

**Heterogenität im Antwortverhalten bei
Kundenzufriedenheitsbefragungen - Ein bayesianischer
Mixture-Modell-Ansatz**

Dissertation

zur Erlangung des Grades

Doktor der Wirtschaftswissenschaft (Dr. rer. pol.)

der Wirtschaftswissenschaftlichen Fakultät Ingolstadt
der Katholischen Universität Eichstätt-Ingolstadt

Datum der mündlichen Prüfung: 7.12.2011

Erstgutachter:

Prof. Dr. Joachim Büschken
Katholische Universität Eichstätt-Ingolstadt

Zweitgutachter:

Prof. Dr. Ulrich Küsters
Katholische Universität Eichstätt-Ingolstadt

vorgelegt von

Dipl.-Kfm. Nino Hardt

Vorwort

[...] the only good statistics is Bayesian statistics. (LINDLEY, 1975, S. 106)

So überzeugt Lindley die bayesianische Statistik vertritt, so spannend und herausfordernd ist der Lernprozess, der mit der Aufnahme meiner Tätigkeit am Lehrstuhl von Professor Dr. Büschken im Jahr 2007 begann. Schon zuvor hatte ich mich mit Fragen der Modellierung im Marketing beschäftigt. Dennoch konnte ich nur erahnen, welche Erkenntnisse noch vor mir liegen würden.

Es hat sich gelohnt. Der Einzug anspruchsvoller quantitativer Methoden in die Betriebswirtschaftslehre ist noch verhältnismäßig jung. Umso spannender ist es, in diesem Bereich zu forschen. Erst nach dem fordernden Weg durch die bayesianische Statistik habe ich einen wirklichen Zugang zur Statistik sowie den aktuellen Publikationen des quantitativen Marketing gefunden. Es macht Spaß!

Ganz besonders danken möchte ich hiermit meinem Doktorvater Professor Dr. Büschken für die stetige Betreuung und hervorragende Qualifizierung. Er hat es mir ermöglicht, in der internationalen Forschung tätig zu sein.

Professor Dr. Küsters möchte ich nicht nur für die Übernahme des Koreferates danken, sondern auch dafür, dass er mir schon im Diplomstudium wichtige Grundlagen vermittelt hat. Sehr gerne erinnere ich mich vor allem an den R-Grundlagenkurs.

Mein spezieller Dank gilt auch Professor Dr. Thomas Otter für seinen lehrreichen Kurs zur bayesianischen Statistik, an dem unser Lehrstuhlteam per Videokonferenz teilnehmen durfte. Mir sind einige Zusammenhänge dadurch noch einmal viel klarer geworden.

Dem Lehrstuhlteam möchte ich für die hervorragende Atmosphäre und Zusammenarbeit danken.

Darüber hinaus möchte ich allen danken, die mich fachlich, persönlich oder auf andere Art und Weise unterstützt und motiviert haben.

Ingolstadt, Dezember 2011

Inhaltsverzeichnis

Abbildungsverzeichnis	V
Tabellenverzeichnis	VI
Abkürzungsverzeichnis	VII
Symbolverzeichnis	VIII
1. Heterogene Kausalstrukturen bei Kundenzufriedenheitsbefragungen	1
1.1. Zielsetzung der Arbeit	2
1.2. Gang der Untersuchung	4
2. Herausforderungen multiattributiver Modelle	6
2.1. Treiberanalyse	6
2.1.1. Direkte Abfrage oder Berechnung von Bedeutungsgewichten	7
2.1.2. Experimente	8
2.1.3. Regressionsanalyse	9
2.1.4. Strukturgleichungsmodelle	10
2.2. Response Bias	14
2.3. Der Halo-Effekt	17
2.3.1. Definition des Halo	17
2.3.2. Beurteilungsprozesse und Halo	22
2.3.3. Operationalisierung des Halo	23
2.3.4. Nachweise des Halo-Effektes	26
2.3.5. Bisherige Ansätze zur Korrektur des Halo-Effektes	29
2.4. Zwischenfazit	31
2.4.1. Herausforderungen	31
2.4.2. Forschungsbedarf	32
2.4.3. MIMIC-Modelle	34
3. Anwendung am Beispiel Kundenzufriedenheit	37
3.1. Zufriedenheit als Ziel von Unternehmen	37

3.1.1.	Zusammenhang mit Unternehmenserfolg	38
3.1.2.	Loyalität und Kundenbindung	40
3.1.3.	Kundenzufriedenheit und wertorientiertes Management	41
3.2.	Definition von Kundenzufriedenheit	42
3.2.1.	Konzeptionelle Grundlagen	44
3.2.2.	Unterschiede der Definitionen	47
3.2.3.	Abgrenzung zu anderen Konstrukten	51
3.2.4.	Definition für diese Arbeit	53
3.3.	Verfahren zur Messung von Kundenzufriedenheit	54
3.3.1.	Differenzierung unterschiedlicher Verfahren	55
3.3.2.	Zufriedenheitsskalen	60
3.3.3.	Zufriedenheitsindizes	64
4.	Entwicklung eines Mixture-Modells von Kundenzufriedenheit	66
4.1.	Mixture von Kausalstrukturen	66
4.1.1.	Heterogenität in Strukturgleichungsmodellen	67
4.1.2.	Alternative Kausalmodelle	68
4.1.3.	MIMIC-Strukturen	70
4.1.4.	Modell-Likelihood	72
4.2.	Bayesianische Inferenz und MCMC	75
4.2.1.	Identifikation	78
4.2.2.	MCMC Methoden	82
4.2.3.	Gibbs-Sampler	84
4.2.4.	Metropolis-Hastings Algorithmus	85
4.2.5.	Adaptive Rejection Sampling	85
4.3.	Strukturgleichungsmodelle	86
4.3.1.	Modellierung	87
4.3.2.	Messmodelle	88
4.3.3.	Konstrukte mit einem Indikator	89
4.3.4.	Äquivalente Modelle	92
4.3.5.	Kausalität	94

4.4.	Ordinale Daten	96
4.4.1.	Probit-Modell	97
4.4.2.	Skalenverwendungsheterogenität	98
4.5.	Schätzung des Mixture-Modells	103
4.6.	Beziehungen zu verwandten Verfahren	105
4.7.	Implementierung	105
4.8.	Simulationsstudie	106
5.	Empirische Anwendung	111
5.1.	Aufgabenstellung der empirischen Anwendung	111
5.2.	Beschreibung des Datensatzes	113
5.3.	Modellaufbau	116
5.4.	Schätzergebnisse	119
6.	Diskussion der Ergebnisse	133
6.1.	Zum Beitrag dieser Arbeit	133
6.2.	Diskussion der empirischen Ergebnisse	135
6.3.	Implikationen für Kundenzufriedenheit	136
6.4.	Einschränkungen	137
6.5.	Anregungen für weitere Forschung	138
	Literatur	140
A.	Anhang	175
A.1.	MCMC Sampler	175
A.1.1.	Latente Factor Scores	175
A.1.2.	LISREL-Modellparameter	175
A.1.3.	Cut-Point-Modell	178
A.1.4.	Skalenverwendungsheterogenitätsmodell	178
A.1.5.	Finite Mixture	181
A.2.	Prior-Spezifikation	181
A.3.	Parameter-Recovery Simulationsstudie	182

A.4. Predictive Fit	187
A.5. Zusammenfassungen der Posteriori-Verteilungen aller Parameter	187

Abbildungsverzeichnis

2.1. Drei unterschiedliche Halo-Modelle	20
2.2. Eine beispielhafte MIMIC-Struktur	35
3.1. Systematisierung unterschiedlicher Messverfahren	56
3.2. Das ACSI-Modell	65
4.1. Heterogenität in Strukturgleichungsmodellen	68
4.2. MIMIC-Struktur Bsp. 2	70
4.3. DAG	74
4.4. Äquivalenz in SEMs	93
4.5. Cut-Point-Modell im bivariaten Fall	99
4.6. Skalenverwendungsheterogenität	100
4.7. Regressions- und Halo-Struktur Bsp.	108
5.1. Korrelationen Modell 1	123
5.2. Zuordnungswahrscheinlichkeiten in Modell 3	125
5.3. Histogramm der Summe der Diagonalwahrscheinlichkeiten	127
5.4. Korrelationen Modell 2	128
5.5. Korrelationen Modell 3 - Halo- und Regressionskomponente	129
5.6. Korrelationen Modell 3 - MIMIC-Komponente	130
5.7. Korrelationen Modell 4 - MIMIC-Komponente	130
5.8. Korrelationen Modell 4	132
A.1. Predictive Likelihood	187

Tabellenverzeichnis

4.1. Korrelation durch Skalenverwendungsheterogenität	102
4.2. Simulationsstudie: Korrelation der latenten Faktoren (Simulationsstudie) .	109
4.3. Simulationsstudie: Ergebnisse Teil 1	110
5.1. Deskriptive Statistiken der Zufriedenheitsdaten	114
5.2. Zahl der abgegebenen Bewertungen	114
5.3. Spearman-Korrelation der Rohdaten	115
5.4. MIMIC-Varianten	117
5.5. Modellvergleich	120
5.6. Treiber in der Regressionskomponente	121
5.7. Ergebnisse Modell 3 - MIMIC-Koeffizienten	131
A.1. Recovery Messmodell	183
A.2. Recovery Strukturfehler und Kovarianzen exogener Konstrukte	185
A.3. Ergebnisse Modell 1	188
A.4. Ergebnisse Modell 2	191
A.5. Ergebnisse Modell 3	195

Abkürzungsverzeichnis

ACSI American Customer Satisfaction Index

ANOVA Analysis of Variance

Bsp. Beispiel

C/D-Paradigma Confirmation-Disconfirmation-Paradigma

CSM Customer Satisfaction Measurement

EAT Earnings After Tax

ECSI European Customer Satisfaction Index

GE Gesamteindruck

HB Hierarchical Bayes (Hierarchische Bayesianische Modelle)

IRT Item Response Theory

iid Independent and identically distributed - identisch verteilt

LISREL Linear Structural RELations

MH-Algorithmus Metropolis-Hastings-Algorithmus

MVN Multivariate Normalverteilung

NPS Net Promoter Score

PM Posterior Mittelwert

RMSEA Root Mean Square Error of Approximation

SEM Structural Equation Model (Strukturgleichungsmodell)

SR Structural Regression Model

Symbolverzeichnis

Dimensionen

- n : Zahl der observierten Fälle
- p : Zahl der Messvariablen
- q : Zahl der latenten Konstrukte, umfasst endogene (q_1) und exogene (q_2) Konstrukte

Modell

- ι : Spaltenvektor, bei dem alle Elemente 1 sind
- Π und Γ : bezeichnen die Koeffizientenmatrix zwischen endogenen bzw. exogenen und endogenen latenten Variablen, wobei Π einer Matrix der Dimension $q_1 \times q_1$ und Γ eine Matrix der Dimension $q_1 \times q_2$ ist.
- Ψ_δ : $q_1 \times q_1$ -Diagonalmatrix der Fehlervarianzen der endogenen latenten Variablen
- Φ : $q_2 \times q_2$ -Kovarianzmatrix der exogenen latenten Variablen
- Λ : $p \times q$ -Matrix der Messkoeffizienten
mit Elementen Λ_{ij} , wobei $i \in \{1 \dots p\}$, $j \in \{1 \dots q\}$
- Ψ_ε : $p \times p$ -Diagonalmatrix der Messfehlervarianzen
- μ_Z : Vektor der Mittelwerte der latenten kontinuierlichen Messvariablen (Länge p)
- Z : $n \times p$ -Matrix der latenten kontinuierlichen Messvariablen
- \tilde{Z} : $n \times p$ -Matrix der latenten kontinuierlichen und Mittelwert zentrierten Messvariablen ($\tilde{Z} = Z - \iota\mu_Z^t$)
- Y : $n \times p$ -Matrix der ordinalen Messvariablen
- \mathbf{c} : Vektor der Cut-Points (Schwellenwerte)
- θ : umfasst alle LISREL-Modellparameter - $\theta = \{\Pi, \Gamma, \Psi_\delta, \Phi, \Lambda, \Psi_\varepsilon, \mu_z\}$
- Ω : $n \times p$ -Matrix der Factor Scores der latenten Konstrukte mit $\Omega = [\omega_1 \dots \omega_n]^T$
und $\omega_i = [\eta_i \ \xi_i]$ (besteht aus endogenen und exogenen Konstrukten)
- Σ : Varianz-Kovarianzmatrix der kontinuierlichen Messvariablen (Dimension $p \times p$)
- Σ_ω : Varianz-Kovarianzmatrix der Factor Scores (Dimension $q \times q$)
- $\{\tau_i\}$: individuelle Mittelwerte der Skalennutzung, $i \in \{1 \dots n\}$
- $\{\sigma_i\}$: individuelle Standardabweichungen der Skalennutzung, $i \in \{1 \dots n\}$

- σ_τ : Verteilungsparameter der individuellen Mittelwerte ($\tau_i \sim N(0, \sigma_\tau)$)
- a_σ : Verteilungsparameter der individuellen Standardabweichungen ($\sigma_i^2 \sim \Gamma(a_\sigma, b_\sigma)$)
- b_σ : festzulegender Verteilungsparameter der individuellen Standardabweichungen, wird nicht geschätzt
- K : Zahl der Komponenten im Finite Mixture-Modell
- IG : Inverse Gamma-Verteilung

Sonstige

- $E(x)$: Erwartungswert von x
- $S(x)$: Standardabweichung von x

1. Heterogene Kausalstrukturen bei Kundenzufriedenheitsbefragungen

Im Marketing und anderen Sozialwissenschaften ist die Anwendung von Strukturgleichungsmodellen zur Analyse von Kausalmodellen gängige Praxis (STEENKAMP und BAUMGARTNER, 2000; DILLON et al., 1997). Abhängig davon, ob man der kausalen Interpretierbarkeit zustimmt (PEARL, 2010a) oder sie ablehnt (KLINE, 2011), spricht man auch von Kausalmodellen bzw. Kovarianzstrukturmodellen. Einen sehr flexiblen Ansatz zur Modellierung von Strukturgleichungen stellt das LISREL-Modell dar (JÖRESKOG und SÖRBOM, 1996). Dieses Modell besteht aus zwei Bestandteilen:

1. der Operationalisierung latenter Konstrukte durch Messmodelle sowie
2. einem Strukturmodell, welches Aussagen zu linearen Beziehungen zwischen endogenen und exogenen latenten Konstrukten abbildet

Kausalmodelle werden eingesetzt, um Zusammenhänge zwischen unterschiedlichen Konstrukten zu untersuchen. In der Praxis werden sehr häufig Zusammenhänge zwischen Teil- und Gesamtzufriedenheiten untersucht, um relevante Treiber von Zufriedenheit zu identifizieren. Kundenzufriedenheit steht bei den meisten Unternehmen weit oben auf der Liste strategischer Managementziele. Wie sich aus Zufriedenheitsbefragungen Erkenntnisse und Rückschlüsse zur Optimierung von Produkten, Prozessen und Serviceleistungen des Unternehmens gewinnen lassen, ist Gegenstand zahlreicher Untersuchungen. Eine Konzentration auf relevante Treiber von Zufriedenheit ermöglicht einen besseren Einsatz knapper Ressourcen. Der Erfolg von Unternehmen hängt daher auch von der Qualität der Analysen ab.

Ziel solcher Modelle ist vor allem die Inferenz bezüglich des Strukturmodells - es beschreibt die Zusammenhänge, welche von primärem Interesse sind.

Obwohl es unterschiedliche Ansätze zum Umgang mit Heterogenität in Strukturgleichungsmodellen gibt (hierarchische Modelle (ANSARI et al., 2000), Finite Mixture Modelle (ZHU und LEE, Mar)), wird in bisherigen Publikationen von homogenen kausalen Strukturen ausgegangen. Während solche Ansätze annehmen, dass die Stärke der (kausalen) Zusammenhänge heterogen ist, berücksichtigen sie nur jeweils eine Theorie zur Struk-

tur dieser Zusammenhänge. Dies manifestiert sich darin, dass die gleichen Aussagen zur Unabhängigkeit von Konstrukten getroffen werden.

Existieren konkurrierende Theorien, so führt dies in i. d. R. zu unterschiedlichen Aussagen über partielle Korrelationen. Dies lässt sich anhand eines Beispiels illustrieren: Das ACSI-Kundenzufriedenheitsmodell (American Customer Satisfaction Index) (FORNELL et al., 1996a) trifft u. a. folgende Annahmen:

1. wahrgenommener Wert treibt Kundenzufriedenheit,
2. Kundenzufriedenheit treibt Loyalität und
3. gegeben 1) und 2) sind wahrgenommener Wert und Loyalität unabhängig

In bisherigen Untersuchungen wurde Heterogenität nur bezüglich der Stärke von Zusammenhängen zwischen Konstrukten berücksichtigt. Betroffen ist also nur die Intensität der Annahmen 1) und 2). Unterschiedliche Aussagen zur bedingten Unabhängigkeit werden jedoch nicht berücksichtigt, da unterschiedliche Theorien zur Begründung notwendig sind. Eine konkurrierende Theorie betrifft Annahme 3 und könnte beinhalten, dass es auch einen direkten Einfluss des wahrgenommenen Wertes auf die Loyalität gibt.

1.1. Zielsetzung der Arbeit

Ziel dieser Arbeit ist es, heterogene Kausalstrukturen explizit zu modellieren. Dabei werden konkurrierende Theorien zur Entstehung von Kundenzufriedenheitsdaten betrachtet, welche eine Modellierung heterogener Kausalstrukturen motivieren. Darüber hinaus sind auch andere Anwendungsfelder denkbar - etwa multiattributive Einstellungsmodelle in der Markenforschung. Bei jedem Strukturgleichungsmodell sollte generell zunächst überprüft werden, ob nicht auch andere plausible Theorien existieren, die zu alternativen Kausalstrukturen führen. Durch die gleichzeitige Modellierung konkurrierender Theorien wird sichtbar, welche Theorien durch die beobachteten Daten am besten unterstützt werden. Wird hingegen nur eine Kausalstruktur modelliert, können die Ergebnisse durch die Daten des „falschen“ Datengenerierungsprozesses verzerrt werden.

Anwendung findet das Modell in dieser Arbeit bei der Kundenzufriedenheitstreiberanalyse.

Kundenzufriedenheit gilt als Indikator für den Erfolg von Unternehmen (MATZLER und BAILOM, 2006, S. 243). Man nimmt an, dass zufriedene Kunden loyaler sind (JONES und SASSER, 1995), sodass Unternehmen mit bestehenden Kunden länger Profit erwirtschaften können.

Aus diesem Grund besteht ein zentrales Erkenntnisinteresse bezüglich der Treiber von Kundenzufriedenheit (MCDUGALL und LEVESQUE, 2000), auch wenn das Potential von Treiberanalysen von einigen Unternehmen noch unterschätzt wird (KATCHER, 2003). Die Identifikation relevanter Treiber von Kundenzufriedenheit ist jedoch wichtig, um zielgerichtete Investitionen zu tätigen, die die Kundenzufriedenheit erhöhen, und um Fehlallokationen von Budgets zu vermeiden. Der Erfolg von Managemententscheidungen hängt daher auch von der Qualität der Treiberanalysen ab.

Die rechnerische Bestimmung von Kundenzufriedenheitstreibern ist der direkten Frage nach Treiber überlegen, da eine direkte Abfrage aus unterschiedlichen Gründen kaum Informationen liefert (KANNING und BERGMANN, 2009, S. 388).

Bei der Berechnung von Zufriedenheitstreibern gibt es jedoch eine Reihe von Schwierigkeiten, die überwunden werden müssen:

- *Multikollinearität*, welche die Durchführung von Regressionsanalysen erschwert (z. B. FINN, 2011, S. 30) und möglicherweise auf unterschiedliche Formen des Response Bias zurückzuführen ist
- *Response Bias*, d. h. die Verzerrung der Zufriedenheitsdaten durch individuell unterschiedlichen Umgang mit Zufriedenheitsfragebögen (PAULHUS, 1991, S. 17)
- Heterogene Kausalstrukturen - abweichend vom Treibermodell, welches unabhängige Bewertungen von Attributen und eine daraus „berechnete“ Gesamtzufriedenheit annimmt, kann die Bewertung aller Attribute allein durch einen Gesamteindruck bestimmt werden. Der Effekt ist in der Literatur als „*Halo-Effekt*“ bekannt (THORNDIKE, 1920). Dabei handelt es sich um eine alternative Theorie zum kognitiven Prozess der Bewertung von Teilzufriedenheiten.

Das zentrale Forschungsziel dieser Arbeit besteht in der Untersuchung von heterogenen Kausalstrukturen. Bei der Treiberanalyse existieren konkurrierende Theorien zum Datengenerierungsprozess der Zufriedenheitsdaten: 1) Ein Treibermodell, welches annimmt,

dass Gesamtzufriedenheit in einem formativen Prozess aus der Gewichtung von Teilzufriedenheiten entsteht und 2) Ein Halo-Modell, welches annimmt, dass sich alle Teilzufriedenheitsbewertungen aus einem Gesamteindruck ergeben. Während diese beiden Komponenten von BÜSCHKEN et al. (2010) adressiert werden, geht diese Arbeit in zwei Punkten darüber hinaus:

1. Es wird eine weitere alternative Kausalstruktur modelliert: diese Struktur entspricht einem Multiple Indicator Multiple Causes-Modell (MIMIC). Im Mixture-Modell kombiniert die *MIMIC*-Komponente Elemente des Treiber- und Halo-Modells. Damit wird die harte Annahme gelockert, dass entweder alle Teilzufriedenheiten differenziert oder alle Zufriedenheiten nur durch einen Gesamteindruck dominiert werden.
2. Das Modell basiert auf dem LISREL-Modell und ist somit in der Lage, Messmodelle zu berücksichtigen. Dabei kann es sich zum einen um psychologische Messkonstrukte oder aber um inhaltlich verwandte Items (sog. Domänen) handeln. Insbesondere bei sehr umfangreichen Fragebögen (etwa mit 20 oder mehr Fragen) lassen sich die Fragen einer geringeren Zahl von Domänen zuordnen.

1.2. Gang der Untersuchung

In **Kapitel 2** wird zunächst auf die Herausforderungen multiattributiver Modelle, insbesondere der Treiberanalyse, eingegangen. Die gängigen Methoden der Bestimmung von Attributsgewichten werden kurz erläutert. Anschließend werden mögliche Beeinträchtigungen durch verschiedene Formen des Response Bias erläutert. Schließlich wird auf den Halo-Effekt eingegangen, der eine konkurrierende Theorie zur Kausalstruktur der Datengenerierung in multiattributiven Modellen motiviert.

Die Untersuchung von Kundenzufriedenheitsdaten eignet sich aufgrund der hohen Relevanz im Marketing sowie der Existenz konkurrierender Theorien, um das in dieser Arbeit entworfene Modell zu demonstrieren. In **Kapitel 3** wird auf die Relevanz des Themas Kundenzufriedenheit, definitorische Grundlagen sowie typische Messmethoden eingegangen.

In **Kapitel 4** wird ein bayesianisches Mixture-Modell vorgestellt, welches zwei bisher ungelöste Herausforderungen der Zufriedenheitsmodellierung adressiert:

1. die Modellierung, Identifikation und Korrektur des Halo-Effekts bei Daten aus Kundenzufriedenheitsbefragungen, welche umfangreiche Item-Batterien verwenden und
2. die Modellierung einer Antwortstruktur, welche den Halo-Fehler nur für einen Teil der Zufriedenheitsattribute annimmt

In diesem Kapitel werden dabei zunächst die methodischen Grundlagen bayesianischer Modelle erklärt. Anschließend werden die Bestandteile des Modells erläutert: das LISREL-Strukturgleichungsmodell, die Modellierung der ordinalen Daten sowie die Finite-Mixture-Modellierung von LISREL-Strukturen. Die Funktionsfähigkeit des Modells wird durch eine Simulationsstudie demonstriert.

In **Kapitel 5** wird das entworfene Modell schließlich auf Daten aus einer Kundenzufriedenheitsbefragung angewendet. Dabei wird in bedeutendem Ausmaß Haloing sowie teilweises Haloing festgestellt.

Die Ergebnisse werden in **Kapitel 6** diskutiert. Die Modelle werden hinsichtlich ihrer Anpassung an die Daten verglichen. Die Relevanz des Halo-Effektes wird ebenso diskutiert wie die Ergebnisse der neu eingeführten MIMIC-Komponente mit der Möglichkeit, partielles Haloing abzubilden. Die Vorteile des entworfenen Modells werden erläutert. Einschränkungen und Ideen für die weitere Forschung von Zufriedenheitstreiberanalysen werden erläutert und diskutiert.

2. Herausforderungen multiattributiver Modelle

Zunächst wird in Abschnitt 2.1 die zentrale Fragestellung multiattributiver (Kundenzufriedenheits-) Modelle erläutert: die Identifikation von Treibern bzw. Bedeutungsgewichten. In Abschnitt 2.2 wird auf die Gefahr der Verzerrung von Befragungsdaten durch unterschiedliche Formen des Response Bias eingegangen. In Abschnitt 2.3 werden Grundlagen und Auswirkungen des Halo-Effektes dargestellt. Der Halo-Effekt stellt die Grundlage für ein Beispiel kausaler Heterogenität dar, welches ein Schwerpunkt dieser Arbeit ist. In Abschnitt 2.4.3 wird eine weitere alternative Kausalstruktur vorgestellt, die im Rahmen dieser Arbeit untersucht wird: die MIMIC-Struktur.

2.1. Treiberanalyse

Für Unternehmen ist weniger ein absoluter Zufriedenheitswert als vielmehr die Kenntnis relevanter Treiber von Kundenzufriedenheit von Bedeutung. So bemängelt KATCHER (2003), dass Zufriedenheitsbefragungen oft nur als reines Performance-Controlling-Tool verwendet werden. Der wahre Wert solcher Befragungen liege aber vielmehr in der Identifikation von Zufriedenheitstreibern. Dies gilt auch für Dienstleistungsunternehmen, wie MCDUGALL und LEVESQUE (2000, S. 392) bemerken: „For service firms the challenge is identifying the critical factors that determine customer satisfaction and loyalty.“ Auf die Bedeutung von Kundenzufriedenheit für Unternehmen wird in Abschnitt 3.1 näher eingegangen.

Die Fragestellung einer Treiberanalyse ist zunächst einmal unabhängig davon, ob Zusammenhänge zwischen Leistung und Teilleistung, Kundenzufriedenheit und Teilzufriedenheit oder einem vergleichbaren Konstrukt und zugehörigen Attributen untersucht werden. Ziel ist es herauszufinden, welche Teilaspekte eines Produktes oder Services die Gesamtbewertung treiben.

So können Kunden Erwartungen bezüglich aller Teilaspekte des Produktes bilden, die nach Betrachtung der Leistung aller Teilaspekte zur Bestimmung von Teilzufriedenheiten herangezogen werden. Alternativ können Kunden auch direkt nach einer Einschätzung der Leistung gefragt werden.

Auf die konzeptionelle Unterscheidung von Erwartungen und Leistungen sowie das Confirmation-Disconfirmation-Paradigma wird in Abschnitt 3.2.1 näher eingegangen. Die Konsequenzen für die praktische Anwendung von Treiberanalysen sind jedoch vermutlich gering. So sehen KANNING und BERGMANN (2009, S. 386) keinen Mehrwert in der separaten Erfassung von Erwartungen und Leistungen.

Analoge Fragestellungen ergeben sich bei der Untersuchung von Einstellungen. So ist auch das Fishbein-Modell (FISHBEIN, 1963) ein multiattributives Modell, welches Einstellung als Summenprodukt von Überzeugungen und Gewichtungen darstellt. Die Problemstellung der Treiberanalyse gilt somit analog auch für andere Modelle wie z. B. gewichtete additive Nutzenmodelle oder Einstellungsmodelle (z. B. COHEN und AHTOLA, 1971; FISHBEIN, 1965).

2.1.1. Direkte Abfrage oder Berechnung von Bedeutungsgewichten

Bei Zweikomponentenverfahren (siehe Abschnitt 3.3.1) sollen Kunden sowohl die Bedeutung als auch den Grad der Erfüllung verschiedener Kriterien bewerten, sodass sich leicht die Verbesserungspotentiale erkennen lassen (HILL et al., 2002, S. 95). Werden nur (Teil-)Zufriedenheiten abgefragt (sogenannte Einkomponentenverfahren), dann werden oft Regressionsmodelle gerechnet, um die Gewichtung der einzelnen Merkmale in Erfahrung zu bringen (MEYER und KANTSPERGER, 2005, S. 225f.).

Zur Durchführung solcher Analysen ist es wichtig, dass Gesamtzufriedenheit separat abgefragt wird. Ohne diese abhängige Variable können ansonsten keine Regressionsmodelle gerechnet werden (VAVRA, 1997, S. 141f.).

Sofern Treiber nicht direkt abgefragt, sondern durch statistische Methoden identifiziert werden sollen, gibt es folgende Möglichkeiten:

1. *kontrollierte Experimente*: potentiell relevante Attribute eines Produktes/Services werden manipuliert. Unterschiede in der gemessenen Gesamtzufriedenheit lassen sich schließlich auf die manipulierten Attribute zurückführen.
2. *Regressionsmodelle bzw. SEM*: Aus der Bewertung von Teil- und Gesamtzufriedenheit werden die Gewichte rechnerisch bestimmt. Je nach Komplexität lassen sich dabei folgende Verfahren unterscheiden:

- a) *multiple Regression*: Teil- und Gesamtzufriedenheit werden direkt, d. h. ohne Messmodell operationalisiert. Zudem sind nur Beziehungen zwischen Teil- und Gesamtzufriedenheit möglich.
- b) *Strukturgleichungsmodelle* erlauben die Verwendung von Messmodellen sowie komplexen Strukturen, die z. B. auch Wiederkaufabsicht und Loyalität abbilden können.

In den folgenden Abschnitten werden diese drei Ansätze kurz vorgestellt.

2.1.2. Experimente

Um einen kausalen Zusammenhang zwischen verschiedenen Leistungskriterien und der Zufriedenheit herzustellen, eignen sich am besten Experimente. In diesem Fall können die Einflussfaktoren vom Forscher determiniert werden, sodass ein Rückschluss auf die Ursachen zulässig ist.

Anwendungen finden sich vor allem im Bereich von Dienstleistungen. WIRTZ und BATESON (1995) manipulieren in einem Experiment eine Leistungsdefinition und beobachten den Einfluss auf die übrigen Bewertungen einer Dienstleistung. McDONNELL (2007) untersucht die Auswirkung von Musik und Duft bei Wartezeiten von Dienstleistungen. KATZ et al. (1991) analysieren den Einfluss von Wartezeiten.

Besonders in früheren Publikationen wurden Experimente durchgeführt, um den Effekt der Manipulation bestimmter Aspekte des Beurteilungsprozesses zu untersuchen. Der Fokus liegt dabei aber i. d. R. nicht auf der Untersuchung von Attributsleistungen bzw. Attributzufriedenheiten und Gesamtzufriedenheit, sondern auf der von Erwartung, Leistung und Zufriedenheit, welche im Diskonfirmationsprozess zu Zufriedenheit oder Unzufriedenheit führen (z.B. CHURCHILL und SURPRENANT, 1982). Auch WIRTZ und BATESON (1999) setzen sich mit dem Einfluss von Erwartungen auf Zufriedenheit auseinander.

Experimente bieten sich jedoch weniger zur kontinuierlichen Überprüfung der Kundenzufriedenheit in der Praxis an. Unternehmen benötigen Informationen über potentielle Treiber in ihrem eigenen Unternehmen und können sich daher nicht nur auf allgemeine Untersuchungen (wie z. B. Experimente zu Wartezeit und Zufriedenheit) verlassen. Zudem müssen die zu manipulierenden Aspekte eines Produktes oder einer Dienstleistung

bekannt sein. Eine Manipulation aller Aspekte würde zu sehr umfangreichen Experimenten führen. In einem Fragebogen können hingegen leicht 20 unterschiedliche Aspekte von Produkten evaluiert werden. Auch ist die Durchführung von Experimenten im Gegensatz zu Befragungen teuer.

Die Durchführung von Experimenten ist nicht Teil dieser Arbeit, daher wird nicht näher auf diese Form der Untersuchung eingegangen.

2.1.3. Regressionsanalyse

Die Durchführung von Experimenten setzt Hypothesen über relevante Treiber voraus. Jede potentiell wirksame Eigenschaft müsste bewusst modelliert und manipuliert werden.

Bestehen noch keine Hypothesen oder sollen über bestehende Hypothesen hinaus neue Zusammenhänge von Teilzufriedenheiten und Gesamtzufriedenheit aufgedeckt werden, bietet sich der Einsatz von Experimenten nicht an.

In diesen Fällen bietet sich der Einsatz von Kundenzufriedenheitsbefragungen an. Um Regressionsanalysen durchzuführen, muss man sowohl Gesamtzufriedenheit als auch potentiell relevante Teilzufriedenheiten erheben. Während derartige Befragungen häufig durchgeführt werden, unterschätzt man das Potential von Treiberanalysen bisweilen (KATCHER, 2003).

Werden Daten im Einkomponentenverfahren erhoben (siehe Abschnitt 3.3.1), dann müssen die Gewichtungen der Attribute bzw. Teilzufriedenheiten aus den Daten berechnet werden.

Häufig führt man Regressionsanalysen durch, die im Ergebnis die bedingte Korrelation der Teilaspekte mit der Gesamtbewertung anzeigen (OLIVER, 2010, S. 33). Somit wird klar, welche Teilaspekte einen eigenständigen Beitrag zur Gesamtzufriedenheit leisten. Dabei wird auch immer wieder untersucht, inwieweit nichtlineare Beziehungen zwischen einzelnen Attributen und der Gesamtzufriedenheit bestehen (u. a. FINN, 2011; TING und CHEN, 2002).

Anwendungen von Regressionsmodellen zur Untersuchung von Kundenzufriedenheit CHEUNG und LEE (2004) verwenden ein Regressionsmodell, um den Einfluss unter-

schiedlicher Attribute auf Zufriedenheit zu untersuchen. Attributskonstrukte werden zwar durch mehrere Fragen operationalisiert, jedoch nicht mit einer konfirmatorischen Faktorenanalyse zusammengefasst. Es werden stattdessen arithmetische Mittelwerte gebildet.

GÓMEZ et al. (2004) beschäftigen sich mit Kundenzufriedenheit im Einzelhandel. Die Untersuchung basiert auf der Idee von HESKETT et al. (1994), wonach Zufriedenheit am Ende der Kette für das Unternehmen erfolgsrelevant wird. Die untersuchten Dimensionen sind Kundenservice, Qualität und Wert. Da 20 Items erhoben worden sind, wird eine Hauptkomponentenanalyse durchgeführt, jedoch werden schließlich Mittelwerte der Items für die drei Konstrukte berechnet und für die Treiberanalyse verwendet.

MATZLER et al. (2004) berechnen die Attributsbedeutungen für eine Importance-Performance-Analyse (MARTILLA und JAMES, 1977) ebenfalls mittels eines multiplen linearen Regressionsmodells.

PARASURAMAN et al. (1988, S. 31) untersuchen den Einfluss unterschiedlicher Service-dimensionen auf die Gesamtbewertung mithilfe eines Regressionsmodells. Die Scores der Servicedimensionen wurden hier durch 22 Fragebogenitems abgefragt.

Werden Kundenzufriedenheit und Loyalität in einem Modell gemeinsam betrachtet, so fungiert Kundenzufriedenheit i. d. R. als Mediator zwischen Attributen und Loyalität. Dies lässt sich mit multipler Regression ebenso wenig darstellen, wie die Zusammenfassung von Item-Batterien (RYAN et al., 1999). Mit der Strukturgleichungsmodellierung wird im nächsten Abschnitt eine Methode vorgestellt, die sich aufgrund dieser Einschränkungen der Regressionsanalyse in der Kundenzufriedenheitsforschung etabliert hat. Das wichtigste Argument für Strukturgleichungsmodellierung scheint zudem die Möglichkeit der Verwendung von Messmodellen zu sein, da einige Anwendungen von Regressionsmodellen bei Treiberanalysen auf eine vorgeschaltete Hauptkomponentenanalyse angewiesen sind.

2.1.4. Strukturgleichungsmodelle

Seit Ende der 1990er Jahre zeichnet sich ein Trend zur Anwendung von Strukturgleichungsmodellen in der Kundenzufriedenheitsforschung ab. In dieser Zeit benennen HOMBURG und HILDEBRANDT (1998, S. 499f.) die Kovarianzstrukturanalyse als das „leistungs-

fähigste“ Verfahren. DILLON et al. (1997) fordern dazu auf, SEMs zur Untersuchung von Kundenzufriedenheit verstärkt einzusetzen. Zentrale Argumente sind die Kontrolle von Messfehlern sowie die gleichzeitige Modellierung von theoriebasierten Pfadstrukturen.

„In the most general terms, SEM provides a holistic framework for simultaneously investigating measurement and structural relationships.“ (DILLON et al., 1997, S. 23)

Zusammenfassend werden im Vergleich zur Regressionanalyse vor allem die folgenden Argumente vorgebracht:

- (Lineare) Regressionsmodelle bilden den Zusammenhang zwischen exogenen und endogenen Items ab. Messfehler können dabei nicht berücksichtigt werden, sodass fehlerfreie Messung aller Items implizit unterstellt wird. PETERSON und WILSON (1992, S. 69) konstatieren jedoch, dass es unmöglich sei, die „wahre“ Zufriedenheit zu messen. Dies impliziert, dass jede Messung von Zufriedenheit fehlerbehaftet ist. Unterstellt man zudem, dass Zufriedenheit per se kontinuierlicher Natur ist, kann eine Messung auf einer ordinalen Rating-Skala per Definition nicht fehlerfrei sein.
- Exogene Variablen werden im linearen Regressionsmodell als unabhängig angenommen. Dies ist aber im Zufriedenheitskontext selten der Fall, insbesondere wenn Item-Batterien verwendet werden. Dabei beziehen sich mehrere Fragen auf die gleichen Teilzufriedenheitsaspekt.
- Strukturen, die über die Beziehung von exogenen zu endogenen Variablen hinausgehen, können mit der Regressionanalyse nicht abgebildet werden. Zwischen endogenen Konstrukten wie Loyalität und Zufriedenheit wird aber regelmäßig eine Abhängigkeit unterstellt. Dies kann mit einem linearen Regressionsmodell nicht erreicht werden.

Messen mit Fehler Ausgangspunkt von Messmodellen ist die Idee, dass Aussagen über die Beziehung von latenten Konstrukten getroffen werden sollen, die nicht fehlerfrei messbar sind. Auch bei „Zufriedenheit“ handelt es sich um ein eher abstraktes Konstrukt, sodass ein direktes „Ablesen“ nicht möglich ist.

Auch in der Testtheorie tragen sogenannte Item Response Models dieser Idee Rechnung.¹² Bei Tests kommen häufig sogenannte „Testlets“ zum Einsatz. Dabei werden bestimmte Fähigkeiten auf unterschiedliche Weise getestet. Die entsprechenden Testelemente sind nicht unabhängig voneinander, da sie sich auf die gleiche zugrunde liegende Fähigkeit beziehen (BRADLOW et al., 1999).

Bei der Untersuchung von Kundenzufriedenheit werden teilweise Item-Batterien eingesetzt, etwa um Gesamtzufriedenheit zu messen. So empfehlen RYAN et al. (1995, S. 12) die Operationalisierung von Gesamtzufriedenheit durch drei Messvariablen (bzw. Fragen). Auch bei der Messung von Einstellungen ist die Kontrolle von Messfehlern bzw. die Beurteilung von Validität und Reliabilität von Skalen ein zentrales Motiv für die Verwendung von Messmodellen mit mehreren Items je Konstrukt (KROSNICK, 1991; ALWIN und KROSNICK, 1991). Zur Überprüfung von Messmodellen werden üblicherweise zunächst explorative und schließlich konfirmatorische Faktoranalysen eingesetzt (GERBING und ANDERSON, 1988). Letztere sind ein Spezialfall eines Strukturgleichungsmodells.

Domänen ANDERSON et al. (2008) untersuchen die Zufriedenheit mit Dienstleistungen im Luftverkehr. Ziel der Untersuchung ist die Identifikation relevanter Komponenten von Dienstleistungen. Zur Schätzung wird LISREL verwendet. Sie ordnen dabei jeweils mehrere Items unterschiedlichen Teildimensionen wie Interaktion, Flugzeug, Personal, Essen, usw. zu. Dabei werden zwei bis sieben Items pro Dimension verwendet. Von Interesse ist die Wichtigkeit dieser Domänen (wie Flugzeug oder Personal), nicht die der einzelnen (hoch miteinander korrelierenden) Items. Cronbach's Alpha nimmt für alle Konstrukte Werte größer als 0.8 an, was auf eine hohe Korrelation der Items innerhalb der Dimensionen (bzw. Konstrukte) deutet. Die Anwendung des Messmodells wird hier folglich durch inhaltliche Domänen motiviert, nicht durch klassische Messkonstrukte (z. B. Gesamtzufriedenheit gemessen durch Leistung, Zufriedenheit, Erfüllung von Erwartungen wie RYAN et al. (1995, S. 12) empfehlen).

¹Einen Überblick über klassische und probabilistische Testtheorie bietet MÜLLER (1999).

²Auf die Verwandtschaft von faktoranalytischen Modellen und testtheoretischen Modellen wird kurz in Abschnitt 4.6 eingegangen.

Anwendungen von SEMs zur Untersuchung von Kundenzufriedenheit Strukturgleichungsmodelle werden seit den 1990er Jahren verstärkt bei der Modellierung von Kundenzufriedenheit eingesetzt. Insbesondere die steigende Popularität von Kundenzufriedenheitsindizes (siehe Abschnitt 3.3.3) hat für ein zunehmendes Interesse an Strukturgleichungsmodellen gesorgt. Der europäische (ECSI) sowie der amerikanische (ACSI) Zufriedenheitsindex basieren auf Strukturgleichungsmodellen (FORNELL et al., 1996b). Das ECSI-Modell beruht auf sieben latenten Variablen, die jeweils durch mehrere Items gemessen werden (EKLOF et al., 1999). Zur Schätzung wird dabei häufig auf das Partial-Least-Squares-Verfahren zurückgegriffen (BAYOL et al., 2000), welches jedoch keine Möglichkeit zur Berücksichtigung von Skalenverwendungsheterogenität bietet. Dennoch ist diese Methode weit verbreitet. BLIEMEL und HASSANEIN (2007) verwenden das Partial-Least-Squares-Verfahren bei der Untersuchung von Zufriedenheit im Gesundheitsbereich, STAN und SAPORTA (2005) auf dem Automobilmarkt. CARDONE-RIPORTELLA et al. (2001) untersuchen die Zufriedenheit von Studenten mit Lehrveranstaltungen.

Auch das INDSAT-Modell (HOMBURG und RUDOLPH, 2001) greift auf Strukturgleichungsmodellierung zurück. Bei der Untersuchung von Kundenzufriedenheit im Industriegüterbereich werden sieben Teilzufriedenheitsdimensionen jeweils durch mindestens drei Messvariablen operationalisiert, während Gesamtzufriedenheit durch zwei Fragen gemessen wird. Auch SHARMA et al. (1999) verwenden ein Strukturgleichungsmodell, um eine große Zahl von Attributskonstrukten (bzw. Domänen) durch jeweils mehrere Messvariablen zu operationalisieren. Insgesamt werden dabei zehn Domänen (wie Lieferung, Installation, Produkt, Wartung, etc.) als potentielle Zufriedenheitstreiber untersucht.

Auch im Dienstleistungsbereich gibt es derartige Anwendungen: FORZA und FILIPPINI (1998) verwenden ein LISREL-Modell, um den Effekt von Total Quality Management auf Kundenzufriedenheit zu untersuchen. Auch bei Modellen, welche sich an das SERVQUAL-Modell (PARASURAMAN et al., 1988) anlehnen, liegt der Einsatz von SEMs nahe (z. B. STEVENS et al., 1995). PARASURAMAN et al. (1988) selbst greifen nicht auf diese Technik zurück, obwohl die betrachteten Servicedimensionen durch eine Reihe von Items repräsentiert werden.

2.2. Response Bias

Eine große Herausforderung für die Modellierung jeglicher Befragungsdaten ist die Berücksichtigung unterschiedlicher Formen des Response Bias.

„A *response bias* is a systematic tendency to respond to a range of questionnaire items on some basis other than the specific item content. [...] To the extent that an individual displays the bias consistently across time and situations, the bias is said to be a *response style*.“ (PAULHUS, 1991, S. 17)

Verzerrendes Antwortverhalten (Response Style) und Informationen bezüglich der Zusammenhänge von originärem Interesse müssen auseinandergelassen werden, da ansonsten falsche Schlüsse gezogen werden können.

„Respondents can vary strongly in the way they use rating scales. Specifically, respondents can exhibit a variety of response styles, which threatens the validity of the responses.“ (VAN ROSMALEN et al., 2010, S. 157)

Bei der Treiberanalyse besteht zwar kein originäres Interesse an der Untersuchung von Response Styles. Jedoch müssen diese berücksichtigt werden, um die „wahren“ Zusammenhänge in den Daten - etwa zwischen Gesamt- und Teilzufriedenheiten - aufzudecken.

Obwohl das Problem der Response Styles schon eine Weile diskutiert wird (z. B. BROEN und WIRST, 1958), stellt es die Modellierung von Einstellungen und Zufriedenheit (PETERSON und WILSON, 1992, S. 67) noch immer vor große Herausforderungen. Der Begriff Response Styles umfasst eine ganze Reihe unterschiedlicher Effekte, welche einzeln oder in Kombination auftreten können. Die am meisten beachteten Effekte benennen BAUMGARTNER und STEENKAMP (vgl. 2001, S. 145) wie folgt:

- *Ja-Sagen / Nein-Sagen*: Allen Aussagen wird tendenziell zugestimmt bzw. alle Aussagen werden abgelehnt.
- „*net acquiescence*“: Bezeichnet die Tendenz positive bzw. negative Antworten zu verstärken.
- *Extreme Response Styles*: Die Antworten bewegen sich nur an den Endpunkten der Skala, sodass nur völlige Zustimmung oder völlige Ablehnung ausgedrückt wird.

Analog wird nur volle Zufriedenheit bzw. völlige Unzufriedenheit bei Zufriedenheitsbefragungen ausgedrückt.

- *Reichweite*: Etwas allgemeiner als die Extreme Response Styles unterscheiden sich Probanden bzw. Kunden bezüglich der verwendeten Skalenbreite.
- *Mittlere Bewertungen*: Alle Attribute werden mit einem Wert Nahe des Skalenmittelpunktes bewertet.
- *Zufällige Antwort*: Die Antworten sind eher zufällig und enthalten somit keine Information über Zusammenhänge. Dies lässt sich u. a. auf mangelnde Motivation zur Beantwortung der Fragen zurückführen.

PAULHUS (1991, S. 17f.) geht zudem auf „soziale Erwünschtheit“ als eine Form des Response Bias ein. Die Tendenz, sozial erwünschte Antworten zu geben (also z. B. sich nur positiv gegenüber Umweltschutz zu äußern), wird in der Literatur ebenfalls ausführlich diskutiert.

Wird nicht berücksichtigt, dass Menschen die Antwortskalen unterschiedlich verwenden, so können berechnete Korrelationen der beobachteten Daten verzerrt sein. In der Folge sind auch Ergebnisse von Zufriedenheitstreiberanalysen irreführend, da sich signifikante Zusammenhänge potentiell nur daraus ergeben, dass sich extreme Ja-Sager und Nein-Sager unter den Teilnehmern einer Befragung befinden. Derartiges Verhalten wird in dem hier entwickelten Modell berücksichtigt. Eine Illustration „inflationärer“ Korrelationen befindet sich in Abschnitt 4.4.2.

Rating-Skalen Bei der Untersuchung von Kundenzufriedenheit kommen i. d. R. *Rating-Skalen* zum Einsatz. Um Regressionsmodelle anzuwenden, werden jedoch kontinuierliche normalverteilte Daten benötigt. Daher werden Modelle eingesetzt, welche die ordinalen Beobachtungen der Ratingskalen in kontinuierliche Informationen „übersetzen“. Solche Modelle werden Cut-Point oder Threshold-Modelle genannt. Einige relevante Response Styles (Ja/Nein-Sagen, Extreme Response Styles, Reichweite) lassen sich auf der Ebene von Cut-Point-Modellen adressieren. Durch die entsprechende Modellierung im Cut-Point-Modell kann schließlich Inferenz auf der Basis von „korrigierten“ Daten erfolgen.

Es lassen sich die folgende Arten der Modellierung von Cut-Points unterscheiden (vgl.

JOHNSON, 2003, S. 567ff.):

- *homogene Cut-Points*: keine Adressierung von Skalenverwendungsheterogenität, d. h. homogene Cut-Points für alle Individuen
- *proportionale Anpassung*: ROSSI et al. (2001) ermöglichen eine proportionale Anpassung der Cut-Points auf Individualebene.
- *diffuse Prior-Struktur individueller Cut-Points*: Diese Möglichkeit erlaubt eine sehr flexible Anpassung an unterschiedliche Skalenverwendungen, jedoch ist die Unsicherheit aufgrund der diffusen Prior-Struktur und der geringen individuen-spezifischen Information groß.
- *logarithmisch normalverteilte Prior-Struktur individueller Cut-Points*: Diesen Ansatz verfolgen z. B. YING et al. (2006); BÜSCHKEN et al. (2010).

Das Skalenheterogenitätsmodell von ROSSI et al. (2001) adressiert bei einer verhältnismäßig sparsamen Parametrisierung einige zentrale Herausforderungen:

- Ein individuen-spezifischer Mittelwert korrigiert individuelle Tendenzen zum Ja-Sagen bzw. Nein-Sagen.
- Ein individuen-spezifischer Varianzparameter korrigiert die Breite der Skalennutzung.

In Abschnitt 4.4.2 wird näher auf die technischen Eigenschaften dieses Modells eingegangen. Informationslose Antworten werden zwar nicht separat adressiert, jedoch kann das hierarchische Modell mit zwei individuen-spezifischen Parametern Unterschiede in der Skalennutzung korrigieren, sodass die Verzerrung von Korrelationen verringert wird. Auch DILLON et al. (2001) nennen dieses Modell vielversprechend, wenden es jedoch noch nicht an. Einen ähnlichen Ansatz verfolgen auch WOLFE und FIRTH (2002). JOHNSON (2003) verallgemeinert unterschiedliche Ansätze und zeigt anhand von Beispielen, dass ein Ignorieren der Response Styles zu verzerrten Ergebnissen führen kann. JAVARAS und RIPLEY (2007) erweitern das Modell von ROSSI et al. (2001), das eine freie Kovarianzstruktur der Variablen erlaubt, und wenden es bei der Messung von Einstellungen an. Auch LENK et al. (2006) kombinieren heterogene Cut-Points mit dem Modell des interpersonalen Circumplex.

Werden der Halo-Effekt und Skalenverwendungsheterogenität gleichzeitig modelliert, so sind beide Effekte nicht immer eindeutig auseinanderzuhalten, insbesondere wenn jedem Individuum nur ein Beobachtungsvektor zugeordnet ist (BÜSCHKEN et al., 2010, S.22f.).

2.3. Der Halo-Effekt

Werden Treiberanalysen (wie in Abschnitt 2.1 vorgestellt) durchgeführt, impliziert dies bestimmte Annahmen zum Datengenerierungsprozess: Die Attributsurteile (Teilzufriedenheiten oder Teilleistungen) werden demnach unabhängig voneinander vorgenommen. Anschließend wird das Summenprodukt der Attributsurteile und der zugehörigen Bedeutungsgewichte berechnet.

Der Halo-Effekt bildet die Grundlage für alternative, d. h. konkurrierende Kausalstrukturen. Demnach werden die Attributs- und Gesamturteile durch einen Gesamteindruck determiniert.

Somit handelt es sich um einen speziellen Fall der Response Substitution (GAL und RUCKER, 2011). Dabei antworten Menschen auf nicht gestellte Fragen. Im Fall des Halo werden alle Attributs- und Gesamturteile abweichend von der eigentlichen Fragestellung beantwortet - der Befragte antwortet konsistent im Sinne seines Gesamteindrucks.

2.3.1. Definition des Halo

Der Halo, Halo-Effekt oder Halo-Fehler wurde in unterschiedlichen Disziplinen beobachtet, so vor allem in der Psychologie und Sozialpsychologie (JUDD et al., 1991; LAY und JACKSON, 1969; MURPHY und JAKO, 1989; NISBETT und WILSON, 1977) oder im Bereich Personal- und Humanressourcenmanagement (THORNDIKE, 1920; FARH et al., 1991; MURPHY und ANHALT, 1992) und später auch bei der Einstellungsmessung (BECKWITH und KUBILIUS, 1978), Markenwahrnehmung (DILLON et al., 2001) und Kundenzufriedenheit (BÜSCHKEN et al., 2010; WIRTZ, 2003; WIRTZ und BATESON, 1995) im Marketing.

Als Erster hat THORNDIKE (1920) den Begriff Halo im Kontext von Personalbewertungen etabliert, nachdem bereits WELLS (1907) und SCOTT (1908) ähnliche Beobachtungen gemacht haben. WELLS (1907) bemerkte bei einer Studie zur Schreibfertigkeit einen konstanten Fehler. Laut Wells würden Bewertende ihren Gesamteindruck durch die Beur-

teilung einzelner Qualitäten beeinflussen lassen. THORNDIKE (1920) beobachtete bei der Personalbewertung im Militär, dass die Einschätzung einzelner Kriterien vom Gesamteindruck dominiert wurde. Dieses „Überstrahlen“ von einem einzigen Gesamteindruck auf alle anderen Teilbewertungen bezeichnete er folglich mit „Halo“.

Seither wurde eine große Zahl an Untersuchungen zum Halo publiziert. BALZER und SULSKY (1992) kritisieren schließlich die uneinheitliche Definition des Halo-Effekts der von ihnen betrachteten Untersuchungen.³ Die Kritik ist zweierlei: Zum einen gebe es unterschiedliche Konzeptionen des Halo, zum anderen aber auch eine sehr große Zahl sehr unterschiedlicher Methoden zur Identifikation des Halo. Einen guten Überblick über unterschiedliche konzeptionelle Definitionen findet sich in FEELEY (2002).

„Falsche“ Korrelation Dabei sind die Symptome des Halo-Effektes unabhängig von der Definition ähnlich:

1. zum einen durch zu hohe Korrelation der einzelnen zur Bewertung stehenden Attribute MURPHY und JAKO (1989) nennen den Halo *überschüssige* Korrelation im Vergleich zur *wahren* Korrelation
2. zum anderen zeigt sich eine zu geringe Varianz bei den Antworten der jeweiligen Subjekte (vgl. COOPER, 1981). Aufgrund von mangelnder Differenzierung zwischen Attributen fallen die Urteile eines Bewertenden sehr ähnlich aus - die Varianz ist geringer als es sich bei einer „richtigen“ Bewertung zeigen würde

BALZER und SULSKY (1992) differenzieren in ihrem Literaturüberblick zunächst zwei unterschiedliche Definitionen des Halo: „General Impression Halo“ und „Dimensional Similarity Halo“. Darüber hinaus unterscheiden FISICARO und LANCE (1990) auch das „Salient Dimension Model“. Die unterschiedlichen Modellstrukturen sind in Abbildung 2.1 dargestellt. Abweichend von der Darstellung in FISICARO und LANCE (1990, S. 420) werden nur die Halo-Strukturen dargestellt. FISICARO und LANCE (1990) gehen hingegen davon aus, dass haloistische und formative Bestandteile gleichzeitig Einfluss auf die beobachteten Bewertungen ausüben. Das in dieser Arbeit entworfene Modell ermöglicht jedoch eine

³BALZER und SULSKY (1992) untersuchten alle Publikationen von 1980 bis 1990 aus sechs ausgewählten akademischen Zeitschriften.

Zuweisung individueller Antworten zu haloistisch und formativ geprägten Komponenten. Die Modellierung eines simultanen Einflusses von Halo und formativer Bewertung wie bei BECKWITH und LEHMANN (1975) und DILLON et al. (2001) wird daher in dieser Arbeit nicht verfolgt.

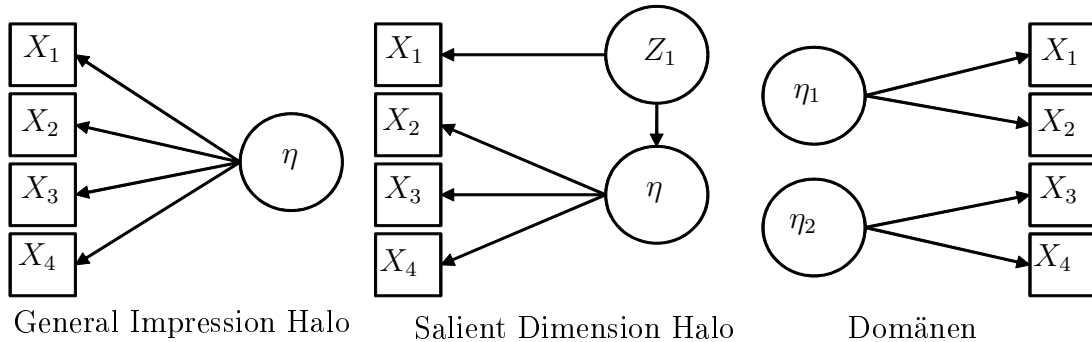
Ausgangspunkt aller Überlegungen zum Halo-Effekt ist, dass Menschen bei der Bewertung einzelner Attribute (oder Charakteristika) nicht in der Weise differenzieren, wie es vom Fragesteller gewünscht ist. Die daraus resultierende Korrelation zwischen den Merkmalen hat somit nicht die gewünschte Aussagekraft. Unterschiedliche Annahmen zu den Ursachen dieser fehlenden Differenzierung (bewusste Annahme von Zusammenhängen durch Bewerter oder generelle Dominanz des Gesamteindrucks sowie der Ausgangspunkt des Halo) führen zu den unterschiedlichen Definitionen.

General Impression Halo Dieses Modell ist der ursprünglichen Definition von THORNDIKE (1920) am nächsten. Attributspezifische Bewertungen werden dabei durch einen Gesamteindruck dominiert. THORNDIKE (1920, S. 25) bezeichnet es als „marked tendency to think of the person in general as rather good or rather inferior and to color the judgements of the qualities by this general feeling“. BALZER und SULSKY (1992) ergänzen, dass dieser Effekt auch von einem neutralen Eindruck ausgehen kann. Dieses Verständnis hat sich bereits früh in der Marketingliteratur durchgesetzt (JAMES und CARTER, 1978).

Salient Dimension Model Beim Salient Dimension Model wird angenommen, dass die attributspezifischen Bewertungen sowie der Gesamteindruck durch eine besonders hervorstechende Teilkomponente dominiert werden. Eine ähnliche Definition nehmen MACCORMICK und TIFFIN (1966) vor. In ähnlicher Weise verstehen LATHAM und WEXLEY (1981) den Halo als unpassende Verallgemeinerung der Leistung einzelner Aspekte auf die übrigen Aspekte.

Domänen Gehören mehrere Items zu einer gleichen inhaltlichen Domäne, wird auch von „Conceptual Similarity“, „Dimensional Similarity Halo“ oder „Inadequate Discrimination“ gesprochen. Dieses Halo-Modell basiert auf der Idee des „Logical Error“, welcher zuerst von NEWCOMB (1931) beschrieben und später auch von GUILFORD (1936) und GHISELLI und

Abbildung 2.1: Drei unterschiedliche Halo-Modelle



- eigene Darstellung -

BROWN (1948) aufgegriffen wurde. Dabei wird angenommen, dass Bewerter Annahmen über Zusammenhänge einzelner Qualitäten bzw. attributsspezifischer Leistungen haben. Nimmt z.B. Bewerter A an, dass die Leistungen von Objekt B bezüglich der Attribute X_1 und X_2 zusammenhängen, so wird er versuchen, eine konsistente Bewertung bezüglich dieser beiden Attribute abzugeben. Daher spricht NEWCOMB (1931) von einem logischen Fehler.

Auftreten kann diese Form des Halo, wenn konzeptionelle Ähnlichkeit zwischen einzelnen Attributen besteht (COOPER, 1981, S. 222). Die Zahl der latenten Faktoren ist dabei größer als eins, jedoch kleiner als die Zahl der beobachteten Attribute (siehe Abbildung 2.1). Wenn die konzeptionelle Ähnlichkeit a priori bekannt ist und grundsätzlich Einigkeit über diese Ähnlichkeit herrscht, dann stellt sich jedoch die Frage, inwieweit die Bezeichnung „Halo-Fehler“ angebracht ist. Das Problem ließe sich durch Zusammenfassung in einem Messmodell oder durch das Streichen quasi-identischer Attribute lösen.

NISBETT und WILSON (1977) unterscheiden die Interpretation des Halo in „stark“ und „schwach“. Verfügen Bewerter über ausreichend Informationen, um eine differenzierte Bewertung durchzuführen, lassen sich aber dennoch von einem Globalurteil beeinflussen, so entspricht dies der „starken“ Interpretation des Halo. Der Halo wirkt hier somit stärker als ein „Logical Error“. Nach der „schwachen“ Interpretation tritt der Halo nur dort auf, wo der Bewerter keine eindeutigen Informationen besitzt.

Halo und verwandte Konzepte BERMAN und KENNY (1976) nennen Halo-Fehler bzw. logical error allgemein „correlational bias“. Sie beschreiben damit generell das Problem, dass systematische Verzerrungen der Korrelation dadurch auftreten, dass Bewertungen unter der Annahme bestimmter Abhängigkeiten zwischen einzelnen Merkmalen durchgeführt werden. Laut NISBETT und WILSON (1977) kann der Halo-Fehler jedoch darüber hinausgehen und auch Korrelationen zwischen logisch trennbaren Konstrukten erzeugen.

GHISELLI und BROWN (1955) beschreiben neben dem Halo zwei Varianten eines „constant error“. Bei dem „leniency error“ befinden sich alle Bewertungen beispielsweise am positiven Ende der Skala. Der „error of central tendency“ hingegen beschreibt eine durchgängig mittlere Bewertung aller Bewertungsobjekte. Diese Bewertungsfehler sind eine Eigenschaft von Bewertern und werden sichtbar, wenn mehrere Bewerber oder Produkte bewertet werden. Diese Art von Fehler wird jedoch im Rahmen des Modells zur Adressierung von Skalenverwendungsheterogenität näher diskutiert (siehe Abschnitt 4.4.2).

GUILFORD (1954) nennt neben dem Halo-Fehler auch den „Rater-trait error“, wobei Attribute aus Zuneigung zum Bewertungsobjekt über- oder untergewichtet werden. In der Marketingliteratur hat dieses Konzept jedoch kaum Eingang gefunden, daher wird auch an dieser Stelle nicht näher darauf eingegangen.

Halo-Definition dieser Arbeit Dieser Arbeit liegt grundsätzlich das Verständnis des Halo im Sinne des „General Impression model“ zugrunde. Analog zur „starken“ Interpretation des Halo wird davon ausgegangen, dass Kunden Teilzufriedenheiten auch dann nicht notwendigerweise differenzieren, wenn genügend Informationen zur Differenzierung vorhanden sind. Dieser Effekt muss jedoch nicht auf alle Teilbewertungen wirken.

Weil in dieser Arbeit Messmodelle verwendet werden, bedarf es zudem einer weiteren Abgrenzung: Das Auftreten konzeptionell ähnlicher Attribute kann mit der Messung psychometrischer Konstrukte verwechselt werden. Dabei fasst man mehrere Items in einem Messmodell zusammen. Die Korrelation der einzelnen Items ist dabei jedoch „gewollt“. Beim „Salient Dimension Model“ lassen sich verschiedene Konstrukte (unbeabsichtigt) nicht differenzieren. In dieser Arbeit werden Messmodelle eingesetzt, um verwandte Items zu Konstrukten zusammenzufassen. Der Halo-Fehler wird jedoch im Sinne des General Impression Modells verstanden.

2.3.2. Beurteilungsprozesse und Halo

LEVY und DUGAN (1960) unterscheiden zwischen der Nutzung von Bewertungsdaten als deskriptive Daten auf der einen Seite und der Analyse des Bewertungsverhaltens an sich auf der anderen Seite. Sie beschäftigen sich, wie auch eine ganze Reihe weiterer Wissenschaftler, mit der Frage nach der Entstehung von Bewertungsdaten.

Während in früheren Untersuchungen meist entweder der Bewertungsprozess untersucht oder eine Analyse der Bewertungsdaten durchgeführt wird, erlauben die modellbasierten Ansätze beides gleichzeitig zu erreichen. So bringen BECKWITH und LEHMANN (1975) in einem simultanen Gleichungssystem attributspezifische Bewertungen und objektspezifische Einstellungen zusammen. Rückschlüsse auf den Anteil des Halo sowie um den Halo korrigierte Gewichtungsfaktoren können gleichzeitig gewonnen werden. Die früheren experimentellen Untersuchungen beschränken sich meist auf den reinen Nachweis der Existenz des Halo.

Auch das in dieser Arbeit entwickelte Modell ermöglicht die Identifikation von haloistischen Bewertungen und gleichzeitig die Identifikation von Treibern von Zufriedenheit.

Wenn Menschen dazu aufgefordert werden, Beurteilungen oder Einschätzungen vorzunehmen, dann müssen sie dazu auf Informationen aus ihrem Gedächtnis zurückgreifen (LYNCH und SRULL, 1982; LICHTENSTEIN und SRULL, 1985). Dabei spielt es zunächst keine Rolle, ob Personen nach Charaktereigenschaften, Einstellungen oder Zufriedenheit befragt werden. Bei Befragungen beurteilen Teilnehmer in der Regel ein oder mehrere Objekte nach mehreren Kriterien.

