

Heft 23 / 2005

EFOplan

Statistisch unterstützte Spezifikationsprüfung:
Die Performance von Tetrad-Test und SEM

Harald Binder
Markus Eberl

**Ludwig-Maximilians-Universität
München**

Institut für Unternehmensentwicklung und
Organisation
Seminar für Empirische Forschung und
Quantitative 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
<http://www.efoplan.de>

STATISTISCH UNTERSTÜTZTE
SPEZIFIKATIONSPRÜFUNG:
DIE PERFORMANCE VON TETRAD-TEST UND SEM

Harald Binder

Klinik und Poliklinik für Psychiatrie, Psychosomatik und Psychotherapie der Universität
Regensburg am Bezirksklinikum

Universitätsstr. 84

93042 Regensburg

Tel.: +49 (0) 941 / 941 – 1624

Fax: +49 (0) 941 / 941 – 1605

eMail: harald.binder@medbo.de

Web: www.bkr-regensburg.de

Markus Eberl

Department für Betriebswirtschaft der Ludwig-Maximilians-Universität München
Institut für Marktorientierte Unternehmensführung (IMM)

Kaulbachstr. 45/I

80539 München

Tel.: +49 (0) 89 / 2180 – 5638

Fax: +49 (0) 89 / 2180 – 99 – 5638

eMail: Eberl@bwl.uni-muenchen.de

Web: www.imm.lmu.de

Zusammenfassung

Bei Pfadmodellen mit latenten Variablen wird zwischen reflektiven und formativen Messmodellen unterschieden. Für Modelle mit formativen Messmodellen bietet sich Partial Least Squares als Verfahren zur Modellanpassung an. Um zu überprüfen, ob dieses Verfahren verwendet werden soll, lassen sich mehrere Tests verwenden. Im Rahmen einer Simulationsstudie werden der Tetrad-Test, das Anpassen eines faktorenanalytischen Messmodells und das Anpassen eines vollständigen Strukturgleichungsmodells verglichen. Es zeigt sich, dass die ersten beiden Verfahren in verschiedenartigen Simulationsszenarien zu ähnlichen Ergebnissen führen, dass das letzte Verfahren aber zusätzliche Informationen liefert. Diese Ergebnisse stützen eine Herangehensweise, die eine Kombination aus einem Tetrad-Test und einer Strukturgleichungsmodell-Anpassung als Test verwendet, bevor Partial Least Squares verwendet wird.

Inhaltsverzeichnis

1	Problemstellung	4
2	Statistische Verfahren im Rahmen der statistisch unterstützten Spezifikationsprüfung	7
3	Datensimulation	9
4	Ergebnisse	12
5	Diskussion	17

1 Problemstellung

Eine Vielzahl betriebswirtschaftlicher Theorien hypothetisiert Zusammenhänge zwischen nicht beobachtbaren Sachverhalten (Konstrukten). Um diese einer empirischen Überprüfung zugänglich zu machen, haben sich die Methodiken der Kausalanalyse als äußerst praktikabel erwiesen (vgl. Homburg und Baumgartner, 1995). Voraussetzung dafür ist zunächst die Operationalisierung der Konstrukte mittels eines Messmodells, das dem interessierenden Konstrukt mehrere beobachtbare Indikatoren zuordnet (vgl. Anderson und Gerbing, 1982, S. 453). Dabei lassen sich Konstrukte mittels eines formativen oder eines reflektiven Messmodells spezifizieren.

Operationalisierung latenter Konstrukte ist zumeist kein Selbstzweck, sondern sieht sich in die Überprüfung von Zusammenhängen zwischen mehreren Latenten mittels Strukturmodellen eingebettet. Wie an anderer Stelle in dieser Ausgabe dargestellt, geht oftmals die Wahl varianz- bzw. kovarianzbasierter Schätzmethoden für diese Strukturgleichungsmodelle mit der Spezifikationswahl einher. Die Anwendung der sehr gebräuchlichen Verfahren der Kovarianzstrukturanalyse (CBSEM) ist aufgrund der grundsätzlich faktorenanalytischen Annahmen im Messmodell potenziell problematisch, sobald formative Konstrukte präsent sind. Wie auch aus den Spezifikationsgleichungen in Tabelle 1 ersichtlich, impliziert ein reflektives Messmodell ebenso viele Gleichungen wie es Indikatoren aufweist, da jeder einzelne Indikator x_j eines Konstrukts als Abhängige in einer Gleichung aufgeführt wird, deren unabhängige Seite u.a. dieses eine Konstrukt ξ beschreibt. Ein formatives Messmodell impliziert dagegen nur eine einzige Bestimmungsgleichung für das Strukturgleichungssystem, da sich die Latente ξ per Definition als Linearkombination („Index“) ihrer Indikatoren bildet. Einem Strukturgleichungssystem mit formativen Indikatoren stehen also deutlich weniger Bestimmungsgleichungen zur Verfügung als einem System mit reflektiven Indikatoren. Damit wird es tendenziell wahrscheinlicher, dass die Zahl der zu schätzenden Parameter zumindest lokal im Vergleich zu den zur Schätzung verfügbaren Gleichungen zu groß wird und das Modell nicht identifizierbar ist. Umgekehrt lässt sich dies so formulieren: bei Präsenz formativer

t

Tabelle 1: Spezifikationsgleichungen reflektiver und formativer Messmodelle (vgl. Edwards und Bagozzi, 2000, S. 161; Bollen und Lennox, 1991, S. 306.) mit beobachteten Variablen x , latente Variablen ξ , Ladungsmatrix Λ_x der reflektiven Indikatoren und Regressionskoeffizienten Π_ξ der formativen Indikatoren.

reflektiv	formativ
$x = \Lambda_x \xi + \delta$	$\xi = \Pi_\xi x + \zeta$

Indikatoren sind tendenziell eher zu viele Parameter im Vergleich zu den verfügbaren Gleichungen vorhanden. Dieser Effekt lässt sich auch als Parameterinflation bezeichnen, der den Verfahren der Varianzstrukturanalyse (Partial Least Squares, PLS) grundsätzlich unbekannt ist, da ein Identifikationsproblem generell nicht besteht (zu den Verfahren im Genaueren vgl. Chin, 1998b).

