zweistufigen Vorgehensweisen

zur Präferenzschätzung und Segmentierung fortgeschrittene simultane Conjointsegmentierungsmethoden, deren Entwicklung seit Mitte der 80er Jahre bis heute stark forciert wurde, näher vorgestellt. Abschnitt D beschließt den Beitrag mit einem Fazit.

C. B. Conjoint-Analyse und Datenaggregation

I. Ranking-, rating- und wahlbasierte Conjoint-Analysen

Im Rahmen der Datenerhebung einer Conjoint-Analyse geben Auskunftspersonen Präferenzurteile für verschiedene Stimuli ab. Unter einem Stimulus versteht man dabei eine bestimmte Kombination von Eigenschaftsausprägungen, die den Befragten in verbaler, bildlicher oder realer Form präsentiert werden. Bei der heute dominierenden, weil realitätsnahen sogenannten Profilmethode sind von den Probanden vollständige Stimulusprofile, bestehend aus der Kombination je einer Ausprägung jeder relevanten Eigenschaft, zu bewerten³.

Zur Erfassung der Präferenzurteile der Auskunftspersonen stehen mehrere Möglichkeiten zur Verfügung. Beim Ranking werden die Stimuli von den Befragten entsprechend des wahrgenommenen Nutzens in eine Reihenfolge gebracht und mit nichtmetrischen Platzziffern versehen.⁴ Eine anspruchsvollere Aufgabe stellt sich den Probanden, wenn sie die Vorziehungswürdigkeit der einzelnen Stimuli auf einer Ratingskala einstufen bzw. im Rahmen von Paarvergleichen Nutzenunterschiede zwischen je zwei Stimuli quantifizieren sollen und damit bereits zu metrischen Präferenzurteilen aufgefordert werden⁵. Bei der Entscheidung für eine von beiden Bewertungsmethoden (Ranking versus Rating) ist unter anderem zwischen der mutmaßlich höheren Reliabilität von ordinalen Rankingdaten und dem potenziell höheren Informationsgehalt metrischer (genau genommen intervallskalierter) Ratingdaten abzuwägen⁶.

Eine Alternative zur Schätzung eines Präferenzmodells auf Basis von Ranking- oder Ratingdaten bietet sich mit der wahlbasierten (choice-based) Conjoint-Analyse⁷. Hier werden die Auskunftspersonen in einem sogenannten Choice-Experiment gebeten, aus wechselnden fiktiven Alternativenmengen (Choice-Sets) den jeweils meistpräferierten Stimulus auszuwählen. Diese Variante wird als die realen Wahlsituationen am nächsten kommende eingestuft. Zudem kann auch der Fall berücksichtigt werden, keine der verfügbaren Alternativen wählen zu wollen⁸. Allerdings unterbleibt hier eine explizite Bewertung

der jeweils nicht gewählten Alternativen, was zu einem vergleichsweise geringeren Informationsgehalt der erhobenen Daten führt⁹.

Im Anschluss an die Datenerhebung werden aus den vorliegenden Präferenzurteilen mit entsprechenden Schätzmethoden metrische (intervallskalierte) Teilnutzenwerte für die einzelnen Stimulusmerkmale gewonnen¹⁰. Das bei Conjoint-Analysen vornehmlich eingesetzte Präferenzmodell ist dabei das linear-additive, kompensatorische Teilnutzenwertmodell, das Unabhängigkeit und Diskretheit der berücksichtigten Eigenschaften voraussetzt und wie folgt dargestellt werden kann¹¹:

$$y_{ij} = \sum_{m=1}^M \sum_{n=1}^{N_m} \beta_{imn} z_{(i)jmn} \quad (1)$$

$i = 1, \dots, I$ Auskunftspersonen;

$j = 1, \dots, J_{(i)}$ von Individuum i zu bewertende Stimuli;

$m = 1, \dots, M$ Eigenschaften; $n = 1, \dots, N_m$ Ausprägungen von Eigenschaft m ;

y_{ij} : Präferenzwert von Individuum i für Stimulus j (bzw. für den bei Bewertungsgelegenheit j zu beurteilenden Stimulus);

β_{imn} : Teilnutzenwert des i -ten Individuums für Ausprägung n von Eigenschaft m ;

$$z_{(i)jmn} = \begin{cases} 1 & \text{falls Stimulus } j \text{ (bzw. der Individuum } i \text{ bei Bewertungsgelegenheit } j \\ & \text{vorgelegte Stimulus) die Ausprägung } n \text{ von Eigenschaft } m \text{ besitzt} \\ 0 & \text{sonst} \end{cases}$$

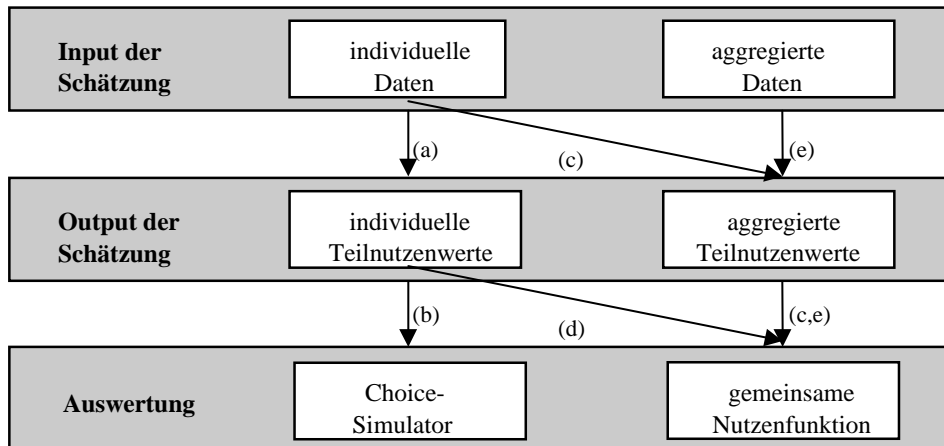
Bei der wahlbasierten (choice-based) Conjoint-Analyse können Teilnutzenwerte *und* Wahlwahrscheinlichkeiten auf Basis der erhobenen dichotomen Auswahlentscheidungen der Auskunftspersonen mit dem multinomialen Logitmodell geschätzt werden. Die Wahrscheinlichkeit $w_{ijq_{ij}}$, mit der ein Stimulus q_{ij} von Auskunftsperson i bei der j -ten Bewertungsgelegenheit aus dem entsprechenden Choice-Set CS_{ij} gewählt wird, ergibt sich dann aus¹²:

$$w_{ijq_{ij}} = \frac{\exp\left(\sum_m \sum_n \beta_{imn} z_{(i)jq_{ij}mn}\right)}{\sum_{k \in CS_{ij}} \exp\left(\sum_m \sum_n \beta_{imn} z_{(i)jkmn}\right)} \quad (2)$$

II. Aspekte der Datenaggregation

Eine grundlegende Entscheidung, die bei der Durchführung von Conjoint-Analysen zu treffen ist, betrifft das Aggregationsniveau der Daten. *Abbildung 1* stellt die verschiedenen Aggregationsmöglichkeiten vereinfacht dar:

Abb. 1: Möglichkeiten der Datenaggregation bei Conjoint-Analysen



Falls für jedes Individuum auf der Grundlage seiner individuellen Präferenzdaten eine Nutzenfunktion geschätzt wird (a), spricht man von Individuallevel-Analysen. Die Ermittlung der individuellen Teilnutzenwerte kann dann bei metrischen Präferenzwerten (Ratingdaten) oder bei berechtigter Annahme der Intervallskalierungseigenschaft von Rangdaten mit Hilfe eines Kleinst-Quadrate-Ansatzes erfolgen¹³:

$$\min_{\beta_{imn}} \sum_{j=1}^{J_{(i)}} \left(y_{ij} - \sum_{m=1}^M \sum_{n=1}^{N_m} \beta_{imn} z_{(i)jmn} \right)^2 \quad (3)$$

Den Anforderungen der Regressionsanalyse entsprechend müssen für jedes Individuum mehr Beobachtungen $J_{(i)}$ vorliegen als Parameter β_{imn} zu schätzen sind. Auch wenn diese technische Voraussetzung erfüllt und eine Schätzung damit grundsätzlich möglich ist, sind bei einer kleinen Datenbasis aufgrund der dann geringen Anzahl an Freiheitsgraden instabile Parameterschätzungen zu erwarten¹⁴. Diese Problematik ineffizienter Parameterschätzungen tritt bei individuellen Conjoint-Analysen fast regelmäßig auf, wird aber bei der Ableitung von Marketing-Implikationen aus den Schätzergebnissen nur selten weiter hinterfragt. Möglichkeiten, die Effizienz von Parameterschätzungen bei Individuallevel-Analysen zu erhöhen, bestehen zum Beispiel in einer sparsameren Parametrisierung oder in

einer Restriktion für die erlaubten Beziehungen zwischen Parametern (wie etwa die plausible Forderung höherer Teilnutzenwerte für niedrigere Preisausprägungen)¹⁵.

Als problematisch bei Individuallevel-Analysen kann auch die Interpretation des Outputs bei einer großen Zahl von Auskunftspersonen gesehen werden. Da Unternehmen in der Regel nicht auf einer vollständig disaggregierten Ebene handeln, sind marketingpolitische Maßnahmen nicht ohne weiteres aus individuellen Teilnutzenwerten zu bestimmen. Hier werden daher häufig Choice-Simulatoren eingesetzt (b), um Marktanteile oder die Profitabilität vorgegebener Stimuli zu prognostizieren¹⁶.

