

# Technische Universität Chemnitz

Steffen Jahn

**Strukturgleichungsmodellierung  
mit LISREL, AMOS und SmartPLS**  
Eine Einführung

WWDP 86/2007  
ISSN 1618-1352



TECHNISCHE UNIVERSITÄT  
CHEMNITZ

**Fakultät  
für  
Wirtschaftswissenschaften**



# Strukturgleichungsmodellierung mit LISREL, AMOS und SmartPLS

## Eine Einführung

Steffen Jahn

Professur für Marketing und Handelsbetriebslehre

Prof. Dr. Cornelia Zanger

Technische Universität Chemnitz

Kontakt:

(+49) 0371-531-35604

[steffen.jahn@wirtschaft.tu-chemnitz.de](mailto:steffen.jahn@wirtschaft.tu-chemnitz.de)

**Impressum:**

**Herausgeber:**

Der Dekan der  
Fakultät für Wirtschaftswissenschaften  
an der Technischen Universität Chemnitz

**Sitz:**

Reichenhainer Str. 39  
09126 Chemnitz

**Postanschrift:**

D-09107 Chemnitz  
Telefon: +49 (0)371 531 – 34208  
Telefax: +49 (0)371 531 – 26019  
E-Mail: [dekanat@wirtschaft.tu-chemnitz.de](mailto:dekanat@wirtschaft.tu-chemnitz.de)

**Internet:**

<http://www.tu-chemnitz.de/wirtschaft/>  
ISSN 1618-1352 (Print)  
ISSN 1618-1460 (Internet)

**Inhaltsverzeichnis**

<b>Abbildungsverzeichnis</b> .....	vi
<b>Tabellenverzeichnis</b> .....	vi
<b>Abkürzungsverzeichnis</b> .....	vii
<b>1 Einleitung</b> .....	1
<b>2 Aufbau eines Strukturgleichungsmodells</b> .....	2
2.1 Grundlagen .....	2
2.2 Das Messmodell .....	4
2.2.1 Empirische Indikatoren und ihre optimale Anzahl .....	4
2.2.2 Reflektive und formative Indikatoren .....	6
2.3 Das Strukturmodell .....	9
<b>3 Die Modellschätzung – ein Vergleich zweier Ansätze</b> .....	11
3.1 Die Kovarianzanalyse .....	11
3.2 Die Varianzanalyse .....	14
3.3 Hinweise zur Wahl des Ansatzes .....	15
<b>4 Modellbeurteilung</b> .....	17
4.1 Erste Betrachtung der Schätzergebnisse .....	17
4.2 Beurteilungskriterien .....	19
4.3 Gütemaße und deren Interpretation .....	20
4.3.1 Lokale Anpassungsmaße .....	21
4.3.1.1 Reflektive Messmodelle .....	21
4.3.1.2 Formative Messmodelle .....	23
4.3.2 Globale Gütemaße für LISREL und AMOS .....	23
4.3.2.1 Absolute Fitindizes .....	24
4.3.2.2 Inkrementelle Fitindizes .....	25
4.3.2.3 Indexselektion .....	25
4.3.3 Globale Gütemaße für SmartPLS .....	28
4.3.4 Hypothesenprüfung .....	30
<b>5 Fazit</b> .....	30
<b>Literaturverzeichnis</b> .....	vii

**Abbildungsverzeichnis**

Abbildung 1: Aufbau eines Strukturgleichungsmodells .....	4
Abbildung 2: Standardisiertes Strukturgleichungsmodell nach der Parameterschätzung.....	18

**Tabellenverzeichnis**

Tabelle 1: Entscheidungshilfe zur Zuordnung formativer vs. reflektiver Indikatoren .....	8
Tabelle 2: Reflektive vs. formative Messmodelle.....	9
Tabelle 3: Kovarianz- vs. Varianzanalyse – Hinweise zur Wahl des Ansatzes .....	16
Tabelle 4: t-Werte zur Signifikanzbeurteilung.....	19
Tabelle 5: Validitätsarten nach Homburg/Giering (1996) .....	20
Tabelle 6: Beurteilung reflektiver Messmodelle.....	22
Tabelle 7: Fitindizes, ihre Verfügbarkeit und Anforderungsniveaus.....	26
Tabelle 8: Fitindizes in LISREL und AMOS und Anforderungsniveaus .....	27
Tabelle 9: Gütemaße für SmartPLS .....	28

***Abkürzungsverzeichnis***

ADF	Asymptotically Distribution Free
AGFI	Adjusted Goodness of Fit-Index
AMOS	Analysis of Moment Structures
AVE	Average Variance Extracted (Durchschnittlich Erfasste Varianz)
BL89	Bollen's Fit Index
CFI	Comparative Fit Index
df	Degrees of Freedom (Freiheitsgrade)
GFI	Goodness of Fit-Index
IFI	Incremental Fit Index
LISREL	Linear Structural Relationships
LVPLS	Latent Variable Partial Least Squares
MANOVA	Multivariate Analysis of Variance
Mc	McDonald's Centrality Index
ML	Maximum Likelihood
NFI	Normed Fit Index
NNFI	Non-Normed Fit Index
NTLI	Normed Tucker-Lewis-Index
PLS	Partial Least Squares
RFI	Relative Fit Index
RMR	Root Mean (Square) Residual
RMSEA	Root Mean Square Error of Approximation
SAS PROC CALIS	Statistical Analysis System Procedure: Covariance Analysis of Linear Structural Equations
SEM	Structural Equation Model(ing)
SPAD PLS	Special Programs and Analysis Division: Partial Least Squares
SRMR	Standardized Root Mean Square Residual
TLI	Tucker-Lewis-Index
ULS	Unweighted Least Squares



## 1 Einleitung

Die **Strukturgleichungsmodellierung**<sup>1</sup> hat sich zu einem bedeutenden statistischen Verfahren<sup>2</sup> zur empirischen Hypothesenprüfung in verschiedenen verhaltens- und sozialwissenschaftlichen Disziplinen, darunter auch im Marketing, entwickelt. Es erlaubt die Überprüfung von Zusammenhängen zwischen hypothetischen Konstrukten auf eine komplexe Weise.

Innerhalb von SEM sind zwei Ansätze zu unterscheiden, die auf verschiedenen Herangehensweisen basieren: die **Kovarianzanalyse** sowie die **Varianzanalyse**. Die Kovarianzanalyse besitzt einen psychometrischen Hintergrund und kann unter Zuhilfenahme von Softwareprogrammen wie *LISREL* (Jöreskog/Sörbom 1996) oder *AMOS* (Arbuckle/Wothke 1999) durchgeführt werden. Als Synonym der stärker ökonometrisch orientierten Varianzanalyse gilt im Rahmen des SEM der **PLS-Ansatz**. Als relativ neue, anwenderfreundliche Software ist *SmartPLS* (Ringle/Wende/Will 2005) zu nennen. An diesen drei Applikationen soll der Prozess der Strukturgleichungsmodellierung in diesem Diskussionspapier näher beschrieben werden.

Dieser Beitrag möchte zum einen das Verständnis der Strukturgleichungsmodellierung verbessern und zum anderen Hilfestellung für eine dem jeweiligen Forschungsanliegen Rechnung tragende Wahl des Ansatzes (Kovarianz- vs. Varianzanalyse) geben. Das Diskussionspapier richtet sich an Studierende der Verhaltens- und Sozialwissenschaften in höheren Semestern sowie Nachwuchswissenschaftler. Es soll zum einen ein Ratgeber beim Anfertigen eigener Analysen sein und zum anderen die Interpretation wissenschaftlicher Arbeiten mit Strukturgleichungsmodellen erleichtern. Ein weiteres Diskussionspapier beschreibt als zusätzliche Unterstützung den Modellprüfungsprozess an je einem Beispiel für LISREL, AMOS und SmartPLS.

Wo es möglich ist, wird auf mathematische Formeln und Matrixalgebra verzichtet. Dadurch verringern sich zwar die Prägnanz und eventuell die Präzision der Ausführungen, allerdings wird eine verbesserte Verständlichkeit des Textes erhofft.

---

<sup>1</sup> Da auch im deutschsprachigen Raum der Begriff ‚Strukturgleichungsmodellierung‘ meist mit dem englischen Akronym ‚SEM‘ (structural equation modeling) abgekürzt wird, geschieht dies auch im vorliegenden Beitrag. Eine alternative Bezeichnung ist ‚Kausalanalyse‘, engl. ‚causal modeling‘.

<sup>2</sup> Einen Überblick zum Verhältnis der Strukturgleichungsmodellierung zu anderen Verfahren wie MANOVA und weiteren Themen gibt Hershberger (2003).

Das vorliegende Diskussionspapier ist wie folgt gegliedert: Zunächst werden Grundlagen sowie Bestandteile von Strukturgleichungsmodellen beschrieben, um auch Neueinsteigern die Möglichkeit zu geben, das Wesen dieser Methode zu verstehen. Eine wichtige theoretische Überlegung betrifft dabei so genannte reflektive und formative Indikatoren, auf die ausführlich im Abschnitt 2.2.2 eingegangen wird. Das Kapitel 3 stellt die beiden Ansätze, mit denen sich Strukturgleichungsmodelle schätzen und beurteilen lassen, gegenüber. Dort finden sich ebenfalls Entscheidungshilfen für die Wahl zwischen den Softwarepaketen LISREL und AMOS auf der einen sowie SmartPLS auf der anderen Seite. Im vierten Kapitel wird die Modellbeurteilung besprochen. Dies ist das zentrale Kapitel des Diskussionspapiers, da auf umfassende Art und Weise die Beurteilungskriterien für LISREL, AMOS und SmartPLS dargestellt werden.

Auf weiterführende Themenstellungen, wie hierarchische Faktoren, Multilevel, Zeitreihenmodelle oder die Betrachtung von Untergruppen wird an dieser Stelle aufgrund des Einführungscharakters in die Thematik verzichtet.

## 2 Aufbau eines Strukturgleichungsmodells

### 2.1 Grundlagen

In der Marketingforschung werden häufig kausale Wirkungsbeziehungen zwischen nicht direkt beobachtbaren hypothetischen Konstrukten, so genannten **latenten Variablen**<sup>3</sup>, untersucht (Jöreskog 1982). Dies ist notwendig, da „an understanding of causality and its role in the theory construction and research phases of scientific is essential if one is to avoid two of the most common faults in contemporary marketing research: blind empiricism and sterile tautologies“ (Bagozzi 1980, S. 29). Ein Verfahren zur Untersuchung kausaler Abhängigkeiten zwischen solchen (latenten) Variablen ist die Strukturgleichungsmodellierung. Über die Verknüpfung hypothetischer Konstrukte mit direkt messbaren (also **manifesten**) Größen – die als deren **Indikatoren** bezeichnet werden – können Kausalitäten zwischen latenten Variablen überprüft werden.

Grundsätzlich hilft SEM dabei, **ein theoretisch fundiertes Hypothesensystem dahingehend zu überprüfen**, „that the relationships you have hypothesized among the latent variables and between the latent variables and the manifest indicators are indeed consistent with the empirical data at hand“ (Diamantopoulos/Siguaw 2000, S. 4; *bei der Kovarianzanalyse*). Im Falle

---

<sup>3</sup> Eine umfassende Darstellung verschiedener Definitionen latenter Variablen findet sich in Bollen (2002).

der Varianzanalyse dient sie der **Ermittlung der Vorhersagerelevanz latenter Variablen** (Jöreskog/Wold 1982, S. 266; Lohmöller 1989, S. 72f.), also wie gut sich Konstruktausprägungen (z.B. hohe vs. niedrige Kaufbereitschaft) durch Ausprägungen anderer Konstrukte (z. B. Höhe der Zufriedenheit) prognostizieren lassen.

Die Form der Modellprüfung besitzt prinzipiell konfirmatorischen Charakter, die Strukturgleichungsmodellierung gehört also zu den **hypothesenprüfenden statistischen Verfahren** (Dependenzanalyse). Das Besondere an Strukturgleichungsmodellen ist, dass sie einerseits (im Unterschied bspw. zu multipler linearer Regression) komplexe Wirkungsbeziehungen simultan überprüfen können, andererseits die Ebene der Messmodelle (mehrere Indikatoren dienen der Messung eines hypothetischen Konstruktes) und dort explizit Messfehler berücksichtigen (Aaker/Bagozzi 1979; Jöreskog 1982). Daher lassen sich Strukturgleichungsmodelle auch als aus zwei Teilen bestehend darstellen<sup>4</sup> (Diamantopoulos/Siguaw 2000, S. 4; Jöreskog 1982, S. 84).

Zum einen gibt es das **Strukturmodell**<sup>5</sup>, in dem die aufgestellten Zusammenhänge zwischen den hypothetischen Konstrukten in einem Pfaddiagramm abgebildet werden. Zum anderen gibt es das **Messmodell**<sup>6</sup>, das die (empirischen) Indikatoren (also die direkt beobachtbaren, *manifesten Variablen*) für die latenten Größen abbildet und den Konstrukten zuordnet. Die Indikatoren dienen der Operationalisierung der nicht direkt beobachtbaren Konstrukte.

