

Hildebrandt, Lutz; Temme, Dirk

Working Paper

Probleme der Validierung mit Strukturgleichungsmodellen

SFB 649 Discussion Paper, No. 2006,082

Provided in Cooperation with:

Collaborative Research Center 649: Economic Risk, Humboldt University Berlin

Suggested Citation: Hildebrandt, Lutz; Temme, Dirk (2006) : Probleme der Validierung mit Strukturgleichungsmodellen, SFB 649 Discussion Paper, No. 2006,082, Humboldt University of Berlin, Collaborative Research Center 649 - Economic Risk, Berlin

This Version is available at:

<https://hdl.handle.net/10419/25165>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

SFB 649 Discussion Paper 2006-082

Probleme der Validierung mit Strukturgleichungs- modellen

Lutz Hildebrandt*
Dirk Temme*



* Institute of Marketing, Humboldt-Universität zu Berlin, Germany

This research was supported by the Deutsche
Forschungsgemeinschaft through the SFB 649 "Economic Risk".

<http://sfb649.wiwi.hu-berlin.de>
ISSN 1860-5664

SFB 649, Humboldt-Universität zu Berlin
Spandauer Straße 1, D-10178 Berlin



SFB 649 ECONOMIC RISK BERLIN

Probleme der Validierung mit Strukturgleichungsmodellen¹

Lutz Hildebrandt, Dirk Temme

Institute of Marketing, Humboldt University Berlin
Spandauer Straße 1, 10099 Berlin
{hildebr, temme}@wiwi.hu-berlin.de

Abstract

Dieser Beitrag setzt sich mit der Leistungsfähigkeit von Strukturgleichungsmodellen bei der Validitätsprüfung von Messmodellen für hypothetische Konstrukte auseinander und geht auf ausgewählte Problembereiche bei der gängigen Anwendung dieser Methodik für die Skalenkonstruktion ein. Insbesondere werden mit der Kontrolle verschiedener Arten von Methodeneffekten Alternativen zur Elimination von Indikatoren ausschließlich auf Basis statistischer Kriterien (z. B. interne Konsistenz) aufgezeigt.

This article evaluates the performance of structural equation models in validating measurement models for hypothetical constructs and deals with specific issues following from the way this methodology is typically applied in scale construction. In particular, controlling for various types of method effects is proposed as an alternative to eliminating indicators based solely on statistical criteria (e.g., internal consistency).

Keywords: Methodeneffekte, reflektive Indikatoren, Reliabilität, Strukturgleichungsmodelle, Validität

JEL-Codes: C31, C51, C52, M31

¹ Financial support by the Deutsche Forschungsgemeinschaft (DFG) through the SFB 649 “Economic Risk” is gratefully acknowledged.

Probleme der Validierung mit Strukturgleichungsmodellen

1. Problemstellung

In der empirischen Marketingforschung hat sich durchgesetzt, Messmodelle mit multiplen Indikatoren zu konzipieren, wenn hypothetische Konstrukte erfasst werden sollen, und diese mit kausalanalytischen Methoden (Strukturgleichungsmodellen) auf Reliabilität und Validität zu prüfen. Diese Vorgehensweise ist vor allem dann sinnvoll, wenn es sich bei den Konstrukten um Phänomene zur Erklärung des individuellen Verhaltens handelt, die z. B. die Entscheidungen zwischen Alternativen – wie Kauf oder Nichtkauf einer Marke – beeinflussen. Dabei wird der Forscher mit Problemen konfrontiert, die sowohl die inhaltlich-theoretische Abbildung seines Konstrukts betreffen, als auch den empirisch-statistischen Nachweis der Güte seines Messmodells. Diese sind nicht unabhängig voneinander, und die Leistungsfähigkeit statistischer Methoden bei den üblichen Validierungsprozeduren wird von vielen Anwendern überschätzt.

