

Heft 21 / 2004

**EFO**plan

DIE ÜBERPRÜFUNG VON SPEZIFIKATION UND  
GÜTE VON STRUKTURGLEICHUNGSMODELLEN:  
VERFAHREN UND ANWENDUNG

Markus Zinnbauer,  
Markus Eberl

**Ludwig-Maximilians-Universität  
München**

Institut für Unternehmensentwicklung und Orga-  
nisation  
Seminar für Empirische Forschung und  
Unternehmensplanung  
Prof. Dr. Manfred Schwaiger

Kaulbachstr. 45 / I  
D-80539 München

Tel.: (089) 2180 5640  
Fax: (089) 2180 5651  
e-mail: [efoplan@bwl.uni-muenchen.de](mailto:efoplan@bwl.uni-muenchen.de)  
<http://www.efoplan.de>

**DIE ÜBERPRÜFUNG VON  
SPEZIFIKATION UND GÜTE VON  
STRUKTURGLEICHUNGSMODELLEN:  
VERFAHREN UND ANWENDUNG**

**Dipl.-Kfm. Markus Zinnbauer**

**Dipl.-Kfm. Markus Eberl, M.B.R.**

Department für Betriebswirtschaft der  
Ludwig-Maximilians-Universität München  
Institut für Unternehmensentwicklung und Organisation  
Seminar für Empirische Forschung und Unternehmensplanung EFOplan

Kaulbachstr. 45/I  
80539 München

Tel.: +49 (0) 89 / 2180 – 5633

Fax: +49 (0) 89 / 2180 – 99 – 5633

eMail: [Zinnbauer@bwl.uni-muenchen.de](mailto:Zinnbauer@bwl.uni-muenchen.de)

[Eberl@bwl.uni-muenchen.de](mailto:Eberl@bwl.uni-muenchen.de)

Web: [www.efoplan.de](http://www.efoplan.de)

---

# Inhaltsverzeichnis

<b>1</b>	<b>VALIDIERUNG VON STRUKTURGLEICHUNGSMODELLEN.....</b>	<b>2</b>
<b>2</b>	<b>DATENVORBEREITUNG UND MODELLSPEZIFIKATION .....</b>	<b>2</b>
2.1	DATENVORBEREITUNG .....	2
2.2	FORMATIVE VS. REFLEKTIVE MESSMODELLE .....	4
<b>3</b>	<b>BEURTEILUNG DER OPERATIONALISIERUNG.....</b>	<b>5</b>
3.1	LOKALE GÜTEKRITERIEN.....	6
3.1.1	Operationalisierung der reflektiven Konstrukte .....	6
3.1.2	Operationalisierung der formativen Konstrukte .....	9
3.2	Globale Gütekriterien .....	9
3.2.1	Globale Gütekriterien mit Referenzwerten .....	9
3.2.2	Relative globale Anpassungsmaße.....	12
3.3	KONSTRUKTERKLÄRUNG .....	13
<b>4</b>	<b>ANWENDUNGSHINWEISE .....</b>	<b>13</b>
<b>5</b>	<b>FALLSTUDIE:</b>	
	<b>STRUKTURGLEICHUNGSMODELLIERUNG MIT AMOS .....</b>	<b>14</b>
5.1	FORMATIVE VS. REFLEKTIVE MODELLSPEZIFIKATION .....	14
5.2	LOKALE GÜTEKRITERIEN.....	15
5.2.1	Operationalisierung reflektiver Konstrukte .....	15
5.2.2	Operationalisierung formativer Konstrukte.....	17
5.3	Globale Gütemaße .....	18
5.4	GESAMTBEURTEILUNG .....	21
	LITERATUR.....	22

## **Zusammenfassung**

Die Kausalanalyse und Strukturgleichungsmodelle gehören mittlerweile zum Standardrepertoire der empirischen Forschung. Ein Hauptproblem bei der Durchführung einer Kausalanalyse stellt dabei regelmäßig die sach- und fachgerechte Spezifikation des Modells und die Evaluation der Güte des Modells dar. Hierzu werden zunächst die wichtigsten Vorbereitungs- und Spezifikations-schritte dargestellt, welche bereits vor der eigentlichen Schätzung und Überprüfung der Hypothesen des Kausalmodells erfolgen. Wichtige Aspekte, wie etwa die Spezifikation eines Konstrukts als formativ oder reflektiv werden hierbei oft vernachlässigt und führen zu unzuverlässigen Ergebnissen. Zudem werden die wesentlichen Gütekriterien auf Indikatoren- als auch auf Modellebene vorgestellt, welche eine Beurteilung der Anpassungsgüte eines Kausalmodells erlauben. Zuletzt wird die Anwendung der dargestellten Überprüfungsschritte mittels der einschlägigen Software SPSS, AMOS und nicht zuletzt PLSGraph anhand einer Fallstudie dargestellt.

## 1 Validierung von Strukturgleichungsmodellen

Strukturgleichungsmodelle mit latenten Variablen lassen sich hauptsächlich auf *Jöreskog* (1973, 1978) bzw. *Jöreskog/Sörbom* (1979, 1982) zurückführen, wobei die von *Jöreskog* entwickelte Softwareapplikation LISREL besondere Bekanntheit erlangte. Im Marketingbereich wurden sie insbesondere von *Bagozzi* (z.B. 1980) weiterentwickelt, wobei das heutige Haupteinsatzgebiet die Überprüfung von theoretisch fundierten Erklärungsmodellen darstellt (vgl. *Steenkamp/Baumgartner*, 2000, S. 197). Statistisch basiert eine Kausalanalyse auf der Schätzung von Abhängigkeitsbeziehungen zwischen latenten Variablen eines Erklärungsmodells auf Grundlage von Varianzen und Kovarianzen zwischen den Modellindikatoren. Ein Strukturgleichungsmodell besteht somit aus zwei Teilmodellen: Einerseits werden in dem Messmodell (auch: konfirmatorisches Faktorenmodell) die beobachtbaren Indikatoren mit den latenten Variablen assoziiert und andererseits werden in dem Strukturmodell die Beziehungen zwischen den exogenen und endogenen (latenten) Variablen erfasst (vgl. z.B. *Leeflang et al.*, 2000, S. 443 f.). Dabei ist eine maßgebliche Eigenschaft, die kausalanalytische Verfahren von der gängigen multiplen Regressionsanalyse abhebt, die Fähigkeit auch komplexe Dependenzstrukturen, wie wechselseitige Beziehungen oder Wirkungsketten, modellieren zu können (vgl. *Homburg/Pflesser*, 1999, S. 636).

Ein Hauptproblem bei der Durchführung einer Kausalanalyse stellt dabei regelmäßig die sach- und fachgerechte Spezifikation des Modells und die Evaluation der Güte des Modells dar. Hierzu werden zunächst die wichtigsten Vorbereitungs- und Spezifikationsschritte dargestellt, welche bereits vor der eigentlichen Schätzung und Überprüfung der Hypothesen des Kausalmodells erfolgen. Wichtige Aspekte, wie etwa die Spezifikation eines Konstrukts als formativ oder reflektiv werden hierbei oft vernachlässigt und führen zu unzuverlässigen Ergebnissen. Im Anschluss werden die wesentlichen Gütekriterien auf Indikatoren- als auch auf Modellebene vorgestellt, welche eine Beurteilung der Anpassungsgüte eines Kausalmodells erlauben.

## 2 Datenvorbereitung und Modellspezifikation

### 2.1 Datenvorbereitung

Für eine kausalanalytische Untersuchung ist vor allem bei Fragebogenerhebungen grundsätzlich die Überprüfung der Daten auf Aspekte wie Skalierung,

fehlende Werte, Ausreißer und die vorliegende Verteilung empfehlenswert bzw. sogar notwendig (vgl. *Baumgartner/Homburg*, 1996).

Insbesondere die **Intervallskalierung** ist eine Grundvoraussetzung zur Schätzung von Strukturgleichungsmodellen (vgl. *Bagozzi*, 1981a, S. 200 und 1998b, S. 380). Die häufig in der betriebswirtschaftlichen und sozialwissenschaftlichen Empirie herangezogenen Ratingskalen dürfen nach der Meinung von Pragmatikern, sofern äquidistant, als quasi-metrisch betrachtet werden (vgl. z.B. *Bortz*, 1999, S. 27 f., *Jaccard/Wan*, 1996, S. 4). Um die kausalanalytischen Annahmen stetiger Variablen trotz diskreter Messung nicht schadhaft zu verletzen, werden aber mindestens fünf (vgl. *Bagozzi*, 1981b, S. 380) bzw. eher sieben Skalenpunkte empfohlen (vgl. *Bagozzi*, 1981a, S. 200).

**Fehlende Werte** führen entweder zu einem Ausschluss der jeweiligen Fälle (vgl. *Hulland et al.*, 1996) oder sie können bei völlig zufälligem Auftreten mittels Imputations- oder Parameterschätzverfahren ersetzt werden (vgl. für eine ausführliche Übersicht *Bankhofer*, 1995, S. 104 f., 154 ff.), um die Fallzahl nicht zu stark reduzieren zu müssen, was die Aussagekraft des Kausalmodells negativ beeinflusst (vgl. *Gerbing/Anderson*, 1993).

Als weiteres Problem ist eine mögliche Verzerrung der Stichprobe gegenüber der Grundgesamtheit aufgrund **Nicht-Antwortender** zu prüfen. Zur Schätzung, ob ein systematisches Problem vorliegt, wird gemeinhin mittels des nicht-parametrischen Mann-Whitney-U-Tests (vgl. *Mann/Whitney*, 1947) untersucht, ob sich früh und spät antwortende Teilnehmer strukturell unterscheiden. Dieses Vorgehen beruht auf der Überlegung, dass spät antwortende Teilnehmer den nicht antwortenden ähnlicher sind (vgl. dazu ausf. *Armstrong/Overton*, 1977).

Schließlich gilt es zu untersuchen, ob die für die kausalanalytische Untersuchung notwendige **Normalverteilung** gegeben ist (vgl. *Hulland et al.*, 1996), wenn das Maximum Likelihood-Prinzip angewandt wird. Diese Voraussetzung ist notwendig, um zur Signifikanzbestimmung t-Test und F-Statistik anwenden zu können (vgl. *Baumgartner/Homburg*, 1996). Da die Schätzung in der Regel auf dieser Annahme basiert, führt eine Verletzung zu einer Verschlechterung der Gütekriterien (vgl. *Hulland et al.*, 1996). Streng genommen verletzen Ratingskalen die Voraussetzungen der Normalverteilung. In der Literatur wird allerdings davon ausgegangen, dass Variablen, deren Werte über Ratingskalen erhoben wurden, unter bestimmten Bedingungen als normalverteilt betrachtet werden dürfen (vgl. z.B. *Hofacker*, 1984, S. 93). Daher bietet sich ein Test auf Normalverteilung oder eine Inaugenscheinnahme anhand von P-P-Verteilungsdiagrammen an (vgl. *D'Agostino/Belanger*, 1990, S. 319 f.).