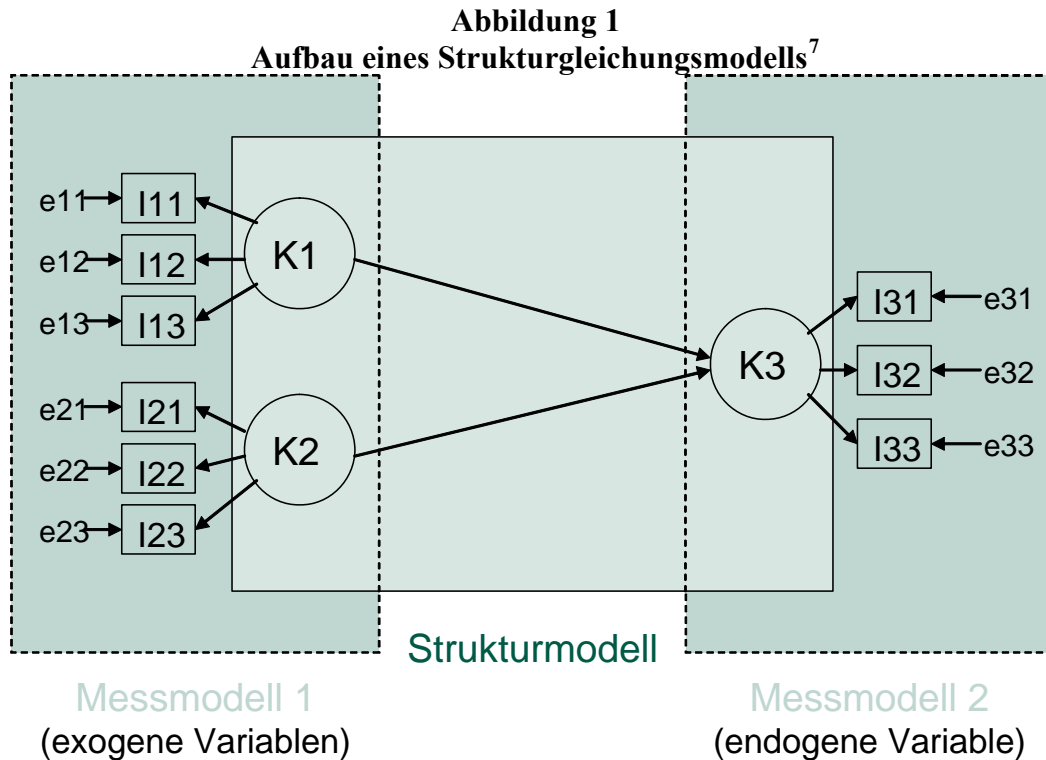
Ausgehend vom Pfaddiagramm werden Messmodelle der *latenten exogenen Variablen* sowie der *latenten endogenen Variablen* formuliert. Bei den exogenen Variablen handelt es sich um jene, die nur Ursache einer Kausalität sind (und folglich im Modell nicht erklärt werden können), während die endogenen diejenigen Variablen sind, die durch andere Größen beeinflusst werden. Variablen, die also ausschließlich unabhängig sind, werden als exogen bezeichnet. Alle anderen Konstrukte sind endogene Variablen. Abbildung 1 zeigt die schematische Darstellung eines Strukturgleichungsmodells zur Illustration. Führt in dieser Abbildung noch ein Pfeil von K1 auf K2, handelte es sich nur bei K1 um eine exogene Variable, während K2 und K3 als endogen bezeichnet würden.

---

<sup>4</sup> Da der LISREL-Ansatz dominiert, orientieren sich die Begriffe zur Beschreibung von Strukturgleichungsmodellen an ihm. Gegebenenfalls abweichende Bezeichnungen bei SmartPLS werden gekennzeichnet.

<sup>5</sup> In der (Smart)PLS-Terminologie als *inneres Modell* bezeichnet

<sup>6</sup> Dies wird im PLS-Ansatz meist *äußeres Modell* genannt



Kx – Konstrukt x (latente Variable) | Ixy – Indikator y für Konstrukt x (manifeste Variable) | e<sub>xy</sub> – Messfehler für Indikator xy

## 2.2 Das Messmodell

### 2.2.1 Empirische Indikatoren und ihre optimale Anzahl

Im Messmodell werden die latenten Variablen mit ihren jeweiligen Indikatoren abgebildet. Bezüglich der Anzahl solcher beobachtbaren Variablen zur Beschreibung eines Konstruktes besteht der Gegensatz darin, einerseits möglichst viele Indikatoren zu verwenden, ohne andererseits mehr Items als nötig zu benutzen. Mit der Zahl der Indikatoren steigt grundsätzlich die Reliabilität (siehe Abschnitt 4.2) der Messung. Außerdem benötigen Programme wie LISREL oder AMOS im Normalfall mindestens drei Indikatoren (eigentlich sogar vier), um ein Strukturgleichungsmodell auf seine Eignung testen zu können. Mit zunehmender Indikatorenzahl erhöht sich aber die Gefahr Artefakte und somit einen Unterfaktor zu produzieren, wodurch das Konstrukt in eine bestimmte Richtung gelenkt wird (Hair et al. 2006, S. 783).

<sup>7</sup> Typischerweise werden die Variablen, Messfehler und Pfade mit griechischen Buchstaben bezeichnet, die der Notation der zugrunde liegenden mathematischen Formeln entsprechen (LISREL Notation). Aus Gründen der besseren Überschaubarkeit wird in dieser Darstellung darauf verzichtet. Die Notationen finden sich aber bspw. bei Mueller (1996, S. xxi f.) oder Diamantopoulos/Siguaw (2000, S. 25).

<b>Beispiel</b>
-----------------

Die Einstellung einem Auto gegenüber könnte über die Indikatoren ‚attraktiv‘, ‚komfortabel‘, ‚sportlich‘, ‚sympathisch‘ und ‚umweltschonend‘ ermittelt werden. Kämen weitere Indikatoren wie ‚dynamisch‘, ‚schnittig‘ oder ‚tolles Design‘ hinzu, bestünde die Gefahr, dass ein Unterfaktor ‚Design‘ bzw. ‚Aussehen‘ gebildet würde, der die Interpretation der Einstellung beeinflusst.
--

Darüber hinaus steigt mit jedem Parameter, der in ein Strukturgleichungsmodell aufgenommen wird, die benötigte Stichprobengröße. Ein dritter Grund, der gegen zu viele Indikatoren spricht, ergibt sich aus forschungsökonomischen und Belastbarkeitsüberlegungen. Je mehr Items pro Konstrukt abgefragt werden, umso länger wird der entsprechende Fragebogen.

Aus den genannten Gründen gilt als Orientierungshilfe eine Mindestindikatorenzahl von drei, besser sind jedoch vier manifeste Variablen pro Konstrukt (Hair et al. 2006, S. 783). Zu beachten ist, dass die verwendeten Indikatoren grundsätzlich eindimensional sein müssen (Anderson/Gerbing/Hunter 1987, S. 432). Das bedeutet, dass sich die Indikatoren eines Konstruktes untereinander ähneln (intern konsistent sind) und nicht auch andere Konstrukte messen. Als Folge der geforderten Eindimensionalität sollten die ausgesuchten Indikatoren eines Konstruktes im Normalfall hoch miteinander korrelieren.

<b>Beispiel</b>
-----------------

In unserem Beispiel könnte der Einfluss der Umweltverträglichkeit von Automobilen auf die Einstellung gemessen werden. Die Umweltverträglichkeit wäre mit Hilfe der Indikatoren ‚verbraucht viel‘, ‚stößt viel CO <sub>2</sub> aus‘ oder ‚belastet die Umwelt‘ möglich. Nun besteht das Problem, dass der Einstellungs-Indikator ‚umweltschonend‘ auch direkt mit ‚Umweltverträglichkeit‘ zusammenhängt. Würde die Operationalisierung beibehalten, bestünde ein Mangel an Eindimensionalität.
--

Neben den zu einem Konstrukt gehörenden Indikatoren wird in einem Messmodell zusätzlich ein so genannter Messfehler berücksichtigt, der jeden einzelnen Indikator bzw. das Konstrukt an sich betrifft. Messfehler ergeben sich daraus, dass eine Erhebung stets von der ‚Realität‘ abweicht. Unproblematisch ist das Vorkommen von zufälligen Messfehlern, da diese um den ‚wahren Wert‘ streuen und ihn somit nicht verzerren. Systematische Fehler hingegen lenken den beobachteten Wert in eine bestimmte Richtung und gefährden damit die Interpretation.

<b>Beispiel</b>
-----------------

Die Probanden kreuzen beim Indikator ‚sympathisch‘ schlechtere Werte an, da sie bei der vorangegangenen Frage nach der Umweltverträglichkeit ebenfalls negativ antworteten und aus sozialem Druck heraus das Automobil als unsympathisch beurteilen (was aber nicht ihrer ‚wahren‘ Einstellung entspricht).
---

### 2.2.2 Reflektive und formative Indikatoren

Bedeutsam ist, zwei Grundtypen von Messmodellen zu unterscheiden – reflektive und formative (Edwards/Bagozzi 2000, S. 155). Zwar ist eine solche Unterteilung nicht neu, dennoch wird dieser Thematik erst in den letzten Jahren verstärkt Aufmerksamkeit geschenkt und es bestehen immer noch Defizite bezüglich des Einsatzes formativer Indikatoren<sup>8</sup> (Albers/Hildebrandt 2006; Jarvis/MacKenzie/Podsakoff 2003).

Ein Indikator ist **reflektiv**, wenn er durch die Konstruktausprägung bedingt wird. Dieser ‚klassische Fall‘ findet sich bspw. bei Einstellungen oder Involvement.

Beispiel
Wir erachten die Adjektive ‚attraktiv‘, ‚komfortabel‘, ‚sportlich‘, ‚sympathisch‘ und ‚umweltschonend‘ als geeignete Indikatoren für die Einstellung zum Automobil. Wenn sich nun unsere Einstellung ändert, führt das theoretisch dazu, dass wir das Auto nun auch als weniger attraktiv, sympathisch usw. beurteilen.

Da reflektive Indikatoren durch das entsprechende Konstrukt bestimmt werden, ändern sie sich, wenn sich die Konstruktausprägung ändert (Diamantopoulos 1999, S. 446). Dies liegt an der Forderung, dass reflektive Indikatoren, die mit einem Konstrukt verbunden werden, auch positiv miteinander korrelieren sollen. Eine solche Perspektive der **internen Konsistenz** ist dominierend in den Sozialwissenschaften. Aufgrund der geforderten hohen Korrelationen sind die Indikatoren (bei gleicher Reliabilität) austauschbar, d.h. die Elimination eines einzelnen Indikators verändert nicht das Wesen des zu Grunde liegenden Konstruktes (Bollen/Lennox 1991, S. 308; Diamantopoulos/Winklhofer 2001, S. 271).

Um eine Messskala auf diesen Annahmen basierend zu überprüfen und gegebenenfalls zu verbessern, ist die Bestimmung der *Item-to-Total-Korrelation* sinnvoll, um weniger geeignete Indikatoren aufzuspüren und zu eliminieren (Churchill 1979, S. 68; Hildebrandt/Temme 2006, S. 619f.). Obwohl die Reliabilitätswerte eines Indikatorensets (z.B. *Cronbachs Alpha*) kleiner werden, wenn sich die Itemzahl in einem Messmodell verringert, bleibt die Konstruktvalidität unverändert, wenn ein einzelner Indikator entfernt wird (der die interne Konsistenz eines Messmodells verschlechtert). Das bedeutet, dass die Qualität der Messung nicht zwangsläufig unter der Löschung eines Indikators leidet. Grund dafür ist, dass alle Facetten eines eindimensionalen Konstruktes durch die verbliebenen Indikatoren repräsentiert werden (Jarvis/MacKenzie/Podsakoff 2003, S. 200). Allerdings ist Vorsicht geboten bei rein mecha-

<sup>8</sup> D.h. es werden häufig Indikatoren als reflektiv behandelt, obwohl sie besser formativ zu spezifizieren wären.

nischer Itemreduktion aufgrund niedriger interner Konsistenz: „Die Sicherung der Inhaltsvalidität sollte daher der kausalanalytischen Reliabilitäts- und Validitätsprüfung in den Vordergrund gestellt werden“ (Hildebrandt/Temme 2006, S. 634). „Es ist diesbezüglich oft hilfreich, bereits im Vorfeld [...] inhaltlich bedeutsame Indikatoren zu identifizieren und diese auch beizubehalten, wenn dies die Konstruktreliabilität ungünstig beeinflusst“ (Homburg/Klarmann 2006, S. 732).

**Formative Indikatoren** unterscheiden sich von den bisher betrachteten reflektiven Indikatoren. Sie werden nicht durch die Konstruktausprägung bedingt, sondern bestimmen selbst das entsprechende Konstrukt: „When a latent variable is defined as a linear sum of a set of measurements [...], the measures are termed formative indicators: the measures produce the construct so to speak“ (Bagozzi 1994, S. 332). Typische Konstrukte mit formativen Indikatoren sind Erfolgsfaktoren, sozioökonomischer Status oder Zufriedenheit.

<b>Beispiel</b>
Die Zufriedenheit mit einem Automobil kann sich aus mehreren Teilzufriedenheiten ergeben. Je nachdem, wie zufrieden ich mit dem Design, dem Fahrverhalten, dem Kraftstoffverbrauch oder der Sicherheit bin, ergibt sich meine Gesamtzufriedenheit. Es ist durchaus vorstellbar, dass ein Individuum mit Design, Fahrverhalten und Sicherheit ist, allerdings sehr unzufrieden mit dem Benzinverbrauch. Die Indikatoren für die Gesamtzufriedenheit müssen folglich nicht notwendigerweise hoch miteinander korrelieren.

Eine Unterscheidung der Messmodelle dient aber nicht nur analytischen Zwecken, sondern wirkt sich fundamental auf das weitere Vorgehen aus. Im formativen Fall ist es gefährlich, einzelne Indikatoren zu löschen. Es kann nämlich passieren, dass zentrale Bestandteile des Konstruktes mit ihnen verloren gehen (Bollen/Lennox 1991, S. 308). Das ergibt sich daraus, dass die Konstruktausprägung aus der Indikatorenausprägung resultiert und somit jedes Item grundsätzlich wichtig für die Beschreibung des Konstruktes ist.

Interne Konsistenz spielt somit aufgrund der Logik formativer Indikatoren für diese keine Rolle, da die Items positiv, negativ oder gar nicht miteinander korrelieren können (Bollen 1984, S. 381; Nunally/Bernstein 1994, S. 489). Ebenso ist es schwierig, für sie die Reliabilität sowie die Konstruktvalidität zu ermitteln (Bagozzi 1994, S. 333; Bollen 1989, S. 222f.; Bollen/Lennox 1991, S. 307; Diamantopoulos/Winklhofer 2001, S. 271; Jarvis/MacKenzie/Podsakoff 2003, S. 200). Wichtiger als das Erreichen dieser klassischen Gütekriterien ist für formative Indikatoren, dass sie die gesamte Bandbreite der latenten Variablen abdecken (Diamantopoulos/Winklhofer 2001, S. 271). Damit entziehen sie sich auch der Forderung nach Eindimensionalität (Bollen/Lennox 1991, S. 312). Eine **Missspezifikation** (fälschliche Be-

trachtung formativer als reflektive Indikatoren) auch nur eines Konstruktes innerhalb eines Strukturgleichungsmodells kann ernsthafte Konsequenzen für die theoretischen Schlussfolgerungen haben, die aus dem Modell gezogen werden. Es könnten Modelle als ‚nicht gut genug‘ beurteilt werden, die eigentlich ‚wahr‘ sind (type I error genannt). Gleichzeitig wäre es möglich, dass ein Modell angenommen wird, obwohl es ‚falsch‘ ist (als type II error bezeichnet) (Jarvis/MacKenzie/Podsakoff 2003, S. 212).