Die kausalanalytische Prüfung soll aufzeigen, ob das Messmodell bei wiederholter Anwendung auch zu annähernd gleichen Ergebnissen kommt, wenn die Umwelt stabil geblieben ist (Reliabilität) und zu welchem Grad ein Messmodell ein zugrundeliegendes Konstrukt repräsentiert (Validität). Als problematisch wird von den Autoren hier die Praxis angesehen, die Konstruktion eines Messmodells mit mehreren Indikatoren über eine inhaltlich nicht begründete, sondern eher mechanisch angewandte Bereinigungsverfahren vorzunehmen, deren Ziel letztendlich darin besteht, das Modell anhand von „Daumenregeln“ für gängige Reliabilitäts- und Validitätskriterien im statistischen Sinne passend zu machen. Die inhaltlich-theoretischen Fragen werden dabei in großen Teilen außer Acht gelassen. Zurückzuführen ist dies u. U. auf die im Zusammenhang mit reflektiven Messmodellen verwendeten Metaphern „Universum von Indikatoren“ und „Domain Sampling“. Werden die Indikatoren eines Konstruktes als austauschbar angesehen, so ist es nur noch ein kleiner Schritt, diese ausschließlich nach Reliabilitäts Gesichtspunkten auszuwählen [1]. Als Beispiel für eine solche Vorgehensweise kann z. B. eine Arbeit von Ewing und Napoli (2005) gelten [2]. In einer ersten Phase der Skalenbereinigung werden dort allein anhand des Kriteriums der korrigierten Item-to-total-Korrelation beinahe 50 % der ursprünglich insgesamt 30 Indikatoren ohne Diskussion eliminiert. Welche Auswirkungen dies auf die Inhaltsvalidität der Konstruktoperationalisierung hat, lässt sich nur erahnen. Wird anschließend die

Strukturgleichungsmethodologie zur Validierung eingesetzt, so sind die Befunde durch die vorangehende Itembereinigung bereits in einem so hohen Maße determiniert, dass nur noch eingeschränkt untersucht werden kann, ob die Indikatoren tatsächlich das zu messende Konstrukt erfassen.

In diesem Beitrag sollen deshalb die spezifischen Fähigkeiten der Strukturgleichungsmethodologie zur Validitätsprüfung von Messmodellen für reflektive Indikatoren noch einmal herausgearbeitet werden. Für die Prüfung der Validität formativer Messmodelle ist diese Methodik nur eingeschränkt geeignet, es werden deshalb nur ausgewählte Aspekte der Konstruktion formativer Messmodelle angesprochen [3]. Primär soll hier verdeutlicht werden, welche Probleme entstehen, wenn der Forscher sich allein von der statistischen Methodik leiten lässt und z. B. einen der in der Literatur vorzufindenden Leitfäden verfolgt, um seine Messmodelle „zu validieren“. Wir werden zunächst noch einmal auf die Kriterien und die verwendete Methodik der Skalenkonstruktion eingehen, sowie die Stufen der Validierung eines Messmodells erörtern. Anschließend werden systematisch besondere Problemkreise der Validierung besprochen und soweit vorhanden Lösungsvorschläge dargestellt.

2. Die Praxis der Skalenkonstruktion und die kausalanalytische Validierung im Marketing

2.1 Das Schema der Skalenkonstruktion

Für den Prozess der Entwicklung valider Skalen existieren zahlreiche Systematiken, die im wesentlichen fünf Phasen umfassen [4].

Phase 1: Hier wird das theoretische Fundament durch Bestimmung der relevanten Konstrukte und ihres Geltungsbereiches gelegt. Wichtig ist die klare und eindeutige Definition der zu untersuchenden Konstrukte anhand ihrer wesentlichen Facetten. Diese umfassen neben dem Objekt der Messung und der Perspektive, aus der die Messung vorgenommen wird (z. B. Kunde, Mitarbeiter), insbesondere die zu messende Eigenschaft [5]. Festzulegen ist, was inhaltlich unter dem Attribut zu verstehen ist, ob es formativen oder reflektiven Charakter besitzt und welche Dimensionen bzw. Komponenten es umfasst. Für viele Konstrukte, wie z. B. Zufriedenheit, Einstellung oder Wertvorstellungen, liegen eingeführte Skalen oder Messmodelle vor [6], die, soweit möglich, übernommen oder modifiziert werden sollten [7].

Problematisch ist die Entwicklung von Messmodellen für als neu erachtete Phänomene und deren Abgrenzung gegenüber existierenden Konstrukten.

Phase 2: In diesem Schritt werden, legt man die klassische Auffassung des „Domain samplings“ zugrunde [8], die Konstrukte und ihre einzelnen Dimensionen inhaltlich durch eine Auswahl von Indikatoren aus einem „Item-Universum“ erfasst, die den gesamten Geltungsbereich repräsentieren sollen. Die Sicherung der Validität wird dann zu der Frage, ob die inhaltliche Abdeckung hinreichend ist und der theoretischen Konzeption des Konstrukts entspricht (Inhaltsvalidität).