Für Kunden ist es einfacher, sich nur an den Gesamteindruck bezüglich eines Produktes oder Services zu erinnern (DECOTHS, 1977; MATTILA, 2003). Sie erinnern sich möglicherweise nicht in dem gewünschten differenzierten Ausmaß an das zu bewertende Objekt oder Ereignis. Die abgefragten Teilzufriedenheiten sind somit nur noch eine Reflektion des Gesamteindrucks. In anderen Fällen verfügen Kunden gar nicht über die notwendigen Informationen für ein differenziertes Urteil, sodass Teilaspekte nicht unterschieden werden können (DILLON et al., 2001).

Die Wurzeln in der Psychologie Die Mehrzahl der früheren Publikationen stammen aus Bereichen der Psychologie, insbesondere zur Bewertung von Personal. So definieren NISBETT und WILSON (1977, S. 250) den Halo-Effekt als „the influence of a global evaluation on evaluations of individual attributes of a person“. Eine ähnliche Definition liegt den meisten betrachteten Untersuchungen zugrunde (so auch abgewandelt im Kontext von Zufriedenheit (z. B. GHISELLI und BROWN, 1955; BALZER und SULSKY, 1992)).

Leicht abweichend beschreiben NEWCOMB (1931) und GUILFORD (1936) einen „logical error“, wobei konsistente Bewertungen für solche Merkmale vergeben werden, die aus Sicht des Bewerbers logisch miteinander verbunden sind. Die Gemeinsamkeit des „logical error“ und des Halo im Sinne von THORNDIKE (1920) liegt darin, dass bei der Bewertung nicht in den vorgegeben Kategorien differenziert bewertet wird. Während der Fragesteller beispielsweise 20 unterschiedliche Teilattribute bewerten lassen möchte, unterscheidet der Befragte gar nicht oder nur zwischen wenigen und damit gröberen Kriterien. Befragte sind also nicht in der Lage, zwischen den Teilattributen zu differenzieren. Die geringere Dimensionalität des Bewertungsprozesses konnte schon früh durch faktoranalytische Ansätze nachgewiesen werden (z. B. LEVY und DUGAN, 1960).

2.3.3. Operationalisierung des Halo

In der Markenforschung und anderen Disziplinen besteht großes Interesse an der Korrelation unterschiedlicher Merkmale. Welche Eigenschaften Marken in welcher Kombination zugebilligt werden, ist von großem Interesse. Aber wann sind die empirisch berechneten Korrelationen „zu hoch“ und wann sind sie „echt“?

Wenn Regressionsanalysen durchgeführt werden, um etwa Treiber von Kundenzufriedenheit zu identifizieren, so ist eine hohe Korrelation der unabhängigen Variablen ein Problem. Zum einen liegt dies in der Modellannahme eines linearen Regressionsmodells begründet, zum anderen besteht das Interesse gerade daran, relevante Treiber zu isolieren. Will man Treiberanalysen durchführen, muss der Halo-Fehler nicht nur identifiziert, sondern auch korrigiert werden. In diesem Abschnitt werden zunächst die bisherigen Ansätze zur Operationalisierung dargestellt. Nicht alle Verfahren eignen sich auch zum „Herausrechnen“ von haloistischen Antworten. Darauf wird in Abschnitt 2.3.5 eingegangen.

In einer frühen Systematisierung der unterschiedlichen Verfahren zur Aufdeckung des Halo unterscheiden JAMES und CARTER (1978) vier unterschiedliche Verfahren zur Identifikation des Halo-Effektes: 1) Korrelation zwischen Attributen, 2) ANOVA, 3) Objektspezifische Varianz und 4) Regressionsmodelle. Ähnliche Unterscheidungen nehmen auch SAAL et al. (1980) und JACOBS und KOZLOWSKI (1985) vor, die Faktoranalyse als spezielleres Verfahren zur Analyse der Korrelation erwähnen. JACOBS und KOZLOWSKI (1985) weisen zudem auf den Ansatz von LANDY et al. (1980) hin, wobei partielle Korrelationen betrachtet werden, um die Halo-Fehler zu kontrollieren. COOPER (1981) bringen in ihrer Systematisierung darüber hinaus Vergleichsansätze (in Experimenten) als weiteres Verfahren ein.

BALZER und SULSKY (1992) finden ein paar Jahre später in ihrem umfassenden Literaturüberblick schließlich 108 unterschiedliche Nachweisverfahren⁴ des Halo, die sie in sechs Kategorien unterteilen. Auch deren Systematisierung ähnelt den vorangegangenen, jedoch werden anders als bei JAMES und CARTER (1978) Regressionsansätze nicht separat aufgeführt. Berücksichtigt man diese zusätzlich, lassen sich folglich sieben unterschiedliche Verfahren unterscheiden:

1. *objektspezifische Varianz*: Ist die Varianz der Bewertung der Teilattribute je Objekt gering, so kann dies auf den Halo hinweisen (BROWN, 1968). Die Festlegung von Schwellenwerten kann hier allerdings nur subjektiv erfolgen. Auch lässt sich schon konzeptionell nicht unterscheiden, ob Skalenverwendungsheterogenität⁵ oder haloistisches Antwortverhalten (oder andere Faktoren) ursächlich für unterschiedliche Antwortvarianz sind.
2. *Korrelation zwischen den Attributen*: Korrelieren die Attribute sehr stark, obwohl es sachlogisch nicht gerechtfertigt erscheint, so sind die Daten möglicherweise vom Halo-Effekt kontaminiert (z. B. BERNARDIN, 1977; SYMONDS, 1925). Dabei wird meist die durchschnittliche Varianz der Bewertungen berechnet. Die Festlegung eines Schwellenwertes ist allerdings subjektiv. Wieviel Korrelation „zu viel“ ist, lässt sich dabei nur schwer festlegen. Eine Möglichkeit besteht in einem Vergleich mit einer

⁴BALZER und SULSKY (1992) nennen es „operative Definitionen“.

⁵Genauer wird darauf in Abschnitt 4.4.2 eingegangen.

- „wahren“ Korrelationsmatrix, wie etwa bei der Untersuchung von NEWCOMB (1931).
3. *ANOVA* über alle Dimensionen bzw. Attribute hinweg: Stellt sich die Interaktion von $\text{Bewerter} \times \text{Objekt}$ als signifikant heraus, dann kann dies als Indiz für den Halo gedeutet werden (GUILFORD, 1954; WILLINGHAM und JONES, 1958; STANLEY, 1961; DICKINSON und TICE, 1977; KING et al., 1980; HEDGE und KAVANAGH, 1988).
 4. *Faktorstruktur der Attribute*: Faktoranalyse oder Hauptkomponentenanalyse der Daten. Dabei zeigt sich, soweit der Halo-Effekt vorhanden ist, ein gemeinsamer Faktor: der Gesamteindruck. Die zugrundeliegende Dimensionalität der Beurteilung durch den Bewerter ist somit geringer als vom Befrager gewünscht (KAFRY et al., 1979; KRAUT, 1975).
 5. *Herausrechnen des Halo*: Dies geschieht z. B. indirekt durch Berücksichtigung einer ebenfalls abgefragten Gesamtbewertung (LANDY et al., 1980; NATHAN und LORD, 1983) oder auf der Basis von partiellen Korrelationen (LANDY et al., 1980). Der letztgenannte Ansatz ist jedoch umstritten (NATHAN und LORD, 1983).
 6. *Vergleich*: Dabei wird mit wahren Bewertungen, z. B. durch Experten erstellte Beurteilungen, die als „Halo-frei“ gelten, verglichen. Die Annahme, dass Experten „wahre Bewertungen“ abgeben können, ist jedoch problematisch. Letztendlich ist ein solches Vorgehen nur in experimentellen Untersuchungen möglich (MURPHY et al., 1993). Dabei werden mitunter auch einzelne Leistungsdimensionen manipuliert und Auswirkungen auf die Bewertung anderer Dimensionen beobachtet (z. B. WIRTZ und BATESON, 1995).
 7. *Regressionsmodelle*: Dabei werden generelle Eindrücke und Teilbewertungen in einem Modell zusammengeführt, sodass der Effekt des Gesamteindrucks kontrolliert werden kann. Die Regressionskoeffizienten geben Aufschluss über den Einfluss von Halo sowie einzelner Attribute.

Anfang der 1990er Jahre waren die gebräuchlichsten Verfahren die Betrachtung der Korrelationen zwischen den Kategorien sowie die Faktoranalyse (BALZER und SULSKY, 1992). Neben den sieben etablierten Verfahren gibt es noch eine weitere neue Möglichkeit, den Halo zu operationalisieren. So schlagen BÜSCHKEN et al. (2010) ein Mixture-Modell

vor, welches auf Basis der Kovarianzinformation Antworten einem Halo-Antwortstil oder einem formativen Antwortstil zuordnet. Dieser Ansatz unterscheidet sich damit grundlegend von den bisherigen Modellierungsansätzen.

2.3.4. Nachweise des Halo-Effektes

Wie im vorherigen Abschnitt dargestellt, gibt es sieben unterschiedliche Gruppen von Nachweisverfahren des Halo. Dabei bemängelten BALZER und SULSKY (1992), dass nicht jede konzeptionelle Definition des Halo mit jeder Operationalisierung kompatibel sei. In der Literatur sei dies nicht immer beachtet worden. Darüber hinaus eignen sich nicht alle Verfahren zur Beurteilung des Halo-Effektes bei allen Daten. Einige Verfahren benötigen „wahre Bewertungen“ oder einen experimentellen Aufbau, um belastbare Ergebnisse produzieren zu können. Die Regressionsmodelle hingegen können auch in nicht-experimentellen Situationen verwendet werden. Dies ist gerade für die Implementierung in Unternehmen wichtig. Daher wird an dieser Stelle zwischen *experimentellen Untersuchungen* und *nicht-experimentellen Untersuchungen* unterschieden. Experimentelle Untersuchungen geben lediglich Aufschluss über Ausmaß und im Experiment untersuchte Treiber des Halo. Besonders die früheren Untersuchungen bauen auf experimentellen Designs auf. In späteren Untersuchungen, insbesondere im Marketing, steht auch die Implementierbarkeit bei einfachen Befragungen im Vordergrund. Nachweise auf nicht-experimenteller Basis sind i. d. R. in einer Form modellbasiert. Insbesondere zu nennen sind die Regressionsmodelle (z. B. BECKWITH und KUBILIUS, 1978; BECKWITH und LEHMANN, 1975) oder das Mixture-Modell (BÜSCHKEN et al., 2010).

Experimentelle Untersuchungen An dieser Stelle werden die wichtigsten Beiträge zur Untersuchung des Halo-Effektes mittels experimenteller Designs zusammengefasst.

NEWCOMB (1931) untersucht Beurteilungen aus einem Boot Camp mit Jugendlichen. Dabei wurden Verhaltensauffälligkeiten in 26 Kategorien sorgfältig protokolliert. Die Korrelation zwischen den Kategorien lag auf dieser Basis bei 0,14. Nach dem Camp wurden Beobachter nach den Verhaltensauffälligkeiten befragt. Die Korrelation zwischen den Kategorien betrug in diesem Fall 0,49. Die Ursache sah Newcomb darin, dass die Beobachter

bestimmte Verhaltensweisen als logisch miteinander verknüpft wahrgenommen haben. In der Folge strahlen die Bewertungen einzelner Kategorien auf andere über.

Einen anderen Weg schlagen LEVY und DUGAN (1960, S. 22f.) ein, die eine ganze Reihe signifikanter Korrelationskoeffizienten zwischen Bewertungen einzelner Charaktereigenschaften finden. Dazu wurden 15 unterschiedliche Charaktereigenschaften untersucht, die auf der Basis von Fotos durch die Probanden evaluiert worden sind. Eine Faktoranalyse zeigt, dass es neben unterschiedlich (differenziert) bewerteten Dimensionen auch einen „Halo-Faktor“ gibt, der die Bewertung eher in den positiven oder negativen Bereich bewegt - je nachdem welchen Eindruck das Foto beim jeweiligen Bewerter hinterlassen hat.

LANDY und SIGALL (1974) finden einen signifikanten Einfluss der physischen Attraktivität einer Autorin auf die Bewertung ihres Aufsatzes. Obwohl die Attraktivität faktisch unabhängig von der Bewertung des Aufsatzes sowie der Schreibfertigkeit ist, hat diese offenbar einen überstrahlenden Einfluss.

NISBETT und WILSON (1977) weisen den Halo-Effekt bei der Bewertung eines Dozenten durch Studenten nach. Die experimentelle Untersuchung verwendet dazu unterschiedliche Videos, wobei die Freundlichkeit des Dozenten manipuliert wurde. Die Bewertung aller einzelnen Kriterien wurde dadurch beeinflusst, sodass der Unterschied auf den Halo zurückgeführt wird. Der Halo wirkt dabei offenbar, obwohl genug Informationen für unabhängige Bewertungen vorhanden sind.

JAMES und CARTER (1978) weisen den Halo-Effekt bei der Bewertung von Städten mittels vier unterschiedlicher Verfahren nach.

MURPHY et al. (1986) untersuchen experimentell den Einfluss des Gedächtnisses auf den Halo-Effekt. Eine später durchgeführte Befragung erhöht den Anteil von Halo-Antworten, da die Testsubjekte stärker auf ihr Erinnerungsvermögen angewiesen sind. Dabei bewerten Studenten Videoaufzeichnungen mit und ohne Zeitverzug. Der Nachweis des Halo erfolgt hier anhand der Korrelation zwischen den Items.

WIRTZ und BATESON (1995) untersuchen die Zufriedenheit mit dem Service einer Bank durch ein Experiment. Dabei werden sowohl Geschwindigkeit als auch Erwartungen manipuliert. Die Manipulation der Geschwindigkeit wirkt auch auf die Bewertung der anderen Attribute, somit kann der Halo-Effekt nachgewiesen werden.

WIRTZ (2000) erweitert die frühere Untersuchung von WIRTZ und BATESON (1995). Dabei werden Reiseagenturen in einem Experimente bewertet. Bei der Manipulation von wichtigeren Teilattributen konnte ein stärkerer Halo beobachtet werden.

Nicht-experimentelle und modellbasierte Ansätze Nicht-experimentelle Ansätze bieten den Vorteil, dass sie sich - zumindest theoretisch - auch in der Praxis einsetzen lassen.

BECKWITH und LEHMANN (1975) beschäftigen sich mit multiattributiven Einstellungsmodellen, die in gleicher Weise wie multiattributiven Zufriedenheitsmodelle vom Halo betroffen sein können. Die Autoren untersuchen die Präferenzen bezüglich 20 unterschiedlicher Fernsehsendungen und lassen dabei sechs unterschiedliche Attribute beurteilen. Weil die Autoren davon ausgehen, dass Einstellungen die Bewertungen von Eigenschaften beeinflussen (Halo-Effekt) und gleichzeitig die Bewertung von Eigenschaften die Einstellung beeinflussen können, modellieren sie beides in einem simultanen Gleichungssystem. Dabei werden starke Einflüsse des Halo gefunden, insbesondere bei den Kategorien, denen a priori weniger Wichtigkeit beigemessen wird. Der Einfluss des Gesamteindrucks lässt sich mit diesem Modell für jede Kategorie separat bestimmen.

Während die überwiegende Mehrzahl der Untersuchungen Halo in bedeutendem Ausmaß feststellen, kommen MOORE und JAMES (1978) bei der Untersuchung der Bewertung von Automobilen zu dem Ergebnis, dass der Halo nur einen geringen Einfluss ausübt. Dabei wurde eine Methode verwendet, die der von BECKWITH und LEHMANN (1975) ähnelt. Aufgrund der Einzigartigkeit dieses Ergebnisses besteht jedoch auch die Möglichkeit, dass die verwendeten Daten oder die Methodologie Schwächen aufweisen.

BECKWITH und KUBILIUS (1978) entwickeln ein Modell für die Untersuchung von Images, in dem die Bewertungen einzelner Attribute als Funktion von Gesamtbewertung, individueller Nutzungsintensität, Gewichtungskoeffizient und wahrer Bewertung dargestellt wird. Dadurch lassen sich durch den Halo-Effekt betroffene Daten bereinigen. Die Untersuchung mit Daten zum Image von Supermärkten ergibt einen erkennbaren Halo-Effekt.

Im Dienstleistungsbereich können CROSBY und STEPHENS (1987) Halo-Fehler bei der Beurteilung von Versicherungen nachweisen. Dabei wird ein zweistufiges Strukturmodell mit elf unterschiedlichen Variablen verwendet, welche Gesamtzufriedenheit erklären. Der

Halo-Einfluss wird auf der Ebene aller Variablen berücksichtigt.

DILLON et al. (2001) unterscheiden bei der Untersuchung von Einstellungen zu Marken zwei Komponenten: markenspezifische Assoziationen und allgemeine Markenwahrnehmung. Die zugrundeliegende Idee ähnelt der von BECKWITH und LEHMANN (1975): Die Bewertung jedes Attributes wird dabei durch allgemeine und differenzierte Urteilskomponenten bestimmt. Je stärker der Einfluss allgemeiner Eindrücke ist, desto stärker wird der Halo eingeschätzt. Methodisch wird hier jedoch mit der Constrained Component Analysis ein faktoranalytisches Modell eingesetzt, um die zwei Komponenten zu trennen. DILLON et al. (2001) finden einen Einfluss des Halo, insbesondere wenn die Marke weniger bekannt ist und weniger Erfahrungen vorliegen.

Auch die Auswertung von Lehrevaluationen kann vom Halo-Effekt betroffen sein. MADDEN et al. (2010) passen das Modell von DILLON et al. (2001) entsprechend an und finden bei der Untersuchung von Lehrevaluationsdaten Halo in großem Ausmaß.

2.3.5. Bisherige Ansätze zur Korrektur des Halo-Effektes

Obwohl der Nachweis des Halo die Wissenschaft nun bereits 100 Jahre beschäftigt, stellt der Umgang damit immer noch eine große Herausforderung dar. Nahezu alle Untersuchungen konnten bisher die Existenz des Halo in sehr unterschiedlichen Bereichen bestätigen. Somit kann bei keiner Befragung davon ausgegangen werden, der Halo würde nicht existieren. Auch wurden unterschiedliche Maßnahmen zur Reduktion des Halo untersucht. Gänzlich „Halo-frei“ werden Befragungsdaten zur Zufriedenheit damit aber auch nicht sein. Bei Zufriedenheitstreiberanalysen muss der Halo folglich berücksichtigt werden - z. B. durch „Herausrechnen“.

Früh kam die Idee auf, den Halo-Effekt durch das „Poolen“ von Bewertern zu kontrollieren (GUILFORD, 1954, Kapitel 11). In einem Experiment von JOHNSON (1963) konnte kein deutlicher Effekt dieser Methode nachgewiesen werden. KENNY und BERMAN (1980) konnten jedoch einen positiven Effekt durch Poolen nachweisen. BECHGER et al. (2010) greifen diese Idee auch wieder auf und schlagen bei Beurteilungen die Bewertung jedes Objektes durch zwei Bewerter vor. Auf dieser Grundlage möchten die Autoren den Halo kontrollieren. Eine Anwendung im Bereich der Kundenzufriedenheit scheint jedoch fraglich,

da in vielen Fällen nur ein Bewertungsobjekt (z. B. ein Service) von vielen heterogenen Kunden bewertet wird.

Regressionsmodelle Praktikabler sind modellgestützte Verfahren, die den Halo explizit berücksichtigen. In dem Modell von BECKWITH und KUBILIUS (1978) werden Bewertungen einzelner Attribute als Funktion von Gesamtbewertung, individueller Nutzungintensität, Gewichtungskoeffizient und wahrer Bewertung dargestellt. Dadurch lassen sich durch den Halo-Effekt betroffene Daten bereinigen. Die Kernidee lässt sich anhand folgender Gleichung erläutern:

$$\text{attitude} = \text{weight} * \text{belief} + \gamma \times \text{attitude}^*$$

Die Einstellung ergibt sich aus den gewichteten Überzeugungen bezüglich unterschiedlicher Eigenschaften. Das Modell unterstellt jedoch Intervallskalenniveau bei den geäußerten Überzeugungen und trägt zudem nicht der Skalenverwendungsheterogenität Rechnung.

Mixture-Modelle Ein völlig neuartiger Ansatz besteht in der individuen-spezifischen Modellierung des Halo in einem Mixture-Modell. BÜSCHKEN et al. (2010) gehen davon aus, dass Kunden heterogen in Bezug auf ihr Antwortverhalten sind. Je nach Zeit, individuellem Informationsstand oder erlebter kritischer Ereignisse können Individuen unterschiedlich stark vom Halo beeinflusst sein. Daher wird auf Basis der individuellen Kovarianz zugeordnet, ob ein Antwortmuster eher aus einem Halo (alle Informationen stammen aus einem Gesamteindruck) oder aus einem formativen Prozess (alle Teilkriterien werden zunächst unabhängig bewertet und tragen daher eigene zusätzliche Informationen) stammt. Somit kann das Modell nur zwischen zwei extremen Formen unterscheiden: Entweder ein Kunde lässt sich ausschließlich von seinem Gesamteindruck leiten *oder* der Kunde bewertet alle Teilkriterien unabhängig voneinander. Die unterschiedlichen Zuordnungswahrscheinlichkeiten zu den Komponenten tragen der Unsicherheit der Zuordnungen Rechnung. Je sicherer die Zuordnung zur Halo-Komponente ist, desto unwahrscheinlicher ist es, dass diese Antwort Informationen für eine Treiberanalyse liefert.

Auch das in dieser Arbeit vorgestellte Modell ermöglicht die individuelle Zuordnung zu unterschiedlichen Prozessen.

2.4. Zwischenfazit

An dieser Stelle werden zunächst noch einmal die Herausforderungen zusammengefasst und der Forschungsbedarf präzisiert.

2.4.1. Herausforderungen

Die sich bei der Analyse von multiattributiven Modellen stellenden Herausforderungen lassen sich wie folgt zusammenfassen:

1. *Response Bias (Abschnitt 2.2)*: Insbesondere der heterogene Umgang von Befragten mit der Beurteilungsskala kann zu einer Verzerrung der Ergebnisse führen. Die Überführung der Informationen ordinaler Rating-Skalen in kontinuierliche Daten erfordert Modelle, die derartige Heterogenität berücksichtigen.
2. *Halo-Fehler (Abschnitt 2.3)*: Entgegen der Annahme unabhängiger Bewertungen von Attributbewertungen urteilen manche Befragte nach dem für sie weniger anstrengenden Halo-Modell. Solche haloistisch geprägten Antworten müssen identifiziert und bei der Berechnung eines Treibermodells herausgefiltert werden.
3. *Domänen bzw. Konstrukte mit mehreren Messindikatoren (Abschnitt 2.3 bzw. 2.1.4)*: Beim Einsatz von Messmodellen (z. B. durch separate Abfrage von Erwartung, Leistung und Zufriedenheit für Teil- und Gesamtzufriedenheiten) bzw. bei der Abfrage inhaltlich eng verwandter Konstrukte (Domänen) können Messmodelle eingesetzt werden. Domänen werden auch unter dem Begriff „Conceptual Similarity“ als Halo-Modell geführt. Oft sind jedoch verwandte Items a priori erkennbar (z. B. lassen sich die Items „Freundlichkeit des Personals“ und „Aufmerksamkeit des Personals“ leicht der Domäne „Personal“ zuordnen). In solchen Fällen sind „Mini“-Halo-Effekte zu erwarten. Während Befragte die Domänen „Personal“ und „Preis/Leistung“ mutmaßlich leicht voneinander trennen können, stellt die Differenzierung der zwei genannten Fragen zum Personal eine deutlich größere Herausforderung bei der Beurteilung dar.

Die in Abschnitt 2.3.5 erwähnten Modelle von BECKWITH und LEHMANN (1975); DILLON et al. (2001); MADDEN et al. (2010) adressieren den Halo-Fehler und ermöglichen dabei eine Korrektur des Halo-Fehlers. Die Ansätze basieren auf Regressions- bzw. Simul-

tangleichungsmodellen. Die Idee derartiger Ansätze ist, dass sowohl ein Gesamteindruck als auch attributspezifische Eigenschaften die Antworten der einzelnen Attribute determinieren.

Das kovarianzbasierte Modell von BÜSCHKEN et al. (2010) ermöglicht die Identifikation und Kontrolle des Halo-Fehlers bei gleichzeitiger Kontrolle der Skalenverwendungsheterogenität. Das Modell kann auf die gesamte Kovarianzinformation zurückgreifen, setzt jedoch eine finite Anzahl von Kovarianzstrukturen voraus: Die Struktur einer Komponente entspricht einem Treibermodell und wird als „formative“ Komponente bezeichnet. Eine zweite Komponente basiert auf dem Halo-Modell.⁶

2.4.2. Forschungsbedarf

Bisher nicht adressiert wird der Halo-Fehler unter Berücksichtigung von Skalenverwendungsheterogenität und *Domänen* bzw. *Messmodellen* mit mehreren Items pro Konstrukt. Das Modell von BÜSCHKEN et al. (2010) unterstützt zwar Halo-Modelle mit mehreren Faktoren, jedoch können psychometrische Messmodelle, wie etwa die Operationalisierung von Gesamtzufriedenheit durch drei Messvariablen, nicht explizit berücksichtigt werden. Gleiches gilt für die Berücksichtigung inhaltlicher Domänen („Mini-Halos“). Die starke Korrelation der zugehörigen Messvariablen eines Faktors kann so zu Fehlallokationen führen. Auch können Items, die einer inhaltlichen Domäne angehören, nicht berücksichtigt werden. Befragte, die zwar die Domänen differenzieren, jedoch innerhalb der Domänen nicht unabhängig bewerten, könnten somit ebenfalls falsch klassifiziert werden. In solchen Fällen kann für jedes Konstrukt nur eine (möglichst repräsentative) Messvariable im Modell berücksichtigt werden. Dabei gehen jedoch Informationen verloren. Die sich daraus ergebende Forschungsfrage lässt sich wie folgt formulieren:

(1) Wie können inhaltliche Domänen bei der Identifikation alternativer Kausalstrukturen berücksichtigt werden?

⁶Darüber hinaus gibt es eine dritte Komponente, welche sogenanntes „straightlining“ adressiert. Von „straightlining“ wird gesprochen, wenn alle Bewertungen eines Befragten identisch sind. Diese Komponente ist jedoch eher technischer Natur.

Die Annahme der zwei sehr unterschiedlichen Datengenerierungsprozesse durch BÜSCHKEN et al. (2010) ist möglicherweise zu restriktiv. Die Autoren gehen davon aus, dass entweder alle Attributbewertungen unabhängig voneinander vorgenommen werden oder dass alle Attributbewertungen durch ein oder zwei Faktoren erklärbar sind. Möglicherweise fällt es manchen Befragten leicht, bestimmte Attribute differenziert (d. h. unabhängig voneinander) zu bewerten, während andere Attribute lediglich den Gesamteindruck reflektieren.

Hingegen nehmen unter anderen BECKWITH und LEHMANN (1975); DILLON et al. (2001); MADDEN et al. (2010) an, dass der Gesamteindruck unterschiedliche Attributbewertungen unterschiedlich stark beeinflusst. Besonders bei sehr umfangreichen Fragebögen ist eine differenzierte Bewertung aller Attribute eine schwierige kognitive Aufgabe für die Befragten. Es ist vorstellbar, dass sich wichtige (möglicherweise Preis, Qualität u. a.) und/oder objektiv wahrnehmbare (z. B. Entfernung) Eigenschaften leichter differenzieren lassen. Eine weitere konkurrierende Theorie könnte auf der Verknüpfung des Halo-Modells mit Elementen des (formativen) Treiber-Modells beruhen. Somit lässt sich die starke Annahme zweier extremer miteinander konkurrierender Theorien lockern. Somit stellt sich eine weitere Forschungsfrage:

(2) Gibt es bei der Untersuchung von Kundenzufriedenheit eine Kausalstruktur, welche Elemente von Haloing und Forming aufweist?

BÜSCHKEN et al. (2010, S. 23) trennen Skalenverwendungsheterogenität und Halo. Darüber hinaus wurde Skalenverwendungsheterogenität bisher nicht gleichzeitig untersucht. Eine Trennung dieser Effekte ist jedoch wünschenswert: Beim Halo-Fehler handelt es sich um eine alternative Theorie zum kognitiven Prozess der Bewertung, während Skalenverwendungsheterogenität lediglich den Umgang mit der Rating-Skala beschreibt. Zumindest konzeptionell handelt es sich folglich um unterschiedliche Effekte. Ein komplexeres Modell, welches Forschungsfragen (1) und (2) berücksichtigt, sollte folglich auch Skalenverwendungsheterogenität berücksichtigen. Daraus ergibt sich für die vorliegende Arbeit noch eine dritte Forschungsfrage, nämlich:

(3) Wie können Skalenverwendungsheterogenität, konkurrierende Kausalmo-

delle und Domänen gleichzeitig modelliert werden?

Dem Verfasser sind keine Beiträge in der bisherigen Literatur bekannt, welche alle genannten Forschungsfragen beantworten. Die Behandlung von Domänen (Frage 1) ist zwar durch SEM möglich, jedoch wurde dies bisher nicht im Zusammenhang mit Skalenverwendungsheterogenität (Frage 3) verwendet.⁷ Partielles Haloing (Frage 2) wurde noch nicht in einem Mixture-Modell untersucht.

Eine Herangehensweise an partielles Haloing (Frage 2) besteht in der Einführung einer MIMIC-Komponente. Bisher wurden MIMIC-Modelle jedoch nicht im Kontext heterogener Kausalstrukturen untersucht. Bisherige MIMIC-Modelle beziehen sich zudem auf das Mess- und nicht das Strukturmodell. Zunächst wird daher auf bisherige Anwendungen von MIMIC-Modellen eingegangen.

2.4.3. MIMIC-Modelle

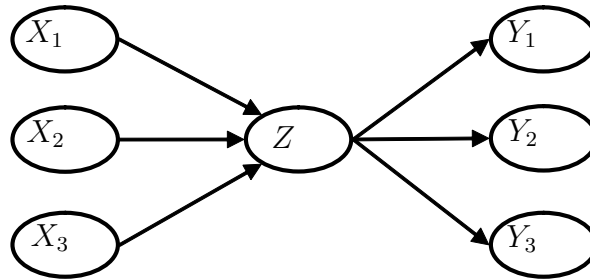
Die Annahme der zwei extremen Komponenten im Modell von BÜSCHKEN et al. (2010) kann sehr restriktiv sein, insbesondere wenn eine große Zahl von Teilbewertungen vorgenommen werden soll. Kunden unterscheiden möglicherweise zwei, drei oder vier Dimensionen im Sinne eines formativen Modells, alle anderen Kriterien können jedoch nicht differenziert werden. Auch DILLON et al. (2001) haben festgestellt, dass der Halo unterschiedlich stark auf unterschiedliche Attribute wirkt. Somit erscheint es plausibel, eine dritte Modellkomponente einzuführen, welche die differenzierbaren Dimensionen im Sinne eines formativen und die schwer differenzierbaren im Sinne eines Halo-Modells berücksichtigt.

MIMIC-Modelle (HAUSER und GOLDBERGER, 1971; JORESKOG und GOLDBERGER, 1975) entsprechen dieser Struktur. An dieser Stelle werden daher zunächst bisherige Anwendungen von MIMIC-Modellen vorgestellt.

Ein beispielhaftes Pfaddiagramm eines MIMIC-Modells ist in Abbildung 2.2 darge-

⁷Ein erster Beitrag zur Skalenverwendungsheterogenität bei LISREL-Modellen stellen HARDT und BUESCHKEN (2010) vor.

Abbildung 2.2: Eine beispielhafte MIMIC-Struktur



- eigene Darstellung -

stellt⁸. Dabei steht das Gesamturteil oder ein anders gearteter Gesamteindruck im Zentrum dieses Modells. Exogene Variablen sind wichtige Zufriedenheitsdimensionen. Die übrigen endogenen Variablen sind nur noch eine Reflektion des Gesamteindrucks, welcher aufgrund der wenigen wichtigen Dimensionen erklärbar ist. Auch BECKWITH und LEHMANN (1975) nehmen an, dass unwichtige Dimensionen potentiell ausschließlich durch den Halo erklärbar sind.

Zentrale Idee von MIMIC-Modellen ist, dass sich eine latente Größe zugleich ursächlich (Multiple Causes) erklären und durch mehrere Indikatoren (Multiple Indicators) messen lässt. Die Verbindung zwischen den Größen auf der „linken Seite“ läuft ausschließlich durch die zentrale latente Variable zu den Größen auf der „rechten Seite“.

ZELLNER (1970) beschreibt als Erster ein MIMIC-Modell. Die Anwendungen beziehen sich zunächst auf ökonomische Fragestellungen, wie etwa das Konzept des permanenten Einkommens.

HAUSER und GOLDBERGER (1971) beschreiben eine MIMIC-Struktur mit einem latenten Faktor. Sie beschäftigen sich mit der Schätzung und Identifikation.

JORESKOG und GOLDBERGER (1975) erläutern allgemein ein MIMIC-Modell und dessen Schätzung. Dabei wird grundsätzlich zwischen Indikatoren bzw. Effekten einer latenten Variablen auf der einen Seite und Ursachen der latenten Variablen auf der anderen

⁸Die Notation in dieser Abbildung (2.2) ist losgelöst von der Notation in der übrigen Arbeit. Die Darstellung dient ausschließlich der prinzipiellen Illustration eines MIMIC-Modells. X sind hier causes, Y indicators und Z ist das fokale Konstrukt.

Seite unterschieden.

ROBINS und WEST (1977) benutzen das Modell von JORES KOG und GOLDBERGER (1975) zur Schätzung von Marktpreisen von Häusern. Als kausale (formative) Indikatoren werden messbare Eigenschaften wie Größe, Zahl der Etagen und Räume, Alter u. a. verwendet. Als reflektive Indikatoren werden Schätzung des Besitzers und andere subjektive Größen benutzt.

FREY und WECK-HANNEMAN (1984) verwenden ein MIMIC-Modell zur Untersuchung der Schattenwirtschaft. BYE et al. (1985) schätzen ein MIMIC-Modell mit ordinalen Daten.

Bei der Untersuchung von Kundenzufriedenheit spielen MIMIC-Modelle keine Rolle. Das von MCDUGALL und LEVESQUE (2000) vorgestellte Modell ähnelt jedoch in etwa einer MIMIC-Struktur. Im Mittelpunkt steht jedoch ein gemessenes Konstrukt (Zufriedenheit). Die weiteren endogenen Variablen sind Wechselabsicht und Loyalität, während Qualität, relative Qualität und wahrgenommener Wert als exogene Variablen auf den zentralen Faktor Zufriedenheit zeigen.

Zur Beschreibung unterschiedlicher kognitiver Prozesse, wie dem Beurteilungsprozess von Produkten oder Dienstleistungen, wurden MIMIC-Modelle bisher nicht eingesetzt.

3. Anwendung am Beispiel Kundenzufriedenheit

In Abschnitt 3.1 wird die Bedeutung der Kundenzufriedenheit für Unternehmen dargestellt. Die definitorischen Grundlagen der Kundenzufriedenheit werden in Abschnitt 3.2 beschrieben. Anschließend wird in Abschnitt 3.3 auf bisher gängige Verfahren zur Messung von Kundenzufriedenheit eingegangen.

3.1. Zufriedenheit als Ziel von Unternehmen

„Consumer satisfaction“ is the central element of the marketing concept. (EREVELLES und LEAVITT, 1992, S. 104)

Kundenzufriedenheit ist heute aus der Unternehmenspraxis nicht mehr wegzudenken. „In der Praxis hat sich Kundenzufriedenheit neben den klassischen ökonomischen Unternehmenszielen etabliert.“ (STAUSS, 1999, S. 5). Inzwischen beschäftigt sich die Mehrheit der Unternehmen mit dem Thema Kundenzufriedenheit. Dies spiegelt sich auch in den Mission Statements großer Unternehmen wider. Auch wenn Zufriedenheit nicht immer direkt genannt wird, so ist ein verwandtes Ziel (z. B. Qualität oder Exzellenz) in den Zielen des Unternehmens genannt (JONES, 2006; ABRAHAMS, 2007; OLIVER, 2010). Nicht zuletzt konnte das sehr einfache Konzept des „Net Promoter Score“ (REICHELDELD, 2003) zur Operationalisierung von Kundenzufriedenheit in Unternehmen beitragen. Zwar sehen MATZLER und BAILOM (2006, S. 243) Defizite bei der Implementierung durch die Unternehmen, jedoch gibt es kaum Zweifel an der großen Bedeutung, die dem Thema Kundenzufriedenheit beigemessen wird. Dabei wird Kundenzufriedenheit jedoch oft nur als Werkzeug zur Performance-Messung verwendet und das Potential der Treiberanalyse vernachlässigt (KATCHER, 2003).

Nahezu alle in einer Studie von STAUSS et al. (2006) befragten Unternehmen sehen Kundenzufriedenheit als zentrales Ziel, welches in fast allen Fällen auch durch die Unternehmensleitung zumindest weitgehend unterstützt wird. Höchste Priorität kommt dabei der Ermittlung eines Zufriedenheitsscores zu (zur reinen Performance-Messung). In der Prioritätenliste folgen die Ziele „Analyse der Kundenanforderungen und -erwartungen“ sowie die „Ermittlung der Bedeutung einzelner Leistungsmerkmale für die Kunden“ (STAUSS

et al., 2009, S. 595). Unternehmen erkennen inzwischen offenbar auch den von KATCHER (2003) angesprochenen Wert von Kundenzufriedenheitsbefragung als Werkzeug, um Verbesserungspotentiale zu identifizieren. Die genaue Kenntnis von Zufriedenheitstreibern erleichtert die zielgerichtete Investition zur Verbesserung der Zufriedenheit.

3.1.1. Zusammenhang mit Unternehmenserfolg

Die langfristige finanzielle Performance von Unternehmen hängt nach überwiegender Auffassung in der Literatur mit der Loyalität der Kunden zusammen (vgl. JONES und SASSER, 1995). Auch MATZLER und BAILOM (2006, S. 243) sehen Kundenzufriedenheit als Indikator für den Erfolg von Unternehmen. Unzufriedenheit (SHARMA et al., 1999) oder zu geringe Zufriedenheit (JONES und SASSER, 1995) gelten als Hauptursache für die Beendigung von Kundenbeziehungen. Die meisten Unternehmen sind jedoch auf Wiederholungskäufe angewiesen, um profitabel zu operieren (OLIVER, 2010, S. 5). Dabei werden unterschiedliche Mediatoren wie insbesondere Kundenloyalität eingeführt. Auch in der Service-Profit-Chain wird von dem Zusammenhang Zufriedenheit → Loyalität → Unternehmenserfolg ausgegangen (HESKETT et al., 1994).

Empirische Untersuchungen Systematische Untersuchungen zum (langfristigen) Zusammenhang von Kundenzufriedenheit mit ökonomischen Kenngrößen wie Marktanteil und Marktkapitalisierung bestätigen einen Zusammenhang. GUPTA und ZEITHAML (2006, S. 725) haben einen Überblick über den Effekt von Zufriedenheit auf finanzielle Performance zusammengestellt. Sie fassen generalisierend zusammen, dass Verbesserungen der Kundenzufriedenheit einen signifikanten positiven Einfluss auf die finanzielle Performance von Unternehmen haben.

ANDERSON et al. (1994) stellen fest, dass Unternehmen mit überdurchschnittlicher Kundenzufriedenheitsbewertung auch überdurchschnittliche Ergebnisse erzielen. Ein Zufriedenheitspunkt (auf einer Zehnerskala) mehr Zufriedenheit wird mit einer Erhöhung des EAT (Earnings After Tax) um 11.5% in Verbindung gebracht. Allerdings, so ANDERSON et al. (1994), würde sich dieser Zusammenhang oft erst langfristig offenbaren. ITTNER und LARCKER (1998) untersuchen aus der Perspektive des Rechnungswesens, inwieweit

sich Kundenzufriedenheit als Indikator für künftiges Kaufverhalten der Kunden eignet und somit auch zur Erklärung des Börsenwertes herangezogen werden kann. Dabei stiftet Kundenzufriedenheit ihren Ergebnissen zufolge einen zusätzlichen Informationswert, der in klassischen Kenngrößen nicht enthalten sei.

GRUCA und REGO (2005) untersuchen anhand von Zeitreihen den Zusammenhang von Zufriedenheit und ökonomischen Größen und kommen zu dem Schluss, dass Kundenzufriedenheit durch Wachstum und Stabilisierung von Cash Flows positiv auf den Unternehmenswert wirke.

Eine Vielzahl von Untersuchungen verwendet den ACSI oder andere Zufriedenheitsindizes als Operationalisierung von Zufriedenheit. Die Unterscheidung von Zufriedenheit und „negativer Zufriedenheit“ als separate Konstrukte konnte sich zwar nicht durchsetzen, dennoch ist die Operationalisierung mittels negativer Zufriedenheitsaspekte ein interessanter Ansatz. LUO (2007) betrachtet den Zusammenhang von Beschwerdehäufigkeit und Börsenwert und findet ebenfalls einen signifikanten Zusammenhang auf der Basis „harter“ Daten. Beschwerden lassen sich zählen, während Zufriedenheitsindizes auf mittels Skalen abgefragten Zufriedenheitsmerkmalen basieren.

GUPTA und ZEITHAML (2006, S. 734) bemerken allerdings auch, dass Kundenzufriedenheit in den allermeisten Untersuchungen als exogen behandelt werde, obwohl die unterschiedliche Behandlung von Kunden je nach Kundenwert bekannt sei. Diese Adressierung von Endogenität ist folglich eine der großen Herausforderungen bei der Beurteilung der finanzwirtschaftlichen Relevanz von Kundenzufriedenheit.

Net Promoter Der Net Promoter Score (NPS), von REICHHELD (2003) als überlegene Metrik angepriesen, ist ebenfalls Grundlage diverser empirischer Untersuchungen zum Zusammenhang von Zufriedenheit und Unternehmenserfolg. KEININGHAM et al. (2007) stellen bei einer Untersuchung mit Zeitreihendaten jedoch fest, dass der NPS kein besserer Prädiktor der Unternehmensperformance ist als Indizes, die nach dem ACSI-Modell⁹ berechnet werden. Der Zusammenhang von Zufriedenheit und Unternehmenserfolg wird dadurch jedoch nicht in Frage gestellt.

⁹Siehe dazu Abschnitt 3.3.3.

Kosten Langfristige Kundenbeziehungen sind jedoch nicht nur auf der Einnahmenseite, sondern auch auf der Kostenseite relevant und können somit zum Unternehmenserfolg beitragen. Insbesondere bei Dienstleistungsunternehmen ist das Bedienen loyaler Kunden profitabler (REICHHELD und SASSER, 1990; REICHHELD, 1996). Die Kette Zufriedenheit → Loyalität → Unternehmenserfolg dürfte vermutlich der wichtigste Grund für das große Interesse von Unternehmen an Kundenzufriedenheit sein.

3.1.2. Loyalität und Kundenbindung

Der Zusammenhang von Zufriedenheit und Loyalität wurde auch unabhängig von einer Betrachtung des Unternehmenserfolgs untersucht, da ein Zusammenhang von Loyalität und Unternehmenserfolg u. a. durch Beiträge von REICHHELD und SASSER (1990); REICHHELD (1996) als gegeben angenommen wird.

Dabei werden auch unterschiedliche moderierende Variablen bei der Betrachtung von Loyalität aufgenommen (z. B. DE RUYTER und BLOEMER, 1999). Grundsätzlich scheint jedoch Konsens darüber zu bestehen, dass Kundenzufriedenheit zumindest eine Determinante von Loyalität ist.

Einige Forscher legen viel Wert auf die Unterscheidung von Loyalität und Kundenbindung (engl. „customer retention“) (u. a. JACOBY und KYNER, 1973; ASSAEL, 1992). Die Diskussion um Loyalität findet dabei vor allem in der Markenliteratur im Kontext von Markentreue bzw. Markenloyalität statt. Ausgangspunkt der Überlegungen ist, dass es sich bei Loyalität um „mehr“ handelt als um reines Wiederkaufverhalten. Somit kann diese Art Loyalität ihrerseits wieder in ihrer Wirkung auf Kundenbindung untersucht werden.

BROWN (1952); KUEHN (1962) machen Loyalität am beobachteten Wiederkaufverhalten fest. Daran bemängelt DAY (1969), dass wahre Loyalität so nicht identifiziert werden könne. Die Einstellung zur Marke könnte sich z. B. in Abwesenheit von Alternativen nicht im beobachteten Verhalten bemerkbar machen. JACOBY und KYNER (1973) weisen schließlich Unterschiede zwischen Loyalität und Wiederkauf nach und stellen dazu sechs Bedingungen für Loyalität auf. Auch ASSAEL (1992, S. 87) beschreibt Markenloyalität als Einstellung, die sich in einem konsistenten Kaufverhalten über die Zeit äußere.

Darüber hinaus kann auch affektive und kognitive Loyalität unterschieden werden (ALLEN und RAO, 2000, S. 8f.). Affektive Loyalität hängt dabei mit emotionalen Aspekten und der Interaktion mit Menschen zusammen, während der kognitive Teil sich aus der Bewertung von Leistungskriterien erklärt.

Auf eine genauere Diskussion des Loyalitätsbegriffs wird an dieser Stelle verzichtet. Unabhängig davon, wie Loyalität und Kundenbindung operationalisiert werden, gilt Kundenzufriedenheit als ursächlich für Loyalität und/oder Kundenbindung.¹⁰ So beschreiben z. B. ALLEN und RAO (2000, S. 132-133) ein Pfadmodell mit exogenen Variablen (Produktqualität, Produktlieferung, Beziehung zum Verkäufer, Preis, Kundendienst) und zwei endogenen Variablen (Gesamtzufriedenheit und Loyalität). Gesamtzufriedenheit stellt sich dabei als loyalitätswirksam heraus. Auch HALLOWELL (1996) weist einen Zusammenhang zwischen Kundenzufriedenheit, Kundenbindung und Profitabilität nach.

Während sich Kundenbindung vor allem bei vertraglichen Geschäftsmodellen als sinnvolles Konzept erweist, wird besonders bei nichtvertraglichen und stärker transaktional geprägten Geschäftsbeziehungen eher der Share of Wallet betrachtet. Analog zur Verbindung von Kundenzufriedenheit und Kundenbindung untersuchen COOIL et al. (2007) den Zusammenhang von Kundenzufriedenheit und Share of Wallet.

Das in dieser Arbeit vorgestellte Modell ist in der Lage, verschiedene endogene Variablen zu berücksichtigen, sofern die entsprechenden Konstrukte abgefragt werden.

3.1.3. Kundenzufriedenheit und wertorientiertes Management

Die Entwicklung hin zu einer wertorientierten Unternehmensführung (COPELAND et al., 1993) wurde zudem durch die Idee des Shareholder-Value-Ansatzes (RAPPAPORT, 1981, 1986) geprägt. Der Effizienzgedanke wurde schließlich im Konzept des Kundenwerts berücksichtigt (MATZLER et al., 2006b, S. 7).

Im Marketing hat in den letzten zwei Jahrzehnten ein Paradigmenwechsel stattgefunden (GELBRICH und MÜLLER, 2006, S. 453f.). Langfristige Kundenbeziehungen und Kundenzufriedenheit standen bereits in den 1980er Jahren im Fokus der Unternehmensführung.

¹⁰Einen entsprechenden Überblick über empirische Untersuchungen haben HOMBURG und STOCK (2003, S. 58-59) zusammengestellt.

Jedoch hat sich schnell die Erkenntnis durchgesetzt, dass nicht alle (treuen) Kunden profitabel sind und zum Unternehmenserfolg beitragen. Nach dem aktuellen Paradigma der Kundenwertorientierung werden Kunden als Investitionsgüter betrachtet (BLATTBERG et al., 2001; GUPTA et al., 2004; BORLE et al., 2008). Dabei hat sich insbesondere das Konzept des „Customer Lifetime Value“ in Theorie und Praxis durchgesetzt.

Bereits RUST und ZAHORIK (1993) geben Ansätze zur Beurteilung des Werts von Kundenzufriedenheit. WOODRUFF (1997) bezeichnen Kundenzufriedenheit als potentielle Quelle von Wettbewerbsvorteilen. In ähnlicher Weise beschäftigen sich RUST et al. (2004) mit dem „Return on Marketing“. Insbesondere in den letzten Jahren gibt es immer mehr Ansätze zur finanzwirtschaftlichen Rechtfertigung der Aktivitäten im Marketing. Auch in der deutschsprachigen Literatur wird der Zusammenhang zwischen Kundenwert und Kundenzufriedenheit ausführlich diskutiert (z. B. STAHL et al., 2006).

3.2. Definition von Kundenzufriedenheit

Bereits die Definition des Begriffs Kundenzufriedenheit nimmt breiten Raum in der Literatur ein. Kundenzufriedenheit wird sehr unterschiedlich definiert. So merkt OLIVER (2010, S. 7) an, dass scheinbar jeder wüsste, was Kundenzufriedenheit sei, bis er um eine Definition gebeten werde.

Eine ganz einfache Definition nimmt FORNELL (1992, S. 11) vor ("[...] satisfaction is an overall postpurchase evaluation."). Kundenzufriedenheit ist demnach eine gesamthafte Bewertung nach einem Kauf. Diese Definition kommt einem intuitiven Verständnis von Zufriedenheit nahe und ist zudem so allgemein gehalten, dass sie prinzipiell unabhängig von etwaigen Untersuchungszielen ist.

Zahlreiche Definitionen basieren auf dem Confirmation-Disconfirmation-Paradigma. Eine diesem Paradigma entspringende Definition von OLIVER (1997, S. 13) nimmt Bezug auf die Idee der „Erfüllung“:

„Satisfaction is the consumer’s fulfillment response. It is a judgment that a product/service feature, or the product or service itself, provided (or is providing) a pleasurable level of consumption-related fulfillment, including levels of under- or overfulfillment.“

Darüber hinaus gibt es noch eine ganze Reihe anderer Definitionen, sodass es in der Literatur keinen Konsens über eine einheitliche Definition gibt. Dies kritisieren u. a. WESTBROOK und REILLY (1983): „One of the major obstacles to the development of theory in consumer satisfaction/dissatisfaction has been a lack of attention to the specification of the construct.“ Auch PETERSON und WILSON (1992, S. 62) sowie GIESE und COTE (2000, S. 1) kritisieren die Uneinheitlichkeit von Definitionen und Methoden in der Kundenzufriedenheitsforschung.

Diese Vielfalt lässt sich schon durch den großen Umfang der Literatur erklären. Anfang der 90er Jahre des letzten Jahrhunderts zählte PETERSON und WILSON (1992, S. 61) bereits mehr als 15.000 Publikationen zur Kundenzufriedenheit. Kundenzufriedenheit wird dabei in vielen Bereichen untersucht: im öffentlichen Sektor (z. B. WISNIEWSKI, 2001), im Bereich der Gesundheit und Pflege (z. B. CROW et al., 2002), in der Bildung (z. B. ATHIYAMAN, 1997) und insbesondere in Unternehmen (KATCHER, 2003). Erschwerend kommt hinzu, dass es neben dem Dissens der konzeptionellen Definition auch unterschiedliche Operationalisierungen von Zufriedenheit gibt, die nicht in einer 1:1-Beziehung stehen.

Schließlich schlagen GIESE und COTE (2000) zur Klärung der Definition ein Bezugssystem vor, welches problembezogen angepasste Definitionen von Zufriedenheit erlaubt, jedoch ein größeres Maß an Vergleichbarkeit herstellt. Zunächst einmal stellen sie Gemeinsamkeiten in der Literatur entlang von drei Kriterien fest:

1. Kundenzufriedenheit ist demnach eine emotionale und/oder kognitive Reaktion unterschiedlicher Intensität.
2. Diese Reaktion bezieht sich auf einen bestimmten Fokus in Bezug auf Kauf oder Kaufentscheidung, d. h. es können unterschiedliche und unterschiedlich viele relevante Teilattribute der Zufriedenheit zur Geltung kommen.
3. Die Antwort wird zu einem bestimmten Zeitpunkt aufgenommen, typischerweise nach dem Konsum oder der Entscheidung für ein bestimmtes Produkt. Je nach Art des Produktes können unterschiedliche Zeitpunkte sinnvoll erscheinen. Sobald Kunden nach ihrer Zufriedenheit gefragt werden, löst dies einen Erinnerungs- und Bewertungsprozess aus. Daher kann die Wahl des Zeitpunktes Auswirkungen auf das Ergebnis dieses Bewertungsprozesses haben - beispielsweise in Abhängigkeit davon,

wie viel Zeit seit dem letzten Konsumerlebnis vergangen ist.

Die Autoren führen zusätzlich Befragungen durch und kommen schließlich zu dem Schluss, dass Kundenzufriedenheit stets als *affektive Reaktion unterschiedlicher Stärke* verstanden werden sollte (GIESE und COTE, 2000, S. 15). Der affektiven Reaktion sei jedoch ein kognitiver Prozess vorgeschaltet, welcher diese Reaktion determiniert. Zur Erklärung der Entstehung von Zufriedenheit könnten so auch kognitive Elemente untersucht werden, Zufriedenheit selbst sei aber affektiv. Dies sei auch konsistent mit OLIVER (1992, S. 242), der darüber schreibt „[...] satisfaction and dissatisfaction reflect the general affective tone.“ In Bezug auf Zeitbezug und Fokus ist der konzeptionelle Rahmen von GIESE und COTE (2000) weitestgehend konsistent mit der Literatur. Die Autoren schlagen vor, alle weiteren Definitionen von Kundenzufriedenheit an den vorgeschlagenen drei Eckpunkten zu orientieren, um Vergleichbarkeit herzustellen.

3.2.1. Konzeptionelle Grundlagen

HOMBURG und KOSCHATE (2007) definieren Kundenzufriedenheit als „das Resultat eines komplexen psychischen Vergleichsprozesses“. Auch dieser Definition liegt die am weitesten verbreitete Theorie zur Entstehung von Zufriedenheit zugrunde: das Confirmation-Disconfirmation-Paradigma (kurz: C/D-Paradigma) (OLIVER, 2010, S. 96f.).

Confirmation-Disconfirmation-Paradigma Mathematisch ausgedrückt ergibt sich Zufriedenheit (Z) als Differenz von Erwartung (E) und Leistung (L) (KANNING und BERGMANN, 2009):

$$Z = E - L$$

Die Erwartung (E) (oder auch Soll-Leistung) wird beeinflusst durch faktische Erwartungen, Ideale, individuelle Normen oder andere Maßstäbe (FOURNIER und MICK, 1999; HOMBURG und KOSCHATE, 2007). Der Kunde bildet somit eine Vorstellung von der „Soll-Leistung“. Die „Ist-Leistung“ (L) wird schließlich vom Kunden beurteilt und mit der Erwartung verglichen.

Erwartungen reflektieren die erwartete Leistung (CHURCHILL und SURPRENANT, 1982, S. 492). MILLER (1977) identifiziert vier Typen von Erwartungen: ideale Leistung, erwar-

tete Leistung, minimal annehmbare Leistung und wünschenswerte Leistung. DAY (1977) unterscheidet zwischen Erwartungen bezüglich des Produktes, Aufwand und Kosten zum Erreichen des Nutzens sowie soziale Kosten und sozialen Nutzen.

Die Ist-Leistung (L) kann als objektive Leistung (z. B. Bearbeitungszeit in Minuten) oder als subjektive Leistung verstanden werden. Die subjektive Leistung ist heterogen, da sie abhängig von der individuellen Wahrnehmung der Leistung ist (RUDOLPH, 1998, S. 20f.). Darüber hinaus können Kunden Bewertungen auf der Grundlage unterschiedlicher situationsspezifischer Standards vornehmen (CADOTTE et al., 1987, S. 313).

Der Vergleichsprozess selbst besteht eigentlich nur aus der Subtraktion $E - L$, sodass der Vergleichsprozess (Diskonfirmation) selbst nur schwer messbar ist (CHURCHILL und SURPRENANT, 1982, S. 493). OLIVER (1977) hingegen operationalisiert auch diesen Prozess separat. In der Folge gibt es drei Möglichkeiten:

1. positive Diskonfirmation, wenn $Z > 0$
2. Konfirmation, wenn $Z = 0$
3. negative Konfirmation, wenn $Z < 0$

Durch Konfirmation oder positive Diskonfirmation entsteht Zufriedenheit. In der Service-Literatur geht z. B. PARASURAMAN et al. (1991) davon aus, dass langfristig nur positive Diskonfirmation zu nachhaltiger und wettbewerbsrelevanter Zufriedenheit führen kann. Das Hauptmotiv der Zufriedenheit liegt in der Befriedigung von Bedürfnissen (OLIVER, 2010, S. 4). Zufriedenheit beginnt daher bereits mit der Konfirmation. CHURCHILL und SURPRENANT (1982) und HOMBURG und STOCK-HOMBURG (2008, S. 20) gehen zudem davon aus, dass die Ist-Leistung auch direkten Einfluss auf die Zufriedenheit nehmen kann.

Assimilations- und Kontrasttheorie Da Kundenzufriedenheitsbefragungen in aller Regel nach dem Kauf (bzw. Serviceerlebnis) durchgeführt werden, ist eine nachträgliche Anpassung der Erwartung nicht auszuschließen, da Kunden bei der Befragung auf Informationen aus ihrem Gedächtnis zurückgreifen müssen (LYNCH und SRULL, 1982; LICHTENSTEIN und SRULL, 1985). In diesem Zusammenhang spielen die Assimilations-Theorie, Kontrast-Theorie und Assimilations-Kontrast-Theorie eine Rolle (HOMBURG und STOCK-HOMBURG, 2008, S. 23f.). Bei der *Assimilationstheorie* geht man davon aus, dass die

Erwartungen oder Wahrnehmung ex-post soweit angepasst werden, dass der Vergleichsprozess den Zustand der Konfirmation erreicht ($Z = 0$). Die *Kontrasttheorie* beschreibt einen gegensätzlichen Effekt: Das Niveau der Diskonfirmation wird durch eine Überbewertung der Differenz $E - L$ vergrößert. Zusammengeführt werden beide Theorien in der *Assimilations-Kontrasttheorie* (HOMBURG und STOCK-HOMBURG, 2008, S. 234). Dabei wird angenommen, dass es einen Toleranzbereich gibt, innerhalb dessen der Vergleichsprozess wie gewohnt zu einem entsprechenden Ergebnis führt. Wie bereits erwähnt gehen z. B. PARASURAMAN et al. (1991) davon aus, dass Kunden nur dauerhaft gebunden werden können, wenn eine starke positive Diskonfirmation gelingt. Jedoch gibt es auch Zweifel an dieser Theorie (YAP und SWEENEY, 2007). Erwartungen können zudem auch bisherige Erfahrungen widerspiegeln ANDERSON et al. (2004).

Nichtlinearität von Zufriedenheit Insbesondere Theorien zum nichtlinearen Zusammenhang von Leistungserbringung und Zufriedenheit werden in der Literatur vielfältig diskutiert. Die Annahme, dass die Über-Erfüllung von Erwartungen zu einem sehr hohen Zufriedenheitsniveau führen kann, findet weite Verbreitung in der Literatur (HOMBURG und STOCK-HOMBURG, 2008, S. 32f.). Eines der bekanntesten Modelle in diesem Kontext ist das Kano-Modell (KANOU, 1984). Dieses Modell unterscheidet zwischen Basis-Merkmalen, Leistungs-Merkmalen und Begeisterungsmerkmalen. Basis-Merkmale entsprechen in etwa den Hygiene-Faktoren aus der Theorie von HERZBERG et al. (1959). Sie werden nur bei negativer Diskonfirmation relevant. Kunden erwarten Leistungsmerkmale und eine positive Diskonfirmation ist möglich. Begeisterungs-Merkmale werden hingegen nicht erwartet und bieten lediglich Potenzial für positive Diskonfirmation. Darüber hinaus kann es auch irrelevante Merkmale und K.O.-Kriterien geben. Einen kurzen Überblick über diesen und weitere Ansätze zur Differenzierung von Basis-, Leistungs- und Begeisterungsfaktoren findet sich im Beitrag von MATZLER et al. (2006a).

Konzeption und Umsetzung In vielen Fällen wird lediglich das Ergebnis des Vergleichsprozesses $Z = E - L$ gemessen: Kunden werden nach ihrer Zufriedenheit befragt. In anderen Fällen wird die Leistung bewertet, ohne den Vergleichsprozess (implizit oder explizit) zu berücksichtigen. Andere Untersuchungen operationalisieren hingegen alle Be-

standteile, sodass Fragen zur Zufriedenheit, zur Erwartung, zur Leistung und zum Vergleichsprozess selbst gestellt werden. Dies ist hilfreich, um den Entstehungsprozess von Zufriedenheit zu erklären. Liegt das Ziel jedoch in der Schätzung relevanter Treiber der Gesamtzufriedenheit, so genügt es, Zufriedenheit als Ergebnis des Vergleichsprozesses zu operationalisieren. Die separate Messung durch Erwartung und Leistung liefert dafür nicht unbedingt zusätzliche Information (KANNING und BERGMANN, 2009, S. 386).

3.2.2. Unterschiede der Definitionen

Unterschiede der Definitionen von Kundenzufriedenheit ergeben sich vor allem, aber nicht ausschließlich, entlang folgender Charakteristika:

- prozessorientierte oder ergebnisorientierte Definitionen
- kognitive oder affektive Definitionen
- transaktionsspezifische oder kumulative Zufriedenheit
- Spezifität
- Dimensionalität

Die Unterschiede in der Definition von Kundenzufriedenheit ergeben sich in vielen Fällen aus der Zielsetzung der zugrundeliegenden Untersuchung.

Auch GIESE und COTE (2000) sehen die Notwendigkeit kontextspezifischer Definitionen, die jedoch entlang verschiedener Kriterien vergleichbar sein sollten. Als Folge der jeweiligen Definition unterscheidet sich auch die Operationalisierung von Zufriedenheit. Einen Überblick über die verschiedenen Messverfahren findet sich bei BEUTIN (2008).

Ergebnis oder Prozess YI (1990, S. 69) unterscheidet Definitionen danach, ob das Ergebnis oder der Prozess im Vordergrund steht. Eher ergebnisorientiert sind die folgenden Definitionen:

„The buyer’s cognitive state of being adequately or inadequately rewarded for the sacrifices he has undergone“ (HOWARD und SHETH, 1969, S. 145)

“Consumer satisfaction, then, is an emotional response to the experiences provided by, or associated with, particular products or services purchased, retail

outlets, or even molar patterns of behavior such as shopping and buyer behavior, as well as the overall marketplace.“ (WESTBROOK und REILLY, 1983, S. 256)

„[...] the summary psychological state resulting when the emotion surrounding disconfirmed expectations is coupled with the consumer’s prior feelings about the consumption experience.“ (OLIVER, 1981, S. 27)

Prozessbetonende Definitionen sind z. B.:

„[...] an evaluation rendered that the experience was at least as good as it was supposed to be.“ (HUNT, 1977, S. 459)

„[...] the consumer’s response to the evaluation of the perceived discrepancy between prior expectations and the actual performance of the product as perceived after its consumption.“ (TSE und WILTON, 1988, S. 204)

Wird eine prozessbezogene Definition gewählt, so geht es zumeist um die Erklärung der Entstehung von Kundenzufriedenheit. Es ist allerdings nicht immer einfach, Konstrukte wie Erwartung, Leistung und Diskonfirmation trennscharf zu definieren (GIESE und COTE, 2000, S. 1). Zielsetzung dieser Arbeit ist nicht, den Entstehungsprozess von Zufriedenheit im Sinne des Confirmation-Disconfirmation-Paradigma zu erklären. Vielmehr wird der Zusammenhang von Teilzufriedenheiten (oder Teilleistungen) mit der Gesamtzufriedenheit (oder einer anders gearteten Gesamtbewertung) untersucht. Daher folgt diese Arbeit der ergebnisorientierten Sichtweise, die sich laut GIESE und COTE (2000, S. 1) inzwischen in den meisten Veröffentlichungen durchgesetzt hat.

Kognitiv oder affektiv Auch in Bezug auf die Einordnung als kognitives oder affektives Konstrukt gibt es unterschiedliche Auffassungen in der Literatur. HOWARD und SHETH (1969, S. 145) beschreiben Kundenzufriedenheit als kognitives Konstrukt: „[...] the buyer’s cognitive state of being adequately or inadequately rewarded for the sacrifices he has undergone.“ Dem gegenüber verstehen HALSTEAD et al. (1994, S. 122) Kundenzufriedenheit als affektive Reaktion:

„Satisfaction has generally been defined as a transaction-specific affective response resulting from the customer’s comparison of product performance to some prepurchase standard.“

SCHARNBACHER und KIEFER (2003, S. 5) definieren Kundenzufriedenheit als „emotionale Reaktion eines Kunden auf eine unternehmerische Leistung.“ PFAFF (1977) hingegen vertritt die Auffassung, dass kognitive und affektive Aspekte zur Beschreibung von Zufriedenheit herangezogen werden können. Das Verständnis von Zufriedenheit manifestiert sich in der Gestaltung der Befragung von Kunden. In vielen Fällen werden auch affektive Aspekte mit aufgenommen. Die in dieser Arbeit untersuchten Effekte können sich ebenfalls auf kognitive und affektive Evaluierungen beziehen. GIESE und COTE (2000) verstehen Zufriedenheit selbst als affektiv. Dem vorgelagert laufen kognitive Prozesse ab, deren Erklärung von besonderem Interesse ist (z. B. mittels Zufriedenheitstreiberanalysen).

Transaktional oder kumulativ Definitionen unterscheiden sich auch in ihrem Zeitbezug. Kundenzufriedenheitsbefragungen können situativ durchgeführt werden. Beispielsweise kann nach einem Serviceereignis die Zufriedenheit mit diesem Ereignis abgefragt werden. Andererseits kann die kumulative Zufriedenheit, welche einem Lernprozess unterliegt, abgefragt werden. Oft wird jedoch diese Unterscheidung nicht explizit getroffen (vgl. BITNER und HUBBERT, 1994, S. 74). Die Gesamtbewertung sowie die Bewertung einer Transaktion sind zudem hoch korreliert (vgl. BITNER und HUBBERT, 1994, S. 72).