## 2.2 Formative vs. reflektive Messmodelle

Die verwendeten Modellkonstrukte stellen i.d.R. a priori nicht direkt messbare, Variablen dar. Um die Wirkungsbeziehungen in einem Strukturgleichungsmodell abbilden zu können, ist deshalb zunächst eine Operationalisierung mittels eines Messmodells nötig (vgl. *Anderson/Gerbing*, 1982, S. 453). Zur Messbarmachung dieser latenten Variablen ist es notwendig, mehrere, wiederum direkt beobachtbare bzw. manifeste Indikatoren einzusetzen, die das Konstrukt möglichst vollständig erklären (vgl. *Homburg/Dobratz*, 1991, S. 214). Bei der konstruktsspezifischen Indikatorenbildung ist allerdings zunächst zu berücksichtigen, ob ein formatives oder reflektives Messmodell spezifiziert werden sollte (vgl. *Bollen/Lennox*, 1991, *Homburg*, 2000, *Law/Wong*, 1999, *Diamantopoulos/Winklhofer*, 2001). Grundsätzlich können die beiden Modellspezifikationen wie folgt unterschieden werden (vgl. dazu ausführlich *Eberl*, 2004 sowie *Eggert/Fassott*, 2003):

Ein **reflektives Messmodell** basiert auf der Auffassung, dass die zugeordneten, direkt beobachtbaren Indikatoren von der latenten Variablen kausal beeinflusst werden (vgl. *Bollen*, 1989, S. 182, *Edwards/Bagozzi*, 2000, S. 157 f.). Exogene Einflüsse auf die latente Variable wirken sich somit theoretisch gleichzeitig indirekt auf die reflektiven Indikatoren aus. Falls die Indikatoren die latente Variable fehlerfrei zu messen imstande sind, sind sie vollständig korreliert und damit praktisch beliebig austauschbar (vgl. *Jarvis et al.*, 2003, S. 2000). Daher sollten in einem reflektiven Messmodell die Indikatoren immer hoch korreliert sein (vgl. *Bollen/Lennox*, 1991, S. 308), einer Überlegung auf der auch *Churchills* (1979a) Vorschlag zur Gütebeurteilung zur Skalenbereinigung auf der Basis von Korrelationen zwischen den reflektiven Items beruht. Indikatoren, die keine hohe Korrelation aufweisen, müssen dementsprechend eliminiert werden, da sie offensichtlich eher anderen, ggf. im Modell auch nicht enthaltenen, Konstrukten zuzuordnen sind.

Im Gegensatz dazu wird in einem **formativen Messmodell** von einer Umkehr der Wirkbeziehung zwischen direkt beobachtbaren Indikatoren und der latenten Variablen ausgegangen. Die Indikatoren verursachen also die Latente (vgl. *MacCallum/Browne*, 1993, S. 533). Allerdings müssen die Indikatoren nicht unbedingt miteinander korreliert sein (vgl. *Jarvis et al.*, 2003, S. 201 f.), da kausal nur eine Beziehung zwischen jedem formativen Indikator und der latenten Variablen bestehen muss. Korrelationen zwischen inhaltlich erklärungsrelevanten, jedoch nicht vollständig überschneidungsfreien Items sind aber durchaus akzeptabel (vgl. *Diamantopoulos/Winklhofer*, 2001, S. 271). Deshalb führt die

Veränderung eines jeden Indikators zu einer Veränderung der kausal von allen relevanten manifesten Indikatoren insgesamt abhängenden latenten Variablen.

Die mögliche Unkorreliertheit der Indikatorvariablen (vgl. *Bollen*, 1984, S. 377) erschwert aber nunmehr die Gütebeurteilung einer Operationalisierung. Schließlich greifen herkömmliche Methoden, die auf der internen Konsistenz und respektive auf dem Zusammenhang der manifesten Indikatoren beruhen, nicht mehr (vgl. *Bollen*, 1989, S. 222, *Rossiter*, 2002, S. 307 f.). Da ein formatives Konstrukt von der Gesamtheit seiner Indikatoren definiert wird, wird die Eliminierung gering korrelierter Indikatoren aus modelltheoretischen Überlegungen heraus nicht erforderlich und ist vor allem aus konzeptionellen Überlegungen nicht unbedingt zulässig (vgl. *Eggert/Fassott*, 2003, S. 6). Die traditionellen Verfahren zur Beurteilung der Güte sollten daher ausschließlich auf reflektiv operationalisierte Konstrukte begrenzt bleiben (vgl. *Diamantopoulos*, 1999, S. 453 f.). Insofern werden entweder die Überlegungen des Forschers anhand von Entscheidungsfragen (vgl. *Chin*, 1998, S. 9 oder *Jarvis et al.*, 2003, S. 203) oder die externe Validität (zumeist im Sinne einer Expertenvalidität) zum Kerngütekriterium für formativ operationalisierte Konstrukte (vgl. z.B. *Rossiter*, 2002).

*Eberl* (2004, S. 16) empfiehlt die Anwendung von folgenden Entscheidungsfragen, um eine Hypothese über die geeignete Spezifikation des Messmodells einer latenten Variablen zu erhalten:

<b>Entscheidungsfragen, die im positiven Fall auf eine reflektive Spezifikation hindeuten</b>
Ist das Konstrukt kausal für die Indikatoren?
Ändern alle Indikatoren die Richtung, wenn ein Indikator die Richtung ändert?
Sind die Indikatoren eines Konstruktes beliebig austauschbar?

**Tabelle 1: Entscheidungsraster zur Bestimmung der Konstruktspezifikation**

### 3 Beurteilung der Operationalisierung

Die Beurteilung der Güte einer Operationalisierung erfolgt sowohl auf lokaler (konstruktbezogener) als auch auf globaler (modellbezogener) Ebene. Speziell bei den lokalen Gütekriterien ist dabei zwischen reflektiven und formativen Konstrukten zu unterscheiden.

### 3.1 Lokale Gütekriterien

#### 3.1.1 Operationalisierung der reflektiven Konstrukte

Da die Indikatoren zunächst lediglich in einem hypothetischen Zusammenhang mit dahinter stehenden, reflektiven Konstrukten stehen, gilt es empirisch zu überprüfen, ob die Indikatoren die entsprechenden Konstrukte auch geeignet messen (vgl. *Schnell et al.*, 1999, S. 125 ff.). Falls Unklarheit darüber besteht, ob es sich um ein einzelnes Konstrukt oder um mehrere Konstrukte handelt, ist es empfehlenswert, zunächst eine explorative Faktorenanalyse durchzuführen, um die zugrunde liegende Faktorstruktur zu untersuchen (vgl. z.B. *Backhaus et al.*, 2003, S. 259 ff.). Eine entsprechende Rotation vereinfacht die inhaltliche Interpretation der Ergebnisse und Zuordnung der Indikatoren zu den einzelnen Konstrukten.

Für eine Gruppe von Indikatoren erfolgt dann zunächst die Überprüfung der Interne-Konsistenz-Reliabilität, wobei das so genannte **Cronbachs Alpha** berechnet und zur Beurteilung herangezogen wird (vgl. z.B. *Cronbach*, 1951, *Peter* 1979). Je höher die Kovarianzen bzw. Korrelationen zwischen den Indikatorvariablen sind, desto mehr nähert sich Cronbachs Alpha, dessen Wertebereich sich von Null bis Eins erstreckt, dem Wert Eins an. Cronbachs Alpha ist definiert als

$$\alpha = \frac{n}{n-1} \left( 1 - \frac{\sum \sigma_i^2}{\sigma_x^2} \right) \quad (1)$$

mit  $n$ : Anzahl der Indikatoren,  $\sigma_i^2$ : Varianz des  $i$ -ten Indikators und  $\sigma_x^2$ : Varianz des gesamten Tests (vgl. *Cronbach*, 1951, S. 299).

Der Grenzwert für eine akzeptable Reliabilität ist umstritten. So wird speziell bei drei und zwei Indikatoren auch schon ein Wert von mindestens 0,4 als akzeptabel erachtet (vgl. *Peter* 1997, S. 180). Allerdings wird meist der Forderung von *Nunnally* (1978, S. 245) gefolgt, der für Cronbachs Alpha seit langem einen Mindestwert von 0,7 postuliert.

Als weiteres Gütekriterium wird die **Item-to-Total-Korrelation** herangezogen. Diese ist als Korrelation einer Indikatorvariablen mit der Summe aller Indikatoren, die demselben Faktor zugeordnet sind, definiert (vgl. z.B. *Homburg/Giering*, 1996, S. 8).

Solange der Reliabilitätswert zu gering ist, wird sukzessive derjenige dem Faktor zugeordnete Indikator mit der niedrigsten Item-to-Total-Korrelation aus der weiteren Analyse ausgeschlossen (vgl. *Churchill*, 1979b, S. 68). Eine derartige

Eliminierung ist unproblematisch, da die Indikatoren prinzipiell austauschbare Messungen der latenten Variablen darstellen (vgl. *Bollen/Lennox*, 1991, S. 308). Gemäß den Empfehlungen von *Bearden et al.* (1989, S. 475) sollte jeder Indikator eine Item-to-Total-Korrelation von mindestens 0,5 aufweisen.

Die reduzierten Skalen werden einer **explorativen Faktorenanalyse** unterzogen, um zu überprüfen, ob die Indikatoren einen einzelnen Faktor bilden, was auf Konvergenzvalidität hinweist (vgl. z.B. *Robinson et al.*, 1991 oder *Bearden et al.*, 1993). Anderenfalls machen eventuell mehrere Faktoren eine Aufspaltung in mehrere Konstrukte erforderlich. Darüber hinaus wird in diesem Zusammenhang gefordert, dass durch den extrahierten Faktor mindestens 50 % der Varianz der zugehörigen Indikatoren erklärt werden und die jeweiligen Faktorladungen mindestens 0,4 je Indikator (bzw. im einfaktoriellen Fall die Kommunalitäten mind. 0,16) aufweisen (vgl. *Peter*, 1997, S. 197 f.).