Um die Zuordnung von Indikatoren als reflektiv oder formativ zu erleichtern, finden sich in Tabelle 1 vier Entscheidungshilfen, die auf die von Jarvis/MacKenzie/Podsakoff (2003, S. 203) beschriebenen Charakteristika formativer Indikatoren aufbauen.

**Tabelle 1**  
**Entscheidungshilfe zur Zuordnung formativer vs. reflektiver Indikatoren**

Charakteristik	Bedingung erfüllt?	
	ja	nein
1. Änderungen in den Indikatoren bedingen vermutlich Veränderungen im Konstrukt	<i>formativ</i>	<i>reflektiv</i>
2. Änderungen im Konstrukt bedingen vermutlich Veränderungen in den Indikatoren	<i>reflektiv</i>	<i>formativ</i>
3. Indikatoren haben nicht notwendigerweise das gleiche Thema zum Gegenstand	<i>formativ</i>	<i>reflektiv</i>
4. Eine Veränderung eines Wertes eines Indikators verändert vermutlich die Werte aller anderen Indikatoren	<i>reflektiv</i>	<i>formativ</i>

Bevor eine empirische Studie durchgeführt wird, muss die Frage nach der Klassifizierung der Indikatoren beantwortet sein, da bspw. das Befragungsdesign darauf abgestimmt werden muss. Nicht unerwähnt bleiben soll an dieser Stelle, dass es nicht immer einfach ist, Indikatoren dem reflektiven oder formativen Fall zuzuordnen. Komplizierend kommt der Umstand hinzu, dass bestimmte Konstrukte sowohl reflektiv als auch formativ spezifiziert werden können. Dazu ein kleiner Exkurs.

#### Exkurs: Reflektive und formative Indikatoren

Wenn man die Trunkenheit einer Person ermitteln möchte, kann man dies anhand eines ggf. glasierten Blickes, schlechter Koordination, verringerter Hemmungen u.ä. erschließen (Fall 1). In einem solchen Fall verläuft die Kausalität vom Konstrukt hin zu dessen Indikatoren. Wir sprechen daher von reflektiven Indikatoren.

Allerdings kann die Trunkenheit noch anders bestimmt werden. Sie hängt nämlich zweifelsfrei auch davon ab, ob die entsprechende Person eine gewisse Menge an Bier, Wein oder Hochprozentigem getrunken hat (Fall 2). Das Besondere an Fall 2 ist, dass eine Änderung eines Indikators (z.B. Trinken von noch einem Glas Wein) eine Änderung des Konstruktes (die Person wird trunkener) bewirkt. Die Kausalität verläuft also exakt entgegengesetzt zum reflektiven Fall, folglich handelt es sich hierbei um formative Indikatoren.

Es besteht nun die Möglichkeit, das Konstrukt Trunkenheit entweder reflektiv oder formativ zu messen. Eine dritte Option ist, sowohl die reflektiven als auch die formativen Indikatoren zu ver-

wenden, um Trunkenheit zu beschreiben. Diese Variante wird als MIMIC-Modell bezeichnet (Hauser/Goldberger 1971). MIMIC bedeutet dabei ‘Multiple Indicators – Multiple Causes’. Allerdings sind in einigen Softwareapplikationen, wie AMOS oder SmartPLS, solche Mischmodelle nicht modellierbar.

Erschwerend kommt hinzu, dass die Wahl der Form des Messmodells stets subjektiv ist. Als Beispiel formativer Indikatoren finden sich weiter oben im Text Teilzufriedenheiten des Konstruktes ‚Zufriedenheit mit einem Automobil‘. Homburg/Klarmann (2006, S. 731) argumentieren entgegengesetzt: Vorstellbar ist, dass Individuen einen allgemeinen, undifferenzierten Zufriedenheitseindruck speichern. Werden sie nun gebeten, ihre Zufriedenheit mit einzelnen Teilfacetten des Gesamtkonstruktes zu äußern, leiten sie diese aus dem Grobeindruck ab. Somit handelte es sich im vorliegenden Fall um reflektive Indikatoren der Gesamtzufriedenheit. Für beide Betrachtungsweisen gibt es Gründe und Gegenpositionen. Der Forscher ist also gefragt, eine intensive Diskussion der Beweggründe für die eine oder andere Konzeptualisierung zu führen.

Als Ergebnis lässt sich festhalten, dass es Messmodelle mit formativen, reflektiven oder sowohl formativen als auch reflektiven Indikatoren gibt. Daher gilt es, vor einer empirischen Erhebung gründlich zu prüfen, welche Art von Messmodell das passendste ist, um dies in der Konzeptualisierungsphase der Konstrukte berücksichtigen zu können. Zusammenfassend stellt Tabelle 2 die Eigenschaften reflektiver und formativer Messmodelle gegenüber.

**Tabelle 2**  
**Reflektive vs. formative Messmodelle**

Reflektives Messmodell	Formatives Messmodell
Kausalität von Konstrukt zu Messwerten	Kausalität von Messwerten zu Konstrukt
Messwerte sollten korreliert sein (sie besitzen interne Konsistenz)	Messwerte müssen nicht korreliert sein
Entfernung eines Indikators ändert nicht die Bedeutung des Konstruktes	Entfernung eines Indikators kann die Bedeutung des Konstruktes ändern
Messfehler auf Item-Ebene	Messfehler auf Konstrukt-Ebene
Konstrukt besitzt „Mehrwert“-Bedeutung	Konstrukt besitzt „Mehrwert“-Bedeutung
Skala repräsentiert nicht adequat das Konstrukt	Skala repräsentiert nicht adäquat das Konstrukt

Quelle: in Anlehnung an Jarvis/MacKenzie/Podsakoff (2003, S. 201)

### 2.3 Das Strukturmodell

Im Strukturmodell werden die eigentlich interessierenden Wirkungsbeziehungen zwischen mehreren latenten Variablen abgebildet. Mit Hilfe der von Wright (1921) entwickelten Pfadanalyse ist es möglich, den Einfluss eines Konstruktes auf andere Variablen zu bestimmen. Das Strukturmodell bildet die aufgestellten Hypothesen in Form kausaler Beziehungen (Pfade) ab. In der grafischen Abbildung bedeutet dies, dass Pfeile von jeweils einer Variablen zu einer anderen führen (siehe Abbildung 1). Diese Pfeile geben die kausale Richtung an. Von einer Variablen können mehrere Pfeile ausgehen, aber auch auf diese zeigen. Nachdem die Richtung der Kausalität festgelegt wurde, kann deren Stärke ermittelt werden. Sie wird durch

sog. Pfadkoeffizienten ausgedrückt, die in der standardisierten Form von -1 bis 1 reichen. Ein Wert von -1 bedeutet, dass eine Vergrößerung der unabhängigen Variablen zu einer Verringerung der abhängigen Variablen im selben Umfang führt. Ein Pfadkoeffizient von +1 zeigt einen uneingeschränkten positiven Einfluss an. Allerdings bedeutet eine solche perfekte Beziehung, dass zweimal das gleiche Konstrukt gemessen worden ist. Folglich ist es wünschenswert, wenn die Pfadstärken möglichst hoch sind, allerdings sind Werte  $> 0,4$  (bzw.  $< -0,4$ ) in komplexeren Modellen schon sehr hoch. Eine Beziehung von null meint, dass die betreffende abhängige nicht von der unabhängigen Variablen beeinflusst wird.

Neben diesen direkten Wirkungsbeziehungen lassen sich noch so genannte indirekte und totale Effekte berechnen (Mueller 1996, S. 141ff.). Dadurch lassen sich die gesamten Einflüsse einer latenten Variablen auf eine andere bestimmen – ein Vorteil der Strukturgleichungsmodellierung, der mit multipler Regressionsanalyse nicht zu bewerkstelligen ist.

<b>Beispiel</b>
<p>In unserem Automobilbeispiel haben wir bisher die Umweltverträglichkeit und die Einstellung betrachtet. Interessant wäre weiterhin zu untersuchen, ob sich die Einstellung auf die Absicht, das betreffende Auto zu kaufen, auswirkt. Wir nehmen folgende kausale Beziehungen an (<b>direkte Effekte</b>):</p> <p>Umweltverträglichkeit → Einstellung  Umweltverträglichkeit → Kaufabsicht  Einstellung → Kaufabsicht</p> <p>Der Einfluss der Umweltverträglichkeit auf die Einstellung ist ein direkter Effekt, genau wie der Einfluss der Einstellung auf die Kaufabsicht sowie der Umweltverträglichkeit auf die Kaufabsicht. Letztere Beziehung ergibt sich aber zusätzlich über die Kette</p> <p>Umweltverträglichkeit → Einstellung → Kaufabsicht</p> <p>Dies ist ein <b>indirekter Effekt</b> und wird durch Multiplikation der Pfadkoeffizienten der Einzelbeziehungen errechnet:</p> <p>Pfadstärke (Umweltverträglichkeit → Einstellung) * Pfadstärke (Einstellung → Kaufabsicht)</p> <p>Der <b>totale Effekt</b> ist die Summe aus dem direkten Effekt und sämtlichen indirekten Effekten, in unserem Beispiel:</p> <p>Pfadstärke (Umweltverträglichkeit → Kaufabsicht) + Pfadstärke (Umweltverträglichkeit → Einstellung) * Pfadstärke (Einstellung → Kaufabsicht)</p>

Wir haben nun die beiden Bestandteile von Strukturgleichungsmodellen – das Struktur- und das Messmodell – kennen gelernt. Wichtig ist die Unterscheidung in reflektive und formative Indikatoren, da davon die Skalenentwicklung sowie der nachfolgende Modellprüfungsprozess abhängen. Auf Ebene der latenten Variablen lassen sich direkte, indirekte und totale Effekte unterscheiden. Je dichter sie an 1 bzw. -1 sind, umso stärker ist die kausale Beziehung zwischen zwei Konstrukten.

Bevor in Kapitel 4 auf die Überprüfung des theoretisch entwickelten Strukturgleichungsmodells und seiner zugrunde liegenden Hypothesen eingegangen wird, stellt das anschließende Kapitel verschiedene Ansätze vor, mit denen die Werte (Parameter) eines Strukturgleichungsmodells ermittelt (geschätzt) werden können. In diesem Zusammenhang wird auf entsprechende Softwareapplikationen eingegangen.

### **3 Die Modellschätzung – ein Vergleich zweier Ansätze**

In der Einleitung wurde angesprochen, dass es zwei verschiedene Ansätze gibt, mit denen Strukturgleichungsmodelle geschätzt und überprüft werden können. Dabei handelt es sich um die Kovarianz- und Varianzanalyse, die im Folgenden kurz vorgestellt werden. Die Beziehungen zwischen latenten und manifesten Variablen können nicht exakt bestimmt werden, da die latenten Variablen in dem Sinne gar nicht ‘existieren’. Noch schwieriger wird es, Kausalitäten zwischen zwei latenten Variablen zu ermitteln, da beide Konstrukte nur über den Umweg ihrer Indikatoren erfasst werden können. Messfehler (auch als Residuen bezeichnet) verkomplizieren die Bestimmung der Wirkungsbeziehungen. Aus diesem Grund können sämtliche Beziehungen zwischen Variablen nur geschätzt werden. Die Kovarianzanalyse und die Varianzanalyse repräsentieren verschiedene Herangehensweisen an die Parameterschätzung. Die Ausführungen verzichten auf technische Beschreibungen der Abläufe einer solchen Schätzung, sondern stellen kurz wesentliche Eigenschaften der Ansätze dar, um sie miteinander vergleichbar zu machen.

#### **3.1 Die Kovarianzanalyse**

Die **Kovarianzanalyse** ist das Verfahren, mit dem am häufigsten Strukturgleichungsmodelle geschätzt und überprüft werden. Mit ihrer Hilfe kann die Struktur beobachtbarer Variablen, die sich aus ihrer Streuungs-(Varianz-Kovarianz)-Matrix ergibt, untersucht werden (Jöreskog/Wold 1982, S. 266). Auf diesem Wege ist es möglich, Beziehungen zwischen latenten Variablen, die durch manifeste Variablen gemessen werden, zu betrachten.

### Beispiel

Die Einstellung gegenüber unserem Auto wurde über die Indikatoren ‚attraktiv‘, ‚komfortabel‘, ‚sportlich‘ und ‚umweltschonend‘ ermittelt. Die Varianz des Indikators ‚attraktiv‘ errechnet sich aus der quadrierten Abweichung jeder Beobachtung der Variablen vom Mittelwert. Bei nur drei Beobachtungen mit den Werten 2, 4, 5 ergäbe sich folgende Varianz:

$$\text{Mittelwert} = (2+4+5)/3 = 3,667$$

$$\text{Quadrierte Abweichungen vom Mittelwert: } (2-3,667)^2 = 2,778; (4-3,667)^2 = 0,111; (5-3,667)^2 = 1,778$$

$$\text{Variation} = 2,778+0,111+1,778 = 4,667$$

$$\text{Varianz} = \text{Variation}/n-1 = 4,667/2 = 2,334$$

Die Kovarianz zeigt die lineare Beziehung von zwei Variablen untereinander, bspw. ‚attraktiv‘ und ‚komfortabel‘. Die auf Werte zwischen -1 und 1 standardisierte Kovarianz wird durch den Korrelationskoeffizienten beschrieben.

In der Varianz-Kovarianz-Matrix sind die Beziehungen unter den Variablen enthalten und diese bilden die Grundlage, um Modelle mit LISREL oder AMOS zu schätzen.

Aus einem Vergleich der postulierten und erhobenen Zusammenhänge erfolgt ein Rückschluss auf die Güte des Modells. Dafür stehen mehrere Maße zur Verfügung, die im Abschnitt 4.3.2 diskutiert werden. Genau diese Gütebeurteilung ist ein wichtiges Ziel der Kovarianzanalyse (Hu/Bentler 1998). Neben der Beurteilung des vorgeschlagenen Modells ist es ebenfalls möglich, verschiedene Varianten eines Modells vergleichbar zu machen.

Die bekanntesten Softwareprogramme zur Durchführung der Kovarianzanalyse sind **LISREL**, **AMOS**, **EQS** und **SAS PROC CALIS**. Eine Illustration der drei erstgenannten Programme liefert Byrne (2001). Ausführliche Darstellungen der Funktionsweise der Kovarianzanalyse, in denen auch die entsprechenden Formeln angegeben werden, finden sich bspw. bei Mueller (1996). Vor allem bei Verwendung von LISREL empfiehlt es sich, die Bezeichnungen der latenten und manifesten exogenen und endogenen Variablen, der Messfehler usw. zu studieren, da die Software nur mit diesen Begriffen operiert.