Steht ein ganzes Kundenpanel zur Verfügung, welches zur kontinuierlichen Erhebung der Zufriedenheit verwendet werden kann, können darüber hinaus dynamische Effekte der Zufriedenheit analysiert werden (siehe z. B. VAN DOORN, 2004). ANDERSON et al. (2004) weisen zudem darauf hin, dass Erwartungen langfristig einen positiven Effekt auf Kundenzufriedenheit haben können, da Erwartungen auch die akkumulierten Erfahrungen aufnehmen. Dieser Effekt lässt sich ebenfalls in einer Untersuchung mit Zeitreihendaten genauer untersuchen. Entsprechend dem Konfirmation-Diskonfirmation-Paradigma resultiert Zufriedenheit aus dem Vergleich von Soll- und Ist-Leistung. Dieser Vergleich ist typischerweise situationsspezifisch (vgl. HUNT, 1977; OLIVER, 1980). Erwartungen bezüglich

einer Kundenbeziehung oder einer daraus resultierenden Einstellung sind weniger intuitiv als Erwartungen bezüglich eines Konsumerlebnisses. BITNER (1990) findet in einem Experiment zudem eine starke Korrelation zwischen der Bewertung eines Serviceereignisses und Servicequalität (die vom Konzept her einer längerfristigen Einstellung ähnelt), sodass diese Fragestellung vor allem konzeptioneller Natur ist.

Für die praktische Anwendung in Unternehmen ist jedoch die kumulative Zufriedenheit von größerem Interesse. Für das CRM, die Bewertung von Kunden und deren Abwanderungsgefahr, ist die Zufriedenheit mit einem einzelnen Erlebnis weniger relevant als die über die gesamte Beziehungsdauer entwickelte Zufriedenheit. Seit den 1990er Jahren findet sich daher vermehrt die kumulative Sicht der Zufriedenheit, wie z. B. in FORNELL (1992); BOULDING et al. (1993); BOLTON und DREW (1991); JOHNSON et al. (1991); SHARMA et al. (1999). Schließlich ist die Kundenzufriedenheit für die Anwendung in Unternehmen nur soweit relevant, wie sie Loyalität und/oder Beziehungsintensität beeinflusst.

Spezifität Definitionen von Zufriedenheit unterscheiden sich in Bezug auf die Spezifität. Nach YI (1990, S. 70) sind typische Beispiele: Produkt, Konsumerlebnis, Erlebnis der Kaufentscheidung, Vertreter, Geschäft, Attribut oder Vorkaufserlebnis.

Dimensionalität Die Faktorstruktur von Zufriedenheit ist auch ein wichtiges Diskussionsfeld in der Literatur (YI, 1990, S. 72f.). Ähnlich zur Zwei-Faktor-Theorie von HERZBERG et al. (1959) über die Arbeitszufriedenheit untersuchen beispielsweise CZEPIEL et al. (1974) die Faktorstruktur von Kundenzufriedenheit. Die Idee dahinter ist, dass Kundenzufriedenheit und Kundenunzufriedenheit zwei unabhängige und unterschiedliche Konstrukte sind. Auch ANDREASEN (1982, S. 184) unterscheidet zudem die Betrachtung von Zufriedenheit gegenüber Unzufriedenheit.

Bestätigt wird die zweifaktorielle Struktur von SWAN und COMBS (1976) und MADDUX (1981). LEAVITT (1977) konnte sie jedoch nicht bestätigen. Laut KAISER (2005, S. 46) konnte sich die Idee Herzbergs in der Kundenzufriedenheitsforschung nicht durchsetzen. In den letzten Jahren wird jedoch die Frage nach nicht-linearen Zusammenhängen von Teilzufriedenheiten zur Gesamtzufriedenheit häufiger diskutiert (z. B. TING und CHEN, 2002; NAGAO et al., 2006; FINN, 2011). Jedoch beschäftigt sich keine der genannten Publikatio-

nen simultan mit der Adressierung von Skalenverwendungsheterogenität, Halo-Fehler und Nichtlinearität. Dies wäre jedoch notwendig, um eventuell vorhandene Nichtlinearitäten eindeutig zu identifizieren. Sofern Konsens über K.O.-Kriterien (oder „Hygiene-Faktoren“ (HERZBERG et al., 1959, S. 64f.)) besteht, stellt sich zudem die Frage, inwieweit sie Teil von Untersuchungen sein müssen. Solche Kriterien scheinen eher Teil der Definition eines zu bewertenden Produktes oder Services zu sein. Daher erscheint die Modellierung von Nichtlinearität nur bei heterogenen Segmenten sinnvoll zu sein.

Neben der Frage, ob Zufriedenheit an sich auf einer eindimensionalen Skala beurteilt werden kann, stellt sich auch die Frage nach der Dimensionalität der Kriterien für Zufriedenheit. So unterscheiden beispielsweise CZEPIEL und ROSENBERG (1977) sieben Dimensionen der Kundenzufriedenheit: Kaufprozess, Entscheidung, Funktionalität, Ästhetik, psychologische Effekte, Service, Umfeld.

Derartige mehrdimensionale merkmalsorientierte Modelle sind inzwischen üblich, um Kundenzufriedenheit zu verstehen. Eine ähnlich komplexe Erhebung von Zufriedenheit ist auch Basis der empirischen Anwendung in dieser Arbeit.

3.2.3. Abgrenzung zu anderen Konstrukten

Neben dem Konstrukt der Kundenzufriedenheit gibt es auch die verwandten Konstrukte der Performance, Dienstleistungsqualität und Einstellung. In diesem Abschnitt wird versucht, Kundenzufriedenheit von anderen Konstrukten abzugrenzen.

Einstellungen Kundenzufriedenheit oder Zufriedenheit im Allgemeinen ist der Einstellung relativ ähnlich. Ähnlich wie bei dem Fishbein-Modell (FISHBEIN, 1963), in dem Einstellung als Gewichtung von Überzeugung bezüglich bestimmter Eigenschaften dargestellt wird, wird Gesamtzufriedenheit in multiattributiven Modellen als gewichtete Summe von Teilzufriedenheiten dargestellt.

„Conceptually, satisfaction is an outcome of purchase and use resulting from the buyer’s comparison of the rewards and costs of the purchase in relation to the anticipated consequences. Operationally, satisfaction is similar to attitude in that it can be assessed as the sum of the satisfaction with the various

attributes of the product or service.“ (CHURCHILL und SURPRENANT, 1982, S. 493)

LATOUR und PEAT (1979) sehen den größten Unterschied zwischen Zufriedenheit und Einstellung auf der Zeitachse: Einstellungen seien demnach vor der Kaufentscheidung relevant, Zufriedenheit hingegen nach dem Kauf. Die Konstrukte seien aber möglicherweise auch gar nicht unterscheidbar:

„[...] we should be careful about attaching any special significance to satisfaction as a construct when it may simply be one measure of an attitude.“ (LATOUR und PEAT, 1979, S. 434)

Einstellungsmodelle und mehrdimensionale Kundenzufriedenheitsmodelle sind sich auch technisch gesehen sehr ähnlich. Zufriedenheitstreibermodelle stellen Gesamtzufriedenheit als gewichtete Summe von Teilzufriedenheiten dar. Ein gewichtetes additives Nutzenmodell (COHEN und AHTOLA, 1971; FISHBEIN, 1965) modelliert Einstellung als gewichtete Summe verschiedener Attribute.

Performance Gelegentlich wird statt der Zufriedenheit (Gesamt- und Teil-) die Leistung (Performance) gemessen - potentiell auch unabhängig davon, ob konzeptionell eigentlich Zufriedenheit erhoben werden sollte.

In der Literatur haben viele Autoren Wert darauf gelegt, dass Leistung und Zufriedenheit zwar eng miteinander verwandt sind, es sich aber dennoch um unterschiedliche Konstrukte handelt. So definieren CHURCHILL und SURPRENANT (1982, S. 492) „Erwartung“, „Leistung“, „Diskonfirmation“ und „Zufriedenheit“ als getrennte Konstrukte. KANNING und BERGMANN (2009) stellen jedoch fest, dass Leistung den größten Erklärungsbeitrag liefert und Erwartungen kaum einen Einfluss auf die Zufriedenheit haben:

„The variable of 'performance' was thus confirmed as the most important predictor of customer satisfaction, with the other two variables ("expectation" and "importance") not providing any noteworthy improvement in the prediction.“ (KANNING und BERGMANN, 2009, S. 386)

Während es also offenbar einfach ist, konzeptionell sehr unterschiedliche Konstrukte zu entwickeln, stellt sich jedoch die Frage, inwieweit solche Unterschiede wirklich operationalisiert werden können. Betrachtet man zudem mögliche Fehlerquellen (etwa die nachträgliche Anpassung von Erwartungen entsprechend der Assimilationstheorie), so ist Vorsicht bei einer konzeptionell sehr feingliedrigen Differenzierung unterschiedlicher Konstrukte angebracht. Relevant für die Entscheidung in Unternehmen ist vielmehr, welche Teilaspekte relevante Treiber der Gesamtzufriedenheit sind.

Dienstleistungsqualität Auch zwischen Kundenzufriedenheit und Dienstleistungsqualität (STAUSS, 1999, S. 5) gibt es Überschneidungen. Das SERVQUAL-Modell (PARASURAMAN et al., 1988) ist ein weit verbreiteter Ansatz zur Erhebung von Dienstleistungsqualität. BITNER und HUBBERT (1994) bezeichnen dieses Konstrukt als abstrakter im Vergleich zur Gesamtzufriedenheit. Sie erkennen aber auch, dass diese Konstrukte empirisch nicht leicht zu trennen sind.

3.2.4. Definition für diese Arbeit

Die vorliegende Arbeit untersucht die Beziehung von Attributs- und Gesamtbewertungen in multiattributiven Modellen. Dabei ist die Gestalt der Attributsurteile zunächst einmal sekundär.

Große Teile der vorgestellten Zufriedenheitsliteratur basieren auf dem Confirmation-Disconfirmation-Paradigma. Häufig wird daher der Zusammenhang von Erwartungen, Leistungen, Diskonfirmationsprozess und Zufriedenheit untersucht. Ziel solcher Untersuchungen ist die Erklärung von Beurteilungsprozessen. Ähnliche Ziele verfolgen Untersuchungen zur Assimilationstheorie, Kontrasttheorie, Assimilations-Kontrast-Theorie oder der Attributionstheorie, die hier nur sehr kurz erwähnt werden.

Die Ergebnisse derartiger Untersuchungen sind kaum praktisch nutzbar. Zudem bekunden KANNING und BERGMANN (2009, S. 388) und STAUSS (1999, S. 13), dass eine direkte Frage nach Zufriedenheit zur Messung von Zufriedenheit völlig ausreichend sei.

Die einfache Definition von FORNELL (1992, S. 11) ("[...] satisfaction is an overall post-purchase evaluation") lässt sich auch auf multiattributive Modelle übertragen. So können

Kunden eine „overall postpurchase evaluation“ auch bezüglich unterschiedlicher Attribute vornehmen. In einem nächsten Schritt können dann Zusammenhänge von Teilzufriedenheiten mit der Gesamtzufriedenheit untersucht werden.

In der Praxis wird jedoch nicht immer nach Zufriedenheit gefragt („Wie zufrieden sind Sie mit Attribut A von Produkt X?“). Es werden auch häufig Leistungsbewertungen vorgenommen („Wie bewerten Sie die Leistungsfähigkeit von Produkt X bezüglich Attribut A?“).

Die Feststellungen von KANNING und BERGMANN (2009); STAUSS (1999) legen nahe, dass die Frage nach der Messung von (Teil-)Zufriedenheiten kaum zu neuen Erkenntnissen führt. Für nähere Ausführungen zu solchen Theorien sei daher auf Beiträge wie von HOMBURG und STOCK (2003, S. 23ff.) verwiesen.

Die Fragestellung dieser Arbeit ist letztendlich losgelöst von der Frage, wie Urteile definiert werden (ob als Leistung oder Zufriedenheit). Die Bewertung von Leistung und Zufriedenheit haben sich in den bisherigen Untersuchungen stets als hoch miteinander korreliert herausgestellt. Wichtiger für Entscheidungen in Unternehmen ist daher die Frage, wie Teilbewertungen und Gesamtbewertungen miteinander zusammenhängen, um Ressourcen an den richtigen Stellen zur Verbesserung der Produkte einsetzen zu können.

3.3. Verfahren zur Messung von Kundenzufriedenheit

REICHELDELD (2003) reduziert die Messung von Kundenzufriedenheit auf eine einzelne Größe - den Net Promoter Score - , die lediglich durch eine Frage operationalisiert ist. Als Werkzeug zur Performance-Messung ist eine derart einfache Messung durchaus ausreichend, wenn auch der NPS entgegen der Behauptung von REICHELDELD (2003) nicht nachweislich eine überlegene Operationalisierung darstellt (KEININGHAM et al., 2007).

Der Wert von Kundenzufriedenheitsuntersuchungen liegt aber in der Identifikation von Treibern der Zufriedenheit (MCDUGALL und LEVESQUE, 2000; KATCHER, 2003). Inzwischen wird dies auch als wichtiges Ziel des Kundenzufriedenheitsmanagements in Unternehmen erkannt (STAUSS et al., 2009, S. 595).

Bei der Untersuchung von Treibern stellt sich nicht nur die Frage nach einer geeigneten Messung von Zufriedenheit als Gesamtzufriedenheit, sondern auch nach der Messung von

Teilzufriedenheiten bezüglich einzelner Leistungsmerkmale.

Darüber hinaus können auch andere Fragestellungen, etwa zum Einfluss von Erwartungen, Markenimage oder Ruf des Unternehmens auf die Zufriedenheit, von Interesse sein.

3.3.1. Differenzierung unterschiedlicher Verfahren

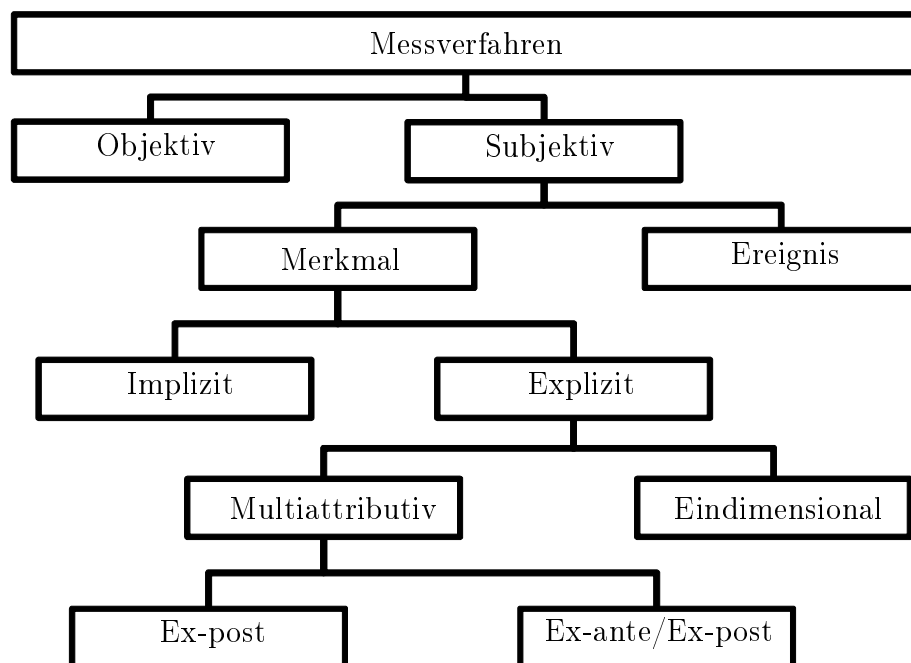
Bei der Messung von Kundenzufriedenheit unterscheiden HOMBURG und FÜRST (u. a. 2008); BEUTIN (u. a. 2008); RAAB (u. a. 2008) folgende Verfahren:

- Wahrnehmung: objektiv oder subjektiv
- Direktheit: implizit oder explizit
- Orientierung: Ereignis oder Merkmal
- Dimensionalität: eine oder mehrere
- Zeitpunkt: ex ante/ex post oder ex post (mit oder ohne separater Messung von Erwartung und Erfüllung)

Die Zusammenhänge sind in Abbildung 3.1 dargestellt. Die häufigsten Methoden sind multiattributive Modelle, welche ex-post Gesamt- und Teilzufriedenheiten messen. Nur bei diesen Modellen ist eine Treiberanalyse möglich.

Objektive oder subjektive Verfahren Zunächst lassen sich demnach Messverfahren in objektive und subjektive Verfahren unterscheiden. Bei *objektiven Verfahren* werden nur beobachtbare Indikatoren herangezogen, sodass keine subjektive Wahrnehmung berücksichtigt wird. Darunter fallen z. B. finanzielle Größen wie Umsatz, Marktanteil, EBIT oder auch Wiederholungskaufraten (BEUTIN, 2008, S. 127f.). Solche Indikatoren sind den Autoren zufolge jedoch kaum von den Determinanten zu trennen. Geht man davon aus, dass Kundenzufriedenheit ursächlich für Unternehmenserfolg ist, so kann sich der Indikator „Marktanteil“ sowohl auf Erfolg als auch auf Kundenzufriedenheit beziehen. Treiberanalysen sind mit objektiven Verfahren also nicht durchführbar. Die Messung finanzieller Größen weist zwar i. d. R. nur geringe Fehler auf, jedoch ist der Rückschluss auf Zufriedenheit kaum möglich. Auch nicht teilnehmende Beobachtungen kann man zu den

Abbildung 3.1: Systematisierung unterschiedlicher Messverfahren



- in Anlehnung an RAAB (2008, S. 78) -

objektiven Verfahren zählen. So wird u. a. die Silent-Shopper-Methode angewendet oder es werden Warentests durchgeführt (KAISER, 2005, S. 122).

Sollen Treiberanalysen gemacht werden, sind *subjektive Verfahren* üblich. Subjektive (Teil-)Zufriedenheiten können Aufschluss darüber geben, welche Verbesserungspotentiale existieren. Die subjektiven Verfahren lassen sich wiederum in Verfahren der impliziten und expliziten Messung unterteilen.

Implizite und explizite Verfahren Die Messung von Zufriedenheit kann direkt oder indirekt erfolgen (HOMBURG und FÜRST, 2008, S. 606). Bei *impliziten Verfahren* wird die Zufriedenheitsinformation indirekt aus vorhandenen Daten gewonnen. So können z. B. Kundenbeschwerden, Reklamationen oder vergleichbare Daten dafür verwendet werden. Implizite Verfahren sind vor allem aufgrund der Nutzbarkeit dieser vorhandenen Daten interessant, jedoch ist es schwierig, darüber hinaus weitere Informationen zu gewinnen. Auch können Beschwerdeinformationen nur dann hilfreich sein, wenn unzufriedene Kunden sich auch tatsächlich beschwerten (BEUTIN, 2008; GÜNTER, 2008). Die Abwesenheit von Beschwerden bedeutet nicht zwangsläufig, dass Kunden zufrieden sind, zumal das Ausbleiben von Beschwerden auch auf mangelnde Aussicht auf Erfolg zurückgehen kann (SCHARNBACHER und KIEFER, 2003, S. 20). Das Ausbleiben von Beschwerden kann auch auf sachliche Gründe zurückgehen: No-Frills-Airlines haben ceteris paribus weniger Beschwerden bezüglich aufgegebenem Gepäck zu erwarten, da weniger Kunden diese (meist aufpreispflichtige) Leistung in Anspruch nehmen. BEUTIN (2008, S. 131) nennt implizite Verfahren schließlich „*unsystematisch und keinesfalls immer repräsentativ*“.

Explizite Verfahren erlauben es, deutlich bessere und zielgerichtete Schlüsse zu ziehen, da alle messbaren potentiellen Treiber und Konsequenzen von Zufriedenheit erhoben werden können. In den meisten Fällen werden daher explizite Verfahren verwendet. Es ist zudem festzustellen, dass multiattributive (und somit explizite) Modelle seit vielen Jahren üblich sind (siehe z. B. BEUTIN (2008, S. 132), BEARDEN und TEEL (1983, S. 25f.)).

Ereignis- oder merkmalsorientierte Verfahren *Ereignisbezogene Verfahren* bieten sich für Untersuchungen kritischer Kontakt Ereignisse (Kauf, Service-Ereignis) an (HOMBURG und WERNER, 1998, S. 128f.). Beispiele sind die Kontaktpunktanalyse, problemorientier-

te Verfahren bzw. Frequenz-Relevanz-Analyse (FRAP) (BRANDT und REFFETT, 1989), Analyse von Standardereignissen oder die Critical Incident Technique (FLANAGAN, 1954). An dieser Stelle wird jedoch nicht näher auf derartige Verfahren eingegangen, denn zur „[...] umfassenden Messung der Kundenzufriedenheit sind derartige Methoden [...] nicht geeignet.“ (HOMBURG und WERNER, 1998, S. 129.).

Merkmalsorientierte Verfahren lassen sich hingegen für eine Vielzahl von Fragestellungen einsetzen. Dabei geht man davon aus, dass unterschiedliche Merkmale des Produktes oder der Dienstleistung in dem Zufriedenheitsurteil münden. (vgl. HOMBURG und FÜRST, 2008, S. 606f.).

Ein- und mehrdimensionale Verfahren *Eindimensionale Verfahren* verwenden nur ein Konstrukt. Dieses Konstrukt wird schließlich oft auch nur mit einer Frage operationalisiert, wie etwa im Beispiel des Net Promoter Score. Häufig werden jedoch *multiattributive Verfahren* eingesetzt. Damit können die Zusammenhänge zwischen Teilzufriedenheiten bzw. Attributen und der Gesamtzufriedenheit untersucht werden (BEUTIN, 2008, S. 131f.). Zudem werden häufig Attribute durch mehrere Fragen operationalisiert, um Messfehler zu kontrollieren.¹¹

Weiter lassen sich multiattributive Verfahren in Ein- und Zweikomponentenansätze unterscheiden. Beim *Zweikomponentenansatz* werden Kunden explizit nach der Bedeutung einzelner Attribute befragt. Diese Art der Messung kann für die Importance-Performance-Analyse (MARTILLA und JAMES, 1977) verwendet werden. Dabei werden in einem Koordinatensystem die Leistungs-Bedeutungs-Kombinationen dargestellt. Somit kann man entsprechende Konsequenzen ziehen. Nachteilig für den Zweikomponentenansatz ist jedoch die Verdoppelung der notwendigen Fragen - zu jedem Attribut müssen Kunden zwei Antworten geben. Auch muss überprüft werden, inwieweit die Auskunft über die Wichtigkeit der Attribute die „wahre“ Gewichtung widerspiegelt. Die Antworten könnten verzerrt sein, z. B. weil sozial erwünschte Antworten gegeben werden. So ist denkbar, dass „Energieeffizienz“ durchgängig als sehr wichtig eingeschätzt wird, weil eine solche Antwort

¹¹So schlagen RYAN et al. (1995, S. 12) vor, das Konstrukt Gesamtzufriedenheit durch drei Fragen zu operationalisieren. Näher wird auf die Frage nach der Skala im nächsten Kapitel eingegangen.

in Zeiten steigenden Umweltbewusstseins mutmaßlich erwartet wird. Zudem sind Kunden nicht mit den Kosten zur Verbesserung der Attribute konfrontiert, sodass die Gefahr besteht, dass alle Attribute als sehr wichtig bewertet werden.

Bei *Einkomponentenansätzen* können die Gewichtungen mittels Regressionsanalysen oder Strukturgleichungsmodellen aus den Angaben zu Attributs- und Gesamtzufriedenheit geschätzt werden. Der Vorteil dabei ist, dass Kunden nur eine Angabe (Zufriedenheit) je Attribut abgeben müssen. Der Response Style der sozial erwünschten Antwort stellt bei Zweikomponentenansätzen potentiell ein größeres Problem dar. Dies lässt sich am Attribut „Umweltverträglichkeit“ verdeutlichen. Die direkte Frage nach der Bedeutung von „Umweltverträglichkeit“ bietet einen größeren Anreiz zur Verzerrung als die Frage nach der Zufriedenheit mit der „Umweltverträglichkeit“ eines Produktes oder Unternehmens.

„It is therefore preferable to infer the weightings indirectly for example by conducting a regression from satisfaction with the individual services/products to general satisfaction.“(KANNING und BERGMANN, 2009, S. 388)

Beim Einkomponentenansatz ist jedoch der Modellierungsaufwand größer, insbesondere wenn man Response Styles und Halo-Effekte kontrollieren will (siehe Abschnitt 2.3).

Der Unterschied zwischen *berechneten* und *direkt abgefragten* Bedeutungen der Attribute ist sowohl in der Zufriedenheits- als auch in der Einstellungsforschung von Interesse. So wird gelegentlich versucht, beide Gewichtungen (berechnete und von Kunden angegebene) in einer Analyse zusammenzubringen (BUSACCA und PADULA, 2005; GRIGOROUDIS und SPYRIDAKI, 2003; MATZLER und SAUERWEIN, 2002). Treiber, die sich nur in der Regressionsanalyse, nicht aber in der direkten Abfrage der Prioritäten zeigen, werden als versteckte Bedürfnisse betrachtet („Excitement Factors“ (MATZLER und SAUERWEIN, 2002, S. 320)). Auf der anderen Seite werden Treiber, die nur explizit, nicht aber im Regressionsmodell identifiziert werden, als Basisfaktoren bezeichnet. Diese Abweichungen sind jedoch methodischer Natur. Konzeptionell sollten beide Befragungsmethoden die gleichen Treiber identifizieren. Die Methoden unterliegen dabei möglicherweise unterschiedlichen Verzerrungen (sozial erwünschtes Antworten bzw. Halo), jedoch sollten sie theoretisch zum gleichen Ergebnis führen. Somit ist fraglich, inwieweit methodische Abweichungen als Grundlage für Managemententscheidungen (etwa Fokussierung auf „Excitement Fac-

tors“) herangezogen werden sollten. Eine genauere Beurteilung ist jedoch erst möglich, wenn Verzerrungen, etwa durch Response Styles, korrigiert werden.

In der Einstellungsforschung haben sich MYERS und ALPERT (1977) mit dem Unterschied zwischen berechneten und berichteten Gewichten beschäftigt. Sie unterscheiden dabei zwischen „Importance“ und „Determinance“. „Importance“ beschreibt Wichtigkeit oder Bedeutung im Wortsinn. „Determinance“ bedeutet, dass hinreichende Varianz der Bewertungen vorhanden ist, sodass K.O.-Kriterien nicht in diesen Bereich fallen. Berechnete Gewichte gehören den Autoren zufolge zur Bedingung der „Determinance“. Dies ergibt sich auch aus den technischen Möglichkeiten - Variablen ohne Varianz können nicht als unabhängige Variable in einem Regressionsmodell verwendet werden.

Ex ante / ex post oder ex post Üblich ist die *ex post*-Messung von Kundenzufriedenheit. Nach einem Konsumerlebnis werden Kunden nach ihrer Zufriedenheit bzw. Einschätzung der Leistung befragt (RUDOLPH, 1998, S. 55). Bei der *ex ante / ex post*-Messung werden in einem ersten Schritt die Erwartungen der Kunden bezüglich unterschiedlicher Attribute einer Leistung erhoben. Nach Erbringung der Leistung wird dann in einem zweiten Schritt abgefragt, inwieweit die Kunden ihre Erwartung erfüllt sehen (BEUTIN, 2008, S. 124).

Eine separate Messung von Erwartung und Erfüllung dieser Erwartungen ist in Experimenten oder mit einem Panel durchaus möglich, ansonsten in der Praxis kaum umsetzbar.

3.3.2. Zufriedenheitsskalen

„It is probably not possible to measure 'true satisfaction'.“ (PETERSON und WILSON, 1992, S. 69) Zur Messung von Kundenzufriedenheit werden meist multiattributive Verfahren eingesetzt. Die Beurteilung von Zufriedenheit, Leistung oder Erwartung erfolgt dabei i. d. R. auf einer Skala.

In der Literatur gibt es sehr viele Beiträge zur Gestaltung von bestmöglichen Beurteilungsskalen. MCREYNOLDS und LUDWIG (1987) verfolgen die Geschichte der Rating-Skalen zurück bis in das Jahr 1692. Oft wird aber auch erst GALTON (1880) als Erfinder dieser Skalen genannt. Auch wenn die Diskussion um die „beste“ Skala inzwischen brei-

ten Raum einnimmt, so scheint die Skalendiskussion für sich genommen wenig zu einer Verbesserung der Zufriedenheitsuntersuchungen beigetragen zu haben, wie an folgendem Zitat deutlich wird:

„Nach einem weiteren Vierteljahrhundert der Skalendiskussion kommt man somit wiederum zu dem „Aschenputtelbefund“ der Zufriedenheitsmessung, nach der die einfachste Methode, die direkte Frage nach der Zufriedenheit, die besten Ergebnisse zu bringen scheint.“ (STAUSS, 1999, S. 13)

Auch KANNING und BERGMANN (2009, S. 388) bestätigen diese Feststellung:

„[...] the best alternative is simply to ask how satisfied the customers are with a particular service or product. Questions regarding “expectation” or “importance” are unlikely to shed any further useful light on the phenomenon.“

Das in dieser Arbeit entwickelte Modell setzt erst nach der Erhebung von Zufriedenheitsinformationen an und hilft dabei, relevante Treiber zu identifizieren. Es ist nicht auszuschließen, dass die Art der Skala auch Einfluss auf den Anteil haloistischer Antworten nehmen kann. Dies muss jedoch in Experimenten geschehen, um den Einfluss der Skala isolieren zu können. Bisher haben Wissenschaftler vor allem die Validität und Reliabilität als Beurteilungskriterien herangezogen. Auch die Schiefe der Verteilungen wurde betrachtet, da zu starke Konzentration auf die Bestnoten den Informationsgehalt der Daten reduziert.

Die zahlreichen Beiträge zur Skalendiskussion drehen sich unter anderem um folgende Punkte:

- Was wird gemessen?
 - Messung von Leistung¹², Diskonfirmation¹³ oder Zufriedenheit¹⁴
- Wie wird gemessen?

¹²Z. B. „Wie bewerten Sie die Leistung von Produkt Y bezüglich Merkmal X?“.

¹³Z. B. „Inwieweit wurden Ihre Erwartungen bezüglich Merkmal X von Produkt Y erfüllt?“.

¹⁴Wie zufrieden sind Sie mit Merkmal X von Produkt Y?“.

- multiattributive Modelle, die auch attributspezifische Zufriedenheit erheben oder
 - eindimensionale Messung, wobei ausschließlich Gesamtzufriedenheit gemessen wird
 - Verwendung von Messmodellen zur Operationalisierung von Partial- und Gesamtbewertungen. Beispielsweise empfehlen RYAN et al. (1995, S. 12) die Operationalisierung von Gesamtzufriedenheit durch drei Fragen.
 - alternativ direkte Abfrage von Attributen ohne Messmodell
- Welche Skala wird verwendet?
 - Art der Skala: numerisch, verbal oder grafisch
 - Breite der Skala (z. B. 7, 9 oder 10 Punkte)
 - ungerade Stufen (z. B. $-2, -1, 0, 1, 2$) oder gerade Stufen (z. B. $-2, -1, 1, 2$)
 - Beschriftung der Ankerpunkte: bipolare oder unipolare Ausrichtung
 - Formulierung: bezogen auf kognitive oder affektive Elemente von Zufriedenheit

Gegenstand der Skala In ihrem vielzitierten Beitrag stellen CHURCHILL und SURPRENANT (1982, S. 495) einen Überblick über verschiedene Skalen zur Messung von Zufriedenheit zusammen. Sie unterscheiden dabei Messungen von Zufriedenheit, Erwartung, Leistung und Diskonfirmation und operationalisieren diese zunächst getrennt. Die Auswertung ergibt, dass es neben dem Diskonfirmationsprozess auch einen direkten Zusammenhang zwischen Leistung und Zufriedenheit gibt.

DANAHER und HADDRELL (1996) vergleichen drei unterschiedliche Messmethoden von Zufriedenheit - Leistung, Diskonfirmation und Zufriedenheit - direkt. Sie stellen fest, dass Zufriedenheit die höchste prädiktive Validität aufweist. Dies ist konsistent mit den späteren Empfehlungen von KANNING und BERGMANN (2009, S. 388) und STAUSS (1999, S. 13), Zufriedenheit abzufragen.

Messverfahren Im vorherigen Abschnitt 3.3.1. wurden bereits unterschiedliche Messverfahren vorgestellt. Die Messung durch Rating-Skalen ist dabei für die üblicherweise gebräuchlichen merkmalsorientierten Verfahren relevant. Üblich sind dabei multiattribu-

tive Messverfahren (BEARDEN und TEEL, 1983). Zudem werden in einigen Fällen auch Messmodelle eingesetzt, um Messfehler zu kontrollieren (RUDOLPH, 1998, S. 97ff.).

Gestaltung der Skala HAUSKNECHT (1990) unterscheidet Skalen nach zwei Dimensionen: kognitiv/affektiv und verbal/grafisch. In einem Überblick unterscheidet VAVRA (1997, S. 177-183) verbale und grafische Skalen für kognitive, affektive und konative Messverfahren von Zufriedenheit. Zunächst ist festzustellen, dass die Messmethoden so ausgewählt werden sollten, dass sie konsistent mit der konzeptionellen Definition von Zufriedenheit sind. Wird Kundenzufriedenheit als rein kognitives Konstrukt verstanden, müssen die Skalenpunkte auch entsprechend beschriftet sein.

Dabei ist darauf zu achten, dass innerhalb eines Fragebogens einheitliche Messskalen verwendet werden, um den Beantwortungsaufwand von Kunden gering zu halten (VAVRA, 1997, S. 146f.).

WESTBROOK und OLIVER (1981) untersuchen verbale, grafische, semantische Differenziale, Likert- und inferentielle Skalen. Die höchste Reliabilität weisen dabei Messmodelle auf, die das semantische Differenzial zur Messung einsetzen. HILL und ALEXANDER (2000, S. 132) bevorzugen eine verbale Skala gegenüber der Likertskala, weil die Fragen nach ihrer Auffassung natürlicher klingen würden. Sie liefern jedoch keine empirische Begründung.

Unabhängig davon, welche Skala verwendet wird, kann nicht zwingend davon ausgegangen werden, dass die Skala einheitlich interpretiert wird.

Zur Verteilung von Zufriedenheitsdaten Die Bewertungen auf Zufriedenheitsskalen folgen in den meisten Fällen schiefen Verteilungen, wobei die meisten Bewertungen sich am positiven Ende der Skala befinden (PETERSON und WILSON, 1992; YI, 1990).

DEVLIN et al. (1993, S. 14) zeigen, dass bei einer Qualitätsbewertung mittels Performance-Skala („excellent“, „above average“, „average“, „below average“, „very poor“) 54% der Befragten mit „excellent“ antworten. Bei der Verwendung einer Konfirmations-Diskonfirmations-Skala („exceeded“, „met“, „nearly met“, „missed“) geben hingegen nur 36% die beste Kategorie „exceeded“ an. Nach dem Kriterium der Schiefe wäre daher die Frage nach der Erfüllung von Erwartungen optimal. In Messmodellen werden folglich häufig unterschiedliche Skalentypen verknüpft, um entsprechende Konstrukte (wie Gesamtzufriedenheit) besser

bezüglich Validität und Reliabilität beurteilen zu können (RYAN et al., 1995).

PETERSON und WILSON (1992) nennen eine Reihe möglicher Erklärungen für die häufige Schiefe von Zufriedenheitsverteilungen:

- Die Kunden sind möglicherweise tatsächlich fast ausschließlich sehr zufrieden.
- Die Beobachtung geht zurück auf die Ursachen (wie Leistung oder Erwartungen), etwa durch konstant besonders gute Leistungen.
- Zufriedenheit weist eine Verteilung auf, die nicht die Normalverteilung ist.
- Die Schiefe geht zurück auf die Untersuchungsmethode, beispielsweise Verzerrung dadurch, dass nur zufriedene Kunden auf die Befragung reagieren, oder aufgrund der Art der Fragestellung.

Zur Herausforderung werden extrem schiefe Verteilungen, weil bei gegebener Skalenbreite einer ordinalen Rating-Skala (z. B. 7-Punkte-Skala) die Varianz der beobachteten Antworten möglicherweise zu gering für aussagekräftige Analysen ist. In vielen früheren Analysen wird davon ausgegangen, dass Zufriedenheit normalverteilt ist. Ist diese Annahme verletzt, so können auch die Ergebnisse von Analysen verzerrt sein (ANDREASSEN et al., 2006).

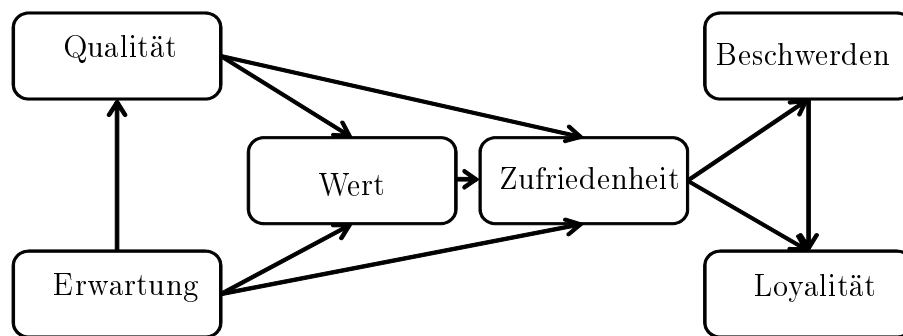
In den meisten Fällen werden Rating-Skalen verwendet, sodass lediglich ordinale Informationen zur Verfügung stehen. Diese lassen sich jedoch in kontinuierliche Informationen übertragen. Darauf wird in Abschnitt 4.4 näher eingegangen. Ein weiteres Problem ist die heterogene Nutzung von Rating-Skalen. Auf diese zentrale Herausforderung zur Modellierung von Zufriedenheit wurde in Abschnitt 2.2 näher eingegangen.

3.3.3. Zufriedenheitsindizes

Der vermutlich bekannteste Zufriedenheitsindex ist der ACSI, der amerikanische Kundenzufriedenheitsindex (FORNELL et al., 1996a). Derartige Indizes sollen zur Standardisierung der Erhebung von Kundenzufriedenheit beitragen. Ausgangspunkt der meisten Indizes ist ein Kausalmodell von Kundenzufriedenheit, welches i. d. R. Antezendenzen und Konsequenzen von Kundenzufriedenheit umfasst. Die einzelnen Komponenten sind zudem meist durch ein Messmodell mit mehreren Items operationalisiert. Die Messitems sind i. d. R. sehr allgemein gehalten, um Vergleiche über mehrere Branchen zu ermöglichen (vgl. BRUHN, 2008, S. 178).

Die Entwicklung von nationalen Zufriedenheitsindizes geht zurück auf das Zufriedenheitsbarometer Schwedens (FORNELL, 1992). Inzwischen gibt es eine Reihe weiterer (supra)nationaler Zufriedenheitsindizes, wie den europäischen ECSI (inzwischen European Performance Satisfaction Index). BRUHN (2008, S. 177f.) hat einen entsprechenden Überblick zusammengestellt. Zur Analyse werden meist Strukturgleichungsmodelle eingesetzt (CIAVOLINO und DAHLGAARD, 2007; EKLOF et al., 1999, z. B.).

Abbildung 3.2: Das ACSI-Modell



- adaptiert aus FORNELL et al. (1996a, S. 8) -

Neben den nationalen Indizes gibt es auch branchenspezifische Indizes von Kundenzufriedenheit. Ein Beispiel im Industriegüterbereich stellt das INDSTAT-Modell dar, welches über 50 Messvariablen zur Operationalisierung der Konstrukte verwendet (RUDOLPH, 1998).

Auch das im Bereich der Dienstleistungen verbreitete SERVQUAL-Modell (PARASURAMAN et al., 1988) funktioniert in ähnlicher Weise, jedoch werden die Bedeutungen der Attribute auch direkt abgefragt.

4. Entwicklung eines Mixture-Modells von Kundenzufriedenheit

In diesem Kapitel wird zunächst die Grundstruktur des hier entwickelten Mixture-Modells konkurrierender Kausalstrukturen dargestellt (Abschnitt 4.1). Dann wird näher auf die Bestandteile des Modells eingegangen. In Abschnitt 4.2 werden einige Grundlagen der bayesianischen Modellierung und der Schätzung mittels MCMC vorgestellt, die die Grundlage für die Schätzung des hier vorgestellte Modells bilden. Anschließend werden die bayesianische Modellierung des LISREL-Modells (Abschnitt 4.3), die Modellierung von ordinalen Daten (Abschnitt 4.4) mittels eines Cut-Point-Modells (Abschnitt 4.4.1) unter Berücksichtigung von Skalenverwendungsheterogenität (Abschnitt 4.4.2) erläutert. Beziehungen zu anderen Verfahren werden kurz in Abschnitt 4.6 dargelegt. Abschnitt 4.7 behandelt die Implementierung des hier eingeführten kausalen Mixture-Modells. Im letzten Abschnitt (4.8) dieses Kapitels wird die Funktionsfähigkeit des vorgestellten Modells mittels einer Simulationsstudie demonstriert.

4.1. Mixture von Kausalstrukturen

In den vorherigen Kapiteln wurden zentrale Herausforderungen für die Modellierung von Kundenzufriedenheit und insbesondere der Untersuchung von Kundenzufriedenheitstreibern dargelegt. Die in Abschnitt 2.4 gestellten Forschungsfragen sind der Ausgangspunkt für das in dieser Arbeit vorgestellte Modell:

1. Wie können inhaltliche Domänen bei der Identifikation alternativer Kausalstrukturen berücksichtigt werden?
2. Gibt es bei der Untersuchung von Kundenzufriedenheit eine Kausalstruktur, welche Elemente von Haloing und Forming aufweist?
3. Wie können Skalenverwendungsheterogenität, konkurrierende Kausalmodelle und Domänen gleichzeitig modelliert werden?

Diese Ziele werden, wie nachfolgend dargestellt, durch die Erweiterung des Modells von BÜSCHKEN et al. (2010) auf LISREL-Modelle (JÖRESKOG und SÖRBOM, 1996) erreicht. Das LISREL-Modell erlaubt zum einen die Implementierung von Messmodellen

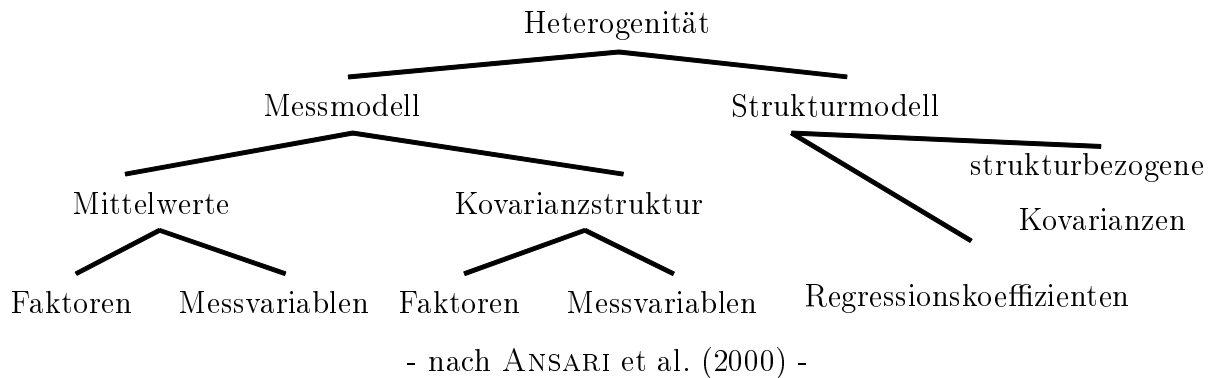
und zum anderen ist es sehr flexibel in Bezug auf modellierbare Kausalstrukturen. Frage (1) lässt sich folglich durch das Messmodell adressieren, während der flexible LISREL-Ansatz die Untersuchung von Frage (2) ermöglicht. Das hier vorgestellte Modell erlaubt es darüber hinaus, bei gleichzeitiger Kontrolle von Skalenverwendungsheterogenität beliebige LISREL-Strukturen in einem Mischmodell zu berücksichtigen (Frage 3). Dazu wird das Modell von ROSSI et al. (2001) verwendet. Somit können auch in anderen Disziplinen konkurrierende Theorien entsprechend berücksichtigt werden.

4.1.1. Heterogenität in Strukturgleichungsmodellen

In früheren Untersuchungen wurden unterschiedliche Kovarianzen in a priori vorgegebenen Gruppen betrachtet (JÖRESKOG, 1971). MUTHEN (1989) wies darauf hin, dass die Annahme homogener Populationen verletzt sein kann. Schließlich verfolgten JEDIDI et al. (1997) einen Finiten Mixture Ansatz, welcher die Identifikation von Gruppen (bzw. Segmenten) mit unterschiedlichen Parametern auf Basis der Beobachtungen erlaubt. Die Anzahl der Gruppen wird dabei üblicherweise vorgegeben.

Im Zuge der sogenannten „HB-Revolution“ (ALLENBY et al., 2004) wurden hierarchische Modelle zur Adressierung von Heterogenität im Marketing immer beliebter. Der Ansatz von ANSARI et al. (2000) greift diesen Gedanken bei SEM auf. Somit werden Random Coefficients auch bei Strukturgleichungsmodellen möglich. Die Systematisierung von unterschiedlichen Ansatzpunkten für die Adressierung von Heterogenität in SEMs nach ANSARI et al. (2000) ist in Abbildung 4.1 dargestellt. Relevant für diese Arbeit ist vor allem die Heterogenität im Strukturmodell, insbesondere bezüglich der Kovarianzstrukturen. Heterogenität im Messmodell wird durch das Skalenverwendungsheterogenitätsmodell berücksichtigt.

Abbildung 4.1: Heterogenität in Strukturgleichungsmodellen



Zur Adressierung von Heterogenität bei SEMs haben sich zwei unterschiedliche Techniken etabliert:

1. Mischmodelle (engl.: Finite Mixture Models) (z. B. JEDIDI et al., 1997; ARMINGER et al., 1999; ZHU und LEE, Mar), welche eine finite Zahl von Gruppen berücksichtigen. Diese Gruppen unterscheiden sich in Bezug auf die Ausprägungen der Parameter. Innerhalb der Gruppen wird weitestgehend Homogenität angenommen.
2. Heterogene Koeffizienten in hierarchischen Modellen (z. B. ANSARI et al., 2000), welche Verteilungsannahmen bezüglich individueller Parameter treffen.

Den Ansätzen ist jedoch gemein, dass Homogenität bezüglich der kausalen Struktur angenommen wird. Heterogenität wird ausschließlich hinsichtlich der Parameterwerte modelliert, nicht bezüglich der Parametrisierung selbst. In dieser Arbeit sollen hingegen explizit alternative Kausalmodelle betrachtet werden.

4.1.2. Alternative Kausalmodelle

Alternative bzw. konkurrierende Theorien, welche alternative Kausalstrukturen implizieren, äußern sich in unterschiedlichen Pfaddiagrammen und somit auch in unterschiedlichen Strukturgleichungen.

So können z. B. zwei Konstrukte A und B in einem Modell 1 durch einen Pfeil verbunden sein, während in einem Modell 2 jedoch angenommen wird, dass A und B unabhängig

voneinander sind. Dadurch unterscheiden sich die Datengenerierungsprozesse von Modell 1 und 2 fundamental voneinander.

In der Ökonomie kann das beispielsweise durch unterschiedliche Produktionsfunktionen verschiedener Unternehmen ausgelöst werden. In dieser Arbeit werden jedoch konkurrierende Kausalstrukturen bezüglich des Antwortverhaltens bei Kundenzufriedenheitsbefragungen modelliert. Das LISREL-Modell eignet sich dabei als Grundlage, um die alternative Kausalstrukturen zu beschreiben. In diesem Kontext bedeutet „alternativ“ Folgendes:

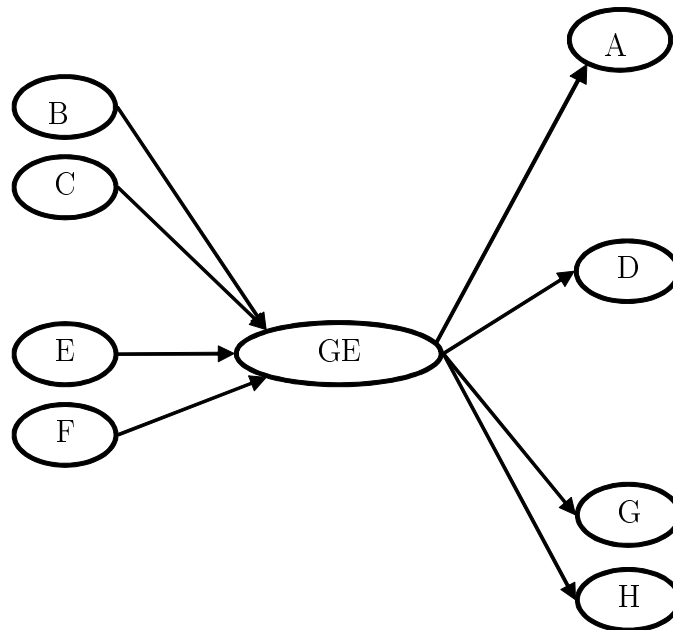
1. Konkurrierende Theorien sind die Grundlage für unterschiedliche Strukturmodelle mit unterschiedlichen Strukturgleichungen.
2. Diese unterschiedlichen Strukturmodelle resultieren in unterschiedlichen konditionalen partiellen Korrelationen. Dies wird dadurch erreicht, dass unterschiedliche Aussagen über Unabhängigkeiten zwischen Konstrukten getroffen werden.

Sind diese Bedingungen nicht erfüllt, sind die Modelle möglicherweise äquivalent (wie in Abschnitt 4.3.4 beschrieben) und können auf Grundlage der Daten nicht voneinander getrennt werden. Ein finites Mixture-Modell kann im Falle äquivalenter Komponenten nur Heterogenität in Parametern adressieren. Allerdings sind dann zusätzliche Restriktionen zur Modellidentifikation notwendig. In dieser Arbeit wird ausschließlich Heterogenität in Kausalstrukturen betrachtet.

Kausalstrukturen im Antwortverhalten Der *formative* Antwortstil folgt der Annahme eines typischen Zufriedenheitstreibermodells. Es werden unabhängige Bewertungen von Teilattributen durchgeführt und anschließend eine sich daraus ergebende Gesamtbewertung abgegeben. Die abgegebenen Bewertungen bezüglich der Teilzufriedenheiten entsprechen tatsächlich einer Bewertung, die sich ausschließlich auf diese Teilzufriedenheiten bezieht. Somit kann der Zusammenhang von Teilzufriedenheiten mit der Gesamtzufriedenheit analysiert werden.

Beim *haloistischen* Antwortstil wird eine völlig andere Kausalstruktur angenommen als beim Treibermodell. Die Bewertungen der Attribute erfolgt dabei nicht unabhängig voneinander. Während des Erinnerungsprozesses wird lediglich eine Gesamtbewertung aus

Abbildung 4.2: MIMIC-Struktur Bsp. 2



- eigene Darstellung -

dem Gedächtnis abgerufen, sodass alle attributsspezifischen Fragen nicht mehr differenziert beantwortet werden können. Um jedoch Antworten in konsistenter Weise abzugeben, spiegeln diese Teilzufriedenheiten nun die Erinnerung an den Gesamteindruck wider.

Der *MIMIC*-Antwortstil verknüpft beide Ideen: Ein Teil der Attribute wird unabhängig evaluiert und zur Bildung eines Gesamturteils herangezogen. Alle übrigen Fragen können anschließend jedoch nicht weiter differenziert werden.

4.1.3. MIMIC-Strukturen

Auf die bisherigen Anwendungen von MIMIC-Modellen wurde bereits in Abschnitt 2.4.3 eingegangen. Die Anwendung als Kausalmodell von Kundenzufriedenheitsbewertungen lässt sich an dem beispielhaften MIMIC-Pfadmodell in Abbildung 4.2 erläutern:

Modellierung von MIMIC Auf der linken Seite des Pfadmodells befinden sich die exogenen Konstrukte B,C,E,F. Die Evaluation dieser vier Teilzufriedenheitskonstrukte treiben den latenten Gesamteindruck (GE), welcher nicht durch ein eigenes Messmodell opera-

tionalisiert ist. Die Konstrukte auf der rechten Seite des Gesamteindrucks werden alleine durch diesen beeinflusst: die *gemessene* Gesamtzufriedenheit (A) sowie weitere latente Teilzufriedenheitskonstrukte D, G, H, die nicht differenziert beurteilt werden. Damit das Modell identifiziert ist, muss die Strukturfehlervarianz von GE auf einen Wert (in dieser Arbeit auf 1) festgelegt werden, da GE nicht durch ein Messmodell operationalisiert ist. Zudem wird der Pfad von $GE \rightarrow A$ auf 1 festgelegt. Somit ist das Modell identifiziert.

Die Trennung der Konstrukte Gesamteindruck (GE) und Gesamtzufriedenheit (A) lässt den Fall zu, dass sich der Gesamteindruck nicht vollständig in der Gesamtzufriedenheit widerspiegelt. Der Strukturfehler von A zeigt an, inwieweit Gesamteindruck und Gesamtzufriedenheit übereinstimmen. Die MIMIC-Struktur ist genau wie die anderen verwendeten Komponenten als LISREL-Modell darstellbar.

Diese Art der MIMIC-Modellierung unterscheidet sich von bisherigen Anwendungen des MIMIC-Modells. Sowohl die „Causes“ als auch die „Indicators“ bestehen im hier vorgestellten Modell aus latenten Konstrukten. In bisherigen Untersuchungen wird nur ein latentes Konstrukt angenommen, welches durch gemessene (nicht latente) „Causes“ bestimmt wird und sich in den gemessenen „Indicators“ reflektiert. Die zentrale Idee bleibt jedoch gleich: Der Zusammenhang zwischen „Indicators“ und „Causes“ lässt sich durch *eine* zentrale latente Variable darstellen.

Diese Idee entspricht genau dem Konzept des partiellen Haloings: Einige (aber nicht alle) Attribute werden zur Bestimmung des Gesamteindrucks herangezogen. Diese Attribute sind gewissermaßen ursächlich („Causes“) für das Konstrukt Gesamteindruck. Alle übrigen Attribute werden nicht differenziert und reflektieren lediglich diesen Gesamteindruck.

Motivation für die MIMIC-Modellierung Die Einteilung als „Cause“ oder „Indicator“ richtet sich mutmaßlich danach, wie leicht eine differenzierte Bewertung für den Befragten ist. Während einer Befragung müssen die Befragten zunächst Informationen aus ihrem Gedächtnis abrufen. Die Bewertung basiert somit auf Informationen aus dem Gedächtnis (LYNCH und SRULL, 1982; ALBA et al., 1991). Die Erinnerungsfähigkeit und somit auch die Urteilsfähigkeit kann sich abhängig vom befragten Attribut unterscheiden.

Relevant für die Urteilsfähigkeit ist vor allem die Verfügbarkeit von Informationen (FELDMAN und LYNCH, 1988). Je einfacher Informationen abgerufen werden können,

desto eher kann differenziert beurteilt werden. „Objektive“ Kriterien sind möglicherweise einfacher aus dem Gedächtnis abrufbar.

So vergessen Menschen abstrakte Wörter schneller als konkrete (WALKER und HULME, 1999). Zufriedenheitsbefragungen umfassen häufig abstrakte (z. B. Ruf, Atmosphäre) und konkrete (z. B. Erreichbarkeit, Öffnungszeiten) Attribute. Möglicherweise gibt es entsprechende Unterschiede zwischen Teilzufriedenheiten.

Konkrete Wörter erzeugen zudem weniger Assoziationen als abstrakte (ALTARRIBA et al., 1999). Die differenzierte Beantwortung von Fragen, welche weniger Assoziationen hervorrufen, erscheint einfacher. Hängen Fragen schon verbal stark zusammen, so erscheint eine differenzierte Beantwortung unwahrscheinlicher.

Auch der „Recency-Effekt“ (WYER und SRULL, 1989, S.289f.) kann eine Rolle spielen. Dabei geht man davon aus, dass länger zurückliegende Ereignisse oder Informationen weniger präsent sind. Sind bei einer Befragung Informationen über die Erfahrungszeiträume verfügbar, so könnten „entferntere“ Aspekte im haloistischen Teil eines MIMIC-Modells aufgenommen werden.

Im Langzeitgedächtnis gibt es den deklarativen (expliziten) und den nicht-deklarativen (impliziten) Bestandteil (SQUIRE, 1992). Die unterschiedliche Neuroanatomie konnte inzwischen durch funktionelle Magnetresonanztomographie nachgewiesen werden (SCHOTT et al., 2005). Zu bewertende Attribute können sich dahingehend unterscheiden, inwieweit auf das explizite Gedächtnis zurückgegriffen wird. Bei der Bewertung von Attributen wie Geschmack oder Atmosphäre wird stärker auf das implizite Gedächtnis zurückgegriffen.

Dem Verfasser sind keine Publikationen bekannt, welche partielles Haloing mit einem MIMIC-Modell modellieren. Unterschiede in der „Urteilsfähigkeit“ aufgrund unterschiedlicher Erinnerungsprozesse können mit dem hier entwickelten Modell explizit untersucht werden.

4.1.4. Modell-Likelihood

Dieses Modell berücksichtigt K unterschiedliche LISREL-Komponenten, i -spezifische (mit $i \in \{1, \dots, n\}$) Skalenverwendungsparameter und gemeinsame Cut-Points zur Überführung der ordinalen in kontinuierliche Beobachtungen. Die gemeinsame Verteilung einer

LISREL-Komponente unter Berücksichtigung von Skalenverwendungsheterogenität lässt sich wie folgt faktorisieren:

$$\begin{aligned}
& p(Y, c, Z, \{\tau_i\}, \{\sigma_i\}, \mu, \Lambda, \Psi_\varepsilon, \Pi, \Gamma, \Psi_\delta, \Phi) \\
&= p(Y|c, Z) p(Z, \mu_Z, \{\tau_i\}, \{\sigma_i\}, \Lambda, \Psi_\varepsilon, \Pi, \Gamma, \Psi_\delta, \Phi, \Omega) p(\Pi, \Gamma, \Psi_\delta, \Phi) p(c) \\
&= p(Y|c, Z) p(Z|\mu_Z, \{\tau_i\}, \{\sigma_i\}, \Lambda, \Psi_\varepsilon, \Omega) p(\Omega|\Pi, \Gamma, \Psi_\delta, \Phi, \mu_Z) p(\Pi, \Gamma, \Psi_\delta, \Phi) p(c) \\
&= \underbrace{p(Y|c, Z)}_{\text{Cut - Point}} \underbrace{p(Z|Z^*, \{\tau_i\}, \{\sigma_i\})}_{\text{Skalenverwendung}} \underbrace{p(Z^*|\mu_Z, \Lambda, \Psi_\varepsilon, \Omega) p(\Omega|\Pi, \Gamma, \Psi_\delta, \Phi, \mu_Z) p(\Pi, \Gamma, \Psi_\delta, \Phi)}_{\text{LISREL}} p(c)
\end{aligned} \tag{4.1}$$

Das finite Mixture-Modell bezieht sich nur auf die LISREL-Parameter, welche darin komponentenspezifisch werden. Auf die LISREL-Modellierung wird in Abschnitt 4.3.1 näher eingegangen. Das Cut-Point-Modell und Skalenverwendungsheterogenität gelten hingegen komponentenübergreifend. Auf diese Bestandteile des Modells wird in den Abschnitten 4.4 und 4.4.2 näher eingegangen. In Abhängigkeit der Gewichtung der Komponenten $\pi = (\pi_1, \dots, \pi_K)$ und der übrigen Modellparameter $\{\theta_k\} = \{\Pi_k, \Gamma_k, \Psi_{\delta k}, \Phi_k, \mu_{Zk}\}$, $\mathbf{c}, \{\tau_i\}, \{\sigma_i\}$ ergibt sich für das Mixture-Modell mit K Komponenten folgende individuelle Likelihoodfunktion:

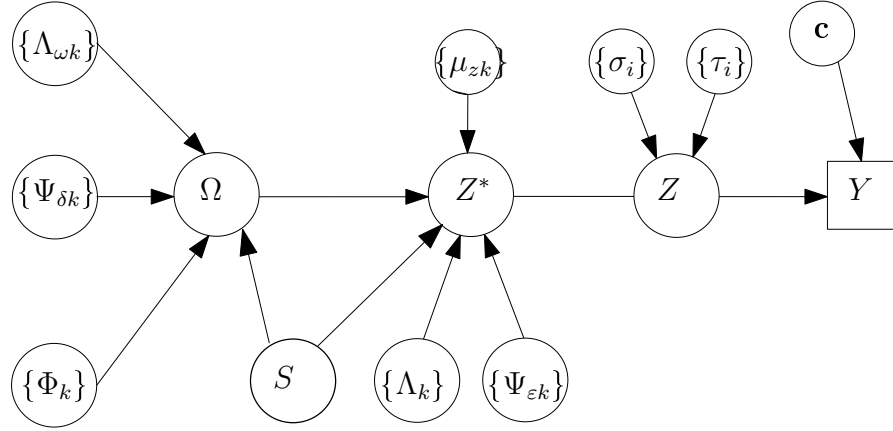
$$p(y_i | \{\theta_k\}, \pi, \mathbf{c}, \{\tau_i\}, \{\sigma_i\}) = \sum_{k=1}^K \pi_k p(y_i | \theta_k, \mathbf{c}, \{\tau_i\}, \{\sigma_i\}) \tag{4.2}$$

Finite Mixture-Modelle lassen sich als hierarchische latente-Variablen-Modelle verstehen, wobei die Beobachtungen Y von einer diskreten, unbeobachteten Indikatorvariable S abhängen (FRÜWIRTH-SCHNATTER, 2006, S. 12). Die gemeinsame Verteilung von beobachteten Daten Y und Indikatorvariable S kann allgemein wie folgt faktorisiert werden:

$$p(Y, S) = p(Y|S) p(S) \tag{4.3}$$

Ist S bekannt, so kann der entsprechende Datengenerierungsprozess der Y „ausgewählt“ werden. $p(S)$ entspricht wiederum der Gewichtung der unterschiedlichen Datengenerierungsprozesse. Bei der Schätzung mittels MCMC wird folglich die Verteilung von S herangezogen, um Schlüsse über die Anteile der Komponenten zu ziehen. In Abschnitt 4.5 wird noch einmal auf die Mixture-Modellierung eingegangen. Analog lässt sich das in

Abbildung 4.3: DAG



- eigene Darstellung -

dieser Arbeit entwickelte Mixture-Modell folgendermaßen faktorisieren:

$$\begin{aligned}
 & p(Y, S, \mathbf{c}, Z, \{\tau_i\}, \{\sigma_i\}, \{\theta_k\}) \\
 & = p(Y|c, Z) \times p(Z|Z^*, \{\tau_i\}, \{\sigma_i\}) \\
 & \times p(Z^*|\{\theta_k\}, S) \times p(\Omega|\{\theta_k\}, S) \times p(\{\theta_k\}) \\
 & \times p(S)
 \end{aligned} \tag{4.4}$$

Die Schätzung der komponentenspezifischen LISREL-Parameter, der Cut-Points und Skalenverwendungsparameter sowie der Allokation lassen sich folglich in unabhängige Schätzprobleme zerlegen (vgl. FRÜWIRTH-SCHNATTER, 2006, S. 30). Der zugehörige DAG (Directed Acyclic Graph) ist in Abbildung 4.3 dargestellt.

Multiple Bewertungsobjekte Befragte können mehr als ein Bewertungsobjekt entlang von p Fragen beurteilen. Das Skalenheterogenitätsmodell von ROSSI et al. (2001) berücksichtigt hingegen nur einen Antwortvektor (also eine Beurteilung entlang von p Fragen) pro befragter Person. Die Likelihood des Modells für einen Befragten mit nur einem Antwortvektor ist in Gleichung 4.5 dargestellt. In Abhängigkeit der Zuordnung S_i zu einer Mixture-Komponente, den zugehörigen LISREL-Parametern θ_k , den personenübergreifenden Cut-Points \mathbf{c} sowie den individuellen Skalennutzungsparametern τ_i, σ_i ergibt sich die Likelihood als multivariates Integral.

$$\begin{aligned}
p(y_i | \{\theta_k\}, S_i, \mathbf{c}, \tau_i, \sigma_i) &= \Phi \left((\sigma_i^2 \Sigma_k)^{-0.5} [c_{y_{ik+1}} - m_{ik}] \right) - \Phi \left((\sigma_i^2 \Sigma_k)^{-0.5} [c_{y_{ik}} - m_{ik}] \right) \\
\text{mit } m_{ik} &= (\mu_k + \tau_i) \sigma_i \\
\text{und } \Sigma_k &= \Lambda_k \begin{bmatrix} \Pi_{0k}^{-1} (\Gamma_k \Phi_k \Gamma_k^T + \Psi_{\delta k}) \Pi_{0k}^{-T} & \Pi_{0k}^{-1} \Gamma_k \Phi_k \\ \Phi_k \Gamma_k^T \Pi_{0k}^{-T} & \Phi_k \end{bmatrix} \Lambda_k + \Psi_{\varepsilon k} \\
\text{und } \Pi_{0k} &= I_{q_{k1}} - \Pi_k
\end{aligned} \tag{4.5}$$

Bei nur einem Antwortvektor pro Person lässt sich die Likelihood des Modells gegeben aller Daten als Produkt über alle $i \in 1n$ Beobachtungen ausdrücken:

$$p(Y | \{\theta_k\}, S, \mathbf{c}, \{\tau_i\}, \{\sigma_i\}) = \prod_{i=1}^n p(y_i | \{\theta_k\}, S_i, \mathbf{c}, \tau_i, \sigma_i) \tag{4.6}$$

Stammen nun aber mehrere Bewertungen von je einer Person, so muss dies berücksichtigt werden. Die Skalenverwendung wird als personenspezifische Eigenschaft aufgefasst, während die Zuordnung zur geeigneten LISREL-Komponente auf Basis des jeweiligen Beobachtungsvektor vorgenommen wird. n sei nun die Zahl der Befragten. Der Vektor A gibt an, welche Objekte von welcher Person bewertet worden sind. Es sei $A_0 = \{\}$. Bewertet Person 1 die Objekte 1 und 2, so gilt $A_1 = \{1, 2\}$. Die Zahl der Beobachtungen ist folglich $\sum_{\alpha=1}^n |A_\alpha|$. S ist dementsprechend auch ein Vektor der Länge $\sum_{\alpha=1}^n |A_\alpha|$, da die Komponentenzuordnung nicht an die Person gebunden ist.¹⁵ Zur Berechnung der Likelihood müssen nun zwei Produkte gebildet werden: eines bezieht sich auf die Personen, das andere auf die Antworten der jeweiligen Person. Für die Likelihood ergibt sich in diesem Fall Gleichung 4.7.

$$p(Y | \{\theta_k\}, S, \mathbf{c}, \{\tau_i\}, \{\sigma_i\}) = \prod_{i=1}^n \prod_{j=1+\sum_{\alpha=0}^{i-1} |A_\alpha|}^{\sum_{\alpha=0}^i |A_\alpha|} p(y_j | \{\theta_k\}, S_j, \mathbf{c}, \tau_i, \sigma_i) \tag{4.7}$$

4.2. Bayesianische Inferenz und MCMC

Zum besseren Verständnis des in dieser Arbeit entwickelten bayesianischen Mixture-Modells wird zunächst die bayesianische Statistik kurz vorgestellt. Neben der Möglichkeit,

¹⁵Die Spezifität der Komponentenzuordnung wird in Abschnitt 5.4 empirisch untersucht.