Neben der Individuallevel-Analyse besteht die Möglichkeit einer gemeinsamen Conjoint-Analyse. Dabei werden die individuell erhobenen Präferenzurteile gepoolt und gemeinsam zur Schätzung *eines* für alle Auskunftspersonen gültigen Teilnutzenwertmodells (c) herangezogen¹⁷. Der Kleinst-Quadrate-Ansatz zur Ermittlung der Teilnutzenwerte lautet in diesem Fall:

$$\min_{\beta_{imn}} \sum_{i=1}^I \sum_{j=1}^{J_{(i)}} \left(y_{ij} - \sum_{m=1}^M \sum_{n=1}^{N_m} \beta_{imn} Z_{(i)jmn} \right)^2 \quad (4)$$

Eine gemeinsame Analyse individueller Erhebungsdaten findet in der Regel auch bei wahlbasierten Conjoint-Modellen statt. Hier scheidet eine Schätzung individueller Teilnutzenwerte aufgrund der beschränkten Datensituation in Form "lediglich" dichotomer Präferenzurteile meist von vornherein aus¹⁸. Die Teilnutzenwerte wahl-basierter Conjoint-Modelle können wie bei multinomialen Logitmodellen üblich mittels Maximum-Likelihood-Schätzung ermittelt werden¹⁹.

Durch das Datenpooling bei der gemeinsamen Conjoint-Analyse wird die Problematik der Überparametrisierung und der damit einhergehenden Ineffizienz der Parameterschätzer vermieden, da bei gleicher Anzahl zu schätzender Teilnutzenwerte eine ungleich höhere Zahl an Beobachtungen ($I \cdot J_{(i)}$) als bei Individuallevel-Analysen (I Bewertungen je Individuum) und damit eine ungleich höhere Anzahl an Freiheitsgraden für die Schätzung vorliegt. Die Anwendbarkeit einer gemeinsamen Conjoint-Analyse hängt jedoch grundlegend davon ab, wie groß die Heterogenität der individuellen Präferenzstrukturen (bzw. der mit der Aggregation verbundene Informationsverlust) und die daraus resultierende Verzerrung

der Parameterschätzer ist. Die Folgen einer Orientierung an verzerrten Teilnutzenwerten für die Ableitung von Implikationen können gravierend sein, so beispielsweise wenn sich die Gestaltung eines neuen Produktes an der Ausprägung mit dem höchsten Teilnutzenwert je Eigenschaft anlehnt, der durch die gepoolten Koeffizienten repräsentierte fiktive Durchschnittskonsument jedoch aufgrund starker Heterogenität der Individuen am Markt nicht oder nur vereinzelt zu finden ist. Die gleichen Vorbehalte gelten im übrigen auch für eine nachträgliche Aggregation individuell geschätzter Teilnutzenwerte durch Mittelwertbildung zu Durchschnittswerten (d). Dass mit Individuallevel-Analysen häufig eine höhere Prognosegenauigkeit als mit gemeinsamen Analysen erzielt werden kann²⁰, liegt daran, dass bei starker Heterogenität der Individuen die Instabilität der Parameterschätzungen weniger schwer als die aus einer Aggregation resultierende Verzerrung der Teilnutzenwerte wiegt.

Schließlich bestünde noch die Möglichkeit einer gemeinsamen Conjoint-Analyse auf Basis bereits zu einem Durchschnittsdatensatz aggregierter Erhebungsdaten (e). Diese Variante birgt allerdings sowohl die Gefahr verzerrter Schätzer (bei vorliegender Heterogenität) als auch ineffizienter Schätzer (durch die im Vergleich zum Pooling stark reduzierte und wie bei Individuallevel-Analysen auf J Beobachtungen beschränkte Datenbasis) und stellt bei vollkommen aggregierter Marktbetrachtung keine wirkliche Alternative dar.

D. C. Conjoint-Analyse und Marktsegmentierung

Mit der Individuallevel-Analyse und der vollständigen Aggregation wurden in Abschnitt B "extreme" Möglichkeiten der Marktbetrachtung aufgezeigt. Unternehmen sind jedoch häufig nicht in der Lage, auf vollständig disaggregierter Ebene zu handeln. Auf der anderen Seite wird die völlige Vernachlässigung heterogener Abnehmerbedürfnisse und die Ausrichtung aller Marketingaktivitäten auf einen fiktiven "Durchschnittskonsumenten" für die meisten Unternehmen keine sinnvolle Strategie darstellen²¹. Seit Einführung des Konzeptes der Marktsegmentierung durch Smith (1956) ist diese ein zentrales Thema in Marketingtheorie und -praxis. Unter Marktsegmentierung versteht man dabei die Aufteilung eines heterogenen Gesamtmarktes von (potenziellen) Abnehmern in homogenere Teilmärkte²².

Traditionelle Vorgehensweisen, die die Conjoint-Analyse und Marktsegmentierung miteinander verbinden, sind dadurch gekennzeichnet, dass die Bildung der Segmente und die

Schätzung von Teilnutzenwerten in zwei getrennten Schritten erfolgt. Die einfachste Möglichkeit stellt dabei die Durchführung segmentspezifischer (gemeinsamer) Conjoint-Analysen für vorab durch andere Verfahren – wie der Clusteranalyse oder subjektiver Beurteilungskriterien – bestimmte Segmente dar²³. Der Erfolg dieser Vorgehensweise hängt allerdings davon ab, in welchem Ausmaß die für die Segmentierung herangezogenen Hintergrundvariablen die Heterogenität der individuellen bzw. segmentspezifischen Präferenzstrukturen bedingen. Die Erfahrung zeigt, dass psychografische und insbesondere demografische Abnehmermerkmale, die gerne für eine solche a priori-Segmentierung herangezogen werden, häufig nicht zur Beschreibung heterogener Präferenzstrukturen geeignet sind²⁴.

Schon sehr früh haben dagegen Studien belegt, dass insbesondere die Nutzen (Benefits), die Konsumenten in Produkten suchen bzw. die Individuen mit den ihnen präsentierten Stimuli verbinden, als Segmentierungsbasis geeignet sind²⁵. Abbildung 2 gibt einen Überblick über Verfahren, die die Conjoint-Analyse zur Benefitsegmentierung nutzen. Diese Verfahren werden im Weiteren als *Conjointsegmentierungsverfahren* bezeichnet und in den folgenden Abschnitten näher vorgestellt. Dabei wird zunächst kurz auf die weit verbreitete, traditionelle zweistufige Benefitsegmentierung eingegangen (Abschnitt C.I), bevor neuere Entwicklungen ausführlicher beleuchtet werden (Abschnitt C.II).

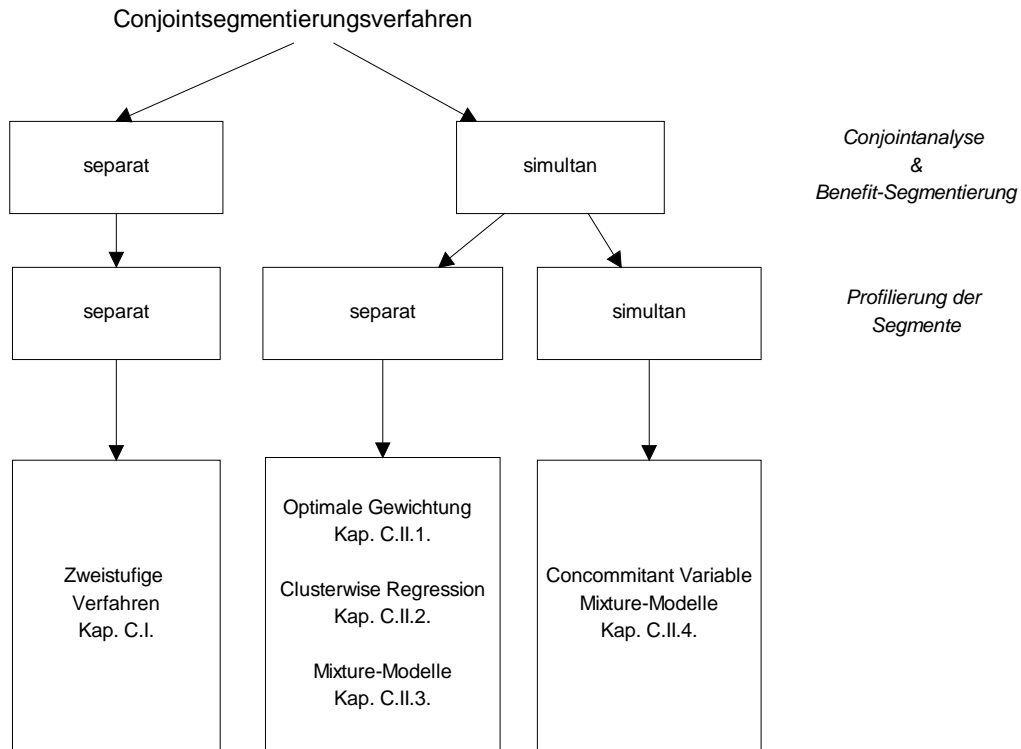
I. Traditionelle zweistufige Conjointsegmentierung

Die zweistufige Conjointsegmentierung nutzt die erhobenen Präferenzdaten oder die ermittelten Teilnutzenwerte der Conjoint-Analyse zur Gruppierung der Auskunftspersonen. Im letzteren Fall spricht man auch von post hoc-Segmentierung²⁶. Bei der post hoc-Segmentierung werden die Konsumenten auf Basis der Ähnlichkeiten ihrer individuell ermittelten Teilnutzenwerte unter Anwendung eines Clusterverfahrens zu in der Regel disjunkten Klassen zusammengefasst²⁷.