Auf viele Strukturgleichungsmodelle können zumindest aufgrund der formalen Voraussetzungen (Identifizierbarkeit im CBSEM-Modell) sowohl kovarianzbasierte als auch varianzbasierte Schätzverfahren gleichermaßen angewendet werden. Dennoch wird die Zahl der Modelle, in denen CBSEM nicht anwendbar ist, realistischerweise überwiegen.

Bisherige Empfehlungen zur Bestimmung der Spezifikationsart stützen sich hauptsächlich auf die Beurteilung von Experten (vgl. Rossiter, 2002, S. 306, 2005 sowie Diamantopoulos und Winklhofer, 2001, S. 271 und Diamantopoulos, 2005) oder subjektive Entscheidungen des Forschers anhand von Entscheidungsfragen (vgl. Chin, 1998a, S. 9 oder Jarvis et al., 2003, S. 203). Dies ist freilich nicht unkritisch zu sehen: Rossiter selbst gibt zu bedenken, dass viele Konstrukte (insbesondere Einstellungen) je nach Kontext formativ als auch reflektiv spezifizierbar sind (vgl. Rossiter, 2002, S. 317 f. sowie Jarvis et al., 2003, S. 203). In vielen Bereichen lassen sich auch begründete Argumente für beide Spezifikationen finden¹. MIMIC (Multiple Indicators Multiple Causes)-Modelle als Spezi-

¹So wird zum Beispiel das Konstrukt „Finanzieller Erfolg im Export“ unter identischer Definition sowohl mit reflektiven als auch formativen Indikatoren gemessen (vgl. Diamantopoulos, 1999, S. 445 f).

alfälle von Strukturgleichungsmodellen, in welchen (im Grundfall) eine latente Variable sowohl mit reflektiven als auch mit formativen Indikatoren gemessen wird. Aufgrund dieser Gegebenheiten sind diese Fälle von Strukturmodellen typischerweise auch in einem CBSEM-Modell identifizierbar (vgl. Jöreskog und Goldberger, 1975). Dies erschwert die Einordnung eines Indikators als formativ oder reflektiv zusätzlich.

Daher scheint es sinnvoll, auch die Tatsache in die Entscheidung mit einzubeziehen, dass formative Indikatoren nicht notwendigerweise korrelieren müssen, reflektive dies jedoch erfüllen sollten. Bollen führt beispielsweise an, dass in Fällen, in denen das *Curtis and Jackson-Paradoxon*² auftritt, reflektive Indikatoren auszuschließen sind (was freilich noch alternative Beziehungen zulässt, bevor von einer formativen Interpretation der Korrelation ausgegangen werden darf) (vgl. Bollen, 1984, S. 382 f.). Wie bereits in Eberl (2004) ausführlich vorgestellt, lassen sich zwei Analyseschritte vorschlagen: der Tetrad-Test (vgl. Bollen und Ting, 2000) sowie eine vergleichende Betrachtung verschiedener Spezifikationen mittels Strukturgleichungsmodellen³.

Ziel dieser Arbeit ist es, die vorgeschlagenen Methoden des Tetrad-Tests und der Spezifikation eines Strukturmodells auf ihre statistische Geeignetheit hin zu untersuchen und dabei Rückschlüsse auf die Geeignetheit der vorgeschlagenen Vorgehensweise der statistisch unterstützten Spezifikationsprüfung zu ziehen. Hierzu wurde eine Vielzahl von Datensätzen simuliert, die realen Gegebenheiten ähneln, aber eine eindeutige Zuordnung zum Bereich formativer bzw. reflektiver Spezifikation erlauben und in der Folge die vorgeschlagenen Verfahren des Tetrad-Test bzw. der Parametrisierung eines Strukturmodells daraufhin testen, inwieweit sie zur Aufdeckung bzw. Bestätigung der vermuteten Spezifikation geeignet sind, also grundsätzlich zur Anwendung geeignet sind.

In der vorliegenden Arbeit wird daher nur kurz auf die statistischen Grundlagen der vorgeschlagene Vorgehensweise zur statistisch unterstützten Spezifikationsprüfung

²Das *Curtis and Jackson-Paradoxon* problematisiert die Tatsache, dass es im reflektiven Weltbild nicht möglich ist, einen negativen Zusammenhang zwischen zwei Größen zu beobachten, die beide positiv mit demselben Konstrukt zusammenhängen (vgl. Bollen, 1984, S. 377 sowie Curtis und Jackson, 1962, S. 195-204).

³Der Vorschlag erweitert dabei die reine CBSEM-Perspektive von Law und Wong, 1999, S. 153 f.

eingegangen (vgl. dazu ausführlich Eberl, 2004). In Kapitel 2 werden die wesentlichen Grundzüge der Verfahren der Strukturgleichungsanalyse dargestellt, soweit sie für die weitere Simulationsstudie von Belang sind. Deren Grundlagen werden in Kapitel 3 dargestellt. Kapitel 4 zeigt die Ergebnisse der Simulationsstudien, bevor im abschließenden Kapitel 5 Implikationen der gewonnenen Erkenntnisse für die statistisch unterstützte Spezifikationsprüfung diskutiert werden.