### Exkurs - Schätzalgorithmen

Innerhalb der Kovarianzanalyse lassen sich mehrere Algorithmen unterscheiden, mit denen die Parameter eines Modells geschätzt werden, diese werden auch als Schätzer bezeichnet. Vor allem unterscheiden sich diese Schätzer hinsichtlich der Verteilungsannahmen der zu Grunde liegenden Daten.

Der gebräuchlichste Algorithmus zur Parameterschätzung ist die **Maximum Likelihood-Methode (ML)**. Sie ist robust, konsistent und liefert umfangreiche und gute Gütemaße (Nevitt/Hancock 2001), weshalb sie am häufigsten eingesetzt wird. Nachteil dieser Methode ist die unterstellte multivariate Normalverteilung der Daten (Jöreskog/Wold 1982).

Problematisch ist in diesem Zusammenhang, dass auch im Marketing normalverteilte Daten selten anzutreffen sind. Die meisten Gütemaße beruhen aber auf einem Chi-Quadrat-Test, der sich ohne eine solche Verteilungsannahme nicht durchführen lässt. Als Konsequenz befassten sich verschiedene Studien mit der Klärung der Frage, inwieweit sich Verletzungen der Verteilungsannahme auf das Schätzergebnis auswirken (Hildebrandt/Görz 1999). Zwar besagen die Ergebnisse grundsätzlich, dass ML verhältnismäßig gut abschneidet (Hoyle/Panter 1995), allerdings konnte noch nicht zweifelsfrei belegt werden, dass eine fehlende Normalverteilung ignoriert werden kann. Da aufgrund dieser Annahmeverletzung sämtliche Folgerungen (z.B. Pfadkoeffizienten, geschätzte Messfehler) verzerrt sein können (Nevitt/Hancock 2001), sind die Ergebnisse mit Vorsicht zu interpretieren, sobald die Daten stark von einer Normalverteilung abweichen.

Keine Verteilungsannahme besitzt die Methode der **Unweighted Least Squares** (ULS). Obwohl der ML-Schätzer der effizientere der beiden ist (Bollen 1989, S. 112), sind beide Verfahren grundsätzlich ähnlich gut geeignet, um die Parameter eines Modells zu schätzen (Bollen 1989, S. 120; Jöreskog/Wold 1982, S. 264). Dies spricht somit eher für die Verwendung von ULS, falls die Daten stark von der Normalverteilung abweichen. Allerdings ergibt sich das umgekehrte Problem aus der fehlenden Verteilungsannahme: ULS kann weniger Gütemaße liefern, als dies mit ML möglich ist.

Daraus ergibt sich ein Dilemma bei nicht-normalverteilten Daten: entweder wird der Verteilung der Daten Rechnung getragen und ULS verwendet oder Maximum Likelihood benutzt, da nur dort die Vorteile der Kovarianzanalyse (v.a. Ermittlung der Gütemaße) richtig zum Tragen kommen.

Eine weitere Methode zur Parameterschätzung ohne Normalverteilungsvoraussetzung ist die der **Asymptotically Distribution Free** (ADF)-Schätzung. Allerdings benötigt ADF sehr große Stichproben – Nevitt/Hancock (2001) sprechen von einem  $n \geq 5000$  – was deren Praxistauglichkeit stark limitiert.

Zwar benötigen sämtliche Methoden der Kovarianzanalyse relativ große Stichproben, die Anforderungen liegen dennoch weit unter denen für ADF. Als Richtwerte gelten die 5-10fache Parameterzahl eines Modells oder Stichproben mit einem  $n > 250$  (Scholderer/Balderjahn 2006). Boomsma (1982, S. 171) rät davon ab, Stichprobengrößen kleiner 100 und stattdessen eine Untergrenze von 200 zu verwenden. Weiterhin fand er heraus,

dass die Parameterschätzung bei kleinen Stichproben robuster wird, je größer die Faktorladungen der Indikatoren sind.

*Zusammenfassend lässt sich festhalten, dass ADF aufgrund der hohen Anforderung an die Stichprobengröße in den meisten Studien nicht zum Einsatz kommen sollte, da selten derart viele Personen befragt werden. Weiterhin sind die erhobenen Daten selten multivariat normalverteilt, was für den Einsatz des ULS-Schätzers spricht. Da dieser aber nur wenige Gütemaße für die Überprüfung des Modells liefern kann und der ML-Algorithmus verhältnismäßig robust gegenüber Abweichungen von der Normalverteilung bleibt, wird an dieser Stelle empfohlen, die Maximum Likelihood-Methode zu verwenden.*

### 3.2 Die Varianzanalyse

Als Vertreter der **Varianzanalyse** ist der **Partial Least Squares** (PLS)-Ansatz in den letzten Jahren erneut ins Blickfeld einiger Forscher gelangt. Dies äußert sich in einer stetig zunehmenden Verwendung in wissenschaftlichen Studien (Hulland 1999; Scholderer/Balderjahn 2006). Als häufig genannte Applikationen sind PLSGraph, SPAD PLS, **SmartPLS** und LVPLS anzuführen. Die Wirkungsweise des PLS-Ansatzes beschreiben bspw. Henseler (2005), Lohmöller (1989) und Wold (1982).

**PLS** optimiert lokal, d.h. es maximiert die Vorhersage jeder abhängigen Variablen. Es ist folglich prognoseorientiert (Jöreskog/Wold 1982). Die Schätzer besitzen allerdings die Eigenschaft der ‚consistency at large‘ (Wold 1982), wodurch eine genaue Parameterbestimmung erst bei hoher Indikatorenzahl erreicht wird. Ist diese nicht gegeben, werden die Pfade im Strukturmodell unter-, und die Ladungen im Messmodell überschätzt<sup>9</sup> (Chin 1995). Dies führt zu der Schlussfolgerung, dass eine latente Variable so viele Indikatoren wie möglich besitzen sollte (Herrmann/Huber/Kressmann 2006, S. 41), dabei aber eine theoretisch bedingte Indikatorenzahlbegrenzung zu berücksichtigen ist (Lohmöller 1989, S. 213ff.). In der Kovarianzanalyse gilt die Forderung nach möglichst vielen Indikatoren nicht, da sie konsistent ist. Die Mindestanzahl an Indikatoren ist dort zwei (wobei vier Indikatoren zu bevorzugen sind, siehe Abschnitt 2.2), wobei ein Indikator nur einmal im Gesamtmodell verwendet werden darf (Anderson/Gerbing/Hunter 1987, S. 435).

---

<sup>9</sup> In der Kovarianzanalyse ist es hingegen genau umgekehrt: dort werden die Pfade über-, und die Ladungen unterschätzt.

Nachdem 1989 die erste (bedeutende) statistische Applikation für PLS<sup>10</sup> erschien, ist das gesteigerte Interesse in der jüngsten Zeit vor allem auf die bereits weiter oben angesprochene Beachtung *formativer Indikatoren* zurückzuführen. In den entsprechenden Programmen, wie PLS-Graph oder SmartPLS, können formative Messmodelle sehr einfach modelliert werden.

Auch die *geringen Anforderungen an die Stichprobengröße* richten das Augenmerk vieler Forscher auf PLS. Chin (1998b, S. 311) nennt eine Heuristik des zehnfachen Betrags der größten Anzahl an Indikatoren entweder a) eines exogenen Messmodells oder b) der abhängigen Variablen, auf die die meisten unabhängigen Variablen zeigen. Folglich wären Stichproben der Größenordnung 50 durchaus handhabbar. Allerdings warnen Marcoulides/Saunders (2006, S. viii) ausdrücklich davor, bei beliebig kleiner Stichprobe PLS einsetzen zu wollen: „PLS is not a silver bullet to be used with samples of any size!“.

Ein dritter Vorteil ist, dass PLS aufgrund der Natur der Kleinste-Quadrate-Schätzung *keine Verteilungsannahmen* trifft. Somit können Modelle mit Hilfe von PLS überprüft werden, ohne dass die Daten multivariat normalverteilt sein müssen. Liegt bspw. eine große Schiefe der Stichprobe vor, beeinflusst das nicht die Parameterschätzung in SmartPLS.

### 3.3 Hinweise zur Wahl des Ansatzes

Erkennbar ist, dass die Vorteile von PLS gerade durch die Schwachstellen der ML entstehen. Andererseits besitzen LISREL und AMOS einige Stärken gegenüber SmartPLS. Diese beziehen sich hauptsächlich auf die Gütemaße, aber auch die Effizienz, Konsistenz und Robustheit der Schätzergebnisse. Der zentrale Schwachpunkt von SmartPLS ist, dass es an Kriterien für die Beurteilung des Gesamtmodells fehlt; noch immer gibt es keinen überzeugenden Goodness-of-Fit-Index für die gemeinsame Beurteilung von Mess- und Strukturmodell im PLS-Ansatz (Hulland 1999). Dies stellt eine Legitimationshürde im Vergleich zu AMOS und LISREL dar. Die bisherigen Ausführungen reflektierend, werden in Tabelle 3 konkrete Hinweise gegeben, wann welches Verfahren – Kovarianz- oder Varianzanalyse – verwendet werden sollte.

---

<sup>10</sup> Lohmöllers LVPLS

**Tabelle 3**  
**Kovarianz- vs. Varianzanalyse – Hinweise zur Wahl des Ansatzes**

Pro Kovarianzanalyse (LISREL, AMOS)	Pro Varianzanalyse (SmartPLS)
Untersuchung <i>etablierter Theorien</i>	<i>Zusammenhänge zwischen Konstrukten noch nicht fundiert</i> : Studie besitzt eher explorativen Charakter <sup>11</sup> .
Auftreten <i>ausnahmslos reflektiver Messmodelle</i> <sup>12</sup>	Vorhandensein <i>formativer Messmodelle</i>
<i>Große Stichprobe</i>	<i>Kleine Stichprobe</i> <sup>13</sup>
Wunsch nach <i>ausdrucksstarken Gütemaßen</i>	Wunsch nach Ermittlung der <i>Prognoserelevanz</i>

Scholderer/Balderjahn (2006) empfehlen die Anwendung von PLS bei Stichprobenumfängen von  $n < 100$  und weniger als vier Indikatoren je latenter Variable oder bei Vorliegen unsicherer Modellhypothesen. Ansonsten raten sie eindringlich zur Verwendung von LISREL (und damit auch AMOS) aufgrund des höheren Leistungs- und Anwendungspotenzials. Dieser Standpunkt scheint aber etwas zu einseitig. Gemessen an den Hürden und einigen Nachteilen bei der Verwendung von LISREL und AMOS, stellt sich SmartPLS als brauchbare Alternative dar (Fornell/Bookstein 1982). Es gilt daher abzuwägen, welcher Ansatz sich besser eignet, anstatt nur danach zu gehen, ob ein bestimmtes Programm verfügbar, gewünscht oder dessen Benutzung bekannt ist. So lässt sich festhalten, dass LISREL/AMOS und SmartPLS eher komplementäre als konkurrierende Ansätze für Pfadmodelle mit latenten Variablen darstellen (Jöreskog/Wold 1982, S. 270).

Der Autor vertritt die Meinung, dass grundsätzlich LISREL (bzw. AMOS, EQS etc.) verwendet werden sollte. Sobald die Bedingungen aber für SmartPLS (bzw. PLS-Graph, SPAD PLS etc.) sprechen (formative Indikatoren, eine zu kleine Stichprobe bspw. aufgrund der Untersuchung mehrerer Untergruppen, unklare Theorielage), sollte nicht notwendigerweise ‚traditionell‘ verfahren, sondern offen mit dem PLS-Ansatz umgegangen werden. Einen interessanten Punkt skizziert Wold (1982, S. 34), der die These aufstellt, dass oftmals das Streben nach Konsistenz des PLS-Ansatzes geeigneter ist als der Optimalitätsanspruch der Kovarianzanalyse in einer echten, nicht-exakten Welt.

<sup>11</sup> Es ist aber „durchaus denkbar, dass der Forscher auf PLS zurückgreifen muss, obwohl dessen ursprüngliche Forschungsintention die Theoriebildung war“ (Hermann/Huber/Kressmann 2006, S. 63).

<sup>12</sup> Die Modellierung rein formativer Konstrukte ist vor allem mit AMOS nur sehr schwer möglich, so dass es dort nicht zu empfehlen ist. Aktuelle Bemühungen um die Einbindung formativer Konstrukte in LISREL (Temme 2006; Temme/Hildebrandt 2007) sind noch nicht hinreichend abgesichert und daher noch risikobehaftet.

<sup>13</sup> Obgleich größere Stichproben auch bessere Schätzergebnisse für den PLS-Algorithmus liefern, da dieser erst asymptotisch korrekt arbeitet, wenn sowohl ‚consistency at large‘ als auch ‚consistency‘ (große Stichprobenzahl) vorliegen (Jöreskog/Wold 1982, S. 266).

Nachdem die Wahl auf eine der beiden methodischen Ansätze gefallen ist, gilt es, das hypothetisch formulierte Strukturgleichungsmodell anhand der empirischen Daten zu überprüfen. Das nächste Kapitel zeigt einen Ablauf, mit dem ein Modell systematisch analysiert werden kann. Die Notwendigkeit einer getrennten Betrachtung je nach Durchführung der Kovarianz- bzw. der Varianzanalyse ergibt sich erst bei den so genannten globalen Gütemaßen in Punkt 4.3.2.

## 4 Modellbeurteilung

### 4.1 Erste Betrachtung der Schätzergebnisse

Unabhängig davon, welches Programm und welche Methode zur Parameterschätzung genutzt werden, folgt die Analyse den gleichen Schritten<sup>14</sup>. Eine erste Betrachtung gilt den Pfadkoeffizienten (siehe Abschnitt 2.3). So ist eine Plausibilitätsprüfung der geschätzten Parameter durchführbar<sup>15</sup> (Hildebrandt/Görz 1999). Die Pfadstärke sollte mindestens 0,2 betragen, damit eine interessante Beziehung vorliegt (Chin 1998a).

Anschließend folgt eine Analyse der Faktorladungen reflektiver Indikatoren. Für sie hat sich ein Mindestwert von 0,7 etabliert, da somit knapp die Hälfte der Varianz des Indikators durch das Konstrukt erklärt werden kann<sup>16</sup>. Da sich erklärte Varianz und Messfehler zu 1 addieren, bedeutet dies im Umkehrschluss, dass bei einer Faktorladung größer 0,7 der Messfehler nicht dominiert. Damit leistet der betreffende Indikator einen Erklärungsbeitrag.