Nach Ermittlung der (Benefit-)Segmente erfolgt gewöhnlich für jedes Segment eine Aggregation der individuellen Teilnutzenwerte durch Mittelwertbildung, um jede Gruppe in bezug auf ihre Präferenzstrukturen zu charakterisieren. Falls im Rahmen der Conjoint-Studie auch Hintergrunddaten wie demografische oder sozioökonomische Charakteristika erhoben worden sind, können diese zur Profilierung der Segmente genutzt werden. Als wesentliche

Kritikpunkte an der post hoc-Segmentierung werden angeführt, dass die potenziell nicht-

Abb. 2: Conjointsegmentierungsverfahren



reliablen individuellen Teilnutzenwerte zu Fehlklassifikationen führen (können) sowie insbesondere, dass der Schätzung und der anschließenden Clusterung der Teilnutzenwerte unterschiedliche Optimierungskriterien zugrunde liegen. So führen die Clusteralgorithmen zugrunde liegenden varianz- bzw. distanzbasierten Fusionierungskriterien nicht zur Maximierung der Prognosegüte innerhalb der Segmente, sondern lediglich zu einer Gruppierung von Individuen mit ähnlichen Teilnutzenwertstrukturen, ohne die Schätzfehler aus der Bestimmung der Teilnutzenwerte zu berücksichtigen²⁸. Das grundsätzliche Manko zweier nicht notwendigerweise kompatibler Optimierungsschritte besteht auch dann, wenn die Individuen zunächst anhand ihrer gesamtheitlichen Präferenzurteile geclustert und im zweiten Schritt für die ermittelten Segmente separate Conjoint-Modelle geschätzt werden. Die Relaxierung bzw. Aufhebung der genannten Beschränkungen der zweistufigen Benefitsegmentierung mit dem Ergebnis einer höheren Prognosegüte ist die eigentliche Intention, die mit der Entwicklung neuerer Verfahren zur simultanen Conjointsegmentierung verfolgt und auch erreicht wird.

II. Neuere Methoden zur simultanen Conjointsegmentierung

1. Die Methode der optimalen Gewichtung

Bei neueren Methoden werden die eigentliche Conjoint-Analyse und die Segmentierung simultan durchgeführt. Die Arbeit von Hagerty (1985) kann dabei als eine der Pionierarbeiten in der Entwicklung dieser Simultanmethoden gesehen werden. Zentraler Ansatzpunkt bei Hagerty ist die plausible Überlegung, dass die Zusammenfassung ähnlicher Individuen zwar zu einem leichten Bias in den Teilnutzenwertschätzungen führt, die Varianz der Schätzungen jedoch durch das Pooling mehrerer Individuen stark verringert und somit insgesamt die Prognosegüte gegenüber Individuallevel-Analysen verbessert werden kann.

Hagerty formuliert dazu ein verallgemeinertes Gewichtungsschema, das jede mögliche Form der Aggregation individueller Präferenzdaten abbilden kann und gleichzeitig in Verbindung mit einer OLS-Regression alle aus der Cluster- wie auch der Faktorenanalyse bekannten Möglichkeiten einer Zuordnung von Individuen zu Klassen bzw. einer Verdichtung von Informationen bietet. Dieses Gewichtungsschema liefert auch einen Bezugsrahmen für die im Weiteren vorgestellten Simultanmethoden. Ausgangspunkt für die mathematische Darstellung des Modells ist eine vollkommen disaggregierte Marktbetrachtung mit dem zugehörigen Kleinst-Quadrate-Ansatz, der in Matrixform wie folgt dargestellt werden kann:

$$B = (X'X)^{-1}X'Y \quad (5)$$

Y : $(J \times I)$ -Matrix aller (standardisierten) Stimulibewertungen. Die ij -te Zelle dieser Matrix enthält das Präferenzurteil von Individuum i für Stimulus j .

X : $(J \times Q)$ -Datenmatrix für das einheitliche Erhebungsdesign der Conjoint-Studie. Die j -te Zeile der Matrix definiert das j -te Stimulusprofil und zeigt dessen Dekomposition in $Q = \sum_m N_m - M$ dummykodierte Eigenschaftsausprägungen.

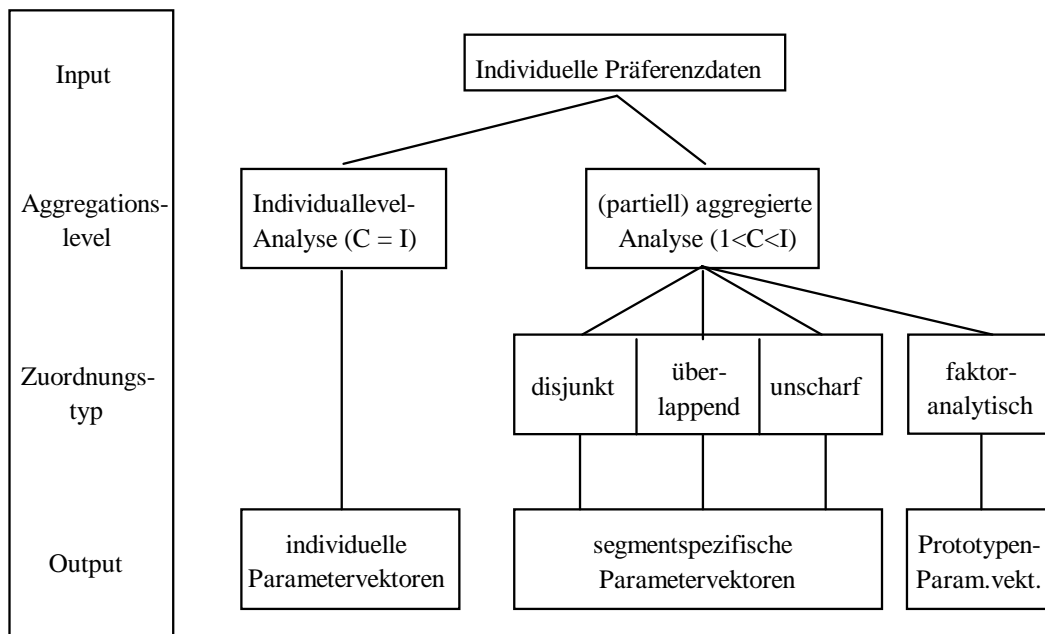
B : $(Q \times I)$ -Matrix der individuellen Parametervektoren $(I \times 1)$. Die iq -te Zelle dieser Matrix enthält den (zu schätzenden) Teilnutzenwert für Individuum i bezüglich des q -ten Prädiktors.

Die Aggregation von Individuen wird nun möglich über die Definition einer Gewichtungsmatrix G der Dimension $(I \times C)$ mit C als Anzahl der Cluster bzw. Faktoren (S_1, S_2, \dots, S_C) . Durch die Festlegung von C lässt sich der Aggregationsgrad steuern, wobei $C = I$ für

Individuallevel-Analysen, $1 < C < I$ für Segmentlevel-Analysen und $C = 1$ für vollständig aggregierte Analysen gilt.

Sollen disjunkte oder überlappende Segmente gebildet werden, so kann dies über die Zuweisung binärer Matrixwerte g_{ic} ($i \in I, c \in C$) erreicht werden. Bei disjunkten Klassen enthält jede Zeile genau einmal den Wert eins, der das i -te Individuum genau einem Segment S_c zuordnet. Sind überlappende Cluster zulässig, ist eine Mehrfachzuordnung von Individuen zu Klassen durch entsprechende Binärkodierungen möglich (d.h. jede Zeile kann mehrmals den Wert eins enthalten). Unscharfe Klassifikationen lassen sich durch reelle Matrixeinträge $g_{ic} \in [0;1]$ repräsentieren, wobei $\sum_{c=1}^C g_{ic} = 1$ gilt und die Matrixelemente g_{ic} die Wahrscheinlichkeiten für die Zugehörigkeit des i -ten Individuums zum c -ten Segment angeben. Im Falle einer faktoranalytischen Partitionierung können die Zellen der Matrix G schließlich jede reelle Zahl enthalten. Die extrahierten Faktoren repräsentieren dann Prototypen von Individuen und die Faktorladungen beschreiben die Ähnlichkeit (den Grad der Zugehörigkeit) eines Individuums zum jeweiligen Prototyp (Segment). *Abbildung 3* fasst die unterschiedlichen "Verarbeitungsmöglichkeiten" von Individualdaten im Ansatz von Hagerty noch einmal zusammen.