2 Statistische Verfahren im Rahmen der statistisch unterstützten Spezifikationsprüfung

Bei der Verwendung von Pfadmodellen mit latenten Variablen wird unterschieden zwischen dem strukturellen Teil und dem Messmodell (siehe z.B. Kaplan, 2000, S. 5). Der strukturelle Teil spezifiziert Verbindungen zwischen latenten Variablen über simultane Gleichungssystem. Das Messmodell stellt die Verbindung zwischen den latenten Variablen und den beobachteten Variablen her. Bei klassischen Strukturgleichungsmodellen wird das Messmodell über eine konfirmatorische Faktorenanalyse realisiert. Die Verbindung zwischen einer latenten Variable $\xi = (\xi_1, \dots, \xi_n)'$ und den zugehörigen beobachteten Variablen $x_j = (x_{1j}, \dots, x_{nj})'$, $j = 1, \dots, p$ ist dabei für Beobachtung i über

$$x_{ij} = \beta_j \xi_i + \delta_i \quad i = 1, \dots, n, j = 1, \dots, p \quad (1)$$

spezifiziert, wobei δ_i ein normalverteilter Fehlerterm ist und $\beta_j, j = 1, \dots, p$ die Parameter des Messmodells für ξ sind, die in der Darstellung von Tabelle 1 in der Matrix Λ_x gesammelt sind. Eine latente Variable, deren Verbindung zu den beobachtbaren Variablen derart spezifiziert ist, wird, wie weiter oben schon beschrieben, reflektives Konstrukt genannt bzw. (1) wird als *reflektives Messmodell* bezeichnet (Jarvis et al., 2003). Pfadmodelle, bei denen alle latenten Variablen über reflektive Messmodelle spezifiziert sind, können mit klassischen Strukturgleichungsmodell-Verfahren (SEM) angepasst werden, z.B. über eine „Maximum Likelihood“-Anpassung der Kovarianzmatrix (siehe z.B. Kaplan, 2000).

Ein wesentlicher Vorteil des „Partial Least Squares“ -Verfahrens (PLS) (Wold, 1985; Lohmöller, 1989) ist, dass neben dem reflektiven Messmodell alternativ auch das in Tabelle 1 beschriebene *formative Messmodell* der Form

$$\xi_i = \sum_{j=1}^p \beta_j x_{ij} + \zeta_i \quad i = 1, \dots, n \quad (2)$$

verwendet werden kann, wobei ζ_i wiederum ein normalverteilter Fehlerterm ist und $\beta_j, j = 0, \dots, p$ die Parameter des Messmodells sind (in Tabelle 1 in der Matrix Π_ξ gesammelt).

Um zu überprüfen, ob die Verwendung eines reflektiven Messmodells für alle latenten Variablen im angestrebten Pfadmodell adäquat ist, oder ob „Partial Least Squares“ mit (teilweise) formativen Messmodellen verwendet werden soll, sind verschiedene Vorgehensweisen denkbar: Einerseits kann ein komplettes Strukturgleichungsmodell angepasst werden und Maße der Anpassungsgüte wie die χ^2 -Statistik und der zugehörige p -Wert geben dann darüber Auskunft, ob ein solches Modell ausreichend gut auf die Daten passt. Alternativ könnte auch nur das Messmodell des Strukturgleichungsmodells selbst angepasst werden. Dies entspricht einer (konfirmatorischen) Faktorenanalyse mit einem Faktor für jede latente Variable (siehe z.B. Kaplan, 2000). Auch hier gibt die χ^2 -Statistik mit zugehörigem p -Wert Auskunft über die Anpassungsgüte.

Für die Überprüfung des Messmodell wurde allerdings auch ein spezieller Test entwickelt, ein sogenannter „vanishing tetrads“-Test oder kurz Tetrad-Test (Bollen und Ting, 2000). Er basiert darauf, dass unter Gültigkeit des reflektiven Messmodells (1) bestimmte Differenzen von Produkten der (Populations-)Kovarianzen der beobachteten Variablen gleich Null sein müssen.

Mit der folgenden Simulationsstudie soll geklärt werden, ob die drei vorgestellten Tests, Tetrad-Test, Anpassung einer Ein-Faktor-Faktorenanalyse und Anpassung eines Strukturgleichungsmodells, sensitiv auf die vorliegende Datenstruktur reagieren und wie sie sich dabei unterscheiden. Für die Überprüfung des Messmodells allein werden hierzu zuerst der Tetrad-Test und die Faktorenanalyse mit einem Faktor verglichen. Anschließend wird untersucht, ob ein Test über das Messmodell hinaus, wie ihn die Anpassung

eines vollständigen Strukturgleichungsmodells darstellt, zusätzliche Informationen liefert.

3 Datensimulation

Für die folgende Untersuchung werden simulierte Daten verwendet. Dies hat den Vorteil, dass die Datenstruktur bekannt ist und systematisch variiert werden kann, um das Verhalten der einzelnen Verfahren zu untersuchen. Um reflektive mit formativen Messmodellen kontrastieren zu können, ist es das Ziel, (beobachtete) Variablen zu erzeugen, die sowohl einen gemeinsamen Anteil als auch einen individuellen Anteil enthalten. Durch die Variation dieser beiden Anteile sollten bei einem Test die resultierenden p -Werte beeinflusst werden können. Bei einem hohen gemeinsamen Anteil sollte ein großer p -Wert darauf Hinweisen, dass ein reflektives Messmodell adäquat ist. Bei einem großen individuellen Anteil sollten kleine p -Werte zur Ablehnung eines reflektiven Messmodells führen, da hier ein formatives Modell adäquater ist.

Um aus Ergebnissen mit simulierten Daten Rückschlüsse für den konkreten Anwendungsfall ziehen zu können, ist es wünschenswert, dass die simulierte Struktur so gut wie möglich der entspricht, die in realen Daten zu finden ist. Im vorliegenden Fall gilt es vor allem, die Kovarianz- bzw. Korrelationsstruktur der beobachteten Variablen realistisch nachzubilden. Abbildung 1 zeigt die Korrelationsstruktur von neun beobachteten Variablen zu einem (inhaltlich begründet) formativem Konstrukt aus einem realen Datensatz. Es fällt auf, dass es einzelne Variablen gibt, deren Korrelation mit anderen Variablen relativ wenig schwankt, also immer eine ähnliche Größe hat (sichtbar durch gleichmäßige „Balken“). Dieses Merkmal könnte als Anhaltspunkt für die Realitätsnähe der simulierten Daten dienen.