Bei formativen Indikatoren werden im PLS-Ansatz nicht deren Faktorladungen, sondern Gewichtungen herangezogen, um sie zu beurteilen. Abbildung 2 zeigt ein geschätztes Strukturgleichungsmodell mit den Pfadkoeffizienten für jeden hypothetischen Zusammenhang sowie zeigt an, ob die Faktorladungen und deren Gewichtung betrachtet werden sollte.

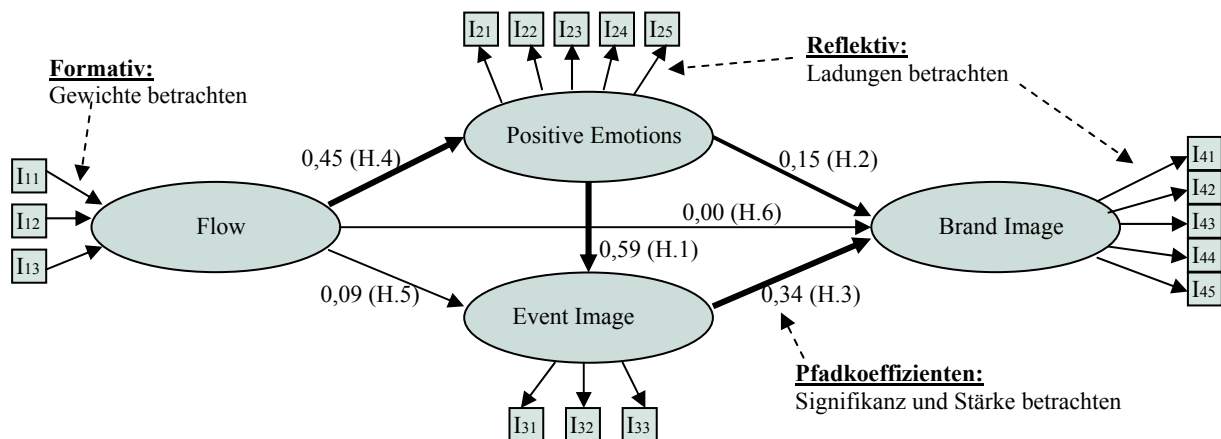
---

<sup>14</sup> Der Einfachheit und Verständlichkeit der Darstellung halber wird dem sog. 'one-step approach' gefolgt.

<sup>15</sup> Ergeben sich unzulässige Ergebnisse, kann dies mehrere Ursachen haben, darunter das Auftreten von „Ausreißern“ (Bollen 1989, S. 282ff.).

<sup>16</sup> Die erklärte Varianz ermittelt sich aus dem Quadrat der Faktorladung.

**Abbildung 2**  
**Standardisiertes Strukturgleichungsmodell nach der Parameterschätzung**



Quelle: in Anlehnung an Dregner/Gaus/Jahn 2007

Allerdings reicht es nicht allein aus, die Stärken der Pfade anzuführen. Besonders im Grenzbereich (0,15-0,25) ist eine zusätzliche Aussage, ob die Wirkungsbeziehungen ‚echt‘ sind, hilfreich. Dies lässt sich mit Hilfe der Signifikanzen bestimmen. Bei Anwendung von Maximum Likelihood ist dies direkt möglich, die Signifikanzen werden von der Software automatisch im Output ausgegeben. Falls eine Methode verwendet wird, die keine Normalverteilungsannahme trifft (ULS, PLS), eignet sich das **Bootstrap Resampling**, um Signifikanzwerte zu ermitteln (Nevitt/Hancock 2001). Hierbei werden aus der Stichprobe zufällig  $n$  Datensätze *mit Zurücklegen* gezogen, daraufhin erfolgt eine Schätzung aller Parameter. Die Prozedur wiederholt sich nun  $m$  mal. Die  $m$  Zufallsziehungen ergeben eine Bootstrapverteilung, auf deren Grundlage sich t-Werte berechnen lassen. Aus ihnen können wiederum Signifikanzniveaus abgeleitet werden.

AMOS gibt direkt ein  $p$  an, das das Signifikanzniveau ausdrückt bzw. weist mit „\*\*\*“ auf signifikant von der Nullhypothese ( $H_0$ : Die Beziehung zwischen zwei Variablen ist gleich null) abweichende Werte hin. Diese Information findet sich in der Tabelle „Standardized Regression Weights“ des Bootstrap-Blocks im AMOS-Output. Der zugrunde gelegte t-Test ist dabei zweiseitig (der Durchschnittswert nach dem Bootstrapping kann größer oder kleiner sein als der ursprüngliche Pfadkoeffizient). Bei Verwendung von SmartPLS muss die Bootstrap-Operation noch spezifiziert werden. In den Einstellungen sind an entsprechender Stelle die Anzahl der Datensätze („cases“,  $n$ ) sowie der Wiederholungen („samples“,  $m$ ) anzugeben. Für das  $n$  ist der Wert der Stichprobengröße zu wählen, im Feld „cases“ empfiehlt sich der Wert 1000, da somit ein t-Test mit 1000 (diese entsprechen in der Interpretation  $\infty$ )

Freiheitsgraden (df) durchgeführt wird. Die t-Werte aus diesem Test ( $df = \infty$ ) sind die gebräuchlichsten und mit ihren dazugehörigen Signifikanzniveaus in Tabelle 4 dargestellt.

**Tabelle 4**  
**t-Werte zur Signifikanzbeurteilung**

Signifikanzniveau	t-Wert (zweiseitig)
0,001	3,291
0,010	2,576
0,050	1,960

Signifikanzen sollten sowohl für die Pfadkoeffizienten als auch für die Gewichte bei formativen Indikatoren berichtet werden. Formal ist dies zwar auch für die Faktorladungen reflektiver Indikatoren sinnvoll, allerdings sind Ladungen größer 0,7 bei den geforderten Stichprobengrößen stets signifikant, was letztere Angabe überflüssig macht. Nachdem ein erster Blick auf das geschätzte Modell erfolgte, gilt es nun, es tiefgründiger anhand verschiedener Gütemaße zu analysieren. Die Kriterien, die letztlich Hintergrund der Entwicklung der Gütemaße sind, werden im kommenden Abschnitt betrachtet.

## 4.2 Beurteilungskriterien

In der Wissenschaft haben sich zwei zentrale Anforderungen herausgebildet, um die Qualität der Erfassung eines Konstruktes durch beobachtbare Variablen sowie eines Modells im Ganzen beurteilen und sicherstellen zu können. Bei ihnen handelt es sich um die **Reliabilität** und die **Validität** (Homburg/Giering 1996, S. 6). Die *Reliabilität* bezieht sich auf die Zuverlässigkeit der Messung. Der Einfluss von *zufälligen Messfehlern* soll so gering wie möglich sein, d.h. es wird der Grad widerspiegelt, mit dem unabhängige, aber vergleichbare Messungen ein und derselben Variablen übereinstimmen (Peter 1981). „Reliable measures allow one to generalize from one particular use of the method to a wide variety of related circumstances“ (Nunally/Bernstein 1994, S. 214). Um nachvollziehen zu können, wie gut das gemessen wurde, was beabsichtigt war zu messen, wurden die Kriterien der *Validität* entwickelt (Nunally/Bernstein 1994, Peter 1981). Hierbei sollen ebenfalls *systematische Fehler* ausgeschlossen werden. Eine notwendige, jedoch keine hinreichende Voraussetzung für Validität sind reliable Messungen.

### Beispiel

Wir ermittelten die Einstellung gegenüber unserem Auto mit Hilfe der Indikatoren ‚attraktiv‘, ‚komfortabel‘, ‚sportlich‘ und ‚umweltschonend‘. Möglich wäre aber auch, die Einstellung darüber zu messen, dass Individuen Steine so weit wie möglich werfen sollten. Bei mehreren Wiederholungen sollten die erbrachten Distanzen eines Individuums hoch miteinander korrelieren. Dies bedeutet, dass diese Messung reliabel ist. Offensichtlich entspricht dieses Vorgehen aber keiner validen Messung der Einstellung zu einem Automobil, sondern eher der Stärke einer Person.

Homburg/Giering (1996, S. 7) unterteilen die Gültigkeit der Messung in vier Validitätsarten<sup>17</sup>:

**Tabelle 5**  
**Validitätsarten nach Homburg/Giering (1996)**

Validitätsart	Beschreibung
<b>Inhaltsvalidität</b>	<ul style="list-style-type: none"> <li>▪ Variablen bilden den Inhalt/die Bedeutung des Konstruktes ab</li> </ul>
<b>Konvergenzvalidität</b>	<ul style="list-style-type: none"> <li>▪ Die Indikatoren, die demselben Faktor zugeordnet sind, sollen eine ausreichend starke Beziehung – also hohe Korrelationen – untereinander aufweisen</li> </ul>
<b>Diskriminanzvalidität</b>	<ul style="list-style-type: none"> <li>▪ Die Assoziation zwischen den Indikatorvariablen, die verschiedenen Faktoren zugeordnet sind, sollen schwächer sein als die Assoziationen zwischen den Indikatoren, die denselben Faktor messen</li> </ul>
<b>Nomologische Validität</b>	<ul style="list-style-type: none"> <li>▪ Einbindung des Konstrukts in einen übergeordneten theoretischen Rahmen, die Ergebnisse sollten theoretisch begründbar sein</li> </ul>

„Although we can never prove validity, we can develop strong support for it“ (Bollen 1989, S. 185). Die Inhalts- und nomologische Validität müssen bei der Indikatorenauswahl, der Herleitung des Hypothesensystems und dessen Interpretation berücksichtigt werden. Für die Ermittlung der Konvergenz- und Diskriminanzvalidität eignen sich Gütemaße, die im nächsten Abschnitt vorgestellt werden. Zuvor soll ein Beispiel das Verständnis für diese Validitätsarten erleichtern.

Beispiel
<p>In unserem Automobilbeispiel können wir die Inhaltsvalidität für die Umweltverträglichkeit bspw. dahingehend prüfen, ob deren ausgewählte Indikatoren wirklich selbige messen:</p> <p>‚verbraucht viel‘, ‚stößt viel CO<sub>2</sub> aus‘ oder ‚belastet die Umwelt‘ wären folglich geeignet, Indikatoren wie ‚lässt sich gut einparken‘ wären nicht inhaltsvalide.</p> <p>Die Konvergenz- und Diskriminanzvalidität wurden in diesem Zusammenhang bereits im Abschnitt 2.2 diskutiert.</p> <p>Wenn ein Ergebnis der Modellanalyse wäre, eine niedrige Umweltverträglichkeit wirke sich positiv auf die Kaufabsicht aus, bestehen Bedenken ob der nomologischen Validität. Der Verdacht ließe sich äußern, dass eigentlich etwas anderes gemessen wurde, sofern nicht andere Begründungen vorliegen (z.B., dass überwiegend Personen befragt wurden, die eine hohe Umweltverträglichkeit mit schlechteren Fahreigenschaften verbinden, bei gleichzeitiger Bevorzugung letzterer und Ausblendung zukünftiger Konsequenzen).</p>

### 4.3 Gütemaße und deren Interpretation

Auf Reliabilitäts- und Validitätskriterien der ersten Generation<sup>18</sup> (Churchill 1979) wird an dieser Stelle nicht eingegangen. Die hier beschriebenen Beurteilungskriterien beziehen sich auf das Messmodell, das Strukturmodell oder Struktur- und Messmodell gemeinsam.

<sup>17</sup> Es gibt verschiedene Unterteilungsmöglichkeiten der Validität. Bagozzi (1980, S. 114ff.) fasst die genannten vier Arten bspw. unter dem Begriff der *‘construct validity’* zusammen.

Die Überprüfung der Messmodelle unterscheidet sich nicht wesentlich zwischen der Kovarianz- und Varianzanalyse. Aus diesem Grund werden diese so genannten lokalen Anpassungsmaße gemeinsam für die drei Softwareapplikationen betrachtet. Aufgrund der im Abschnitt 2.2.2 beschriebenen Unterschiede zwischen formativen und reflektiven Indikatoren, differieren auch die Methoden für deren Reliabilitäts- und Validitätsprüfung.

Die globale Beurteilung eines Strukturgleichungsmodells unterscheidet sich erheblich, je nach Verwendung von LISREL/AMOS oder SmartPLS. Daher werden die spezifischen Gütemaße in getrennten Abschnitten erläutert.

#### 4.3.1 Lokale Anpassungsmaße

##### 4.3.1.1 *Reflektive Messmodelle*

Lokale Gütemaße beziehen sich auf die Eignung der Messmodelle. Zum einen werden die Indikatoren an sich und zum anderen ihre Beziehung zur latenten Variablen beurteilt.

Zunächst werden die **Faktorladungen** betrachtet, die einen Mindestwert von 0,707 annehmen sollten, damit die Hälfte der Varianz in den Indikatoren mit dem Konstrukt verbunden werden kann (Johnson/Herrmann/Huber 2006, S. 126).

Zur Bestimmung der internen Konsistenz kann die *Konstruktreliabilität* („**composite reliability**“) (Chin 1998b, S. 320; Hair et al. 2006, S. 777) herangezogen werden. Sie ist genauer als Cronbachs Alpha, da letzteres auf die Anzahl der Indikatoren reagiert (Chin 1998b, S. 320). Die Konstruktreliabilität sollte größer als 0,7 sein.

Letztlich eignet sich die *durchschnittlich erfasste Varianz (AVE)* als weiteres Maß für die Beurteilung eines Indikatorensets. Ihr Mindestwert wird mit 0,5 angegeben (Fornell/Larcker 1981, S. 45f.).

---

<sup>18</sup> Z.B. Cronbachs Alpha oder die Item-to-Total-Korrelation

### Exkurs – Berechnung der Konstruktreliabilität und AVE

Die Konstruktreliabilität und AVE werden direkt im Output von SmartPLS ausgegeben. Bei Verwendung von LISREL oder AMOS müssen diese über folgende Formeln berechnet werden (Fornell/Larcker 1981, S. 45f.):

$$(1) \text{ Konstruktreliabilität} = \frac{(\text{Summe der standardisierten Ladungen})^2}{(\text{Summe der standardisierten Ladungen})^2 + \text{Summe des Messfehlers}}$$

$$(2) \text{ Durchschnittlich erfasste Varianz} = \frac{\text{Summe der quadrierten standardisierten Ladungen}}{\text{Summe der quadr. standardisierten Ladungen} + \text{Summe des Messfehlers}}$$

Die standardisierten Ladungen sind direkt ablesbar: in LISREL in den Output-Tabellen der *completely standardized solution* ‚LAMBDA-Y‘ und ‚LAMBDA-X‘ und in AMOS in der Tabelle ‚Standardized Regression Weights‘ bzw. ‚Standardized Direct Effects‘. Der Messfehler jedes Indikators lässt sich entweder direkt ablesen (LISREL: Tabelle ‚THETA-EPS‘ im *completely standardized solution*-Bereich) oder leicht über  $1 - (\text{standardisierte Ladung}^2)$  berechnen.