Prior-Informationen im Modell aufzunehmen, ist es vor allem die einfache Schätzung hochkomplexer Modelle mittels Markov Chain Monte Carlo-Methoden, die zur Anwendung der bayesianischen Statistik motiviert. Die Einführung von MCMC in den 1980er Jahren wurde begünstigt durch die Entwicklung immer leistungsfähigerer Computer. Die Nutzbarkeit unterschiedlicher Sampling-Algorithmen für bayesianische Analyse (GELFAND und SMITH, 1990) trug ebenfalls zur Popularität dieser Methoden bei. Die Berechnung einiger bayesianischer Modelle wurde überhaupt erst durch solche Sampling-Algorithmen möglich (FIENBERG, 2006, S. 27). Alternativ können Näherungsverfahren eingesetzt werden, die jedoch numerisch sehr anspruchsvoll sind und somit nur von wenigen Wissenschaftlern eingesetzt werden konnten (LYNCH, 2007, S. 79). Die Aufteilung hochkomplexer Modelle in iterativen Verfahren erleichtert die Schätzung. Insbesondere der Gibbs-Sampler (GEMAN und GEMAN, 1984) erlaubt eine verhältnismäßig schnelle Approximation der Posterior-Verteilungen. Zudem ermöglicht die bayesianische Statistik eine intuitivere Interpretierbarkeit, insbesondere bei den Konfidenz- bzw. Vertrauensintervallen (CONGDON, 2003, S. 1). Dies liegt insbesondere in einem anderen Verständnis der Parameter begründet: Während in der klassischen Statistik angenommen wird, dass Daten zufällig, Parameter aber fix sind, geht man in der bayesianischen Statistik davon aus, dass auch die Parameter aus stochastischen Verteilungen stammen (CONGDON, 2003, S. 2).

„Bayesianische Statistik“ als Begriff ist verhältnismäßig jung und erst seit den 1970er Jahren gebräuchlich (FIENBERG, 2006). Die Anfänge um den grundlegenden Satz von Bayes gehen möglicherweise zurück auf BAYES et al. (1763), dies ist jedoch umstritten (STIGLER, 1983). Heute spricht man von einem „Bayesianer“ (engl.: „Bayesian“) als „person who thinks it makes sense to treat observables as random variables and to assign probability distributions to them.“ (FIENBERG, 2006, S. 1)

Prior-Information Während bei klassischen Verfahren, wie z. B. Maximum Likelihood, alleine die Daten zur Inferenz herangezogen werden, spielen bei bayesianischen Verfahren A-Priori-Informationen oder kurz Prior-Informationen auch eine Rolle. CONGDON (2003, S. 2f.) fasst dies prägnant zusammen: Die Likelihood der observierten Daten y gegeben Parameter θ sei $f(y|\theta)$ bzw. $L(\theta|y)$. Diese wird verwendet, um ein „Update“ der A-Priori Erwartung (dargestellt durch $\pi(\theta)$) durchzuführen. Daraus ergibt sich die A-Posteriori-

Verteilung $\pi(\theta|y)$, welche die Informationen aus Prior und Likelihood zusammenführt. Es gilt:

$$f(y, \theta) = f(y|\theta) \pi(\theta) = \pi(\theta|y) m(y) \quad (4.8)$$

Somit lässt sich, entsprechend dem Satz von Bayes, die Posteriori-Verteilung wie folgt darstellen:

$$\pi(\theta|y) = f(y|\theta) \pi(\theta) / m(y) \quad (4.9)$$

wobei $m(y)$ die marginale Likelihood der Daten darstellt:

$$m(y) = \int f(y|\theta) \pi(\theta) d\theta \quad (4.10)$$

Die marginale Likelihood wird zwar für Modellselektion benötigt, jedoch kann sie in vielen anderen Fällen vernachlässigt werden, sodass sich die Approximation der Posteriori-Verteilung erheblich vereinfacht.

Zur Rolle von A-Priori-Informationen Zunächst erscheint es sehr vorteilhaft, wenn A-Priori-Wissen für Analysen verwendet werden kann. Vorherige Durchführungen von Analysen oder Expertenmeinungen können so in kohärenter Weise verarbeitet werden. Jedoch ist auch Vorsicht geboten, insbesondere wenn kein A-Priori-Wissen vorhanden ist. In diesem Fall scheint sich ein nicht-informativer-Prior anzubieten, beispielsweise eine Gleichverteilung zwischen $-\infty$ und $+\infty$ oder Jeffreys Prior:

$$\pi(\theta) \propto \det \{I(\theta)\}^{0.5} \quad (4.11)$$

wobei $I(\theta)$ der erwarteten Informationsmatrix entspricht (vgl. CONGDON, 2003, S. 3). Uninformative Prior-Verteilungen bergen jedoch die Gefahr, dass Identifikationsprobleme auftreten (GELFAND und SAHU, 1999), sodass zumindest leicht informative Prior-Verteilungen vorgezogen werden.¹⁶ Nicht immer ist jedoch sofort ersichtlich, ob eine Prior-Verteilung wirklich schwach-informativ ist. Die Definition der Prior-Verteilungen erfordert folglich sehr viel Sorgfalt. Auch gibt es Gründe, informative Prior-Verteilungen vorzuziehen - insbesondere bei hierarchischen Modellen (z. B. LENK und ORME, 2009). Das

¹⁶Auf Identifikation bei Bayesianischen Modellen wird im nächsten Abschnitt noch genauer eingegangen.

hier entwickelte Modell hat nur einen hierarchischen Modellbestandteil - das Skalenverwendungsheterogenitätsmodell. Bei wiederholten Untersuchungen sind hier auch stärker informative Prior-Verteilungen denkbar.

Ebenso ist bei Reparametrisierungen mit zusätzlichen latenten Variablen, die zur Vereinfachung der Berechnung eingeführt werden, Vorsicht geboten. Die Prior-Verteilungen solcher latenter Größen können Rückwirkungen auf das gesamte Modell haben (GELMAN, 2004).

Konjugate Verteilungen Häufig werden in bayesianischen Analysen konjugate Prior-Verteilungen angenommen. In diesem Fall verbinden sich die Kernel der Prior-Verteilung mit dem Kernel der Likelihood, sodass die Posteriori-Verteilung zur gleichen Familie wie die Prior-Verteilung gehört. Dies vereinfacht die Berechnung deutlich, sodass in einem Gibbs-Sampler nur Zufallszahlen aus bekannten Verteilungen generiert werden müssen. Wird keine konjugate Prior-Verteilung verwendet, so ist dies nicht möglich. Erst in den letzten Jahren ist man aufgrund der gestiegenen Rechnerkapazitäten nicht mehr dazu gezwungen, konjugate Prior-Verteilungen annehmen zu müssen. Zuvor waren die numerischen Beschränkungen jedoch ganz real (DRÈZE, 1972, S. 8). Inzwischen stehen auch Verfahren wie Slice-Sampling (NEAL, 2003) oder das sehr allgemein anwendbare Adaptive-Rejection-Sampling (GILKS, 1992) zur Verfügung, mit denen Zufallszahlen auch aus Nicht-Standard-Verteilungen generiert werden können.

4.2.1. Identifikation

„In passing it might be noted that unidentifiability causes no real difficulty in the Bayesian approach“ (LINDLEY, 1972, S. 46)

Das Zitat von Lindley mag dazu verleiten, sich nicht ausführlich mit dem Thema Identifikation auseinanderzusetzen. In diversen Grundlagenbüchern zur bayesianischen Statistik wird Modellidentifikation auch nicht näher problematisiert (z. B. ROBERT und CASELLA, 2004; KOCH, 2007; BOLSTAD, 2007). Dies liegt möglicherweise auch darin begründet, dass rein technisch gesehen Identifikation kein Problem bei bayesianischen Analysen darstellt (ROSSI et al., 2006, S. 19).

Ein Modell ist nicht identifiziert, wenn für zwei unterschiedliche Parameter θ und θ^* $p(y|\theta) = p(y|\theta^*)$ für fast alle $y \in Y$ gilt (z. B. FRÜWIRTH-SCHNATTER, 2006, S. 15). Insbesondere bei Strukturgleichungsmodellen können Identifikationsprobleme auftreten. Bezogen auf Strukturgleichungsmodelle definieren STERN und JEON (2004, S. 233) ein Modell als identifiziert, wenn es möglich ist, alle Parameter aus der Varianz-Kovarianz-Matrix der Beobachtungen zu identifizieren.

Eine einfache „bayesianische“ Antwort auf Identifikationsprobleme liegt in dem Verweis auf Prior-Verteilungen. So schlagen ROSSI et al. (2006, S. 19-20) vor, Identifikationsprobleme mit informativen Prior-Verteilungen zu lösen:

„Rather than imposing some sort of constraint on the parameter space, the Bayesian can deal with lack of identification through an informative prior. With the proper informative prior, the posterior may not have any region of constancy.“

Dem zugrunde liegt das Verständnis von ZELLNER (1971, S. 254). Danach ist die Prior-Verteilung in das Verständnis von Identifikation mit einzubeziehen.

Verfolgt man jedoch die Diskussion um Identifikation in bayesianischen Modellen etwas weiter zurück, so findet man eine kontroverse Diskussion vor. Umfassende Überblicke über die Entwicklung des Identifikationsbegriffs in der bayesianischen Statistik finden sich bei ALDRICH (2002) und POIRIER (1998). Die Diskussion drehte sich vor allem um die Frage, ob die bayesianische Theorie eine eigene - und sich damit von der klassischen Definition abweichende - Definition von Identifikation erforderlich macht. HSIAO (1983, S. 272) nennt dies noch 1983 eine ungeklärte Frage.

Entwicklung des Identifikationsbegriffs Das Konzept von „Identifikation“ kam in den 1940er Jahren auf (ALDRICH, 2002, S. 79). Frühe Abhandlungen hat KOOPMANS (1949, 1953) verfasst. So definieren KOOPMANS et al. (1953, S. 36) Identifikation wie folgt:

„We shall call a structure S permitted by the model (uniquely) identifiable within that model if there is no other equivalent structure S^\oplus contained in the model.“

FISHER (1966) widmet schließlich ein ganzes Buch der Auseinandersetzung mit Identifikation. ROTHENBERG (1971) stellt Theoreme für die Identifikation von parametrischen Modellen auf. Seine zentrale Idee besteht darin, dass die Fisher-Informationsmatrix vollen Rang besitzen muss, wenn das Modell identifiziert ist (ROTHENBERG, 1971, S. 577). Ein Identifikationsproblem sieht er als Mangel an Informationen, wodurch keine Unterscheidung alternativer Parameterstrukturen möglich ist.

Identifikation und bayesianische Statistik Mit dem Aufkommen der bayesianischen Statistik musste diskutiert werden, inwieweit Konzepte wie Likelihood und Identifikation aus der klassischen Statistik übernommen werden können. Die anschließende Diskussion wurde kontrovers geführt.

KADANE (1977, S. 175) vertritt die Auffassung, dass Identifikation eine Eigenschaft der Likelihood sei und dass es folglich keinen Unterschied zwischen bayesianischer und klassischer Theorie geben könne. ZELLNER (1971, S. 254) hingegen möchte ein breiteres Verständnis von Identifikation etablieren, welches die Prior-Verteilungen berücksichtigt. Beim Einsatz diffuser Prior-Verteilung gäbe es dementsprechend keinen Unterschied zwischen bayesianischer und klassischer Analyse. Jedoch könnten klassisch nicht identifizierte Modelle in bayesianischer Analyse identifiziert sein, wenn Prior-Verteilungen informativ sind. Auch MORALES (1971, S. 20) sieht ein Modell als identifiziert an, wenn die Posterior-Verteilung keine „flachen Stellen“ aufweist. Diese Sicht wird jedoch von ROTHENBERG (1973, S. 158) kritisiert. Bei unimodalen glockenförmigen Verteilungen könne so fast immer Identifikation erreicht werden, obwohl die Daten dafür völlig irrelevant seien. Somit sei es besser, in diesem Fall von „Schätzbarkeit“ (Estimability) denn von „Identifikation“ zu sprechen.

DRÈZE (1972, S. 8-9) wiederum findet besonders die Möglichkeit, stochastische statt starre Restriktionen zum Erreichen von Identifikation heranzuziehen, sehr nützlich. Starre Restriktionen könne man zudem als Spezialfall ansehen - die Masse der Prior-Verteilung ist an einem Punkt konzentriert.¹⁷ Ähnlich beschreibt es auch ZELLNER (1971, S. 254). DRÈ-

¹⁷Dem Weg der „starren“ Restriktion wird auch im empirischen Teil dieser Arbeit gefolgt. Die Ergebnisse der untersuchten Modelle unterscheiden sich kaum im Vergleich zu einer stark informativen Prior-Verteilung.

ZE (1977, S. 167) stellt schließlich klar, dass er die Ansicht von KADANE (1977, S. 175), dass Identifikation eine Eigenschaft der Likelihood sei, teile. Das Wort „Identifikation“ hält im Zusammenhang mit der Definition der Prior-Verteilungen auch für missverständlich. Vielmehr solle nach Festlegung aller Restriktionen die Likelihood überprüft werden.

POIRIER (1998, S. 483) beantwortet schließlich die Frage von HSIAO (1983, S. 272). Demnach ist das Konzept von „Identifikation“ identisch, egal ob bayesianische oder klassische Analysen durchgeführt werden.

Ähnlich wie ROTHENBERG (1973, S. 158), der kritisch anmerkt, dass bei informativen Prior-Verteilungen nahezu jedes Modell „schätzbar“ sei, schreibt auch POIRIER (1998, S. 483), dass alle nicht-identifizierte Modelle mit einer entsprechenden Prior-Spezifikation analysiert werden könnten. Jedoch gäbe es diese Möglichkeit nicht „geschenkt“:

„There is, however, no Bayesian free lunch. The „price“ is that there exist quantities about which the data are uninformative, i.e., their marginal prior and posterior distributions are identical.“

ALDRICH (2002, S. 96) sieht inzwischen Konsens darüber, dass sich der Identifikationsbegriff nicht zwischen klassischer und bayesianischer Statistik unterscheidet. Vielmehr drehe sich die Diskussion um den Informationsgehalt der beobachteten Daten.

Die Beiträge von ROTHENBERG (1971, 1973) und anderen bleiben damit auch in der bayesianischen Statistik relevant. Jedoch lassen sich Identifikationsprobleme (bzw. „Schätzbarkeitsprobleme“) leicht lösen, wenn es eine Grundlage für informative Prior-Verteilungen gibt - z.B. bei der wiederholten Durchführung von Experimenten oder Befragungen.

Bei bayesianischen Analysen muss jedoch nicht immer Identifikation sichergestellt werden. Rein technisch funktionieren MCMC-Methoden auch bei nicht-identifizierten Modellen. Jedoch wird damit das Problem nur verschoben. Die Beschreibung der Posteriori-Verteilungen ist erst nach sogenannten Post-Processing möglich. Nach der Generierung der Züge werden diese „nachbehandelt“, sodass eine Beschreibung der Parameterverteilungen möglich ist. Diese Variante wird insbesondere bei Probit-Modellen häufig eingesetzt (z.B. McCULLOCH und ROSSI, 1994; McCULLOCH et al., 2000).

4.2.2. MCMC Methoden

Die gemeinsame Posteriori-Verteilung von komplexen und hochparametrischen und somit hochdimensionalen Modellen stellt eine große Herausforderung für klassische Schätzverfahren dar. Eine zuverlässige Maximum-Likelihood-Schätzung durch sogenannte Hill-climbing-Algorithmen ist schwierig, da diese oft anfällig für Multimodalität sind und somit viele verschiedene Startwerte ausprobiert werden müssen, um Konvergenz sicherzustellen. Bei bayesianischen Analysen kann hingegen auf Markov Chain Monte Carlo-Methoden zurückgegriffen werden, die auch komplexe Modelle zuverlässig schätzen können. Die Grundlage von Markov Chain Monte Carlo-Methoden ist zweierlei: Zum einen handelt es sich um ein Monte Carlo-Verfahren, bei dem versucht wird, aus der gemeinsamen Posteriori-Verteilung Zufallszahlen zu generieren, die schließlich zur Beschreibung der Verteilung herangezogen werden können. Zum anderen basiert MCMC auf der Konstruktion einer Markov-Kette, deren stationäre Verteilung der gesuchten Verteilung entspricht (vgl. ROSSI et al., 2006, S. 50). Die Idee dahinter ist, dass sich durch Konditionierung in einer Markov-Kette leichter Zufallszahlen generieren lassen, da die gemeinsame Posteriori-Verteilung keiner bekannten Form entspricht (vgl. HASTINGS, 1970, S. 97). Die Markov-Eigenschaft bezieht sich darauf, dass beim Sampling nur auf das jeweils letzte Sampling-Ergebnis konditioniert wird.¹⁸

Die Schwierigkeit, komplexe gemeinsame Verteilungen zu analysieren, lässt sich bereits an einem einfachen Beispiel verdeutlichen: Bei einem linearen Regressionsmodell lässt sich die gemeinsame Verteilung von Fehlervarianz und Regressionskoeffizienten nicht einfach erfassen. Die Varianzen stammen aus einer inversen Gammaverteilung, die Regressionskoeffizienten aus einer Normalverteilung. Die Kombination aus beiden entspricht jedoch keiner Standardverteilung.¹⁹ Bei hierarchischen Modellen ist die gemeinsame Posteriori-

¹⁸MCMC stellt nur eine Anwendung des Konzepts von Markov-Ketten dar. Eine grundsätzliche Einführung in Markov-Ketten und Markov-Prozesse findet sich bei MEYN und TWEEDIE (2009). Im Rahmen der Anwendung in MCMC geht man von der stationären Verteilung (in diesem Fall der Posteriori-Verteilung) aus und konstruiert eine Markov-Kette, welche zu dieser stationären Verteilung führt.

¹⁹Bei entsprechender Gestaltung der Prior-Verteilung lässt sich zwar ein iid-Sampler herleiten. An dieser Stelle geht es jedoch zunächst um die Komplexität gemeinsamer Verteilungen aller Parameter.

Verteilung der Parameter in aller Regel nicht der Ausgangspunkt der Modellspezifikation, sondern die Kombination von Verteilungen (vgl. CONGDON, 2006; CARLIN und LOUIS, 2000, S. 8f., S. 137).

Formale Darstellung von MCMC Die Sequenz von Zufallsvariablen $\{X_n\}_{n \geq 0}$ ist eine Markovkette für jedes n . Gegeben dem aktuellen Wert X_n sind Vergangenheit und Zukunft unabhängig (vgl. GHOSH et al., 2006, S. 216). Eine Markov-Kette ist gewissermaßen „gedächtnislos“. Die gemeinsame Wahrscheinlichkeitsverteilungen von $X_j|_{j \in 0, \dots, n}$ für ein beliebiges n lässt sich im diskreten State Space wie folgt darstellen (vgl. GHOSH et al., 2006, S. 216):

$$\begin{aligned}
 & p(X_0 = i_0, X_1 = i_1, \dots, X_{n-1} = i_{n-1}, X_n = i_n) \\
 &= p(X_n = i_n | X_0 = i_0, \dots, X_{n-1} = i_{n-1}) \\
 & \times p(X_0 = i_0, X_1 = i_1, \dots, X_{n-1} = i_{n-1}) \\
 &= p(X_0 = i_0) p_{i_0 i_1} p_{i_1 i_2} \dots p_{i_{n-1} i_n}
 \end{aligned} \tag{4.12}$$

Die Wahrscheinlichkeitsverteilung ist stationär, wenn die Übergangswahrscheinlichkeiten sich im Zeitablauf nicht ändern. Es gilt dann Folgendes:

$$\begin{aligned}
 p(X_1 = j) &= \sum_i p(X_1 = j | X_0 = i) p(X_0 = i) \\
 &= \sum_i \pi_i p_{ij} \\
 &= p(x_0 = j) = \pi_j
 \end{aligned} \tag{4.13}$$

Bei der Anwendung von MCMC gilt es also herauszufinden, wann die Markov-Kette ihre stationäre Verteilung erreicht hat. Die Phase bis zur Konvergenz wird auch „Burn-in“ genannt. In der Praxis werden die Züge aus der Burn-in-Phase außer Acht gelassen und nur die Züge nach dem Burn-in zur Approximation der Posteriori-Verteilungen herangezogen (vgl. CARLIN und LOUIS, 2000, S. 295). Lässt man den Sampler jedoch sehr lange laufen, so ist dies überflüssig.

Beim Gibbs-Sampling konvergiert der Sampler in vielen Fällen bereits nach sehr wenigen Iterationen, sodass der „Burn-in“ nicht betrachtet werden muss. Bei der Durchführung von Schätzungen ist die Anzahl der Iterationen ein Entscheidungsparameter (ROBERT und CASELLA, 2004, S. 268). Der Anwender muss zunächst sicherstellen, dass der Sampler

konvergiert und anschließend eine ausreichende Anzahl von Zügen generieren, die die gewünschte Genauigkeit der Approximation ermöglicht.

4.2.3. Gibbs-Sampler

Der Gibbs-Sampler ist eine der einfachsten in der bayesianischen Statistik verwendeten Methoden. Ursprünglich wurde der Gibbs-Sampler in der Physik eingesetzt. Später folgte in den 1990er Jahren durch GELFAND und SMITH (1990) die Einführung in der bayesianischen Statistik (LYNCH, 2007, S. 88). Neben dem Beitrag von CASELLA und GEORGE (1992), welcher sich dediziert der Erklärung des Gibbs-Samplers widmet, finden sich auch gute einfache Einführungen bei GEWEKE (2005, S. 120f.) sowie LYNCH (2007, S. 88f.). Ausführlichere Abhandlungen finden sich bei TANNER (1997, S. 137f.) und GHOSH et al. (2006, S. 200f.). Auf deren Beiträgen basiert dieser Abschnitt.

Der Gibbs-Sampler ist ein Spezialfall des Metropolis-Hastings-Algorithmus, wobei die Annahmewahrscheinlichkeit stets 100% entspricht. LYNCH (2007, S. 89f.) stellt den Ablauf eines Gibbs-Samplers wie folgt dar ²⁰:

0. Startwerte S für den Parametervektor einsetzen, sodass $\theta^{j=0} = S$
1. Iterationsschritt $j = j + 1$ setzen
2. Konditionaler Zug $(\theta_1^j | \theta_2^{j-1}, \theta_3^{j-1} \dots \theta_k^{j-1})$
3. Konditionaler Zug $(\theta_2^j | \theta_1^j, \theta_3^{j-1} \dots \theta_k^{j-1})$
-
- k. Konditionaler Zug $(\theta_k^j | \theta_1^j, \theta_2^j \dots \theta_{k-1}^j)$
- k+1 Zurück zu Schritt 1

Jedes j stellt eine Iteration des Samplers dar. Die Zahl der durchzuführenden Iterationen richtet sich nach der gewünschten Genauigkeit der Approximation.

Die Beliebtheit des Gibbs-Sampling liegt unter anderem daran, dass die voll-konditionalen Verteilungen (auf Englisch kurz „full conditionals“) in den Schritten 1 bis k lediglich bis auf eine Konstante bekannt sein müssen (vgl. LYNCH, 2007, S. 89). Bei der Herleitung

²⁰ θ soll in diesem Abschnitt einen generischen Parametervektor darstellen.

der voll-konditionalen Verteilungen können daher konstante Terme vernachlässigt werden. Dies vereinfacht die Herleitung erheblich.

Nicht immer ist es jedoch möglich, alle konditionalen Verteilungen so herzuleiten, dass sie einer bekannten Form entsprechen. Für einzelne Parameter muss dann z.B. der Metropolis-Hastings-Algorithmus oder Adaptive-Rejection-Sampling verwendet werden.

4.2.4. Metropolis-Hastings Algorithmus

Der Metropolis-Hastings-Algorithmus (METROPOLIS et al., 1953; HASTINGS, 1970) setzt an der Konstruktion einer Markov-Kette an. Gegeben Zustand θ_j in der j -ten Iteration des Samplers stellt sich die Frage, welcher Wert für θ_{j+1} gelten soll.

Bei gegebener Ziel-Dichte π auf X erzeugt der MH-Algorithmus eine Markov-Kette \mathbf{X} mit stationärer Verteilung π . Zu jedem Iterationsschritt wird konditional auf $X_{k-1} = x$ ein Kandidat Y_k vorgeschlagen, der aus einer festzulegenden Vorschlagsdichte $q(x, \cdot)$ generiert wird. Anschließend wird der neue Wert nach folgendem Mechanismus ausgewählt: $X_k = Y_k$ wird mit der konditionalen Wahrscheinlichkeit $\alpha(x, y) = \min \{1, \pi(y)q(y, x)/\pi(x)q(x, y)\}$ angenommen. Wenn nicht, bleibt der Wert beim vorherigen stehen $X_k = X_{k-1}$, der Vorschlag wird abgelehnt (vgl. MEYN und TWEEDIE, 2009, S. 522f.). Der MH-Algorithmus funktioniert mit allen univariaten und multivariaten Verteilungen und ist somit universell einsetzbar. Dieser Flexibilität steht jedoch die Schwierigkeit der Definition einer Vorschlagsdichte gegenüber. CHIB und GREENBERG (1995); TIERNEY (1994) stellen diese Herausforderungen ausführlich dar.

4.2.5. Adaptive Rejection Sampling

Zuerst wurde Adaptive Rejection Sampling von WILD und GILKS (1992) entwickelt. In der ersten Variante benötigt man jedoch die Ableitung der Ziel-Dichte. Für die Variante von GILKS (1992) ist schließlich keine Ableitung mehr nötig. Hier spricht man auch vom „Derivative-free Adaptive Rejection Sampling“. Adaptive Rejection Sampling funktioniert bei univariaten log-konkaven Dichten und stellt eine nützliche Erweiterung des Gibbs-Samplings dar. In einigen Fällen können voll-konditionale Verteilungen eine unbekannte Form annehmen, für die kein Zufallszahlengenerator existiert. In diesem Fall kann

Adaptive Rejection Sampling eingesetzt werden, ohne dass „Tuning“ nötig ist. Aus diesem Grund wird es auch bei der verbreiteten Software *WinBUGS* eingesetzt, welche aber inzwischen nicht mehr weiterentwickelt wird (LUNN et al., 2009).

Adaptive Rejection Sampling erzeugt eine envelope-Funktion der log-Dichte. Diese envelope-Funktion wird für Rejection Sampling verwendet. Sobald ein Vorschlag abgelehnt wird, wird die envelope-Funktion verbessert, sodass sie näher an die wahre log-Dichte herankommt. Somit verringert sich die Wahrscheinlichkeit von Ablehnungen in folgenden Zügen. Das Sampling wird somit im Verlauf der Iterationen immer effizienter.

In einer Vergleichsuntersuchung stellen SON und OH (2006) fest, dass Adaptive Rejection Sampling für das Sampling von Shape-Parametern der Gamma-Verteilung effizienter funktioniert als der Metropolis-Hastings-Algorithmus. Aus diesem Grund wird für das hier entwickelte Modell auch auf Adaptive Rejection Sampling zurückgegriffen. Alternativ bietet sich auch Slice Sampling (NEAL, 2003) als Verfahren zum Sampling aus Nicht-Standard-Verteilungen an. Es gibt allerdings noch keine zu SON und OH (2006) vergleichbare Untersuchung für die hier benötigte Anwendung.

4.3. Strukturgleichungsmodelle

Strukturgleichungsmodelle spielen inzwischen eine große Rolle in der Betriebswirtschaft, Ökonomie und in anderen Disziplinen. Insbesondere im Marketing wird SEM häufig eingesetzt, um Beziehungen zwischen gemessenen, aber latenten Variablen zu quantifizieren. HERSHBERGER (2003) hat einen historischen Überblick über die Entwicklung der Strukturgleichungsmodellierung von 1994-2001 zusammengestellt. Einen neueren Überblick über die auch „Latent Variable Models“ genannten Modelle findet sich bei SKRONDAL und RABE-HESKETH (2007). Weitere Standardwerke sind die Bücher von SKRONDAL und RABE-HESKETH (2004), BARTHOLOMEW und KNOTT (1999) bzw. BARTHOLOMEW et al. (2011) sowie EVERITT (1984). SHAH und GOLDSTEIN (2006) bieten einen Überblick über die Anwendung von Strukturgleichungsmodellen in der Operations Management-Literatur von 1984 bis 2003. Sie schätzen SEM als eine sehr leistungsfähige Methode ein, finden jedoch unterschiedliche methodologische Probleme in den untersuchten Publikationen.

Besonders das LISREL Modell (JÖRESKOG und SÖRBOM, 1996) ist sehr verbreitet.

PALOMO et al. (2007, S. 165) nennen es das normale lineare SEM. Die Verbreitung lässt sich mutmaßlich auch durch den Aufruf, Regressionsanalysen durch SEM zu ersetzen, erklären. Unter anderen empfiehlt MACKENZIE (2001), SEM als dominierendes Analysewerkzeug zu verwenden. Das wichtigste Argument für die Verwendung von SEM liegt in der Kontrolle von Messfehlern, die etwa durch mehrdeutige Interpretationen von Fragen entstehen.

Regressionanalyse ist ein Spezialfall eines Strukturgleichungsmodells, wobei alle Messungen ohne Fehler durchgeführt werden. Die Messfehlerkontrolle wird durch die Einführung latenter Faktoren möglich, die durch mehrere Items gemessen werden. Darüber hinaus sind mit SEM und dem LISREL-Modell im Speziellen ganz unterschiedliche Modelle möglich: neben regressionsartigen Strukturen auch Faktormodelle oder ökonomische Pfadmodelle.

4.3.1. Modellierung

Bayesianische Implementierungen haben unter anderem STERN und JEON (2004) und LEE (2007) vorgestellt. Die Ausführungen zum Sampling der SEM-Parameter orientieren sich an LEE (2007). Dem LISREL-Modell liegt folgende Beziehung zwischen Messvariablen Y und latenten Faktoren Ω zugrunde:

$$y_i = \mu + \Lambda\omega_i + \varepsilon_i \quad (4.14)$$

wobei folgende Verteilungsannahmen getroffen werden:

$$\begin{aligned} y_i &\sim N(\mu + \Lambda\omega_i, \Psi_\varepsilon) \\ \omega_i &\sim N(0, \Sigma_\omega) \end{aligned} \quad (4.15)$$

Die latenten Faktoren bestehen aus zwei Teilen: endogene (η_i) und exogene (ξ_i) latenten Variablen. Somit gilt für die latenten Variablen $\omega_i = \begin{bmatrix} \eta_i & \xi_i \end{bmatrix}$. Die Beziehungen der latenten Faktoren werden im Strukturmodell beschrieben:

$$\eta_i = \Pi\eta_i + \Gamma\xi_i + \delta_i \quad (4.16)$$

Daraus ergibt sich folgende Kovarianzmatrix der latenten Faktoren:

$$\Sigma_\omega = \begin{bmatrix} \Pi_0^{-1} (\Gamma\Phi\Gamma^T + \Psi_\delta) \Pi_0^{-T} & \Pi_0^{-1}\Gamma\Phi \\ \Phi\Gamma^T\Pi_0^{-T} & \Phi \end{bmatrix} \quad (4.17)$$

wobei $\Pi_0 = I_{q_1} - \Pi$ definiert ist. Eine weitere zentrale Annahme des Modells ist, dass die Strukturfehlervarianzmatrix Ψ_δ sowie die Matrix der Messfehlervarianzen Ψ_ε diagonal sind, d.h. die Fehler werden als unkorreliert angenommen (EVERITT, 1984, S. 34).

So ergibt sich folgende konditionale Likelihood-Funktion mit konditionalem Mittelwert $\Lambda\omega_i$ und Varianz Ψ_ε :

$$p(Y|\Lambda, \Psi_\varepsilon, \Omega) \propto |\Psi_\varepsilon|^{-n/2} \exp \left[-\frac{1}{2} \sum_{i=1}^n (y_i - \Lambda\omega_i)^T \Psi_\varepsilon^{-1} (y_i - \Lambda\omega_i) \right] \quad (4.18)$$

Unkonditional entspricht die Likelihood einer multivariaten Normalverteilung (MVN)

$$y_i = N(\mu, \Sigma) \quad (4.19)$$

wobei für die Kovarianzmatrix gilt:

$$\Sigma = \Lambda\Sigma_\omega\Lambda^T + \Psi_\varepsilon \quad (4.20)$$

4.3.2. Messmodelle

Bei der Untersuchung von Kundenzufriedenheit werden häufig Messmodelle eingesetzt (siehe Abschnitt 3.3.2).

Die Verwendung mehrerer Indikatoren erlaubt die Schätzung von Messfehlern und führt somit zu einer Anpassung der geschätzten Pfadkoeffizienten (BOLLEN, 1989, S. 151ff.). Die Empfehlungen zur Anzahl der notwendigen Messvariablen sind unterschiedlich - bisweilen werden drei oder vier Messvariablen empfohlen (QUINTANA und MAXWELL, 1999, S. 489f.).

Messvariablen müssen konsistente Maße eines unidimensionalen Konstruktes sein (BOLLEN und LENNOX, 1991, S. 312). Oft wird zudem verlangt, dass die Korrelation von Messvariablen, die dem gleichen Konstrukt zugeordnet sind, größer sein sollte als die Korrelation zu anderen Messvariablen. BOLLEN und LENNOX (1991, S. 312) halten dies jedoch für problematisch und teilweise irreführend.

Wichtiger als die reine Anzahl der Indikatoren pro Konstrukt ist aber eine theoretisch aussagekräftige Repräsentation des Konstruktes mit den jeweiligen Indikatoren (QUINTANA und MAXWELL, 1999, S. 490). Messfehlervarianzen geben schließlich Aufschluss darüber, wie valide die Operationalisierung eines Konstruktes ist.

BOLLEN und LENNOX (1991) unterscheiden zudem reflektive und kausale (oftmals auch formativ genannte) Messmodelle. Formative Messmodelle werden jedoch zunehmend kritisch betrachtet (HOWELL et al., Jun; WILCOX et al., 2008; EDWARDS, 2011) und eignen sich im Kontext von (Teil)zufriedenheitskonstrukten nicht zur Operationalisierung.

4.3.3. Konstrukte mit einem Indikator

Schon auf dem Einband von WEIBER und MÜHLHAUS (2010) ist ein Strukturgleichungsmodell abgebildet, bei dem die latente Variable „Preis“ mit nur einem Faktor gemessen wird. Unter dem Stichwort „single-item measures“ ist dieser Fall Thema in der Literatur. Laut SACKETT und LARSON (1990, S. 468) ist dies kein Problem bei eindeutig definierten und einfachen Konstrukten. Jedoch bestünde bei komplexeren Konstrukten die Gefahr, dass ein einfaches Item schwerer interpretierbar sei. Letztendlich kommen sie zu dem Schluss, dass single-item measures generell vermieden werden sollten, solange es keine hinreichenden Belege dafür gebe.

Auch WANOUS et al. (1997) untersuchen single-item measures im Kontext von Arbeitszufriedenheit. In ihrer Meta-Analyse konstatieren sie schließlich:

„It should be interpreted, however, as a case for the acceptability of single-item measures when either the research question implies their use or when situational constraints limit or prevent the use of scales.“ (WANOUS et al., 1997, S.250)

Bezüglich der Anwendung in Strukturgleichungsmodellen äußern sich QUINTANA und MAXWELL (1999, S. 489) wie folgt: „It is possible, albeit undesirable, to have a construct indicated by a single instrument.“. Auch die Definition von BOLLEN (1989, S. 182) („A measurement model specifies a structural model connecting latent variables to one or more measures or observed variables“) schließt die Messung von Konstrukten mit nur einer Variable nicht aus.

Umgang mit Single-Item measures Zunächst ist es durchaus einleuchtend, dass auch bei nur einer verfügbaren Messvariable für ein Konstrukt eine gedankliche Trennung von Konstrukt und Messung möglich sein soll. Schließlich ist bei Strukturgleichungsmodellen

vor allem das Strukturmodell von Interesse. Darüber hinaus kann man auch in diesem Fall nicht immer davon ausgehen, dass die Messvariable das Konstrukt fehlerfrei misst. Konzeptionell ist es also durchaus wünschenswert, den Messfehler zu quantifizieren. Jedoch ist ein solches Modell nicht identifiziert.

Eine gängige Lösung ist, die Fehlervarianz auf einen Wert festzulegen (FUCHS und DIAMANTOPOULOS, 2009, S. 197). Wird die Messfehlervarianz auf 0 festgelegt, so entspricht dies der Annahme einer perfekten Messung des Konstruktes durch die Messvariable. Alternativ kann, z.B. basierend auf früheren Erfahrungen, ein anderer fester Wert angenommen werden (KLINE, 2011; FUCHS und DIAMANTOPOULOS, 2009, S. 276). Beispielhaft werden 10% und 20% der Indikatorvarianz genannt. Auch KLINE (2011, S. 276) empfiehlt ein derartiges Vorgehen:

„At times a researcher has only one measure of some construct. Scores from a single indicator are unlikely to be both perfectly reliable and valid. There is an alternative to representing a single indicator in the structural part of an SR model as one would in path analysis (i.e., without a measurement error term). This alternative requires an a priori estimate of the proportion of variance in a single indicator that is due to measurement error (10%, 20%, etc.).“²¹

Das gleiche Vorgehen schlagen auch (GODBOLD und PFAU, 2000, S. 424) vor:

„To test the structural model, the error variance for the indicator was fixed to correct for measurement error. Following Bollen (1989), the variance of the error was fixed to 1-alpha multiplied by the variance of the indicator.“

In einem bayesianischen Modell können zudem informative Prior-Verteilungen eingesetzt werden, um frühere Erfahrungen abzubilden. Somit ist das Modell schätzbar, auch wenn die Likelihood nicht identifiziert ist. Identifikation bei bayesianischen Modellen wurde in Abschnitt 4.2.1 ausführlicher behandelt.

Messen ohne Messfehler Wird die Messfehlervarianz auf 0 festgelegt, so gibt es keinen Unterschied mehr zwischen Konstrukt und Messvariable. Dies lässt sich an einem einfachen Beispiel mit einem endogenen Konstrukt mit einfachem Indikator verdeutlichen.

²¹Mit SR model wird ein Structural Regression Model bezeichnet.

Das Modell besteht aus 2 exogenen Konstrukten, die mit je zwei Messvariablen gemessen werden. Das endogene Konstrukt hat nur eine Messvariable. Das Messmodell ist gegeben durch:

$$\Lambda = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \lambda_1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & \lambda_2 \end{bmatrix}$$

sowie der diagonalen Matrix der Messfehlervarianzen Ψ_ε . Das Strukturmodell ist gegeben durch:

$$\Gamma = \begin{bmatrix} \gamma_1 & \gamma_2 \end{bmatrix}$$

sowie der Kovarianz der exogenen Variablen Φ und die Fehlervarianz π . Ohne Pfade zwischen endogenen Faktoren vereinfacht sich die Kovarianzmatrix zu:

$$\Sigma_\omega = \begin{bmatrix} \Gamma\Phi\Gamma^T + \psi_\delta & \Gamma\Phi \\ \Phi\Gamma^T & \Phi \end{bmatrix}$$

In diesem Fall ergibt sich damit:

$$\Sigma = \begin{bmatrix} \Gamma\Phi\Gamma^T + \pi & \gamma_1\phi_{11} + \gamma_2\phi_{12} & \gamma_2\phi_{12} + \gamma_2\phi_{22} \\ \gamma_1\phi_{11} + \gamma_2\phi_{21} & \phi_{11} & \phi_{12} \\ \gamma_2\phi_{12} + \gamma_2\phi_{22} & \phi_{21} & \phi_{22} \end{bmatrix}$$

Die Kovarianzmatrix der Messvariablen berechnet sich folgendermaßen: $\Lambda\Sigma\Lambda^T + \Psi_\varepsilon$. In diesem Fall ergibt sich:

$$\begin{bmatrix} \Gamma\Phi\Gamma^T + \pi + \psi_1 & \gamma_1\phi_{11} + \gamma_2\phi_{21} & \lambda_1(\gamma_1\phi_{11} + \gamma_2\phi_{21}) & \gamma_1\phi_{12} + \gamma_2\phi_{22} & \lambda_2(\gamma_1\phi_{12} + \gamma_2\phi_{22}) \\ \gamma_1\phi_{11} + \gamma_2\phi_{21} & \phi_{11} + \psi_2 & \lambda_1\phi_{11} & \phi_{12} & \lambda_2\phi_{12} \\ \lambda_1(\gamma_1\phi_{11} + \gamma_2\phi_{21}) & \lambda_1\phi_{11} & \lambda_1\lambda_1\phi_{11} + \psi_3 & \lambda_1\phi_{12} & \lambda_2\lambda_1\phi_{12} \\ \gamma_1\phi_{12} + \gamma_2\phi_{22} & \phi_{21} & \lambda_1\phi_{21} & \phi_{22} + \psi_4 & \lambda_2\phi_{22} \\ \lambda_2(\gamma_1\phi_{12} + \gamma_2\phi_{22}) & \lambda_2\phi_{21} & \lambda_1\lambda_2\phi_{21} & \lambda_2\phi_{22} & \lambda_2\lambda_2\phi_{22} + \psi_5 \end{bmatrix}$$

Das Identifikationsproblem wird nun deutlich: Der Term $\Gamma\Phi\Gamma^T + \pi + \psi_1$ besteht aus drei Summanden, wobei nur die Summe $\pi + \psi_1$ identifiziert ist. Dies lässt sich durch Festlegen der Messfehlervarianz auf 0 beheben. Allerdings ist dies identisch mit einem Modell,

bei dem die endogene Variable als gemeinsamer Messindikator der exogenen Variablen modelliert wird. Das entsprechende Messmodell umfasst nun die Parameter l_1 und l_2 :

$$\Lambda = \begin{bmatrix} l_1 & l_2 \\ 1 & 0 \\ \lambda_1 & 0 \\ 0 & 1 \\ 0 & \lambda_2 \end{bmatrix}$$

Das Strukturmodell besteht nur noch aus zwei latenten Faktoren, sodass $\Sigma = \Phi$ gilt. Die resultierende Kovarianzmatrix der Messvariablen ähnelt der zuvor gezeigten:

$$\begin{bmatrix} l_1(\phi_{11}l_1 + \phi_{21}l_2) + \psi_1 & l_1\phi_{11} + l_2\phi_{21} & \lambda_1(l_1\phi_{11} + l_2\phi_{21}) & l_1\phi_{12} + l_2\phi_{22} & \lambda_2(l_1\phi_{12} + l_2\phi_{22}) \\ l_1\phi_{11} + l_2\phi_{21} & \phi_{11} + \psi_2 & \lambda_1\phi_{11} & \phi_{12} & \lambda_2\phi_{12} \\ \lambda_1(l_1\phi_{11} + l_2\phi_{21}) & \lambda_1\phi_{11} & \lambda_1\lambda_1\phi_{11} + \psi_3 & \lambda_1\phi_{12} & \lambda_2\lambda_1\phi_{12} \\ \gamma_1\phi_{12} + l_2\phi_{22} & \phi_{21} & \lambda_1\phi_{21} & \phi_{22} + \psi_4 & \lambda_2\phi_{22} \\ \lambda_2(l_1\phi_{12} + l_2\phi_{22}) & \lambda_2\phi_{21} & \lambda_1\lambda_2\phi_{21} & \lambda_2\phi_{22} & \lambda_2\lambda_2\phi_{22} + \psi_5 \end{bmatrix}$$

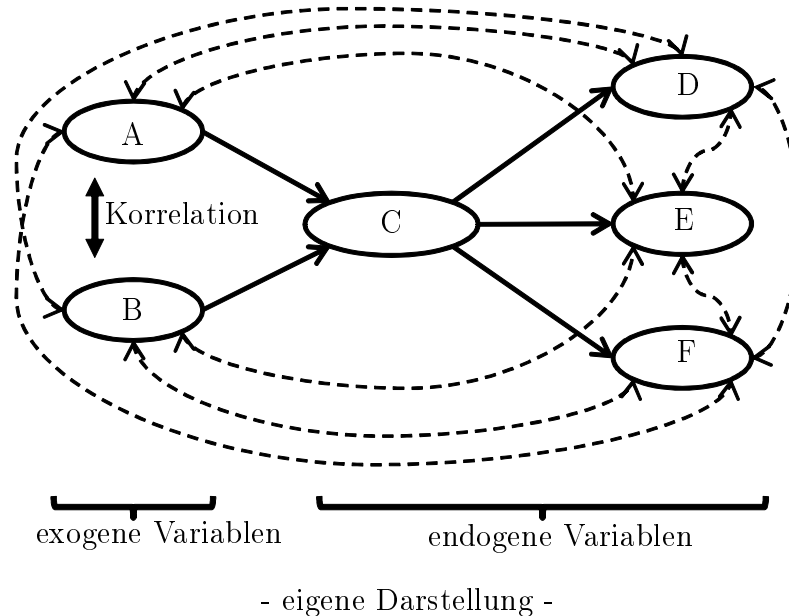
$\{l_1, l_2\}$ rücken an die Stelle von $\{\gamma_1, \gamma_2\}$. Zudem ist in Σ_{11} keine Strukturfehlervarianz mehr enthalten, das Modell ist identifizierbar.

Auch wenn beide Varianten äquivalent zu sehen sind, wird aufgrund der einfachen Interpretation die erste Variante gewählt: Obwohl keine Messung durch mehrere Konstrukte möglich ist, wird eine latente Variable eingeführt. Die Interpretation der Ergebnisse kann dann i. d. R. allein anhand des Strukturmodells erfolgen.

4.3.4. Äquivalente Modelle

Obwohl äquivalente Modelle ein großes Problem darstellen können, stellen MACCALLUM et al. (1993) fest, dass nur sehr wenige Forscher sich damit beschäftigen. Bereits STELZL (1986) erkannte, dass unterschiedliche Pfadmodelle die gleiche Kovarianzmatrix aufweisen können. Sofern alle möglichen Pfade definiert sind und somit ein gesättigtes Modell vorliegt, ist dies intuitiv nachvollziehbar. Im simplen Fall eines Regressionsmodells mit zwei Knoten ist stets ein Pfad definiert. Die Richtung des Pfades ist jedoch beliebig. Ganz

Abbildung 4.4: Äquivalenz in SEMs



analog lässt sich auch bei komplexeren gesättigten Pfadmodellen die Pfadrichtung beliebig definieren. Selbst wenn Pfadmodelle nicht vollständig gesättigt sind, ist es möglich, dass äquivalente Modelle existieren. Das heißt, dass alle mit Struktur *A* darstellbaren Kovarianzmatrizen auch durch Struktur *B* abgebildet werden können.

In dem beispielhaften Pfadmodell in Abbildung 4.4 sind sechs Variablen dargestellt. Es sind sechs Pfade definiert (Pfeile mit durchgezogenen Linien sowie der Doppelpfeil zwischen den zwei exogenen Variablen), die neun übrigen der 15 möglichen Pfade stellen die stärkere Annahme dar: Die zehn Pfeile mit gestrichelten Linien stellen Annahmen über Unabhängigkeiten dar. Sie entsprechen (durchgehenden) Pfeilen, deren Koeffizienten auf 0 festgelegt sind.

Das Konzept der Äquivalenz wurde später von LEE und HERSHBERGER (1990) aufgegriffen. Sie versuchen, die Regeln von STELZL (1986) zur Identifikation von äquivalenten Modellen zu vereinfachen. Auch RAYKOV und PENEV (2001) widmen sich der Problematik, jedoch nimmt das Problem äquivalenter Modelle insgesamt nur sehr wenig Raum in der Literatur ein.

Besondere Relevanz bekommt dieses Problem jedoch bei Mixture-Modellen. Werden bei

einem Mixture-Modell zwei identische Komponenten angenommen, um Heterogenität der Parameter zu ermöglichen, so ist das Mixture-Modell nicht identifiziert. Die Bedeutungen von Komponente 1 und Komponente 2 sind austauschbar. Um Identifikationsrestriktionen zu finden, kann z. B. der Random Permutation Sampler eingesetzt werden (FRÜWIRTH-SCHNATTER, 2006, S. 81f.). Sollen jedoch unterschiedliche Strukturen modelliert werden, so muss sichergestellt sein, dass es sich nicht um äquivalente Strukturen handelt. Sind die Strukturen identisch, so wird lediglich Heterogenität in Parametern adressiert und Identifikation muss durch Restriktionen sichergestellt werden. Sind die Strukturen jedoch unterschiedlich, sind keine weiteren Restriktionen zur Identifikation notwendig.

4.3.5. Kausalität

„[...] the causal interpretation of SEM is still a mystery to most SEMs researchers, leaders, educators, and practitioners.“ (PEARL, 2010a, S. 369)

Die kausale Interpretation von Strukturgleichungsmodellen wird in der Literatur teils kontrovers diskutiert. Eine umfassende Diskussion würde den Rahmen dieser Arbeit sprengen, jedoch ist die Interpretation von Strukturkoeffizienten ein wichtiger Bestandteil von Analysen mit SEM. Daher werden wesentliche Elemente der Diskussion aufgegriffen. Nicht alle Autoren stellen sich dieser Diskussion, sodass oft nur von „Zusammenhängen“ zwischen latenten Größen gesprochen wird. Die Art und Weise, wie SEMs verwendet werden, scheint sich dennoch kaum zu unterscheiden - unabhängig davon, ob Wissenschaftler der kausalen Interpretierbarkeit folgen oder nicht.²²

In der deutschen Literatur schreiben z.B. HOMBURG und HILDEBRANDT (1998, S. 17), dass der Begriff „Kausalanalyse“ nicht unproblematisch sei, da dies die Möglichkeit suggeriere, „mit Hilfe eines statistischen Verfahrens Kausalität zu untersuchen“, obwohl dies „im strengen wissenschaftstheoretischen Sinn jedoch nur mittels der bereits erwähnten kontrollierten Experimente möglich ist.“ Die Methode wird schließlich „Kovarianz-Strukturanalyse“ genannt, sodass das Wort „Kausalität“ vermieden wird. Auch KLINE (2011, S. 8)

²²Auch wenn die kausale Interpretierbarkeit in Publikationen verneint wird, klingen die Texte so, als würde man doch einer kausalen Interpretation zustimmen.

bezeichnet „causal modelling“ als veralteten Ausdruck und verneint die kausale Interpretierbarkeit von Strukturgleichungsmodellen.

Dieses Vorgehen ist jedoch inkonsequent. Die gezeichneten Pfaddiagramme laden geradezu zu einer kausalen Interpretation ein. Eine „quasi-kausale“ Interpretation des Strukturmodells gefolgt von einer Erklärung, dass es sich „im strengen wissenschaftstheoretischen Sinn“ aber doch nicht um Kausalität handelt, scheint zumindest ein fragwürdiger Umgang mit dem Modell zu sein.

Einer der bekanntesten Vertreter der kausalen Interpretierbarkeit ist Judea Pearl. Sehr deutlich nimmt er in nicht-begutachteten Manuskripten (PEARL, 2007, 2010b) und im Kapitel „Reflections, Elaborations, and Discussions with Readers“ des Nachdrucks von „Causality“ (PEARL, 2010a, S. 331 ff.) Stellung. Pearl beruft sich in verschiedenen Beiträgen auf die Anfänge von Strukturmodellen.

Die Ursprünge der Strukturmodelle (anfangs ohne Messmodell) gehen zurück auf Beiträge in der Biologie (WRIGHT, 1921) und Ökonometrie (HAAVELMO, 1943). Laut PEARL (1995, S. 244f.) sahen die frühen Anwender von SEM das Gleichheitszeichen als asymmetrisches Symbol, soweit Strukturgleichungen betroffen sind. Neben der rein mathematischen Beschreibung des Modells stehen somit weitere Annahmen. Dies verdeutlicht WRIGHT (1923) wie folgt „Prior knowledge of the causal relations is assumed as a prerequisite.“.

PEARL (1995, S. 243) nennt zwei Bedingungen für kausale Interpretierbarkeit: 1) Fehlende Pfeile zwischen X und Y im Pfadmodell entsprechen der Annahme, dass c.p. X keinen Einfluss auf Y hat und 2) dass die Abwesenheit eines Doppelpfeils zwischen X und Y bedeutet, dass es keine gemeinsamen Faktoren von X und Y gibt. Im Abschnitt „Defending the Causal Interpretation of SEM (or, SEM Survival Kit)“ stellt PEARL (2010a, S. 368ff.) dar, dass Analysen mit SEM kausale Schlüsse bei verhältnismäßig milden Annahmen zulassen:

- Die kausalen A-priori-Informationen sind qualitativer Natur, d.h. es wird nur die Richtung, nicht die Stärke des Zusammenhangs angegeben.
- „Gesunder Menschenverstand“ ist Startpunkt für die Vorgaben.
- Annahmen sind (theoretisch) mit Experimenten überprüfbar.
- Anpassungsmaße können zur Beurteilung von Modellen herangezogen werden.

Ein großer Teil der Kontroverse dreht sich folglich darum, inwieweit Vorab-Annahmen bezüglich des Pfadmodells akzeptiert werden. So stellt BOLLEN (1989, S. 44 ff.) zwar die kausale Interpretierbarkeit nicht grundsätzlich in Abrede, formuliert jedoch zusätzliche Bedingungen. Pearl hingegen sieht die kausale Interpretierbarkeit in den vorab definierten Annahmen und widerspricht dieser Auffassung (PEARL, 1995, S. 242).

Eine abschließende Bewertung fällt an dieser Stelle schwer, da sehr unterschiedliche Auffassungen aufeinandertreffen. Die Argumentation von Pearl und die kausale Interpretation auf Basis von vorab getroffenen Annahmen zum Strukturmodell erscheinen konsequent. Werden Treiberanalysen zur Kundenzufriedenheit durchgeführt, so impliziert bereits das Wort „Treiber“ eine Richtung und damit eine kausale Beziehung. Für jede im SEM untersuchte Beziehung ließe sich ein Experiment konstruieren, um die Kausalität zu überprüfen.

4.4. Ordinale Daten

Das klassische LISREL-Modell basiert auf der Annahme normalverteilter Messvariablen. Diese Annahme ist jedoch in aller Regel verletzt, wenn ordinale Rating-Skalen als Messvariablen verwendet werden. Zum einen sind nur diskrete Ausprägungen auf Rating-Skalen möglich, die Normalverteilung ist aber eine kontinuierliche Verteilung. Zufriedenheitsskalen sind zudem fast immer nicht symmetrisch, stattdessen folgen sie einer schiefen Verteilung (PETERSON und WILSON, 1992; YI, 1990).

Häufig werden ordinalskalierte Daten als „quasi-metrisch“ uminterpretiert (HOMBURG und KROHMER, 2003). Als eine Voraussetzung dafür werden äquidistante Abstände zwischen den Skalenelementen gesehen, sodass man hier auch von „Intervallskalenniveau“ spricht (KUMAR et al. (2002, S. 251), GREEN et al. (1988, S. 287), BEREKOVEN et al. (2000, S. 74), BACKHAUS et al. (2003, S. 6)). Äquidistante Ausprägungen sind jedoch eine sehr starke Annahme, sodass u.a. auch GREEN et al. (1988, S. 287) sie in Frage stellen. Durch entsprechende Beschriftung der Skalen hoffen einige Wissenschaftler, die Annahme des Intervallskalenniveaus rechtfertigen zu können (HAMMANN und ERICHSON (2000, S. 340f.), BORTZ und DÖRING (2002, S. 180)). Aufgrund der heute zur Verfügung stehenden methodischen Möglichkeiten gibt es jedoch keine Veranlassung, derart problematische Annahmen zu treffen.

Auch bei der Schätzung von SEMs mit Maximum-Likelihood-Methoden können kontinuierliche und ordinale Variablen durch die Einführung von latenten Messvariablen berücksichtigt werden (vgl. PALOMO et al., 2007, S. 163).

Die Idee, ordinale Variablen explizit zu modellieren (MCKELVEY und ZAVOINA, 1975), hat bereits sehr früh Eingang in die Strukturgleichungsmodellierung gefunden (MUTHEN, 1984; ARMINGER und KÜSTERS, 1988). Umso verwunderlicher ist die Tatsache, dass die Modellierung ordinaler Daten als „quasi-metrisch“ noch immer in der Literatur vorzufinden ist.

Dabei ist jedoch nicht einmal klar, ob die Bewertungen einer Rating-Skala überhaupt äquidistant (um Intervallskalenniveau zu erreichen) zu interpretieren sind. So weisen STAHL et al. (2009) auf ein nicht-lineares Verständnis der Zufriedenheitsskala hin. In einer Montecarlo-Studie zeigen COENDERS et al. (1997), dass die Annahme quasimetrischer Daten aus ordinalen Skalen zu verzerrten Ergebnissen führen kann. Darüber hinaus sind ordinale Daten oft von Skalenverwendungsheterogenität betroffen (siehe Abschnitt 2.2). Ordinale Rating-Skalen werden demnach unterschiedlich interpretiert, sodass Korrelationen im Aggregat verzerrt sein können.

Zunächst wird aber in Abschnitt 4.4.1 das Cut-Point-Modell vorgestellt, bevor in Abschnitt 4.4.2 auf die Adressierung von Skalenverwendungsheterogenität eingegangen wird.

4.4.1. Probit-Modell

Bei der Modellierung mittels eines Cut-Point-Modells wird angenommen, dass jeder Befragte latente kontinuierliche Bewertungen der einzelnen attributbezogenen Fragen vornimmt. Diese Bewertungen sind jedoch wie gesagt latent, also unbeobachtet. Beobachtet wird nur eine ordinale Antwort, die durch den Abgleich mit einem Vektor von Schwellenwerten („Cut-Points“) generiert wird. Durch das Cut-point-Modell wird eine Verteilung der möglichen latenten Bewertungen generiert. Diese latenten Bewertungen entsprechen wiederum der Annahme des LISREL-Modells - kontinuierlich und normalverteilt. Formal lässt sich das wie folgt ausdrücken: y_i sei der Vektor der ordinalen Antworten von Subjekt i , wobei gilt:

$$y_i = k \text{ wenn } c_{k-1} \leq z_{y,i} < c_k \quad (4.21)$$

Dabei gibt es $k+1$ Cut-Points $\{c_k : c_{k-1} \leq c_k, k = 1, \dots, K\}$, wobei $c_0 = -\infty$ und $c_k = \infty$. Zählt man minus Unendlich und plus Unendlich nicht zu den Cut-Points, so ergeben sich $k-1$ Cut-Points für k ordinale Kategorien. Da $c_0 = -\infty$ und $c_k = \infty$ unveränderlich gelten, erscheint dies sinnvoll, für die Notation ist die Einbeziehungen der Cut-Points im Unendlichen jedoch einfacher.

Die ordinalen Beobachtungen entsprechen einem ordered probit-Modell, wie es bei MCKELVEY und ZAVOINA (1975) vorgestellt wird.

Die Cut-Points sind somit Modellparameter. Jedoch müssen zwei Cut-Points (oder zwei weitere Cut-Points, wenn man $c_0 = -\infty$ und $c_k = \infty$ mitzählt) fixiert werden, damit das Modell identifiziert ist. Ein einfacher Ansatz ist es, c_1 und c_{K-1} auf Werte festzulegen, die der Anwendung der kumulativen empirischen Verteilung der ordinalen Daten auf Quantile der Standardnormalverteilung entsprechen (siehe auch BÜSCHKEN et al., 2010).

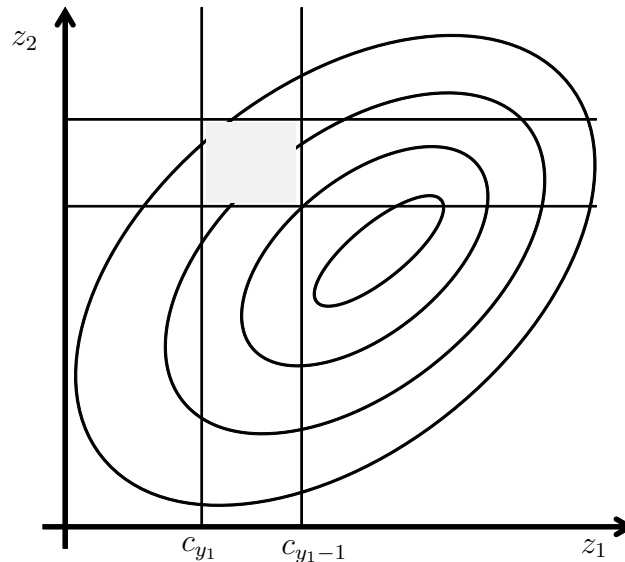
In Abbildung 4.5 ist die gemeinsame Verteilung zweier latenter kontinuierliche Variablen z_1 und z_2 dargestellt. Die Konturlinien deuten eine multivariate Normalverteilung von $Z = \begin{pmatrix} z_1 & z_2 \end{pmatrix}^T$ an. Die ordinalen Beobachtungen y_1 und y_2 steuern nun, welche Cut-Points für Variable 1 bzw. Variable 2 relevant sind. Die latenten Messvariablen werden anschließend aus dem markierten Bereich der multivariaten Normalverteilung generiert.

4.4.2. Skalenverwendungsheterogenität

Wie bereits in Abschnitt 2.2 dargelegt, verwenden Menschen vorgegebene Skalen in Fragebögen auf unterschiedliche Weise. „Ja-Sager“ bewerten alle Items tendenziell positiver, „Nein-Sager“ nutzen lieber das untere Ende der Skala (ROSSI et al., 2001). Diese Unterschiede sind in Abbildung 4.6 oben dargestellt. Dort bewertet Kunde 2 konsistent alle sechs Attribute besser als Kunde 1.

Auch nutzen manche Menschen nur einen Teil der Skala, sodass sich die Bewertung aller Items nur wenig um eine mittlere Bewertung herumbewegt. Andere wiederum nutzen die gesamte Skala vollständig aus. In Abbildung 4.6 unten sind Antwortmuster mit geringer und mittlerer Skalennutzung dargestellt. Kunde 2 differenziert hier in ähnlicher Weise wie Kunde 1, jedoch sind die Ausschläge bei Kunde 2 stärker, da dieser einen größeren Teil der zur Verfügung stehenden Skala nutzt.

Abbildung 4.5: Cut-Point-Modell im bivariaten Fall



- in Anlehnung an ROSSI et al. (2001, S. 23) -

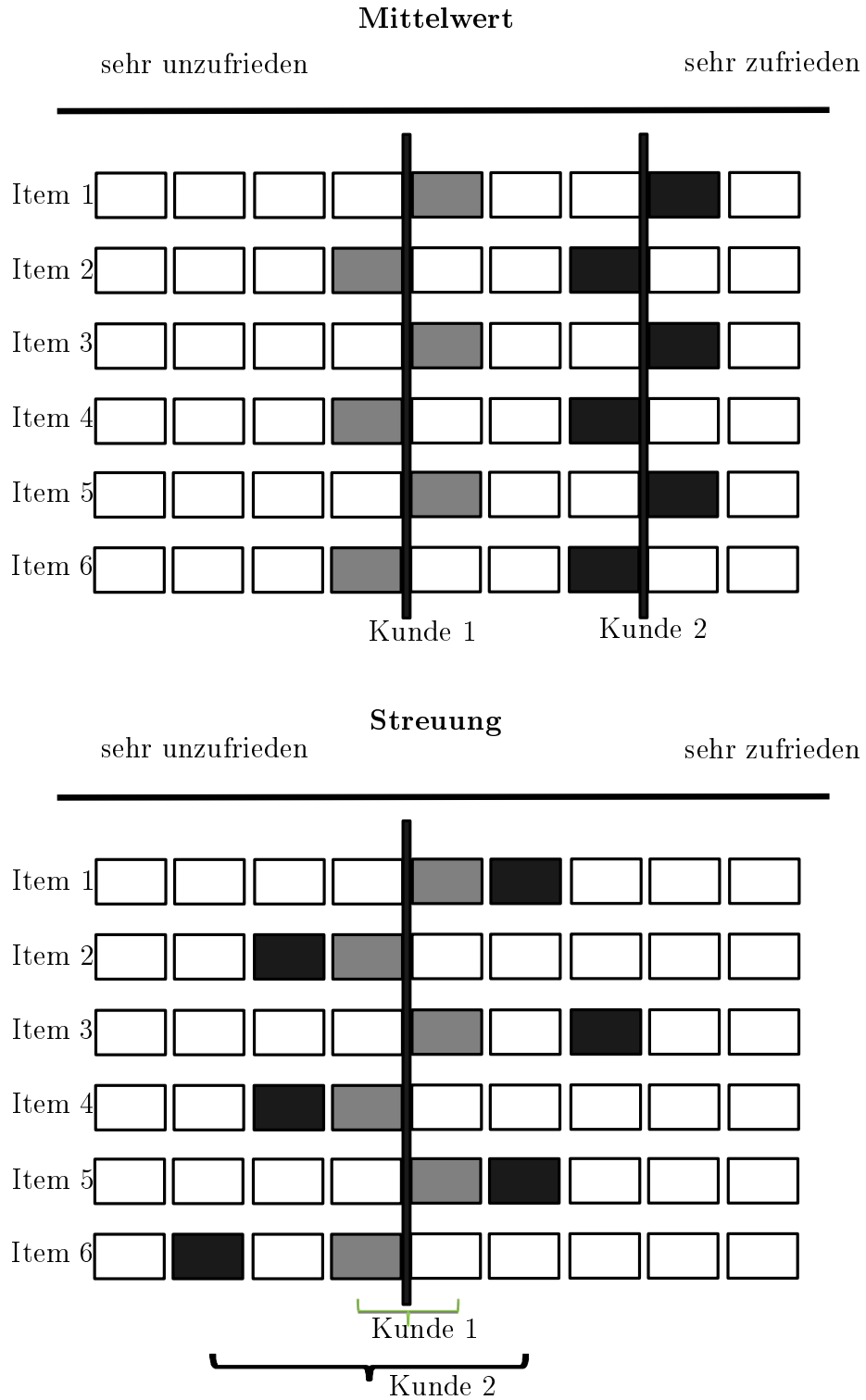
Diese Form der Heterogenität lässt sich auf zwei Arten begreifen:

1. als Individualisierung der mittleren Bewertung und der Streuung der individuellen Bewertungen
2. als Individualisierung der Cut-Points, die somit individuenspezifisch verschoben und gestaucht werden können

Wird die Skalenverwendungsheterogenität ignoriert, so werden „inflationierte“ Korrelationen in Kauf genommen. Zur Erläuterung soll ein Extrembeispiel dienen: Von 200 Umfrageteilnehmern bewerten 100 alle Teilaspekte mit 9-10 auf einer 10er-Skala. Die Bewertungen der übrigen Teilnehmer bewegen sich im Bereich von 1-2. Wird nun die Korrelation aller Items berechnet, so scheint es sehr starke Zusammenhänge zwischen allen diesen Items zu geben - allein durch die unterschiedlichen Mittelwerte der Bewertungen.