Die restlichen, identifizierten Indikatoren eines Faktors werden anschließend mit Hilfe der **konfirmatorischen Faktorenanalyse** untersucht (vgl. *Aaker/Bagozzi* 1979, S. 149), wobei das entsprechende Modell eine einfaktorielle Struktur postuliert. Die lokalen Anpassungsmaße Indikator- und Faktorreliabilität beziehen sich auf das Messmodell und beurteilen hauptsächlich Reliabilitäts- und Validitätsaspekte der Messung der hypothetischen Konstrukte durch die jeweiligen Indikatorvariablen (vgl. *Homburg/Baumgartner*, 1995, S.170).

Die **Reliabilität eines Indikators**  $x_i$  beschreibt, wie gut dieses Item die latente Variable  $\xi_j$  misst und ist nach *Bagozzi* (1982, S. 156) definiert als

$$rel(x_i) = \frac{\lambda_{ij}^2 \phi_{jj}}{\lambda_{ij}^2 \phi_{jj} + \theta_{ii}} \quad (2)$$

mit  $\lambda_{ij}$  : geschätzte Faktorladung,  $\phi_{jj}$  : geschätzte Varianz der latenten Variablen  $\xi_j$  und  $\theta_{ii}$  : geschätzte Varianz des zugehörigen Messfehlers der Indikatorvariablen. Die **Faktorreliabilität** und die **durchschnittlich erfasste Varianz** (DEV) geben an, wie gut eine latente Variable durch alle ihr zugeordneten Indikatoren gemessen wird. Die Faktorreliabilität und die DEV einer latenten Variablen  $\xi_j$  errechnen sich nach den Formeln

$$rel(\xi_j) = \frac{\left( \sum_{i=1}^k \lambda_{ij} \right)^2 \phi_{jj}}{\left( \sum_{i=1}^k \lambda_{ij} \right)^2 \phi_{jj} + \sum_{i=1}^k \theta_{ii}} \quad (3) \quad \text{sowie} \quad DEV(\xi_j) = \frac{\sum_{i=1}^k \lambda_{ij}^2 \phi_{jj}}{\sum_{i=1}^k \lambda_{ij}^2 \phi_{jj} + \sum_{i=1}^k \theta_{ii}} \quad (4)$$

wobei die Summation jeweils über alle  $k$  zugeordneten Indikatoren erfolgt (vgl. z.B. *Fornell/Larcker*, 1981, S. 45 f.; *Balderjahn*, 1986, S. 118; *Bagozzi/Yi*, 1988, S. 80).

Darüber hinaus empfiehlt sich auch im einfaktoriellen Fall die Betrachtung des Gütekriteriums  $\chi^2$ /Freiheitsgrade, das u.a. im Folgenden diskutiert wird.

Als Resultat dieser Untersuchungsstufe enthält jeder einzelne Faktor eine unter den Gesichtspunkten der Reliabilität und der **Konvergenzvalidität** bereinigte Indikatormenge.

Darauf aufbauend wird im Rahmen der Analyse der **Diskriminanzvalidität** überprüft, inwieweit die entwickelten und einzeln operationalisierten, reflektiven Konstrukte tatsächlich auch eigenständige Faktorgebilde darstellen.

Dazu wird zunächst eine explorative Faktorenanalyse über sämtliche Indikatoren der reflektiven exogenen, moderierenden und endogenen Konstrukte durchgeführt. Im günstigen Fall reproduzieren die explorativ ermittelten Faktoren exakt jene reflektiven Konstrukte, die theoretisch vorgegeben und operationalisiert wurden.

Um die Diskriminanzvalidität weitergehend zu untersuchen, werden jeweils zweifaktorielle konfirmatorische Faktorenanalysen durchgeführt (vgl. *Anderson/Gerbing*, 1988, S. 416). Dabei wird zunächst jedes paarweise Modell unrestringiert, d.h. ohne Einschränkung bezüglich möglicher interfaktorieller Korrelationen, geschätzt und dann in seiner restringierten Form (Korrelation zwischen den Konstrukten=1) erneut auf seine, aufgrund der Restriktion freilich schlechteren, Güte überprüft.

Die Beurteilung der Diskriminanzvalidität erfolgt u.a. anhand des  $\chi^2$ -**Differenztests** (vgl. z.B. *Bagozzi*, 1980, S. 144, *Bentler/Bonett*, 1980, S. 593 f. oder *Jöreskog*, 1971, S. 117 ff.). Die Differenz zwischen den  $\chi^2$ -Gütewerten der beiden Modellvarianten liefert eine Einschätzung, inwieweit sich jeweils zwei Konstrukte voneinander unterscheiden. Der maßgebliche Grenzwert, der für eine Aussagekraft zum üblichen 5 %-Niveau überschritten werden sollte, liegt bei 3,841 (vgl. *Homburg*, 2000, S. 93 f.).

Um die Diskriminanzvalidität weitergehend zu prüfen wird darüber hinaus auf das prinzipiell strengere **Fornell/Larcker-Kriterium** zurückgegriffen. Danach wird gefordert, dass die durchschnittlich erfasste Varianz eines Faktors stets größer als jede quadrierte Korrelation dieses Faktors mit einem anderen Faktor sein muss (vgl. *Fornell/Larcker*, 1981, S. 46).

### 3.1.2 Operationalisierung der formativen Konstrukte

Bei der Operationalisierung der formativen Konstrukte besteht das besondere Problem darin, dass faktisch keine geeigneten statistischen Methoden existieren, um die Operationalisierungsgüte zu beurteilen. Vielmehr bildet hier, wie weiter oben dargelegt, die **Validierung durch Experten** bzw. den Forscher selbst ein maßgebliches Kriterium für die inhaltlich vollständige Erfassung eines Konstruktes mittels mehrerer Indikatoren bei gleichzeitig forschungsökonomisch begrenzter Indikatorenanzahl und weitgehender Überschneidungsfreiheit mit anderen Konstrukten.

Um einen Hinweis auf die Güte der Indikatoren zur formativen Erklärung des jeweiligen Konstruktes zu erhalten, können z.B. mittels des **PLS-Algorithmus** in Verbindung mit einer Bootstrapping-Prozedur Standardfehler und t-Statistik für die geschätzte Ladung jedes Indikators auf die Latente berechnet werden (vgl. *Tenenhaus*, 2003).

PLS ermittelt die Schätzparameter, respektive die Ladungen der formativen Indikatoren auf die latente Variable, durch den Einsatz von Gewichten als Hilfsvariablen. Die im vorliegenden Fall einzelne latente Variable wird aus einer gewichteten Linearkombination ihrer Indikatoren gebildet (vgl. *Lohmöller*, 1989, S. 29 f.). Der Algorithmus wird für ein Messmodell dann iterativ durchlaufen, um die Residuen zu minimieren und das Modell möglichst optimal an die Realdaten anzupassen, also Konvergenz zu erreichen (vgl. *Chin/Newsted*, 1999, S. 316, *Cassel et al.*, 1999, S. 438 oder *Fornell/Cha*, 1994, S. 64-69).

## 3.2 Globale Gütekriterien

Zur Beurteilung des gesamten Strukturmodells werden so genannte globale Gütekriterien herangezogen. Die Grundidee besteht bei allen globalen Gütekriterien darin, die vom Modell reproduzierte Varianz-/Kovarianzmatrix ( $\Sigma$ ) mit der empirischen Varianz-/Kovarianzmatrix ( $S$ ) der Stichprobe zu vergleichen. Die Anpassung des Modells an den zugrunde liegenden Datensatz wird umso positiver beurteilt, je geringer diese Abweichung ist.

### 3.2.1 Globale Gütekriterien mit Referenzwerten

Grundsätzlich ist zwischen globalen Gütekriterien mit Referenzwerten und solchen ohne Referenzwerte (Vergleichsstandards) zu unterscheiden. Die Existenz solcher Referenzwerte ermöglicht dem Anwender die Interpretation des berechneten Wertes. Sie werden daher auch als Stand-Alone-Kriterien bezeichnet. Gütekriterien, für die keine Referenzwerte existieren, bezeichnet man als relativ.

### 3.2.1.1 Stand-Alone-Gütekriterien

**Inferenzstatistische Gütekriterien** beurteilen die Anpassungsgüte eines Modells mit Hilfe von statistischen Tests. Die beiden bekanntesten Statistiken sind der  $\chi^2$ -**Anpassungstest** sowie der **RMSEA** (vgl. *Homburg/Baumgartner*, 1995, S. 166).

Unter der Voraussetzung, dass die Stichprobe ausreichend groß ist, entspricht das  $\chi^2$ - **Anpassungsmaß** einer Likelihood-Ratio-Teststatistik, um die Nullhypothese ( $H_0$ ), dass die modelltheoretische Kovarianz-Matrix ( $\Sigma$ ) der empirischen Kovarianz-Matrix ( $S$ ) entspricht, gegen die Alternativhypothese ( $H_1$ ), dass  $S$  einer beliebigen, positiv definiten Matrix entspricht, zu testen (vgl. *Backhaus*, 2003, S. 373). Es berechnet sich als  $\chi^2 = (n - 1)F(S, \Sigma)$ . (5)

Der  $\chi^2$ -Wert ist umso kleiner, je geringer die Differenz ( $S - \Sigma$ ) ist. Die sich aus dieser Gleichung ergebende Prüfgröße ist  $\chi^2$ -verteilt mit  $df = \frac{1}{2}(p + q)(p + q + 1) - t$  Freiheitsgraden (vgl. *Jöreskog/Sörbom*, 1982, S. 407), was eine Beurteilung des  $\chi^2$ -wertes mittels klassischer Testtheorie zulässt.

Der  $\chi^2$ -Anpassungstest ist allerdings mit einer Reihe von Schwachstellen behaftet und wird für gewöhnlich als hochproblematisch angesehen. Harte Verteilungsannahmen und rigide Voraussetzungen stellen meist ein Problem für praktische Anwendung der Inferenz dar. Bei ausreichend großem Stichprobenumfang wird zudem nahezu jedes Modell abgelehnt. (vgl. *Jöreskog/Sörbom*, 1982, S.408, *Bagozzi*, 1981, S. 380). Darüber hinaus kann die Gütefunktion der  $\chi^2$ -Statistik und damit der Fehler 2. Art nicht angegeben werden (vgl. *Förster et al.*, 1984, S. 361).