Abb. 3: Möglichkeiten der Datenaggregation im Ansatz von Hagerty



Unter diesen Voraussetzungen lassen sich die Parametervektoren für jedes beliebige vorgegebene Zuordnungsschema G mit folgendem modifizierten Kleinst-Quadrate-Ansatz schätzen:

$$B = (X'X)^{-1} X'YG(G'G)^{-1} G' \quad (6)$$

Nach Bestimmung der gepoolten Schätzer kann die Prognosegüte für das zugrunde liegende Gewichtungsschema G durch Kreuzvalidierung bestimmt werden, indem die erwartete mittlere quadratische Abweichung (EMQA) zwischen den vom Modell prognostizierten (\hat{y}_{ij}) und den von den Auskunftspersonen tatsächlich abgegebenen Bewertungen (y_{ij}) für eine Menge nicht zur Schätzung herangezogener Stimuli ($j = J + 1, \dots, R$) eines Holdout-Samples berechnet wird:

$$EMQA = E \left[\frac{1}{I} \sum_{i=1}^I \frac{1}{(R-J)} \sum_{j=J+1}^R (y_{ij} - \hat{y}_{ij})^2 \right] \quad (7)$$

Dieses Fehlermaß berücksichtigt neben dem modellimmanenten Stichprobenfehler sowohl die Standardfehler (Varianzen) der geschätzten Teilnutzenwerte als auch den durch eine Segmentierung bzw. Faktorisierung entstehenden Informationsverlust²⁹ und trägt somit dem in Kapitel B erläuterten Trade-off zwischen der Ineffizienz von Parameterschätzungen, die auf disaggregierter Ebene am stärksten ausgeprägt ist, und der durch die (partielle) Aggregation der Individuen bedingten Verzerrung der Parameterschätzungen, die je nach Heterogenität der individuellen Präferenzstrukturen stärker oder schwächer ausfällt, geeignet Rechnung.

Unter der Beschränkung zueinander orthogonaler Spalten(vektoren) der Gewichtungsmatrix G , wodurch die Bildung unscharfer sowie überlappender Klassifikationen ausgeschlossen wird, zeigt Hagerty, dass die EMQA minimiert werden kann und die optimale Gewichtungsmatrix G^* einer Hauptkomponentenlösung mit C^* extrahierten Faktoren ($C^* \leq Q$) entspricht³⁰. Die Spaltenvektoren der Matrix G^* repräsentieren dann die c^* Eigenvektoren zur Korrelationsmatrix $Y'Y$ und das ci -te Matrixelement die Faktorladung (Korrelation) für das i -te Individuum auf den c -ten Faktor.

Sofern die Hauptkomponentenlösung auf die Existenz einer Klassenstruktur hinweist (was im Idealfall durch eine Einfachstruktur der Gewichtungsmatrix G^* gegeben wäre³¹), lassen

sich entsprechend C* Prototypen (Zentroide) für die Segmente durch Aggregation der auf die jeweiligen Faktoren hoch ladenden Individuen bilden. Kritisiert werden kann die Methode der optimalen Gewichtung deshalb vor allem, wenn keine Klassenstruktur erkennbar ist, da in diesem Fall das Ziel einer maximalen Prognosegüte zu Lasten der Interpretierbarkeit der Lösung geht. Die Zahl der extrahierten Faktoren kann dann auch nicht als adäquater Indikator für die Zahl der Segmente herangezogen werden und die entsprechenden Faktorladungen nicht als Grad der Zugehörigkeit zu einem Segment oder zu einem Prototyp interpretiert werden³². Gerade im Hinblick auf die für unternehmenspolitische Zwecke im Vordergrund stehende Identifizierbarkeit von (mehr oder weniger) voneinander abgrenzbaren Segmenten sind dann clusterbildende Simultanverfahren vorzuziehen. Das Ausmaß des damit gegebenenfalls verbundenen Verlustes an Prognosegüte ist davon abhängig, inwieweit die Annahme der Existenz einer Klassenstruktur gerechtfertigt ist.

2. Clusterwise Regression

Unter Clusterwise Regression lassen sich in Anlehnung an das von Hagerty eingeführte verallgemeinerte Gewichtungsschema alle Simultanansätze subsumieren, die von der Existenz einer Segmentstruktur ausgehen und eine Klassifikation von Individuen anstreben. Unabhängig davon, ob disjunkte, überlappende oder unscharfe Klassen gebildet werden sollen, wird hier die Zielsetzung verfolgt, die Individuen einer bestimmten Anzahl C von Segmenten (S_1, S_2, \dots, S_C) so zuzuordnen und damit die Gewichtungsmatrix G so zu bestimmen, dass die Anpassungsgüte (als Indikator für die erwartete Prognosegüte) maximiert wird. Dies ist gleichbedeutend mit der Minimierung der (gewichteten) gepoolten Fehlerquadratsumme aus C linearen Regressionsanalysen³³

$$\min_{\beta_{cmn}} \sum_{c=1}^C \sum_{i=1}^I \sum_{j=1}^{J(i)} g_{ic}^\delta \left(y_{ij} - \sum_{m=1}^M \sum_{n=1}^{N_m} \beta_{cmn} z_{(i)jmn} \right)^2 \quad (8)$$

mit g_{ic}^δ ($i \in I, c \in C$) als Elementen der Gewichtungsmatrix G mit je nach Klassifikationstyp festgelegten Wertebereichen und dem Exponenten δ mit $\delta = 1$ bei disjunkten und überlappenden Klassifikationen bzw. $\delta > 1$ bei unscharfen Clustern.

Die vorliegenden Lösungsansätze für diese Problemstellung lassen sich sowohl nach dem zugrunde liegenden Klassifikationstyp als auch nach der verwendeten Technik zur Klassifi-

zierung differenzieren. Für den Fall disjunkter Klassen schlägt Kamakura (1988) eine hierarchische agglomerative Vorgehensweise vor. Ausgehend von einer vollkommen disaggregierten Partition ($C = I$) erfolgt hier eine schrittweise Fusionierung der Individuen. In jedem Fusionierungsschritt werden dabei diejenigen Individuen (Klassen von Individuen) zusammengefasst, deren Vereinigung zum geringsten Anstieg der gepoolten Residuenquadratsumme (8) führt. Im Rahmen zweier Studien mit synthetischen bzw. empirischen Daten zeigt Kamakura (1988, S. 162 ff.), dass die *hierarchical clusterwise regression* für unterschiedlichste Aggregationsgrade konsistent bessere Prognosen als die traditionelle zweistufige Benefitsegmentierung ermöglicht. Als Kritikpunkte an der hierarchischen Vorgehensweise lassen sich dennoch anführen, dass Fehlklassifikationen infolge noch relativ instabiler Parameterschätzungen in frühen Fusionierungsschritten auf höhere Ebenen mitgeführt werden und Modelle, die auf individueller Ebene überparametrisiert sind ($J_{(i)} \leq Q$), der Methode nicht zugänglich sind³⁴. Eine im Grundprinzip zum Ansatz von Kamakura (1988) analoge hierarchische Simultanmethode hat Ogawa (1987) für ordinale Ranking-Daten vorgestellt.

Bei Verwendung eines partitionierenden Verfahrens für disjunkte Klassen wird dagegen bereits von einer vorgegebenen Zerlegung in C Segmente, die zufällig gewählt wird oder auch das Ergebnis einer vorgeschalteten *hierarchical clusterwise regression* sein könnte, ausgegangen. Mittels eines Austauschverfahrens werden dann die Individuen unter Beibehaltung der Anzahl der Klassen und unter Berücksichtigung der Restriktion $\sum_{i \in S_c} J_i \geq Q$ ($c \in C$) so umgruppiert, dass die Anpassungsgüte maximiert wird. Späth (1979, 1981, 1982) hat in diesem Zusammenhang erstmals eine Methode zur Clusterwise Regression vorgestellt. Dabei wird sukzessive für jedes Individuum untersucht, ob seine Verschiebung in ein anderes Segment die gepoolte Residuenquadratsumme (8) reduziert. Falls dies der Fall ist, wird das Individuum in jenes Cluster verschoben, für das die maximale Verminderung erzielt werden kann. Der Algorithmus terminiert, wenn sich innerhalb eines Durchlaufs (I Iterationen) keine Veränderungen mehr ergeben. Der ursprüngliche Algorithmus von Späth kann problemlos auf die Conjointsegmentierung übertragen werden, indem beachtet wird, dass jeweils alle $J_{(i)}$ von einem Individuum i abgegebenen Präferenzurteile gleichzeitig verschoben werden müssen.

Wedel und Kistemaker (1989) verwenden anstelle der Methode von Späth den Austauschalgorithmus von Banfield und Bassill (1977). Hier folgt auf die "Transferphase", die mit dem Algorithmus von Späth identisch ist, eine "Swapphase", bei der alle Möglichkeiten, durch den wechselseitigen Tausch zweier Individuen zwischen verschiedenen Clustern eine weitere Verbesserung der Anpassungsgüte zu erzielen, enumeriert werden. Beide Phasen werden solange iteriert, bis die gepoolte Residuenquadratsumme nicht weiter reduziert werden kann. Auch Wedel und Kistemaker belegen mit einer empirischen Studie die höhere Prognosevalidität der *clusterwise regression* gegenüber der zweistufigen post hoc-Segmentierung. Mit der *clusterwise logistic regression* hat Katahira (1987) einen zu Wedel und Kistemaker analogen Ansatz für die wahlbasierte Conjoint-Analyse vorgestellt.

Für den Fall überlappender Klassen haben bislang lediglich DeSarbo, Oliver und Rangaswamy (1989) eine Methode publiziert. DeSarbo et al. setzen dabei Simulated Annealing ein, um ausgehend von einer zufällig erzeugten Gewichtungsmatrix G mit entsprechenden binären Zeilenvektoren iterativ eine verbesserte Klassenlösung (mit jeweils niedrigerer gepoolter Residuenquadratsumme) zu erzeugen. Die Methode kann allerdings dahingehend kritisiert werden, dass sie nur für kleine Datenmengen geeignet ist³⁵.