Für die Erzeugung von n Beobachtungen mit $p = 5$ (beobachteten) Variablen $x_j = (x_{1j}, \dots, x_{nj})'$, $j = 1, \dots, p$ werden für den individuellen Anteil $m = 5$ zugrundeliegende Variablen $x_{ind,l} = (x_{ind,1l}, \dots, x_{ind,nl})'$, $l = 1, \dots, m$ mit

$$x_{ind,il} \sim N(0, 1) \quad i = 1, \dots, n$$

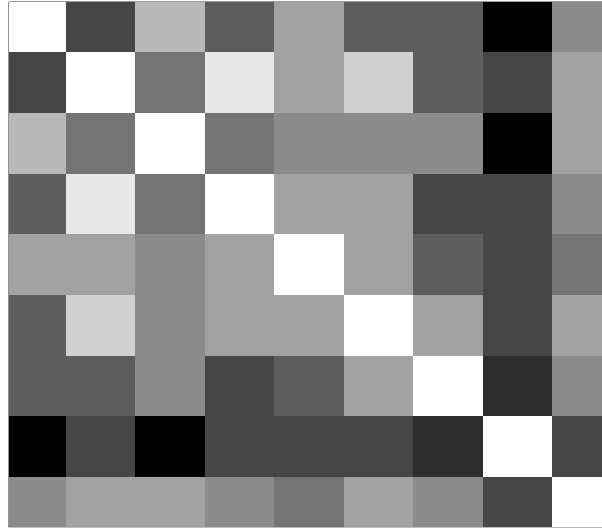


Abbildung 1: Korrelationsstruktur realer beobachteter Variablen eines formativen Konstrukts (hellere Grautöne bedeuten größere Korrelation).

aus einer Normalverteilung gezogen. Für jede zu erzeugende Variable x_j werden aus der Menge der Indices $\{1, \dots, m\}$ 5 Elemente $\{v_{j1}, \dots, v_{j5}\}$ mit Zurücklegen gezogen. Zusammen mit der gemeinsamen Komponente $x_{gem,i} \sim N(0, 1), i = 1, \dots, n$ ergibt sich x_j aus

$$x_{ij} = (1 - c_{ind}) \cdot x_{gem,i} + c_{ind} \cdot \frac{1}{5} \sum_{k=1}^5 x_{ind,iv_{jk}} + \epsilon_i \quad i = 1, \dots, n, \quad j = 1, \dots, p$$

wobei die Verteilung von $\epsilon_i \sim N(0, \sigma_{stn})$ so gewählt ist, dass sich ein festes Signal-Stör-Verhältnis c_{stn} ergibt. Die Konstante c_{ind} (mit Werten zwischen 0 und 1), auf die im Folgenden das Hauptaugenmerk gerichtet wird, steuert das Verhältnis des individuellen und des gemeinsamen Anteils der Variablen x_j . Abbildung 2 zeigt die Korrelationsstruktur beispielhaft erzeugter beobachteter Variablen. Die vorhandene Ähnlichkeit mit der in Abbildung 1 gezeigten Korrelationsstruktur realer Daten ist ein Hinweis darauf, dass die simulierten Daten von ihrer Struktur her eine Übertragung von mit ihnen gefundenen Ergebnissen auf reale Anwendungen erlauben.

Für den Vergleich des Tetrad-Test mit der Anpassung einer Ein-Faktor-Faktorenanalyse

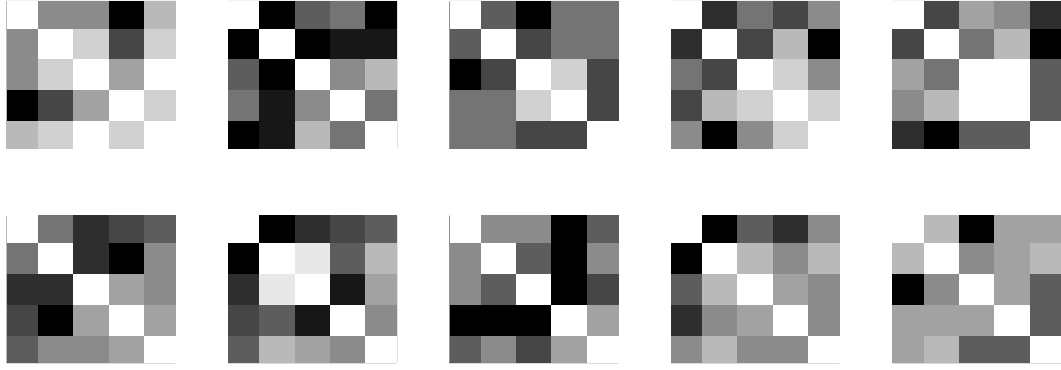


Abbildung 2: Korrelationsstruktur beispielhaft simulierter Daten (hellere Grautöne bedeuten größere Korrelation).

werden nur die wie eben beschrieben erzeugten (beobachteten) Variablen benötigt. Für die Anpassung eines Strukturgleichungsmodells wird allerdings noch mindestens eine zweite Gruppe von Variablen benötigt, die mit den x_j s zusammenhängt. Dafür werden zwei latente Variablen ξ_1 und ξ_2 mit

$$\xi_{1i} = \sum_{j=1}^p \beta_j x_{ij} + \zeta_{1i} \quad i = 1, \dots, n$$

und

$$\xi_{2i} = \xi_{1i} + \zeta_{2i} \quad i = 1, \dots, n$$

erzeugt, wobei die Fehlerterme ζ_{1i} und ζ_{2i} jeweils normalverteilt sind, mit Mittelwert 0 und einer Standardabweichung, die zu einem Signal-Stör-Verhältnis von 1 führt. Die Parameter $\beta_j, j = 1, \dots, p$ sind aus einer Normalverteilung mit Mittelwert 1 und Standardabweichung σ_β gezogen. Wie sich zeigen wird, hat σ_β wesentlichen Einfluss auf die Anpassung des Strukturgleichungsmodells. Aus ξ_2 werden die fünf reflektiven Variablen $x_{res,j} = (x_{res,1j}, \dots, x_{res,nj})', j = 1, \dots, 5$ mit

$$x_{res,ij} = \frac{1}{15} \xi_{2i} + \epsilon_{res,i} \quad i = 1, \dots, n$$

erzeugt, wobei ϵ_{res} normalverteilt ist, mit Mittelwert 0 und Standardabweichung 1.