Um die Probleme des  $\chi^2$ -Maßes zu umgehen, empfehlen *Browne/Cudeck* (1993) die Verwendung des ursprünglich von *Steiger/Lind* (1980) entwickelten **RMSEA** (Root Mean Square Error of Approximation):

$$\text{RMSEA} = \sqrt{\frac{\hat{F}_0}{df}} = \sqrt{\text{Max}\left\{\left(\frac{\hat{F}}{df} - \frac{1}{n}\right), 0\right\}} \quad (6)$$

*Browne/Cudeck* (1993) unterscheiden drei Diskrepanztypen: (1) die auf die Approximation bezogene, (2) die auf die Schätzung bezogene, sowie (3) die auf den Gesamtfehler bezogene Diskrepanz. Von besonderem Interesse ist dabei die erste Diskrepanz, weswegen der RMSEA auch besonderes Augenmerk darauf legt: Er ist ein Maß für die Diskrepanz zwischen  $\Sigma_0$  und  $\tilde{\Sigma}_0$  (ausgedrückt

in der Diskrepanzfunktion  $\hat{F}_0$ ) pro Freiheitsgrad. Der RMSEA hat einen Definitionsbereich zwischen 0 und 1, wobei Werte kleiner als 0,08 im Allgemeinen für eine gute Modellanpassung sprechen.

Im Gegensatz zu den inferenzstatistischen Anpassungsmaßen erfolgt die Beurteilung der Anpassungsgüte mittels **deskriptiver Gütekriterien** nicht auf der Basis von statistischen Tests, sondern anhand von Faustregeln (vgl. *Homburg/Baumgartner*, 1995, S. 166). Prinzipiell lassen sich deskriptive Gütemaße dahingehend unterscheiden, ob sie die Zahl der Freiheitsgrade des Modells berücksichtigen oder nicht. Letztere Gruppe hat jedoch eine „sehr begrenzte Aussagekraft“ (*Homburg/Baumgartner*, 1995, S. 166), da ggf. überparametrisierte Modelle bevorzugt würden. Die Maße dieser Kategorie (GFI, Scaled Likelihood Ratio) werden hier deshalb nicht vorgestellt und sollten in der Anwendung nur äußerst zurückhaltend eingesetzt werden.

Der  $\chi^2$ -Wert ist zwar nicht als inferenzstatistisches Gütekriterium brauchbar, jedoch nach Standardisierung mit der Zahl der Freiheitsgrade als deskriptives Anpassungsmaß sehr gut anwendbar (vgl. *Jöreskog/Sörbom*, 1982, S. 408). Von einem gut angepassten Modell kann dann ausgegangen werden, wenn das **Verhältnis von  $\chi^2$ -Wert zu den Freiheitsgraden** kleiner oder gleich 2,5 ist (vgl. *Homburg/Baumgartner*, 1995, S. 169.).

Der **adjusted Goodness-of-Fit Index (AGFI)** ist eine Erweiterung des hier nicht in Tiefe diskutierten GFI um die Freiheitsgrade des Modells. Der AGFI ist ein Maß für die im Modell erklärte Varianz (vgl. *Jöreskog/Sörbom*, 1981, S.1.41.) und berechnet sich als

$$AGFI = 1 - \left[ \frac{(p+q)(p+q+1)}{2df} \right] (1 - GFI) \quad (7)$$

Der Wertebereich des AGFI liegt zwischen 0 und 1. Der Fit des Modells ist umso besser, je mehr sich der AGFI an 1 annähert (vgl. *Jöreskog/Sörbom*, 1981, S. 1.40).

### 3.2.1.2 Inkrementelle Anpassungsmaße

Bei diesen Gütekriterien wird untersucht, inwieweit sich die Anpassungsgüte beim Übergang eines sog. Basismodells, welches keine inhaltliche Plausibilität aufweist, zum relevanten Modell ( $M_k$ ) verbessert. Als Basis- oder auch Nullmodell ( $M_1$ ) dient ein besonders schlecht angepasstes Modell, in dem alle ge-

messenen Variablen als unkorreliert angenommen werden. Dagegen ist das sog. saturierte Modell ( $M_s$ ) das am Besten an die Daten angepasste Modell.

Der **Normed Fit Index (NFI)** setzt die  $\chi^2$ -Werte des relevanten Modells ( $\chi_k^2$ ) zu den  $\chi^2$ -Werten des Basismodells ( $\chi_i^2$ ) in Relation:

$$\text{NFI} = \frac{\chi_i^2 - \chi_k^2}{\chi_i^2} \quad (8)$$

Der NFI ist auf das Intervall (0,1] normiert. *Bentler/Bonnett* (1980, S. 600) verdeutlichen, dass Werte ab 0,9 erfahrungsgemäß als zufrieden stellend betrachtet werden können. Auf *James/Mulaik/Brett* (1982) geht der Ansatz zurück, den NFI mit dem Verhältnis der Freiheitsgrade des relevanten Modells und des Basismodells zu multiplizieren und so zum Gütemaß **PNFI (Parsimony Normed Fit Index)** zu gelangen. Der PNFI ist ein gutes Maß um verschiedene Modelle, die sich in der Anzahl der Modellparameter und Freiheitsgrade unterscheiden, zu vergleichen. Daneben lässt sich auch die Anwendung des um den Stichprobenumfang bereinigten Maßes **NNFI (Non-Normed Fit Index)** beobachten, das auch als TLI (Tucker-Lewis Index) bekannt ist (vgl. *Tucker/Lewis*, 1973). Es ist jedoch in Anwendung und Interpretation nicht unumstritten.

### 3.2.2 Relative globale Anpassungsmaße

Relative globale Anpassungsmaße vergleichen mehrere, alternative Kausalmodelle untereinander und weniger, wie gut ein einzelnes Modell an die Daten angepasst ist. Damit können sie auch zur Modellselektion eingesetzt werden.

Mit Hilfe von **Akaike's Informationskriterium (AIC)** wird aus einer Reihe von alternativen Modellen das Modell ausgewählt, welches bei gleichem Genauigkeitsgrad die geringste Komplexität aufweist (vgl. *Bozdogan*, 1987, S. 345 f.):  $AIC = \chi^2 + 2t$  (vgl. *Akaike*, 1987). Von mehreren alternativen Modellen ist dasjenige mit dem geringsten AIC zu bevorzugen (vgl. *Homburg/Baumgartner*, 1995, S. 169).

**Bayes' Informationskriterium (BIC)** zieht sparsame Modellspezifikationen gegenüber Modellen mit vielen, teils irrelevanten, Parametern vor (vgl. *Raftery*, 1993, S. 164 f.). Die Größe ist dabei eine Approximation für den sog. Bayes-Faktor. Werden zwei konkurrierende Modelle  $M_0$  und  $M_1$  bei gegebenem Datensatz  $D$  betrachtet, gibt er an, wie gut das Modell  $M_0$  im Vergleich zum Modell  $M_1$  die Daten vorhersagt. Nach *Raftery* (1995, S. 134) können mit BIC zwei Modelle leicht verglichen werden: Das Modell mit dem kleineren BIC-Wert wird

bevorzugt. Der BIC- Wert kann neben der Berechnung in Relation zum saturierten Modell, auch in Relation zu einem Nullmodell berechnet werden.

### 3.3 Konstrukterklärung

Zur abschließenden Beurteilung werden die erklärten Varianzanteile der endogenen latenten Konstrukte betrachtet. Mit Hilfe der quadrierten multiplen Korrelation ist eine Beurteilung einzelner Gleichungen des Strukturgleichungsmodells wie folgt möglich:

$$qmk(\eta_j) = 1 - \frac{\Psi_{jj}}{\text{var}(\eta_j)} \quad (9)$$

mit  $\text{var}(\eta_j)$ : geschätzte Varianz des Faktors  $\eta_j$ ,  $\Psi_{jj}$ : geschätzte Varianz der Fehlervariablen  $\zeta_j$  (vgl. *Homburg/Baumgartner*, 1995, S. 170f). Üblicherweise werden Aufklärungswerte von mindestens 0,4 zur Konstrukterklärung gefordert. Falls der Zweck der Untersuchung jedoch ausschließlich in der Überprüfung der nomologischen Validität liegt, ist dieses Gütemaß nicht von sonderlicher Relevanz (vgl. *Homburg/Pflesser* 1999, S. 652).

## 4 Anwendungshinweise

Die Kausalanalyse bietet ein umfangreiches und äußerst hilfreiches Instrumentarium zur Überprüfung von Zusammenhängen zwischen latenten Größen. Mit den vielen Möglichkeiten geht jedoch auch eine unerschöpfliche Menge an denkbaren Fehlspezifikationen einher. Schon bei der Modellbildung ist auf die richtige Spezifikation der Konstrukte zu achten. Sind diese Voraussetzungen für die Anwendung jedoch erfüllt, stellt sich das Problem, die Anpassungsgüte der empirischen Daten an das entwickelte Modell zu überprüfen. Eine akzeptable Güte ist die unbedingte Voraussetzung, um schließlich Pfade zwischen den Konstrukten, also die eigentlich interessierenden Kausalbeziehungen zu überprüfen. Zur Beurteilung der Güte des Modells stehen im Wesentlichen die in diesem Beitrag vorgestellten Gütemaße zur Verfügung. Je nach Ziel der vorgenommenen Untersuchung ist dabei aus dem Spektrum der Gütemaße dasjenige bzw. diejenigen auszuwählen, die für die jeweilige Untersuchung sinnvoll sind. Dies sollte sinnvollerweise geschehen, **bevor** die Auswertung beginnt und nicht, nachdem das eigentlich favorisierte Gütemaß zu einer negativen Beurteilung einzelner Konstrukte oder des Gesamtmodells gelangt.

## 5 Fallstudie: Strukturgleichungsmodellierung mit AMOS

In den vorherigen Kapiteln wurde dargestellt, welche grundsätzlichen Schritte notwendig sind und welche Gütekriterien es zu unterscheiden gilt. In diesem Tutorial wird die konkrete Anwendung anhand der Softwareapplikationen AMOS 5, SPSS 12.0 und PLS-Graph 3.0 vorgestellt. AMOS lässt grundsätzlich die Spezifikation formativer und reflektiver Konstrukte innerhalb eines Modells zu.

Als Fallbeispiel wird in einem vereinfachten Modell untersucht, wie sich die Übernahme gesellschaftlicher Verantwortung durch ein Unternehmen auf die Reputationsdimensionen Kompetenz und Sympathie auswirkt. Der Datensatz basiert auf einer repräsentativen Befragung 6.868 zufällig ausgewählter Haushalte aus D, UK und USA. Auf die inhaltliche Theorie zur Reputation soll an dieser Stelle zugunsten der methodischen Vorgehensweise nicht eingegangen werden (hierzu sei auf *Schwaiger, 2004* verwiesen).