Die *fuzzy clusterwise regression* mit $0 \leq g_{ic} \leq 1$ ($i \in I, c \in C$), $\sum_{c=1}^C g_{ic} = 1$ und $\sum_{i=1}^I g_{ic} > 0$ erlaubt schließlich auch unscharfe Zuordnungen und damit partielle Zugehörigkeiten von Individuen zu Segmenten. Wie bei partitionierenden Verfahren wird auch hier von einer vorgegebenen Anzahl C an Klassen ausgegangen³⁶. Zusätzlich ist auch ein Unschärfeparameter δ ($\delta > 1$) geeignet voreinzustellen, was von Befürwortern der Methode als hohe Flexibilität, von Kritikern dagegen als Nachteil eingestuft wird³⁷. Mit diesem Parameter lässt sich der Grad der Unschärfe, zu dem ein Individuum mehr als nur einem Cluster angehören kann, beeinflussen. Werte von δ nahe 1 liefern relativ scharfe Zuordnungen mit individuellen Zugehörigkeiten g_{ic} ($c \in C$) nahe 0 oder 1, während für große Werte von δ sehr unscharfe Klassifikationen mit Zugehörigkeiten $g_{ic} \approx 1/C$ resultieren³⁸. Für die Fuzzy-Clusterwise Regression haben Wedel und Steenkamp (1989) einen Algorithmus vorgestellt. Dabei werden ausgehend von der gewählten Startklassifikation und nach Fixierung des Unschärfeparameters δ in einem ersten Schritt zunächst mit C gewichteten

Kleinst-Quadrate-Schätzungen die C klassenspezifischen Parametervektoren ermittelt. In einem zweiten Schritt werden dann für jede Person auf Basis ihrer ungewichteten Residuen

$\hat{r}_{ic} = \sum_{j=1}^{J_i} (y_{ij} - \sum_{m=1}^M \sum_{n=1}^{N_m} \beta_{cmn} z_{(i)jmn})^2$ bezüglich jeden Segments ($c \in C$) neue Segmentzugehörigkeiten \hat{g}_{ic} geschätzt³⁹:

$$\hat{g}_{ic} = \left[\sum_{w=1}^C (\hat{r}_{ic} / \hat{r}_{iw})^{\frac{1}{z-1}} \right]^{-1} \quad (9)$$

Die Anpassung der Gewichte mittels Ausdruck (9) bewirkt, dass Individuen eine stärkere (schwächere) Zugehörigkeit für Segmente erhalten, in denen sie vergleichsweise geringe (hohe) Residuen aufweisen. Beide Schritte (Schätzung und Anpassung) werden iterativ solange wiederholt, bis Konvergenz hinsichtlich der gewichteten Residuenquadratsumme (8) erreicht ist bzw. bis sich die Parameterschätzungen oder die Segmentzugehörigkeiten stabilisieren.

Die Methode von Wedel und Steenkamp (1989) ist nach geeigneter Datentransformation auch für wahlbasierte Conjoint-Daten anwendbar⁴⁰. Ebenso wie der Ansatz von Wedel und Kistemaker (1989) für disjunkte Klassen kann auch der Ansatz von Wedel und Steenkamp (1989) dahingehend modifiziert werden, dass individuelle Conjoint-Designs, bei denen Auskunftspersonen nicht notwendigerweise identische Stimulisets vorgelegt werden, verarbeitet werden können⁴¹.

3. Mixture-Modelle

Eine Alternative zur Clusterwise Regression stellt die Anwendung von Mixture-(Mischverteilungs-)modellen dar. Mixture-Modelle bauen auf der Wahrscheinlichkeitstheorie auf und zeichnen sich damit im Gegensatz zur Clusterwise Regression durch eine statistische Fundierung aus. Grundlegende Annahme ist hier zunächst, dass sich die Grundgesamtheit, aus der Daten erhoben werden, aus zwei oder mehreren latenten Segmenten (S_1, S_2, \dots, S_C) mit unbekanntem relativen Größen ($\alpha_1, \alpha_2, \dots, \alpha_C$) zusammensetzt. Die relativen Segmentgrößen können als a priori-Wahrscheinlichkeiten, mit der eine bestimmte Beobachtung i ($i = 1, \dots, I$) aus dem Segment c ($c = 1, \dots, C$) stammt, interpretiert werden. Ferner wird für jedes Segment eine spezifische Dichtefunktion $f_c(y_i | \theta_c)$ mit unbekanntem Parametervektor θ_c unterstellt, mit der die bedingte Wahrscheinlichkeit des Auftretens der Beobach-

tung y_i bei gegebener Zugehörigkeit zum Segment c beschrieben werden kann. Bei kontinuierlichen Variablen, wie sie bei der metrischen Conjoint-Analyse vorliegen, wird dabei am häufigsten eine normalverteilte Dichte zugrunde gelegt. Bei Anwendung des Mixture-Modells auf Daten aus einer Conjoint-Analyse ist zudem wieder zu beachten, dass für jedes Individuum i nicht nur eine Beobachtung y_i , sondern ein Vektor $\tilde{y}_i = (y_{i1}, \dots, y_{iJ})$ von J Stimulibewertungen vorliegt. \tilde{y}_i folgt dann einer multivariaten Normalverteilung und der segmentspezifische Parametervektor θ_c beinhaltet neben den zu schätzenden Teilnutzenwerten $\beta_c = (\beta_{c11}, \dots, \beta_{cmm})$ auch die $J \times J$ -Kovarianzmatrix Σ_c für Segment c ⁴². Die segmentspezifischen Dichtefunktionen lassen sich damit – mit X wiederum als Datenmatrix für ein einheitliches Erhebungsdesign der Conjoint-Studie – wie folgt formulieren⁴³:

$$f_c(\tilde{y}_i | X, \beta_c, \Sigma_c) = (2\pi)^{-J/2} |\Sigma_c|^{-1/2} \exp \left[-\frac{1}{2} (\tilde{y}_i - X\beta_c)^T \Sigma_c^{-1} (\tilde{y}_i - X\beta_c) \right] \quad (10)$$

Die unbedingte Wahrscheinlichkeit für den Vektor \tilde{y}_i der Präferenzurteile von Individuum i ergibt sich schließlich als Mischung (Mixture) der segmentspezifischen Dichtefunktionen:

$$f(\tilde{y}_i | \Phi) = f(\tilde{y}_i | \alpha_1, \dots, \alpha_C, X, \beta_1, \dots, \beta_C, \Sigma_1, \dots, \Sigma_C) = \sum_{c=1}^C \alpha_c f_c(\tilde{y}_i | X, \beta_c, \Sigma_c) \quad (11)$$

Unter Verwendung des Bayes-Theorems können nun a posteriori-Wahrscheinlichkeiten \hat{g}_{ic} ($i \in I, c \in C$) für die Zugehörigkeit des Individuums i zu Segment c mit $\sum_c \hat{g}_{ic} = 1$ und $0 \leq \hat{g}_{ic} \leq 1$ ermittelt werden:

$$\hat{g}_{ic} = \frac{\alpha_c f_c(\tilde{y}_i | X, \beta_c, \Sigma_c)}{\sum_{c=1}^C \alpha_c f_c(\tilde{y}_i | X, \beta_c, \Sigma_c)} \quad (12)$$

Die Bestimmung der Parameter im spezifizierten Mischverteilungsansatz erfolgt durch die Maximierung der zugehörigen Log-Likelihoodfunktion:

$$\ln L = \sum_{i=1}^I \ln \left[\sum_{c=1}^C \alpha_c f_c(\tilde{y}_i | X, \beta_c, \Sigma_c) \right] \quad (13)$$

DeSarbo et al. (1992) verwenden zur Maximierung der Log-Likelihoodfunktion den Expectation-Maximization (EM)-Algorithmus⁴⁴. In der E-Phase werden für gegebene Parameter α_c, β_c und Σ_c ($c \in C$) die a posteriori-Wahrscheinlichkeiten \hat{g}_{ic} nach Ausdruck (12) berechnet, während in der M-Phase durch gewichtete Kleinstquadratschätzungen die

Koeffizienten für die Teilnutzenwerte β_c ermittelt werden. Das Verfahren weist insofern Ähnlichkeit zur Fuzzy-Clusterwise-Regression auf, da iterativ in einem ersten Schritt die Teilnutzenwerte innerhalb der Segmente ermittelt und im zweiten Schritt die Zugehörigkeiten der Individuen zu den Segmenten neu justiert werden, bis die Anpassungsgüte im Gesamtmodell, repräsentiert durch den Log-Likelihood-Wert, nicht mehr verbessert werden kann.

In einer empirischen Anwendung des beschriebenen Modells zur Gestaltung eines schlüssellosen Türöffnungssystems für Automobile können DeSarbo et al. (1992, 279 ff.) nicht nur zeigen, dass sich die aus dem Mischverteilungsansatz resultierende Segmentlösung deutlich von entsprechenden Segmentlösungen aus der Anwendung der zweistufigen post hoc-Segmentierung (mit Single Linkage, Complete Linkage, Average Linkage, Median, Centroid, Ward und KMEANS als alternativen Clusteralgorithmen) unterscheidet, sondern auch zu einer wesentlich höheren Anpassungsgüte führt⁴⁵. DeSarbo, Ramaswamy und Cohen (1995) haben darüber hinaus auch einen Mischverteilungsansatz für die wahlbasierte Conjoint-Analyse vorgestellt. Die unbedingte Wahrscheinlichkeit, dass ein Individuum i bei der j -ten Bewertungsgelegenheit aus seinem Choice-Set CS_{ij} einen bestimmten Stimulus auswählt, wird dann durch eine Mischung von C konditionalen Logitmodellen operationalisiert⁴⁶. Durch Anwendung der "Rank Explosion"-Regel können mit dem Mischverteilungsansatz für wahlbasierte Daten schließlich auch Teilnutzenwerte bei Vorliegen von Rangordnungsdaten geschätzt werden⁴⁷.