Die Daten müssen folglich erst „dekontaminiert“ werden, d. h. individueller Mittelwert und individuelle Standardabweichung der Antworten werden zur Standardisierung der Daten verwendet. Jegliche Inferenz sollte sich dabei auf die dekontaminierten Beobachtungen beziehen.

Abbildung 4.6: Skalenverwendungsheterogenität



- eigene Darstellung -

In vielen Fällen werden Daten mittelwertzentriert und teils auch standardisiert. Dieses Vorgehen ist jedoch nicht völlig unproblematisch (CLEMANS, 1956). Das Modell von ROSSI et al. (2001) hingegen basiert auf theoretischen Annahmen zum individuellen Antwortverhalten. Durch die Integration in das LISREL-Modell und die Schätzung durch MCMC kann zudem der Unsicherheit der Parameter zur Dekontaminierung Rechnung getragen werden. Formal lässt sich die „Dekontaminierung“ wie folgt beschreiben:

$$\mathbf{z}_i^* = \mathbf{z}_i / \sigma_i - \tau_i \boldsymbol{\iota} \quad (4.22)$$

Dabei ist $\boldsymbol{\iota}$ ein Vektor, bei dem alle Elemente gleich 1 sind.

Während ROSSI et al. (2001) eine multivariate a-Priori-Verteilung annehmen, wird hier von unabhängigen a-Priori-Verteilungen ausgegangen:

$$\begin{aligned} \tau_i &\sim N(0, \sigma_\tau) \\ \sigma_i^2 &\sim \text{Gamma}(a_\sigma, b_\sigma) \end{aligned} \quad (4.23)$$

In den bisher untersuchten Daten konnte keine deutliche Korrelation zwischen den Parametern festgestellt werden, sodass die Annahme unabhängiger A-Priori-Verteilungen gerechtfertigt scheint. Auch BÜSCHKEN et al. (2010) nehmen unabhängige A-Priori-Verteilungen an.

Beispiel Zur Demonstration des Skalenheterogenitätsmodells soll folgendes synthetisches Beispiel dienen: Für $n = 500$ individuenpezifische Beobachtungen in $p = 12$ Kategorien werden zunächst jeweils ein Mittelwert τ_i und eine Varianz σ_i^2 entsprechend folgender Verteilungen generiert:

$$\begin{aligned} \mu_i &\sim N(0, 1) \\ \sigma_i^2 &\sim \text{Gamma}(3, 3) \end{aligned} \quad (4.24)$$

Anschließend wird für jedes $i \in \{1, \dots, n\}$ ein Vektor der Länge $p = 12$ entsprechend einer Normalverteilung $N(\mu_i, \sigma_i)$ simuliert.

Die resultierende Pearson-Korrelationsmatrix ist in Tabelle 4.1 dargestellt. Obwohl kein Individuum eine Auskunft über Zusammenhänge (Korrelationen) gibt - die individuellen Beobachtungen teilen sich je einen Mittelwert - ergeben sich im Aggregat hohe Korrelationen (um 50%).

Tabelle 4.1: Korrelation durch Skalenverwendungsheterogenität

1	2	3	4	5	6	7	8	9	10	11	12
1											
0.46	1										
0.46	0.50	1									
0.53	0.49	0.42	1								
0.52	0.49	0.56	0.47	1							
0.42	0.45	0.44	0.45	0.44	1						
0.51	0.48	0.49	0.49	0.54	0.43	1					
0.48	0.44	0.49	0.45	0.50	0.45	0.49	1				
0.53	0.45	0.44	0.46	0.50	0.45	0.45	0.51	1			
0.49	0.48	0.48	0.47	0.51	0.49	0.52	0.53	0.49	1		
0.54	0.47	0.48	0.50	0.50	0.44	0.53	0.48	0.49	0.50	1	
0.51	0.50	0.52	0.51	0.47	0.44	0.51	0.46	0.49	0.50	0.48	1

Skalenverwendungsheterogenität im SEM Unter Berücksichtigung der Skalenverwendungsheterogenität ergibt sich für die Verteilung der latenten Messvariablen im SEM folgende abgeschnittene Normalverteilung:

$$[z_i | \mu_z, \mathbf{c}, \Lambda, \Psi_\varepsilon, \omega_i, \tau_i, \sigma_i, y_i] \sim MVN_{[c_{y_i}, c_{y_i+1}]} \left(\sigma_i (\tau_i + \mu_z + \Lambda \omega_i), (\sigma_i^2 \Psi_\varepsilon)^{0,5} \right) \quad (4.25)$$

Die Cut-Points c_{y_i}, c_{y_i+1} ergeben sich aus dem gemeinsamen Cut-Point-Vektor \mathbf{c} und dem individuellen (ordinalen) Beobachtungsvektor y_i .

Die univariate abgeschnittene MVN mit Schnittpunkten a, b lässt sich wie folgt darstellen:

$$N_{[a,b]}(\mu, \sigma) = \frac{\frac{1}{\sigma} \phi\left(\frac{x-\mu}{\sigma}\right)}{\phi\left(\frac{b-\mu}{\sigma}\right) - \phi\left(\frac{a-\mu}{\sigma}\right)} \quad (4.26)$$

wobei ϕ die kumulative Standardnormalverteilung ist. Die Kovarianzmatrix in Formel 4.25 ist aufgrund der Annahme unkorrelierter Messfehler diagonal, daher können alle p latenten Messvariablen unabhängig voneinander aus univariaten abgeschnittenen Normalverteilungen gezogen werden.

Das Modell erlaubt ebenfalls den Fall, dass die Messkoeffizienten in Λ Werte nahe 0 annehmen, sodass die Varianz ausschließlich durch Skalenverwendungsheterogenität entsteht.

Skalenhomogenität als Spezialfall Das ordinale Strukturgleichungsmodell ohne Berücksichtigung der Skalenverwendungsheterogenität ist ein Spezialfall, wobei $\sigma_i = 1, \tau_i = 0$

$\forall i \in \{1, \dots, n\}$ gilt.

Individualisierte Cut-Points Soll Heterogenität in der Skalenverwendung über eine Verschiebung von Mittelwert und Varianz hinaus modelliert werden, können auch individuelle Cut-Points implementiert werden (YING et al., 2006). Eine Verwendung dieses Cut-Point-Modells für die Schätzung von LISREL-Modellen ist jedoch aufgrund der großen Zahl von individuellen Parametern nicht trivial. Dem Verfasser dieser Arbeit ist keine Publikation bekannt, welche das Modell von YING et al. (2006) mit dem LISREL-Modell verknüpfen würde.

Das hier eingesetzte Modell von ROSSI et al. (2001) verwendet lediglich je zwei Parameter pro Individuum. Auch dieses Modell kann man als Individualisierung von Cut-Points interpretieren: Das „Dekontaminieren“ ist äquivalent zu einer Anpassung der Cut-Points entsprechend folgender Gleichung:

$$\mathbf{c}^* = \sigma_i^{-1} \mathbf{c} - \tau_i \quad \forall i \quad (4.27)$$

Personenspezifische Eigenschaft Es wird angenommen, dass die Art der Skalenverwendung eine personenspezifische Eigenschaft ist. Ist für einen Kunden beispielsweise 6 auf einer Siebenerkala eine gefühlt sehr gute Bewertung, so sollte dies für alle bewerteten Objekte (Produkte, Services,...) gelten. Werden also mehrere Objekte bewertet, so stehen mehr Informationen zur Schätzung der Skalenverwendungsparameter zur Verfügung.²³

4.5. Schätzung des Mixture-Modells

Sei y_i der Vektor der ordinalen Informationen von Individuum i und K die Anzahl der Komponenten, dann lässt sich die korrespondierende Likelihood-Funktion wie folgt ausdrücken:

$$f(y_i | \{\theta_k\}, \pi, \mathbf{c}, \sigma_i, \tau_i) = \sum_{k=1}^K \pi_k f(y_i | \theta_k, \mathbf{c}, \sigma_i, \tau_i) \quad \forall i \in \{1, \dots, n\} \quad (4.28)$$

Die komponentenspezifische Funktion $f(y_i | \theta_k, \mathbf{c}, \sigma_i, \tau_i)$ konditioniert dabei auf folgende Elemente des Modells:

²³Die technischen Details dazu sind in Abschnitt A.1.4 im Anhang dargestellt.

1. der k -spezifischen SEM-Komponente mit resultierender Kovarianzmatrix Σ_k der latenten dekontaminierten Messvariablen sowie den k -spezifischen Mittelwerten dieser Messvariablen μ_k
2. den gemeinsamen Vektor der Cut-Points \mathbf{c}
3. den Skalenheterogenitätsparametern σ_i und τ_i

Die Kovarianzmatrizen der dekontaminierten latenten Messvariablen unterscheiden sich nun k -spezifisch:

$$\Sigma_k = \Lambda_k \Sigma_{\omega k} \Lambda_k^T + \Psi_{\varepsilon k} \quad (4.29)$$

wobei für die Kovarianzmatrix der latenten Konstrukte der Komponente k Folgendes gilt:

$$\Sigma_{\omega k} = \begin{bmatrix} \Pi_{0k}^{-1} (\Gamma_k \Phi_k \Gamma_k^T + \Psi_{\delta k}) \Pi_{0k}^{-T} & \Pi_{0k}^{-1} \Gamma_k \Phi_k \\ \Phi_k \Gamma_k^T \Pi_{0k}^{-T} & \Phi_k \end{bmatrix} \quad (4.30)$$

mit $\Pi_{0k} = I_{q_{k1}} - \Pi_k$

Integrieren über $f(y_i | \{\theta_k\}, \pi, \mathbf{c}, \sigma_i, \tau_i)$ wäre sehr aufwendig. Analog zu DIEBOLT und ROBERT (1994) wird daher auf die Technik der Data Augmentation (TANNER und WONG, 1987) zurückgegriffen.

Die Wahrscheinlichkeit in einem Iterationsschritt des MCMC-Samplers für den Fall i zur Gruppe k zugeordnet zu werden $p(S_i = k | y_i, \{\theta_k\}, \mathbf{c}, \tau_i, \sigma_i)$, lässt sich durch folgenden Quotienten darstellen:

$$p(S_i = k | y_i, \{\theta_k\}, \mathbf{c}, \tau_i, \sigma_i) = \frac{p(y_i | \theta_k, \mathbf{c}, \tau_i, \sigma_i)}{p(y_i | \theta_1, \mathbf{c}, \tau_i, \sigma_i) + \dots + p(y_i | \theta_k, \mathbf{c}, \tau_i, \sigma_i)} \quad (4.31)$$

Eine konjugate Prior-Verteilung für ein Modell mit k Komponenten ist die Dirichlet-Verteilung (bzw. die Beta-Verteilung im Fall von nur zwei Gruppen):

$$Dir(\alpha = \{\alpha_1, \dots, \alpha_k\}) \quad (4.32)$$

Somit ergibt sich für die Posteriori-Verteilung:

$$p(S_i = k | y_i, \{\theta_k\}, \mathbf{c}, \tau_i, \sigma_i, \alpha) = \frac{p(y_i | \theta_k, \mathbf{c}, \tau_i, \sigma_i) \alpha_k}{p(y_i | \theta_1, \mathbf{c}, \tau_i, \sigma_i) \alpha_1 + \dots + p(y_i | \theta_k, \mathbf{c}, \tau_i, \sigma_i) \alpha_k} \quad (4.33)$$

Die Beobachtungen y_i sind ordinal, daher ist $p(y_i | \theta_k, \mathbf{c}, \tau_i, \sigma_i)$ eine Wahrscheinlichkeit von multivariaten Rechtecken einer MVN, welche durch den Cut-Point-Vektor \mathbf{c} definiert sind. Zur Berechnung wird der GHK-Simulator (GEWEKE, 1991; HAJIVASSILIOU und MCFADDEN, 1998) verwendet.

4.6. Beziehungen zu verwandten Verfahren

Probabilistische Testmodelle (sogenannte „Item Response“-Modelle) ähneln dem hier vorgestellten ordinalen LISREL-Modell.

So wenden beispielsweise PAGANI und ZANAROTTI (2010) das weit verbreitete Rasch-Modell auf Kundenzufriedenheitsdaten an. Ursprünglich wurden diese Modelle für die Auswertung von Tests (etwa im didaktischen Bereich) verwendet. Dabei werden Parameter für die Schwierigkeit einzelner Aufgaben oder Testfragen sowie Parameter für die Fähigkeit von Probanden verwendet. Das Interesse liegt dabei nicht nur in der Schwierigkeit der Aufgaben, sondern auch auf der Fähigkeitsbewertung der Probanden. Werden solche Modelle zur Modellierung von Zufriedenheit eingesetzt, dann ist die individuen-spezifische „Fähigkeit“ kaum relevant. Das Hauptinteresse liegt bei der „Schwierigkeit“, welche im Kontext von Zufriedenheit als Gewichtung verstanden werden kann (vgl. GHOSH et al., 2000). TAKANE und LEEUW (1987) zeigen, dass das häufig verwendete zweiparametrische Normal-Ogiven-Modell äquivalent zu einer Faktoranalyse dichotomisierter Variablen ist. Dies lässt sich auch auf den Fall von ordinalen Daten erweitern. Auch HOIJTINK et al. (1999) beschäftigen sich mit der Äquivalenz der zwei Modelle.

Diese Arbeit knüpft an das LISREL-Modell an, daher wird auf die Beziehung zur IRT an dieser Stelle nicht weiter eingegangen.

4.7. Implementierung

Die Implementierung des Modells erfolgte in der Programmiersprache R (R DEVELOPMENT CORE TEAM, 2011).

4.8. Simulationsstudie

Um die Funktionsfähigkeit des vorgestellten Mixture Modells zu demonstrieren, wird eine Simulationsstudie durchgeführt. Dabei werden Beobachtungen entsprechend der drei vorgestellten Komponenten simuliert:

- MIMIC-Komponente (entsprechend Abbildung 4.2)
- Halo-Komponente (entsprechend dem Halo-Modell in Abbildung 4.7)
- Regressions-Komponente (entsprechend dem Regressions-Modell in Abbildung 4.7)

Zu jeder Komponente werden 500 Fälle zufällig generiert. Die Struktur der Messmodelle ist bei allen Komponenten identisch.

Alle relevanten Parameter können aus den simulierten Daten geschätzt werden. Die wichtigsten Parameter sind in Tabelle 4.3 dargestellt: die wahren und geschätzten Strukturkoeffizienten, Mixture-Anteile sowie Cut-Points und Skalenverwendungsparameter. Zwei Cut-Points müssen zur Sicherstellung der Identifikation festgelegt werden. In der Simulationsstudie wurde diese nicht bei wahren Werten festgelegt. Die anderen Cut-Points sowie die Skalenverwendungsparameter weichen daher leicht von den wahren Werten ab, die wichtigen Strukturparameter sind davon jedoch nicht betroffen. Dies ist wichtig, da bei empirischen Daten ebenfalls keine Informationen bezüglich der „wahren“ Cut-Points vorhanden sind. Der Pfad $A \leftarrow GE$ wurde im MIMIC-Modell zwecks Identifikation festgelegt.

²⁴ Die Wiederherstellung der Parameter des Messmodells ist im Anhang (Tabelle A.1) zu finden. Dort ist ebenfalls in Tabelle A.2 die Recovery der Strukturfehlervarianzen und der Kovarianz der exogenen Konstrukte abgebildet.

Die Startwerte wurden dabei abweichend von den wahren Werten festgelegt, die äußeren Cut-Points wurden gegenüber den wahren Werten etwas weiter „auseinandergezogen“, um die Robustheit des Cut-Point-Modells zu demonstrieren. Im Skalenheterogenitätsmodell weicht a_σ im Posteriori Mittelwert leicht vom wahren Wert ab, was sich unter anderem auf zwei Ursachen zurückführen lässt: 1) Die äußeren Cut-Points wurden abweichend von

²⁴Neben der Restriktion der Strukturfehlervarianz der zentralen latenten Variable (Gesamteindruck) muss auch ein Pfad festgelegt werden. Eine Skalierung der Parameter sowie der Gesamtvarianz der latenten Variable Gesamteindruck würden ansonsten likelihood-äquivalente Modelle erzeugen.

den wahren Cut-Points definiert und 2) das Skalenheterogenitätsmodell besteht aus zwei individuellen Parametern, welche lediglich auf der Basis von 24 Beobachtungen geschätzt werden können.

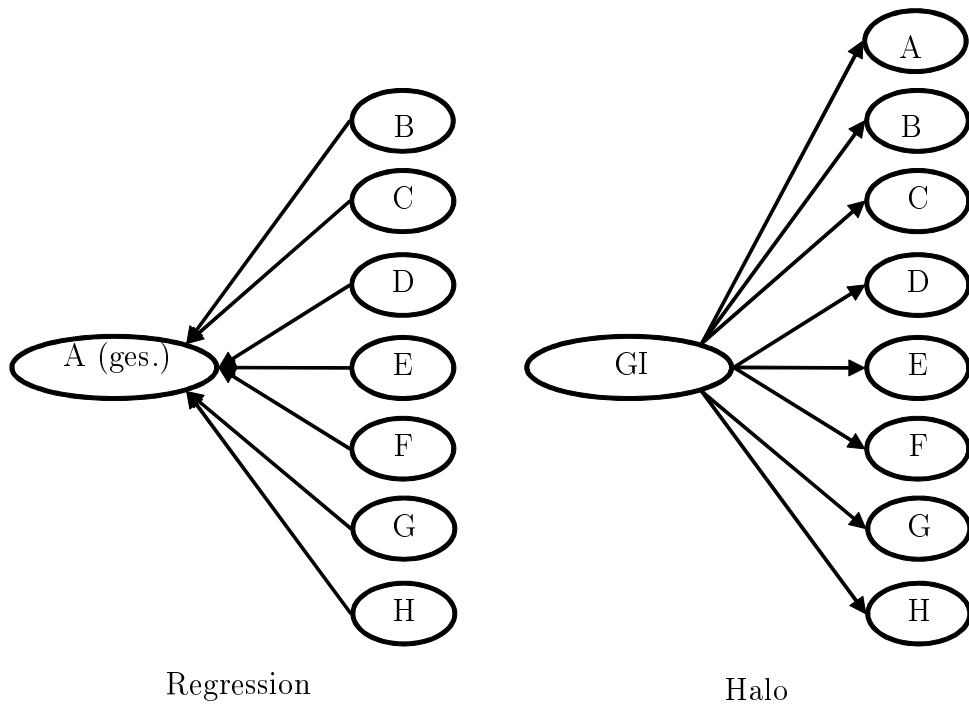
Die aus den unterschiedlichen Strukturmodellen resultierenden Korrelationen der latenten Konstrukte sind in Tabelle 4.2 dargestellt. Zwei Komponenten basieren auf neun latenten Konstrukten: MIMIC (Komponente 1) als auch Halo (Komponente 2). Ein latentes Konstrukt repräsentiert bei diesen Modellen den latenten und nicht gemessenen Gesamteindruck. In der Regressionskomponente (Komponente 3) gibt es nur acht latente Konstrukte, da kein latenter Gesamteindruck modelliert wird. Die dritte Komponente bildet lediglich den Zusammenhang von Gesamtzufriedenheit und Teilzufriedenheiten ab.

Die Trennung der Komponenten erfolgt auf Basis der Kovarianzinformation. Die Korrelationen der simulierten Konstrukte unterscheiden sich abhängig von der Komponente sehr deutlich voneinander, sodass die meisten Fälle der richtigen Komponente zugewiesen werden können. In Tabelle 4.2 ist zudem erkennbar, welche Formen Korrelationsmatrizen entsprechend der drei Modelle annehmen können:

- Die Korrelationsmatrix der Konstrukte in der MIMIC-Komponente lässt zwei Bereiche erkennen. Die Konstrukte 1-6 korrelieren allesamt sehr stark miteinander. Der Gesamteindruck (Konstrukt 1) erklärt die Varianz der Konstrukte 2-5 fast vollständig, sodass die Korrelationen sehr hohe Werte annehmen. Die Konstrukte 6-9 hingegen korrelieren kaum miteinander:
- Die Korrelationsmatrix der Konstrukte der Halo-Komponente besteht nur aus sehr großen Werten.
- Die Korrelationsmatrix der Konstrukte der Regressionskomponente zeigt nur schwache Korrelationen zwischen den erklärenden Konstrukten.

Die unterschiedliche Zahl differenzierter (miteinander schwach korrelierter) und nicht-differenzierter (stark korrelierter) Konstrukte ist Grundlage zur Unterscheidung der unterschiedlichen Modelle im Mixture-Modell.

Abbildung 4.7: Regressions- und Halo-Struktur Bsp.



- eigene Darstellung -

Tabelle 4.2: Simulationsstudie: Korrelation der latenten Faktoren (Simulationsstudie)

Komponente 1

	1	2	3	4	5	6	7	8	9
1	1.00	0.98	0.98	0.99	0.98	0.57	0.48	0.44	0.50
2	0.98	1.00	0.97	0.97	0.97	0.56	0.48	0.43	0.49
3	0.98	0.97	1.00	0.97	0.97	0.56	0.48	0.43	0.49
4	0.99	0.97	0.97	1.00	0.97	0.56	0.48	0.43	0.49
5	0.98	0.97	0.97	0.97	1.00	0.56	0.48	0.43	0.49
6	0.57	0.56	0.56	0.56	0.56	1.00	0.05	0.05	0.05
7	0.48	0.48	0.48	0.48	0.48	0.05	1.00	0.05	0.05
8	0.44	0.43	0.43	0.43	0.43	0.05	0.05	1.00	0.05
9	0.50	0.49	0.49	0.49	0.49	0.05	0.05	0.05	1.00

Komponente 2

	1	2	3	4	5	6	7	8	9
1	1.00	0.95	0.95	0.95	0.96	0.93	0.93	0.96	0.98
2	0.95	1.00	0.96	0.95	0.96	0.93	0.93	0.97	0.98
3	0.95	0.96	1.00	0.95	0.96	0.93	0.93	0.97	0.98
4	0.95	0.95	0.95	1.00	0.95	0.92	0.92	0.96	0.97
5	0.96	0.96	0.96	0.95	1.00	0.94	0.94	0.97	0.98
6	0.93	0.93	0.93	0.92	0.94	1.00	0.91	0.94	0.95
7	0.93	0.93	0.93	0.92	0.94	0.91	1.00	0.94	0.95
8	0.96	0.97	0.97	0.96	0.97	0.94	0.94	1.00	0.99
9	0.98	0.98	0.98	0.97	0.98	0.95	0.95	0.99	1.00

Komponente 3

	1	2	3	4	5	6	7	8
1	1.00	0.37	0.37	0.37	0.37	0.37	0.37	0.37
2	0.37	1.00	0.00	0.00	0.00	0.00	0.00	0.00
3	0.37	0.00	1.00	0.00	0.00	0.00	0.00	0.00
4	0.37	0.00	0.00	1.00	0.00	0.00	0.00	0.00
5	0.37	0.00	0.00	0.00	1.00	0.00	0.00	0.00
6	0.37	0.00	0.00	0.00	0.00	1.00	0.00	0.00
7	0.37	0.00	0.00	0.00	0.00	0.00	1.00	0.00
8	0.37	0.00	0.00	0.00	0.00	0.00	0.00	1.00

Tabelle 4.3: Simulationsstudie: Ergebnisse Teil 1

	wahr	PMW	2,5%	97,5%
MIMIC Strukturkoeffizienten				
A \leftarrow GE*	1	1	1	1
D \leftarrow GE	1.07	1.05	1.01	1.1
G \leftarrow GE	1.12	1.17	1.12	1.22
H \leftarrow GE	1.04	1.08	1.03	1.13
GE \leftarrow B	1.44	1.61	1.37	1.82
GE \leftarrow C	1.18	1.26	1.05	1.48
GE \leftarrow E	1.05	1.10	0.89	1.31
GE \leftarrow F	1.22	1.16	0.94	1.37
HALO Strukturkoeffizienten				
A \leftarrow GE	1.83	1.78	1.69	1.87
B \leftarrow GE	1.89	1.86	1.78	1.95
C \leftarrow GE	1.97	1.96	1.88	2.05
D \leftarrow GE	1.67	1.62	1.53	1.71
E \leftarrow GE	2.19	2.15	2.06	2.25
F \leftarrow GE	1.30	1.26	1.18	1.34
G \leftarrow GE	1.30	1.27	1.19	1.36
H \leftarrow GE	2.85	2.81	2.69	2.93
Regression Strukturkoeffizienten				
A \leftarrow B	2.00	1.90	1.58	2.23
A \leftarrow C	2.00	1.72	1.39	2.04
A \leftarrow D	2.00	1.99	1.70	2.28
A \leftarrow E	2.00	1.94	1.64	2.27
A \leftarrow F	2.00	2.21	1.87	2.52
A \leftarrow G	2.00	2.20	1.83	2.55
A \leftarrow H	2.00	1.82	1.52	2.13

*Pfad A \leftarrow GE wurde auf 1 festgelegt.

	wahr	PMW	2,5%	97,5%
Finite Mixture				
k = 1	0.33	0.33	0.29	0.36
k = 2	0.33	0.33	0.31	0.36
k = 3	0.33	0.34	0.31	0.37
Cut-Points				
cut_0	-2.68	-3.07	-3.07	-3.07
cut_1	-1.89	-1.90	-1.93	-1.85
cut_2	-1.21	-1.13	-1.16	-1.09
cut_3	-0.61	-0.55	-0.59	-0.52
cut_4	-0.02	-0.06	-0.10	-0.03
cut_5	0.60	0.47	0.43	0.50
cut_6	1.22	1.04	1.00	1.07
cut_7	1.86	1.74	1.70	1.77
cut_8	2.75	2.85	2.85	2.85
Skalenheterogenität				
σ_τ^2	0.25	0.24	0.21	0.27
a_σ	3.00	3.33	3.16	3.48
b_σ	3.00	3.00	3.00	3.00

5. Empirische Anwendung

In Abschnitt 5.1 wird für die beispielhafte empirische Anwendung des kausalen Mixture-Modells die Aufgabenstellung und Zielsetzung der Treiberanalyse dargelegt, in Abschnitt 5.2 der Datensatz vorgestellt, anschließend der verwendete Modellaufbau in Abschnitt 5.3 erläutert und schließlich werden die Ergebnisse in Abschnitt 5.4 dargelegt.

5.1. Aufgabenstellung der empirischen Anwendung

Kundenzufriedenheit besitzt eine sehr große Bedeutung für Unternehmen (siehe Abschnitt 3.1). Eine Steigerung der Kundenzufriedenheit wirkt sich i. d. R. positiv auf die Loyalität von Kunden aus und wird somit erfolgswirksam. Um knappe Ressourcen optimal zur Verbesserung der Zufriedenheit einzusetzen, ist die Kenntnis *relevanter Treiber von Zufriedenheit* wichtig.

Treiberanalyse mit empirischen Daten Für Dienstleistungsunternehmen - und damit auch für die Ketten von Schnellrestaurants - bringen MCDUGALL und LEVESQUE (2000, S. 392) die Aufgabe prägnant auf den Punkt: „For service firms the challenge is identifying the critical factors that determine customer satisfaction and loyalty.“ Restaurants stehen für diskrete Konsumerlebnisse mit hohen Erfahrungs- und Sucheigenschaften, sodass sie relativ leicht zu beurteilen sind (MCDUGALL und LEVESQUE, 2000, S. 397).

Alle vier untersuchten Schnellrestaurants bieten vor allem verschiedene Pizzen an. Über die angebotenen Produkte alleine ist somit kaum eine Differenzierung möglich. Daher wurden in der Befragung mehr als 20 Teilaspekte abgefragt, die potenziell relevant für die Gesamtzufriedenheit sind.

Die Restaurants ähneln sich in Bezug auf das Angebot und wurden daher entlang der gleichen Kriterien bewertet. Daher wird angenommen, dass prinzipiell die gleichen Teilaspekte für die Gesamtzufriedenheit relevant sind - unabhängig davon, welches Restaurant bewertet wird. Da (bis zu) vier Restaurants jeweils von dem gleichen Kunden bewertet worden sind, ist davon auszugehen, dass diesem Kunden ein Teilattribut (z.B. Geschmack) gleich wichtig ist, egal welches Restaurant er bewertet. Eine Segmentierung wird nicht vorgenommen, zumal alle Restaurants zur Kategorie der Pizza-Schnellrestaurants gehören

und alle Teilnehmer der Befragung Erfahrungen mit den bewerteten Restaurants gesammelt haben.

Halo-Effekt Aufgrund der zahlreichen empirischen Nachweise des Halo-Effekts in einer Reihe sehr unterschiedlicher Untersuchungen (siehe Abschnitt 2.3.4) kann auch hier nicht davon ausgegangen werden, dass alle Kunden unabhängige Bewertungen aller Teilaspekte vornehmen. Vielmehr ist davon auszugehen, dass ein bedeutender Teil der Kunden Bewertungen abgibt, welche lediglich eine Reflektion des jeweiligen Gesamteindrucks sind. Eine Analyse relevanter Treiber von Kundenzufriedenheit sollte sich daher nur auf die Antworten beziehen, die tatsächlich Informationen über relevante Treiber liefern.

MIMIC-Struktur Aufgrund der großen Anzahl von zu bewertenden Attributen ist die Annahme von BÜSCHKEN et al. (2010), dass entweder alle Bewertungen unabhängig voneinander vorgenommen werden oder alle Bewertungen lediglich eine Reflektion eines Gesamteindrucks sind, möglicherweise zu restriktiv. Selbst wenn Kunden sich bemühen, differenzierte Bewertungen vorzunehmen, so gelingt ihnen dies möglicherweise nur zum Teil.

Daher sollen zunächst inhaltliche Domänen zusammengefasst werden. Darüber hinaus soll untersucht werden, inwieweit partielles Haloing eine Rolle spielt. Die unterschiedlichen Aspekte der Restaurants stellen die Befragten vor unterschiedliche Herausforderungen: Die Bewertung der Erreichbarkeit basiert stärker auf Fakten (explizites Gedächtnis), während die Bewertung der Atmosphäre oder des Geschmacks stärker durch implizites Wissen beeinflusst werden kann. Somit gelingt es einem Teil der Befragten möglicherweise nur für einen Teil der Attribute, differenzierte Urteile zu bilden (entsprechend des formativen Modells). Dies soll durch die Modellierung unterschiedlicher MIMIC-Strukturen im Mixture-Modell getestet werden. Zumindest für einen Teil der Attribute könnten somit auch Treiberinformationen gewonnen werden, auch wenn nicht alle Attribute unabhängig bewertet werden. Aufgrund der großen Zahl möglicher MIMIC-Strukturen sollen zunächst nur einige wenige Varianzen getestet werden.

Wichtigstes Ziel der empirischen Anwendung ist jedoch die *Identifikation relevanter Treiber* unter der Berücksichtigung des Antwortverhaltens. Die genaue Struktur des Halo-Modells ist für Unternehmen weniger von Bedeutung, wenn die Aufgabe in der Identi-

fikation von Zufriedenheitstreibern besteht. Zur Lösung dieser Aufgabe ist lediglich die Regressionskomponente sowie der formative Teil der MIMIC-Komponente von Bedeutung.

5.2. Beschreibung des Datensatzes

Die Daten stammen aus der Bewertung vier unterschiedlicher Ketten von Schnellrestaurants durch 283 Kunden. Jeder Kunde wurde aufgefordert, soweit möglich ²⁵ alle vier Restaurantketten nach unterschiedlichen Kriterien zu bewerten.

Die Befragung umfasste teilweise Domänen wie etwa Atmosphäre und Geschmack, welche mehrere Items zugeordnet werden können. Weiterhin gibt es Attribute wie z.B. die Wartezeit, die nur mit je einer Frage erhoben worden sind.²⁶ Aufgrund der Vielzahl der untersuchten Einflussfaktoren und des faktischen Charakters (wie z. B. Erreichbarkeit) ist die Verwendung von single-item Konstrukten nachvollziehbar. Die Items sowie die Zusammenfassung zu Domänen bzw. Konstrukten sind in Tabelle 5.1 dargestellt.

In Tabelle 5.1 sind zudem die relevanten deskriptiven Statistiken dargestellt. Insgesamt wurden 865 Bewertungen über vier unterschiedliche Restaurants abgegeben. Die Zahl der abgegebenen Bewertungen pro Kunde variiert, da nicht alle der 283 Kunden mit allen Restaurants vertraut waren. In Tabelle 5.2 ist ersichtlich, wie viele Kunden jeweils wie viele Bewertungen abgegeben haben.

Weitere Informationen zu den genauen Formulierungen der Fragen, zum Ort und zu den genauen Namen der Lokale können aus Datenschutzgründen an dieser Stelle leider nicht erwähnt werden. Die Spearman-Korrelation der Rohdaten ist in Tabelle 5.3 ersichtlich. Auf Basis der Korrelationen können allerdings noch keine Aussagen zum Haloing getroffen werden, da es keine „wahren“ Korrelationen für einen etwaigen Vergleich gibt und zum anderen, da Skalenverwendungsheterogenität die Daten möglicherweise kontaminiert. Dennoch fällt auf, dass die Variablen keinesfalls unkorreliert sind. Zu stark korrelierte erklärende Variablen stellen jedoch ein Problem für die Durchführung von Treiberanalysen dar.

²⁵D. h. sofern Erfahrungen mit den jeweiligen Restaurants gemacht worden sind.

²⁶Auf die Schwierigkeit der Modellierung von Konstrukten im LISREL-Modell, welche nur durch je eine Frage operationalisiert werden, wurde bereits in Abschnitt 4.3.3 eingegangen.

Tabelle 5.1: Deskriptive Statistiken der Zufriedenheitsdaten

Faktor¹	α^2	Item	E(x)	S(x)	Median
		(1) Gesamt	4.77	1.27	5.00
Geschmack	0.8737	(2) Geschmack	5.07	1.54	5.00
		(3) Aroma	5.06	1.36	5.00
		(4) Qualität	5.05	1.39	5.00
		(5) Nährwert	3.91	1.45	4.00
Gesundheit	0.7613	(6) Gesundheit	4.01	1.39	4.00
		(7) Lieferung (Zeit)	4.60	1.23	5.00
Lieferservice	0.71945	(8) Lieferung (Qualität)	4.73	1.26	5.00
		(9) Service	4.79	1.26	5.00
Umfeld	0.7953	(10) Sauberkeit	4.77	1.23	5.00
		(11) Atmosphäre	4.69	1.31	5.00
		(12) Ruf	5.20	1.29	5.00
		(13) Mengen	4.97	1.30	5.00
		(14) Preis	3.69	1.69	4.00
		(15) Wert	4.55	1.35	5.00
		(16) Abwechslung	4.98	1.31	5.00
		(17) Auswahl (Belag)	5.34	1.34	6.00
		(18) Verpackung	4.82	1.31	5.00
		(19) Teig	4.35	1.55	4.00
		(20) Weitere Produkte	5.03	1.32	5.00
		(21) Wartezeit	4.75	1.20	5.00
		(22) Erreichbarkeit	4.56	1.50	5.00

¹ bzw. Domäne

² Cronbach's alpha

Tabelle 5.2: Zahl der abgegebenen Bewertungen

Zahl der Bewertungen	1	2	3	4
Kunden	30	45	87	121

Tabelle 5.3: Spearman-Korrelation der Rohdaten

1	1.00	0.67	0.60	0.68	0.38	0.37	0.45	0.66	0.55	0.50	0.50	0.62	0.56	0.07	0.60	0.44	0.48	0.48	0.44	0.37	0.32	0.48
2	0.67	1.00	0.71	0.72	0.24	0.20	0.24	0.50	0.50	0.32	0.35	0.56	0.50	-0.07	0.53	0.41	0.48	0.39	0.34	0.27	0.17	0.30
3	0.60	0.71	1.00	0.66	0.23	0.20	0.30	0.50	0.57	0.31	0.31	0.50	0.53	0.00	0.50	0.41	0.46	0.39	0.33	0.28	0.21	0.29
4	0.68	0.72	0.66	1.00	0.34	0.32	0.31	0.56	0.55	0.39	0.40	0.55	0.52	0.02	0.65	0.49	0.47	0.47	0.40	0.31	0.25	0.29
5	0.38	0.24	0.23	0.34	1.00	0.62	0.29	0.31	0.31	0.27	0.35	0.24	0.30	0.25	0.40	0.23	0.16	0.28	0.28	0.17	0.19	0.24
6	0.37	0.20	0.20	0.32	0.62	1.00	0.33	0.34	0.27	0.29	0.35	0.22	0.30	0.19	0.33	0.22	0.21	0.31	0.33	0.22	0.22	0.23
7	0.45	0.24	0.30	0.31	0.29	0.33	1.00	0.59	0.40	0.41	0.42	0.35	0.29	0.20	0.37	0.31	0.24	0.36	0.35	0.32	0.52	0.43
8	0.66	0.50	0.50	0.56	0.31	0.34	0.59	1.00	0.50	0.49	0.44	0.55	0.53	0.05	0.50	0.39	0.43	0.49	0.44	0.41	0.39	0.49
9	0.55	0.50	0.57	0.55	0.31	0.27	0.40	0.50	1.00	0.49	0.49	0.44	0.43	0.19	0.52	0.42	0.35	0.38	0.30	0.37	0.41	0.37
10	0.50	0.32	0.31	0.39	0.27	0.29	0.41	0.49	0.49	1.00	0.68	0.46	0.33	0.10	0.39	0.29	0.30	0.39	0.39	0.48	0.48	0.36
11	0.50	0.35	0.31	0.40	0.35	0.35	0.42	0.44	0.49	0.68	1.00	0.43	0.35	0.11	0.41	0.28	0.33	0.41	0.37	0.40	0.41	0.35
12	0.62	0.56	0.50	0.55	0.24	0.22	0.35	0.55	0.44	0.46	0.43	1.00	0.47	-0.11	0.43	0.43	0.47	0.44	0.44	0.35	0.27	0.48
13	0.56	0.50	0.53	0.52	0.30	0.30	0.29	0.53	0.43	0.33	0.35	0.47	1.00	-0.03	0.43	0.42	0.53	0.50	0.34	0.45	0.24	0.32
14	0.07	-0.07	0.00	0.02	0.25	0.19	0.20	0.05	0.19	0.10	0.11	-0.11	-0.03	1.00	0.29	0.01	-0.13	-0.02	-0.03	-0.02	0.21	0.12
15	0.60	0.53	0.50	0.65	0.40	0.33	0.37	0.50	0.52	0.39	0.41	0.43	0.43	0.29	1.00	0.37	0.32	0.38	0.36	0.27	0.33	0.32
16	0.44	0.41	0.41	0.49	0.23	0.22	0.31	0.39	0.42	0.29	0.28	0.43	0.42	0.01	0.37	1.00	0.45	0.39	0.35	0.30	0.22	0.29
17	0.48	0.48	0.46	0.47	0.16	0.21	0.24	0.43	0.35	0.30	0.33	0.47	0.53	-0.13	0.32	0.45	1.00	0.42	0.34	0.36	0.21	0.28
18	0.48	0.39	0.39	0.47	0.28	0.31	0.36	0.49	0.38	0.39	0.41	0.44	0.50	-0.02	0.38	0.39	0.42	1.00	0.43	0.46	0.28	0.29
19	0.44	0.34	0.33	0.40	0.28	0.33	0.35	0.44	0.30	0.39	0.37	0.44	0.34	-0.03	0.36	0.35	0.34	0.43	1.00	0.46	0.27	0.37
20	0.37	0.27	0.28	0.31	0.17	0.22	0.32	0.41	0.37	0.48	0.40	0.35	0.45	-0.02	0.27	0.30	0.36	0.46	0.46	1.00	0.43	0.31
21	0.32	0.17	0.21	0.25	0.19	0.22	0.52	0.39	0.41	0.48	0.41	0.27	0.24	0.21	0.33	0.22	0.21	0.28	0.27	0.43	1.00	0.34
22	0.48	0.30	0.29	0.29	0.24	0.23	0.43	0.49	0.37	0.36	0.35	0.48	0.32	0.12	0.32	0.29	0.28	0.29	0.37	0.31	0.34	1.00

5.3. Modellaufbau

Aufgrund der korrelierenden Items innerhalb der Fragebatterien zu den Konstrukten bietet sich die Regressionanalyse als Werkzeug nicht an. Stark korrelierte exogene Variablen könnten dabei zu (Multi-)Kollinearitätsproblemen führen. Zudem sollten a priori bekannte Gruppierungen bzw. Domänen nicht ignoriert werden. Folgende Domänen wurden als latente Konstrukte modelliert (Zahl der Items in Klammern):

- Geschmack (3 Items)
- Gesundheit (2 Items)
- Lieferservice (2 Items)
- Atmosphäre (3 Items)

In der klassischen Statistik wird Cronbach's alpha zur Beurteilung der Konstruktreliabilität herangezogen. Werte ab $\alpha = 0,7$ gelten für Konstrukte als akzeptabel (FIELD, 2005, S. 666 ff.).²⁷ Die berechneten Werte für die Konstrukte befinden sich auch in Tabelle 5.1. Die Items „Preis“ und „Wert“ wurden nicht zu einer Komponente zusammenfasst, weil deren Cronbach's alpha nur $\alpha = 0.4614823$ beträgt. Obwohl eine Zusammenfassung der Items „Preis“ und „Wert“ inhaltlich sinnvoll erscheint, wurde aufgrund der geringen Reliabilität des Konstruktes auf eine Zusammenfassung verzichtet und das Modell mit der Zusammenfassung dieser Items wieder verworfen. Auch nach Korrektur von Skalenheterogenität ist eine Zusammenfassung nicht sinnvoll möglich, da die Messfehlervarianz sehr große Werte annimmt²⁸.

Mit Hilfe der LISREL-Komponenten im Mixture-Modell können die vier durch mehrere Items operationalisierten Konstrukte berücksichtigt werden. Konstrukte mit je nur einer Messvariable werden auch als Konstrukte (d. h. als latente Variablen) modelliert, jedoch muss man hier die Messfehlervarianz festlegen (siehe dazu Abschnitt 4.3.3). Die

²⁷ Angesichts der Problematik heterogener Response Styles ist es durchaus fraglich, inwieweit Cronbach's alpha noch eine Bedeutung hat. Jedoch ist eine „Dekontamination“ der Daten erst möglich, wenn ein Modell angewendet wird. Daher wird Cronbach's alpha zunächst nur als grobe Indikation verstanden.

²⁸ Bei unterschiedlichen Modellvarianten wurden Werte von 3 regelmäßig überschritten.

Messfehlervarianz wurde hier auf einen kleinen Wert von 0.1 festgelegt.²⁹ Aufgrund der Erhebung der Daten auf ordinalem Skalenniveau und der Annahme kontinuierlicher latenter Messvariablen ist ein Messfehler von 0 konzeptionell unmöglich. Die Festlegung auf einen kleinen arbiträren Wert ist zudem konsistent mit in Abschnitt 4.3.3 dargestellten Beiträgen der Literatur.

Die Komponenten Während bei der *Halo*-Komponente davon ausgegangen wird, dass alle Zufriedenheitskonstrukte durch einen latenten Faktor erklärt werden können, werden in der *Regressionskomponente* alle Teilzufriedenheitskonstrukte als (potentielle) Treiber der Gesamtzufriedenheit modelliert. Die Pfadmodelle ähneln denen der Simulationsstudie (Abbildung 4.7); aufgrund der großen Anzahl an Variablen wird an dieser Stelle auf eine separate Darstellung des Pfadmodells verzichtet. An Stelle der *MIMIC*-Komponente wurden verschiedene Varianten gerechnet.

MIMIC-Struktur Das Pfadmodell der *MIMIC*-Komponente ähnelt der beispielhaften Darstellung in Abbildung 4.2. Auf der Seite der Treiber befinden sich in Abhängigkeit der betrachteten Variante die folgenden Variablen:

Tabelle 5.4: MIMIC-Varianten

	MIMIC Variante 1	MIMIC Variante 2	MIMIC Variante 3
			Wert
Treiber		Portionen	Portionen
	Auswahl Belag	Auswahl Belag	Auswahl Belag
	Verpackungsdesign	Verpackungsdesign	Verpackungsdesign
	Auswahl Teig	Auswahl Teig	Auswahl Teig
	Auswahl Sonst.	Auswahl Sonst.	Auswahl Sonst.
	Wartezeit	Wartezeit	Wartezeit
	Erreichbarkeit	Erreichbarkeit	Erreichbarkeit

Alle übrigen Konstrukte werden als Reflektion der Gesamteindrucks betrachtet. Diese Komponente stellt somit einen Mittelweg zwischen reinem „Haloing“ und reinem „For-

²⁹Eine informative Prior-Verteilung wie $IG(1, 10)$ führt zu sehr ähnlichen Ergebnisse. Die starre Festlegung auf 0.1 wurde aus Gründen der Einfachheit gewählt.

ming“ - also unabhängiger Treiberbewertung - dar. Auf eine mögliche Motivierung dieser Art Modellierung wurde im Absatz „Motivation für die MIMIC-Modellierung“ auf Seite 71 eingegangen. Unterschiedliche kognitive Herausforderungen sowie der Rückgriff auf explizites bzw. implizites Wissen könnten Ursachen für partielles Haloing sein. Ziel dieser Arbeit ist jedoch nicht die Berechnung aller möglichen MIMIC-Strukturen. Vielmehr wird demonstriert, dass ein MIMIC-Bewertungsstil prinzipiell vorstellbar und modellierbar ist. Die Suche nach dem oder den „richtigen“ MIMIC-Modell(en) ist hingegen eine neue Herausforderung, welche effiziente Suchalgorithmen voraussetzt.

Modellvergleich Um den Einfluss der Modellierung des Halo-Effektes sowie der MIMIC-Komponente auf den Modellfit zu beurteilen, werden folgende Modelle geschätzt:

- *Einkomponentenmodell*, welches nur die Regressionskomponente umfasst, also den Halo-Effekt ignoriert
- *Zweikomponentenmodell*, welches zusätzlich eine reine Halo-Komponente beinhaltet
- *Dreikomponentenmodell*, umfasst zusätzlich eine MIMIC-Komponente. Darüber hinaus wurden die in Tabelle 5.4 dargestellten Varianten verwendet. Je weniger Treiber („linksseitige“ Variablen im Pfadmodell) die MIMIC-Komponente umfasst, desto ähnlicher wird sie der Halo-Komponente.

Zum Vergleich wird marginale Likelihood verwendet. Zur Berechnung wird auf die Methode von NEWTON und RAFTERY (1994) zurückgegriffen. Dieses Vorgehen und eine knappe Diskussion der Schwächen des Newton-Raftery-Verfahrens finden sich auch in ROSSI et al. (2006, S. 166ff.). Der Vorteil des "Herausintegrierens" der Parameter ist, dass Modelle ohne die Berücksichtigung weiterer Strafterme vergleichbar gemacht werden.

Skalenverwendungsheterogenität Zudem wird Skalenverwendungsheterogenität adressiert. Grundlage ist das Modell von ROSSI et al. (2001), jedoch werden unkorrelierte Prior-Verteilung für Mittelwert und Varianz angenommen.

A-Priori-Annahmen Die Prior-Verteilungen werden schwach informativ gewählt - die genaue Spezifikation befindet sich im Anhang. Modelle ohne Skalenverwendungsheteroge-

nität wurden nicht geschätzt, da in der Literatur inzwischen Konsens über die Notwendigkeit herrscht (siehe Abschnitt 2.2).

Schätzung mit MCMC Zur Schätzung werden zunächst 5000 Iterationen bis zum „burn-in“ durchgeführt, um Konvergenz zu erreichen, obwohl bereits nach weniger als 1000 Iterationen auf Zeitreihenplots keine Instationarität der gezogenen Parameter mehr erkennbar ist. Anschließend werden 5000 Züge generiert, die zur Approximation der Posteriori-Verteilungen der Parameter herangezogen werden.³⁰

Zusammenfassung der Modelle

Modell 1: Treibermodell, alle Attribute werden als potenzielle Treiber der Gesamtzufriedenheit modelliert

Modell 2: Mixture-Modell mit folgenden Komponenten: Treibermodell und Halo-Modell

Modell 3: Mixture-Modell mit folgenden Komponenten: Treibermodell und Halo-Modell, MIMIC-Modell Variante 1

Modell 4: Mixture-Modell mit folgenden Komponenten: Treibermodell und Halo-Modell, MIMIC-Modell Variante 2

Modell 5: Mixture-Modell mit folgenden Komponenten: Treibermodell und Halo-Modell, MIMIC-Modell Variante 3

5.4. Schätzergebnisse

Anpassung Der Modellfit sowie die Anteile der geschätzten Komponenten der unterschiedlichen Modelle sind in Tabelle 5.5 dargestellt.

Erwartungsgemäß weist das Einkomponentenmodell (*Modell 1*) die schlechteste Anpassung an die Daten (geringste marginale Likelihood) auf. Die Berücksichtigung der haloistisch geprägten Antworten in *Modell 2* verbessert die marginale Likelihood deutlich.

³⁰Die Durchführung von 10.000 MCMC-Iterationen benötigt ca. 3-4 Tage auf einem Computer mit Intel Core i7-Prozessor.

Tabelle 5.5: Modellvergleich

Modell	1	2	3	4	5
log Marginal Likelihood	-26204.65	-25348.15	-24864.16	-24935.78	-24988.58
Anteil MIMIC			36%	16%	14%
Anteil Halo		54%	24%	45%	46%
Anteil Regression	100%	46%	40%	39%	39%

Dabei werden 54% der Antworten als haloistisch identifiziert. Die zusätzliche Einführung insbesondere der MIMIC-Komponente in Variante 1 (*Modell 3*) führt zu einer weiteren deutlichen Verbesserung der Anpassung an die Daten (die marginale Likelihood steigt um 484 gegenüber Modell 2). Der Anteil der rein formativ geprägten Daten geht dabei auf 40% zurück. 94% der in Modell 3 als „formativ“ identifizierten Fälle werden auch in Modell 2 als „formativ“ identifiziert.³¹ 36% der Fälle werden in Modell 3 der MIMIC-Komponente zugeordnet, die meisten davon wurden in Modell 2 als „Halo“ identifiziert.

Dies lässt sich auch anhand der Korrelationsmatrizen in Abbildung 5.6 bzw. Abbildung 5.4 (oben) nachvollziehen: Die MIMIC-Komponente in Modell 3 ähnelt der Halo-Komponente in Modell 2.

Insgesamt lässt sich feststellen, dass nur wenige (40%) Kunden tatsächlich alle Dimensionen unabhängig voneinander bewerten. Dies hat auch Auswirkungen auf die im Regressionsmodell identifizierten Treiber von Kundenzufriedenheit.

Auswirkungen auf das Regressionsmodell In Tabelle 5.6 wird deutlich, welchen Einfluss die Berücksichtigung des Halo-Fehlers auf die Identifikation von Treibern hat.³² Nur die Fälle, die der Regressionskomponente zugeordnet werden, liefern Informationen für die Treiberanalyse.

In Modell 1 - ohne Berücksichtigung der unterschiedlichen Kausalstrukturen - werden

³¹Zuordnung auf Basis der Posteriori-Mittelwerte der Zuordnungswahrscheinlichkeiten.

³²An dieser Stelle werden nur die Pfadkoeffizienten dargestellt, da bei Treiberanalysen sich das Interesse vor allem auf die Gewichtung der Treiber richtet. Statistiken der anderen Parameter finden sich im Anhang.

Tabelle 5.6: Treiber in der Regressionskomponente

	Strukturkoeffizienten der jeweiligen Regressionskomponente											
	Modell 1			Modell 2			Modell 3			Modell 4		
	PMW	2.5%	97.5%	PMW	2.5%	97.5%	PMW	2.5%	97.5%	PMW	2.5%	97.5%
Geschmack	0.40	0.29	0.51	0.51	0.34	0.69	0.63	0.38	0.83	0.57	0.36	0.83
Gesundheit	0.16	0.06	0.26	0.26	0.08	0.46	0.35	0.08	0.56	0.32	0.07	0.57
Lieferservice	0.27	0.15	0.40	0.20	0.03	0.37	0.26	0.01	0.49	0.32	0.06	0.63
Umfeld	0.15	0.04	0.27	0.06	-0.08	0.19	-0.03	-0.23	0.15	0.00	-0.21	0.20
Ruf	0.10	0.03	0.16	0.19	0.08	0.30	0.31	0.11	0.44	0.37	0.15	0.56
Menge	0.05	-0.02	0.12	0.03	-0.09	0.14	0.00	-0.13	0.14	-0.01	-0.14	0.13
Preis	0.03	-0.01	0.07	0.10	0.02	0.17	0.17	0.05	0.25	0.13	0.06	0.26
Wert	0.03	-0.06	0.11	-0.03	-0.16	0.10	-0.04	-0.19	0.11	-0.02	-0.18	0.12
Portionen	0.01	-0.04	0.06	-0.01	-0.09	0.07	-0.06	-0.15	0.06	0.00	-0.13	0.11
Auswahl Belag	0.00	-0.05	0.05	0.10	0.00	0.20	0.18	0.02	0.30	0.13	-0.02	0.32
Verp. Des.	-0.05	-0.11	0.02	0.02	-0.08	0.11	0.09	-0.06	0.20	0.12	-0.04	0.24
Auswahl Teig	0.02	-0.03	0.07	0.00	-0.07	0.07	-0.01	-0.10	0.07	-0.01	-0.10	0.08
Auswahl Sonst.	-0.02	-0.08	0.04	0.03	-0.06	0.14	0.04	-0.07	0.16	0.05	-0.08	0.19
Wartezeit	-0.01	-0.09	0.06	0.10	-0.01	0.22	0.07	-0.09	0.23	0.05	-0.13	0.23
Erreichbarkeit	0.08	0.03	0.13	0.05	-0.01	0.12	0.03	-0.04	0.11	0.04	-0.06	0.13

fettgedruckte Koeffizienten: Posteriori-Dichte von 0 verschieden (95%)

folgende Treiber identifiziert: *Geschmack*, *Lieferservice*, *Gesundheit*, *Umfeld*, *Ruf*, *Erreichbarkeit*

In der Regressionskomponente von Modell 2 ändern sich die Ergebnisse. Nun werden *Umfeld* und *Erreichbarkeit* nicht mehr als „signifikanter“³³ Treiber identifiziert. Stattdessen sind nun die Aspekte *Preis* und *Auswahl Belag* relevante Treiber. Zudem ändert sich die Reihenfolge der Gewichtungen: Während *Geschmack* offenbar der wichtigste Treiber bleibt, bekommt nun der Punkt *Gesundheit* eine stärkere Bedeutung.

In Modell 3 ändern sich die Ergebnisse der Regressionskomponente nur wenig, denn die Fälle der Regressionskomponenten in Modell 2 und 3 überdecken sich größtenteils. Die Gewichtung des Attributs *Ruf* scheint jedoch stärker zu sein, wenn man die Regressionskomponente von Modell 3 betrachtet. Die Identifikation der relevanten Treiber ist jedoch ansonsten identisch.

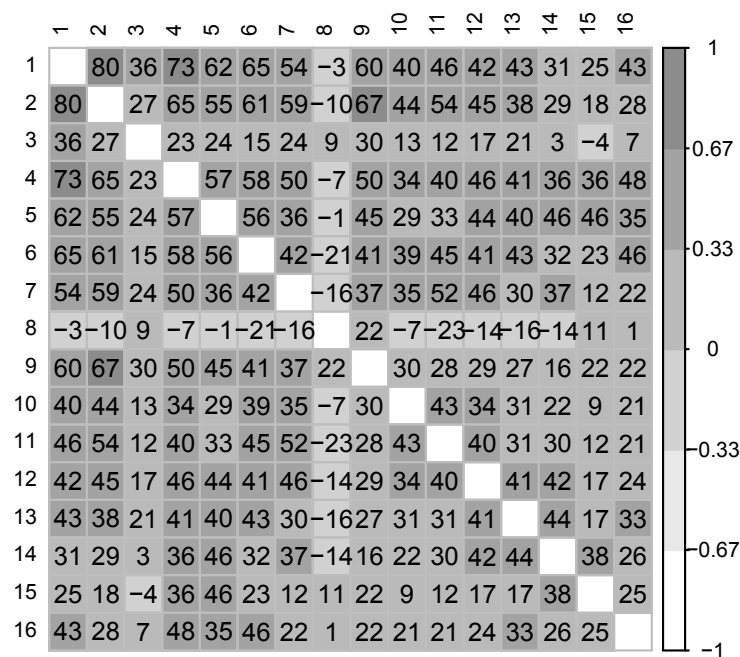
Ergebnisse des MIMIC-Modells Auch aus der MIMIC-Komponente lassen sich Informationen zu relevanten Zufriedenheitstreibern gewinnen, obgleich die Dimensionalität des Treibermodells geringer ist. In Tabelle 5.7 sind die MIMIC-Pfadkoeffizienten für Modell 3 dargestellt. Erkennbare Treiber des Gesamteindrucks sind Verpackungsdesign, Teigauswahl, Wartezeit und Erreichbarkeit. Die Auswahl von Belägen und sonstigen Produkten können jedoch nicht als eindeutige Treiber des Gesamteindrucks identifiziert werden (dem 95%-Glaubwürdigkeitsintervall der Parameterschätzer folgend). Der Gesamteindruck wiederum treibt die Bewertung aller übrigen Zufriedenheitsdimensionen.

Ergebnisse des Halo-Modells Die Modellierung eines Halo-Modells ergibt nur soweit Sinn, wie sich auch von Null verschiedene Pfadkoeffizienten ergeben. Der Halo-Effekt definiert sich schließlich dadurch, dass sich alle Bewertungen durch einen latenten Faktor erklären lassen. In der Konsequenz ergibt sich eine Korrelationsstruktur mit zahlreichen starken Korrelationskoeffizienten. Betrachtet man die Korrelationsmatrizen der Modelle 2 und 3 (Abbildungen 5.4 bzw. 5.5), so lässt sich feststellen, dass die Hinzunahme der

³³Der Begriff „signifikant“ gehört zur klassischen Statistik. An dieser Stelle soll „signifikant“ bedeuten, dass sich ein großer Teil der Fläche unter der Posteriori-Verteilung entweder im positiven oder negativen Bereich befindet. Dazu werden das 2,5%- und 97,5%-Quantil herangezogen.

MIMIC-Komponente im Mixture-Modell dazu führt, dass die Korrelationen in der Halo-Komponente deutlich abnehmen.³⁴ In Modell 3 wird die Halo-Komponente durch Fälle dominiert, bei denen es keinen Zusammenhang zwischen den latenten Faktoren mehr gibt. Streng genommen handelt es sich also nicht mehr um ein Halo-Modell.

Korrelationen In Abbildung 5.1 ist die Korrelation der Konstrukte in Modell 1 abgebildet. Hier ist erkennbar, dass insbesondere die ersten sieben Konstrukte relativ stark miteinander korrelieren. Die übrigen Faktoren korrelieren teilweise auch erkennbar miteinander, jedoch auf einem geringen Niveau. Da in Modell 1 nur die Regressionskomponente berücksichtigt wird, könnten die Korrelationen durch die Überlagerung von Fällen entstehen, die eher einem Halo oder MIMIC-Modell entsprechen. Die unterschiedlichen



Mixture-Modelle (Modelle 2-5) zerlegen die Korrelationsstruktur in mehrere Komponenten. Die Dekomposition der Korrelationsstruktur in zwei Komponenten durch Modell 2 ist in Abbildung 5.4 ersichtlich. Die Korrelationen innerhalb der Regressionskomponente sind nun geringer, während in der Halo-Komponente alle Konstrukte erkennbar miteinander

³⁴Die zugehörigen Pfadkoeffizienten sind im Anhang dargestellt.

korrelieren. Die resultierenden Korrelationsmatrizen für Modell 3 sind in den Abbildungen 5.5 und 5.6, die für Modell 4 in den Abbildungen 5.7 und 5.7 dargestellt.

Allokation Anhand von Abbildung 5.2 lässt sich die „Stabilität“ der Allokation in Modell 1 darstellen. Dabei wurde wie folgt verfahren:

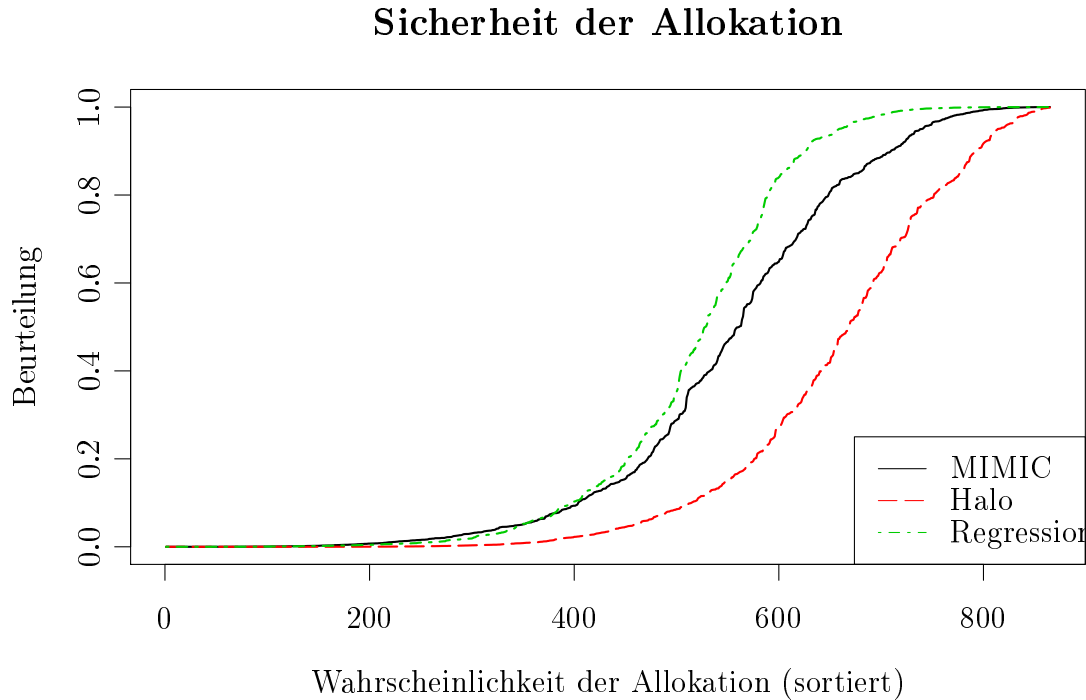
- Für die 865 Antworten wurde die Wahrscheinlichkeiten der Allokation zu jeder der drei Komponenten berechnet.
- Die Wahrscheinlichkeiten wurde für jede der drei Komponenten sortiert.
- Die drei sortierten Vektoren mit Wahrscheinlichkeiten der 865 Antworten sind in Abbildung 5.2 als Linien dargestellt.

Je steiler der Übergang der Kurven zwischen den Wahrscheinlichkeiten 0 und 1 verläuft, desto „stabiler“ ist die Allokation. Für 109 Antworten ist die Wahrscheinlichkeit der Allokation zur MIMIC-Komponente 80% oder größer. Zur Halo-Komponente werden mit 80%-iger Wahrscheinlichkeit 223 Antworten gezählt, zur Regressionskomponente sogar 251 Antworten. Akzeptiert man also eine Zuordnungswahrscheinlichkeit von 80% als relativ sicher, so lässt sich die Mehrheit der Antworten relativ „stabil“ einer Komponente zuordnen. Die Daten scheinen also die Theorie zu unterstützen, dass es neben den extremen Komponenten (Halo und Regression) auch einen „Zwischentyp“ gibt: „partielles“ Haloing dargestellt durch eine MIMIC-Komponente. Dieser Schluss ist konsistent mit der deutlichen Verbesserung der marginalen Likelihood (Tabelle 5.5)

Spezifität der kausalen Heterogenität Skalenverwendungsheterogenität stellt eine *kundenspezifische* Eigenschaft dar (siehe Abschnitt 2.2). Das subjektive Verständnis von einer Skala hängt - der Definition nach - nicht davon ab, welches Produkt oder welcher Service bewertet wird. So spiegelt das Modell der Skalenverwendungsheterogenität die personenbezogene Tendenzen zum Ja-/Nein-Sagen sowie zur Streuung/Mittelwertorientierung der Antworten wider.