### 5.1 Formative vs. reflektive Modellspezifikation

Zunächst gilt es zu prüfen, ob die Konstrukte formativ oder reflektiv zu spezifizieren sind. Dies erfolgt exemplarisch anhand des Konstruktes „Verantwortung“.

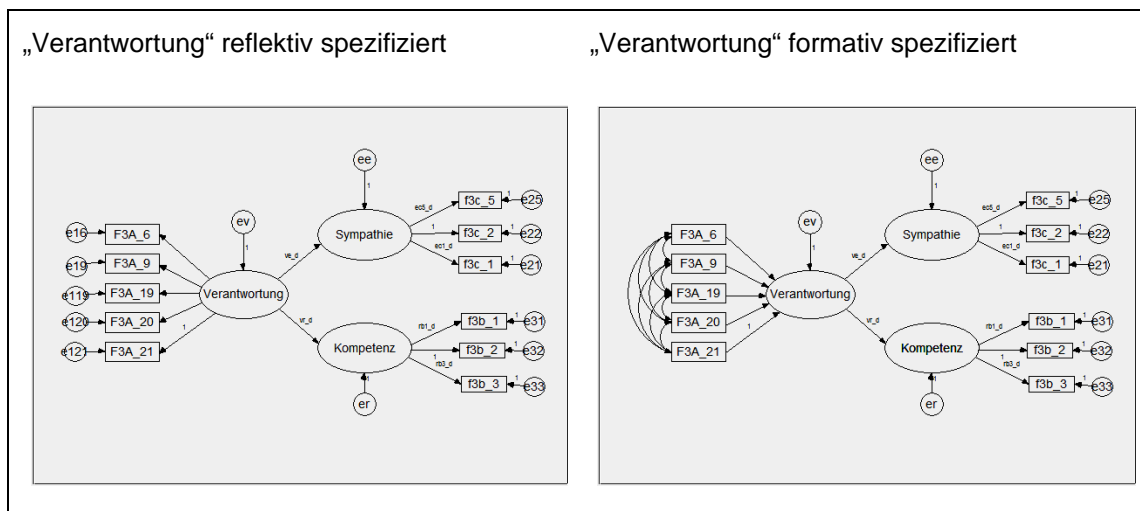


Abb. 1: Reflektive vs. formative Spezifikation des Konstruktes Übernahme gesellschaftl. Verantwortung

Bei der reflektiven Spezifikation werden die Indikatoren regelmäßig von der latenten Variablen beeinflusst, wohingegen im formativen Fall das Konstrukt von den Indikatoren determiniert wird. Abbildung 1 zeigt die Umsetzung der beiden Spezifikationsalternativen. Ob es sich um ein reflektiv oder formativ zu spezifizierendes Konstrukt handelt, ist u.a. anhand von dem beschriebenen **Entscheidungsraaster** möglich.

Dabei gilt es zu beurteilen, ob das Konstrukt kausal für die Indikatoren (vgl. Tab. 1) ist. Im vorliegenden Fall ist diese Frage nicht eindeutig zu beantworten. Allerdings zeigt sich bei den Überlegungen, ob regelmäßig alle Indikatoren die Richtung gleichzeitig ändern bzw. ob alle Indikatoren eines Konstruktes beliebig austauschbar sind, dass diese typischen reflektiven Merkmale nicht gegeben sind. Schlüssigerweise muss bspw. ein fairer Umgang mit dem Wettbewerb nicht mit dem Umweltengagement eines Unternehmens korreliert sein. Insofern ist von einer formativen Spezifikation auszugehen. Kompetenz und Sympathie sind gemäß diesem Raster jeweils reflektiv zu spezifizieren.

Code	Indikatoren für „Verantwortung“
F3A_6	Ich habe den Eindruck, dass Unternehmen XXX die Öffentlichkeit aufrichtig informiert
F3A_9	Ich habe den Eindruck, dass sich Untern. XXX gegenüber Wettbewerbern fair verhält.
F3A_19	Ich habe das Gefühl, dass Unternehmen XXX nicht nur an den Profit denkt
F3A_20	Unternehmen XXX verhält sich gegenüber der Gesellschaft verantwortungsbewusst.
F3A_21	Unternehmen XXX engagiert sich auch für den Erhalt der Umwelt.

**Tabelle 2: Indikatoren des Konstruktes Übernahme gesellschaftlicher Verantwortung**

## 5.2 Lokale Gütekriterien

### 5.2.1 Operationalisierung reflektiver Konstrukte

Die Überprüfung soll anhand des Beispiels des Konstruktes „Sympathie“ erfolgen. Zur Operationalisierung dieses endogenen Konstruktes wurden in einem ersten Schritt aus der Literatur und eigenen Vorstudie die in Tabelle 3 dargestellten fünf Indikatoren entwickelt. Die Berechnung von **Cronbachs Alpha** als erstem Prüfungsschritt der Operationalisierung ergibt jedoch zunächst einen unbefriedigenden Wert von nur 0,358, was deutlich unter dem geforderten Kriterium von 0,7 liegt (vgl. Tabelle 8).

Die fünf entwickelten Indikatoren eignen sich also relativ schlecht dazu, das Konstrukt Sympathie reliabel zu messen. Eine Verbesserung wird erreicht, wenn das Messinstrument bereinigt wird, also schrittweise die Indikatoren mit der geringsten **Item-to-Total-Korrelation** entfernt werden. In diesem Fall werden die beiden Indikatoren F3C\_3 und F3C\_4 entfernt. Durch die Entfernung zeigt sich im bereinigten Messinstrument mit drei Indikatoren eine deutliche Verbesserung der Gütemaße: Cronbachs Alpha liegt nun bei 0,772, die Item-to-Total Korrelationen sind zufrieden stellend.

Code	Indikatoren für „Sympathie“
F3C_1	Unternehmen XXX halte ich für ein sympathisches Unternehmen.
F3C_2	U. XXX ist ein Untern., mit dem ich mich mehr identifizieren kann als mit anderen Untern.
F3C_3	Unternehmen XXX hebt sich positiv von anderen Unternehmen ab.
F3C_4	Unternehmen XXX ist mir sympathischer als andere Unternehmen.
F3C_5	U. XXX ist ein U. bei dem ich es mehr als bei a. U. bedauere, wenn es nicht mehr besteht.

**Tabelle 3: Indikatoren des Konstruktes Übernahme gesellschaftlicher Verantwortung**

Code	Item-to-Total-Korrelation	Cronbachs Alpha	Kommunalitäten	Indikatorreliabilität	Faktorreliabilität	DEV
F3C_1	0,768	0,772	0,623	0,400	0,786	0,580
F3C_2	0,868		0,754	0,712		
F3C_5	0,852		0,693	0,523		

**Tabelle 4: Ergebnisse der Reliabilitätsprüfung und der konfirmatorischen Faktorenanalyse für „Sympathie“**

Um die Faktorstruktur weiter zu prüfen, wird basierend auf der bereinigten Skala eine **explorative Faktorenanalyse** durchgeführt. Mit dem Kaiser-Kriterium lässt sich eine Komponente extrahieren, welche 69,02 % der Ausgangsvarianz erklärt und damit auf eine gute Reproduktion deutet. Die Kommunalitäten liegen weit über den geforderten 0,16. Die Voraussetzungen für die konfirmatorische Faktorenanalyse sind also gegeben.

Die **konfirmatorische Betrachtung** des unterstellten einfaktoriellen Messmodells erfolgt mittels AMOS. Dabei werden die Parameter des Messmodells, also die Faktorladungen, geschätzt und es lassen sich die lokalen Anpassungsmaße errechnen (vgl. Tabelle 4). Die minimal geforderten Werte nach Tabelle 8 sind nun vollständig erfüllt. Die Bereinigung der Konstruktmessung um zwei Indikatoren hat damit zu einer Verbesserung der Operationalisierung geführt: Die drei Indikatoren stellen nun eine **reliable und konvergenzvalide** Operationalisierung von „Sympathie“ dar.

Um eine Messung jedoch als valide einzustufen, muss darüber hinaus die Diskriminanzvalidität geprüft werden, d.h. ob das operationalisierte Konstrukt Sympathie tatsächlich ein vom Konstrukt Kompetenz verschiedenes Faktorgebilde darstellt.

Die **Diskriminanzvalidität** kann mit dem  $\chi^2$ -Differenztest untersucht werden, bei dem ein unrestringiertes Modell mit einem Modell verglichen wird, bei dem

die Korrelation zwischen beiden Konstrukten auf 1 fixiert ist. Tabelle 5 zeigt, dass dieses restringierte Modell hochsignifikant schlechter ist. Das 5%-Niveau ist bereits bei einer Verschlechterung von 3,841 erfüllt. Strenger ist das **Fornell/Larcker-Kriterium**, das wie Tabelle 6 zeigt, nur knapp erfüllt wird: die DEV von „Kompetenz“ und „Sympathie“ ist jeweils größer als die quadrierte Korrelation zwischen den Konstrukten.

	Sympathie
Kompetenz	1069,169

Tabelle 5:  $\chi^2$ -Differenztest

	DEV	Sympathie
DEV		0,580
Kompetenz	0,577	0,572

Tabelle 6: Fornell/Larcker-Kriterium

### 5.2.2 Operationalisierung formativer Konstrukte

Zur Operationalisierung der formativen Konstrukte stehen keine geeigneten statistischen Methoden zur Verfügung. Neben der Validierung durch Experten bzw. den Forscher selbst können mittels des PLS-Algorithmus in Verbindung mit einer **Bootstrap-Prozedur** Standardfehler und t-Statistik für die geschätzte Ladung jedes Indikators auf die Latente berechnet werden. Exemplarisch wird dies mit PLS-Graph 3.0 dargestellt (vgl. dazu ausführlich *Chin*, 2004). Das reflektiv basierte Prozedere der Reduzierung der Skalen muss unterbleiben, da keine Indikatoren entfernt werden dürfen, die die Latente sachlogisch determinieren. PLS-Graph schätzt bei der Bootstrap-Prozedur dasselbe formative Faktormodell mehrfach und berechnet daraus die Zuverlässigkeit der im Mittel errechneten Koeffizientenschätzer. Diese Koeffizientenschätzer können auch als Validitätskoeffizienten bezeichnet werden. Tabelle 7 zeigt die Standardfehler und t-Werte (nach 300 Durchläufen, sämtlich signifikant zu 5% (vgl. auch Tabelle 7)).