4. Concomitant-Variable-Mixture-Modelle

Für eine segmentspezifische Marktbearbeitung reichen Informationen, die sich auf die Existenz und relative Größe von Segmenten beschränken, in der Regel nicht aus. Die Segmente müssen zusätzlich durch beobachtbare Charakteristika – wie typischerweise demografische und sozioökonomische Hintergrundvariablen (concomitant variables) – profiliert werden, um die Zielgruppen beschreiben und mit entsprechend ausgestalteten Marketing-Maßnahmen erreichen zu können.

Üblicherweise geht dem Profilierungsschritt die Bestimmung der Segmente bzw. im vorliegenden Kontext die simultane Conjointsegmentierung voraus. Wurde eine disjunkte Segmentlösung ermittelt, kann zur Profilierung etwa die Diskriminanzanalyse eingesetzt werden. Eine Möglichkeit zur Profilierung bei unscharfen Zuordnungen besteht in einer Regression der logittransformierten Zugehörigkeitswahrscheinlichkeiten g_{ic} ($i \in I, c \in C$) auf die Hintergrundvariablen⁴⁸. Ein Nachteil dieser getrennten Vorgehensweise, die die simultane Segmentierung und Schätzung der Teilnutzenwerte in einem ersten Schritt und die Profilierung erst in einem nachgelagerten zweiten Schritt vornimmt, liegt jedoch darin, dass in beiden Schritten wiederum unterschiedliche Kriterien optimiert werden⁴⁹. Zudem wird bei unscharfen Segmentlösungen der Schätzfehler der im ersten Schritt ermittelten Zugehörigkeitswahrscheinlichkeiten im zweiten Schritt ignoriert⁵⁰.

Zur Beseitigung der genannten Nachteile haben Kamakura, Wedel und Agrawal (1994) ein sogenanntes *concomitant variable mixture model*, mit dem sich simultan Teilnutzenwerte schätzen sowie Segmente bilden und profilieren lassen, entwickelt. Die simultane Profilierung wird dabei in Erweiterung eines "gewöhnlichen" Mischverteilungsmodells durch Reparametrisierung der a priori-Zugehörigkeitswahrscheinlichkeiten $\alpha_{c|h}$ ($c \in C$) in Abhängigkeit von den zu berücksichtigenden Hintergrundvariablen ($h = 1, \dots, H$), mit a_{ih} als Ausprägung der h-ten Hintergrundvariablen bei Individuum i ($h \in H, i \in I$) und γ_{hc} als Parameter für den Einfluss der h-ten Profilierungsvariablen auf die a priori-Zugehörigkeitswahrscheinlichkeit zum Segment c ($h \in H, c \in C$), möglich⁵¹:

$$\alpha_{c|h} = \frac{\exp\left(\sum_{h=1}^H \gamma_{hc} a_{ih}\right)}{\sum_{c=1}^C \exp\left(\sum_{h=1}^H \gamma_{hc} a_{ih}\right)} \quad (14)$$

In einer von Kamakura, Wedel und Agrawal (1994) durchgeführten empirischen Anwendung für Bankdienstleistungen übertrifft das Concomitant-Variable-Mixture-Modell hinsichtlich der Log-Likelihood-Werte in Schätz- und Validierungssample sowohl ein aggregiertes Conjoint-Modell als auch ein Mixture-Modell ohne Berücksichtigung von Hintergrundvariablen⁵².

D. Zusammenfassung

Im Zusammenhang mit kommerziellen Anwendungen der Conjoint-Analyse in Deutschland stellt die Marktsegmentierung heute eine der wichtigsten Zielsetzungen von Unternehmen dar. Dies wird durch die empirischen Recherchen von Baier (1999) bei den wichtigsten deutschen Anbietern (Marktforschungs- und Marketingberatungsunternehmen) für Conjoint-Studien dokumentiert. Die Untersuchung von Baier zeigt aber auch, dass die Unternehmenspraxis hier (noch immer) in starkem Maße auf den Einsatz zweistufiger Vorgehensweisen, mit denen die Schätzung von Präferenzstrukturen entweder im Vorfeld einer Segmentierung (post hoc-Segmentierung) oder erst im Anschluss an eine Segmentierung (a priori-Segmentierung) erfolgt, setzt. Ein Grund dafür könnte sein, dass in den Unternehmen die neueren Ansätze, die eine simultane Durchführung von Conjoint-Analyse und Marktsegmentierung ermöglichen, noch nicht ausreichend bekannt sind. Der vorliegende Artikel will deshalb vor allem einen Beitrag zu einer höheren Transparenz dieser Methoden leisten, denn die Ergebnisse der bisher durchgeführten, vorwiegend metrische Conjoint-Modelle betreffenden Vergleichsstudien deuten – wie auch in den einzelnen Abschnitten angeführt – auf eine höhere Prognosevalidität dieser Simultan-Ansätze gegenüber den zweistufigen Vorgehensweisen hin. Vriens, Wedel und Wilms (1996) haben im Rahmen einer Monte-Carlo-Simulationsstudie für metrische Conjoint-Daten zudem eine deutliche Überlegenheit sowohl des Mischverteilungsansatzes als auch der clusterwise und fuzzy clusterwise regression, was die Aufdeckung der "wahren" Koeffizienten (Teilnutzenwerte) und Segmentzugehörigkeiten anbetrifft, gegenüber zweistufigen post hoc-Segmentierungen aufzeigen können. Die Leistungsunterschiede zwischen Simultan- und zweistufigen Ansätzen bezüglich dieser beiden Kriterien werden dabei mit steigender Segmentanzahl immer größer.

Für wahlbasierte Conjoint-Modelle liegen derart umfangreiche Vergleichsstudien unseres Wissens bislang nicht vor. Lediglich DeSarbo et al. (1995) erzielen in einer empirischen Anwendung für ihr wahlbasiertes Mixture-Modell eine deutlich höhere Anpassungsgüte gegenüber einer zweistufigen Vorgehensweise mit einer a priori-Segmentierung auf Basis von Wahlhäufigkeiten. Teichert (2000b) vergleicht den Mischverteilungsansatz von DeSarbo et al. (1995)⁵³ und ein aggregiertes Conjoint-Modell anhand zweier empirischer Datensätze mit dem Ergebnis einer (teilweise deutlich) höheren Prognosegüte für das

latente Klassenmodell. In einem direkten empirischen Vergleich von wahlbasierter und metrischer Conjoint-Analyse erzielen Vriens et al. (1998) schließlich mit dem wahlbasierten Mixture-Modell von DeSarbo et al. (1995) eine höhere Prognosegüte als mit dem Mixture-Modell von DeSarbo et al. (1992) für metrische Conjoint-Daten.

Aus Anwendersicht ist festzuhalten, dass heute für alle praktisch relevanten Erhebungsformen der Conjoint-Analyse und Klassifikationstypen für die Segmentierung eine Simultanmethode als Pendant zur zweistufigen Vorgehensweise vorliegt. Insbesondere für Mixture-Modelle sind darüber hinaus auch kommerzielle Softwarepakete, vertrieben beispielsweise von Sawtooth Software, Inc., erhältlich. Für die Clusterwise und Fuzzy Clusterwise Regression werden entsprechende Softwarepakete von den jeweiligen Autoren (vgl. Abschnitt C.II.3) zur Verfügung gestellt.

Grundsätzlich wird die Wahl der Methode vor allem von der Frage, ob die angestrebte Segmentlösung durch eine disjunkte, nichtdisjunkte oder unscharfe Klassifikation repräsentiert werden soll, abhängen⁵⁴. Als Vorteile der Gruppe der Clusterwise-Regression-Verfahren gegenüber Mixture-Modellen können diesbezüglich gesehen werden, dass auch disjunkte und überlappende Segmente ermittelbar sind beziehungsweise für eine unscharfe Segmentlösung die gewünschte Unschärfe voreingestellt werden kann. Letzteres kann wie bereits angeführt aber auch als Nachteil ausgelegt werden, wenn der Anwender keine Vorstellung über eine erwünschte oder sinnvolle Unschärfe besitzt. Auch ist es möglich, nachträglich aus unscharfen Zuordnungen disjunkte Segmente zu bilden, indem jedes Individuum dem Segment zugeordnet wird, für das es die stärkste Zugehörigkeit aufweist.

Ein Vorteil von Mixture-Modellen gegenüber der Clusterwise Regression liegt darin, dass im Rahmen der Maximum-Likelihood-Schätzung die Kovarianzmatrix bestimmt werden kann und damit ohne wesentlichen Aufwand t-Statistiken für Teilnutzenwerte ermittelt werden können. Bei den Clusterwise-Regression-Methoden ist dies nur durch aufwändigere Verfahren wie beispielsweise Monte-Carlo-Tests möglich⁵⁵. Der Ansatz von Hagerty bietet dagegen keine Möglichkeit zur Durchführung von Signifikanztests⁵⁶. Nicht zuletzt erlauben Mixture-Modelle bei zusätzlicher Reparametrisierung der a priori-Zugehörigkeitswahrscheinlichkeiten in Abhängigkeit von relevanten Hintergrundvariablen als einzige Modellgruppe auch die simultane Profilierung von Segmenten.