Phase 3: Die generierte Ausgangsmenge an Indikatoren sollte dann von Experten sowie einer Auswahl von Personen aus der Grundgesamtheit (z. B. Konsumenten) daraufhin überprüft werden, ob sie tatsächlich dem Geltungsbereich des Konstruktes zuzuordnen sind („Face“-Validität) und ob der intendierte semantische Gehalt der Indikatoren auch von den Befragten verstanden wird. Diejenigen Indikatoren, die sich als problematisch erweisen, sollten in dieser Phase entweder modifiziert oder aus dem Pool der Indikatoren eliminiert werden.

Phase 4: Die vierte Stufe steht für einen Prozess, der in der Literatur auch mit „Item purification“ beschrieben wird und idealerweise eine Serie von Vorstudien umfasst [9]. Unter der Prüfung von statistischen Kriterien (z. B. durchschnittliche Itemkorrelation, Cronbachs α , Faktorladungen, Faktormuster, Modelfit) wird die Skala soweit optimiert, dass die Reliabilität und die Validität des konzipierten Messinstrumentes gesichert sind.

Phase 5: Eine abschließende Validierung des entwickelten Messinstrumentes erfolgt auf der Basis einer erneuten Datenerhebung. Im Prinzip kann hier auf die bereits im vierten Schritt eingesetzten Verfahren zurückgegriffen werden, wobei die quantitative Prüfung mehrerer Validitätskriterien im Vordergrund steht. Dabei sollten nochmals modifizierte Messmodelle (z. B. durch Itemelimination) kreuzvalidiert werden [10].

Die ersten drei Phasen des Schemas umfassen im Wesentlichen die inhaltliche Aufgabe, die vorab entwickelten theoretischen Vorstellungen über ein Konstrukt in eine Menge von Items zu überführen. Das Design des Messmodells ist dabei abhängig zum einen von der semantischen Konzeption eines Konstrukts in Form von Items bzw. Fragebatterien und den Möglichkeiten der statistischen Prüfung dieser Indikatoren in den verschiedenen Phasen der

Entwicklung des Messmodells. Die inhaltliche Validität wird üblicherweise von den Prozeduren in Vorstudien und der Verwendung von Expertenurteilen bestimmt. Statistische Tests, die über einfache Itemanalysen und Reliabilitätsprüfungen von Skalen hinausgehen, sind meist aufgrund der eingeschränkten Datenbasis in den Vorstudien nicht durchzuführen. Erst das Vorhandensein großer Datenmengen ist die Voraussetzung zur Anwendung komplexer statistischer Methoden (insbesondere der konfirmatorischen Faktorenanalyse), wie sie in den Phasen 4 und 5 erwartet wird.

2.2 Konzeptionelle Probleme der Skalenkonstruktion

Die im vorangehenden Abschnitt beschriebenen Phasen der Skalenkonstruktion beziehen sich (mit Ausnahme der Phase 1) auf Messmodelle, bei denen die Items (fehlerbehaftete) Repräsentanten eines dahinterliegenden Konstrukts sind (reflektive Indikatoren). Eine Veränderung dieses Konstrukts wird in diesem Fall einen kausalen Effekt auf die Ausprägung jedes Items haben. Streng genommen muss bei der Skalenkonstruktion jeder Indikator ein Stellvertreter des Konstrukts sein, so dass zwangsläufig eine relativ hohe Korrelation der Items zu fordern ist. Eine geringe Korrelation wird daher als Indiz für Messfehler bzw. eine mangelnde Reliabilität der Indikatoren angesehen.

Im Gegensatz dazu beruht die Skalenkonstruktion bei formativ konzeptualisierten Konstrukten auf der Vorstellung, dass jedes Item nur einen bestimmten Teil des Konstrukts repräsentiert und dieses erst durch die Gesamtheit aller Indikatoren definiert ist. Einzelne Indikatoren sind daher auch nicht austauschbar, ohne die Bedeutung des Konstruktes selbst substantiell zu verändern [11]. Gleichzeitig besteht damit auch keine Notwendigkeit einer Korrelation der Indikatoren. Die Indikatoren in formativen Messmodellen werden darüber hinaus i. d. R. ohne Messfehler konzipiert [12]. Die Fehlerkomponente, in der konzeptionell die nicht berücksichtigten Facetten eines formativen Konstruktes erfasst werden, ist hier vielmehr auf der Ebene des Konstrukts angesiedelt [13].