Die Konstruktreliabilität und die durchschnittlich erfasste Varianz eignen sich als Prüfgrößen für die **Konvergenzvalidität** der einem Faktor zugeordneten Indikatoren (Hair et al. 2006, S. 776f.; Homburg/Giering 1996, S. 11). Um die Reliabilitäts- und Validitätsbetrachtungen zu komplettieren, erfolgt eine Beurteilung der **Diskriminanzvalidität** anhand des **Fornell/Larcker-Kriteriums** (Fornell/Larcker 1981, S. 46). Es besagt, dass die durchschnittlich erfasste Varianz eines Faktors größer sein muss als jede quadrierte Korrelation zwischen diesem und einem weiteren Konstrukt. Ein Beispiel zur Darstellung der Analyse reflektiver Messmodelle gibt Tabelle 6.

**Tabelle 6**  
**Beurteilung reflektiver Messmodelle**

Konstrukt/ Indikator (Anforderung)	Faktor- ladungen ( $\geq 0.707$ )	Konvergenz- validität		Diskriminanz- validität
		Konstrukt- reliabilität ( $\geq 0.7$ )	AVE ( $\geq 0.5$ )	Fornell/Larcker ( $\text{AVE} > \text{Korr}^2$ )
<b>Emotionen</b>		0.88	0.588	$0.59 > 0.41$
erfreut	0.745			
zufrieden	0.709			
begeistert	0.763			
glücklich	0.843			
fröhlich	0.768			

Legende: AVE – Average Variance Extracted (Durchschnittlich erfasste Varianz),  
Korr<sup>2</sup> – Höchste quadrierte Korrelation zwischen den Konstrukten

Quelle: Auszug in Anlehnung an Drenchner/Gaus/Jahn (2007)

#### 4.3.1.2 Formative Messmodelle

Für formative Indikatoren ist eine Betrachtung der internen Konsistenz und der Konvergenzvalidität (und somit ebenfalls der Diskriminanzvalidität) nicht maßgeblich (Bollen/Lennox 1991; Jarvis/MacKenzie/Podsakoff 2003; Nunally/Bernstein 1994). Aus diesem Grund werden bei ihnen lediglich die Gewichtungen, die signifikant sein sollen, betrachtet (Chin 1998b, S. 324).

Da in formativen Messmodellen eine multiple Regression durchgeführt wird, darf keine Multikollinearität<sup>19</sup> vorliegen, da diese die Bestimmung der Indikatorvalidität erschwert (Diamantopoulos/Winklhofer 2001, S. 272). Um dennoch überprüfen zu können, ob die Operationalisierung richtig gewählt wurde, lässt sich ein Test auf nomologische Validität durchführen. Dabei wird untersucht, ob die Beziehung zwischen dem formativ spezifizierten Konstrukt und einer davon abhängigen latenten Variablen die theoretische Erwartung widerspiegelt (ebenda). Allerdings ist es dazu notwendig, dass diese Beziehung nicht nur theoretisch postuliert, sondern schon empirisch überprüft wurde. Deckt sich der Befund mit den früheren Untersuchungen, spricht dies für die nomologische Validität und eine gelungene Operationalisierung des Konstruktes.

Erfüllen Faktorladungen, Konstruktvalidität, durchschnittlich erfasste Varianz sowie der Vergleich der AVE mit den Korrelationen mit anderen Konstrukten die Anforderungen an die reflektiven Messmodelle, ist von einer reliablen und validen Messung auszugehen. Im Anschluss daran ist es nun möglich, das Strukturmodell zu überprüfen. Wird die Annahme unterstützt, das untersuchte Modell hält der empirischen Überprüfung stand, können dann die Hypothesen überprüft werden. Die Untersuchung, ob die erhobenen Daten den durch das Modell ausgedrückten Zusammenhängen widersprechen oder nicht, wird mit Hilfe der globalen Gütemaße durchgeführt.

#### 4.3.2 Globale Gütemaße für LISREL und AMOS

Für die Beurteilung der Strukturgleichungsmodelle steht in LISREL und AMOS eine Vielzahl so genannter Fitindizes zur Verfügung. Da es nicht *den* besten Index für die Gesamtbeurteilung eines Modells gibt, sollten mehrere Indizes zur Gütebeurteilung herangezogen werden (Bollen/Long 1993, S. 8; Marsh/Balla/Hau 1996, S. 315).

---

<sup>19</sup> Wechselseitige Abhängigkeit der formativen Indikatoren.

#### 4.3.2.1 Absolute Fitindizes

**Absolute Fitindizes** bewerten, wie gut ein a-priori-Modell die Daten der Stichprobe reproduziert. Um diese Übereinstimmung statistisch zu testen, lässt sich ein  $\chi^2$ -Test durchführen. Allerdings besteht das Problem, dass ein Modell immer nur eine Annäherung an die Realität sein kann und kein Abbild ihrer: „Since a null hypothesis that a model fits exactly in some population is known a priori to be false, it seems pointless even to try to test whether it is true“ (Browne/Cudeck 1993, S. 137). Die hohen Annahmen (perfekte Abbildung der Realität durch das Modell) führen dazu, dass die Nullhypothese mit hoher Wahrscheinlichkeit abgelehnt wird, sobald die Stichprobenanzahl groß genug ist (Bagozzi/Baumgartner 1994, S. 399; Browne/Cudeck 1993, S. 146). Es ist daher angebrachter,  $\chi^2$  weniger als harte Teststatistik, sondern vielmehr als Gütemaß zu interpretieren (Jöreskog 1993). Eine Möglichkeit besteht darin  $\chi^2$  durch die Anzahl der Freiheitsgrade zu dividieren. In LISREL werden  $\chi^2$  und Freiheitsgrade angegeben, die Division muss aber per Hand erfolgen. AMOS gibt das Maß direkt als ‚CMIN/DF‘ aus. Das Modell ist umso besser, je kleiner  $\chi^2/\mathbf{df}$  ist. Allerdings herrscht kein Konsens darüber, wann ‚guter‘ Fit erreicht wird. Hier schwanken die Empfehlungen von 3, 2 oder kleiner bis hin zu 5 (Arbuckle/Wothke 1999, S. 399f.; Bollen 1989, S. 278).

Häufig verwendet werden weiterhin der Goodnes of Fit-Index (**GFI**) sowie der die Freiheitsgrade berücksichtigende Adjusted Goodness of Fit-Index (**AGFI**). Im Gegensatz zum Chi-Quadrat zeigen Werte nahe 1 ein gutes Modell an. Als Daumenregel sollten diese Fitindizes (sowie die im Anschluss vorgestellten inkrementellen Indizes) größer 0,9 sein (Hair et al. 2006, S. 747; Hoyle/Panter 1995, S. 164). Hu/Bentler (1998) empfehlen, GFI und AGFI *nicht* zu verwenden, da diese sensibel auf die Stichprobengröße reagieren. Darüber hinaus verringert sich ihr Wert in größeren Modellen, wodurch diese ungerechtfertigter Weise abgelehnt werden könnten. (Anderson/Gerbing 1984, S. 166).

Als Maß für die in einem Modell durchschnittlich nicht erklärten Kovarianzen kann der **RMR-Index** (Root Mean Squared Residual) herangezogen werden (Fan/Sivo 2005, S. 360), der dem Standardfehler im Rahmen der Regressionsanalyse entspricht. Von diesem Maß existiert auch eine standardisierte Variante, der **SRMR**. Ein Modell ist umso besser, je kleiner RMR bzw. SRMR sind. Deren Werte gelten als gut, wenn sie unter 0,05 liegen.

**RMSEA** (Root Mean Square Error of Approximation) betrachtet ebenfalls den mangelnden Fit des Modells im Verhältnis zur Kovarianzmatrix der Stichprobe. Das bedeutet, dass RMSEA die Abweichung zwischen durch das Modell postulierten und tatsächlich erhobenen Daten ermittelt. Diese Diskrepanz wird anschließend durch die Anzahl der Freiheitsgrade

dividiert. Dadurch berücksichtigt RMSEA die Sparsamkeit (als Gegenstück der Komplexität) eines Modells und bestraft damit nicht die einfachen, leicht zu interpretierenden Modelle<sup>20</sup>. Werte für RMSEA sollten kleiner als 0,08 sein, um ein akzeptables Modell anzuzeigen. Ein guter Fit in Relation zu den Freiheitsgraden wird durch Werte kleiner 0,05 ausgedrückt. Ist RMSEA größer als 0,1, spricht dies gegen das Modell (Browne/Cudeck 1993, S. 144).

#### 4.3.2.2 Inkrementelle Fitindizes

**Inkrementelle Fitindizes** ergeben sich aus dem Vergleich des postulierten Modells mit einem Nullmodell, das stärker restringiert ist als das Zielmodell, da seine Variablen nicht miteinander korrelieren dürfen (Hu/Bentler 1998). Marsh/Balla/Hau (1996, S. 346ff.) empfehlen als Ergebnis ihrer Analyse mehrerer inkrementeller Indizes die Verwendung von RNI bzw. dessen genormter Variante **CFI** (Comparative-Fit-Index) sowie von NNFI (auch: **TLI** – Tucker-Lewis Index genannt) bzw. dessen genormter Variante NTLI.

Der Normed-Fit-Index (**NFI**) sowie der Relative-Fit-Index (**RFI**) – beides häufiger verwendete Gütemaße – sind kritisch zu betrachten und sollten nicht in die Analyse einbezogen werden. Sie neigen nämlich zu starken Verzerrungen, hervorgerufen vor allem durch die Stichprobengröße (Marsh/Balla/Hau 1996, S. 318, 342ff.). Dies gilt auch für den **IFI** (Incremental Fit Index).

#### 4.3.2.3 Indexselektion

Hu/Bentler (1998, S. 446f.) empfehlen in ihrem zentralen Artikel zu Gütemaßen der Kovarianzanalyse den **SRMR** (Standardized Root Mean Square Residual) in Kombination mit **TLI**, **BL89**, **RNI**, **CFI**, **gamma hat**, **Mc** und **RMSEA**<sup>21</sup>. Fan/Sivo (2005) fanden allerdings Anzeichen dafür, die gegen eine solche ‚Zwei-Index-Strategie‘ sprechen. Dennoch erachten sie grundsätzlich die Betrachtung mehrerer Indizes als sinnvoll, da die Stärken und Schwächen der einzelnen Fitindizes noch immer nicht genau bekannt sind (ebenda, S. 367).

Die empfohlenen Indizes der genannten Autoren werden nun als Basis für eine solide Modellprüfung angesehen und dahingehend überprüft, ob AMOS und LISREL diese ausgeben. Aufgrund der Nicht-Eignung des IFI (=BL89) wird dieser nicht berücksichtigt. Der AGFI sollte trotz Hu/Bentlers (1998) Bedenken grundsätzlich angeführt werden, da aufgrund dessen

---

<sup>20</sup> Dies ist bspw. bei  $\chi^2$  der Fall: „The difficulty is that the fit of the model can usually be improved by increasing the number of parameters” (Browne/Cudeck 1993, S. 136).

<sup>21</sup> Da nach ihren Analysen SRMR als einziges Gütemaß sensibel auf Fehler im Strukturmodell reagiert; die anderen Indizes reagieren am sensibelsten auf Fehler in den Messmodellen.

Popularität viele Gutachter von Abschlussarbeiten und wissenschaftlichen Artikeln erwarten, dass dieser berichtet wird. Sivo et al. (2006, S. 285) konnten zeigen, dass so genannte ‚parsimonious indexes‘, die sparsame Modelle<sup>22</sup> (mit vielen Restriktionen) belohnen, nicht interpretierbar sind, da für sie keine optimalen Werte bestimmt werden können. Tabelle 7 fasst die Ausführungen noch einmal zusammen.

**Tabelle 7**  
**Fitindizes und ihre Verfügbarkeit in LISREL und AMOS**

Index	Benötigte Schätzmethode	
	LISREL 8.7	AMOS 6
RMSEA	ML <sup>23</sup>	ML
SRMR	ML	Nicht verfügbar
RMR	ML, ULS	ML, ULS
CFI	ML	ML
TLI (=NNFI)	ML	ML
CMIN/DF	ML	ML
AGFI	ML, ULS	ML, ULS
RNI	Nicht verfügbar	Nicht verfügbar
Gamma hat	Nicht verfügbar	Nicht verfügbar
Mc	Nicht verfügbar	Nicht verfügbar

Als Mindestzahl der berichteten Gütemaße empfehlen Hair et al. (2006, S. 752) drei Indizes. Unter ihnen sollten sich ein absoluter und ein inkrementeller sowie ein „badness-of-fit-index“, vorzugsweise RMSEA, befinden.

Der Autor empfiehlt die unbedingte Verwendung von **RMSEA**, **CFI**, **TLI (NNFI)** sowie **(S)RMR** (siehe Tabelle 8). Zusätzlich sollten der Vollständigkeit halber  $\chi^2/df$  bzw. *CMIN/DF* und *AGFI* berichtet werden.

<sup>22</sup> Ein Modell wird durch zusätzliche Restriktionen sparsamer: Ließen wir in unserem Automobilbeispiel den kausalen Pfeil von Umweltverträglichkeit auf die Kaufabsicht weg, bedeutet dies implizit, dass wir diese Beziehung gleich null setzen, was einer Restriktion entspricht.

<sup>23</sup> Zwar werden sämtliche Gütemaße selbst bei Durchführung des ULS-Schätzers angegeben, beruhen aber auf multivariater Normalverteilung, woraus sich diesbezüglich kein Unterschied/Vorteil gegenüber der Wahl von ML ergibt.