Code	F3A_6	F3A_9	F3A_19	F3A_20	F3A_21
Standardfehler	0,0164	0,015	0,0148	0,0145	0,0206
t-Wert	61,6904	80,1211	75,8265	81,5862	26,0056

Tabelle 7: PLS-Bootstrap-Standardfehler des Konstrukts Übernahme gesellschaftlicher Verantwortung

### 5.3 Globale Gütemaße

Nachdem die lokalen Gütemaße überprüft und die Messung der Modellkonstrukte wo nötig angepasst wurde, wird nun die Anpassungsgüte des gesamten Erklärungsmodells an die empirischen Daten evaluiert. Im AMOS 5 – Output lassen sich die globalen Gütemaße unter „Model Fit“ ablesen. Für die meisten der Kriterien werden neben unserem Modell (hier: „Modell“) auch entsprechende Gütemaße für ein saturiertes Modell (mit allen denkbaren Beziehungen) sowie für ein Unabhängigkeitsmodell ausgewiesen.

CMIN				
Model	NPAR	CMIN	DF	P CMIN/DF
Modell	34	1684,779	32,000	52,649
Saturated model	66	,000	0	
Independence model	11	31334,532	55,000	569,719

Abb. 2: Globale Gütemaße: CMIN

RMR, GFI				
Model	RMR	GFI	AGFI	PGFI
Modell	,117	,954	,905	,463

Abb. 3: Globale Gütemaße: RMR, GFI

Im Bereich **CMIN** ( $\chi^2$ ) werden nach der Zahl der im Modell zu schätzenden Parameter (**NPAR**) die  $\chi^2$ -basierten Statistiken ausgewiesen. Der absolute  $\chi^2$ -Wert ist nur für Modellvergleiche von Relevanz (bspw. bei  $\chi^2$ -Differenzentests). **CMIN/DF**, also  $\chi^2$  geteilt durch die Zahl der Freiheitsgrade (df) bringt mit kleineren Werten bessere Anpassung zum Ausdruck. Werte kleiner als 3 werden i.d.R. als akzeptabel betrachtet. Der ausgewiesene p-Wert ist die Wahrscheinlichkeit dafür, bei der Ablehnung der folgenden Nullhypothese einen Fehler zu begehen: das theoretische Modell ist eine akzeptable Anpassung an die Realität. Da der Test jedoch auf dem umstrittenen  $\chi^2$ -Wert basiert, gilt er gemeinhin nur als eingeschränkt brauchbar. Mit steigendem Stichprobenumfang steigt der  $\chi^2$ -Wert überproportional, so dass der  $\chi^2$ -Anpassungstest so gut wie jedes Modell als vermeintlich unbrauchbar ablehnen würde. Auch in unserem Beispiel ist dies der Fall.

**RMR** steht für „root mean square residual“ und ist ein Maß für die Abweichung der empirischen Varianz-/Kovarianzgrößen von den theoretischen modellbasierten Größen. Die Größen GFI und AGFI sind auf [0;1] definiert. Unser Modell erreicht sehr gute Anpassungswerte, die über 0,9 liegen. **AGFI** ist eine korrigierte („adjusted“) Version des GFI, die eher verwendet werden sollte. Der **PGFI**-Wert korrigiert den **GFI**-Wert mit der Zahl der Freiheitsgrade und ist daher wiederum eher für Modellvergleiche geeignet.

Baseline Comparisons					
Model	NFI	RFI	IFI	TLI	CFI
	Delta1	rho1	Delta2	rho2	
Modell	,946	,908	,947	,909	,947

**Abb. 4: Globale Gütemaße:  
NFI, RFI, CFI, TLI**

Parsimony-Adjusted Measures			
Model	PRATIO	PNFI	PCFI
Modell	,582	,551	,551

**Abb. 5: Globale Gütemaße:  
PRatio, PNFI, PCFI**

Die „Baseline Comparisons“ sind Maße, die das zu testende Modell mit dem schlechteren Unabhängigkeitsmodell vergleichen. Dieses unterstellt keine Beziehung zwischen den Modellkonstrukten. Es ist daher keine Überraschung, dass dieses naturgemäß eine schlechte Anpassung hat. NFI, RFI, IFI, TLI und CFI variieren in der Regel zwischen 0 und 1. Unser Modell erreicht bei allen fünf Maßen die geforderte Mindestgrenze von 0,9 und stellt damit ein deutlich besseres Modell als das Unabhängigkeitsmodell dar, was freilich über ein „besser als nichts“ nicht hinausgeht.

PNFI und PCFI basieren auf einem Vergleich der Zahl der zu schätzenden Parameter in unserem Modell zu der Zahl im Unabhängigkeitsmodell. PRATIO selbst ist nicht als Gütemaß interpretierbar. PNFI und PCFI sind die damit korrigierten Versionen des NFI bzw. des CFI. Modelle, die wenige Parameter schätzen („sparsame“ Modelle) werden durch diese Gütemaße belohnt, es gelten jedoch dieselben Einschränkungen.

NCP			
Model	NCP	LO 90	HI 90
Modell	1652,779	1522,032	1790,893
Saturated model	,000	,000	,000
Independence model	31279,532	30700,602	31864,740

**Abb. 6: Globale Gütemaße in AMOS: NCP**

FMIN				
Model	FMIN	F0	LO 90	HI 90
Modell	,245	,241	,222	,261
Saturated model	,000	,000	,000	,000
Independence model	4,563	4,555	4,471	4,640

**Abb. 7: Globale Gütemaße in AMOS: FMIN**

Die Werte, welche unter den Überschriften NCP sowie FMIN ausgegeben werden, sind für sich genommen weniger als Gütemaß interpretierbar. NCP und Fmin gehen jedoch in die Berechnung des wichtigen Gütemaßes RMSEA ein.

RMSEA				
Model	RMSEA	LO 90	HI 90	PCLOSE
Modell	,087	,083	,090	,000

**Abb. 8: Globale Gütemaße in AMOS: RMSEA**

Der RMSEA ist eines der üblichsten Gütemaße. Streng genommen handelt es sich jedoch nur um eine Schätzung für den nicht beobachtbaren „wahren“ RMSEA des Modells. Je näher bei 0, desto besser die Anpassungsgüte des Modells. Üblicherweise werden Werte kleiner als 0,08 als akzeptabel gesehen. Unser Modell erreicht dieses wichtige Kriterium leider nicht. Auch das 90%-Konfidenzintervall für den RMSEA in unserem Modell umfasst [0,83;0,90] und schließt damit gute Fit-Werte zum 90%-Niveau aus. Pclose ist die Fehlerwahrscheinlichkeit, beim Ablehnen der Nullhypothese „der wahre RMSEA im Modell ist kleiner als 0,05“ einen Fehler zu machen. Diese Wahrscheinlichkeit ist hier extrem klein ( $<0,01$ ), was den Rest der Ergebnisse bestätigt.

AIC				
Model	AIC	BCC	BIC	CAIC
Modell	1752,779	1752,898	1985,156	2019,156

**Abb. 9: Globale Gütemaße in AMOS: AIC, BIC**

ECVI				
Model	ECVI	LO 90	HI 90	MECVI
Modell	,255	,236	,275	,255

**Abb. 10: Globale Gütemaße in AMOS: ECVI**

Weiterhin gibt AMOS auch informationstheoretische Gütemaße aus, wobei die meisten auf Akaike's Informationskriterium (AIC) basieren: BCC (Browne-Cudeck), BIC, CAIC. Auch der in AMOS eigens aufgeführte ECVI-Wert ist lediglich eine Variante des AIC, MECVI von BCC. Da es sich um relative Maße handelt, sind sie für die Beurteilung eines einzelnen Modells nicht interpretierbar.

HOELTER		
Model	HOELTER	HOELTER
	.05	.01
Modell	189	219
Independence model	17	19

**Abb. 11: Globale Gütemaße in AMOS: HOELTER**

Zum Abschluss der Gütemaße gibt AMOS Hoelter's kritisches N aus. Noch einmal wird hier auf den  $\chi^2$ -Anpassungstest abgestellt, der wie beschrieben stark von der Stichprobengröße abhängt. Das kritische N für 0,05 ist diejenige Stichprobengröße, bei der das Modell mit diesem  $\chi^2$ -Wert zum 5%-Niveau akzeptiert werden würde. In unserem Fall dürfte die Stichprobe höchstens 189 Beobachtungen umfassen, damit das Modell als signifikant ausgewiesen werden würde. Dies ist freilich deutlich niedriger als die hier eingegangenen 6.868 Fälle. *Hoelter* (1983) selbst argumentiert, dass Werte von 200 oder besser auf guten Fit hindeuten, was jedoch umstritten ist. Analoges gilt für das ebenso ausgewiesene 1%-Niveau.

## 5.4 Gesamtbeurteilung

Tabelle 8 stellt nochmals sämtliche Gütekriterien zusammen und gibt Literaturhinweise zur jeweiligen Fundierung der Entscheidung. Zusammenfassend lässt sich festhalten, dass bei der Betrachtung der Gütekriterien eines Modells zwar eine gewisse Gewichtung durchaus angebracht ist, jedoch keinesfalls zu einer Beliebigkeit bei der Auswahl der Kriterien führen darf. Der Forscher sollte sich im Zusammenspiel der Gütekriterien ein Gesamtbild der Anpassung eines Modells machen. Vollständige Dokumentation aller Kriterien ist durchaus angebracht (vgl. *Bagozzi/Yi*, 1988, S. 82). Obwohl die wichtigen Gütekriterien GFI und AGFI eher für eine gute Anpassung sprechen, müsste im Fall unseres Modells der niedrige RMSEA-Wert zu größter Vorsicht bei der Interpretation der Ergebnisse führen, was freilich dennoch möglich ist.