Literaturverzeichnis

- Anderson, S.; de Palma, A.; Thisse, J.-F.* (1992): Discrete Choice Theory of Product Differentiation, Cambridge, London.
- Backhaus, K.; Erichson, B.; Plinke, W.; Weiber, R.* (2000): Multivariate Analysemethoden, 9. Auflage. Springer Verlag, Berlin, Heidelberg, New York.
- Baier, D.* (1999): Methoden der Conjoint-Analyse in der Marketingforschungs- und Marketingpraxis. In: Gaul, W. (Hrsg.): Mathematische Methoden der Wirtschaftswissenschaften: Festschrift für Otto Opitz, S. 197–206.
- Banfield, C.F.; Bassil, L.C.* (1977): A Transfer Algorithm for Non-hierarchical Classification. In: Applied Statistics, Vol. 16, S. 206–210.
- Batsell, R.; Louviere, J.J.* (1991): Experimental Analysis of Choice. In: Marketing Letters, Vol. 2, S. 199–214.
- Cattin, P.* (1981): Some Findings on the Estimation of Continuous Utility Functions in Conjoint Analysis. In: Advances in Consumer Research, Vol. 9, S. 367–372.
- Cattin, P.; Wittink, D.R.* (1982): Commercial Use of Conjoint Analysis: A Survey. In: Journal of Marketing, Vol. 46, S. 44–53.
- Chapman, R.G.; Staelin, R.* (1982): Exploiting Rank Ordered Choice Set Data Within the Stochastic Utility Model. In: Journal of Marketing Research, Vol. 19, S. 288–301.
- Dempster, A.P.; Laird, N.M.; Rubin, D.B.* (1977): Maximum Likelihood from Incomplete Data via the EM-Algorithm. In: Journal of the Royal Statistical Society, Series B39, S. 1–38.
- DeSarbo, W.S.; Cron, W.L.* (1988): A Maximum Likelihood Methodology for Clusterwise Linear Regression. In: Journal of Classification, Vol. 5, S. 249–282.
- DeSarbo, W.S.; Oliver, R.L.; Rangaswamy, A.* (1989): A Simulated Annealing Methodology for Clusterwise Linear Regression. In: Psychometrika, Vol. 54, S. 707–736.
- DeSarbo, W.S.; Ramaswamy, V.; Cohen, S.H.* (1995): Market Segmentation with Choice-Based Conjoint Analysis. In: Marketing Letters, Vol. 6, S. 137–147.
- DeSarbo, W.S.; Ramaswamy, V.; Chatterjee, R.* (1995): Analyzing Constant-Sum Multiple Criterion Data: A Segment-Level Approach. In: Journal of Marketing Research, Vol. 32, S. 222–232
- DeSarbo, W.S.; Wedel, M.; Vriens, M.; Ramaswamy, V.* (1992): Latent Class Metric Conjoint Analysis. In: Marketing Letters, Vol. 3, S. 273–288.
- Elrod, T.; Louviere, J.J.; Davey K.S.* (1992): An Empirical Comparison of Ratings-Based and Choice-Based Models. In: Journal of Marketing Research, Vol. 29, S. 368–377.
- Freter, H.* (1983): Marktsegmentierung, Verlag W. Kohlhammer, Stuttgart, Berlin, Köln.
- Green, P.E.* (1977): A New Approach to Market Segmentation, In: Business Horizons, Vol. 20, S. 61–73.
- Green, P.E.; Carroll, D.; Goldberg S.M.* (1981): A General Approach to Product Design Optimization via Conjoint Analysis. In: Journal of Marketing, Vol. 45, S. 17–37.

- Green, P.E.; Krieger, A.M.* (1991): Segmenting Markets with Conjoint Analysis. In: *Journal of Marketing*, Vol. 55, S. 20–31.
- Green, P.E.; Krieger, A.M.* (1993): Conjoint Analysis with Product Positioning Applications. In: *Eliashberg, J.; Lilien, G.J.: Handbooks in Operations Research and Management Science*, Vol. 5, North Holland, S. 467–516.
- Green, P.E.; Srinivasan, V.* (1978): Conjoint Analysis in Consumer Research: Issues and Outlook. In: *Journal of Consumer Research*, Vol. 5, S. 103–123.
- Green, P.E.; Srinivasan, V.* (1990): Conjoint Analysis in Consumer Marketing: New Developments with Implications for Research and Practice. In: *Journal of Marketing*, Vol. 54, S. 3–19.
- Hagerty, M.R.* (1985): Improving the Predictive Power of Conjoint Analysis: The Use of Factor Analysis and Cluster Analysis. In: *Journal of Marketing Research*, Vol. 22, S. 168–184.
- Haley, R.I.* (1968): Benefit Segmentation: A Decision-Oriented Research Tool, In: *Journal of Marketing*, Vol. 32, S. 30–35.
- Haley, R.I.* (1984): Benefit Segments: Backwards and Forwards. In: *Journal of Advertising Research*, Vol. 24, S. 19–25.
- Jain, A.K.; Malhotra, N.K.; Mahajan, V.* (1979): A Comparison of Internal Validity of Alternative Parameter Estimation Methods in Decompositional Multiattribute Preference Models. In: *Journal of Marketing Research*, Vol. 16, S. 313–322.
- Kamakura, W.A.* (1988): A Least Squares Procedure for Benefit Segmentation with Conjoint Experiments. In: *Journal of Marketing Research*, Vol. 25, S. 157–167.
- Kamakura, W.A., Wedel, M.; Agrawal, J.* (1994): Concomitant Variable Latent Class Models for Conjoint Analysis. In: *International Journal of Research in Marketing*, Vol. 11, S. 451–464.
- Katahira, H.* (1987): A Discrete Choice Model and Heterogeneous Preferences. In: *Journal of Econometrics (University of Tokyo)*, Vol. 53, S. 31–45.
- Louviere, J.J.* (1991): Experimental Choice Analysis: Introduction and Review. In: *Journal of Business Research*, Vol. 23, S. 291–297.
- Louviere, J.J.; Woodworth, G.* (1983): Design and Analysis of Simulated Consumer Choice or Allocation Experiments: An Approach Based on Aggregate Data. In: *Journal of Marketing Research*, Vol. 20, S. 350–367.
- Maier, G.; Weiss, P.* (1998): *Modelle diskreter Entscheidungen*, Springer Verlag, Wien, New York.
- Moore, W.I.* (1980): Levels of Aggregation in Conjoint Analysis: An Empirical Comparison. In: *Journal of Marketing Research*, Vol. 17, S. 516–523.
- Ogawa, K.* (1987): An Approach to Simultaneous Estimation and Segmentation in Conjoint Analysis. In: *Marketing Science*, Vol. 6, S. 66–81.
- Smith, W.R.* (1956): Product Differentiation and Market Segmentation as Alternative Marketing Strategies. In: *Journal of Marketing*, Vol. 21, S. 3–8.
- Späth, H.* (1979): Clusterwise Linear Regression. In: *Computing*, Vol. 22, (4), S. 367–373.
- Späth, H.* (1981): Clusterwise Linear Regression. In: *Computing*, Vol. 26 (26), S. 275.

- Späth, H.* (1982): A Fast Algorithm for Clusterwise Linear Regression. In: *Computing*, Vol. 29, S. 175–181.
- Srinivasan V.; Jain, A.K.; Malhotra, N.K.* (1983): Improving Predictive Power of Conjoint Analysis by Constrained Estimation, In: *Journal of Marketing Research*, Vol. 20, S. 433–438.
- Stallmeier, C.* (1993): Die Bedeutung der Datenerhebungsmethode und des Untersuchungsdesigns für die Ergebnisstabilität der Conjoint-Analyse, Regensburg.
- Teichert, T.* (1998): Schätzgenauigkeit von Conjoint-Analysen. In: *Zeitschrift für Betriebswirtschaft*, Vol. 68, S. 1245–1266.
- Teichert, T.* (2000a): Auswirkungen von Verfahrensalternativen bei der Erhebung von Präferenzurteilen. In: *Marketing ZFP*, Heft 2, S. 145–160.
- Teichert, T.* (2000b): Das Latent-Class Verfahren zur Segmentierung von wahlbasierten Conjoint-Daten: Befunde einer empirischen Anwendung. In: *Marketing ZFP*, Heft 3, S. 227–239.
- Vriens, M.; Oppewal, H.; Wedel, M.* (1998): Rating-Based versus Choice-Based Latent Class Conjoint Models: An Empirical Comparison. In: *Journal of the Market Research Society*, Vol. 40, S. 237–248.
- Vriens, M.; Wedel, M.; Wilms, T.* (1996): Metric Conjoint Segmentation Methods: A Monte Carlo Comparison. In: *Journal of Marketing Research*, Vol. 33, S. 73–85.
- Wedel, M.; Kamakura, W.A.* (1998): *Market Segmentation. Conceptual and Methodological Foundations.* International Series in Quantitative Marketing, Kluwer Academic Publishers, Boston, Dordrecht, London.
- Wedel, M.; Kistemaker, C.* (1989): Consumer Benefit Segmentation Using Clusterwise Linear Regression. In: *International Journal of Research in Marketing*, Vol. 6, S. 45–59.
- Wedel, M.; Steenkamp, J-B.* (1989): A Fuzzy Clusterwise Regression Approach to Benefit Segmentation. In: *International Journal of Research in Marketing*, Vol. 6, S. 241–258.
- Wind, Y.* (1978): Issues and Advances in Segmentation Research. In: *Journal of Marketing Research*, Vol. 15, S. 317–337.
- Wittink, D.R.; Cattin, P.* (1981): Alternative Estimation Methods for Conjoint Analysis: A Monte Carlo Study. In: *Journal of Marketing Research*, Vol. 28, S. 101–106.
- Wittink, D.R.; Cattin, P.* (1989): Commercial Use of Conjoint Analysis: An Update. In: *Journal of Marketing*, Vol. 53, S. 91–96.
- Wittink, D.R.; Vriens, M.; Burhenne, W.* (1994): Commercial Use of Conjoint Analysis in Europe: Results and Critical Reflections. In: *International Journal of Research in Marketing*, Vol. 11, S. 41–52.
- Zwerina, K.* (1997): *Discrete Choice Experiments in Marketing*, Physika Verlag, Heidelberg.
-