Der Vergleich der drei Verfahren Tetrad-Test, Ein-Faktor-Faktorenanalyse und Strukturgleichungsmodell wird in der Statistikumgebung „R“ (R Development Core Team, 2005) durchgeführt. Für den Tetrad-Test wird eine Eigenimplementierung verwendet, die sich eng an die Implementierung von Ting (1995) anlehnt. Für die Faktorenanalyse wird die R-Funktion „factanal“ verwendet und für das Strukturgleichungsmodell das R-Paket „sem“. Das Strukturgleichungsmodell wird mit zwei latenten Variablen mit jeweils fünf zugehörigen Variablen und einer einseitigen Wirkung der einen latenten Variable auf die andere spezifiziert. Durchgänge, in denen die Schätzung nicht konvergiert, werden ausgeschlossen (was aber nur sehr selten der Fall ist). Die p -Werte für die Faktorenanalyse und das Strukturgleichungsmodell basieren auf der klassischen χ^2 -Statistik (siehe z.B. Kaplan, 2000).

4 Ergebnisse

Bei den folgenden Untersuchungen geht vorrangig nicht darum, zu testen, ob die untersuchten Verfahren in einem bestimmten Szenario absolut gesehen ein reflektives Mesmodell verwerfen, da es für die verwendeten Simulationsszenarien keine objektiven Kriterien gibt, wann ein reflektives und wann ein formatives Modell angebracht ist. Vielmehr soll es um den Vergleich der von den Verfahren gelieferten p -Werte untereinander gehen.

Für einen ersten Vergleich des Tetrad-Tests und der Anpassung einer Ein-Faktor-Faktorenanalyse werden Simulationsdaten mit einer Fallzahl von $n = 1000$ verwendet. Abbildung 3 zeigt die p -Werte beider Verfahren für simulierte Daten mit verschiedenen Parametern jeweils für 20 Datensätze.

Wie zu sehen ist, lassen sich die p -Werte durch die Größe des individuellen Anteils an den Variablen beeinflussen. Dies deutet auf die Eignung beider Test für die Detektion individueller Anteile der Variablen hin, die nicht auf zufälliges Rauschen zurückgehen. Das Signal-Stör-Verhältnisses wirkt moderierend in dem Sinn, dass bei einem größeren Signal-Anteil ein größerer individueller Anteil besser detektiert werden kann (rechter oberer Graph vs. linker oberer Graph). In allen Szenarien liegen die Punkte fast ge-

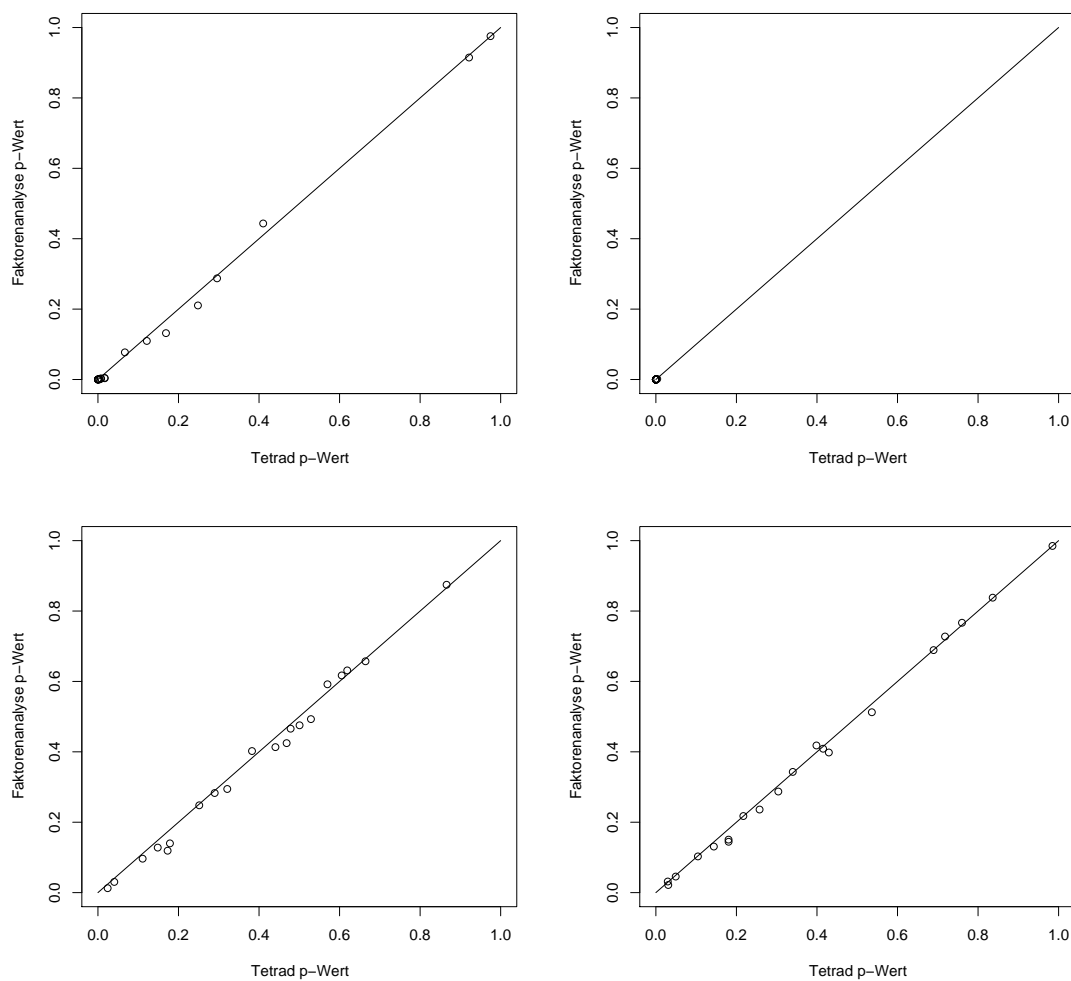


Abbildung 3: p -Werte für den Tetrad-Test und die Anpassung einer Faktorenanalyse mit einem Faktor für simulierte Datensätze mit jeweils 1000 Beobachtungen und verschiedene Datensatzparameter (großer ($c_{ind} = 0,5$) vs. kleiner ($c_{ind} = 0,1$) individueller Anteil der Variablen: obere vs. untere Graphen; kleines ($c_{stn} = 0,6$) vs. großes ($c_{stn} = 2$) Signal-Stör-Verhältnis: linke vs. rechte Graphen).

nau auf der Winkelhalbierenden, d.h. die p -Werte, die vom Tetrad-Test und von der Faktorenanalyse-Anpassung geliefert werden, sind faktisch identisch.