	Gütemaß	Schwellenwert	Quelle
Lokale Anpassungsmaße für reflektive Konstrukte	<b>Reliabilität</b>		
	Cronbachs Alpha	≥ 0,7	<i>Nunnally</i> (1978, S. 245)
	bei 2-3 Indikatoren	≥ 0,4	<i>Peter</i> (1997, S. 180)
	Item-to-Total Korrelation	≥ 0,5	<i>Bearden et al.</i> (1989, S. 475)
	Indikatorreliabilität	≥ 0,4	<i>Bagozzi/Baumgartner</i> (1994, S. 402)
	bei großen Stichproben (n>1.000)	≥ 0,1 bis 0,2	<i>Balderjahn</i> (1986, S. 117)
	<b>explorative Faktorenanalyse</b>		
	erklärter Varianzanteil	≥ 0,5	<i>Peter</i> (1997, S. 180)
	Faktorladung	≥ 0,5	<i>Backhaus et al.</i> (2003, S. 331)
	<b>konfirmatorische Faktorenanalyse</b>		
	Faktorreliabilität	≥ 0,6	<i>Bagozzi/Yi</i> (1988, S. 82)
	Durchschnittlich erfasste Varianz	≥ 0,5	<i>Bagozzi/Yi</i> (1988, S. 82)
	Signifikanztest der Faktorladungen	t ≥ 2,00 b. 1,96	<i>Backhaus et al.</i> (2003, S. 74)
	abh. von Fallzahl hier: 5%-Niveau 60 bis 1000		
<b>Diskriminanzvalidität</b>			
$\chi^2$ -Differenztest 5%-Niveau	≥ 3,841	<i>Homburg</i> (2000, S. 94)	
Fornell/Larcker-Kriterium	$DEV(\xi_i), DEV(\xi_j) > r^2(\xi_i, \xi_j)$	<i>Fornell/Larcker</i> (1981, S. 46)	
Globale Anpassungsmaße	$\chi^2/df$	≤ 2,0	<i>Byrne</i> (1989, S. 55)
		≤ 3,0	<i>Homburg/Giering.</i> (1996, S. 13)
		≤ 5,0	<i>Wheaton et al.</i> (1977, S. 84 ff.)
	RMSEA	≤ 0,08	<i>Browne/Cudeck</i> (1993, S. 144)
	GFI 363)	≥ 0,9	<i>Homburg/Baumgartner</i> (1998, S.
	AGFI	≥ 0,9	<i>Bagozzi/Yi</i> (1988, S. 82)
	NFI	≥ 0,9	<i>Bentler/Bonett</i> (1980, S. 600)
	CFI 363)	≥ 0,9	<i>Homburg/Baumgartner</i> (1998, S.
	AIC	n/a	<i>Akaike</i> (1987, S. 317)
	Hoelter's N	≥ 200	<i>Hoelter</i> (1983, S. 325)

Tabelle 8: Lokale und globale Gütekriterien und die entsprechenden Schwellenwerte

## Literatur

- Aaker, D. A.; Bagozzi, R. P.*, Unobservable Variables in Structural Equation Models with an Application in Industrial Selling, in: *Journal of Marketing Research*, Vol. 16, No. 2 (1979), S. 147-158.
- D'Agostino, R. B., A. Belanger*, A suggestion for using powerful and informative tests of normality, in: *American Statistician*, Vol. 44, No. 4 (1990), S. 316-321.
- Akaike, H.*, Factor analysis and AIC, *Psychometrika*, Vol. 52 (1987), S. 317-332.
- Anderson, J. C.; Gerbing, D. W.*, Some Methods for Respecifying Measurement Models to Obtain Unidimensional Construct Measurement, in: *Journal of Marketing Research*, Vol. 19, No. 4 (1982), S. 453-460.
- Anderson, J. C.; Gerbing, D. W.*, Structural Equation Modelling in Practice: A Review and Recommended Two-Step Approach, in: *Psychological Bulletin*, Vol. 103, No. 3 (1988), S. 411-423.
- Armstrong, J. S., T. S. Overton*, Estimating nonresponse bias in mail survey, in: *Journal of Marketing Research*, Vol. 14 (1977), S. 385-399.
- Backhaus, K., B. Erichson, W. Plinke, R. Weiber*, *Multivariate Analysemethoden - Eine anwendungsorientierte Einführung*, 10. erw. Auflage, Berlin et al. 2003.
- Bagozzi, R. P.*, *Causal Models in Marketing*, New York et al. 1980.
- Bagozzi, R. P.*, Causal Modeling: A General Method for Developing and Testing Theories in Consumer Research, in: *Advances in Consumer Research*, Vol. 8 (1981a), S. 195-202.
- Bagozzi, R. P.*, Evaluating Structural Equation Models with Unobservable Variables and Meas Error: A Comment, in: *Journal of Marketing Research*, Vol. 18, Nr. 3 (1981b), S. 375-382.
- Bagozzi, R. P.*, An Examination of the Validity of Two Models of Attitude, in: C. Fornell (Hrsg.), *A Second Generation of Multivariate Analysis*, Vol. 2 – Measurement and Evaluation, New York 1982, S. 145-184.
- Bagozzi, R. P., H. Baumgartner*, The Evaluation of Structural Equation Models and Hypothesis Testing, in R. P. Bagozzi (Hrsg.), *Principles in Marketing Research*, Cambridge, 1994, S. 386-422.

- Bagozzi, R. P., Y. Yi*, On the Evaluation of Structural Equation Models, in: Journal of the Academy of Marketing Science, Vol. 16, No. 1 (1988), S. 74-94.
- Balderjahn, I.*, Das umweltbewusste Konsumentenverhalten, Berlin 1986.
- Bankhofer, U.*, Unvollständige Daten- und Distanzmatrizen in der multivariaten Datenanalyse, Köln 1995.
- Baumgartner, H., Ch. Homburg*, Applications of structural equation modeling in marketing and consumer research - A review, in: International Journal of Research in Marketing, Vol. 13 (1996), S. 139-161.
- Bearden, W. O., R. G. Netemeyer, M. F. Mobley*, Handbook of marketing scales: Multi item measures for marketing and consumer behavior research, Newbury Park 1993.
- Bearden, W. O., R. G. Netemeyer, J. E. Teel*, Measurement of consumer susceptibility influence, in: Journal of Consumer Research, Vol. 15, March (1989), S. 473-481.
- Bentler, P. M., D. G. Bonett*, Significance Tests and Goodness of Fit in the Analysis of Covariance Structures, in: Psychological Bulletin, Vol. 88, No. 3 (1980), S. 588-606.
- Bollen, K. A.*, Multiple Indicators: Internal Consistency or No Necessary Relationship?, in: Quality and Quantity, Vol. 18, No. 4 (1984), S. 377-385.
- Bollen, K. A.*, Structural Equations with Latent Variables, New York et al. 1989.
- Bollen, K. A., R. Lennox*, Conventional Wisdom in Measurement: A Structural Equation Perspective, in: Psychological Bulletin, Vol. 110, No. 2 (1991), S. 305-314.
- Bortz, J.*, Statistik für Sozialwissenschaftler, 5. vollst. überarb. Aufl., Berlin et al. 1999.
- Bozdogan, H.*, Model Selection and Akaike's Information Criterion (AIC): The General Theory and its Analytical Extensions, in: Psychometrika, 52. Jg. (1987), S. 345-370.
- Browne, M. W., R. Cudeck*, Alternative Ways of Assessing Equation Model Fit, in: *K. Bollen, J. S. Long* (Hrsg.), Testing Structural Equation Models, Newbury Park 1993, S.136-162.
- Byrne, B. M.*, A primer of LISREL, New York 1989.

- Cassel, C., P. Hackl., A. Westlund, Robustness of Partial Least-Squares Method for Estimating Latent Variable Quality Structures, in: Journal of Applied Statistics, Vol. 26, No. 4 (1999), S. 435-446.
- Chin, W. W., Issues and Opinion on Structural Equation Modeling, in: MIS Quarterly, Vol. 22, No. 1 (1998), S. 7-16.
- Chin, W. W., P. R. Newsted, Structural Equation Modeling Analysis With Small Samples Using Partial Least Squares, in: R. H. Hoyle (Hrsg.), Statistical Strategies for Small Sample Research, Thousand Oaks, CA 1999, S. 307-341.
- Chin, W. W., *PLS-Graph 3.0 User's Guide*, Houston 2004.
- Churchill, G. A., A Paradigm for Developing Better Measures of Marketing Constructs, in: Journal of Marketing Research, Vol. 16, No. 1 (1979a), S. 64-73.
- Churchill, G. A., *Marketing Research: Methodological Foundations*, 4th ed., Fort Worth 1979b.
- Cronbach, L., Coefficient Alpha and the Internal Structure of Tests, *Psychometrika*, Vol. 16, No. 3 (1951), S. 297-334.
- Diamantopoulos, A., Export Performance Measurement: Reflective versus Formative Indicators, in: *International Marketing Review*, Vol. 16, No. 6 (1999), S. 444-457.
- Diamantopolous, A., H. M. Winklhofer, Index Construction with Formative Indicators: An alternative in Scale Development, in: *Journal of Marketing Research*, Vol. 38, No. 2 (2001), S. 269-277.
- Eberl, M., Formative und reflektive Indikatoren im Forschungsprozess: Entscheidungsregeln und die Dominanz des reflektiven Modells, in: *Schriften zur Empirischen Forschung und Quantitativen Unternehmensplanung*, Heft 19 (2004), online: <http://www.efoplan.de/pdf/ap1904.pdf> [aufgerufen: 12.08.2004].
- Edwards, J. R., R. P. Bagozzi, On the Nature and Direction of Relationships between Constructs and Measures, in: *Psychological Methods*, Vol. 5, No. 2 (2000) S. 155-174.
- Eggert, A., G. Fassot, Zur Verwendung formativer und reflektiver Indikatoren in Strukturgleichungsmodellen. Ergebnisse einer Metaanalyse und Anwendungsempfehlungen, in: 65. VHB-Pfingsttagung, Zürich 2003, online: <http://www.pfingsttagung.unizh.ch/Papers/Eggert-Fassott.pdf> [aufgerufen: 12.08.2004].