Anmerkungen

- ¹ Vgl. dazu insbesondere die empirischen Studien von Cattin/Wittink (1982), Wittink/Cattin (1989), Wittink et al. (1994) und Baier (1999) für den amerikanischen, europäischen bzw. deutschen Markt.
- ² Vgl. Wittink/Cattin (1989); Wittink et al. (1994). Nach den Recherchen von Baier (1999, S. 204) ist die Marktsegmentierung mittlerweile die am häufigsten involvierte Zielsetzung in kommerziellen Conjoint-Studien.
- ³ Zu Vor- und Nachteilen unterschiedlicher Erhebungsdesigns vgl. beispielsweise Stallmeier (1993, S. 45 ff.); Backhaus (2000, S. 571 ff.).
- ⁴ Vgl. dazu beispielsweise Stallmeier (1993, S. 58).
- ⁵ Einen Überblick über Skalierungsmöglichkeiten in diesem Zusammenhang gibt Stallmeier (1993, S. 60 ff.).
- ⁶ Vgl. dazu Green/Srinivasan (1978, S. 111 f.).
- ⁷ Vgl. Louviere/Woodworth (1983); Batsell/Louviere (1991); Louviere (1991).
- ⁸ Vgl. Anderson et al. (1992, S. 38 f.); Backhaus et al. (2000, S. 612).
- ⁹ Zu den möglichen Auswirkungen alternativer Erhebungsformen von Präferenzurteilen (metrisch, ordinal, dichotom) vgl. Teichert (2000a).
- ¹⁰ Green/Srinivasan (1978, 1990) diskutieren die in Abhängigkeit vom jeweiligen Skalierungsniveau der erhobenen Präferenzdaten anwendbaren Schätzmethoden. Teichert (1998) geht ausführlich auf die Problematik der Verwendung metrischer Schätzverfahren bei Vorliegen nichtmetrischer Rangurteile ein.
- ¹¹ Die Modellformulierung sieht in der vorliegenden Form die Möglichkeit der Präsentation individuell unterschiedlicher Stimulussets für die Befragungspersonen vor. Vgl. dazu auch Baier (1999).
- ¹² Einen Vergleich von rating- und choice-basierter Conjoint-Analyse nehmen Elrod et al. (1992) vor.
- ¹³ Zur Eignung der OLS-Regression bei Rating- bzw. Rankingdaten vgl. Wittink/Cattin (1981) bzw. Jain et al. (1979); Teichert (1998).
- ¹⁴ Vgl. beispielsweise Wedel/Kistemaker (1989, S. 46); Wedel/Steenkamp (1989, S. 242).
- ¹⁵ Vgl. Cattin (1981); Srinivasan et al. (1983).
- ¹⁶ Vgl. Green et al. (1981); Green/Krieger (1991, S. 29); Green/Krieger (1993, S. 482 ff.); Kamakura/Wedel (1998, S. 305).
- ¹⁷ Vgl. Backhaus et al. (2000, S. 590).
- ¹⁸ Vgl. Teichert (2000b, S. 228). Vgl. Zwerina (1997, S. 75 ff.) zu Möglichkeiten, individuelle Teilnutzenwerte auch bei wahlbasierten Modellen zu schätzen.
- ¹⁹ Vgl. dazu insbesondere Maier/Weiss (1990, S. 111 ff.). Louviere/Woodworth (1983, S. 354 ff.) schlagen ein gewichtetes Kleinstquadratverfahren vor.
- ²⁰ Vgl. beispielsweise Moore (1980, S. 521); Zwerina (1997, S. 85 ff.).
- ²¹ Vgl. beispielsweise Freter (1983, S. 16 ff.).
- ²² Vgl. Wind (1978).
- ²³ Vgl. beispielsweise Green/Krieger (1991, S. 22).

-
- ²⁴ Vgl. Moore (1980, S. 521); Green/Krieger (1991, S. 21); DeSarbo et al. (1992, S. 274).
- ²⁵ Vgl. beispielsweise Haley (1968, 1984); Wind (1978).
- ²⁶ Vgl. beispielsweise Green (1977); Green/Krieger (1991, S. 21); DeSarbo et al. (1992, S. 274); Wedel/Kamakura (1998, S. 25).
- ²⁷ Vgl. Wedel/Kamakura (1998, S. 24 f.); Backhaus (2000, S. 591).
- ²⁸ Vgl. Kamakura (1988, S. 158 ff.); Wedel/Kistemaker (1989, S. 46); Wedel/ Steenkamp (1989, S. 242); DeSarbo et al. (1992, S. 274).
- ²⁹ Vgl. Hagerty (1985, S. 171).
- ³⁰ Vgl. Hagerty (1985, S. 171 f. und S. 183). Damit liefert die Hauptkomponentenlösung eine höhere Prognosegüte als eine disjunkte Segmentlevel-Lösung und eine Individual-level-Analyse, was von Hagerty auch empirisch demonstriert wird. Vgl. Hagerty (1985, S. 173 ff.).
- ³¹ Vgl. dazu beispielsweise Backhaus et al. (2000, S. 291).
- ³² Vgl. Kamakura (1988, S. 158).
- ³³ Vgl. Kamakura (1988, S. 160).
- ³⁴ Vgl. Kamakura (1988, S. 166); Wedel/Steenkamp (1989, S. 246).
- ³⁵ Vgl. Wedel/Kamakura (1998, S. 312).
- ³⁶ Die Startklassifikation kann hier durch die zufällige Wahl der Klassenzugehörigkeiten der Individuen oder etwa auch ausgehend vom Ergebnis einer vorgeschalteten disjunkten Clusterwise Regression (bzw. Clusterung der Präferenzurteile) und anschließender Addition eines Zufallsfehlers zu den 0-1-Zugehörigkeiten festgelegt werden. Vgl. Wedel/Steenkamp (1989, S. 247 ff.).
- ³⁷ Vgl. Wedel/Steenkamp (1989, S. 246); Wedel/Kamakura (1998, S. 25).
- ³⁸ Empfehlungen für eine geeignete Wahl des Unschärfeparameters δ geben Wedel/Steenkamp (1989, S. 244 und S. 256 f.).
- ³⁹ Ausdruck (9) ergibt sich unter Berücksichtigung der Summenrestriktion für die individuellen Klassenzugehörigkeiten aus den Bedingungen erster Ordnungen des Lagrange-Ansatzes zur Minimierung des gewichteten Kleinst-Quadrate-Ansatzes (8). Vgl. Wedel/Steenkamp (1989, S. 257).
- ⁴⁰ Vgl. Wedel/Steenkamp (1989, S. 257).
- ⁴¹ Vgl. Wedel/Steenkamp (1989, S. 246).
- ⁴² Das Modell läßt in dieser Form auch Korrelationen der Fehlerterme bezüglich der Stimulusbewertungen eines Individuums zu und stellt somit eine Verallgemeinerung der gewöhnlichen Kleinstquadrateschätzung zur Bestimmung der Teilnutzenwerte dar. Vgl. Wedel/Kamakura (1989, S. 312).
- ⁴³ Vgl. DeSarbo et al. (1992, S. 276 f.).
- ⁴⁴ Vgl. zum EM-Algorithmus auch Dempster et al. (1977); DeSarbo/Cron (1988); Wedel/Kamakura (1998, S. 120 ff.).
- ⁴⁵ Vgl. dazu DeSarbo et al. (1992, S. 287).
- ⁴⁶ Vgl. DeSarbo et al. (1995, S.139 f.).
- ⁴⁷ Vgl. dazu Chapman/Staelin (1982).
- ⁴⁸ Vgl. beispielsweise DeSarbo/Ramaswamy/Chatterjee (1995).

-
- ⁴⁹ Vgl. dazu analog die Kritik an der zweistufigen Vorgehensweise zur Segmentierung und Schätzung von Teilnutzenwerten in Abschnitt C.1 sowie Kamakura et al. (1994, S. 453).
- ⁵⁰ Vgl. Wedel/Kamakura (1998, S. 148).
- ⁵¹ Vgl. Kamakura et al. (1994, S. 454).
- ⁵² Vgl. Kamakura et al. (1994, S. 458 ff.).
- ⁵³ Vgl. Abschnitt C.2.4.
- ⁵⁴ Während die einfachere Interpretation bei eindeutiger Zuordnung der Individuen zu Segmenten zugunsten einer disjunkten Segmentlösung spricht, wird unscharfen oder überlappenden Segmenten eine größere Realitätsnähe zugeschrieben. Vgl. dazu beispielsweise Vriens et al. (1996, S. 76).
- ⁵⁵ Vgl. Wedel/Steenkamp (1989 S. 245); Wedel/Kistemaker (1989, S. 48 f.).
- ⁵⁶ Vgl. Wedel/Steenkamp (1989 S. 251).