Um eine größere Bandbreite an Anwendungsszenarien abzudecken lohnt es sich, den

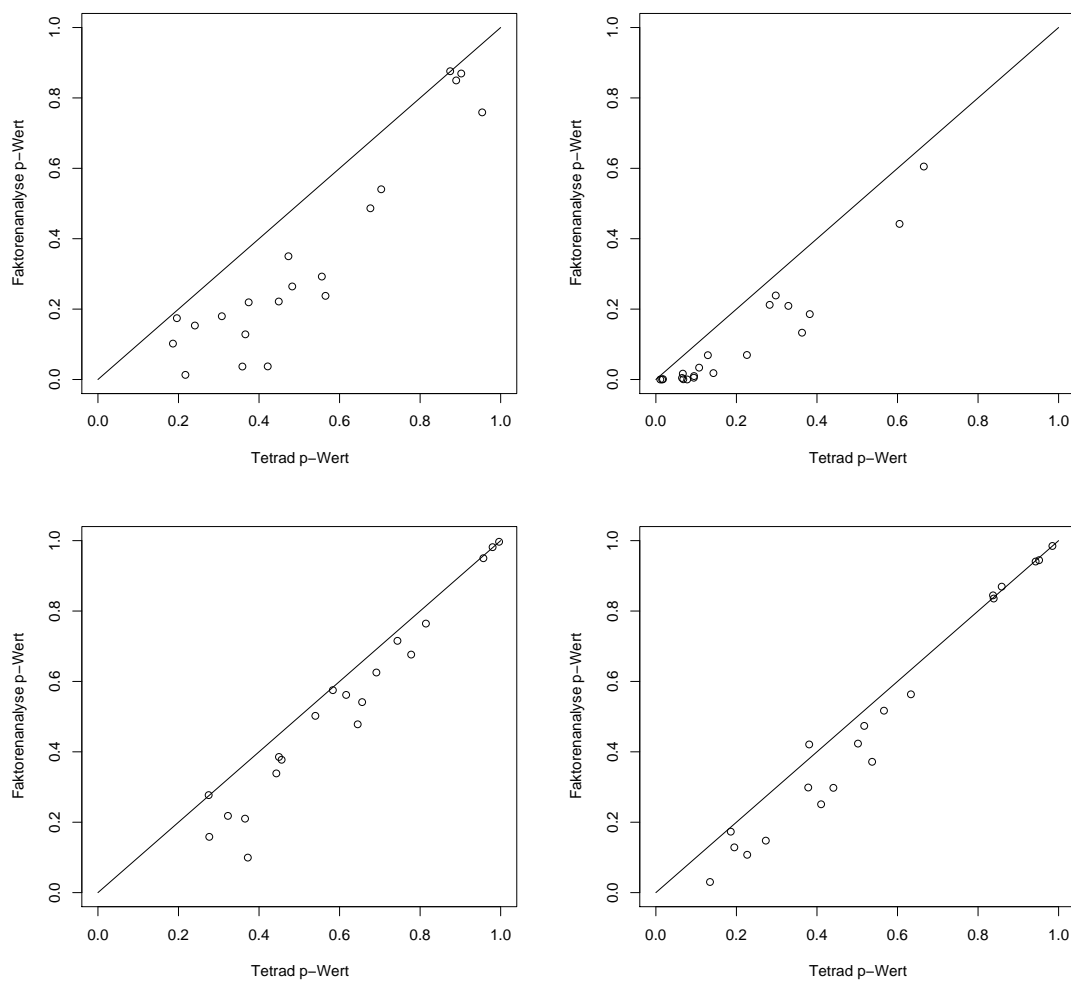


Abbildung 4: p -Werte für den Tetrad-Test und die Anpassung einer Faktorenanalyse mit einem Faktor für simulierte Datensätze mit jeweils 100 Beobachtungen und verschiedene Datensatzparameter (großer ($c_{ind} = 0,5$) vs. kleiner ($c_{ind} = 0,1$) individueller Anteil der Variablen; obere vs. untere Graphen; kleines ($c_{stn} = 0,6$) vs. großes ($c_{stn} = 2$) Signal-Stör-Verhältnis: linke vs. rechte Graphen).

Tetrad-Test und die Anpassung einer Faktorenanalyse mit einem Faktor auch für den Fall kleiner Stichproben zu vergleichen. Abbildung 4 zeigt die Ergebnisse für Datenbeispiele, bei denen alle Parameter gleich denen in Abbildung 3 sind, nur dass statt 1000

jeweils nur 100 Beobachtungen verwendet werden. Der Einfluss des individuellen Anteils der Variablen und des Signal-Stör-Verhältnisses ist ähnlich wie in den vorangegangenen Beispielen. Zu bemerken ist, dass die p -Werte für einen kleinen gemeinsamen Anteil (obere Graphen) größer sind als im Beispiel mit 1000 Beobachtungen. Dies resultiert aus der geringeren Power zur Detektion des individuellen Anteils aufgrund der kleineren Anzahl von Beobachtungen. Vergleicht man den Tetrad-Test mit der Ein-Faktor-Faktorenanalyse so zeigt sich mit nur 100 Beobachtungen ein Unterschied: Die p -Werte des Tetrad-Tests sind für fast alle Datensätze größer als die Faktorenanalyse- p -Werte. Da abgesehen von der Zahl der Beobachtungen alle Parameter der Beispiele gleich geblieben sind und die Faktorenanalyse- p -Werte nicht niedriger ausfallen als bei den korrespondierenden Beispielen mit 1000 Beobachtungen, also keine Verzerrung nach unten vorliegt, scheint es so, als hätte der Tetrad-Test weniger Power zum Verwerfen eines reflektiven Messmodells. Da die Tetrad- p -Werte aber immer noch einigermaßen proportional zu den Faktorenanalyse- p -Werten sind, scheint es auch bei einer kleinen Zahl von Beobachtungen keine strukturellen Unterschiede zwischen dem Tetrad-Test und der Anpassung einer Ein-Faktor-Faktorenanalyse zu geben. Da der Tetrad-Test direkt für den Test auf ein reflektives Messmodell entwickelt wurde, wird im weiteren Verlauf dieser als Referenz verwendet.