Die kausale Struktur ist jedoch *kunden- und produktspezifisch*. Der Erinnerungsprozess und die anschließende Bewertung und ggf. Gewichtung der Attribute können von Bewertungsobjekt zu Bewertungsobjekt variieren. Zentrales Motiv für Haloing ist die

Abbildung 5.2: Zuordnungswahrscheinlichkeiten in Modell 3



Vereinfachung des Bewertungsprozesses (siehe Abschnitt 2.3.2). Je weniger Informationen bezüglich eines Bewertungsobjektes zur Verfügung stehen, desto schwieriger wird es für den Kunden, die Attribute differenziert zu bewerten. Somit steigt auch die Wahrscheinlichkeit, haloistisch zu bewerten. Neben dem Grad der Vertrautheit mit dem Bewertungsobjekt sind auch weitere Ursachen für objektspezifisches Haloing denkbar (u. a. kritische Ereignisse). Ist das alleinige Ziel eine Treiberanalyse, sind die Ursachen für Haloing zunächst einmal irrelevant. Jedoch muss der Möglichkeit Rechnung getragen werden, dass kausale Strukturen $Kunde \times Objekt$ -spezifisch sind.

Inwieweit diese Annahme empirisch zutrifft, lässt sich anhand von Abbildung 5.3 feststellen. Dargestellt ist die Verteilung der Posteriori-Mittelwerte der Summe der Diagonalmatrixwahrscheinlichkeiten über die 253 Kunden für die jeweils erste und letzte Bewertung.³⁵ 95 der 253 Kunden (38%) bewerten mit einer Wahrscheinlichkeit von 50% oder mehr in

³⁵Da die Kunden unterschiedlich viele Bewertungen abgegeben haben (siehe Tabelle 5.2), wurde jeweils nur die erste und letzte Bewertung betrachtet, sofern mindestens 2 Restaurants bewertet worden sind. 30 Kunden haben nur eine Bewertung abgegeben und wurden daher nicht berücksichtigt.

konsistenter Weise.

Eine Kreuztabelle der a posteriori wahrscheinlichsten Klassifizierung ließe zwar sehr schnell erkennen, inwieweit konsistent ein Strukturmodell verwendet wurde, jedoch geht dabei die Unsicherheit der Klassifizierung verloren. Sind die Zuordnungswahrscheinlichkeiten 34%,32%,32% für Antwort 1 und 32%, 32%, 34% für Antwort 2, so würde die große Unsicherheit über die Zuordnung missachtet. Multipliziert man hingegen die Zuordnungswahrscheinlichkeiten $p(s_i)^t * p(s_i)$, so gibt die Spur Auskunft über die Wahrscheinlichkeit, dass ein und dasselbe Strukturmodell hinter den betrachteten Antworten steht. Das Histogramm der Spur zeigt, dass nur bei wenigen Personen je zwei Bewertungen zum selben Strukturmodell zugeordnet werden. Folgendes Beispiel einer sicher nicht-konsistenten Zuordnung illustriert diese Vorgehensweise:

$$\begin{bmatrix} 0.01 & 0.01 & 0.01 \end{bmatrix} \times \begin{bmatrix} 0.98 \\ 0.01 \\ 0.01 \end{bmatrix} = \begin{bmatrix} 0.0098 & 0.0001 & 0.0001 \\ 0.0098 & 0.0001 & 0.0001 \\ 0.9604 & 0.0098 & 0.0098 \end{bmatrix}$$

Je größer die Spur der Matrix, desto wahrscheinlicher ist eine konsistente Klassifikation.

Abbildung 5.3: Histogramm der Summe der Diagonalwahrscheinlichkeiten

Histogramm der Summe der Diagonalwahrscheinlichkeiten

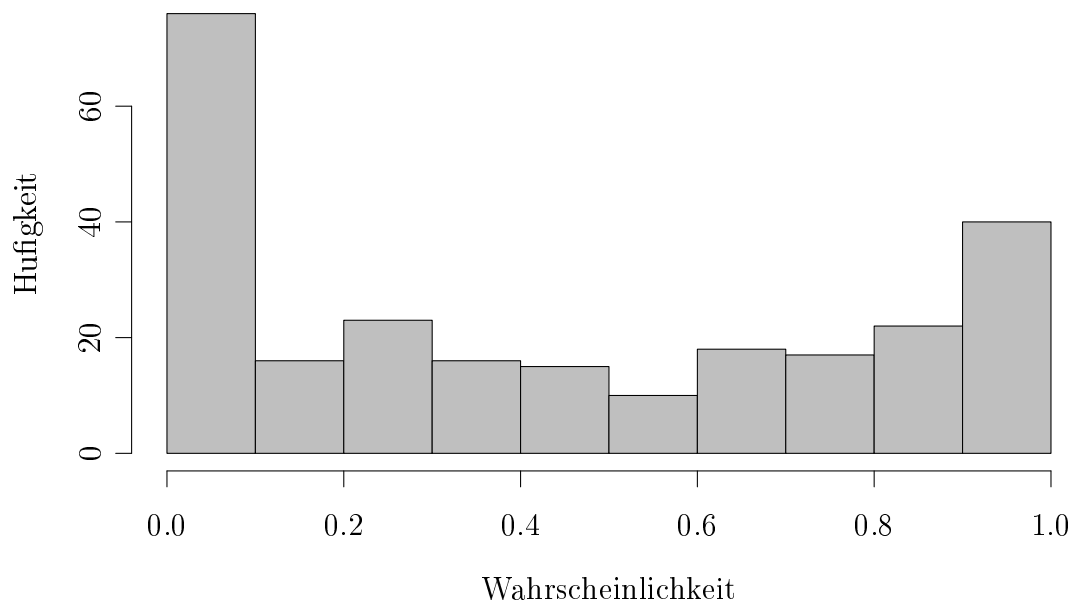


Abbildung 5.4: Korrelationen Modell 2

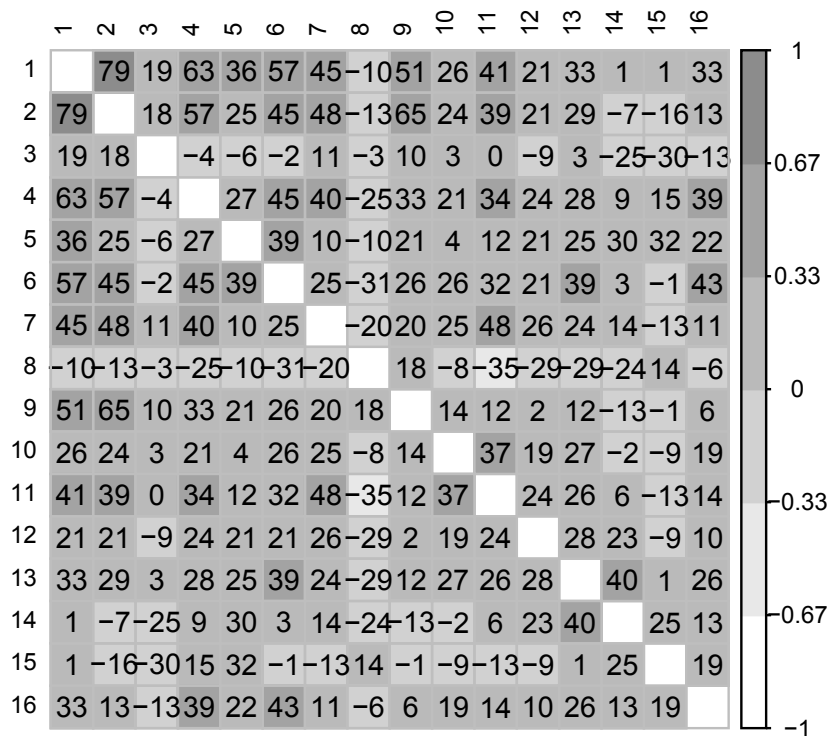
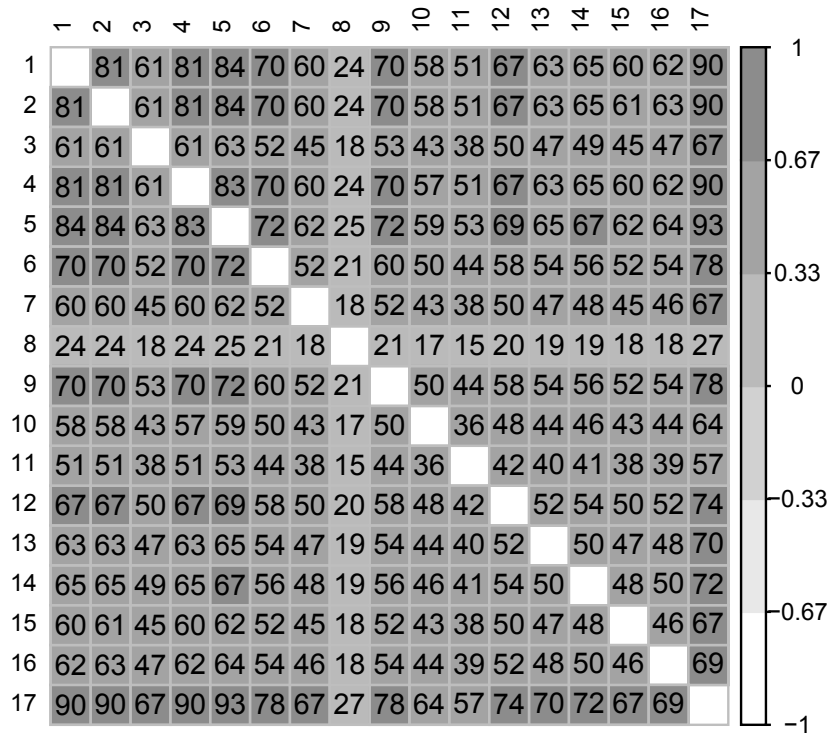


Abbildung 5.5: Korrelationen Modell 3 - Halo- und Regressionskomponente

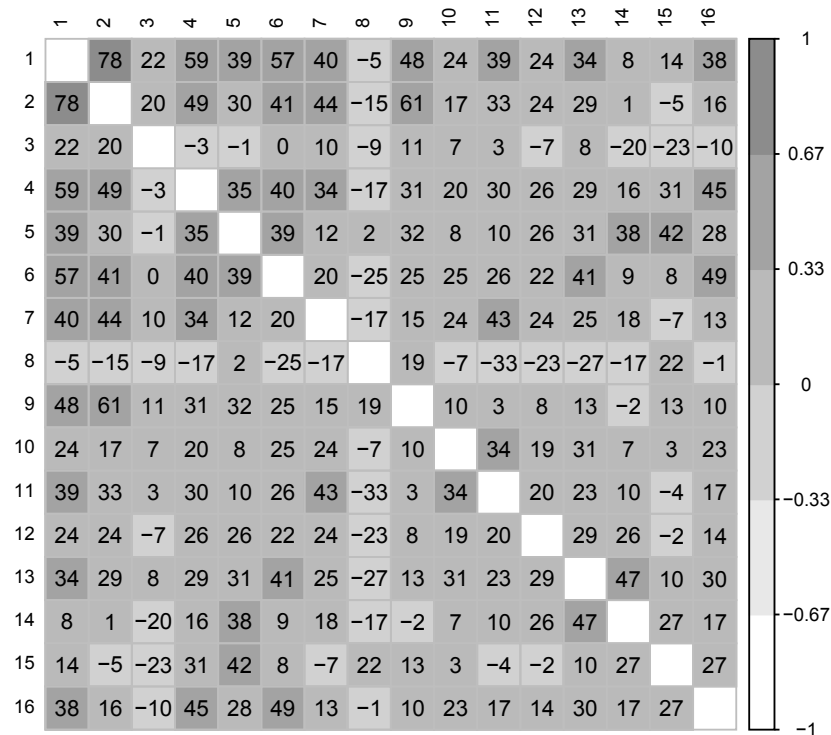
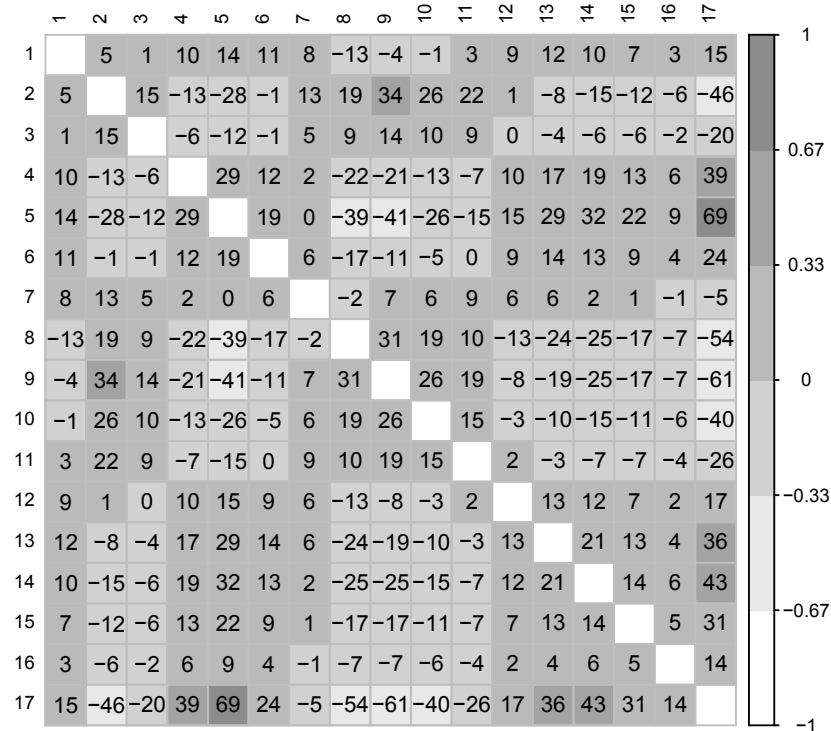


Abbildung 5.6: Korrelationen Modell 3 - MIMIC-Komponente

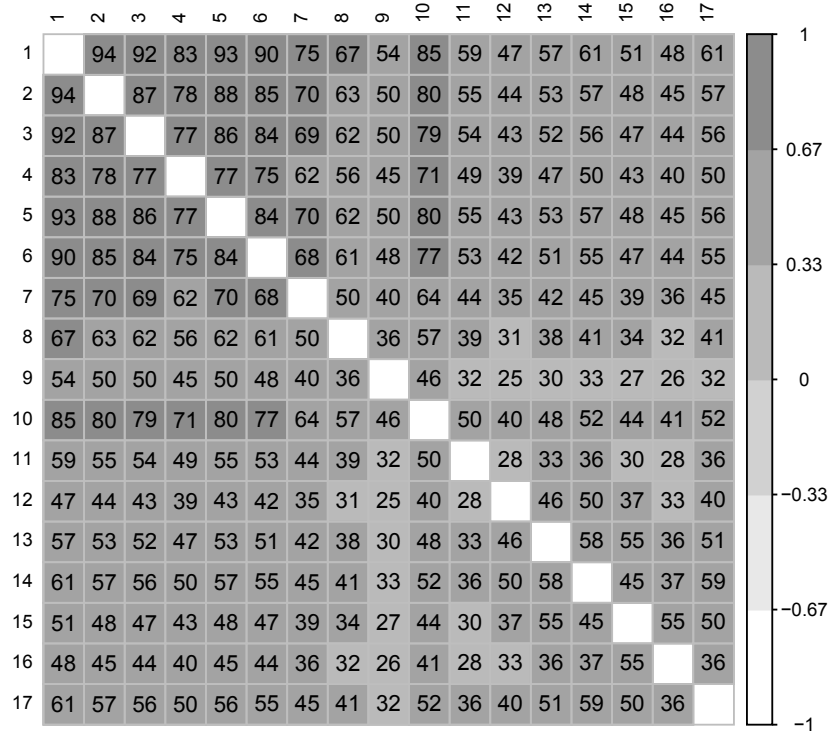


Abbildung 5.7: Korrelationen Modell 4 - MIMIC-Komponente

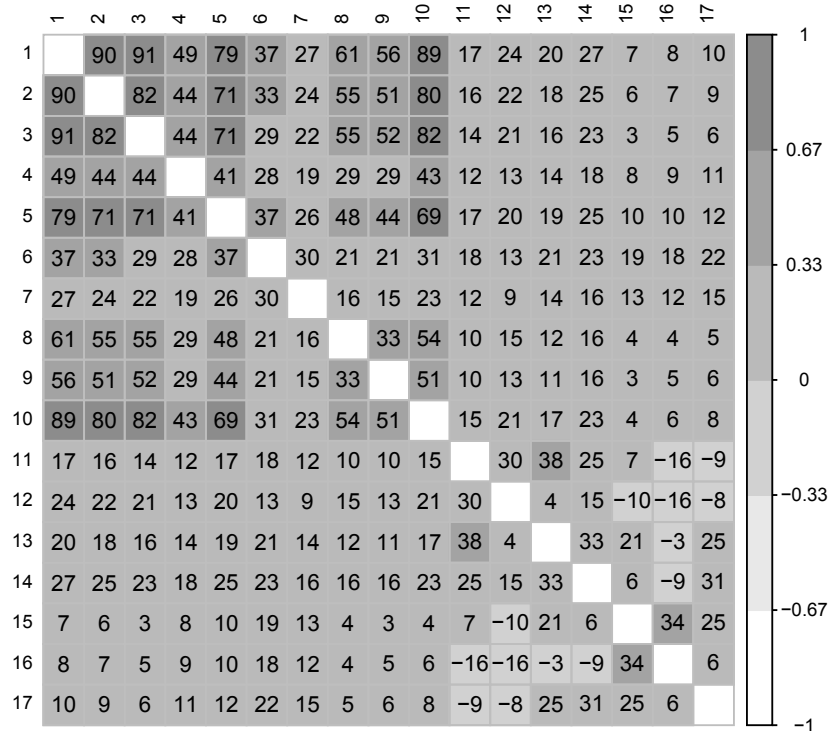
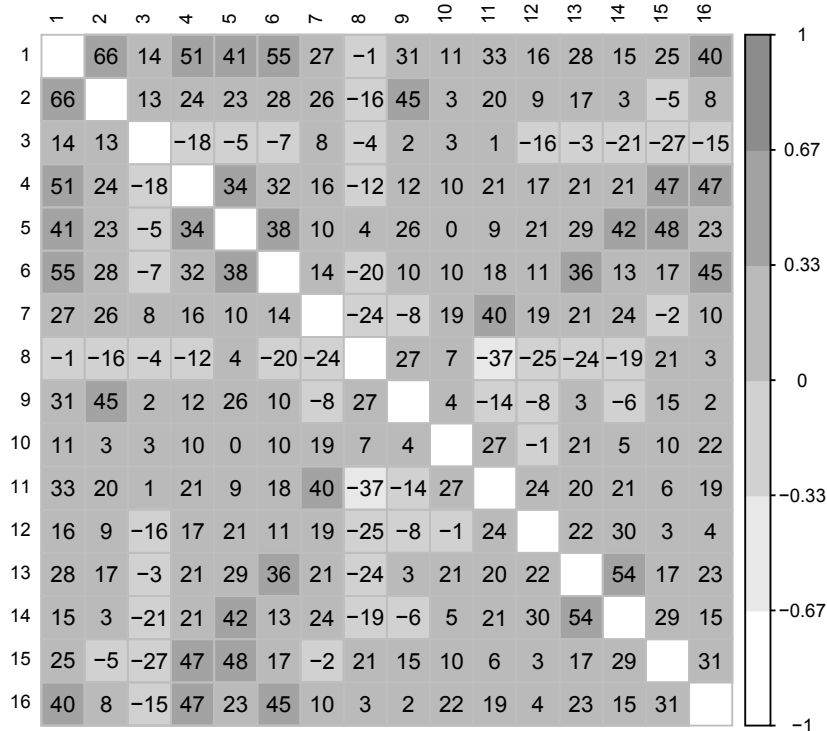
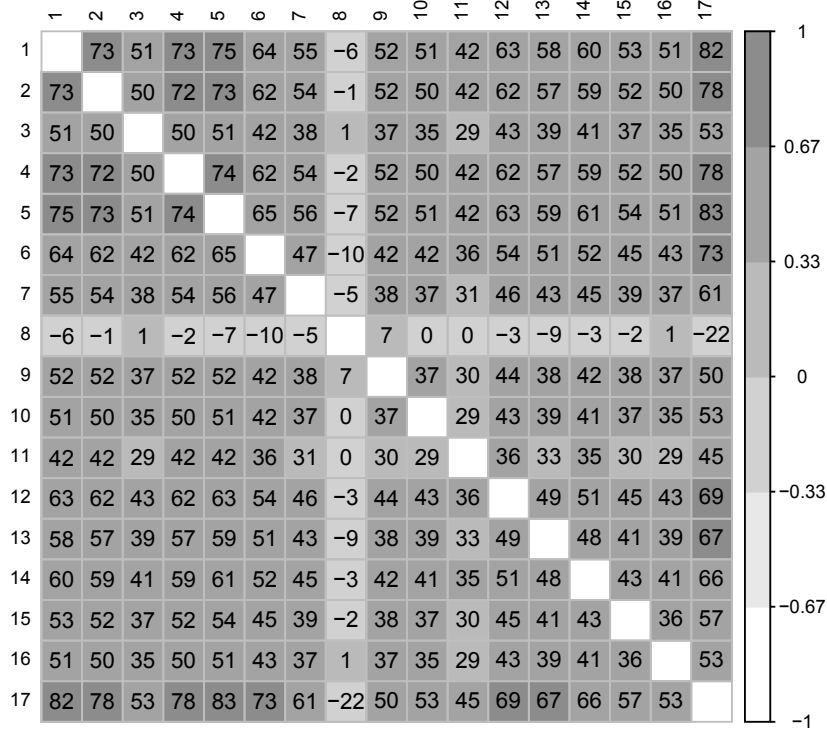


Tabelle 5.7: Ergebnisse Modell 3 - MIMIC-Koeffizienten

	PM	2.5%	97.5%
Pfadkoeffizienten			
Gesamt \leftarrow GE	1.00	1.00	1.00
Geschmack \leftarrow GE	0.86	0.66	0.97
Gesundheit \leftarrow GE	0.63	0.48	0.77
Lieferservice \leftarrow GE	0.80	0.69	0.91
Umfeld \leftarrow GE	0.70	0.59	0.84
Ruf \leftarrow GE	0.67	0.56	0.81
Menge \leftarrow GE	0.59	0.48	0.71
Preis \leftarrow GE	0.58	0.40	0.73
Wert \leftarrow GE	0.86	0.69	0.96
Portionen \leftarrow GE	0.49	0.37	0.65
GE \leftarrow Ausw. Belag	0.10	-0.04	0.24
GE \leftarrow Verp. Des.	0.19	0.02	0.34
GE \leftarrow Ausw. Teig	0.25	0.03	0.38
GE \leftarrow Ausw. Sonst.	0.03	-0.09	0.22
GE \leftarrow Wartezeit	0.22	0.08	0.37
GE \leftarrow Erreichb.	0.28	0.13	0.50

Abbildung 5.8: Korrelationen Modell 4



6. Diskussion der Ergebnisse

In Abschnitt 6.1 wird zunächst der Beitrag dieser Arbeit zusammengefasst. Die empirischen Ergebnisse werden in Abschnitt 6.2 diskutiert. In Abschnitt 6.3 werden anschließend die Implikationen für die Untersuchung von Kundenzufriedenheit reflektiert. Abschließend werden Einschränkungen der Arbeit aufgezeigt (Abschnitt 6.4) sowie Anknüpfungspunkte für weitere Forschung auf dem Gebiet (Abschnitt 6.5) benannt.

6.1. Zum Beitrag dieser Arbeit

Das durch drei Forschungsfragen konkretisierte Forschungsziel dieser Arbeit (siehe Abschnitt 2.4.2) konnte durch die Entwicklung des kausalen Mixture-Modells wie folgt erreicht werden:

- Die Berücksichtigung inhaltlicher Domänen (Frage 1) wird durch den Rückgriff auf LISREL-Modellierung erreicht.
- Darüber hinaus ermöglichen die LISREL-Strukturen in den Mixture-Komponenten sehr flexible Kausalstrukturen, mit denen eine MIMIC-Struktur modelliert werden kann, welche Elemente von Haloing und Forming kombiniert (Frage 2).
- Außerdem ist das eingeführte Modell in der Lage, Skalenverwendungsheterogenität, konkurrierende Kausalmodelle und Domänen gleichzeitig zu berücksichtigen (Frage 3).

Die Idee von finite Mixture-Modellen (Abschnitt 4.5) wurde in dieser Arbeit einen Schritt weitergeführt. Während man in den meisten Modellen davon ausgeht, dass es *eine* kausale Struktur gibt, innerhalb derer sich nur die *Stärke der Zusammenhänge* unterscheiden kann, können hier völlig *unterschiedliche* Kausalstrukturen modelliert werden. Sofern sichergestellt ist, dass diese Strukturen tatsächlich unterschiedlich (also nicht äquivalent) sind (siehe Abschnitt 4.3.4), müssen keine weiteren Identifikationsrestriktionen eingeführt werden.

Der Beitrag dieser Arbeit liegt insbesondere in der Kombination unterschiedlicher Modellierungstechniken: LISREL, Skalenheterogenitätsmodell und Mixture-Modellierung. Die

Zusammenführung in einem Modell ermöglicht es nun, bei zahlreichen Fragestellungen in Sozial- und Wirtschaftswissenschaften konkurrierende Theorien zu überprüfen und gleichzeitig typische, bei Befragungen auftretende Verzerrungen zu kontrollieren und deren Einfluss auf die Ergebnisse zu verringern. Es konnte empirisch nachgewiesen werden, dass Modelle, welche kausale Heterogenität berücksichtigen, sich besser an die Daten anpassen.

Die gängige Praxis der Strukturgleichungsmodellierung läuft dem gegenüber Gefahr, Hypothesen zu vorschnell zu bestätigen bzw. zu verwerfen. Dafür gibt es zwei Gründe:

- Skalenverwendungsheterogenität, welche Korrelationen potenziell inflationiert, kann zur Bestätigung zahlreicher Zusammenhänge führen, die gar nicht existieren (siehe auch HARDT und BUESCHKEN, 2010). Gleichzeitig können Multikollinearitätsprobleme bei Regressionsanalysen auftreten, sodass Regressionskoeffizienten verzerrt werden.
- Kausale Heterogenität kann ebenfalls zu einer falschen Einschätzung von Zusammenhängen führen, da sich unterschiedliche Kausalstrukturen überlagern.

Beide Probleme lassen sich mit dem in dieser Arbeit entwickelten Modell lösen. Dadurch ergibt sich auch ein Auftrag an Anwender von SEMs: Es genügt nicht mehr die alleinige theoretische Begründung *eines* Kausalmodells. Es sollte zukünftig stets überprüft werden, ob auch *konkurrierende* Theorien existieren oder sich sachlogisch herleiten lassen.

Die Bedeutung dieser Aufgabe wurde am Beispiel Kundenzufriedenheit deutlich demonstriert: Wer nur *ein* Treibermodell untersucht, bemerkt nicht, dass ein großer Teil der untersuchten Fälle diesem Modell gar nicht entsprechen. Die konkurrierenden Theorien (Treibermodell und Halo) sind jedoch bereits lange bekannt und erscheinen bei Zufriedenheits- und Einstellungsuntersuchungen plausibel. Auch das für multiattributive Modelle neu eingeführte MIMIC-Modell lässt sich mittels des Kriteriums der unterschiedlichen „Bewertbarkeit“ von Attributen begründen.

Das hier entwickelte Modell wurde in der verbreiteten Programmiersprache R implementiert und lässt sich daher nun auch praktisch auf sehr unterschiedliche Fragestellungen anwenden.

6.2. Diskussion der empirischen Ergebnisse

Es konnte nachgewiesen werden, dass die Einführung einer MIMIC-Komponente zu einer erkennbaren Verbesserung der Anpassung an die Daten führt (Tabelle 5.5). Schlüsse über relevante Zufriedenheitstreiber ändern sich sowohl durch Berücksichtigung von haloistisch und teilweise haloistisch geprägten (MIMIC-) Antworten (Tabelle 5.6). Folglich hat die Modellierung alternativer Kausalstrukturen praktische Relevanz: Zwingt man die „falschen“ Fälle einem Treibermodell auf, so werden nicht die richtigen Treiber identifiziert. Dies kann zu ineffizienter Allokation von Ressourcen führen: Die Auswahl der Pizzabeläge würde nicht als relevanter Zufriedenheitstreiber erkannt. Darüber hinaus würde möglicherweise zu viel Geld in Verbesserung des Umfelds investiert, obwohl keine Verbesserung der Zufriedenheit dadurch zu erwarten ist.

Die Idee des Mixture-Modells lässt sich noch einmal anhand der Dekomposition der Korrelationsmatrizen erläutern. Ohne Mixture-Modellierung überlagern sich die Beobachtungen, welche eigentlich aus unterschiedlichen Datengenerierungsprozessen stammen (siehe Abbildung 5.1).

Durch die Berücksichtigung von Halo- und Regressionskomponenten *in Modell 2* werden unterschiedliche Korrelationsstrukturen sichtbar (siehe Abbildung 5.4).³⁶

Innerhalb der Regressionskomponente weisen nun die erklärenden Variablen eine deutlich geringere Korrelationen untereinander auf - die Attribute werden unabhängig voneinander bewertet. Nur diejenigen Kunden, welche im Sinne des Treibermodells denken und antworten, werden dabei zur Schätzung des Treibermodells herangezogen. Somit ändern sich auch die Ergebnisse bezüglich der relevanten Treiber.

Die identifizierten Korrelationsstrukturen verändern sich durch die Einführung der MIMIC-Komponente in *Modell 3*. Wie Abbildung 5.5 zu entnehmen ist, hat dies insbesondere Konsequenzen für die Halo-Komponente. Sie ist in diesem Modell um etwa 30%-Punkte kleiner. Die Korrelationen sind gering. Nach Korrektur um Skalenheterogenität wird ersichtlich, dass bei den Fällen dieser Komponente keine Zusammenhänge mehr existieren. Zu einem ähnlichen Ergebnis kommen auch BÜSCHKEN et al. (2010, S. 20) bei der Untersuchung eines Datensatzes. Die Komponente entspricht damit nicht mehr der ei-

³⁶Die Übersicht der unterschiedlichen Modelle befindet sich in Abschnitt 5.3.

gentlichen Idee eines Halo-Modells, vielmehr beschreibt sie „informationsfreie“ Antworten. Die MIMIC-Komponente übernimmt allerdings einen großen Teil der Fälle. Die Korrelationen dieser Komponente (Abbildung 5.6) bewegen sich allesamt auf einem sehr hohen Niveau. Die Korrelation zwischen den endogenen Größen ist jedoch noch einmal deutlich größer - diese sind offenbar am schwierigsten differenziert zu bewerten. Eine mögliche Erklärung dafür ist die Art der Konstrukte: Insbesondere die ersten sieben latenten Faktoren können nach Rückgriff auf implizites Wissen bewertet werden. Die Faktoren *Geschmack*, *Gesundheit*, *Umfeld*, *Ruf* sind zudem abstrakter als beispielsweise *Erreichbarkeit* oder *Wartezeit*. Genauen Aufschluss über die Hintergründe kann zwar nur eine neurologische Untersuchung geben, jedoch erscheint diese Unterscheidung plausibel. Je konkreter und genauer die Fragestellung ist, desto einfacher scheint auch eine unabhängige Bewertung zu sein.

6.3. Implikationen für Kundenzufriedenheit

Nahezu alle bisherigen Publikationen (siehe Abschnitt 2.3.4) bestätigen die Existenz des Halo-Effektes bei Zufriedenheits- oder Einstellungsuntersuchungen. Es gibt daher bereits a priori Grund zu der Annahme, dass zumindest eine zum Treibermodell konkurrierende Theorie durch die Daten bestätigt werden könnte. Dies bestätigt sich in der vorliegenden Arbeit. Der Halo-Effekt konnte auch bei der Bewertung der Pizza-Restaurants nachgewiesen werden. Allerdings scheint es Unterschiede bezüglich der Reichweite des Halo zu geben. Eine Mehrheit der Kunden ist zumindest in der Lage, einen Teil der unterschiedlichen Aspekte differenziert (d. h. unabhängig) zu bewerten. Eine ausschließliche Fokussierung auf Treiberinformationen von den Kunden, welche *alle* Dimensionen differenziert bewerten, bedeutet möglicherweise den Verlust von Informationen. Durch die Einführung der MIMIC-Komponente kann berücksichtigt werden, dass nur ein Teil der Bewertungen unabhängig voneinander vorgenommen wird. Somit können zusätzlich Erkenntnisse über Treiber gewonnen werden. Der Einfluss auf die Schätzung des Treibermodells war bei der Untersuchung der Pizza-Restaurants gering. In anderen Fällen können sich die Ergebnisse der Regressionskomponente stärker verändern. Daher sind weitere Untersuchungen notwendig, um die Konsequenzen der MIMIC-Modellierung bei multiattributiven Model-

len näher zu studieren. Jede Vernachlässigung von kausaler Heterogenität wäre jedoch - ebenso wie die Vernachlässigung von Skalenverwendungsheterogenität - fahrlässig, da die Identifikation falscher Treiber zu unrentablen Fehlinvestitionen führen kann.

6.4. Einschränkungen

Die größten Schwachstellen der durchgeführten Untersuchungen lassen sich wie folgt zusammenfassen:

- Das verwendete Cut-Point-Modell ist möglicherweise zu restriktiv. Die Ergebnisse in BÜSCHKEN et al. (2010) deuten darauf hin, dass sich Modelle mit heterogenen Cut-Points deutlich besser an die Daten anpassen.
- Die MIMIC-Strukturen werden nur aufgrund begründeter Annahmen erstellt (explizite und implizite Attribute). Diese Annahmen können - wie bei jedem Strukturgleichungsmodell - kritisch diskutiert werden. Es werden in dieser Arbeit nicht alle technisch machbaren Modelle ausprobiert, denn alle betrachteten konkurrierenden Strukturen sollten auch theoretisch begründbar sein. Die Berücksichtigung konkurrierender Modelle zwingt dabei zu einer genauen Auseinandersetzung mit den betrachteten Theorien.
- Die verwendeten Daten entsprechen nicht denen klassischer psychometrischer Messmodelle. Vielmehr wurden inhaltliche Domänen zusammengefasst. Dieses Vorgehen ist nicht unüblich (siehe Abschnitt 2.1.4). Dennoch wäre es interessant, die Heterogenität bei der Zusammensetzung der Domänen zu untersuchen.
- Eine Validierung der Ergebnisse findet nicht statt. Dazu wären neurologische Untersuchung notwendig. Der vorgestellte Ansatz beschränkt sich darauf, bei gegebenen Daten konkurrierende Theorien zu modellieren.
- Strukturgleichungsmodellierung kommt nicht ohne - mehr oder weniger schwer zu begründende - Annahmen aus. So schreiben BARTHOLOMEW et al. (2011, S. 228) „[...] everything is conditional on the model being true [...]“ und stellen den Nutzen latenter Variablenmodelle insgesamt in Frage. Eine gründliche theoretische Fundierung sämtlicher dargestellter Hypothesen ist daher unumgänglich.

6.5. Anregungen für weitere Forschung

In dieser Arbeit wurden die MIMIC-Strukturen auf der Basis plausibler Annahmen erstellt und in das Mixture-Modell aufgenommen. Nicht immer lässt sich jedoch die Zahl der plausiblen Modelle soweit reduzieren, dass alle Modelle geschätzt werden können. Der nächste logische Schritt ist daher die Entwicklung eines *Suchalgorithmus*, welcher die optimale(n) MIMIC-Struktur identifiziert. Somit ließen sich auch automatisch Erkenntnisse über die Dimensionalität des Urteilsprozesses der Befragten gewinnen.

Auch die Adressierung weiterer *Response Styles* erscheint wünschenswert. So erlaubt das Cut-Point-Modell von YING et al. (2006) sehr unterschiedliche individuelle Skalenerständnisse. Aufgrund der Komplexität dieses Modells wurde in dieser Arbeit darauf verzichtet. Dennoch wäre es für die weitere Untersuchung von LISREL-Modellen interessant, welche Verbesserungspotentiale sich durch ein verfeinertes Cut-Point-Modell ergeben. BÜSCHKEN et al. (2010) konnten durch ein derartiges Cut-Point-Modell die Anpassung bedeutend verbessern.

Eine weitaus größere Herausforderung stellt die *Validierung* der Ergebnisse dar. Dazu müssten allerdings neurologische Untersuchungen durchgeführt werden. Nur so ließe sich herausfinden, ob sich Kunden wirklich in ihrem Denkprozess bei der Beantwortung von Zufriedenheitsfragebögen unterscheiden. Dem Verfasser sind allerdings zur Zeit keine Publikationen bekannt, die den Halo-Fehler auf diese Art untersuchen.

Weiteren Forschungsbedarf gibt es auch bei der Untersuchung von Determinanten der unterschiedlichen Antwortmuster. Zum einen sollten Umstände, die die Skalennutzung beeinflussen, untersucht werden. Einen ersten Schritt in diese Richtung geht ISHWARAN (2000), der Kovariate in einem Cut-Point-Modell berücksichtigt. Darüber hinaus können auch Kovariate bei der Allokation zur Halo- bzw. Regressionsstruktur eingebunden werden. Um möglichst sichere Ergebnisse zu erzielen, sollten jedoch Experimente durchgeführt werden, sodass Determinanten eindeutig festgestellt werden können. Für Praktiker ist dabei besonders relevant, inwieweit der Halo-Effekt reduziert werden kann, um möglichst aussagekräftige Treiberanalysen durchführen zu können.

Auch eine Übertragung der Idee alternativer Kausalmodelle auf andere Disziplinen wäre interessant. Das vorgestellte Modell basiert auf dem LISREL-Modell und lässt sich somit

auf eine ganze Reihe weiterer Fragestellungen anwenden.

Literatur

- ABRAHAMS, JEFFREY (2007). *101 mission statements from top companies: Plus guidelines for writing your own mission statement*. Ten Speed Press, Berkeley.
- ALBA, JOSEPH W., W. J. HUTCHINSON und J. G. LYNCH (1991). *Memory and Decision Making*. In: ROBERTSON, THOMAS S. und H. H. KASSARJIAN, Hrsg.: *Handbook of consumer behavior*. Prentice-Hall, Englewood Cliffs, NJ.
- ALDRICH, JOHN (2002). *How Likelihood and Identification Went Bayesian*. International Statistical Review / Revue Internationale de Statistique, 70(1):79–98.
- ALLEN, DEREK R. und T. R. RAO (2000). *Analysis of customer satisfaction data: A comprehensive guide to multivariate statistical analysis in customer satisfaction, loyalty, and service quality research*. ASQ Quality Press, Milwaukee, WI.
- ALLENBY, GREG M, D. G. BAKKEN und P. E. ROSSI (2004). *The HB revolution*. Marketing Research, 16(2):20–25.
- ALTARRIBA, JEANETTE, L. M. BAUER und C. BENVENUTO (1999). *Concreteness, context availability, and imageability ratings and word associations for abstract, concrete, and emotion words*. Behavior Research Methods, Instruments, & Computers, 31(4):578–602.
- ALWIN, DUANE F. und J. A. KROSNICK (1991). *The Reliability of Survey Attitude Measurement: The Influence of Question and Respondent Attributes*. Sociological Methods & Research, 20(1):139–181.
- ANDERSON, EUGENE W., C. FORNELL und D. R. LEHMANN (1994). *Customer Satisfaction, Market Share, and Profitability: Findings from Sweden*. The Journal of Marketing, 58(3):53–66.
- ANDERSON, EUGENE W., C. FORNELL und S. K. MAZVANCHERYL (2004). *Customer Satisfaction and Shareholder Value*. The Journal of Marketing, 68(4):172–185.

- ANDERSON, SHANNON, L. K. PEARO und S. K. WIDENER (2008). *Drivers of Service Satisfaction: Linking Customer Satisfaction to the Service Concept and Customer Characteristics*. *Journal of Service Research*, 10(4):365–381.
- ANDREASEN, ALAN R. (1982). *Verbraucherpolitik als ein Beurteilungsmaßstab für die unternehmerische Marktleistung*. In: HANSEN, URSULA, B. STAUSS und M. RIEMER, Hrsg.: *Marketing und Verbraucherpolitik*. C.E. Poeschel, Stuttgart.
- ANDREASSEN, TOR, B. LORENTZEN und U. OLSSON (2006). *The Impact of Non-Normality and Estimation Methods in SEM on Satisfaction Research in Marketing*. *Quality & Quantity*, 40(1):39–58.
- ANSARI, ASIM, K. JEDIDI und S. JAGPAL (2000). *A Hierarchical Bayesian Methodology for Treating Heterogeneity in Structural Equation Models*. *Marketing Science*, 19(4):328–347.
- ARMINGER, GERHARD und U. KÜSTERS (1988). *Latent trait models with indicators of mixed measurement level*. In: LANGEHEINE, ROLF und J. ROST, Hrsg.: *Latent trait and latent class models*, S. 51–73. Plenum Press, New York.
- ARMINGER, GERHARD, P. STEIN und J. WITTENBERG (1999). *Mixtures of conditional mean- and covariance-structure models*. *Psychometrika*, 64(4):475–494.
- ASSAEL, HENRY (1992). *Consumer behavior and marketing action*. PWS-KENT, Boston, MA, 4. Aufl.
- ATHIYAMAN, ADEE (1997). *Linking student satisfaction and service quality perceptions: the case of university education*. *European Journal of Marketing*, 31(7):528–540.
- BACKHAUS, KLAUS, B. ERICHSON, W. PLINKE und R. WEIBER (2003). *Multivariate Analysemethoden: Eine anwendungsorientierte Einführung*. Springer, Berlin, 10. Aufl.
- BALZER, WILLIAM K. und L. M. SULSKY (1992). *Halo and performance appraisal research: A critical examination*. *Journal of Applied Psychology*, 77(6):975–985.

- BARTHOLOMEW, DAVID J., M. KNOTT und I. MOUSTAKI (2011). *Latent variable models and factor analysis: A unified approach*. Wiley, Chichester, West Sussex, 3. Aufl.
- BARTHOLOMEW, DAVID JOHN und M. KNOTT (1999). *Latent variable models and factor analysis*. Oxford University Press, London, 2. Aufl.
- BAUMGARTNER, HANS und J.-B. E. M. STEENKAMP (2001). *Response Styles in Marketing Research: A Cross-National Investigation*. Journal of Marketing Research, 38(2):143–156.
- BAYES, THOMAS, R. PRICE und J. CANTON (1763). *An Essay towards solving a Problem in the Doctrine of Chances*. C. Davis Printer to the Royal Society of London, London.
- BAYOL, MARIE-PAULE, A. DE LA FOYE, C. TELLIER und M. TENENHAUS (2000). *Use of PLS Path Modelling to estimate the European Consumer Satisfaction Index (ECSI) model*. Statistica Applicata, 12(3):361–375.
- BEARDEN, WILLIAM O. und J. E. TEEL (1983). *Selected determinants of consumer satisfaction and complaint reports*. Journal of Marketing Research, 20(1):21–28.
- BECHGER, TIM M., G. MARIS und Y. P. HSIAO (2010). *Detecting Halo Effects in Performance-Based Examinations*. Applied Psychological Measurement, 34(8):607–619.
- BECKWITH, NEIL E. und U. V. KUBILIUS (1978). *Empirical Evidence of Halo Effects in Store Image Research by Estimating True Location*. Advances in Consumer Research, 5(1):485–493.
- BECKWITH, NEIL E. und D. R. LEHMANN (1975). *The importance of halo effects in multi-attribute attitude models*. Journal of Marketing Research, 12(3):265–275.
- BEREKOVEN, LUDWIG, W. ECKERT und P. ELLENRIEDER (2000). *Marktforschung: Methodische Grundlagen und praktische Anwendung*. Gabler, Wiesbaden, 8. Aufl.
- BERMAN, JEFFREY S. und D. A. KENNY (1976). *Correlational bias in observer ratings*. Journal of Personality and Social Psychology, 34(2):263–273.

- BERNARDIN, H. JOHN (1977). *Behavioral expectation scales versus summated scales: A fairer comparison*. Journal of Applied Psychology, 62(4):422–427.
- BEUTIN, NIKOLAS (2008). *Verfahren zur Messung der Kundenzufriedenheit im Überblick*. In: HOMBURG, CHRISTIAN, Hrsg.: *Kundenzufriedenheit*, S. 121–171. Gabler, Wiesbaden.
- BITNER, MARY J. und A. R. HUBBERT (1994). *Encounter Satisfaction Versus Overall Satisfaction Versus Quality: The Customer's Voice*. In: RUST, ROLAND T. und R. L. OLIVER, Hrsg.: *Service quality*. Sage Publications, Thousand Oaks.
- BITNER, MARY JO (1990). *Evaluating Service Encounters: The Effects of Physical Surroundings and Employee Responses*. The Journal of Marketing, 54(2):69–82.
- BLATTBERG, ROBERT C., G. GETZ und J. S. THOMAS (2001). *Customer equity: Building and managing relationships as valuable assets*. Harvard Business School Press, Boston, MA.
- BLIEMEL, MICHAEL und K. HASSANEIN (2007). *Consumer Satisfaction with Online Health Information Retrieval: A Model and Empirical Study*. e Service Journal, 5(2):53–83.
- BOLLEN, KENNETH und R. LENNOX (1991). *Conventional wisdom on measurement: A structural equation perspective*. Psychological Bulletin, 110(2):305–314.
- BOLLEN, KENNETH A. (1989). *Structural equations with latent variables*. Wiley, New York, NY u.a.
- BOLSTAD, WILLIAM M. (2007). *Introduction to Bayesian statistics*. John Wiley, Hoboken, N.J., 2. Aufl.
- BOLTON, RUTH N. und J. H. DREW (1991). *A multistage model of customers' assessments of service quality and value*. Journal of Consumer Research, 17(4):375–384.
- BORLE, SHARAD, S. S. SIDDHARTH und D. C. JAIN (2008). *Customer Lifetime Value Measurement*. Management Science, 54(1):100–112.

- BORTZ, JÜRGEN und N. DÖRING (2002). *Forschungsmethoden und Evaluation: Für Human- und Sozialwissenschaftler*. Springer, Berlin, 3. Aufl.
- BOULDING, WILLIAM, A. KALRA, R. STAELIN und V. A. ZEITHAML (1993). *A dynamic process model of service quality: From expectations to behavioral intentions*. Journal of Marketing Research, 30(1):7–27.
- BRADLOW, ERIC T., H. WAINER und X. WANG (1999). *A Bayesian random effects model for testlets*. Psychometrika, 64(2):153–168.
- BRANDT, D. RANDALL und K. L. REFFETT (1989). *Focusing on customer problems to improve service quality*. Journal of Services Marketing, 3(4):5–14.
- BROEN, WILLIAM E. JR. und R. D. WIRST (1958). *Varieties of response sets*. Journal of Consulting Psychology, 22(3):237–240.
- BROWN, EVA METZGER (1968). *Influence of Training, Method, and Relationship on the Halo Effect*. Journal of Applied Psychology, 52(3):195–199.
- BROWN, GEORGE H. (1952). *Brand loyalty- fact or fiction?*. Advertising Age, 23:53–55.
- BRUHN, MANFRED (2008). *Nationale Kundenzufriedenheitsindizes*. In: HOMBURG, CHRISTIAN, Hrsg.: *Kundenzufriedenheit*, S. 174–201. Gabler, Wiesbaden.
- BUSACCA, BRUNO und G. PADULA (2005). *Understanding the relationship between attribute performance and overall satisfaction: Theory, measurement and implications*. Marketing Intelligence & Planning, 23(6):543–561.
- BÜSCHKEN, JOACHIM, T. OTTER und G. M. ALLENBY (2010). *Do We Halo or Form? A Bayesian Mixture Model for Customer Satisfaction Data*. SSRN eLibrary.
- BYE, BARRY V., S. J. GALLICCHIO und J. M. DYKACZ (1985). *Multiple-Indicator, Multiple-Cause Models for a Single Latent Variable with Ordinal Indicators*. Sociological Methods & Research, 13(4):487–509.

- CADOTTE, ERNEST R., R. B. WOODRUFF und R. L. JENKINS (1987). *Expectations and Norms in Models of Consumer Satisfaction*. Journal of Marketing Research, 24(3):305–314.
- CARDONE-RIPORTELLA, CLARA, N. LADO und P. RIVERA-TORRES (2001). *Measurement and Effects of Teaching Quality: An Empirical Model Applied to Masters Programs*.
- CARLIN, BRADLEY P. und T. A. LOUIS (2000). *Bayes and empirical bayes methods for data analysis*. Texts in statistical science. Chapman & Hall/CRC, Boca Raton ; London, 2. Aufl.
- CASELLA, GEORGE und E. I. GEORGE (1992). *Explaining the Gibbs Sampler*. The American Statistician, 46(3):167–174.
- CHEUNG, CHRISTY M. K. und M. K. O. LEE (2004). *The Asymmetric Effect of Web Site Attribute Performance on Web Satisfaction: An Empirical Study*. E-Service Journal, 3(Summer):265–278.
- CHIB, SIDDHARTHA und E. GREENBERG (1995). *Understanding the Metropolis-Hastings Algorithm*. The American Statistician, 49(4):327–335.
- CHURCHILL, GILBERT A. JR und C. SURPRENANT (1982). *An Investigation into the Determinants of Customer Satisfaction*. Journal of Marketing Research, 19(4):491–504.
- CIAVOLINO, ENRICO und J. J. DAHLGAARD (2007). *ECSI - Customer Satisfaction Modelling and Analysis: A Case Study*. Total Quality Management & Business Excellence, 18(5):545–554.
- CLEMANS, WILLIAM VANCE (1956). *An analytical and empirical examination of some properties of ipsative measures*. Doktorarbeit, University of Washington, Seattle.
- COENDERS, GERMA, A. SATORRA und W. E. SARIS (1997). *Alternative approaches to structural modeling of ordinal data: A Monte Carlo study*. Structural Equation Modeling, 4(4):261–282.

- COHEN, JOEL B. und O. T. AHTOLA (1971). *An expectancy X value analysis of the relationship between consumer attitude and behavior*. In: GARDNER, DAVID M., Hrsg.: *Proceedings*, S. 344–364.
- CONGDON, PETER (2003). *Applied Bayesian modelling*. Wiley, Chichester, England u.a.
- CONGDON, PETER (2006). *Bayesian statistical modelling*. Wiley series in probability and statistics. Wiley, Chichester, England u.a., 2. Aufl.
- COOIL, BRUCE, T. L. KEININGHAM, L. AKSOY und M. HSU (2007). *A Longitudinal Analysis of Customer Satisfaction and Share of Wallet: Investigating the Moderating Effect of Customer Characteristics*. *Journal of Marketing*, 71(1):67–83.
- COOPER, WILLIAM H. (1981). *Ubiquitous halo*. *Psychological Bulletin*, 90(2):218–244.
- COPELAND, THOMAS E., T. KOLLER und J. MURRIN (1993). *Unternehmenswert: Methoden und Strategien für eine wertorientierte Unternehmensführung*. Campus-Verlag, Frankfurt am Main u.a.
- CROSBY, LAWRENCE A. und N. STEPHENS (1987). *Effects of Relationship Marketing on Satisfaction, Retention, and Prices in the Life Insurance Industry*. *Journal of Marketing Research*, 24(4):404–411.
- CROW, ROSEMARIE, H. GAGE, S. HAMPSON, J. HART, A. KIMBER, L. STOREY und H. THOMAS (2002). *The measurement of satisfaction with healthcare: implications for practice from a systematic review of the literature*. *Health technology assessment*, 6(32):1–244.
- CZEPIEL, JOHN und L. ROSENBERG (1977). *Consumer satisfaction: Concept and measurement*. *Journal of the Academy of Marketing Science*, 5(4):403–411.
- CZEPIEL, JOHN A., L. J. ROSENBERG und A. AKERELE (1974). *Perspectives on consumer satisfaction*. New York University, Graduate School of Business Administration, New York.

- DANAHER, PETER J. und V. HADDRELL (1996). *A comparison of question scales used for measuring customer satisfaction*. International Journal of Service Industry Management, 7(4).
- DAY, GEORGE S. (1969). *A two-dimensional concept of brand loyalty*. Journal of Advertising Research, 9(3):29–31,34–35.
- DAY, RALPH L. (1977). *Toward a Process Model of Consumer Satisfaction*. In: HUNT, H. KEITH, Hrsg.: *Conceptualization and Measurement of Consumer Satisfaction and Dissatisfaction*. Marketing Science Institute, Cambridge, MA.
- DECOTHS, THOMAS A. (1977). *An analysis of the external validity and applied relevance of three rating formats*. Organizational Behavior and Human Performance, 19(2):247–266.
- DEVLIN, SUSAN J., H. K. DONG und M. BROWN (1993). *Selecting a Scale for Measuring Quality*. Marketing Research, 15(3):13–16.
- DICKINSON, TERRY L. und T. E. TICE (1977). *The Discriminant Validity of Scales Developed by Retranslation*. Personnel Psychology, 30(2):217–228.
- DIEBOLT, JEAN und C. P. ROBERT (1994). *Estimation of Finite Mixture Distributions through Bayesian Sampling*. Journal of the Royal Statistical Society. Series B (Methodological), 56(2):363–375.
- DILLON, WILLIAM R., T. J. MADDEN, A. KIRMANI und S. MUKHERJEE (2001). *Understanding What's in a Brand Rating: A Model for Assessing Brand and Attribute Effects and Their Relationship to Brand Equity*. Journal of Marketing Research, 38(4):415–429.
- DILLON, WILLIAM R., J. B. WHITE, V. R. RAO und D. FILAK (1997). *'Good science'*. Marketing Research, 9(4):22–31.
- DOORN, JENNY VAN (2004). *Zufriedenheitsdynamik: Eine Panelanalyse bei industriellen Dienstleistungen*. Dt. Univ.-Verl., Wiesbaden, 1. Aufl.

- DRÈZE, JACQUES H. (1972). *Econometrics and Decision Theory*. *Econometrica*, 40(1):1–17.
- DRÈZE, JACQUES H. (1977). *Bayesian Theory of Identification in Simultaneous Equation Models*. In: FIENBERG, STEPHEN E., A. ZELLNER und L. J. SAVAGE, Hrsg.: *Studies in bayesian econometrics and statistics*. North - Holland, Amsterdam u.a.
- EDWARDS, JEFFREY R. (2011). *The Fallacy of Formative Measurement*. *Organizational Research Methods*, 14(2):370–388.
- EKLOF, JAN A., P. HACKL und A. WESTLUND (1999). *On measuring interactions between customer satisfaction and financial results*. *Total Quality Management*, 10(4):514–522.
- EREVELLES, SUNIL und C. LEAVITT (1992). *A Comparison of Current Models of Consumer Satisfaction/Dissatisfaction*. *Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behavior*, 5:104–114.
- EVERITT, BRIAN (1984). *An introduction to latent variable models*. Chapman and Hall, London.
- FARH, JIING-LIH, A. A. CANNELLA und A. G. BEDEIAN (1991). *The Impact of Purpose On Rating Quality and User Acceptance*. *Group & Organization Management*, 16(4):367–386.
- FEELEY, THOMAS HUGH (2002). *Comment on Halo Effects in Rating and Evaluation Research*. *Human Communication Research*, 28(4):578–586.
- FELDMAN, JACK M. und J. G. LYNCH (1988). *Self-generated validity and other effects of measurement on belief, attitude, intention, and behavior*. *Journal of Applied Psychology*, 73(3):421–435.
- FIELD, ANDY (2005). *Discovering statistics using SPSS for Windows: and sex, drugs and rock 'n' roll*. Sage, London, 2. Aufl.

- FIENBERG, STEPHEN (2006). *When did Bayesian inference become “Bayesian”?*. BAYESIAN ANALYSIS, 1(1):1–41.
- FINN, ADAM (2011). *Investigating the non-linear effects of e-service quality dimensions on customer satisfaction*. Journal of Retailing and Consumer Services, 18(1):27–37.
- FISHBEIN, MARTIN (1963). *An Investigation of the Relationships between Beliefs about an Object and the Attitude toward that Object*. Human Relations, 16(3):233–239.
- FISHBEIN, MARTIN (1965). *A Consideration of Beliefs, Attitudes, and Their Relationships*. In: STEINER, I.D, M. FISHBEIN und SOCIETY FOR THE PSYCHOLOGICAL STUDY OF SOCIAL ISSUES, Hrsg.: *Current studies in social psychology*, S. 107–120. Holt, Rinehart and Winston, New York, NY.
- FISHER, FRANKLIN MARVIN (1966). *The identification problem in econometrics*. Economics handbook series. McGraw-Hill, New York.
- FISICARO, SEBASTIANO A. und C. E. LANCE (1990). *Implications of Three Causal Models for the Measurement of Halo Error*. Applied Psychological Measurement, 14(4):419–429.
- FLANAGAN, JOHN C. (1954). *The critical incident technique*. Psychological Bulletin, 51(4):327–358.
- FORNELL, CLAES (1992). *A national customer satisfaction barometer: The Swedish experience*. Journal of Marketing, 56(1):6–21.
- FORNELL, CLAES, C. FORNELL, M. D. JOHNSON, E. W. ANDERSON und J. CHA (1996a). *The American Customer Satisfaction Index: Nature, purpose, and findings*. Journal of Marketing, 60(4):7–18.
- FORNELL, CLAES, M. D. JOHNSON, E. W. ANDERSON, J. CHA und B. E. BRYANT (1996b). *The American Customer Satisfaction Index: Nature, Purpose, and Findings*. The Journal of Marketing, 60(4):7–18.

- FORZA, CIPRIANO und R. FILIPPINI (1998). *TQM impact on quality conformance and customer satisfaction: A causal model*. International Journal of Production Economics, 55(1):1–20.
- FOURNIER, SUSAN und D. G. MICK (1999). *Rediscovering Satisfaction*. The Journal of Marketing, 63(4):5–23.
- FREY, BRUNO S. und H. WECK-HANNEMAN (1984). *The hidden economy as an "unobserved" variable*. European Economic Review, 26(1-2):33–53.
- FRÜWIRTH-SCHNATTER, SYLVIA (2006). *Finite Mixture and Markov-Switching Models*. Springer.
- FUCHS, CHRISTOPH und A. DIAMANTOPOULOS (2009). *Using single-item measures for construct measurement*. Die Betriebswirtschaft, 69(2):195–210.
- GAL, DAVID und D. D. RUCKER (2011). *Answering the Unasked Question: Response Substitution in Consumer Surveys: Journal of Marketing Research*. Journal of Marketing Research, 48(1):185–195.
- GALTON, FRANCIS (1880). *Statistics of Mental Imagery*. Mind, 5(19):301–318.
- GELBRICH, KATJA und S. MÜLLER (2006). *Kundenwert: Hintergrund, Konzeptualisierung und Messmethoden*. In: SCHWEICKART, NIKOLAUS und A. TÖPFER, Hrsg.: *Wertorientiertes Management*, S. 449–487. Springer Berlin Heidelberg.
- GELFAND, ALAN E. und S. K. SAHU (1999). *Identifiability, Improper Priors, and Gibbs Sampling for Generalized Linear Models*. Journal of the American Statistical Association, 94(445):247–253.
- GELFAND, ALAN E. und A. F. M. SMITH (1990). *Sampling-Based Approaches to Calculating Marginal Densities*. Journal of the American Statistical Association, 85(410):398–409.
- GELMAN, ANDREW (2004). *Parameterization and Bayesian Modeling*. Journal of the American Statistical Association, 99(466):537–545.

- GEMAN, STUART und D. GEMAN (1984). *Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images*. IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-6(6):721–741.
- GERBING, DAVID W. und J. C. ANDERSON (1988). *An Updated Paradigm for Scale Development Incorporating Unidimensionality and Its Assessment*. Journal of Marketing Research, 25(2):186–192.
- GEWEKE, JOHN (1991). *Efficient Simulation from the Multivariate Normal and Student-t Distributions Subject to Linear Constraints and the Evaluation of Constraint Probabilities*.
- GEWEKE, JOHN (2005). *Contemporary Bayesian econometrics and statistics*. John Wiley, Hoboken, N.J.
- GHISELLI, EDWIN E. und C. W. BROWN (1948). *Personnel and Industrial Psychology*. McGraw-Hill, New York.
- GHISELLI, EDWIN E. und C. W. BROWN (1955). *Personnel and Industrial Psychology*. McGraw-Hill, New York, 2 Aufl.
- GHOSH, JAYANTA, M. K. DELAMPADY und T. SAMANTA (2006). *An Introduction to Bayesian Analysis*. Springer Science+Business Media, LLC, New York, NY.
- GHOSH, MALAY, A. GHOSH, M.-H. CHEN und A. AGRESTI (2000). *Noninformative priors for one-parameter item response models*. Journal of Statistical Planning and Inference, 88(1):99–115.
- GIESE, JOAN L und J. A. COTE (2000). *Defining Consumer Satisfaction*. Academy of Marketing Science Review, (1).
- GILKS, WALLY R. (1992). *Derivative-free Adaptive Rejection Sampling for Gibbs Sampling*. In: BERNARDO, J. M., J. BERGER, A. P. DAWID und A. F. M. SMITH, Hrsg.: *Bayesian statistics 4*, S. 641–649. Clarendon Press and Oxford University Press, Oxford, New York.

- GODBOLD, LINDA C. und M. PFAU (2000). *Conferring Resistance to Peer Pressure Among Adolescents: Using Inoculation Theory to Discourage Alcohol Use*. *Communication Research*, 27(4):411–437.
- GÓMEZ, MIGUEL I., E. W. MCCLAUGHLIN und D. R. WITTINK (2004). *Customer satisfaction and retail sales performance: an empirical investigation*. *Journal of Retailing*, 80(4):265–278.
- GREEN, PAUL E., D. S. TULL und G. ALBAUM (1988). *Research for marketing decision*. PHI Learning, New Delhi, Indien, 5. Aufl.
- GRIGOROUDIS, EVANGELOS und O. SPYRIDAKI (2003). *Derived vs. stated importance in customer satisfaction surveys*. *Operational Research*, 3(3):229–247.
- GRUCA, THOMAS S. und L. L. REGO (2005). *Customer Satisfaction, Cash Flow, and Shareholder Value*. *The Journal of Marketing*, 69(3):115–130.
- GUILFORD, JOY P. (1936). *Psychometric Methods*. McGraw-Hill, New York, NY.
- GUILFORD, JOY P. (1954). *Psychometric methods*. McGraw-Hill series in psychology. McGraw-Hill, New York, NY, 2. Aufl.
- GÜNTER, BERND (2008). *Beschwerdemanagement als Schlüssel zur Kundenzufriedenheit*. In: HOMBURG, CHRISTIAN, Hrsg.: *Kundenzufriedenheit*. Gabler, Wiesbaden.
- GUPTA, SUNIL, D. R. LEHMANN und J. A. STUART (2004). *Valuing Customers*. *Journal of Marketing Research*, 41(1):7–18.
- GUPTA, SUNIL und V. ZEITHAML (2006). *Customer Metrics and Their Impact on Financial Performance*. *Marketing Science*, 25(6):718–739.
- HAAVELMO, TRYGVE (1943). *The Statistical Implications of a System of Simultaneous Equations*. *Econometrica*, 11(1):1–12.
- HAJIVASSILIOU, VASSILIS A. und D. L. MCFADDEN (1998). *The Method of Simulated Scores for the Estimation of LDV Models*. *Econometrica*, 66(4):863–896.

- HALLOWELL, ROGER (1996). *The relationships of customer satisfaction, customer loyalty, and profitability: an empirical study*. International Journal of Service Industry Management, 7(4):27–42.
- HALSTEAD, DIANE, D. HARTMAN und S. L. SCHMIDT (1994). *Multisource Effects on the Satisfaction Formation Process*. Journal of the Academy of Marketing Science, 22(2):114–129.
- HAMMANN, PETER und B. ERICHSON (2000). *Marktforschung*. Lucius und Lucius, Stuttgart, 4. Aufl.
- HARDT, NINO und J. BUESCHKEN (2010). *Estimating SEM with Ordinal Data and Scale Usage Heterogeneity*. SSRN eLibrary.
- HASTINGS, W. KEITH (1970). *Monte Carlo Sampling Methods Using Markov Chains and Their Applications*. Biometrika, 57(1):97–109.
- HAUSER, ROBERT M. und A. S. GOLDBERGER (1971). *The Treatment of Unobservable Variables in Path Analysis*. Sociological Methodology, 3:81–117.
- HAUSKNECHT, DOUGLAS (1990). *Measurement scales in consumer satisfaction/dissatisfaction*. Journal of Consumer Satisfaction/Dissatisfaction and Complaint Behavior, (3):1–11.
- HEDGE, JERRY W. und M. J. KAVANAGH (1988). *Improving the accuracy of performance evaluations: Comparison of three methods of performance appraiser training*. Journal of Applied Psychology, 73(1):68–73.
- HERSBERGER, SCOTT L. (2003). *The growth of structural equation modeling: 1994–2001*. Structural Equation Modeling, 10(1):35–46.
- HERZBERG, FREDERICK, B. MAUSNER und B. SNYDERMAN (1959). *The motivation to work*. John Wiley, Oxford, 2 Aufl.
- HESKETT, JAMES L., T. O. JONES, G. W. LOVEMAN, W. E. J. SASSER und L. A. SCHLESINGER (1994). *Putting the Service-Profit Chain to Work*. Harvard Business Review, 72(2):164–170.