Die Entscheidung zwischen einem reflektiven und einem formativen Messmodell ist letztlich nicht so sehr vom Konstrukt selbst sondern von der zugrunde liegenden Forschungsfrage abhängig. Bei einem formativen Messmodell besteht das Interesse darin, das Zustandekommen einer bestimmten Konstruktausprägung zu erklären. Dabei ist zunächst zu identifizieren, welche Komponenten wesentlich sind. Die Konzeptualisierung kann sich hier z. B. auf den facettentheoretischen Ansatz stützen (siehe hierzu Abschnitt 3.1). Für die

Zufriedenheit mit einer Flugreise lässt sich z. B. annehmen, dass diese sich aus der Zufriedenheit mit einzelnen Teilleistungen (z. B. Ticketing, Boardservice, Pünktlichkeit) zusammensetzt. Deren Bedeutung im Hinblick auf die Gesamtzufriedenheit zu ermitteln stellt dann ein weiteres Erkenntnisziel dar. Die Zufriedenheit lässt sich aber andererseits auch durch ein reflektives Messmodell erfassen, in dem affektive und kognitive Indikatoren die Ausprägung dieses Konstruktes widerspiegeln.

Die in den Phasen 4 und 5 geforderten Prozeduren zur Prüfung von Reliabilität und Validität, die auf der Vorstellung einer durch das zugrunde liegende Konstrukt verursachten Kovariation der Indikatoren beruhen, sind bei formativen Messmodellen nicht einsetzbar; die Angemessenheit des Messmodells ist hier durch Expertenurteile oder theoretisch-deduktiv zu begründen. Darüber hinaus lässt sich die Validität u. U. auch durch Außenkriterien überprüfen [14].

Im folgenden werden wir uns auf die statistische Methodik zur Validierung von Messmodellen konzentrieren, wodurch der Fokus primär auf reflektive Messmodelle gerichtet ist.

2.3 Statistische Tests zum Nachweis von Reliabilität und Validität

Die Marketingforschung folgt im Allgemeinen bei Validierungen von reflektiven Messmodellen der Auffassung der sogenannten „True score“-Theorie [15]. Eine empirische Messung wird als Variable aufgefasst, die sich aus einem wahren Wert und einem zufälligen Fehler zusammensetzt [16]:

$$x_i = T_i + \varepsilon_i, \quad (1)$$

mit x_i als beobachtetem Wert der i -ten Messung, T_i als wahren Wert und ε_i als Zufallsfehler. Die Reliabilität als Grad, zu dem eine Messung frei von zufälligen Fehlern ist, wird über eine Zerlegung der Varianz von x_i ermittelt und resultiert aus dem Verhältnis der Fehlervarianz zur Gesamtvarianz. Unterschiedliche Designs zur Erfassung der Fehlervarianz führen zu den verschiedenen Konzeptionen von Reliabilität, wie Test-Retest-Reliabilität, Parallel-Test-Reliabilität oder interne Konsistenz [17].

Für Messmodelle in Form von Itembatterien hat besonders die Messung interner Konsistenz über Cronbachs α Bedeutung erlangt [18]. Ausgehend von der Konzeption Cronbachs (1951) sind eine Reihe von erweiterten Kriterien entwickelt worden, die auf der Basis von

Varianzzerlegungen unterschiedliche Formen von Konsistenz prüfen und über konfirmatorische Faktorenmodelle berechnet werden können [19].

Im Gegensatz zur Reliabilität wird mit dem Konzept der Validität das Ausmaß verstanden, in dem eine Skala sowohl von zufälligen als auch von systematischen Fehlern frei ist [20]. Üblicherweise setzt die Prüfung der Validität voraus, dass bereits eine feste Vorstellung darüber existiert, inwieweit ein theoretisches Konstrukt durch eine Anzahl von Indikatoren erfasst werden kann und in welcher Beziehung das Konstrukt zu anderen Konstrukten steht. Der Nachweis der Validität findet dann je nach Phase der Skalenkonstruktion oder der vorhandenen Daten über die Prüfkriterien Inhaltsvalidität, Kriteriumsvalidität (Konkurrent-, Prognose-, „Known-Groups“-Validität) oder Konstruktvalidität (Konvergenz- und Diskriminanzvalidität sowie nomologische Validität) statt [21]. Die Konzeption der Validität, die ein Untersucher bei der Prüfung zu Grunde legt, bestimmt zumeist auch die Form der statistischen Prüfung, wobei die kausalanalytische Prüfung sich am Kriterium der Konstruktvalidität orientiert.