Der Vergleich Tetrad-Test vs. Ein-Faktor-Faktorenanalyse zeigt, dass keines der beiden Verfahren Information über die Datenstruktur liefert, die nicht schon durch das andere zur Verfügung stehen würde. Für den folgenden Vergleich von Tetrad-Test vs. Anpassung eines Strukturgleichungsmodells wird erwartet, dass letzteres Verfahren andere Information im Vergleich zum Tetrad-Test liefert. Dies ist allerdings nicht selbstverständlich, da mit der Ein-Faktor-Faktorenanalyse der wesentliche Teil des Messmodells des Strukturgleichungsmodells schon abgedeckt ist. Die zusätzliche Information muss also aus der Anpassung über das Messmodell hinaus entstehen. Die Unabhängigkeit der SEM- p -Werte von den Tetrad- p -Werten wird im Folgenden durch drei Konstellationen belegt: 1) Die p -Werte variieren unabhängig voneinander. 2) Die Tetrad- p -Werte sind sehr klein während die SEM- p -Werte variieren. 3) Die SEM- p -Werte sind sehr klein

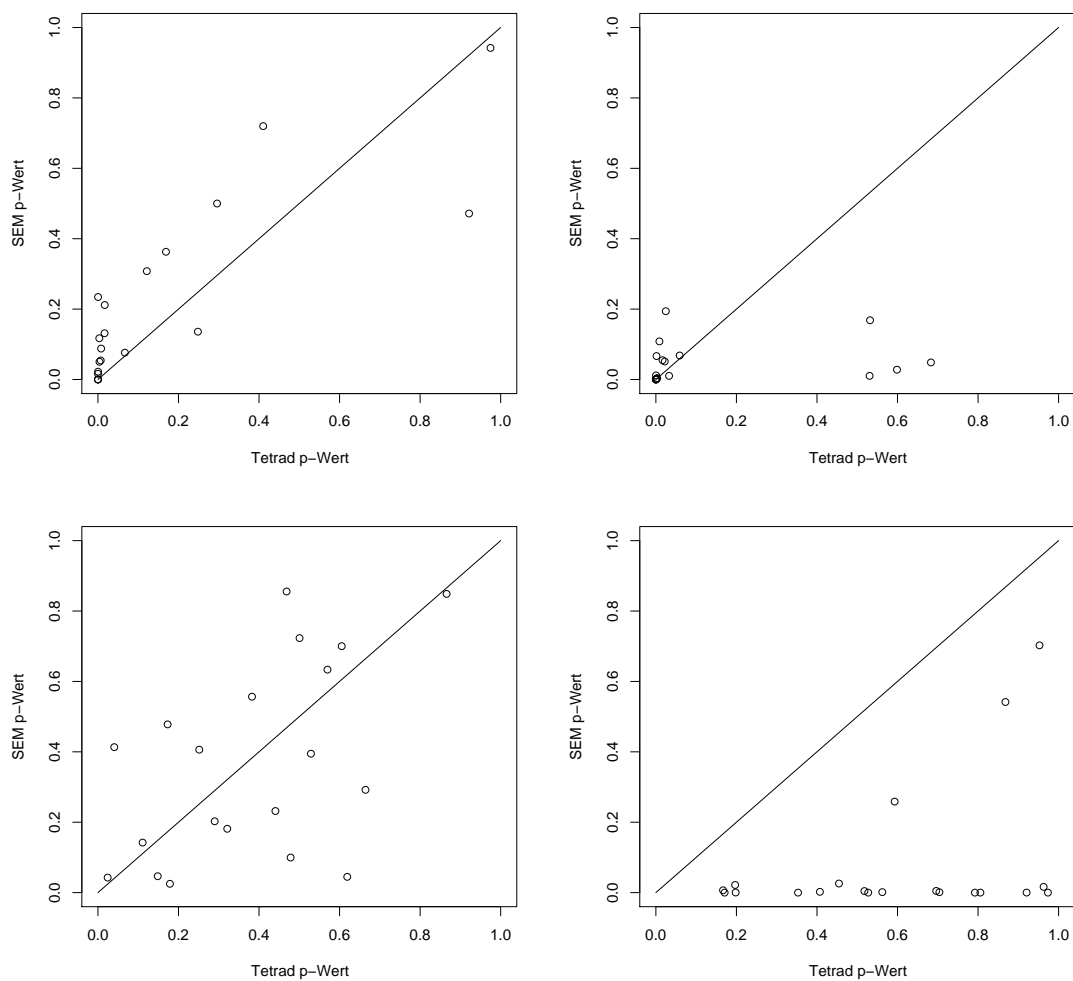


Abbildung 5: p -Werte für den Tetrad-Test und die Anpassung einer Strukturgleichungsmodells für simulierte Datensätze mit jeweils 1000 Beobachtungen und verschiedene Datensatzparameter (großer ($c_{ind} = 0,5$) vs. kleiner ($c_{ind} = 0,1$) individueller Anteil der Variablen: obere vs. untere Graphen; kleine ($\sigma_{\beta} = 0,01$) vs. große ($\sigma_{\beta} = 1$) Varianz der Koeffizienten der latenten Variable: linke vs. rechte Graphen).

während die Tetrad- p -Werte variieren.