- HILL, NIGEL und J. ALEXANDER (2000). *Handbook of customer satisfaction and loyalty measurement*. Gower, Aldershot, 2. Aufl.
- HILL, NIGEL, B. SELF und G. ROCHE (2002). *Customer satisfaction measurement for ISO 9000: 2000*. Butterworth-Heinemann, Oxford.
- HOIJTINK, HERBERT, G. ROOKS und F. W. WILMINK (1999). *Confirmatory Factor Analysis of Items With a Dichotomous Response Format Using the Multidimensional Rasch Model*. *Psychological Methods*, 4(3):300–314.
- HOMBURG, CHRISTIAN und A. FÜRST (2008). *Überblick über die Messung von Kundenzufriedenheit und Kundenbindung*. In: BRUHN, MANFRED und C. HOMBURG, Hrsg.: *Handbuch Kundenbindungsmanagement*, S. 607–642. Gabler, Wiesbaden.
- HOMBURG, CHRISTIAN und L. HILDEBRANDT (1998). *Die Kausalanalyse: Bestandsaufnahme, Entwicklungsrichtungen, Problemfelder*. In: HILDEBRANDT, LUTZ und C. HOMBURG, Hrsg.: *Die Kausalanalyse*. Schäffer-Poeschel, Stuttgart.
- HOMBURG, CHRISTIAN und N. KOSCHATE (2007). *Kundenzufriedenheit und Kundenbindung*. In: ALBERS, SÖNKE und A. HERRMANN, Hrsg.: *Handbuch Produktmanagement*, S. 843–867. Gabler, Wiesbaden.
- HOMBURG, CHRISTIAN und H. KROHMER (2003). *Marketingmanagement*. Gabler, Wiesbaden, 1. Aufl.
- HOMBURG, CHRISTIAN und B. RUDOLPH (2001). *Customer satisfaction in industrial markets: Dimensional and multiple role issues*. *Journal of Business Research*, 52(1):15–33.
- HOMBURG, CHRISTIAN und R. STOCK-HOMBURG (2008). *Theoretische Perspektiven zur Kundenzufriedenheit*. In: HOMBURG, CHRISTIAN, Hrsg.: *Kundenzufriedenheit*, S. 17–51. Gabler, Wiesbaden.
- HOMBURG, CHRISTIAN und R. STOCK (2003). *Theoretische Perspektiven zur Kundenzufriedenheit*. In: HOMBURG, CHRISTIAN, Hrsg.: *Kundenzufriedenheit*. Gabler, Wiesbaden.

- HOMBURG, CHRISTIAN und H. WERNER (1998). *Kundenorientierung mit System: Mit Customer-Orientations-Management zu profitabilem Wachstum*. Campus, Frankfurt.
- HOWARD, JOHN A. und J. N. SHETH (1969). *The theory of buyer behavior*. Wiley & Sons, New York.
- HOWELL, ROY D., E. BREIVIK und J. B. WILCOX (Jun, 2007). *Reconsidering formative measurement*. *Psychological Methods*, 12(2):205–218.
- HSIAO, CHENG (1983). *Chapter 4 Identification*. In: ZVI GRILICHES AND MICHAEL D. INTRILIGATOR, Hrsg.: *Handbook of Econometrics*, Bd. Volume 1, S. 223–283. Elsevier.
- HUNT, H. KEITH (1977). *CS/D - Overview and Future Research Direction*. In: HUNT, H. KEITH, Hrsg.: *Conceptualization and Measurement of Consumer Satisfaction and Dissatisfaction*. Marketing Science Institute, Cambridge, MA.
- ISHWARAN, HEMANT (2000). *Univariate and multivariate ordinal cumulative link regression with covariate specific cutpoints*. *Can J Statistics*, 28(4):715–730.
- ITTNER, CHRISTOPHER D. und D. F. LARCKER (1998). *Are Nonfinancial Measures Leading Indicators of Financial Performance? An Analysis of Customer Satisfaction*. *Journal of Accounting Research*, 36:1–35.
- JACOBS, RICK und S. W. J. KOZLOWSKI (1985). *A Closer Look at Halo Error in Performance Ratings*. *The Academy of Management Journal*, 28(1):201–212.
- JACOBY, JACOB und D. B. KYNER (1973). *Brand Loyalty vs. Repeat Purchasing Behavior*. *Journal of Marketing Research*, 10(1):1–9.
- JAMES, WILLIAM und F. S. CARTER (1978). *Halo Effects and Location Preferences*. *Advances in Consumer Research*, 5(1):474–476.
- JAVARAS, KRISTIN N. und B. D. RIPLEY (2007). *An “Unfolding” Latent Variable Model for Likert Attitude Data*. *Journal of the American Statistical Association*, 102(478):454–463.

- JEDIDI, KAMEL, H. S. JAGPAL und W. S. DESARBO (1997). *Finite-Mixture Structural Equation Models for Response-Based Segmentation and Unobserved Heterogeneity*. Marketing Science.
- JOHNSON, DONALD M. (1963). *Reanalysis of experimental halo effects*. Journal of Applied Psychology, 47(1):46–47.
- JOHNSON, MICHAEL D., M. D. JOHNSON und C. FORNELL (1991). *A framework for comparing customer satisfaction across individuals and product categories*. Journal of Economic Psychology, 12(2):267–286.
- JOHNSON, TIMOTHY R. (2003). *On the use of heterogeneous thresholds ordinal regression models to account for individual differences in response style*. Psychometrika, 68(4):563–583.
- JONES, MICHAEL A. (2006). *A Content Analysis of Customer Satisfaction in Annual Reports*. Journal of Consumer Satisfaction, Dissatisfaction & Complaining Behavior, 19:59–75.
- JONES, THOMAS O. und W. E. J. SASSER (1995). *Why Satisfied Customers Defect*. Harvard Business Review, 73(6):88–91.
- JÖRESKOG, KARL und D. SÖRBOM (1996). *LISREL 8: Structural Equation Modeling with the SIMPLIS Command Language*. SSI Scientific Software International.
- JÖRESKOG, KARL G. (1971). *Simultaneous factor analysis in several populations*. Psychometrika, 36(4):409–426.
- JÖRESKOG, KARL G. und A. S. GOLDBERGER (1975). *Estimation of a Model with Multiple Indicators and Multiple Causes of a Single Latent Variable*. Journal of the American Statistical Association, 70(351):631–639.
- JUDD, CHARLES M., R. A. DRAKE, J. W. DOWNING und J. A. KROSNICK (1991). *Some dynamic properties of attitude structures: Context-induced response facilitation and polarization*. Journal of Personality and Social Psychology, 60(2):193–202.

- KADANE, JOSEPH B. (1977). *The Role of Identification in Bayesian Theory*. In: FIENBERG, STEPHEN E., A. ZELLNER und L. J. SAVAGE, Hrsg.: *Studies in bayesian econometrics and statistics*. North - Holland, Amsterdam u.a.
- KAFRY, DITSA, R. JACOBS und S. ZEDECK (1979). *Discriminability in Multidimensional Performance Evaluations*. *Applied Psychological Measurement*, 3(2):187–192.
- KAISER, MARC-OLIVER (2005). *Erfolgsfaktor Kundenzufriedenheit: Dimensionen und Messmöglichkeiten*. Erich Schmidt, Berlin, 2. Aufl.
- KANNING, UWE PETER und N. BERGMANN (2009). *Predictors of customer satisfaction: testing the classical paradigms*. *Managing Service Quality*, 19(4):377–390.
- KANO, NORIAKI (1984). *Attractive Quality and Must-be Quality*. *Journal of the Japanese Society for Quality Control*, 14(2):39–48.
- KATCHER, BRUCE L. (2003). *Make more Strategic Use of Customer Satisfaction Surveys*. *Journal of Business Strategy*, 24(1):34.
- KATZ, KAREN L., B. M. LARSON und R. C. LARSON (1991). *Prescription for the Waiting-in-Line Blues: Entertain, Enlighten, and Engage*. *Sloan Management Review*, 32(2):44–53.
- KEININGHAM, TIMOTHY L., B. COOIL, T. W. ANDREASSEN und L. AKSOY (2007). *A Longitudinal Examination of Net Promoter and Firm Revenue Growth*. *Journal of Marketing*, 71(3):39–51.
- KENNY, DAVID A. und J. S. BERMAN (1980). *Statistical approaches to the correction of correlational bias*. *Psychological Bulletin*, 88(2):288–295.
- KING, LARRY M., J. E. HUNTER und F. L. SCHMIDT (1980). *Halo in a multidimensional forced-choice performance evaluation scale*. *Journal of Applied Psychology*, 65(5):507–516.
- KLINE, REX B. (2011). *Principles and practice of structural equation modeling*. Guilford Press, New York, 3. Aufl.

- KOCH, KARL-RUDOLF (2007). *Introduction to Bayesian statistic*. Springer, Berlin, 2. Aufl.
- KOOPMANS, TJALLING C. (1949). *Identification Problems in Economic Model Construction*. *Econometrica*, 17(2):125–144.
- KOOPMANS, TJALLING C. (1953). *Identification Problems in Economic Model Construction*. In: KOOPMANS, TJALLING C., W. C. HOOD und COWLES COMMISSION RESEARCH STAFF MEMBERS, Hrsg.: *Studies in econometric method*, Bd. 14 d. Reihe *Monographs of the Cowles Commission for Research in Economics*, S. 27–48. Wiley & Sons, New York.
- KOOPMANS, TJALLING C., W. C. HOOD und COWLES COMMISSION RESEARCH STAFF MEMBERS, Hrsg. (1953). *Studies in econometric method*, Bd. 14 d. Reihe *Monographs of the Cowles Commission for Research in Economics*. Wiley & Sons, New York.
- KRAUT, ALLEN I. (1975). *Prediction of managerial success by peer and training-staff ratings*. *Journal of Applied Psychology*, 60(1):14–19.
- KROSNICK, JON A. (1991). *The Stability of Political Preferences: Comparisons of Symbolic and Nonsymbolic Attitudes*. *American Journal of Political Science*, 35(3):547–576.
- KUEHN, ALFRED A. (1962). *Consumer Brand Choice as a Learning Process*. *Journal of Advertising Research*, 2(4):10–17.
- KUMAR, V., D. A. AAKER und G. S. DAY (2002). *Essentials of marketing research*. Wiley, New York, 2. Aufl.
- LANDY, DAVID und H. SIGALL (1974). *Beauty is talent: Task evaluation as a function of the performer's physical attractiveness*. *Journal of Personality and Social Psychology*, 29(3):299–304.
- LANDY, FRANK J., R. J. VANCE, J. L. BARNES-FARRELL und J. W. STEELE (1980). *Statistical control of halo error in performance ratings*. *Journal of Applied Psychology*, 65(5):501–506.

- LATHAM, GARY P. und K. N. WEXLEY (1981). *Increasing productivity through performance appraisal*. Addison-Wesley, Reading, MA.
- LATOUR, STEPHEN A. und N. C. PEAT (1979). *Conceptual and Methodological Issues in Consumer Satisfaction Research*. *Advances in Consumer Research*, 6(1):431–437.
- LAY, CLARRY H. und D. N. JACKSON (1969). *Analysis of the generality of trait-inferential relationships*. *Journal of Personality and Social Psychology*, 12(1):12–21.
- LEAVITT, CLARK (1977). *Consumer Satisfaction and Dissatisfaction: Bipolar or Independent*. In: HUNT, H. KEITH, Hrsg.: *Conceptualization and Measurement of Consumer Satisfaction and Dissatisfaction*. Marketing Science Institute, Cambridge, MA.
- LEE, SIK-YUM (2007). *Structural equation modeling: A Bayesian approach*. Wiley series in probability and statistics. Wiley, Chichester.
- LEE, SOONMOOK und S. HERSHBERGER (1990). *A Simple Rule for Generating Equivalent Models in Covariance Structure Modeling*. *Multivariate Behavioral Research*, 25(3):313–334.
- LENK, PETER und B. ORME (2009). *The Value of Informative Priors in Bayesian Inference with Sparse Data*. *Journal of Marketing Research (JMR)*, 46(6):832–845.
- LENK, PETER, M. WEDEL und U. BÖCKENHOLT (2006). *Bayesian Estimation of Circumplex Models Subject to Prior Theory Constraints and Scale-Usage Bias*. *Psychometrika*, 71(1):33–55.
- LEVY, LBON H. und R. D. DUGAN (1960). *A constant error approach to the study of dimensions of social perception*. *Journal of Abnormal and Social Psychology*, 61(1):21–24.
- LEYDOLD, JOSEF und W. HÖRMANN (2011). *Runuran: R interface to the UNU.RAN random variate generators*.
- LICHTENSTEIN, MERYL und T. K. SRULL (1985). *Conceptual and Methodological Issues in Examining the Relationship Between Consumer Memory and Judgment*. In: ALWITT,

- LINDA F. und A. A. MITCHELL, Hrsg.: *Psychological processes and advertising effects*, S. 113–128. L. Erlbaum Associates, Hillsdale, NJ.
- LINDLEY, D. V. (1975). *The Future of Statistics: A Bayesian 21st Century*. *Advances in Applied Probability*, 7:106–115.
- LINDLEY, DENNIS V. (1972). *Bayesian statistics: A review*, Bd. 2 d. Reihe *Regional conference series in applied mathematics*. Society for Industrial and Applied Mathematics, Philadelphia, PA.
- LUNN, DAVID, D. SPIEGELHALTER, A. THOMAS und N. BEST (2009). *The BUGS project: Evolution, critique and future directions*. *Statistics in Medicine*, 28(25):3049–3067.
- LUO, XUEMING (2007). *Consumer Negative Voice and Firm-Idiosyncratic Stock Returns*. *Journal of Marketing*, 71(3):75–88.
- LYNCH, JOHN G. und T. K. SRULL (1982). *Memory and attentional factors in consumer choice: Concepts and research methods*. *Journal of Consumer Research*, 9(1):18–37.
- LYNCH, SCOTT M. (2007). *Introduction to applied Bayesian statistics and estimation for social scientists*. Springer, New York.
- MACCALLUM, ROBERT C., D. T. WEGENER, B. N. UCHINO und L. R. FABRIGAR (1993). *The problem of equivalent models in applications of covariance structure analysis*. *Psychological Bulletin*, 114(1):185–199.
- MACCORMICK, ERNEST J. und J. TIFFIN (1966). *Industrial psychology*. Allen & Unwin, London, 3. Aufl.
- MACKENZIE, SCOTT B. (2001). *Opportunities for improving consumer research through latent variable structural equation modeling*. *Journal of Consumer Research*, 28(1):159–166.
- MADDEN, THOMAS J., W. R. DILLON und R. L. LEAK (2010). *Students' Evaluation of Teaching: Concerns of Item Diagnosticity*. *Journal of Marketing Education*, 32(3):264–274.

- MADDOX, R. NEIL (1981). *Two-factor Theory and Consumer Satisfaction: Replication and Extension*. The Journal of Consumer Research, 8(1):97–102.
- MARTILLA, JOHN A. und J. C. JAMES (1977). *Importance-Performance Analysis*. The Journal of Marketing, 41(1):77–79.
- MATTILA, ANNA S. (2003). *The Impact of Cognitive Inertia on Postconsumption Evaluation Processes*. Journal of the Academy of Marketing Science, 31(3):287–299.
- MATZLER, KURT und F. BAILOM (2006). *Messung von Kundenzufriedenheit*. In: HINTERHUBER, HANS H. und K. MATZLER, Hrsg.: *Kundenorientierte Unternehmensführung*, S. 241–270. Gabler, Wiesbaden.
- MATZLER, KURT, F. BAILOM, H. H. HINTERHUBER, B. RENZL und J. PICHLER (2004). *The asymmetric relationship between attribute-level performance and overall customer satisfaction: a reconsideration of the importance-performance analysis*. Industrial Marketing Management, 33(4):271–277.
- MATZLER, KURT und E. SAUERWEIN (2002). *The factor structure of customer satisfaction: An empirical test of the importance grid and the penalty-reward-contrast analysis*. International Journal of Service Industry Management, 13(4):314–332.
- MATZLER, KURT, E. SAUERWEIN und C. STARK (2006a). *Methoden zur Identifikation von Basis-, Leistungs- und Begeisterungsfaktoren*. In: HINTERHUBER, HANS H. und K. MATZLER, Hrsg.: *Kundenorientierte Unternehmensführung*, S. 289–313. Gabler, Wiesbaden.
- MATZLER, KURT, H. K. STAHL und H. H. HINTERHUBER (2006b). *Die Customer-based View der Unternehmung*. In: HINTERHUBER, HANS H. und K. MATZLER, Hrsg.: *Kundenorientierte Unternehmensführung*. Gabler, Wiesbaden.
- MCCULLOCH, ROBERT und P. E. ROSSI (1994). *An exact likelihood analysis of the multinomial probit model*. Journal of Econometrics, 64(1/2):207–240.

- MCCULLOCH, ROBERT E., N. G. POLSON und P. E. ROSSI (2000). *A Bayesian analysis of the multinomial probit model with fully identified parameters*. *Journal of Econometrics*, 99(1):173–193.
- MCDONNELL, JOHN (2007). *Music, scent and time preferences for waiting lines*. *International Journal of Bank Marketing*, 25(4):223–237.
- MCDOUGALL, GORDON H.G und T. LEVESQUE (2000). *Customer satisfaction with services: putting perceived value into the equation*. *Journal of Services Marketing*, 14(5):392.
- MCKELVEY, RICHARD und W. ZAVOINA (1975). *A statistical model for the analysis of ordinal level dependent variables*. *The Journal of Mathematical Sociology*, 4(1):103–120.
- MCREYNOLDS, PAUL und K. LUDWIG (1987). *On the history of rating scales*. *Personality and Individual Differences*, 8(2):281–283.
- METROPOLIS, NICHOLAS, A. ROSENBLUTH, M. ROSENBLUTH, A. TELLER und E. TELLER (1953). *Equation of State Calculations by Fast Computing Machines*. *The Journal of Chemical Physics*, 21(6):1087–1092.
- MEYER, ANTON und R. KANTSPERGER (2005). *Kundenzufriedenheit*. In: *Wirtschaftspsychologie*. Beltz, Weinheim.
- MEYN, SEAN P. und R. L. TWEEDIE (2009). *Markov chains and stochastic stability*. Cambridge University Press, Cambridge, 2. Aufl.
- MILLER, JOHN A. (1977). *Studying satisfaction, modifying models, eliciting expectations, posing problems and making meaningful measurements*. In: HUNT, H. KEITH, Hrsg.: *Conceptualization and Measurement of Consumer Satisfaction and Dissatisfaction*. Marketing Science Institute, Cambridge, MA.
- MOORE, WILLIAM L. und W. L. JAMES (1978). *A Study of the Amount of Halo in the Perceptions of Automobiles*. *Advances in Consumer Research*, 5(1):481–484.

- MORALES, JUAN-ANTONIO (1971). *Bayesian full information structural analysis*. Springer-Verlag, Berlin, New York.
- MÜLLER, HANS (1999). *Probabilistische Testmodelle für diskrete und kontinuierliche Ratingskalen: Einführung in die Item-Response-Theorie für abgestufte und kontinuierliche Items ; mit 4 Tabellen*. Huber, Bern u.a., 1. Aufl.
- MURPHY, KEVIN R. und R. L. ANHALT (1992). *Is halo error a property of the rater, rates, or the specific behaviors observed?*. Journal of Applied Psychology, 77(4):494–500.
- MURPHY, KEVIN R. und R. JAKO (1989). *Under What Conditions Are Observed Intercorrelations Greater Or Smaller Than True Intercorrelations?*. Journal of Applied Psychology, 74(5):827.
- MURPHY, KEVIN R., R. A. JAKO und R. L. ANHALT (1993). *Nature and Consequences of Halo Error: A Critical Analysis*. Journal of Applied Psychology, 78(2):218–225.
- MURPHY, KEVIN R., K. R. MURPHY und W. K. BALZER (1986). *Systematic distortions in memory-based behavior ratings and performance evaluations: Consequences for rating accuracy*. Journal of Applied Psychology, 71(1):39–44.
- MUTHEN, BENGT (1984). *A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators*. Psychometrika, 49(1):115–132.
- MUTHEN, BENGT O. (1989). *Latent Variable Modeling in Heterogeneous Populations*. Psychometrika, 54(4):557–585.
- MYERS, JAMES H. und M. I. ALPERT (1977). *Semantic Confusion in Attitude Research: Salience vs. Importance vs. Determinance*. Advances in Consumer Research, 4(1):106–110.
- NAGAO, KAORI, S. L. VARGO und F. W. MORGAN (2006). *Satisfiers, Dissatisfiers, Criticals, and Neutrals: A Review of Their Relative Effects on Customer (Dis)Satisfaction*. AMA Winter Educators' Conference Proceedings, 17:71–72.

- NATHAN, BARRY R. und R. G. LORD (1983). *Cognitive categorization and dimensional schemata: A process approach to the study of halo in performance ratings*. Journal of Applied Psychology, 68(1):102–114.
- NEAL, RADFORD M. (2003). *Slice sampling*. The Annals of Statistics, 31(3):705–767.
- NEWCOMB, THEODOR M. (1931). *An experiment designed to test the validity of a rating technique*. Journal of Educational Psychology, 22(4):279–289.
- NEWTON, MICHAEL A. und A. E. RAFTERY (1994). *Approximate Bayesian Inference with the Weighted Likelihood Bootstrap*. Journal of the Royal Statistical Society. Series B (Methodological), 56(1):3–48.
- NISBETT, RICHARD E. und T. D. WILSON (1977). *The halo effect: Evidence for unconscious alteration of judgments*. Journal of Personality and Social Psychology, 35(4):250–256.
- OLIVER, RICHARD L. (1977). *A Theoretical Reinterpretation of Expectation and Disconfirmation Effects on Posterior Product Evaluation: Experiences in the Field*. In: DAY, RALPH, Hrsg.: *Consumer Satisfaction, Dissatisfaction, and Complaining Behavior*, S. 2–9.
- OLIVER, RICHARD L. (1980). *A Cognitive Model of the Antecedents and Consequences of Satisfaction Decisions*. Journal of Marketing Research, 17(4):460–469.
- OLIVER, RICHARD L. (1981). *Measurement and evaluation of satisfaction processes in retail settings*. Journal of Retailing, 57(3):25–48.
- OLIVER, RICHARD L. (1992). *An Investigation of the Attribute Basis of Emotion and Related Affects in Consumption: Suggestions for a Stage-Specific Satisfaction Framework*. Advances in Consumer Research, 19:237–244.
- OLIVER, RICHARD L. (1997). *Satisfaction: A behavioral perspective on the consumer*. McGraw Hill, New York.

- OLIVER, RICHARD L. (2010). *Satisfaction: A behavioral perspective on the consumer*. M.E. Sharpe, Armonk, NY, 2. Aufl.
- PAGANI, LAURA und M. C. ZANAROTTI (2010). *Some uses of Rasch models parameters in customer satisfaction data analysis*. *Quality Technology & Quantitative Management*, 7(1):83–95.
- PALOMO, JESUS, D. B. DUNSON und K. BOLLEN (2007). *Bayesian Structural Equation Modeling*. In: LEE, SIK-YUM, Hrsg.: *Handbook of latent variable and related models*, S. 163–188. North-Holland, Amsterdam u.a.
- PARASURAMAN, A., L. L. BERRY und V. A. ZEITHAML (1991). *Understanding Customer Expectations of Service*. *Sloan Management Review*, 32(3):39–48.
- PARASURAMAN, A., V. A. ZEITHAML und L. L. BERRY (1988). *SERVQUAL: A Multiple-Item Scale for Measuring Consumer Perceptions of Service Quality*. *Journal of Retailing*, 64(1):12–40.
- PAULHUS, DELROY L. (1991). *Measurement and Control of Response Bias*. In: ROBINSON, JOHN P., P. R. SHAVER, L. S. WRIGHTSMAN und F. M. ANDREWS, Hrsg.: *Measures of personality and social psychological attitudes*, S. 17–59. Academic Press, San Diego.
- PEARL, JUDEA (1995). *On the testability of causal models with latent and instrumental variables*. In: BESNARD, PHILIPPE, Hrsg.: *Uncertainty in artificial intelligence*. Morgan Kaufmann.
- PEARL, JUDEA (2007). *On The Foundation Of Structural Equation Models or When Can We Give Causal Interpretation To Structural Coefficients?*.
- PEARL, JUDEA (2010a). *Causality: Models, reasoning, and inference*. Univ. Press, Cambridge, 2. Aufl.
- PEARL, JUDEA (2010b). *The Causal Foundations of Structural Equation Modeling*.

- PETERSON, ROBERT A. und W. R. WILSON (1992). *Measuring customer satisfaction: Fact and artifact*. Journal of the Academy of Marketing Science, 20(1):61–71.
- PEAFF, MARTIN (1977). *The index of consumer satisfaction: Measurement problems and opportunities*. In: HUNT, H. KEITH, Hrsg.: *Conceptualization and Measurement of Consumer Satisfaction and Dissatisfaction*. Marketing Science Institute, Cambridge, MA.
- POIRIER, DALE J. (1998). *Revising Beliefs in Nonidentified Models*. Econometric Theory, 14(4):483–509.
- QUINTANA, STEPHEN M. und S. E. MAXWELL (1999). *Implications of Recent Developments in Structural Equation Modeling for Counseling Psychology*. The Counseling Psychologist, 27(4):485–527.
- R DEVELOPMENT CORE TEAM (2011). *R: A Language and Environment for Statistical Computing*.
- RAAB, GERHARD (2008). *Customer relationship management: A global perspective*. Gower, Aldershot, Hampshire, England u.a.
- RAPPAPORT, ALFRED (1981). *Selecting strategies that create shareholder value*. Harvard Business Review, 59(3):139–149.
- RAPPAPORT, ALFRED (1986). *Creating shareholder value: The new standard for business performance*. Free Press and Collier Macmillan, New York, London.
- RAYKOV, TENKO und S. PENEV (2001). *The Problem of Equivalent Structural Equation Models: An Individual Residual Perspective*. In: MARCOULIDES, GEORGE A und R. E. SCHUMACKER, Hrsg.: *New developments and techniques in structural equation modeling*. Erlbaum, Mahwah, N.J u.a.
- REICHHELD, FREDERICK F. (1996). *The loyalty effect: The hidden force behind growth, profits, and lasting value*. Harvard Business School Press, Boston, MA.

- REICHHELD, FREDERICK F. (2003). *The One Number You Need to Grow*. Harvard Business Review, 81(12):46–54.
- REICHHELD, FREDERICK F. und W. E. J. SASSER (1990). *Zero defections: Quality comes to services*. Harvard Business Review, 68(5):105–111.
- ROBERT, CHRISTIAN P. und G. CASELLA (2004). *Monte Carlo statistical methods*. Springer texts in statistics. Springer, New York, 2. Aufl.
- ROBINS, PHILIP K. und R. W. WEST (1977). *Measurement Errors in the Estimation of Home Value*. Journal of the American Statistical Association, 72(358):290–294.
- ROSMALLEN, JOOST VAN, H. VAN HERK und P. J. F. GROENEN (2010). *Identifying Response Styles: A Latent-Class Bilinear Multinomial Logit Model*. Journal of Marketing Research, 47(1):157–172.
- ROSSI, PETER E. (2011). *bayesm: Bayesian Inference for Marketing/Micro-econometrics*.
- ROSSI, PETER E, G. M. ALLENBY und R. MCCULLOCH (2006). *Bayesian statistics and marketing*. Wiley series in probability and statistics. Wiley, Chichester.
- ROSSI, PETER E., Z. GILULA und G. M. ALLENBY (2001). *Overcoming Scale Usage Heterogeneity: A Bayesian Hierarchical Approach*. Journal of the American Statistical Association, 96(453):20–31.
- ROTHENBERG, THOMAS J. (1971). *Identification in Parametric Models*. Econometrica, 39(3):577–591.
- ROTHENBERG, THOMAS J. (1973). *Efficient estimation with a priori information*. Yale University Press, New Haven, London.
- RUDOLPH, BETTINA (1998). *Kundenzufriedenheit im Industriegüterbereich*. Gabler, Wiesbaden.
- RUST, ROLAND T., K. N. LEMON und V. A. ZEITHAML (2004). *Return on Marketing: Using Customer Equity to Focus Marketing Strategy*. Journal of Marketing, 68:109–127.

- RUST, ROLAND T. und A. J. ZAHORIK (1993). *Customer satisfaction, customer retention, and market share*. Journal of Retailing, 69(2):193–215.
- RUYTER, KO DE und J. BLOEMER (1999). *Customer loyalty in extended service settings*. International Journal of Service Industry Management, 10(3):320.
- RYAN, MICHAEL J., T. BUZAS und V. RAMASWAMY (1995). *Making CSM a power tool*. Marketing Research, 7(3):10–16.
- RYAN, MICHAEL J., R. RAYNER und A. MORRISON (1999). *Diagnosing Customer Loyalty Drivers*. Marketing Research, 11(2):19–26.
- SAAL, FRANK E., R. G. DOWNEY und M. A. LAHEY (1980). *Rating the ratings: Assessing the psychometric quality of rating data*. Psychological Bulletin, 88(2):413–428.
- SACKETT, PAUL R. und J. R. J. LARSON (1990). *Research strategies and tactics in industrial and organizational psychology*. In: DUNNETTE, MARVIN D. und L. M. HOUGH, Hrsg.: *Handbook of industrial and organizational psychology*, Bd. 1, S. 419–489. Consulting Psychologists Press, Palo Alto, CA.
- SCHARNBACHER, KURT und G. KIEFER (2003). *Kundenzufriedenheit: Analyse, Messbarkeit und Zertifizierung*. Oldenbourg, München u.a., 3. Aufl.
- SCHOTT, BJÖRN H., R. N. HENSON, A. RICHARDSON-KLAVEHN, C. BECKER, V. THOMA, H.-J. HEINE und E. DÜZEL (2005). *Redefining implicit and explicit memory: The functional neuroanatomy of priming, remembering, and control of retrieval*. Proceedings of the National Academy of Sciences, 102(4):1257–1262.
- SCOTT, WALTER D. (1908). *The psychology of advertising in theory and practice*. Small, Maynard & Company, Boston, MA.
- SHAH, RACHNA und S. M. GOLDSTEIN (2006). *Use of structural equation modeling in operations management research: Looking back and forward*. Journal of Operations Management, 24(2):148–169.

- SHARMA, SUBBASH, R. W. NIEDRICH und G. DOBBINS (1999). *A Framework for Monitoring Customer Satisfaction: An Empirical Illustration*. *Industrial Marketing Management*, 28(3):231–243.
- SKRONDAL, ANDERS und S. RABE-HESKETH (2004). *Generalized latent variable modeling: Multilevel, longitudinal and structural equation models*. Chapman & Hall/CRC, Boca Raton (Fla.).
- SKRONDAL, ANDERS und S. RABE-HESKETH (2007). *Latent Variable Modelling: A Survey*. *Scandinavian Journal of Statistics*, 34(4):712–745.
- SON, YOUNG und M. OH (2006). *Bayesian Estimation of the Two-Parameter Gamma Distribution*. *Communications in Statistics: Simulation and Computation*, 35(2):285–293.
- SQUIRE, LARRY R. (1992). *Declarative and nondeclarative memory: Multiple brain systems supporting learning and memory*. *J. Cognitive Neuroscience*, 4(3):232–243.
- STAHL, HEINZ K., H. H. HINTERHUBER, S. A. FRIEDRICH VON DEN EICHEN und K. MATZLER (2009). *Kundenzufriedenheit und Kundenwert*. In: HINTERHUBER, HANS H., Hrsg.: *Kundenorientierte Unternehmensführung*, S. 248–266. Gabler, Wiesbaden.
- STAHL, HEINZ K., H. H. HINTERHUBER, S. A. FRIEDRICH EICHEN und K. MATZLER (2006). *Kundenzufriedenheit und Kundenwert*. In: HINTERHUBER, HANS H. und K. MATZLER, Hrsg.: *Kundenorientierte Unternehmensführung*, S. 221–240. Gabler, Wiesbaden.
- STAN, VALENTINA und G. SAPORTA (2005). *Customer satisfaction and PLS structural equation modeling. An application to automobile market*. In: JANSSEN, JACQUES und P. LENCA, Hrsg.: *Applied Stochastic Models and Data Analysis*, S. 756–763. Groupe de linguistique appliquée des télécommunications.
- STANLEY, JULIAN (1961). *Analysis of unreplicated three-way classifications, with applications to rater bias and trait independence*. *Psychometrika*, 26(2):205–219.

- STAUSS, BERND (1999). *Kundenzufriedenheit*. Marketing ZFP, 21(1):5–24.
- STAUSS, BERND, F. DORNACH und C. COENEN (2006). *Zufriedenheitsmanagement in Deutschland: Von der Messung zur Unternehmenssteuerung*. Servicebarometer AG, München, 1. Aufl.
- STAUSS, BERND, F. DORNACH und C. COENEN (2009). *Zufriedenheitsmanagement — Konzept und Realisierung*. In: HINTERHUBER, HANS H. und K. MATZLER, Hrsg.: *Kundenorientierte Unternehmensführung*, S. 587–603. Gabler, Wiesbaden.
- STEENKAMP, JAN-BENEDICT E. M. und H. BAUMGARTNER (2000). *On the use of structural equation models for marketing modeling*. International Journal of Research in Marketing, 17(2-3):195–202.
- STELZL, INGBORG (1986). *Changing a causal hypothesis without changing the fit: Some rules for generating equivalent path models*. Multivariate Behavioral Research, 21(3).
- STERN, HAL S. und Y. JEON (2004). *Applying structural equation models with incomplete data*. In: RUBIN, DONALD B., A. GELMAN und X.-L. MENG, Hrsg.: *Applied Bayesian modeling and causal inference from incomplete-data perspectives*. Wiley, Chichester, u.a.
- STEVENS, PETE, B. KNUTSON und M. PATTON (1995). *Dineserv: A Tool for Measuring Service Quality in Restaurants*. Cornell Hotel and Restaurant Administration Quarterly, 36(2):56–60.
- STIGLER, STEPHEN M. (1983). *Who Discovered Bayes's Theorem?*. The American Statistician, 37(4):290–296.
- SWAN, JOHN E. und L. J. COMBS (1976). *Product Performance and Consumer Satisfaction: A New Concept*. The Journal of Marketing, 40(2):25–33.
- SYMONDS, PERCIVAL (1925). *Notes on Ratings*. Journal of Applied Psychology, 4:25–29.
- TAKANE, YOSHIO und J. D. LEEUW (1987). *On the relationship between item response theory and factor analysis of discretized variables*. Psychometrika, 52(3):393–408.

- TANNER, MARTIN A. und W. H. WONG (1987). *The Calculation of Posterior Distributions by Data Augmentation*. Journal of the American Statistical Association, 82(398):528–540.
- TANNER, MARTIN ABBA (1997). *Tools for statistical inference: Methods for the exploration of posterior distributions and likelihood functions*. Springer, New York, 3. Aufl.
- THORNDIKE, EDWARD L. (1920). *A constant error in psychological ratings*. Journal of Applied Psychology, 4(1):25–29.
- TIERNEY, LUKE (1994). *Markov Chains for Exploring Posterior Distributions*. The Annals of Statistics, 22(4):1701–1728.
- TING, SHUEH-CHIN und C.-N. CHEN (2002). *The asymmetrical and non-linear effects of store quality attributes on customer satisfaction*. Total Quality Management & Business Excellence, 13(4):547–569.
- TSE, DAVID K. und P. C. WILTON (1988). *Models of Consumer Satisfaction Formation: An Extension*. Journal of Marketing Research, 25(2):204–212.
- VAVRA, TERRY G. (1997). *Improving your measurement of customer satisfaction: A guide to creating, conducting, analyzing, and reporting customer satisfaction measurement programs*. ASQ Quality Press, Milwaukee, WI.
- WALKER, IAN und C. HULME (1999). *Concrete words are easier to recall than abstract words: Evidence for a semantic contribution to short-term serial recall*. Journal of Experimental Psychology: Learning, Memory, and Cognition, 25(5):1256–1271.
- WANOUS, JOHN P., A. E. REICHERS und M. J. HUDY (1997). *Overall Job Satisfaction: How Good Are Single-Item Measures?*. Journal of Applied Psychology, 82(2):247–252.
- WEIBER, ROLF und D. MÜHLHAUS (2010). *Strukturgleichungsmodellierung: Eine anwendungsorientierte Einführung in die Kausalanalyse mit Hilfe von AMOS, SmartPLS und SPSS*. Springer, Berlin.

- WELLS, FREDERIC L. (1907). *A Statistical Study of Literary Merit*. Archives of Psychology, (7):5–30.
- WESTBROOK, ROBERT A und R. L. OLIVER (1981). *Developing Better Measures of Consumer Satisfaction: Some Preliminary Results*. Advances in Consumer Research, 8(1):94–99.
- WESTBROOK, ROBERT A und M. D. REILLY (1983). *Value-Percept Disparity: An Alternative to the disconfirmation of Expectations Theory of Consumer Satisfaction*. Advances in Consumer Research, 10:256–261.
- WILCOX, JAMES B., R. D. HOWELL und E. BREIVIK (2008). *Questions about formative measurement: Formative Indicators - Formative Indicators*. Journal of Business Research, 61(12):1219–1228.
- WILD, PASCAL und W. R. GILKS (1992). *Adaptive Rejection Sampling from Log-concave Density Functions*. Journal of the Royal Statistical Society: Series C (Applied Statistics), 42(4):701.
- WILLINGHAM, WARREN W. und M. B. JONES (1958). *On the Identification of Halo Through Analysis of Variance*. Educational and Psychological Measurement, 18(2):403–407.
- WIRTZ, JOCHEN (2000). *An examination of the presence, magnitude and impact of halo on consumer satisfaction measures*. Journal of Retailing and Consumer Services, 7(2):89–99.
- WIRTZ, JOCHEN (2003). *Halo in customer satisfaction measures: The role of purpose of rating, number of attributes and customer involvement*. International Journal of Service Industry Management, 14(1):96.
- WIRTZ, JOCHEN und J. E. BATESON (1995). *An experimental investigation of halo effects in satisfaction measures of service attributes*. International Journal of Service Industry Management, 6(3):84.

- WIRTZ, JOCHEN und J. E. BATESON (1999). *Introducing uncertain performance expectations in satisfaction models for services*. International Journal of Service Industry Management, 10(1):82–99.
- WISNIEWSKI, MIK (2001). *Using SERVQUAL to assess customer satisfaction with public sector services*. Managing Service Quality, 11(6):380–388.
- WOLFE, RORY und D. FIRTH (2002). *Modelling Subjective Use of an Ordinal Response Scale in a Many Period Crossover Experiment*. Journal of the Royal Statistical Society. Series C (Applied Statistics), 51(2):245–255.
- WOODRUFF, ROBERT (1997). *Customer value: The next source for competitive advantage*. Journal of the Academy of Marketing Science, 25(2):139–153–153.
- WRIGHT, SEWALL (1921). *Correlation and Causation*. Journal of Agricultural Research, 20:557–585.
- WRIGHT, SEWALL (1923). *The Theory of Path Coefficients: A Reply to Niles's Criticism*. Genetics, 8(3):239–255.
- WYER, ROBERT S. und T. K. SRULL (1989). *Memory and cognition in its social context*. L. Erlbaum Associates, Hillsdale, NJ.
- YAP, KENNETH B. und J. C. SWEENEY (2007). *Zone-of-tolerance moderates the service quality-outcome relationship*. Journal of Services Marketing, 21(2):137–148.
- YI, YOUJAE (1990). *A Critical Review of Consumer Satisfaction*. In: ZEITHAML, VALERIE A., Hrsg.: *Review of marketing 1990*, Review of marketing. American Marketing Association.
- YING, YUANPING, F. FEINBERG und M. WEDEL (2006). *Leveraging Missing Ratings to Improve Online Recommendation Systems*. Journal of Marketing Research, 43(3):355–365.
- ZELLNER, ARNOLD (1970). *Estimation of Regression Relationships Containing Unobservable Independent Variables*. International Economic Review, 11(3):441–454.

ZELLNER, ARNOLD (1971). *An introduction to Bayesian inference in econometrics*. A Wiley-Interscience publication. Wiley, New York, NY.

ZHU, HONG TU und S. Y. LEE (Mar, 2001). *A Bayesian analysis of finite mixtures in the Lisrel model*. *Psychometrika*, 66(1):133–152.

A. Anhang

A.1. MCMC Sampler

Der MCMC-Sampler ist wie folgt konstruiert:

1. $[\Omega_k | \theta, \mathbf{c}, Z]$ - Factor Scores der latenten Konstrukte
2. $[\theta_k | \Omega, \mathbf{c}, Z]$ - Modellparameter des SEM
3. $[\mathbf{c}, Z | \theta, \Omega, Y]$ - Cut-Points und latente Messvariablen
4. Modell der Skalenverwendungsheterogenität mit vier Komponenten:
 - a) $[\{\tau_i\} | \{\theta_k\}, \mathbf{c}, Z, Y]$ - individuelle Skalenmittelwerte
 - b) $[\{\sigma_i\} | \{\theta_k\}, \mathbf{c}, Z, Y]$ - individuelle Skalenstandardabweichung
 - c) $[a_\sigma | \{\theta_k\}, \{\sigma_i\}, \mathbf{c}, Z, Y]$ - Heterogenität der Varianz
 - d) $[\sigma_\tau^2 | \{\theta_k\}, \{\tau_i\}, \mathbf{c}, Z, Y]$ - Heterogenität der Mittelwerte
5. $[S | \{\theta_k\}, \mathbf{c}, Y, \{\tau_i\}, \{\sigma_i\}, \alpha]$ - Allokation der Mixture-Komponenten

A.1.1. Latente Factor Scores

Der Zug der latenten Factor Scores $\{\omega_i\}$ erfolgt konditional auf die unkontaminierten latenten Messvariablen Z^* sowie den Parametern des LISREL-Modells θ :

$$p(\omega_i | Z_i^*, \theta) \propto N\left(\left(\Sigma_\omega^{-1} + \Lambda^T \Psi_\varepsilon^{-1} \Lambda\right)^{-1} \left(\Lambda^T \Psi_\varepsilon^{-1} (Z_i^* - \mu_z)\right), \left(\sigma_\omega^{-1} + \Lambda^T \Psi_\varepsilon^{-1} \Lambda\right)^{-1}\right) \quad (\text{A.1})$$

A.1.2. LISREL-Modellparameter

Messkoeffizienten Bei gegebenen Faktor Scores Ω entspricht der Zug der Messkoeffizientenmatrix Λ dem eines Regressionsmodells. Es sollen in diesem Schritt jedoch nur die freien Elemente der Messkoeffizientenmatrix Λ gezogen werden. Einige Elemente in Λ sind jedoch auf 0 oder 1 festgelegt. Λ_k^* enthält nur die freien Elemente der k -ten Zeile von Lambda. Freie und festgelegte Elemente werden in Matrix F festgehalten. F ist eine $p \times q$ Matrix mit Elementen f_{kj} , wobei $k \in \{1, \dots, p\}$, $j \in \{1, \dots, q\}$. Diese geben die freien Elemente in Λ an. Freie Elemente werden mit 1 gekennzeichnet, festgelegte mit 0. Festgelegt werden folgende Elemente:

- Analog zu LEE (2007, S. 95) wird die Identifikation des LISREL-Modells durch Restriktionen in Λ erreicht. In jeder Spalte von Λ wird dazu ein Koeffizient auf 1 festgelegt. In der Konsequenz übernimmt der zugehörige latente Faktor ω_j Mittelwert und Varianz der Messvariable. Sofern keine weiteren Informationen zur Verfügung stehen, wird stets das letzte von Null verschiedene Element in der jeweiligen Spalte von Λ auf 1 festgelegt.
- Besteht keine Beziehung zwischen z_{ik}^* and ω_{ij} , dann wird λ_{kj} auf 0 festgelegt.

Darüber hinaus sei Ω_k^* eine Submatrix von Ω , welche nur die für die Ziehung von Λ_k notwendigen Elemente enthält (angegeben in F_k).

Für die Ziehung der Messkoeffizienten Λ_k^* kann schließlich folgende Normalverteilung herangezogen werden:

$$\left[\Lambda_k^* | \tilde{Z}_k, \psi_{\varepsilon k}^{-1}, \Omega^* \right] \sim N(a_k, \psi_{\varepsilon k} A_k) \quad (\text{A.2})$$

mit

$$A_k = \left(\Omega^{*T} \Omega^* + (H_{0\Lambda k}^*)^{-1} \right)^{-1}$$

$$a_k = A_k \left[(\Lambda_{0k}^* H_{0\Lambda k}^*) + \left(\Omega^{*T} \tilde{Z}_k \right) \right]$$

Mittelwerte der Messvariablen Bei der Ziehung werden zusätzlich Baselines hinzugefügt, die separat als μ_Z bezeichnet werden.

Messfehlervarianzen Die Elemente der diagonalen Messfehlermatrix Ψ_ε werden aus der konjugaten inversen Gamma-Verteilung gezogen:

$$\left[\psi_{\varepsilon k}^{-1} | Y_k^*, \Omega^* \right] \sim \text{Gamma} \left(\frac{n}{2} + \alpha_{0\varepsilon k}, \beta_{\varepsilon k} + \frac{1}{2} b_k \right) \quad (\text{A.3})$$

mit

$$\beta_{\varepsilon k} = \beta_{0\varepsilon k} + 2^{-1} b_k$$

$$b_k = \tilde{Z}_k^T \tilde{Z}_k - \tilde{Z}_k^T \Omega^* \left(\Omega^{*T} \Omega^* \right)^{-1} \Omega^{*T} \tilde{Z}_k = \tilde{Z}_k^T \tilde{Z}_k - \tilde{Z}_k^T a_k^T A_k^{-1} a_k$$

Strukturkoeffizienten $\Lambda_\omega = [\text{III}]$ sei die Matrix aller Strukturkoeffizienten. Für jedes Individuum (oder jede Antwort) i enthält der zugehörige Vektor der latenten Faktoren $\omega_i = [\eta_i \xi_i]$ endogene und exogene latente Faktoren. Nun sei Ω_k^* eine Submatrix von Ω

welche nur die Spalten enthält, auf die sich die k -te Zeile von Λ_ω bezieht. Als Indexmatrix dient $L_\omega = l_{\omega kj}$ für $k = 1, \dots, q_1$, $j = 1, \dots, q$. $l_{\omega kj} = 0$ weist auf festgelegte Parameter hin, während $l_{\omega kj} = 1$ auf freie Parameter deutet.

Die konjugaten Prior-Verteilungen für Λ_ω und Ψ_δ sind

$$[\Lambda_{\omega k}^* | \psi_{\delta k}] \sim N(\Lambda_{\omega 0k}^*, \psi_{\delta k} \mathbf{H}_{0\omega k}^*) \quad (\text{A.4})$$

bzw.

$$[\psi_{\delta k}^{-1}] \sim \text{Gamma}(\alpha_{0\delta k}, \beta_{0\delta k}) \quad (\text{A.5})$$

Die Posteriori-Verteilung von $\Lambda_{\omega k}^*$ stellt sich somit wie folgt dar:

$$[\Lambda_{\omega k}^* | \Omega_{\eta k}^*, \psi_{\delta k}^{-1}, \Omega^*] \sim N(c_M, \psi_{\delta k} C_M^{-1}) \quad (\text{A.6})$$

mit

$$\begin{aligned} C_M &= (\Omega^{*T} \Omega^* + (\mathbf{H}_{0\omega k}^*)^{-1})^{-1} \\ c_M &= C_M \times (\Omega^{*T} \Omega_{\eta k}^* + \mathbf{H}_{0\omega k}^{*-1} \Lambda_{\omega 0k}^*) \end{aligned} \quad (\text{A.7})$$

Strukturfehlervarianzen Die Elemente der diagonalen Strukturfehlervarianz Ψ_δ der endogenen latenten Konstrukte werden aus einer konjugaten inversen Gammaverteilung gezogen:

$$[\psi_{\delta k}^{-1} | \Omega_{\eta k}^*, \Omega^*] \sim \text{Gamma}\left(\frac{n}{2} + \alpha_{0\delta k}, d_k\right) \quad (\text{A.8})$$

mit

$$d_k = \beta_{0\delta k} + 2^{-1} (\Omega_{\eta k}^{*T} \Omega_{\eta k}^* - c_M^T C_M^{-1} c_M + \Lambda_{0\omega k}^T \mathbf{H}_{0\omega k}^{-1} \Lambda_{0\omega k}) \quad (\text{A.9})$$

Kovarianz der exogenen Konstrukte Für die Kovarianz der exogenen Konstrukte wird eine konjugate inverse Wishart Verteilung mit Parametern ν_o und V_o angenommen. Die Posteriori-Verteilung von Φ ergibt sich somit als inverse Wishart-Verteilung mit folgenden Parametern:

$$p(\Phi | \xi) \sim IW(\nu_o + n, V_o + \Omega_\xi^T \Omega_\xi) \quad (\text{A.10})$$

A.1.3. Cut-Point-Modell

Da es sich bei der Matrix Ψ_ε um eine diagonale Matrix handelt und die Cut-Points unabhängig vom Item konstant sind, lässt sich die konditionale Verteilung des Cut-Point-Vektors wie folgt ausdrücken:

$$p(\mathbf{c}|Y_k, \theta, \Omega, \{\tau_i\}, \{\sigma_i\}) \propto \prod_{i=1}^n \prod_{k=1}^p \left\{ \Phi^* \left[(\sigma_i \psi_{\varepsilon k})^{-1/2} (c_{y_{ik+1}} - m_{ik}) \right] - \Phi^* \left[(\sigma_i \psi_{\varepsilon k})^{-1/2} (c_{y_{ik}} - m_{ik}) \right] \right\} \quad (\text{A.11})$$

$$\text{mit } m_{ik} = (\mu_k + \Lambda_k \omega_i^T + \tau_i) \sigma_i$$

Cut-Points Da es keine konjugate Prior-Verteilung gibt, wird der MH-Algorithmus eingesetzt. Es wird eine gleichverteilte Prior-Verteilung angenommen. Der Annahmequotient lässt sich wie folgt darstellen:

$$R = \prod_{k=1}^p \prod_{i=1}^n \frac{\Phi \left((\sigma_i^2 \psi_{\varepsilon k})^{-0.5} [c_{y_{ik+1}} - m_{ik}] \right) - \Phi \left((\sigma_i^2 \psi_{\varepsilon k})^{-0.5} [c_{y_{ik}} - m_{ik}] \right)}{\Phi \left((\sigma_i^2 \psi_{\varepsilon k})^{-0.5} [\tilde{c}_{y_{ik+1}} - m_{ik}] \right) - \Phi \left((\sigma_i^2 \psi_{\varepsilon k})^{-0.5} [\tilde{c}_{y_{ik}} - m_{ik}] \right)} \times \prod_{d=2}^{b-1} \frac{\Phi ([\tilde{c}_{d+1} - \tilde{c}_d] / \sigma_c) - \Phi ([c_{d-1} - \tilde{c}_d] / \sigma_c)}{\Phi ([c_{d+1} - c_d] / \sigma_c) - \Phi ([\tilde{c}_{d-1} - c_d] / \sigma_c)} \quad (\text{A.12})$$

$$\text{mit } m_{ik} = (\mu_k + \Lambda_k \omega_i^T + \tau_i) \sigma_i$$

Φ steht in diesem Fall für die kumulative Standardnormalverteilung.

Latente Messvariablen Die konditionale Posteriori-Verteilung der latenten Messvariablen entspricht einer abgeschnittenen MVN. Da Ψ_ε jedoch eine diagonale Matrix ist, können für alle $k = \{1, \dots, p\}$ Items unabhängige Züge aus univariaten abgeschnittenen Normalverteilungen erfolgen:

$$p(Z_{ik} | \omega_i, \theta, c, y_{ik}, \tau_i, \sigma_i) = N_{[c_{\mathbf{z}_{yk}}, c_{\mathbf{z}_{yk+1}}]} \left((\mu_{zk} + \Lambda_{zk}^T \omega_i + \tau_i) \sigma_i, \sigma_i^2 \Psi_{\varepsilon k} \right) \quad (\text{A.13})$$

A.1.4. Skalenverwendungsheterogenitätsmodell

Die Implementierung des Skalenheterogenitätsmodell ähnelt der in HARDT und BUESCHKEN (2010).

Individuelle Mittelwerte $\{\tau_i\}$ Es wird folgende Prior-Verteilung der individuellen Mittelwerte $\{\tau_i\}$ angenommen:

$$p(\tau_i) \propto \sigma_l \tau^{-1} \exp\left(-2^{-1} (\tau_i - 0)^T \sigma_\tau^{-2} (\tau_i - 0)\right) \quad (\text{A.14})$$

Aufgrund der konjugaten Eigenschaft der Prior-Verteilung nimmt τ_i die konditionale Posteriori-Verteilung wieder die Form einer Normalverteilung an:

$$p(\tau_i|\cdot) \propto N\left(\left((\iota^T \Psi_\varepsilon^{-1} \iota) + \sigma_\tau^{-2}\right)^{-1} (\iota^T \Psi_\varepsilon^{-1} Z_i^+), \left((\iota^T \Psi_\varepsilon^{-1} \iota) + \sigma_\tau^{-2}\right)^{-1}\right) \quad (\text{A.15})$$

mit

$$Z_i^+ = Z/\sigma_i - \iota \mu_Z \quad (\text{A.16})$$

Dabei ist ι ein Spaltenvektor der Länge p , der nur 1en enthält.

Individuelle Standardabweichungen σ_i Züge aus der Posteriori-Verteilung von σ_i werden mit Hilfe des MH-Algorithmus erzeugt. Der Annahmequotient R ist in diesem Fall gegeben durch:

$$R_i = \prod_{k=1}^p \frac{\Phi\left((\sigma_i^2 \psi_{\varepsilon k})^{-0.5} [c_{y_{ik}+1} - m_{ik}]\right) - \Phi\left((\sigma_i^2 \psi_{\varepsilon k})^{-0.5} [c_{y_{ik}} - m_{ik}]\right)}{\Phi\left((\sigma_i^2 \psi_{\varepsilon k})^{-0.5} [c_{y_{ik}+1} - \tilde{m}_{ik}]\right) - \Phi\left((\sigma_i^2 \psi_{\varepsilon k})^{-0.5} [c_{y_{ik}} - \tilde{m}_{ik}]\right)} \times \frac{\varphi(\tilde{\sigma}_i - \sigma_i / \sigma_{\text{cand}}) I_{(0,\infty)}}{\varphi(\sigma_i - \tilde{\sigma}_i / \sigma_{\text{cand}}) I_{(0,\infty)}} \quad (\text{A.17})$$

mit

$$m_{ik} = (\mu_k + \Lambda_k \omega_i^T + \tau_i) \sigma_i \text{ und}$$

$$\tilde{m}_{ik} = (\mu_k + \Lambda_k \omega_i^T + \tau_i) \tilde{\sigma}_i \text{ (enthält den Kandidaten von } \sigma_i)$$

φ und Φ stehen für die Dichte bzw. kumulative Verteilung der Standardnormalverteilung

Verteilung der Mittelwerte Der Hyperparameter von τ_i , σ_τ^2 kann aus einer konjugaten Gammaverteilung gezogen werden:

$$\sigma_\tau^2 \sim \text{Gamma}\left(n/2 + \alpha_\tau, \tau_i^T \tau_i / 2 + \beta_\tau\right) \quad (\text{A.18})$$

Verteilung der Standardabweichungen Da b_σ auf einen Wert festgelegt wird, muss nur ein Hyperparameter von $\{\sigma_i\}$ geschätzt werden: a_σ . In Ermangelung konjugater Prior-Verteilungen kann auch hier der MH-Algorithmus eingesetzt werden. Der Annahmequotient ist in diesem Fall:

$$R = \frac{\prod_{i=1}^n \frac{b_\sigma^{\tilde{a}_\sigma}}{\Gamma(\tilde{a}_\sigma)} \sigma_i^{\tilde{a}_\sigma - 1} \exp(b_\sigma \sigma_i)}{\prod_{i=1}^n \frac{b_\sigma^{a_\sigma}}{\Gamma(a_\sigma)} \sigma_i^{a_\sigma - 1} \exp(b_\sigma \sigma_i)} \times \frac{\phi(\tilde{a}_\sigma - a_\sigma / \sigma_{\text{cand}}) I_{(0, \infty)}}{\phi(a_\sigma - \tilde{a}_\sigma / \sigma_{\text{cand}}) I_{(0, \infty)}} \quad (\text{A.19})$$

SON und OH (2006) zeigen jedoch, dass der derivative-free adaptive rejection -Algorithmus (GILKS, 1992) einen effizienten Weg darstellt. Zur Anwendung dieses Algorithmus genügt die unnormalisierte log-Dichte der Zielverteilung. Implementiert ist dieser Algorithmus im Paket *Runuran* (LEYDOLD und HÖRMANN, 2011).

Mehrere Antworten je Individuum Der Umgang mit Skala (Ja-Sagen / Nein-Sagen usw.) wird als individuenspezifisch angenommen. Alle Antworten eines Befragten weisen daher die gleichen Skalenverwendungsheterogenitätsparameter auf: τ_i und σ_i sind identisch für alle A_i Antworten eines Befragten. Allgemein stellt sich die Likelihood wie folgt dar:

$$p(z_{ik}^+ | \tau_i, \Sigma) \propto |\Sigma|^{-0.5} \exp(-0.5 (z_{ik}^+ - \tau_i \iota) \Sigma^{-1} (z_{ik}^+ - \tau_i \iota)) \quad (\text{A.20})$$

Berücksichtigt man nun, dass für i je A_i Antworten die gleichen Skalenverwendungsparameter gelten, so ergibt sich:

$$p(z_{ik}^+ | \tau_i, \Sigma) \propto \prod_{k=1}^{A_i} |\Sigma|^{-0.5} \exp\left(-0.5 \sum_{k=1}^{A_i} (z_{ik}^+ - \tau_i \iota) \Sigma^{-1} (z_{ik}^+ - \tau_i \iota)\right) \quad (\text{A.21})$$

Unter Verwendung der gleichen Prior-Verteilung (Formel A.14) ergibt sich somit folgende Verteilung:

$$p(\tau_i | \cdot) \propto N\left(\left(\sum_{k=1}^{A_i} \iota^T \Sigma^{-1} + \sigma_\tau^{-2}\right)^{-1} \left(\sum_{k=1}^{A_i} z_{ik}^{+T} \Sigma^{-1} \iota\right), \left(\sum_{k=1}^{A_i} \iota^T \Sigma^{-1} + \sigma_\tau^{-2}\right)^{-1}\right) \quad (\text{A.22})$$

Für die Ziehung von $\{\sigma_i\}$ werden die entsprechenden Likelihood-Bestandteile der zusammengehörigen Beobachtungen einfach aufsummiert. Die Verteilungen von $\{\tau_i\}$ und $\{\sigma_i\}$ berücksichtigen, dass es nur ein Parameter Tupel τ_i, σ_i je Befragtem gibt.

A.1.5. Finite Mixture

Als Prior-Verteilung der Gruppenzugehörigkeit wird eine konjugate Dirichlet-Verteilung angenommen:

$$Dir(\alpha = \{\alpha_1, \dots, \alpha_k\}) \quad (\text{A.23})$$

Somit ergibt sich als Posteriori-Verteilung:

$$p(S_i = k | y_i, \{\theta_k\}, \mathbf{c}, \tau_i, \sigma_i, \alpha) = \frac{p(y_i | \theta_k, \mathbf{c}, \tau_i, \sigma_i) \alpha_k}{p(y_i | \theta_1, \mathbf{c}, \tau_i, \sigma_i) \alpha_1 + \dots + p(y_i | \theta_k, \mathbf{c}, \tau_i, \sigma_i) \alpha_k} \quad (\text{A.24})$$

Da die Beobachtungen Y auf ordinalem Skalenniveau erfolgten, lässt sich die Wahrscheinlichkeit $p(y_i | \theta_k, \mathbf{c}, \tau_i, \sigma_i)$ nur approximativ durch den GHK-Simulator bestimmen. Dieser ist in im Paket *bayesm* in der C-Funktion *ghk* implementiert (ROSSI, 2011).

A.2. Prior-Spezifikation

Für die empirischen Modelle wird folgende a-priori-Annahmen getroffen:

LISREL-Parameter Bezüglich der Pfad- und Messkoeffizienten werden schwach informative Prior-Verteilungen angenommen: $\sim N(0, 100)$

Auch bezüglich der Mess- und Strukturfehlervarianzen werden schwach informative Prior-Verteilungen angenommen: $\sim IG(1, 1)$

Aufgrund der konjugaten Eigenschaft wird eine Inverse Wishart-Verteilung mit folgenden Parametern als Prior-Verteilung der Kovarianz der exogenen Konstrukte angenommen: $\sim IW(\nu, N)$ mit $\nu = 3 + q_2$, $N = \nu I_{q_2}$

Cut-Points Es ist keine konjugate Prior-Verteilung für die Cut-Points bekannt. An dieser Stelle wird aus Gründen der Einfachheit eine gleichverteilte Prior-Verteilung angenommen.

Skalenheterogenitätsmodell Die individuellen Varianzparameter sind in diesem Modell mit $\Gamma(a_{\sigma_i}, b_{\sigma_i})$ verteilt. Um Identifikation zu erreichen, muss b_{σ_i} auf einen Wert festgelegt werden, während a_{σ_i} geschätzt werden kann. Es wird angenommen, dass: $b_{\sigma_i} = 3$

Für a_{σ_i} wird folgende Prior-Verteilung angenommen: $p(a_{\sigma_i}) = N(3, 10)$ Die Annahme eines Mittelwertes von 3 bedeutet folgendes: Da $b_{\sigma_i} = 3$ festgelegt ist, führt dies zu einer

Zentrierung der Varianzparameter bei 1. Aufgrund der hohen Varianz von 10 ist dieser Prior nur geringfügig informativ.

A.3. Parameter-Recovery Simulationsstudie

Die Recovery weiterer Parameter der Simulationsstudie wird in den folgenden zwei Tabellen dargestellt. Sind PM, 2.5% und 97.5% identisch, so handelt es sich um festgelegte Werte.