- Förster, F., W. Fritz, G. Silberer, H. Raffée*, Der LISREL- Ansatz der Kausalanalyse und seine Bedeutung für die Marketing-Forschung, in: *ZfB*, 54. Jg., Heft 4 (1984), S. 346-367.
- Fornell, C., J. Cha*, Partial Least Squares, in: *R. P. Bagozzi* (Hrsg.), *Advanced Methods of Marketing Research*, Cambridge, MA 1994, S. 52-78.
- Fornell, C., D. F. Larcker*, Evaluating Structural Equation Models with Unobservable Variables and Measurement Error, in: *Journal of Marketing Research*, Vol. 18, No. 2 (1981), S. 39-50.
- Gerbing, D., J. Anderson*, Monte Carlo evaluations of Goodness-of-fit Indices, in: *K. A. Bollen, J. S. Long* (Hrsg.), *Testing structural equation models*, Newbury Park 1993, S. 40-65.
- Hoelter, J. W.*, The analysis of covariance structures: Goodness-of-fit indices, in: *Sociological Methods and Research*, Vol. 11, No. 2 (1983), S. 325-344.
- Hofacker, Ch. F.*, Categorical Judgment Scaling with Ordinal Assumptions, in: *Multivariate Behavioral Research*, Vol. 19, No. 1 (1984), S. 91-106.
- Homburg, Ch.*, *Kundennähe von Industriegüterunternehmen: Konzeption - Erfolgsauswirkungen - Determinanten*, 3. Aufl., Wiesbaden 2000.
- Homburg, Ch., H. Baumgartner*, Beurteilung von Kausalmodellen – Bestandsaufnahme und Anwendungsempfehlungen, in: *Marketing ZfP*, Vol. 17, Heft 3 (1995), S. 162-176.
- Homburg, Ch., A. Dobratz*, Iterative Modellselektion in der Kausalanalyse, in: *Zeitschrift für betriebswirtschaftliche Forschung*, 48. Jg., Heft 3 (1991), S. 213-237.
- Homburg, Ch., A. Giering*, Konzeptualisierung und Operationalisierung komplexer Konstrukte: Ein Leitfaden für die Marketingforschung, in: *Marketing - Zeitschrift für Forschung und Praxis*, 18. Jg., Heft 1 (1996), S. 5-24.
- Homburg, Ch., Ch. Pflesser*, Strukturgleichungsmodelle mit latenten Variablen: Kausalanalyse, in: *A. Herrmann, Ch. Homburg* (Hrsg.), *Marktforschung*, Wiesbaden 1999, S. 633-659.
- Hulland, J., Y. H. Chow, S. Lam*, Use of causal models in marketing research: A review, in: *International Journal of Research in Marketing*, Vol. 13 (1996), S. 181-197.
- Jaccard, J., C. K. Wan*, *LISREL approaches to interaction effects in multiple regression*, Thousand Oaks, CA 1996.

- James, L., S. Mulaik, J. Brett*, Causal analysis: Assumptions, models and data, Beverly Hills, CA (1982).
- Jarvis, C. B., S. B. Mackenzie, P. M. Podsakoff*, A Critical Review of Construct Indicators and Measurement Model Misspecification in Marketing and Consumer Research, in: *Journal of Consumer Research*, Vol. 30, No. 3 (2003), S. 199-218.
- Jöreskog, K. G.*, Statistical Analysis of Sets of Congeneric Tests, in: *Psychometrika*, Vol. 36, No. 2 (1971), S. 109-133.
- Jöreskog, K. G.*, A General Method for Estimating a Linear Structural Equation System, in: *A. S. Goldberger, O. D. Duncan* (Hrsg.), *Structural Equation Models in Mathematical Psychology*, Vol. II, New York 1973, S. 85-112.
- Jöreskog, K. G.*, Statistical Analysis of Covariance and Correlation Matrices, in: *Psychometrika*, Vol. 43 (1978), S. 443-477.
- Jöreskog, K. G., D. Sörbom*, *Advances in Factor Analysis and Structural Equation Models*, Cambridge, MA 1979.
- Jöreskog, K. G., D. Sörbom*, Recent Developments in Structural Equation Modeling, *Journal of Marketing Research*, Vol. 19, No. 4 (1982), S. 404-416.
- Law, K. S., C.-S. Wong*, Multidimensional Constructs in Structural Equation Analysis: An Illustration using the Job Perception and Job Satisfaction Constructs, in: *Journal of Management*, Vol. 25, No. 2 (1999), S. 143-154.
- Leeflang, P. S. H., D. R. Wittink, M. Wedel, Ph. A. Naert*, *Building Models for Marketing Decisions*, Boston 2000.
- Lohmöller, J.-B.*, *Latent Variables Path Modeling with Partial Least Squares*, Heidelberg 1989.
- MacCallum, R. R., M. W. Browne*, The Use of Causal Indicators in Covariance Structure Models: Some Practical Issues, in: *Psychological Bulletin*, Vol. 114, No. 3 (1993), S. 533-541.
- Mann, H., D. Whitney*, On a test of whether one of two variables is stochastically larger than the other, in: *Annals of Mathematical Statistics*, Vol. 18 (1947), S. 50-60.
- Nunnally, J. C.*, *Psychometric Theory*, 2nd ed., New York 1978.

- Peter, J. P.*, Reliability: A Review of Psychometric Basics and Recent Marketing Practices, in: *Journal of Marketing Research*, Vol. 16, May (1979), S. 6-24.
- Peter, S. I.*, Kundenbindung als Marketingziel: Identifikation und Analyse zentraler Determinanten, Wiesbaden 1997.
- Raftery, A.*, Bayesian Model Selection in Structural Equation Models, in: *K. Bollen, J. S. Long* (Hrsg.), *Testing Structural Equation Models*, Newbury Park 1993, S.163-180.
- Raftery, A.*, Bayesian Model Selection in Social Research, in: *P. Marsden* (Hrsg.), *Sociological Methodology*, San Francisco 1995, S.111-163.
- Robinson J. P., P. R. Shaver, L. S. Wrightsman*, Measures of personality and social psychological attitudes, San Diego 1991.
- Rossiter, J. R.*, The C-OAR-SE Procedure for Scale Development in Marketing, in: *International Journal of Marketing Research*, Vol. 19, No. 4 (2002), S. 305-335.
- Schnell, R.; P. B. Hill, E. Esser*, Methoden der empirischen Sozialforschung, 6. überarb. u. erw. Auflage, München 1999.
- Schwaiger, M.*, Components and Parameters of Corporate Reputation – an Empirical Study, in: *Schmalenbach Business Review*, Vol. 56, Januar (2004), S. 46-71.
- Steenkamp, J.-B. E. M., H. Baumgartner*, On the use of structural equation models for marketing modelling, in: *International Journal of Research in Marketing*, Vol. 17 (2000), S. 195-202.
- Steiger, J. H., J. C. Lind*, Statistically-based tests for the number of common factors. Paper presented at the annual Spring Meeting of the Psychometric Society in Iowa City, May 30, 1980.
- Tenenhaus, M.*, Comparison between PLS and LISREL approaches for Structural Equation Modeling: Application to the measure of Customer Satisfaction, in: *PLS '03, 3rd international Symposium on PLS and Related Methods*, September 15-17, 2003.
- Tucker, L. R., C. Lewis*, A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, Vol. 38 (1973), S. 1-10.

## Schriften zur Empirischen Forschung und Quantitativen Unternehmensplanung

- Heft 1/1999 Rennhak, Carsten H.: *Die Wirkungsweise vergleichender Werbung unter besonderer Berücksichtigung der rechtlichen Rahmenbedingungen in Deutschland*
- Heft 2/2000 Rennhak, Carsten H. / Kapfelsberger, Sonja: *Eine empirische Studie zur Einschätzung vergleichender Werbung durch Werbeagenturen und werbetreibende Unternehmen in Deutschland*
- Heft 3/2001 Schwaiger, Manfred: *Messung der Wirkung von Sponsoringaktivitäten im Kulturbereich – Zwischenbericht über ein Projekt im Auftrag des AKS / Arbeitskreis Kultursponsoring*
- Heft 4/2001 Zinnbauer, Markus / Bakay, Zoltàn: *Preisdiskriminierung mittels internetbasierter Auktionen*
- Heft 5/2001 Meyer, Matthias, / Weingärtner, Stefan / Jahke, Thilo / Lieven, Oliver: *Web Mining und Personalisierung in Echtzeit*
- Heft 6/2002 Meyer, Matthias / Müller, Verena / Heinold, Peter: *Internes Marketing im Rahmen der Einführung von Wissensmanagement*
- Heft 7/2002 Meyer, Matthias / Brand, Florin: *Kundenbewertung mit Methoden des Data Mining*
- Heft 8/2002 Schwaiger, Manfred: *Die Wirkung des Kultursponsoring auf die Mitarbeitermotivation – 2. Zwischenbericht über ein Projekt im Auftrag des AKS / Arbeitskreis Kultursponsoring*
- Heft 9/2002 Schwaiger, Manfred: *Die Zufriedenheit mit dem Studium der Betriebswirtschaftslehre an der Ludwig-Maximilians- Universität München – eine empirische Untersuchung*
- Heft 10/2002 Eberl, Markus / Zinnbauer, Markus / Heim, Martina: *Entwicklung eines Scoring-Tools zur Messung des Umsetzungsgrades von CRM-Aktivitäten – Design des Messinstrumentes und Ergebnisse der Erstmessung am Beispiel des deutschen Automobilmarktes –*
- Heft 11/2002 Festge, Fabian / Schwaiger, Manfred: *Direktinvestitionen der deutschen Bau- und Baustoffmaschinenindustrie in China – eine Bestandsaufnahme*
- Heft 12/2002 Zinnbauer, Markus / Eberl, Markus: *Bewertung von CRM-Aktivitäten aus Kundensicht*
- Heft 13/2002 Zinnbauer, Markus / Thiem, Alexander: *e-Paper: Kundenanforderungen an das Zeitungsmedium von morgen – eine empirische Studie*
- Heft 14/2003 Bakay, Zoltàn / Zinnbauer, Markus: *Der Einfluss von E-Commerce auf den Markenwert*
- Heft 15/2003 Meyer, Matthias / Lüling, Max: *Data Mining in Forschung und Lehre in Deutschland*
- Heft 16/2003 Steiner-Kogrina, Anastasia / Schwaiger, Manfred: *Eine empirische Untersuchung der Wirkung des Kultursponsorings auf die Bindung von Bankkunden*
- Heft 17/2003 Numberger, Siegfried / Schwaiger, Manfred: *Cross Media, Print, and Internet Advertising: Impact of Medium on Recall, Brand Attitude, and Purchase Intention*

- Heft 18/2004 Unterreitmeier, Andreas / Schwinghammer, Florian: *Die Operationalisierung von Unternehmenskultur – Validierung eines Messinstruments (Arbeitstitel)*
- Heft 19/2004 Eberl, Markus: *Formative und reflektive Indikatoren im Forschungsprozess: Entscheidungsregeln und Dominanz des reflektiven Modells*
- Heft 20/2004 Eberl, Markus / Schwaiger, Manfred: *Die wahrgenommene Übernahme gesellschaftlicher Verantwortung als Determinante unternehmerischer Einstellungsziele – ein internationaler kausalanalytischer Modellvergleich*
- Heft 21/2004 Zinnbauer, Markus / Eberl, Markus: *Die Überprüfung von Spezifikation und Güte von Strukturgleichungsmodellen: Verfahren und Anwendung*

ISSN 1862-9059