Abbildung 5 zeigt das Verhältnis von Tetrad- p -Werten und SEM- p -Werten für Beispiele mit verschiedenen Parametern. Wie zu sehen ist, lassen sich die Tetrad- p -Werte,

wie schon in den vorangegangenen Beispielen, durch den Anteil der individuellen Komponente an den Variablen kontrollieren (obere vs. untere Graphen). Dieser Parameter scheint auf die SEM- p -Werte (im Gegensatz zu den Ein-Faktor-Faktorenanalyse- p -Werten) keinen Einfluss zu haben. Die SEM- p -Werte lassen sich stattdessen durch σ_β , die Streuung der Koeffizienten der latenten Variable x_{L1} , beeinflussen, ein Parameter, der sich naturgemäß nicht auf die Tetrad- p -Werte auswirken kann. Konstellation 1 findet sich also im linken unteren Graphen, Konstellation 2 im linken oberen und Konstellation 3 im rechten unteren Graphen.

5 Diskussion

Die dargestellten Ergebnisse belegen, dass der Tetrad-Test, die Anpassung einer Ein-Faktor-Faktorenanalyse und die Anpassung eines Strukturgleichungsmodells geeignet sind, Struktureigenschaften von Datensätzen in Bezug auf das zu verwendende Messmodell aufzudecken. Der Tetrad-Test und die Ein-Faktor-Faktorenanalyse sind dabei weitgehend identisch in Bezug auf die Information, die ihr p -Wert liefert. Der Tetrad-Test führt zwar bei kleineren Datensätzen zu weniger Ablehnungen eines reflektiven Messmodells, er ist allerdings genau für diese Fragestellung konstruiert worden. Deshalb ist abzuwägen, ob er tendenziell aus inhaltlichen Gesichtspunkten vorzuziehen ist.

Die Anpassung eines Strukturgleichungsmodells liefert in mehreren Szenarien p -Werte, die sich deutlich von denen des Tetrad-Tests unterscheiden. Da dies schon bei dem verwendeten, sehr einfachen Modell mit nur zwei latenten Variablen der Fall ist, ist zu erwarten, dass sich für komplexere Modelle noch größere Unterschiede ergeben. Deshalb ist es sinnvoll, dieses Instrument zusätzlich zu einem Tetrad-Test zu verwenden. Im Fazit lässt sich also eine kombinierte Strategie aus Tetrad-Tests und der Anpassung eines Strukturgleichungsmodells empfehlen, um zu überprüfen, ob es angebracht ist, das „Partial Least Squares“-Verfahren mit (teilweise) formativem Messmodell zu verwenden.

Literatur

- Anderson, J. C. und Gerbing, D. W. (1982). Some methods for respecifying measurements models to obtain unidimensional construct measurement. *Journal of Marketing Research*, 19(4):453–460.
- Bollen, K. A. (1984). Multiple indicators: Internal consistency or no necessary relationship? *Quality and Quantity*, 18(4):377–385.
- Bollen, K. A. und Lennox, R. (1991). Conventional wisdom in measurement: A structural equation perspective. *Psychological Bulletin*, 110(2):305–314.
- Bollen, K. A. und Ting, K.-F. (2000). A tetrad test for causal indicators. *Psychological Methods*, 5(1):3–22.
- Chin, W. W. (1998a). Issues and opinion on structural equation modeling. *MIS Quarterly*, 22(1):7–16.
- Chin, W. W. (1998b). The partial least squares approach for structural equation modeling. In Marcoulides, G. A., editor, *Modern Methods for Business Research*.
- Curtis, R. F. und Jackson, E. F. (1962). Multiple indicators in survey research. *American Journal of Sociology*, 68(2):195–204.
- Diamantopoulos, A. (1999). Viewpoint – export performance measurement: Reflective versus formative indicators. *International Marketing Review*, 16(6):444–457.
- Diamantopoulos, A. (2005). The C-OAR-SE procedure for scale development in marketing: A comment. *International Journal of Research in Marketing*, 22(1):1–9.
- Diamantopoulos, A. und Winklhofer, H. M. (2001). Index construction with formative indicators: An alternative to scale development. *Journal of Marketing Research*, 38(2):269–277.

- Eberl, M. (2004). Formative und reflektive Indikatoren im Forschungsprozess: Entscheidungsregeln und die Dominanz des reflektiven Modells. Technical report, Schriftenreihe zur empirischen Forschung und quantitativen Unternehmensplanung, No. 19/2004.
- Edwards, J. R. und Bagozzi, R. P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods*, 5(2):155–174.
- Homburg, C. und Baumgartner, H. (1995). Beurteilung von Kausalmodellen – Bestandsaufnahme und Anwendungsempfehlungen. *Marketing ZFP*, 17(3):162–176.
- Jarvis, C. B., MacKenzie, S. B., und Podsakoff, P. M. (2003). A critical review of construct indicators and measurement model misspecification in marketing and consumer research. *Journal of Consumer Research*, 30:199–218.
- Jöreskog, K. G. und Goldberger, A. S. (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association*, 70(351):631–639.
- Kaplan, D. (2000). *Structural Equation Modeling*. Sage Publications, Thousand Oaks.
- Law, K. S. und Wong, C.-S. (1999). Multidimensional constructs in structural equation analysis. *Journal of Management*, 25(2):143–154.
- Lohmöller, J.-B. (1989). *Latent Variable Path Modeling with Partial Least Squares*. Physica-Verlag, Heidelberg.
- R Development Core Team (2005). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
- Rossiter, J. R. (2002). The C-OAR-SE procedure for scale development in marketing. *International Journal of Research in Marketing*, 19(4):305–335.
- Rossiter, J. R. (2005). Reminder: A horse is a horse. *International Journal of Marketing*, 22(1):23–25.

Ting, K.-F. (1995). Confirmatory tetrad analysis in SAS. *Structural Equation Modeling*, 2:163–171.

Wold, H. (1985). Partial least squares. In Kotz, S. and Johnson, N. L., editors, *Encyclopedia of Statistical Sciences*. Wiley, New York.

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*
- 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*
- Heft 22/2005 Festge, Fabian / Meyer, Matthias / Schwaiger, Manfred: *Die Zufriedenheit mit dem Studium der Betriebswirtschaftslehre an der Technischen Universität Dresden*
- Heft 23/2005 Binder, Harald / Eberl, Markus: *Statistisch unterstützte Spezifikationsprüfung: Die Performance von Tetrad-Test und SEM*

ISSN 1862-9059