Tabelle A.1: Recovery Messmodell

Parameter	wahr	PM	2.5%	97.5%	Parameter	wahr	PM	2.5%	97.5%
Messkoeffizienten					Q_23 ← H	0.92	0.89	0.85	0.93
Komponente 1					Q_3 ← A	1.00	1.00	1.00	1.00
Q_1 ← A	0.69	0.69	0.66	0.72	Q_6 ← B	1.00	1.00	1.00	1.00
Q_2 ← A	0.69	0.74	0.70	0.77	Q_9 ← C	1.00	1.00	1.00	1.00
Q_10 ← D	0.58	0.58	0.55	0.62	Q_12 ← D	1.00	1.00	1.00	1.00
Q_11 ← D	0.64	0.66	0.62	0.69	Q_15 ← E	1.00	1.00	1.00	1.00
Q_19 ← G	0.57	0.57	0.54	0.60	Q_18 ← F	1.00	1.00	1.00	1.00
Q_20 ← G	1.01	1.04	0.99	1.09	Q_21 ← G	1.00	1.00	1.00	1.00
Q_22 ← H	0.83	0.81	0.77	0.85	Q_24 ← H	1.00	1.00	1.00	1.00
Q_23 ← H	0.92	0.93	0.89	0.97	Komponente 3				
Q_4 ← B	0.88	0.87	0.79	0.96	Q_1 ← A	0.69	0.68	0.65	0.71
Q_5 ← B	1.26	1.28	1.18	1.38	Q_2 ← A	0.69	0.66	0.63	0.70
Q_7 ← C	0.49	0.50	0.44	0.58	Q_4 ← B	0.88	0.86	0.74	0.99
Q_8 ← C	0.96	0.99	0.88	1.10	Q_5 ← B	1.26	1.38	1.21	1.57
Q_13 ← E	0.86	0.84	0.76	0.94	Q_7 ← C	0.49	0.44	0.33	0.55
Q_14 ← E	0.49	0.52	0.45	0.59	Q_8 ← C	0.96	0.89	0.72	1.06
Q_16 ← F	0.93	0.89	0.80	1.00	Q_10 ← D	0.58	0.56	0.45	0.67
Q_17 ← F	0.53	0.55	0.48	0.63	Q_11 ← D	0.64	0.58	0.47	0.70
Q_3 ← A	1.00	1.00	1.00	1.00	Q_13 ← E	0.86	0.84	0.71	0.97
Q_12 ← D	1.00	1.00	1.00	1.00	Q_14 ← E	0.49	0.52	0.42	0.63
Q_21 ← G	1.00	1.00	1.00	1.00	Q_16 ← F	0.93	0.88	0.73	1.03
Q_24 ← H	1.00	1.00	1.00	1.00	Q_17 ← F	0.53	0.51	0.41	0.62
Q_6 ← B	1.00	1.00	1.00	1.00	Q_19 ← G	0.57	0.55	0.44	0.67
Q_9 ← C	1.00	1.00	1.00	1.00	Q_20 ← G	1.01	1.02	0.88	1.18
Q_15 ← E	1.00	1.00	1.00	1.00	Q_22 ← H	0.83	0.88	0.75	1.02
Q_18 ← F	1.00	1.00	1.00	1.00	Q_23 ← H	0.92	0.92	0.78	1.06
Komponente 2					Q_3 ← A	1.00	1.00	1.00	1.00
Q_1 ← A	0.69	0.68	0.63	0.72	Q_6 ← B	1.00	1.00	1.00	1.00
Q_2 ← A	0.69	0.69	0.64	0.74	Q_9 ← C	1.00	1.00	1.00	1.00
Q_4 ← B	0.88	0.85	0.80	0.90	Q_12 ← D	1.00	1.00	1.00	1.00
Q_5 ← B	1.26	1.26	1.20	1.33	Q_15 ← E	1.00	1.00	1.00	1.00
Q_7 ← C	0.49	0.47	0.43	0.51	Q_18 ← F	1.00	1.00	1.00	1.00
Q_8 ← C	0.96	0.89	0.85	0.94	Q_21 ← G	1.00	1.00	1.00	1.00
Q_10 ← D	0.58	0.57	0.53	0.62	Q_24 ← H	1.00	1.00	1.00	1.00
Q_11 ← D	0.64	0.63	0.58	0.68	Messfehlervarianzen				
Q_13 ← E	0.86	0.85	0.81	0.90	Komponente 1				
Q_14 ← E	0.49	0.47	0.43	0.51	Psi_1	0.30	0.31	0.24	0.38
Q_16 ← F	0.93	0.94	0.88	1.00	Psi_2	0.30	0.34	0.26	0.42
Q_17 ← F	0.53	0.51	0.45	0.56	Psi_3	0.30	0.32	0.23	0.42
Q_19 ← G	0.57	0.53	0.48	0.59	Psi_4	0.30	0.29	0.23	0.36
Q_20 ← G	1.01	1.03	0.97	1.10	Psi_5	0.30	0.25	0.17	0.33
Q_22 ← H	0.83	0.82	0.78	0.86	Psi_6	0.30	0.35	0.28	0.43
					Psi_7	0.30	0.24	0.19	0.30
					Psi_8	0.30	0.41	0.32	0.52

Parameter	wahr	PM	2.5%	97.5%	Parameter	wahr	PM	2.5%	97.5%
Psi_9	0.30	0.26	0.18	0.35	Psi_3	0.30	0.32	0.21	0.44
Psi_10	0.30	0.34	0.27	0.42	Psi_4	0.30	0.28	0.23	0.34
Psi_11	0.30	0.39	0.31	0.49	Psi_5	0.30	0.21	0.14	0.29
Psi_12	0.30	0.29	0.20	0.39	Psi_6	0.30	0.30	0.24	0.37
Psi_13	0.30	0.25	0.18	0.33	Psi_7	0.30	0.26	0.22	0.32
Psi_14	0.30	0.29	0.23	0.35	Psi_8	0.30	0.32	0.24	0.41
Psi_15	0.30	0.33	0.24	0.43	Psi_9	0.30	0.28	0.20	0.37
Psi_16	0.30	0.32	0.24	0.41	Psi_10	0.30	0.30	0.25	0.37
Psi_17	0.30	0.33	0.27	0.40	Psi_11	0.30	0.30	0.24	0.36
Psi_18	0.30	0.28	0.19	0.37	Psi_12	0.30	0.21	0.14	0.30
Psi_19	0.30	0.29	0.23	0.37	Psi_13	0.30	0.28	0.22	0.35
Psi_20	0.30	0.34	0.23	0.45	Psi_14	0.30	0.25	0.21	0.31
Psi_21	0.30	0.40	0.30	0.52	Psi_15	0.30	0.23	0.17	0.31
Psi_22	0.30	0.32	0.25	0.40	Psi_16	0.30	0.34	0.26	0.42
Psi_23	0.30	0.32	0.24	0.42	Psi_17	0.30	0.23	0.19	0.29
Psi_24	0.30	0.33	0.24	0.43	Psi_18	0.30	0.23	0.17	0.31
Komponente 2					Psi_19	0.30	0.29	0.24	0.35
Psi_1	0.30	0.27	0.22	0.33	Psi_20	0.30	0.24	0.17	0.31
Psi_2	0.30	0.31	0.25	0.37	Psi_21	0.30	0.29	0.22	0.37
Psi_3	0.30	0.32	0.24	0.40	Psi_22	0.30	0.28	0.22	0.35
Psi_4	0.30	0.31	0.25	0.38	Psi_23	0.30	0.31	0.24	0.39
Psi_5	0.30	0.28	0.22	0.37	Psi_24	0.30	0.27	0.20	0.35
Psi_6	0.30	0.29	0.23	0.36	Mittelwerte				
Psi_7	0.30	0.32	0.27	0.39	Komponente 1				
Psi_8	0.30	0.30	0.24	0.37	mu_1	-0.21	-0.18	-0.35	-0.02
Psi_9	0.30	0.23	0.17	0.29	mu_2	-0.15	-0.12	-0.31	0.04
Psi_10	0.30	0.32	0.26	0.38	mu_3	-0.25	-0.15	-0.40	0.06
Psi_11	0.30	0.30	0.24	0.37	mu_4	-0.20	-0.18	-0.29	-0.07
Psi_12	0.30	0.28	0.21	0.36	mu_5	-0.18	-0.19	-0.32	-0.05
Psi_13	0.30	0.31	0.25	0.39	mu_6	-0.01	-0.02	-0.14	0.11
Psi_14	0.30	0.38	0.31	0.46	mu_7	0.26	0.30	0.21	0.39
Psi_15	0.30	0.29	0.21	0.37	mu_8	0.12	0.13	0.00	0.25
Psi_16	0.30	0.28	0.22	0.36	mu_9	-0.18	-0.15	-0.27	-0.03
Psi_17	0.30	0.34	0.28	0.42	mu_10	-0.04	-0.11	-0.27	0.04
Psi_18	0.30	0.25	0.19	0.32	mu_11	-0.29	-0.30	-0.47	-0.14
Psi_19	0.30	0.31	0.25	0.37	mu_12	-0.21	-0.17	-0.42	0.06
Psi_20	0.30	0.28	0.21	0.35	mu_13	0.11	0.07	-0.05	0.19
Psi_21	0.30	0.28	0.21	0.35	mu_14	-0.30	-0.29	-0.38	-0.19
Psi_22	0.30	0.26	0.20	0.33	mu_15	-0.26	-0.27	-0.41	-0.15
Psi_23	0.30	0.26	0.20	0.33	mu_16	-0.16	-0.17	-0.28	-0.06
Psi_24	0.30	0.33	0.25	0.43	mu_17	-0.27	-0.27	-0.36	-0.18
Komponente 3					mu_18	0.17	0.14	0.02	0.26
Psi_1	0.30	0.25	0.19	0.32	mu_19	-0.17	-0.12	-0.29	0.04
Psi_2	0.30	0.26	0.20	0.34	mu_20	0.15	0.22	-0.07	0.49

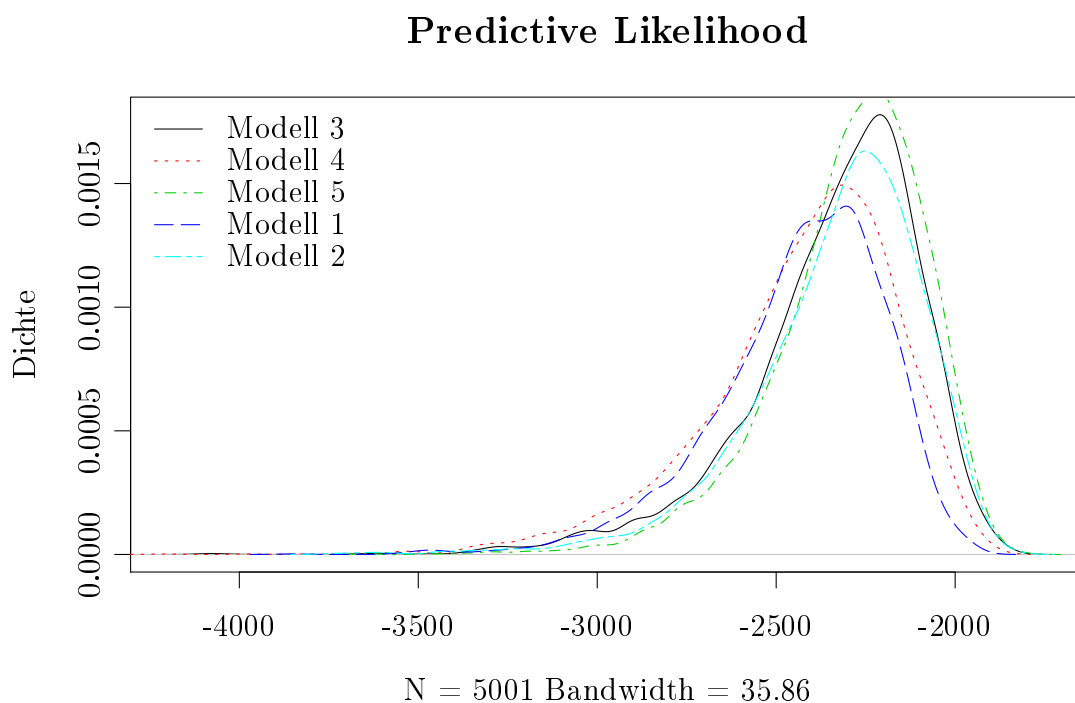
Parameter	wahr	PM	2.5%	97.5%	Parameter	wahr	PM	2.5%	97.5%
mu_21	0.22	0.35	0.07	0.61	mu_15	0.13	0.03	-0.09	0.13
mu_22	-0.12	-0.17	-0.38	0.02	mu_16	-0.04	-0.07	-0.18	0.03
mu_23	-0.13	-0.19	-0.44	0.03	mu_17	0.11	0.05	-0.04	0.14
mu_24	-0.08	-0.02	-0.27	0.22	mu_18	0.24	0.27	0.17	0.38
Komponente 2					mu_19	-0.16	-0.16	-0.26	-0.06
mu_1	0.05	-0.06	-0.19	0.10	mu_20	0.09	-0.01	-0.12	0.11
mu_2	0.13	0.13	-0.00	0.30	mu_21	-0.09	-0.11	-0.22	0.00
mu_3	-0.16	-0.19	-0.38	0.04	mu_22	0.06	0.05	-0.06	0.15
mu_4	-0.08	-0.15	-0.30	0.07	mu_23	0.27	0.22	0.11	0.33
mu_5	0.07	0.01	-0.20	0.33	mu_24	0.23	0.18	0.07	0.29
mu_6	0.26	0.24	0.07	0.50	<hr/>				
mu_7	0.14	0.14	0.03	0.26	Tabelle A.2: Recovery Strukturfehler und Kovarianzen exogener Konstrukte				
mu_8	0.05	-0.01	-0.18	0.23	Parameter	wahr	PM	2.5%	97.5%
mu_9	0.10	0.06	-0.12	0.32	<hr/>				
mu_10	0.04	0.07	-0.05	0.20	Strukturfehlervarianzen				
mu_11	-0.26	-0.26	-0.39	-0.12	<hr/>				
mu_12	0.28	0.29	0.12	0.51	Komponente 1				
mu_13	0.12	0.07	-0.11	0.32	1	1.00	1.00	1.00	1.00
mu_14	0.27	0.23	0.11	0.37	2	0.30	0.35	0.24	0.48
mu_15	-0.27	-0.23	-0.42	0.07	3	0.30	0.28	0.19	0.39
mu_16	0.08	0.05	-0.09	0.21	4	0.30	0.38	0.26	0.51
mu_17	-0.26	-0.27	-0.38	-0.18	5	0.30	0.36	0.24	0.50
mu_18	-0.19	-0.25	-0.40	-0.08	Komponente 2				
mu_19	-0.23	-0.23	-0.34	-0.13	1	0.20	0.23	0.16	0.30
mu_20	-0.24	-0.24	-0.39	-0.05	2	0.20	0.15	0.11	0.20
mu_21	-0.19	-0.25	-0.40	-0.07	3	0.20	0.19	0.13	0.26
mu_22	0.21	0.17	-0.04	0.49	4	0.20	0.23	0.16	0.31
mu_23	-0.12	-0.09	-0.32	0.26	5	0.20	0.19	0.13	0.26
mu_24	-0.02	-0.03	-0.29	0.35	6	0.20	0.25	0.18	0.31
Komponente 3					7	0.20	0.23	0.17	0.30
mu_1	-0.16	-0.17	-0.42	0.06	8	0.20	0.29	0.20	0.39
mu_2	0.22	0.17	-0.07	0.41	Komponente 3				
mu_3	0.02	-0.00	-0.37	0.32	1	0.40	0.49	0.13	1.27
mu_4	-0.06	-0.06	-0.16	0.05	Kovarianz (exogene Konstrukte)				
mu_5	-0.23	-0.24	-0.36	-0.13	<hr/>				
mu_6	0.10	0.02	-0.09	0.12	Komponente 1				
mu_7	-0.20	-0.25	-0.35	-0.17	Phi_1_1	1.05	0.93	0.76	1.13
mu_8	-0.22	-0.27	-0.38	-0.16	Phi_2_1	0.05	0.05	-0.06	0.17
mu_9	-0.27	-0.37	-0.48	-0.26	Phi_3_1	0.05	-0.06	-0.17	0.05
mu_10	0.18	0.13	0.02	0.22	Phi_4_1	0.05	-0.07	-0.18	0.05
mu_11	0.07	0.03	-0.08	0.13	Phi_2_2	1.05	1.02	0.83	1.25
mu_12	0.11	0.05	-0.06	0.16	Phi_3_2	0.05	-0.00	-0.13	0.13
mu_13	0.13	0.02	-0.09	0.12	Phi_4_2	0.05	0.07	-0.05	0.20
mu_14	0.20	0.12	0.02	0.22	Phi_3_3	1.05	1.10	0.88	1.36

Parameter	wahr	PM	2.5%	97.5%
Phi_4_3	0.05	0.03	-0.10	0.17
Phi_4_4	1.05	1.03	0.83	1.26
Komponente 2				
Phi_1_1	1.00	1.00	1.00	1.00
Komponente 3				
Phi_1_1	0.50	0.39	0.31	0.49
Phi_2_1	0.00	0.03	-0.02	0.09
Phi_3_1	0.00	-0.01	-0.06	0.05
Phi_4_1	0.00	0.03	-0.02	0.09
Phi_5_1	0.00	-0.01	-0.07	0.04
Phi_6_1	0.00	-0.03	-0.08	0.02
Phi_7_1	0.00	-0.01	-0.06	0.04
Phi_2_2	0.50	0.47	0.37	0.59
Phi_3_2	0.00	0.00	-0.06	0.07
Phi_4_2	0.00	0.01	-0.05	0.07
Phi_5_2	0.00	-0.03	-0.09	0.03
Phi_6_2	0.00	0.01	-0.05	0.07
Phi_7_2	0.00	0.02	-0.04	0.07
Phi_3_3	0.50	0.52	0.42	0.64
Phi_4_3	0.00	-0.03	-0.10	0.03
Phi_5_3	0.00	0.03	-0.03	0.09
Phi_6_3	0.00	-0.03	-0.09	0.03
Phi_7_3	0.00	0.00	-0.06	0.06
Phi_4_4	0.50	0.47	0.37	0.58
Phi_5_4	0.00	0.04	-0.02	0.09
Phi_6_4	0.00	-0.03	-0.09	0.03
Phi_7_4	0.00	-0.01	-0.07	0.05
Phi_5_5	0.50	0.46	0.37	0.57
Phi_6_5	0.00	-0.02	-0.08	0.04
Phi_7_5	0.00	-0.01	-0.07	0.04
Phi_6_6	0.50	0.45	0.35	0.56
Phi_7_6	0.00	0.00	-0.05	0.06
Phi_7_7	0.50	0.45	0.36	0.56

A.4. Predictive Fit

Zur Berechnung des Predictive Fit wurden zunächst 50 Befragte zufällig ausgewählt. Anschließend wurde je eine der maximal vier möglichen Antworten zum „Holdout-Sample“ zugeordnet. Diese 50 Fälle werden nicht zu Schätzung, sondern nur zu Berechnung der Predictive Likelihood verwendet. Die Modelle 3 und 5 (mit MIMIC-Komponente) haben mehr Masse bei höheren Likelihood-Werten, sind folglich auch prädiktiv überlegen.

Abbildung A.1: Predictive Likelihood



A.5. Zusammenfassungen der Posteriori-Verteilungen aller Parameter

Die Posteriori-Verteilungen der Modelle 1-3 werden in den folgenden Tabellen durch PM sowie zwei Quantile beschrieben. Sind PM, 2.5% und 97.5% identisch, so handelt es sich um festgelegte Werte.

Tabelle A.3: Ergebnisse Modell 1

	PM	2.5%	97.5%		PM	2.5%	97.5%
				Phi_8_1	1.15	0.96	1.35
				Phi_9_1	0.82	0.66	1.01
				Phi_10_1	1.16	0.96	1.39
Mixture				Phi_11_1	0.77	0.62	0.94
k=1 (MIMIC)	1.00	1.00	1.00	Phi_12_1	0.82	0.64	1.01
k=2 (Halo)	1.00	1.00	1.00	Phi_13_1	0.53	0.38	0.71
k=3 (Regression)	1.00	1.00	1.00	Phi_14_1	0.26	0.14	0.39
Cut-Points				Phi_15_1	0.55	0.39	0.73
cut_1	-2.54	-2.54	-2.54	Phi_2_2	0.77	0.56	1.00
cut_2	-1.76	-1.82	-1.71	Phi_3_2	0.22	0.12	0.34
cut_3	-1.10	-1.15	-1.04	Phi_4_2	0.22	0.12	0.33
cut_4	-0.30	-0.35	-0.26	Phi_5_2	0.19	0.07	0.33
cut_5	0.49	0.45	0.52	Phi_6_2	0.28	0.16	0.40
cut_6	1.51	1.51	1.51	Phi_7_2	0.15	0.00	0.31
Skalenverwendung				Phi_8_2	0.33	0.22	0.46
σ_τ^2	0.50	0.41	0.61	Phi_9_2	0.16	0.04	0.28
a_σ	2.15	1.98	2.33	Phi_10_2	0.17	0.04	0.31
b_σ	3.00	3.00	3.00	Phi_11_2	0.20	0.08	0.32
LISREL-Parameter				Phi_12_2	0.30	0.15	0.45
				Phi_13_2	0.04	-0.08	0.17
Pfadkoeffizienten				Phi_14_2	-0.04	-0.13	0.06
Gesamt \leftarrow Geschmack	0.40	0.29	0.51	Phi_15_2	0.10	-0.02	0.23
Gesamt \leftarrow Gesundheit	0.16	0.06	0.26	Phi_3_3	1.21	1.02	1.42
Gesamt \leftarrow Lieferservice	0.27	0.15	0.40	Phi_4_3	0.65	0.53	0.80
Gesamt \leftarrow Umfeld	0.15	0.04	0.27	Phi_5_3	0.93	0.76	1.11
Gesamt \leftarrow Ruf	0.10	0.03	0.16	Phi_6_3	0.72	0.57	0.88
Gesamt \leftarrow Menge	0.05	-0.02	0.12	Phi_7_3	-0.14	-0.30	0.04
Gesamt \leftarrow Preis	0.03	-0.01	0.07	Phi_8_3	0.70	0.56	0.85
Gesamt \leftarrow Wert	0.03	-0.06	0.11	Phi_9_3	0.51	0.38	0.66
Gesamt \leftarrow Portionen	0.01	-0.04	0.06	Phi_10_3	0.71	0.54	0.89
Gesamt \leftarrow Ausw.Belag	-0.00	-0.05	0.05	Phi_11_3	0.65	0.51	0.80
Gesamt \leftarrow Verp.Des.	-0.05	-0.11	0.02	Phi_12_3	0.72	0.56	0.89
Gesamt \leftarrow Ausw.Teig	0.02	-0.03	0.07	Phi_13_3	0.55	0.41	0.71
Gesamt \leftarrow Ausw.Sonst.	-0.02	-0.08	0.04	Phi_14_3	0.43	0.31	0.56
Gesamt \leftarrow Wartezeit	-0.01	-0.09	0.06	Phi_15_3	0.79	0.63	0.96
Gesamt \leftarrow Erreichb.	0.08	0.03	0.13	Phi_4_4	1.07	0.89	1.27
Strukturfehlervarianz				Phi_5_4	0.84	0.68	1.02
Gesamt	0.34	0.28	0.41	Phi_6_4	0.49	0.35	0.62
Phi				Phi_7_4	-0.02	-0.18	0.14
Phi_1_1	1.79	1.53	2.08	Phi_8_4	0.60	0.47	0.74
Phi_2_1	0.32	0.21	0.45	Phi_9_4	0.42	0.30	0.56
Phi_3_1	0.96	0.80	1.14	Phi_10_4	0.54	0.39	0.71
Phi_4_1	0.76	0.62	0.93	Phi_11_4	0.58	0.45	0.72
Phi_5_1	1.21	1.01	1.43	Phi_12_4	0.65	0.50	0.82
Phi_6_1	1.03	0.86	1.22				
Phi_7_1	-0.23	-0.43	-0.04				

	PM	2.5%	97.5%		PM	2.5%	97.5%
Phi_13_4	0.66	0.52	0.82	Phi_12_9	0.67	0.49	0.86
Phi_14_4	0.52	0.41	0.65	Phi_13_9	0.42	0.26	0.59
Phi_15_4	0.54	0.39	0.70	Phi_14_9	0.13	0.01	0.26
Phi_5_5	2.15	1.85	2.47	Phi_15_9	0.43	0.27	0.61
Phi_6_5	0.81	0.64	1.00	Phi_10_10	2.53	2.18	2.91
Phi_7_5	-0.57	-0.79	-0.36	Phi_11_10	0.81	0.64	1.01
Phi_8_5	0.77	0.60	0.95	Phi_12_10	0.79	0.59	1.02
Phi_9_5	0.79	0.61	0.98	Phi_13_10	0.65	0.46	0.86
Phi_10_5	1.05	0.85	1.28	Phi_14_10	0.20	0.06	0.35
Phi_11_5	0.76	0.60	0.94	Phi_15_10	0.50	0.30	0.70
Phi_12_5	1.01	0.81	1.22	Phi_11_11	1.63	1.40	1.88
Phi_13_5	0.65	0.48	0.84	Phi_12_11	0.82	0.65	1.01
Phi_14_5	0.36	0.23	0.51	Phi_13_11	0.75	0.59	0.92
Phi_15_5	1.02	0.81	1.24	Phi_14_11	0.23	0.12	0.36
Phi_6_6	1.71	1.47	1.97	Phi_15_11	0.46	0.31	0.63
Phi_7_6	-0.38	-0.58	-0.20	Phi_12_12	2.51	2.19	2.85
Phi_8_6	0.63	0.48	0.78	Phi_13_12	0.96	0.77	1.18
Phi_9_6	0.63	0.48	0.80	Phi_14_12	0.30	0.16	0.45
Phi_10_6	1.08	0.88	1.30	Phi_15_12	0.80	0.60	1.00
Phi_11_6	0.77	0.62	0.94	Phi_13_13	1.93	1.68	2.23
Phi_12_6	0.62	0.45	0.80	Phi_14_13	0.59	0.45	0.74
Phi_13_6	0.67	0.52	0.84	Phi_15_13	0.54	0.37	0.72
Phi_14_6	0.17	0.05	0.30	Phi_14_14	1.22	1.05	1.41
Phi_15_6	0.44	0.28	0.61	Phi_15_14	0.42	0.28	0.57
Phi_7_7	3.30	2.88	3.75	Phi_15_15	2.26	1.97	2.57
Phi_8_7	0.51	0.32	0.70	Messkoeffizienten			
Phi_9_7	-0.17	-0.36	0.03	Q_2 ←Geschmack	1.28	1.18	1.38
Phi_10_7	-0.66	-0.91	-0.43	Q_3 ←Geschmack	0.92	0.85	1.01
Phi_11_7	-0.33	-0.52	-0.14	Q_5 ←Gesundheit	1.17	0.92	1.47
Phi_12_7	-0.47	-0.70	-0.25	Q_7 ←Lieferservice	0.60	0.52	0.69
Phi_13_7	-0.35	-0.56	-0.15	Q_9 ←Umfeld	0.80	0.69	0.91
Phi_14_7	0.22	0.06	0.38	Q_10 ←Umfeld	0.93	0.84	1.03
Phi_15_7	0.02	-0.19	0.23	Q_1 ←Gesamt	1.00	1.00	1.00
Phi_8_8	1.64	1.42	1.88	Q_4 ←Geschmack	1.00	1.00	1.00
Phi_9_8	0.54	0.39	0.71	Q_6 ←Gesundheit	1.00	1.00	1.00
Phi_10_8	0.56	0.39	0.75	Q_8 ←Lieferservice	1.00	1.00	1.00
Phi_11_8	0.47	0.33	0.62	Q_11 ←Umfeld	1.00	1.00	1.00
Phi_12_8	0.54	0.38	0.71	Q_12 ←Ruf	1.00	1.00	1.00
Phi_13_8	0.29	0.14	0.44	Q_13 ←Menge	1.00	1.00	1.00
Phi_14_8	0.31	0.19	0.44	Q_14 ←Preis	1.00	1.00	1.00
Phi_15_8	0.42	0.27	0.59	Q_15 ←Wert	1.00	1.00	1.00
Phi_9_9	1.92	1.66	2.20	Q_16 ←Portionen	1.00	1.00	1.00
Phi_10_9	0.94	0.75	1.15	Q_17 ←Ausw.Belag	1.00	1.00	1.00
Phi_11_9	0.60	0.46	0.76	Q_18 ←Verp.Des.	1.00	1.00	1.00

	PM	2.5%	97.5%		PM	2.5%	97.5%
Q_19 ←Ausw.Teig	1.00	1.00	1.00	mu_17	0.76	0.61	0.92
Q_20 ←Ausw.Sonst.	1.00	1.00	1.00	mu_18	0.05	-0.08	0.20
Q_21 ←Wartezeit	1.00	1.00	1.00	mu_19	-0.45	-0.60	-0.28
Q_22 ←Erreichb.	1.00	1.00	1.00	mu_20	0.31	0.17	0.46
Messfehler				mu_21	-0.04	-0.18	0.10
Psi_1	0.10	0.10	0.10	mu_22	-0.22	-0.36	-0.06
Psi_2	0.71	0.57	0.85				
Psi_3	0.82	0.70	0.96				
Psi_4	0.54	0.45	0.64				
Psi_5	0.72	0.46	0.94				
Psi_6	0.74	0.54	0.93				
Psi_7	0.78	0.67	0.89				
Psi_8	0.26	0.17	0.37				
Psi_9	0.92	0.79	1.06				
Psi_10	0.53	0.43	0.64				
Psi_11	0.58	0.48	0.69				
Psi_12	0.10	0.10	0.10				
Psi_13	0.10	0.10	0.10				
Psi_14	0.10	0.10	0.10				
Psi_15	0.10	0.10	0.10				
Psi_16	0.10	0.10	0.10				
Psi_17	0.10	0.10	0.10				
Psi_18	0.10	0.10	0.10				
Psi_19	0.10	0.10	0.10				
Psi_20	0.10	0.10	0.10				
Psi_21	0.10	0.10	0.10				
Psi_22	0.10	0.10	0.10				
Item-Mittelwerte							
mu_1	-0.03	-0.16	0.12				
mu_2	0.47	0.31	0.65				
mu_3	0.34	0.20	0.50				
mu_4	0.35	0.21	0.51				
mu_5	-1.00	-1.14	-0.84				
mu_6	-0.88	-1.03	-0.73				
mu_7	-0.22	-0.35	-0.08				
mu_8	-0.07	-0.21	0.08				
mu_9	-0.01	-0.15	0.13				
mu_10	-0.05	-0.18	0.09				
mu_11	-0.11	-0.24	0.04				
mu_12	0.54	0.40	0.70				
mu_13	0.21	0.07	0.35				
mu_14	-1.20	-1.39	-1.02				
mu_15	-0.27	-0.41	-0.12				
mu_16	0.24	0.10	0.39				

Tabelle A.4: Ergebnisse Modell 2

	PM	2.5%	97.5%	PM	2.5%	97.5%	
				Preis	2.40	1.77	3.04
				Wert	0.61	0.45	0.81
Mixture				Portionen	0.86	0.66	1.09
k=2 (Halo)	0.54	0.48	0.61	Ausw.Belag	1.17	0.92	1.47
k=3 (Regression)	0.46	0.39	0.52	Verp.Des.	0.65	0.48	0.83
Cut-Points				Ausw.Teig	0.92	0.71	1.14
cut_1	-2.54	-2.54	-2.54	Ausw.Sonst.	0.75	0.55	0.96
cut_2	-1.73	-1.78	-1.68	Wartezeit	0.60	0.45	0.76
cut_3	-1.07	-1.11	-1.00	Erreichb.	0.68	0.50	0.88
cut_4	-0.29	-0.33	-0.24	Messkoeffizienten			
cut_5	0.49	0.45	0.53	Q_2 ← Geschmack	1.29	1.08	1.50
cut_6	1.51	1.51	1.51	Q_3 ← Geschmack	1.09	0.93	1.27
Skalenverwendung				Q_5 ← Gesundheit	1.14	0.87	1.49
σ_τ^2	0.58	0.47	0.70	Q_7 ← Lieferservice	0.80	0.67	0.94
a_σ	2.08	1.91	2.26	Q_9 ← Umfeld	1.03	0.88	1.21
b_σ	3.00	3.00	3.00	Q_10 ← Umfeld	0.84	0.69	1.00
LISREL-Parameter				Q_1 ← Gesamt	1.00	1.00	1.00
Komponente 1 (k=2, Halo)				Q_4 ← Geschmack	1.00	1.00	1.00
Pfadkoeffizienten				Q_6 ← Gesundheit	1.00	1.00	1.00
Pfadkoeffizienten	0.00	0.00	0.00	Q_8 ← Lieferservice	1.00	1.00	1.00
Gesamt ← GE	1.09	0.93	1.25	Q_11 ← Umfeld	1.00	1.00	1.00
Geschmack ← GE	0.98	0.81	1.15	Q_12 ← Ruf	1.00	1.00	1.00
Gesundheit ← GE	0.59	0.38	0.81	Q_13 ← Menge	1.00	1.00	1.00
Lieferservice ← GE	0.95	0.79	1.10	Q_14 ← Preis	1.00	1.00	1.00
Umfeld ← GE	0.94	0.79	1.11	Q_15 ← Wert	1.00	1.00	1.00
Ruf ← GE	1.04	0.86	1.24	Q_16 ← Portionen	1.00	1.00	1.00
Menge ← GE	0.81	0.65	0.98	Q_17 ← Ausw.Belag	1.00	1.00	1.00
Preis ← GE	0.43	0.13	0.73	Q_18 ← Verp.Des.	1.00	1.00	1.00
Wert ← GE	0.97	0.80	1.14	Q_19 ← Ausw.Teig	1.00	1.00	1.00
Portionen ← GE	0.77	0.61	0.93	Q_20 ← Ausw.Sonst.	1.00	1.00	1.00
Ausw.Belag ← GE	0.75	0.56	0.94	Q_21 ← Wartezeit	1.00	1.00	1.00
Verp.Des. ← GE	0.91	0.74	1.06	Q_22 ← Erreichb.	1.00	1.00	1.00
Ausw.Teig ← GE	0.93	0.75	1.12	Messfehler			
Ausw.Sonst. ← GE	0.89	0.71	1.08	Psi_1	0.10	0.10	0.10
Wartezeit ← GE	0.69	0.54	0.86	Psi_2	1.00	0.77	1.28
Erreichb. ← GE	0.78	0.61	0.97	Psi_3	0.56	0.41	0.73
Strukturfehler	0.00	0.00	0.00	Psi_4	0.58	0.44	0.73
Gesamt	0.27	0.19	0.37	Psi_5	0.69	0.44	0.96
Geschmack	0.21	0.14	0.30	Psi_6	0.81	0.59	1.05
Gesundheit	0.41	0.22	0.65	Psi_7	0.37	0.28	0.49
Lieferservice	0.21	0.13	0.31	Psi_8	0.35	0.23	0.48
Umfeld	0.14	0.09	0.21	Psi_9	0.53	0.41	0.68
Ruf	0.71	0.54	0.91	Psi_10	0.74	0.57	0.93
Menge	0.81	0.64	1.00	Psi_11	0.67	0.52	0.84

	PM	2.5%	97.5%		PM	2.5%	97.5%
Psi_12	0.10	0.10	0.10	Gesamt ← Portionen	-0.01	-0.09	0.07
Psi_13	0.10	0.10	0.10	Gesamt ← Ausw.Belag	0.10	0.00	0.20
Psi_14	0.10	0.10	0.10	Gesamt ← Verp.Des.	0.02	-0.08	0.11
Psi_15	0.10	0.10	0.10	Gesamt ← Ausw.Teig	0.00	-0.07	0.07
Psi_16	0.10	0.10	0.10	Gesamt ← Ausw.Sonst.	0.03	-0.06	0.14
Psi_17	0.10	0.10	0.10	Gesamt ← Wartezeit	0.10	-0.01	0.22
Psi_18	0.10	0.10	0.10	Gesamt ← Erreichb.	0.05	-0.01	0.12
Psi_19	0.10	0.10	0.10	Strukturfehler			
Psi_20	0.10	0.10	0.10	Gesamt	0.35	0.24	0.47
Psi_21	0.10	0.10	0.10	Φ			
Psi_22	0.10	0.10	0.10	Phi_1_1	2.01	1.56	2.52
Mittelwerte				Phi_2_1	0.22	0.05	0.41
mu_1	-0.31	-0.55	-0.07	Phi_3_1	0.90	0.65	1.18
mu_2	-0.29	-0.62	0.08	Phi_4_1	0.37	0.15	0.60
mu_3	-0.31	-0.57	-0.02	Phi_5_1	0.93	0.65	1.24
mu_4	-0.21	-0.47	0.07	Phi_6_1	0.86	0.62	1.14
mu_5	-0.93	-1.12	-0.73	Phi_7_1	-0.38	-0.73	-0.05
mu_6	-0.78	-0.98	-0.58	Phi_8_1	1.22	0.92	1.54
mu_7	-0.26	-0.45	-0.06	Phi_9_1	0.54	0.28	0.83
mu_8	-0.30	-0.52	-0.06	Phi_10_1	0.86	0.56	1.20
mu_9	-0.40	-0.63	-0.15	Phi_11_1	0.39	0.17	0.63
mu_10	-0.19	-0.41	0.03	Phi_12_1	0.78	0.48	1.11
mu_11	-0.25	-0.48	-0.03	Phi_13_1	-0.15	-0.42	0.10
mu_12	-0.04	-0.31	0.29	Phi_14_1	-0.29	-0.53	-0.07
mu_13	-0.31	-0.54	-0.06	Phi_15_1	0.36	0.05	0.68
mu_14	-0.78	-1.02	-0.49	Phi_2_2	0.77	0.56	1.04
mu_15	-0.43	-0.66	-0.20	Phi_3_2	-0.04	-0.18	0.10
mu_16	-0.18	-0.42	0.06	Phi_4_2	-0.05	-0.19	0.08
mu_17	-0.06	-0.31	0.19	Phi_5_2	-0.02	-0.19	0.16
mu_18	-0.28	-0.50	-0.04	Phi_6_2	0.12	-0.02	0.29
mu_19	-0.53	-0.76	-0.27	Phi_7_2	-0.06	-0.32	0.18
mu_20	-0.13	-0.37	0.12	Phi_8_2	0.11	-0.05	0.28
mu_21	-0.21	-0.42	-0.00	Phi_9_2	0.05	-0.13	0.25
mu_22	-0.35	-0.56	-0.13	Phi_10_2	0.01	-0.18	0.21
Komponente 3 (k=3, Regression)				Phi_11_2	-0.11	-0.27	0.06
Pfadkoeffizienten				Phi_12_2	0.04	-0.17	0.26
Gesamt ← Geschmack	0.51	0.34	0.69	Phi_13_2	-0.32	-0.50	-0.15
Gesamt ← Gesundheit	0.26	0.08	0.46	Phi_14_2	-0.32	-0.48	-0.17
Gesamt ← Lieferservice	0.20	0.03	0.37	Phi_15_2	-0.22	-0.44	-0.01
Gesamt ← Umfeld	0.06	-0.08	0.19	Phi_3_3	1.24	0.98	1.53
Gesamt ← Ruf	0.19	0.08	0.30	Phi_4_3	0.31	0.13	0.51
Gesamt ← Menge	0.03	-0.09	0.14	Phi_5_3	0.74	0.51	0.98
Gesamt ← Preis	0.10	0.02	0.17	Phi_6_3	0.57	0.37	0.79
Gesamt ← Wert	-0.03	-0.16	0.10	Phi_7_3	-0.55	-0.85	-0.27

	PM	2.5%	97.5%		PM	2.5%	97.5%
Phi_8_3	0.49	0.28	0.70	Phi_10_7	-1.07	-1.53	-0.67
Phi_9_3	0.37	0.16	0.60	Phi_11_7	-0.73	-1.06	-0.44
Phi_10_3	0.58	0.35	0.84	Phi_12_7	-1.10	-1.55	-0.67
Phi_11_3	0.35	0.15	0.55	Phi_13_7	-0.69	-1.05	-0.35
Phi_12_3	0.59	0.32	0.88	Phi_14_7	0.35	0.06	0.64
Phi_13_3	0.15	-0.07	0.38	Phi_15_7	-0.24	-0.64	0.17
Phi_14_3	0.20	-0.02	0.42	Phi_8_8	1.73	1.43	2.09
Phi_15_3	0.82	0.53	1.13	Phi_9_8	0.29	0.05	0.54
Phi_4_4	1.09	0.83	1.39	Phi_10_8	0.24	-0.01	0.52
Phi_5_4	0.59	0.39	0.82	Phi_11_8	0.04	-0.17	0.25
Phi_6_4	0.13	-0.04	0.31	Phi_12_8	0.28	0.01	0.57
Phi_7_4	-0.21	-0.48	0.06	Phi_13_8	-0.26	-0.49	-0.03
Phi_8_4	0.29	0.10	0.50	Phi_14_8	-0.02	-0.23	0.18
Phi_9_4	0.07	-0.13	0.27	Phi_15_8	0.15	-0.14	0.44
Phi_10_4	0.19	-0.02	0.41	Phi_9_9	2.45	2.00	2.96
Phi_11_4	0.27	0.10	0.46	Phi_10_9	0.92	0.61	1.28
Phi_12_4	0.49	0.25	0.75	Phi_11_9	0.38	0.14	0.62
Phi_13_4	0.45	0.25	0.68	Phi_12_9	0.79	0.47	1.15
Phi_14_4	0.41	0.22	0.61	Phi_13_9	-0.04	-0.30	0.22
Phi_15_4	0.43	0.17	0.70	Phi_14_9	-0.18	-0.42	0.05
Phi_5_5	2.13	1.74	2.58	Phi_15_9	0.56	0.24	0.90
Phi_6_5	0.46	0.23	0.71	Phi_10_10	2.43	1.94	3.01
Phi_7_5	-0.88	-1.27	-0.53	Phi_11_10	0.49	0.23	0.76
Phi_8_5	0.50	0.25	0.75	Phi_12_10	0.76	0.40	1.13
Phi_9_5	0.60	0.34	0.90	Phi_13_10	0.14	-0.14	0.43
Phi_10_5	0.73	0.45	1.07	Phi_14_10	-0.25	-0.49	-0.02
Phi_11_5	0.39	0.16	0.63	Phi_15_10	0.43	0.09	0.79
Phi_12_5	1.07	0.72	1.44	Phi_11_11	1.65	1.34	2.02
Phi_13_5	0.06	-0.20	0.33	Phi_12_11	0.67	0.39	0.97
Phi_14_5	-0.02	-0.26	0.21	Phi_13_11	0.43	0.20	0.69
Phi_15_5	1.20	0.85	1.59	Phi_14_11	-0.15	-0.36	0.06
Phi_6_6	1.61	1.30	1.95	Phi_15_11	0.25	-0.03	0.53
Phi_7_6	-0.50	-0.83	-0.21	Phi_12_12	3.53	2.90	4.26
Phi_8_6	0.34	0.13	0.56	Phi_13_12	1.10	0.76	1.50
Phi_9_6	0.51	0.28	0.76	Phi_14_12	0.01	-0.28	0.30
Phi_10_6	0.95	0.69	1.25	Phi_15_12	0.92	0.51	1.35
Phi_11_6	0.42	0.22	0.64	Phi_13_13	2.15	1.76	2.63
Phi_12_6	0.56	0.28	0.86	Phi_14_13	0.45	0.21	0.71
Phi_13_6	0.26	0.04	0.50	Phi_15_13	0.36	0.04	0.69
Phi_14_6	-0.21	-0.41	-0.02	Phi_14_14	1.47	1.19	1.81
Phi_15_6	0.25	-0.03	0.53	Phi_15_14	0.45	0.15	0.75
Phi_7_7	3.88	3.19	4.68	Phi_15_15	3.58	2.94	4.39
Phi_8_7	0.47	0.14	0.80	Messkoeffizienten			
Phi_9_7	-0.24	-0.60	0.10	Q_2 ← Geschmack	1.23	1.11	1.36

	PM	2.5%	97.5%		PM	2.5%	97.5%
Q_3 ← Geschmack	0.80	0.69	0.92	Mittelwerte			
Q_5 ← Gesundheit	1.30	0.99	1.62	mu_1	0.18	0.01	0.34
Q_7 ← Lieferservice	0.53	0.38	0.69	mu_2	1.09	0.83	1.31
Q_9 ← Umfeld	0.54	0.38	0.68	mu_3	0.91	0.69	1.12
Q_10 ← Umfeld	1.05	0.92	1.20	mu_4	0.80	0.58	1.00
Q_1 ← Gesamt	1.00	1.00	1.00	mu_5	-1.11	-1.31	-0.91
Q_4 ← Geschmack	1.00	1.00	1.00	mu_6	-1.01	-1.19	-0.83
Q_6 ← Gesundheit	1.00	1.00	1.00	mu_7	-0.23	-0.38	-0.07
Q_8 ← Lieferservice	1.00	1.00	1.00	mu_8	0.08	-0.08	0.25
Q_11 ← Umfeld	1.00	1.00	1.00	mu_9	0.29	0.14	0.46
Q_12 ← Ruf	1.00	1.00	1.00	mu_10	0.04	-0.11	0.19
Q_13 ← Menge	1.00	1.00	1.00	mu_11	-0.02	-0.17	0.14
Q_14 ← Preis	1.00	1.00	1.00	mu_12	1.00	0.77	1.20
Q_15 ← Wert	1.00	1.00	1.00	mu_13	0.65	0.44	0.83
Q_16 ← Portionen	1.00	1.00	1.00	mu_14	-1.67	-1.93	-1.35
Q_17 ← Ausw.Belag	1.00	1.00	1.00	mu_15	-0.18	-0.34	0.00
Q_18 ← Verp.Des.	1.00	1.00	1.00	mu_16	0.60	0.39	0.80
Q_19 ← Ausw.Teig	1.00	1.00	1.00	mu_17	1.48	1.21	1.77
Q_20 ← Ausw.Sonst.	1.00	1.00	1.00	mu_18	0.30	0.12	0.46
Q_21 ← Wartezeit	1.00	1.00	1.00	mu_19	-0.44	-0.67	-0.24
Q_22 ← Erreichb.	1.00	1.00	1.00	mu_20	0.66	0.49	0.84
Messfehler				mu_21	0.07	-0.09	0.24
Psi_1	0.10	0.10	0.10	mu_22	-0.15	-0.36	0.06
Psi_2	0.41	0.25	0.60				
Psi_3	1.05	0.85	1.29				
Psi_4	0.51	0.36	0.67				
Psi_5	0.66	0.34	1.02				
Psi_6	0.73	0.47	0.97				
Psi_7	1.16	0.94	1.40				
Psi_8	0.25	0.13	0.40				
Psi_9	1.19	0.98	1.43				
Psi_10	0.25	0.15	0.38				
Psi_11	0.55	0.40	0.73				
Psi_12	0.10	0.10	0.10				
Psi_13	0.10	0.10	0.10				
Psi_14	0.10	0.10	0.10				
Psi_15	0.10	0.10	0.10				
Psi_16	0.10	0.10	0.10				
Psi_17	0.10	0.10	0.10				
Psi_18	0.10	0.10	0.10				
Psi_19	0.10	0.10	0.10				
Psi_20	0.10	0.10	0.10				
Psi_21	0.10	0.10	0.10				
Psi_22	0.10	0.10	0.10				

Tabelle A.5: Ergebnisse Modell 3

	PM	2.5%	97.5%		PM	2.5%	97.5%
				Menge	1.03	0.79	1.32
				Preis	1.88	1.42	2.52
				Wert	0.55	0.35	0.80
Mixture				Portionen	1.51	1.19	1.90
k=1 (MIMIC)	0.36	0.31	0.41	Φ			
k=2 (Halo)	0.24	0.20	0.29	Phi_1_1	1.93	1.47	2.45
k=3 (Regression)	0.40	0.35	0.45	Phi_2_1	0.83	0.50	1.24
Cut-Points				Phi_3_1	1.02	0.66	1.46
cut_1	-2.54	-2.54	-2.54	Phi_4_1	0.72	0.37	1.15
cut_2	-1.68	-1.74	-1.64	Phi_5_1	0.51	0.19	0.94
cut_3	-1.01	-1.05	-0.96	Phi_6_1	0.61	0.29	1.04
cut_4	-0.25	-0.28	-0.21	Phi_2_2	1.93	1.49	2.44
cut_5	0.50	0.47	0.53	Phi_3_2	1.20	0.83	1.64
cut_6	1.51	1.51	1.51	Phi_4_2	1.07	0.70	1.53
Skalenverwendung				Phi_5_2	0.59	0.29	0.97
σ_{τ}^2	0.53	0.44	0.63	Phi_6_2	0.82	0.51	1.21
a_{σ}	2.08	1.91	2.25	Phi_3_3	2.39	1.84	3.01
b_{σ}	3.00	3.00	3.00	Phi_4_3	0.94	0.50	1.49
LISREL-Parameter				Phi_5_3	0.63	0.26	1.14
Komponente 1				Phi_6_3	1.07	0.72	1.50
Pfadkoeffizienten				Phi_4_4	2.29	1.73	2.92
Gesamt \leftarrow GE	1.00	1.00	1.00	Phi_5_4	1.05	0.71	1.45
Geschmack \leftarrow GE	0.86	0.66	0.97	Phi_6_4	0.87	0.52	1.31
Gesundheit \leftarrow GE	0.63	0.48	0.77	Phi_5_5	1.66	1.22	2.14
Lieferservice \leftarrow GE	0.80	0.69	0.91	Phi_6_5	0.54	0.28	0.87
Umfeld \leftarrow GE	0.70	0.59	0.84	Phi_6_6	1.53	1.19	1.93
Ruf \leftarrow GE	0.67	0.56	0.81	Messkoeffizienten			
Menge \leftarrow GE	0.59	0.48	0.71	Q_2 \leftarrow Geschmack	1.11	0.94	1.30
Preis \leftarrow GE	0.58	0.40	0.73	Q_3 \leftarrow Geschmack	0.95	0.80	1.13
Wert \leftarrow GE	0.86	0.69	0.96	Q_5 \leftarrow Gesundheit	1.05	0.88	1.26
Portionen \leftarrow GE	0.49	0.37	0.65	Q_7 \leftarrow Lieferservice	0.79	0.66	0.93
GE \leftarrow Ausw. Belag	0.10	-0.04	0.24	Q_9 \leftarrow Umfeld	0.99	0.81	1.21
GE \leftarrow Verp. Des.	0.19	0.02	0.34	Q_10 \leftarrow Umfeld	0.84	0.68	1.01
GE \leftarrow Ausw. Teig	0.25	0.03	0.38	Q_1 \leftarrow Gesamt	1.00	1.00	1.00
GE \leftarrow Ausw. Sonst.	0.03	-0.09	0.22	Q_4 \leftarrow Geschmack	1.00	1.00	1.00
GE \leftarrow Wartezeit	0.22	0.08	0.37	Q_6 \leftarrow Gesundheit	1.00	1.00	1.00
GE \leftarrow Erreichb.	0.28	0.13	0.50	Q_8 \leftarrow Lieferservice	1.00	1.00	1.00
Strukturfehlervarianzen				Q_11 \leftarrow Umfeld	1.00	1.00	1.00
GE	1.00	1.00	1.00	Q_12 \leftarrow Ruf	1.00	1.00	1.00
Gesamt	0.31	0.19	0.45	Q_13 \leftarrow Menge	1.00	1.00	1.00
Geschmack	0.24	0.15	0.36	Q_14 \leftarrow Preis	1.00	1.00	1.00
Gesundheit	0.50	0.32	0.72	Q_15 \leftarrow Wert	1.00	1.00	1.00
Lieferservice	0.20	0.11	0.31	Q_16 \leftarrow Portionen	1.00	1.00	1.00
Umfeld	0.33	0.19	0.51	Q_17 \leftarrow Ausw. Belag	1.00	1.00	1.00
Ruf	0.95	0.73	1.21				

	PM	2.5%	97.5%		PM	2.5%	97.5%
Q_18 ←Verp.Des.	1.00	1.00	1.00	mu_16	-0.34	-0.59	-0.06
Q_19 ←Ausw.Teig	1.00	1.00	1.00	mu_17	-0.07	-0.39	0.21
Q_20 ←Ausw.Sonst.	1.00	1.00	1.00	mu_18	-0.34	-0.58	-0.11
Q_21 ←Wartezeit	1.00	1.00	1.00	mu_19	-0.70	-0.94	-0.42
Q_22 ←Erreichb.	1.00	1.00	1.00	mu_20	0.08	-0.32	0.40
Messfehlervarianzen				mu_21	-0.05	-0.35	0.23
Psi_1	0.10	0.10	0.10	mu_22	-0.44	-0.65	-0.20
Psi_2	1.03	0.76	1.36	Komponente 2			
Psi_3	0.75	0.54	1.00	Pfadkoeffizienten			
Psi_4	0.58	0.41	0.80	Gesamt ←GE	0.03	-0.19	0.51
Psi_5	0.41	0.21	0.64	Geschmack ←GE	-0.27	-0.65	0.23
Psi_6	0.57	0.38	0.80	Gesundheit ←GE	-0.18	-0.38	0.17
Psi_7	0.59	0.45	0.77	Lieferservice ←GE	0.16	0.02	0.41
Psi_8	0.53	0.36	0.71	Umfeld ←GE	0.31	0.20	0.57
Psi_9	0.72	0.52	0.96	Ruf ←GE	0.15	-0.10	0.74
Psi_10	1.04	0.78	1.33	Menge ←GE	-0.15	-0.37	0.55
Psi_11	0.87	0.63	1.14	Preis ←GE	-0.81	-1.81	-0.37
Psi_12	0.10	0.10	0.10	Wert ←GE	-0.57	-0.83	-0.17
Psi_13	0.10	0.10	0.10	Portionen ←GE	-0.31	-0.58	0.02
Psi_14	0.10	0.10	0.10	Ausw.Belag ←GE	-0.42	-0.73	0.42
Psi_15	0.10	0.10	0.10	Verp.Des. ←GE	0.02	-0.22	0.52
Psi_16	0.10	0.10	0.10	Ausw.Teig ←GE	0.19	-0.09	0.84
Psi_17	0.10	0.10	0.10	Ausw.Sonst. ←GE	0.21	-0.01	0.57
Psi_18	0.10	0.10	0.10	Wartezeit ←GE	0.15	-0.02	0.35
Psi_19	0.10	0.10	0.10	Erreichb. ←GE	0.15	-0.12	0.42
Psi_20	0.10	0.10	0.10	Strukturfehlervarianzen			
Psi_21	0.10	0.10	0.10	Gesamt	0.27	0.15	0.41
Psi_22	0.10	0.10	0.10	Geschmack	0.18	0.10	0.28
Item-Mittelwerte				Gesundheit	0.44	0.17	1.02
mu_1	-0.52	-0.79	-0.20	Lieferservice	0.19	0.11	0.30
mu_2	-0.80	-1.11	-0.42	Umfeld	0.12	0.07	0.21
mu_3	-0.52	-0.76	-0.24	Ruf	0.63	0.42	0.88
mu_4	-0.53	-0.79	-0.24	Menge	0.56	0.36	0.79
mu_5	-0.92	-1.15	-0.70	Preis	2.23	1.43	3.12
mu_6	-0.68	-0.91	-0.45	Wert	0.41	0.21	0.66
mu_7	-0.25	-0.44	-0.02	Portionen	0.38	0.21	0.58
mu_8	-0.50	-0.74	-0.22	Ausw.Belag	0.89	0.55	1.25
mu_9	-0.49	-0.73	-0.22	Verp.Des.	0.37	0.23	0.53
mu_10	-0.16	-0.40	0.10	Ausw.Teig	0.61	0.40	0.85
mu_11	-0.29	-0.53	-0.02	Ausw.Sonst.	0.32	0.20	0.48
mu_12	-0.22	-0.49	0.08	Wartezeit	0.27	0.14	0.43
mu_13	-0.36	-0.59	-0.13	Erreichb.	0.81	0.54	1.14
mu_14	-0.63	-0.88	-0.39	Messkoeffizienten			
mu_15	-0.69	-0.97	-0.44	Q_2 ←Geschmack	1.40	0.65	2.34

	PM	2.5%	97.5%		PM	2.5%	97.5%
Q_3 ←Geschmack	1.39	0.66	2.29	Item-Mittelwerte			
Q_5 ←Gesundheit	1.59	0.57	2.71	mu_1	-0.26	-0.51	0.17
Q_7 ←Lieferservice	0.60	0.19	1.04	mu_2	0.37	-0.00	1.03
Q_9 ←Umfeld	-0.59	-1.15	-0.06	mu_3	0.11	-0.19	0.57
Q_10 ←Umfeld	1.03	0.68	1.42	mu_4	0.18	-0.14	0.79
Q_1 ←Gesamt	1.00	1.00	1.00	mu_5	-0.99	-1.29	-0.60
Q_4 ←Geschmack	1.00	1.00	1.00	mu_6	-1.01	-1.31	-0.67
Q_6 ←Gesundheit	1.00	1.00	1.00	mu_7	-0.30	-0.51	-0.11
Q_8 ←Lieferservice	1.00	1.00	1.00	mu_8	-0.29	-0.48	-0.01
Q_11 ←Umfeld	1.00	1.00	1.00	mu_9	-0.18	-0.44	0.16
Q_12 ←Ruf	1.00	1.00	1.00	mu_10	-0.32	-0.53	-0.07
Q_13 ←Menge	1.00	1.00	1.00	mu_11	-0.26	-0.48	0.07
Q_14 ←Preis	1.00	1.00	1.00	mu_12	0.21	-0.08	0.79
Q_15 ←Wert	1.00	1.00	1.00	mu_13	-0.24	-0.49	0.32
Q_16 ←Portionen	1.00	1.00	1.00	mu_14	-1.04	-1.48	-0.67
Q_17 ←Ausw.Belag	1.00	1.00	1.00	mu_15	-0.24	-0.52	0.20
Q_18 ←Verp.Des.	1.00	1.00	1.00	mu_16	0.07	-0.22	0.50
Q_19 ←Ausw.Teig	1.00	1.00	1.00	mu_17	0.15	-0.16	0.80
Q_20 ←Ausw.Sonst.	1.00	1.00	1.00	mu_18	-0.18	-0.43	0.22
Q_21 ←Wartezeit	1.00	1.00	1.00	mu_19	-0.51	-0.77	-0.05
Q_22 ←Erreichb.	1.00	1.00	1.00	mu_20	-0.19	-0.44	0.06
Messfehlervarianzen				mu_21	-0.29	-0.50	-0.07
Psi_1	0.10	0.10	0.10	mu_22	-0.41	-0.66	-0.13
Psi_2	0.67	0.38	1.00	Komponente 3			
Psi_3	0.39	0.20	0.64	Pfadkoeffizienten			
Psi_4	0.48	0.24	0.72	Gesamt ←Geschmack	0.63	0.38	0.83
Psi_5	0.79	0.21	1.57	Gesamt ←Gesundheit	0.35	0.08	0.56
Psi_6	1.08	0.49	1.58	Gesamt ←Lieferservice	0.26	0.01	0.49
Psi_7	0.22	0.13	0.33	Gesamt ←Umfeld	-0.03	-0.23	0.15
Psi_8	0.18	0.11	0.29	Gesamt ←Ruf	0.31	0.11	0.44
Psi_9	0.43	0.28	0.62	Gesamt ←Menge	0.00	-0.13	0.14
Psi_10	0.20	0.12	0.31	Gesamt ←Preis	0.17	0.05	0.25
Psi_11	0.19	0.11	0.30	Gesamt ←Wert	-0.04	-0.19	0.11
Psi_12	0.10	0.10	0.10	Gesamt ←Portionen	-0.06	-0.15	0.06
Psi_13	0.10	0.10	0.10	Gesamt ←Ausw.Belag	0.18	0.02	0.30
Psi_14	0.10	0.10	0.10	Gesamt ←Verp.Des.	0.09	-0.06	0.20
Psi_15	0.10	0.10	0.10	Gesamt ←Ausw.Teig	-0.01	-0.10	0.07
Psi_16	0.10	0.10	0.10	Gesamt ←Ausw.Sonst.	0.04	-0.07	0.16
Psi_17	0.10	0.10	0.10	Gesamt ←Wartezeit	0.07	-0.09	0.23
Psi_18	0.10	0.10	0.10	Gesamt ←Erreichb.	0.03	-0.04	0.11
Psi_19	0.10	0.10	0.10	Strukturfehlervarianzen			
Psi_20	0.10	0.10	0.10	Gesamt	0.21	0.11	0.34
Psi_21	0.10	0.10	0.10	Φ			
Psi_22	0.10	0.10	0.10	Phi_1_1	1.42	0.77	2.58

	PM	2.5%	97.5%		PM	2.5%	97.5%
Phi_2_1	0.19	-0.01	0.46	Phi_7_4	0.03	-0.32	0.39
Phi_3_1	0.48	0.07	1.25	Phi_8_4	0.42	0.17	0.74
Phi_4_1	0.37	0.10	0.78	Phi_9_4	0.13	-0.11	0.39
Phi_5_1	0.58	0.15	1.30	Phi_10_4	0.14	-0.13	0.44
Phi_6_1	0.53	0.11	1.20	Phi_11_4	0.38	0.15	0.62
Phi_7_1	-0.41	-0.79	-0.04	Phi_12_4	0.74	0.37	1.12
Phi_8_1	0.85	0.39	1.66	Phi_13_4	0.75	0.36	1.15
Phi_9_1	0.25	-0.12	0.81	Phi_14_4	0.67	0.35	0.98
Phi_10_1	0.45	-0.01	1.27	Phi_15_4	0.69	0.33	1.05
Phi_11_1	0.29	-0.04	0.79	Phi_5_5	1.83	1.33	2.44
Phi_12_1	0.64	0.21	1.24	Phi_6_5	0.24	-0.08	0.68
Phi_13_1	0.03	-0.24	0.31	Phi_7_5	-0.65	-1.10	-0.23
Phi_14_1	-0.06	-0.34	0.18	Phi_8_5	0.35	0.00	0.84
Phi_15_1	0.36	-0.03	0.85	Phi_9_5	0.47	0.10	0.91
Phi_2_2	0.78	0.52	1.13	Phi_10_5	0.40	-0.06	1.09
Phi_3_2	-0.07	-0.24	0.13	Phi_11_5	0.34	0.05	0.69
Phi_4_2	-0.03	-0.20	0.15	Phi_12_5	1.15	0.70	1.62
Phi_5_2	-0.04	-0.25	0.20	Phi_13_5	0.25	-0.13	0.63
Phi_6_2	0.09	-0.09	0.29	Phi_14_5	0.22	-0.18	0.58
Phi_7_2	-0.21	-0.56	0.12	Phi_15_5	1.39	0.94	1.88
Phi_8_2	0.07	-0.14	0.33	Phi_6_6	1.42	1.01	1.96
Phi_9_2	0.10	-0.12	0.33	Phi_7_6	-0.45	-0.84	-0.06
Phi_10_2	0.01	-0.20	0.27	Phi_8_6	0.10	-0.27	0.67
Phi_11_2	-0.10	-0.28	0.08	Phi_9_6	0.42	0.13	0.78
Phi_12_2	0.11	-0.17	0.41	Phi_10_6	0.69	0.32	1.20
Phi_13_2	-0.26	-0.51	-0.04	Phi_11_6	0.33	0.07	0.64
Phi_14_2	-0.27	-0.47	-0.09	Phi_12_6	0.56	0.18	0.98
Phi_15_2	-0.20	-0.47	0.07	Phi_13_6	0.38	0.06	0.69
Phi_3_3	0.99	0.65	1.55	Phi_14_6	-0.05	-0.35	0.20
Phi_4_3	0.40	0.21	0.63	Phi_15_6	0.30	-0.05	0.68
Phi_5_3	0.49	0.18	0.97	Phi_7_7	4.12	3.20	5.19
Phi_6_3	0.33	0.04	0.77	Phi_8_7	0.50	0.10	0.92
Phi_7_3	-0.29	-0.68	0.06	Phi_9_7	-0.13	-0.65	0.37
Phi_8_3	0.31	0.00	0.83	Phi_10_7	-0.92	-1.44	-0.43
Phi_9_3	0.31	0.09	0.58	Phi_11_7	-0.64	-1.03	-0.28
Phi_10_3	0.37	0.07	0.84	Phi_12_7	-1.15	-1.76	-0.52
Phi_11_3	0.30	0.07	0.60	Phi_13_7	-0.56	-1.02	-0.12
Phi_12_3	0.56	0.21	0.98	Phi_14_7	0.56	0.18	0.96
Phi_13_3	0.32	-0.00	0.61	Phi_15_7	-0.02	-0.58	0.58
Phi_14_3	0.50	0.12	0.80	Phi_8_8	1.59	1.14	2.22
Phi_15_3	0.98	0.64	1.37	Phi_9_8	0.16	-0.13	0.49
Phi_4_4	1.28	0.95	1.68	Phi_10_8	-0.07	-0.48	0.53
Phi_5_4	0.59	0.34	0.88	Phi_11_8	0.06	-0.24	0.46
Phi_6_4	0.15	-0.05	0.37	Phi_12_8	0.28	-0.11	0.72

	PM	2.5%	97.5%		PM	2.5%	97.5%
Phi_13_8	-0.03	-0.30	0.24	Q_13 ←Menge	1.00	1.00	1.00
Phi_14_8	0.25	-0.02	0.52	Q_14 ←Preis	1.00	1.00	1.00
Phi_15_8	0.23	-0.13	0.62	Q_15 ←Wert	1.00	1.00	1.00
Phi_9_9	2.52	1.94	3.22	Q_16 ←Portionen	1.00	1.00	1.00
Phi_10_9	0.70	0.25	1.31	Q_17 ←Ausw.Belag	1.00	1.00	1.00
Phi_11_9	0.32	-0.03	0.73	Q_18 ←Verp.Des.	1.00	1.00	1.00
Phi_12_9	1.00	0.56	1.48	Q_19 ←Ausw.Teig	1.00	1.00	1.00
Phi_13_9	0.21	-0.16	0.60	Q_20 ←Ausw.Sonst.	1.00	1.00	1.00
Phi_14_9	0.16	-0.30	0.55	Q_21 ←Wartezeit	1.00	1.00	1.00
Phi_15_9	0.81	0.35	1.29	Q_22 ←Erreichb.	1.00	1.00	1.00
Phi_10_10	2.00	1.31	3.10	Messfehlervarianzen			
Phi_11_10	0.34	0.03	0.69	Psi_1	0.10	0.10	0.10
Phi_12_10	0.61	0.12	1.19	Psi_2	0.32	0.17	0.50
Phi_13_10	0.29	-0.17	0.72	Psi_3	1.09	0.82	1.40
Phi_14_10	0.01	-0.44	0.39	Psi_4	0.48	0.32	0.67
Phi_15_10	0.50	0.08	0.97	Psi_5	0.74	0.36	1.15
Phi_11_11	1.67	1.28	2.16	Psi_6	0.62	0.35	0.90
Phi_12_11	0.75	0.38	1.15	Psi_7	0.74	0.45	1.08
Phi_13_11	0.60	0.20	0.99	Psi_8	0.43	0.26	0.61
Phi_14_11	0.01	-0.29	0.30	Psi_9	1.06	0.84	1.33
Phi_15_11	0.37	0.01	0.75	Psi_10	0.23	0.13	0.37
Phi_12_12	4.25	3.32	5.23	Psi_11	0.63	0.45	0.85
Phi_13_12	1.63	0.94	2.33	Psi_12	0.10	0.10	0.10
Phi_14_12	0.35	-0.17	0.81	Psi_13	0.10	0.10	0.10
Phi_15_12	1.29	0.71	1.94	Psi_14	0.10	0.10	0.10
Phi_13_13	2.55	1.93	3.24	Psi_15	0.10	0.10	0.10
Phi_14_13	0.59	0.24	0.96	Psi_16	0.10	0.10	0.10
Phi_15_13	0.61	0.09	1.13	Psi_17	0.10	0.10	0.10
Phi_14_14	1.67	1.31	2.06	Psi_18	0.10	0.10	0.10
Phi_15_14	0.78	0.31	1.24	Psi_19	0.10	0.10	0.10
Phi_15_15	4.48	3.45	5.66	Psi_20	0.10	0.10	0.10
Messkoeffizienten				Psi_21	0.10	0.10	0.10
Q_2 ←Geschmack	1.23	1.02	1.48	Psi_22	0.10	0.10	0.10
Q_3 ←Geschmack	0.74	0.53	0.97	Item-Mittelwerte			
Q_5 ←Gesundheit	1.23	0.88	1.65	mu_1	0.58	0.17	0.90
Q_7 ←Lieferservice	1.25	0.93	1.58	mu_2	1.57	0.96	2.00
Q_9 ←Umfeld	0.62	0.47	0.77	mu_3	1.26	0.87	1.60
Q_10 ←Umfeld	1.02	0.87	1.17	mu_4	1.22	0.71	1.59
Q_1 ←Gesamt	1.00	1.00	1.00	mu_5	-1.02	-1.29	-0.77
Q_4 ←Geschmack	1.00	1.00	1.00	mu_6	-0.93	-1.14	-0.71
Q_6 ←Gesundheit	1.00	1.00	1.00	mu_7	-0.08	-0.29	0.14
Q_8 ←Lieferservice	1.00	1.00	1.00	mu_8	0.47	0.13	0.75
Q_11 ←Umfeld	1.00	1.00	1.00	mu_9	0.55	0.26	0.81
Q_12 ←Ruf	1.00	1.00	1.00	mu_10	0.26	0.02	0.49

	PM	2.5%	97.5%
mu_11	0.20	-0.10	0.45
mu_12	1.39	0.91	1.71
mu_13	1.02	0.69	1.30
mu_14	-1.83	-2.21	-1.26
mu_15	0.11	-0.26	0.44
mu_16	0.90	0.55	1.18
mu_17	1.88	1.50	2.22
mu_18	0.58	0.29	0.84
mu_19	-0.16	-0.59	0.21
mu_20	0.87	0.59	1.14
mu_21	0.15	-0.05	0.36
mu_22	0.17	-0.16	0.49