**Tabelle 8**  
**Fitindizes in LISREL und AMOS und Anforderungsniveaus**

LISREL 8.7	AMOS 6	Anforderung
RMSEA	RMSEA	<0,08
SRMR		<0,05
	RMR	<0,05 (nicht >0,1)
CFI	CFI	>0,90
NNFI	TLI	>0,90
CHI-SQUARE/df	CMIN/DF	<3
AGFI	AGFI	>0,90

Quellen: Hair et al. (2006); Hu/Bentler (1999); Marsh/Balla/Hau (1996)

Bezüglich der Beurteilungsrichtwerte fordern Hu/Bentler (1999, S. 27) eine Erhöhung dieser ‚cutoff values‘ von der Daumenregel 0,90 auf mindestens 0,95. Marsh/Hau/Wen (2004, S. 349) betonen allerdings: „Hence, we strongly encourage researchers, textbook authors, reviewers, and journal editors not to overgeneralize the Hu and Bentler (1998,1999) results, transforming heuristic findings based on a very limited sample of misspecified models into golden rules of fit that are broadly applied without the cautions recommended by Hu and Bentler (1999).“ Beachtet werden sollte außerdem, dass die Fitindizes in ML- und ULS-Schätzungen unterschiedlich berechnet werden und bei ULS tendenziell höher ausfallen. Grundsätzlich scheinen die Fitindizes aus ML besser geeignet zur Modellbeurteilung (Hu/Bentler 1998). Auf ihnen beruhen auch die abgeleiteten cutoff-Werte (Hu/Bentler 1999). Zusammengefasst bedeutet dies, dass die Fitindizes der Maximum-Likelihood-Schätzung berichtet werden sollten und diese mindestens einen Wert von 0,9 (das gilt für CFI, NNFI und AGFI) beziehungsweise kleiner 0,08 (RMSEA), respektive 0,05 (SRMR, RMR) aufweisen sollten.

Die besprochenen Fitindizes eignen sich ebenfalls, um verschiedene Varianten eines Modells zu vergleichen. Diese Alternativen sollten aber ähnlich komplex sein um die Vergleichbarkeit zu gewährleisten. Besonders bieten sich so genannte ‚nested models‘ an. Hier werden lediglich die Beziehungen zwischen Konstrukten verändert, d.h. es werden im Modell Pfade hinzugefügt oder entfernt. Verbessert sich bspw. der CFI nach einer solchen Veränderung, ist diese Alternative geeigneter (Hair et al. 2006, S. 752). Allerdings ist davor zu warnen, lediglich alle Beziehungen so lange zu ändern, bis ein guter Fit ‚entsteht‘. Wie schon im Abschnitt 2.1 erwähnt, wird in der Kausalanalyse ein fundiertes Hypothesensystem überprüft. Ein Pfad muss folglich theoretisch begründet sein. „Strong grounding in prior theory should guard against capitalization on chance, and consideration of alternate theoretical frameworks should

make it less likely for relevant relationships to go undetected“ (Bagozzi/Baumgartner 1994, S. 390). Sind verschiedene kausale Beziehungen zwischen den Variablen denkbar, ist ein Alternativenvergleich ein probates Mittel, um ein Strukturgleichungsmodell weiter zu verbessern. Jöreskog (1993, S. 305ff.) sowie Browne/Cudeck (1993, S. 147ff.) beschreiben weitere Möglichkeiten und Maße (wie AIC, CAIC, CVI und ECVI) zur Kreuzvalidierung eines Modells sowie zum Vergleich mehrerer Modellvarianten.

Nachdem ein Modell verändert wurde (z.B. Entfernung von Indikatoren, Änderung der strukturellen Beziehungen), sollte es an einem neuen Datensatz überprüft werden (Kreuzvalidierung). Somit kann nachvollzogen werden, ob bestimmte Beziehungen nur in der einen Stichprobe bestanden. Hat man nur eine Stichprobe zur Verfügung und sollte dennoch kleinere Modifikationen am Modell vornehmen (z.B. nachdem die Hauptuntersuchung bereits zu einem Termin stattfand, der nicht mehr wiederholbar ist), ist es bei ausreichender Größe der Stichprobe ratsam, selbige zufällig in zwei Hälften aufzuteilen. Eine Gruppe kann nun dazu genutzt werden, das Modell zu optimieren („calibration sample“), während die zweite der Kreuzvalidierung dient („validation sample“) (Diamantopoulos/Siguaw 2000, S. 129).

#### 4.3.3 Globale Gütemaße für SmartPLS

Fitindizes wie bei LISREL oder AMOS sind aufgrund der Natur der Varianzanalyse mit SmartPLS nicht bestimmbar. Für die PLS-Schätzung gelten somit andere Gütemaße, die mit denen einer linearen Regression vergleichbar sind. Nachfolgende Tabelle 9 stellt die globalen Gütemaße dar, die für den PLS-Ansatz generell gelten. Die aktuelle Version von SmartPLS gibt sämtliche Maße aus.

**Tabelle 9**  
**Gütemaße für SmartPLS**

Gütemaße (global)	Anforderungen
$R^2$ (Bestimmtheitsmaß)	>0,4
$Q^2$ (Prognoserelevanz)	>0
$f^2$ (Effektstärke)	>0,35 (large) >0,15 (medium) ~0,02 (small)
Pfadkoeffizienten	>0,2

Quellen: Chin (1998a; 1998b); Ringle (2004)

$R^2$  beschreibt den Anteil der Varianzaufklärung und stellt somit das Bestimmtheitsmaß dar. Entsprechende Werte sollten größer als 0,4 sein (Chin 1998b). Die Effektstärke ( $f^2$ ) der unabhängigen (exogenen) Variablen auf ihre jeweils abhängigen (endogenen) latenten Variablen liefert Hinweise für die Bedeutung einer exogenen auf eine bestimmte endogene Variable. Sie

ergibt sich aus der Differenz der Bestimmtheitsmaße der endogenen Variablen bei Vorhandensein und Ausschluss der exogenen Variablen. Akzeptable Einflüsse liegen bei Werten ab 0,15 vor (Chin 1998b, S. 316f.; Ringle 2004, S. 15f.). SmartPLS gibt die Effektstärke nicht explizit aus, sie muss über eine einfache Formel berechnet werden:

$$(3) f^2 = \frac{R_{included}^2 - R_{excluded}^2}{1 - R_{included}^2}$$

Dazu wird das  $R^2$  einer endogenen Variablen notiert ( $R_{included}^2$ ) und noch einmal berechnet, nachdem die exogene Variable aus dem Modell entfernt wurde ( $R_{excluded}^2$ ). Das Resultat ist die Effektstärke der exogenen auf die endogene Variable.

Die kreuzvalidierte Kommunalität (cv-communality) misst die Fähigkeit des Pfadmodells, die manifesten Variablen (Indikatoren) aus den berechneten latenten Variablen vorherzusagen (Tenenhaus et al. 2005). Der Durchschnittswert der cv-communalities kann für die globale Beurteilung der Messmodelle verwendet werden. Das Stone-Geisser Kriterium,  $Q^2$ , evaluiert die Vorhersagerelevanz, also wie gut die abhängigen von ihren unabhängigen Variablen im Strukturmodell beschrieben werden (Chin 1998b; Ringle 2004). SmartPLS gibt diesen Wert unter der Bezeichnung ‚Construct Crossvalidated Redundancy‘ aus, er sollte größer als null sein (Fornell/Cha 1994). Die kreuzvalidierte Kommunalität und Redundanz werden durch das Verfahren des Blindfolding bestimmt<sup>24</sup>.

Um zu einem Globalurteil ähnlich GFI bei LISREL/AMOS zu kommen, schlagen Tenenhaus et al. (2005) folgenden Goodness-of-Fit-Index vor:

$$(4) \text{GoF} = \sqrt{\text{durchschnittliche Kommunalität} * \text{durchschnittliches } R^2} \text{ bzw.}$$

$$(5) \text{GoF} = \sqrt{AVE * \text{durchschnittliches } R^2}$$

Dieser Index muss aber aus theoretischer Sicht weiter verbessert werden, da er leicht zu manipulieren ist. Außerdem ist er nicht anwendbar, sobald formative Indikatoren vorkommen. Darüber hinaus existieren keine Richtwerte, ab welchem GoF ein Modell als gut beurteilt wird. Aus diesen Gründen kann momentan dieser Wert noch nicht sinnvoll eingesetzt werden.

---

<sup>24</sup> Die Funktionsweise der Blindfoldingprozedur wird bspw. in Chin (1998b, S. 317) näher beschrieben.

#### 4.3.4 Hypothesenprüfung

Ergeben die lokalen und globalen Prüfungen ein akzeptables – also reliables und valides – Modell, können im letzten Schritt die aufgestellten Hypothesen getestet werden. Ist die Beziehung zwischen zwei Konstrukten signifikant von null verschieden und verläuft in der postulierten Richtung (positiver oder negativer Zusammenhang), kann die durch sie ausgedrückte Hypothese bestätigt<sup>25</sup> werden. Weiterhin hilft die Betrachtung totaler Effekte, die Wirkungszusammenhänge in einem komplexen Modell besser zu verstehen. Ein totaler Effekt ist dabei der gesamte Einfluss einer Variablen auf eine andere über alle denkbaren Beziehungen mit weiteren Konstrukten (mediierende Variablen).

## 5 Fazit

Ziel dieses Beitrags war es, ein besseres Verständnis von Strukturgleichungsmodellen zu ermöglichen und für eigene wissenschaftliche Untersuchungen Orientierung zu bieten. Wichtig ist, dass Strukturgleichungsmodelle unterteilt werden können in die Ebene der Messung und Repräsentation hypothetischer Konstrukte (latente Variablen mit ihren Indikatoren: **Messmodell**) sowie die Ebene der kausalen Beziehungen zwischen den Konstrukten (**Strukturmodell**). Die Messung der Konstrukte kann wiederum über **reflektive**, **formative** oder reflektive *und* formative **Indikatoren** erfolgen. Diese Art der Messung determiniert den gesamten Forschungsprozess, weshalb die Diskussion hinsichtlich der Art des Messmodells von entscheidender Bedeutung ist.

Durch die Gegenüberstellung von **Kovarianz-** und **Varianzanalyse** wird eine Entscheidungshilfe gegeben, welcher Analyseansatz (und damit welche Software, bspw. **LISREL** und **AMOS** vs. **SmartPLS**) im jeweiligen Forschungsprojekt Verwendung finden sollte. Es empfiehlt sich, diese Wahl aufgrund des konkreten Ziels sowie des Datenmaterials zu treffen.

Die Systematik des vierten Kapitels (Modellbeurteilung) hilft dabei, ein eigenes Strukturgleichungsmodell einer umfassenden Prüfung zu unterziehen. **Globale** und **lokale Gütemaße** helfen dabei, Schwachstellen eines Modells zu identifizieren. Sie sollten stets auch in vorgeschlagenen Fülle benutzt werden. Während sich die Gütemaße für die Messmodelle bei SmartPLS, LISREL und AMOS nicht unterscheiden, sind unterschiedliche Bewertungsverfahren hinsichtlich des Gesamtmodells notwendig.

---

<sup>25</sup> Üblicherweise werden in Aufsätzen Hypothesen als *bestätigt* statt, statistisch korrekter, als *nicht abgelehnt* bezeichnet.

Bei LISREL und AMOS gilt es, aus dem Kanon verfügbarer Fitindizes die passenden auszuwählen. Zu empfehlen sind **RMSEA**, **SRMR** (bzw. RMR bei AMOS), **CFI** und **NNFI** (TLI bei AMOS), eingeschränkt der  $\chi^2$ -Wert (durch die Anzahl der Freiheitsgrade dividiert) sowie der Vollständigkeit halber der AGFI. Für SmartPLS lassen sich **R<sup>2</sup>**, **f<sup>2</sup>** und, für die Prognoserelevanz der strukturellen Beziehungen, **Q<sup>2</sup>** bestimmen.

An dieser Stelle sei noch einmal betont, dass Gütemaße neben der Überprüfung auch der Verbesserung eines Modells dienlich sind. Ihnen blind zu folgen und theoretische Überlegungen hintenan zu stellen, nur um einen besseren Fit zu erzielen, entspricht allerdings nicht dem wissenschaftlichen Ziel solcher Untersuchungen. Werden theoretisch fundierte Hypothesen in einem Strukturgleichungsmodell abgebildet, dieses adäquat gemessen und anschließend gründlich hinsichtlich Reliabilität und Validität geprüft, steht dem Anwender freilich ein leistungsstarkes Instrument zur Verfügung, um tiefere Einblicke in Wirkungsbeziehungen zwischen den betrachteten Konzepten zu erlangen.

<b>Literaturempfehlungen</b>
<b>Grundlegendes</b>
Nunally/Bernstein (1994) zum psychometrischen Ansatz
Bagozzi (1994), Bollen (1989) und Hair et al. (2006) zur Kovarianzanalyse
Byrne (2001) zur Darstellung von AMOS und LISREL
Bagozzi/Baumgartner (1994) zur Beurteilung von SEM in der Kovarianzanalyse
Fornell/Larcker (1981), Hu/Bentler (1998) und Marsh/Balla/Hau (1996) zu Gütemaßen für LISREL/AMOS
Chin (1998b), Henseler (2005), Lohmöller (1989) und Tenenhaus et al. (2005) zum PLS-Ansatz
Fornell/Bookstein (1982), Herrmann/Huber/Kressmann (2006), Jöreskog/Wold (1982) und Scholderer/Balderjahn (2005) zum Vergleich der Kovarianz- und Varianzanalyse
Hoyle/Panter (1995) zur Darstellung der Ergebnisse einer Strukturgleichungsmodellierung
<b>Vertiefendes</b>
Churchill (1979) zum klassischen Paradigma der Skalenbildung, Gerbing/Anderson (1988) zu einer Weiterentwicklung
Rossiter (2002) zu einer Kritik und Alternative zum ‚Churchill-Paradigma‘
Bollen/Lennox (1991), Jarvis/MacKenzie/Podsakoff (2003) und MacCullum/Browne (1993) zu formativen Indikatoren
Diamantopoul/Winklhofer (2001) zur Indexkonstruktion mit formativen Indikatoren
Temme (2006) und Temme/Hildebrandt (2007) zur Spezifikation und Identifikation formativer Konstrukte in LISREL
Hu/Bentler (1999) und Sivo et al. (2006) zu Cutoff-Kriterien der Fitindizes bei LISREL/AMOS
Yuan/Bentler (2001) zum Einfluss von Ausreißern in der Kovarianzanalyse



**Literaturverzeichnis**

- Aaker, David A. und Richard P. Bagozzi (1979), „Unobservable Variables in Structural Equation Models with an Application in Industrial Selling,” *Journal of Marketing Research*, 16 (Mai), 147-158.
- Albers, Sönke und Lutz Hildebrandt (2006), „Methodische Probleme bei der Erfolgsfaktorenforschung – Messfehler, formative versus reflektive Indikatoren und die Wahl des Strukturgleichungs-Modells,” *Zeitschrift für betriebswirtschaftliche Forschung*, 58 (Februar), 2-33.
- Anderson, James C. und David W. Gerbing (1984), „The Effect of Sampling Error on Convergence, Improper Solutions, and Goodness-of-Fit Indices for Maximum Likelihood in Confirmatory Factor Analysis,” *Psychometrika*, 49 (2), 155-173.
- , ——— und John E. Hunter (1987), „On the Assessment of Unidimensional Measurement: Internal and External Consistency, and Overall Consistency Criteria,” *Journal of Marketing Research*, 24 (November), 432-437.
- Arbuckle, James L. und Werner Wothke (1999), *AMOS 4.0 User's Guide*. Chicago: Small Waters.
- Bagozzi, Richard P. (1984), „A Prospectus for Theory Construction in Marketing,” *Journal of Marketing*, 48 (Winter), 11-29.
- (1980), *Causal Models in Marketing*. New York u.a.: John Wiley & Sons.
- (1994), „Structural Equation Models in Marketing Research: Basic Principles“ in *Principles of Marketing Research*, Richard P. Bagozzi (Hrsg.). Cambridge, MA: Blackwell, 317-386.
- und Hans Baumgartner (1994), „The Evaluation of Structural Equation Models and Hypothesis Testing” in *Principles of Marketing Research*, Richard P. Bagozzi (Hrsg.). Cambridge, MA: Blackwell, 387-422.
- Bollen, Kenneth A. (2002), „Latent Variables in Psychology and the Social Sciences,” *Annual Review of Psychology*, 53, 605-634.
- (1984), „Multiple Indicators: Internal Consistency or No Necessary Relationship?,” *Quality and Quantity*, 18, 377-385.
- (1989), *Structural Equations with Latent Variables*. New York u.a.: Wiley.
- und Richard Lennox (1991), „Conventional Wisdom on Measurement: A Structural Equation Perspective,” *Psychological Bulletin*, 110 (2), 305-314.

- und J. Scott Long (1993), „Introduction“ in *Testing Structural Equation Models*, Kenneth A. Bollen und J. Scott Long (Hrsg.). Newbury Park: Sage, 1-9.
- Boomsma, Anne (1982), „The Robustness of LISREL against Small Sample Size in Factor Analysis Models“ in *Systems Under Direct Observations: Causality, Structure, Prediction*, Karl G. Jöreskog und Herman Wold (Hrsg.). Amsterdam: North-Holland, Teil 1, 149-174.
- Browne, Michael W. und Robert Cudeck (1993), „Alternative Ways of Assessing Model Fit“ in *Testing Structural Equation Models*, Kenneth A. Bollen und J. Scott Long (Hrsg.). Newbury Park: Sage, 136-161.
- Byrne, Barbara M. (2001), „Structural Equation Modeling with AMOS, EQS, and LISREL: Comparative Approaches to Testing for the Factorial Validity of a Measuring Instrument,“ *International Journal of Testing*, 1 (1), 55-86.
- Chin, Wynne W. (1998a), „Issues and Opinions on Structural Equation Modeling,“ *MIS Quarterly*, 22 (1), VII-XVI.
- (1995), „Partial Least Squares is to LISREL as Principal Components Analysis is to Common Factor Analysis,“ *Technology Studies*, 2, 315-319.
- (1998b), „The Partial Least Squares Approach to Structural Equation Modeling“ in *Modern Business Research Methods*, George A. Marcoulides (Hrsg.). Mahwah: Lawrence Erlbaum Associates, 295-336.
- Churchill, Gilbert A. (1979), „A Paradigm for Developing Better Measures of Marketing Constructs,“ *Journal of Marketing Research*, 16 (February), 64-73.
- Diamantopoulos, Adamantios (1999), „Viewpoint: Export Performance Measurement: Reflective versus Formative Indicators,“ *International Marketing Review*, 16 (6), 444-457.
- und Judy A. Siguaw (2000), *Introducing LISREL*. London u.a.: Sage.
- und Heidi M. Winklhofer (2001), „Index Construction with Formative Indicators: An Alternative to Scale Development,“ *Journal of Marketing Research*, 38, 269-277.
- Drengner, Jan, Hansjörg Gaus und Steffen Jahn (2007), „Image Effects of Marketing Events: The Impact of Flow Experiences,“ in *Proceedings of the AMA Winter Educators' Conference 2007*, Andrea Dixon und Karen A. Machleit (Hrsg.). San Diego (16th -19th February 2007), 87-88.
- Edwards, Jeffrey R. und Richard P. Bagozzi (2000), „On the Nature and Direction of Relationships Between Constructs and Measures,“ *Psychological Methods*, 5 (2), 155-174.
- Fan, Xitao und Stephen A. Sivo (2005), „Sensitivity of Fit Indexes to Misspecified Structural or Measurement Model Components: Rationale of Two-Index Strategy Revisited,“ *Structural Equation Modeling*, 12 (3), 343-367.

- Fornell, Claes und Fred L. Bookstein (1982), „Two Structural Equation Models: LISREL and PLS Applied to Consumer Exit-Voice Theory,” *Journal of Marketing Research*, 19 (November), 440-452.
- und Jaesung Cha (1994), „Partial Least Squares,” in *Advanced Methods of Marketing Research*, Richard P. Bagozzi (Hrsg.). Cambridge: Blackwell, 52-78.
- und David F. Larcker (1981), „Evaluating Structural Equation Models with Unobservable Variables and Measurement Error,” *Journal of Marketing Research*, 18 (February), 39-50.
- Gerbing, David W. und James C. Anderson (1988), „An Updated Paradigm for Scale Development Incorporating Unidimensionality and its Assessment,” *Journal of Marketing Research*, 25 (May), 186-192.
- Hair, Joseph F., William C. Black, Barry J. Babin, Rolph E. Anderson und Ronald L. Tatham (2006), *Multivariate Data Analysis* (6th ed.). Upper Saddle River, N.J.: Prentice-Hall.
- Hauser, Robert M. und Arthur S. Goldberger (1971), „The Treatment of Unobservable Variables in Path Analysis,” in *Sociological Methodology 1971*, Herbert L. Costner (Hrsg.). San Francisco: Jossey-Bass, 81-117.
- Henseler, Jörg (2005), „Einführung in die PLS-Pfadmodellierung,” *WiSt*, 2 (Februar), 70-75.
- Herrmann, Andreas, Frank Huber und Frank Kressmann (2006), „Varianz- und kovarianzbasierte Strukturgleichungsmodelle – Ein Leitfaden zu deren Spezifikation, Schätzung und Beurteilung,” *Zeitschrift für betriebswirtschaftliche Forschung*, 58 (Februar), 34-66.
- Hershberger, Scott L. (2003), „The Growth of Structural Equation Modeling: 1994-2001,” *Structural Equation Modeling*, 10 (1), 35-46.
- Hildebrandt, Lutz und Nicole Görz (1999), „Zum Stand der Kausalanalyse mit Strukturgleichungsmodellen – Methodische Trends und Software-Entwicklungen,” Sonderforschungsbereich 373: Quantification and Simulation of Economic Processes 46, ISSN: 1436-1086, Humboldt-Universität zu Berlin.
- und Dirk Temme (2006), „Probleme der Validierung mit Strukturgleichungsmodellen,” *DBW*, 66 (6), 618-639.
- Homburg, Christian und Annette Giering (1996), „Konzeptualisierung und Operationalisierung komplexer Konstrukte. Ein Leitfaden für die Marketingforschung,” *Marketing ZFP*, 18 (1), 5-24.
- und Martin Klarmann (2006), „Die Kausalanalyse in der empirischen betriebswirtschaftlichen Forschung – Problemfelder und Anwendungsempfehlungen,” *DBW*, 66 (6), 727-748.

- Hoyle, Rick H. und Abigail T. Panter (1995), „Writing About Structural Equation Models“ in *Structural Equation Modeling. Concepts, Issues, and Applications*, Rick H. Hoyle (Hrsg.). Thousand Oaks u.a.: Sage, 158-176.
- Hu, Li-tze und Peter M. Bentler (1998), „Fit Indices in Covariance Structure Modeling: Sensitivity to Underparameterized Model Misspecification“, *Psychological Methods*, 3 (4), 424-453.
- und——(1999), „Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria versus New Alternatives“, *Structural Equation Modeling*, 6, 1-55.
- Hulland, John (1999), „Use of Partial Least Squares (PLS) in Strategic Management Research: A Review of Four Recent Studies“, *Strategic Management Journal*, 20, 195-204.
- Jarvis, Cheryl B., Scott B. MacKenzie und Philip M. Podsakoff (2003), „A Critical Review of Construct Indicators and Measurement Model Misspecification in Marketing and Consumer Research“, *Journal of Consumer Research*, 30, 199-218.
- Jöreskog, Karl G. (1993), „Testing Structural Equation Models“ in *Testing Structural Equation Models*, Kenneth A. Bollen und J. Scott Long (Hrsg.). Newbury Park: Sage, 294-316.
- (1982), „The LISREL Approach to Causal Model-Building in the Social Sciences“ in *Systems Under Direct Observations: Causality, Structure, Prediction*, Karl G. Jöreskog und Herman Wold (Hrsg.). Amsterdam: North-Holland, Teil 1, 81-100.
- und Dag Sörbom (1996), *LISREL 8 User's Reference Guide*. Chicago, IL: Scientific Software.
- und Herman Wold (1982), „The ML and PLS Techniques for Modeling with Latent Variables: Historical and Comparative Aspects“ in *Systems Under Direct Observations: Causality, Structure, Prediction*, Karl G. Jöreskog und Herman Wold (Hrsg.). Amsterdam: North-Holland, part 1, 263-270.
- Johnson, Michael D., Andreas Herrmann und Frank Huber (2006), „The Evolution of Loyalty Intentions“, *Journal of Marketing*, 70 (April), 122-132.
- Krafft, Manfred, Oliver Götz und Kerstin Liehr-Gobbers (2005), „Die Validierung von Strukturgleichungsmodellen mit Hilfe des Partial-Least-Squares (PLS)-Ansatzes“ in *Handbuch PLS-Pfadmodellierung. Methode, Anwendung, Praxisbeispiele*, Friedhelm Bliemel, Andreas Eggert, Georg Fassott und Jörg Henseler (Hrsg.). Stuttgart: Schäffer-Poeschel, 71-86.
- Lohmöller, Jan-Bernd (1989), *Latent Variable Path Modeling with Partial Least Squares*. Heidelberg: Physica.
- MacCullum, Robert C. und Michael W. Browne (1993), „The Use of Causal Indicators in Covariance Structure Models: Some Practical Issues“, *Psychological Bulletin*, 114 (3), 533-541.

- Marsh, Herbert W., John R. Balla und Kit-Tai Hau (1996), „An Evaluation of Incremental Fit Indices: A Clarification of Mathematical and Empirical Properties“ in *Advanced Structural Equation Modeling*, George A. Marcoulides und Randall E. Schuhmacker (Hrsg.). Mahwah, NJ: Lawrence Erlbaum Associates, 315-353.
- , Kit-Tai Hau und Zhonglin Wen (2004), „In Search of Golden Rules: Comment on Hypothesis-Testing Approaches to Setting Cutoff Values for Fit Indexes and Dangers in Overgeneralizing Hu and Bentler’s (1999) Findings,“ *Structural Equation Modeling*, 11 (3), 320-341.
- Mueller, Ralph O. (1996), *Basic Principles of Structural Equation Modeling. An Introduction to LISREL and EQS*. New York u.a.: Springer.
- Nevitt, Jonathan und Gregory R. Hancock (2001), „Performance of Bootstrapping Approaches to Model Test Statistics and Parameter Standard Error Estimation in Structural Equation Modeling,“ *Structural Equation Modeling*, 8 (3), 353-377.
- Nunally, Jum C. und Ira H. Bernstein (1994), *Psychometric Theory* (3rd Ed.). New York u.a.: McGraw-Hill.
- Peter, J. Paul (1981), „Construct Validity: A Review of Basic Issues and Marketing Practices,“ *Journal of Marketing Research*, 18 (Mai), 133-145.
- Ringle, Christian M. (2004), „Gütemaße für den Partial Least Squares-Ansatz zur Bestimmung von Kausalmodellen,“ *Arbeitspapier Nr. 16*, Universität Hamburg.
- , Sven Wende und Alexander Will (2005), „*SmartPLS 2.0 (beta)*,“ <http://www.smartpls.de>, Abfrage: 27. Juni 2006.
- Rossiter, John R. (2002), „The C-OAR-SE Procedure for Scale Development in Marketing,“ *International Journal of Research in Marketing*, 19, 305-335.
- Scholderer, Joachim und Ingo Balderjahn (2005), „PLS versus LISREL: Ein Methodenvergleich“ in *Handbuch PLS-Pfadmodellierung. Methode, Anwendung, Praxisbeispiele*, Friedhelm Bliemel, Andreas Eggert, Georg Fassott und Jörg Henseler (Hrsg.). Stuttgart: Schäffer-Poeschel, 87-986.
- und———(2006), „Was unterscheidet harte und weiche Strukturgleichungsmodelle nun wirklich? Ein Klärungsversuch zur LISREL-PLS-Frage,“ *Marketing ZFP*, 28 (1), 57-70.
- Sivo, Stephen A., Xitao Fan, E. Lea Witta und John T. Willse (2006), „The Search for “Optimal” Cutoff Properties: Fit Index Criteria in Structural Equation Modeling,“ *The Journal of Experimental Education*, 74 (3), 267-288.
- Temme, Dirk (2006), „Die Spezifikation und Identifikation formativer Messmodelle der Marketingforschung in Kovarianzstrukturanalysen,“ *Marketing ZFP*, 28 (3), 183-196.

- und Lutz Hildebrandt (2007), „Formative Measurement Models in Covariance Structure Analysis: Specification and Identification,“ *Proceedings of the 36th EMAC Conference, Reykjavik (22-25 May 2007)*.
- Tenenhaus, Michel, Vincenzo E. Vinzi, Yves-Marie Chatelin und Carlo Lauro (2005), „PLS Path Modeling,“ *Computational Statistics & Data Analysis*, 48, 159-205.
- Wold, Herman (1982), „Soft Modeling: The Basic Design and Some Extensions” in *Systems Under Direct Observations: Causality, Structure, Prediction*, Karl G. Jöreskog und Herman Wold (Hrsg.). Amsterdam: North-Holland, Teil 2, 1-54.
- Wright, Sewall (1921), „Correlation and Causation,“ *Journal of Agricultural Research*, 20, 557-585.
- Yuan, Ke-Hai und Peter M. Bentler (2001), „Effect of Outliers on Estimators and Tests in Covariance Structure Analysis,“ *British Journal of Mathematical and Statistical Psychology*, 54, 161-175.