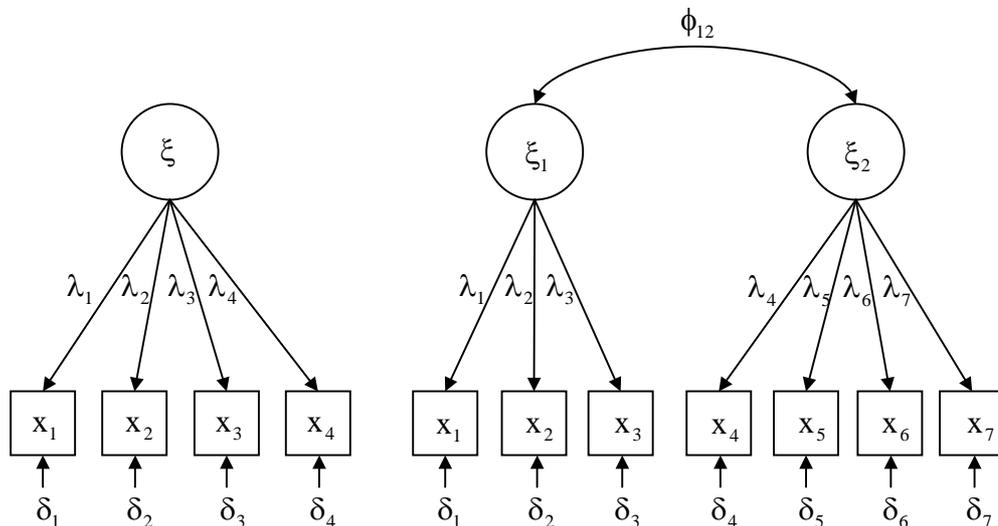
Unter Konstruktvalidität wird die Eigenschaft eines Messmodells verstanden, erstens alle Eigenschaften und deren positive oder negative Ausprägungen eines Konstrukts abzubilden, zweitens nur die Merkmale zu erfassen, die eine Bedeutung für das Konstrukt haben [22], und drittens die Relation zu anderen Konstrukten widerzuspiegeln [23]. Da Konstruktvalidität eine theoretische Konzeption für die Gültigkeit eines Messmodells ist, wird der Grad der Validität mit statistischen Tests über das Vorhandensein von Konvergenzvalidität und Diskriminanzvalidität nachgewiesen [24]. Einige Forscher fordern zusätzlich auch die Prüfung der Messkonzeption im Kontext eines größeren Modells zum Nachweis von nomologischer Validität [25].

Die Prüfung von Konvergenz- und Diskriminanzvalidität kann über das Modell der konfirmatorischen Faktorenanalyse erfolgen. Dieses Modell hat die Struktur

$$\mathbf{x} = \Lambda \boldsymbol{\xi} + \boldsymbol{\delta}, \quad (2)$$

wobei \mathbf{x} der Vektor der beobachteten Indikatoren, $\boldsymbol{\xi}$ der Vektor der Faktoren, Λ die Ladungsmatrix der Indikatoren und $\boldsymbol{\delta}$ der Vektor der Fehlerterme ist. Das Modell spiegelt die theoretisch postulierten Korrespondenzregeln für die Beziehungen zwischen den Konzepten der theoretischen Ebene und der Beobachtungsebene wider [26]. Für den Validitätsnachweis ist neben dem Modellfit die Größe und das Muster der Faktorladungen in Λ , die Größe der

Fehlervarianzen θ_{δ} und die Beziehung zwischen den latenten Variablen maßgeblich [27]. Die Struktur der Modelle zur Prüfung von Konvergenz- und Diskriminanzvalidität wird durch ihre grafische Repräsentation in Abbildung 1 deutlich, die Prüfung nomologischer Validität erfordert dagegen ein größeres Hypothesengeflecht auf der Ebene latenter Variablen.



Konvergenzvalidität

z. B. mit $\lambda_1, \lambda_2, \lambda_3, \lambda_4 \geq 0,7$
und signifikant

Diskriminanzvalidität

z. B. mit $\lambda_1, \lambda_2, \lambda_3; \lambda_4, \lambda_5, \lambda_6, \lambda_7 \geq 0,7$ und
signifikant sowie ϕ_{12} signifikant kleiner 1

Abb. 1: Die Überprüfung der Konstruktvalidität über Konvergenz- und Diskriminanzvalidität

Sowohl Reliabilität als auch Validität werden in empirischen Studien nicht vollständig zu sichern sein, sondern nur graduell. Wird die oben dargestellte Methodik verwendet, so ist die Beantwortung der inhaltlichen Frage an die Generierung einer akzeptablen Lösung mit einer komplexen statistischen Methode geknüpft. Die Schaffung einer nach statistischen Kriterien validen Faktorenlösung durch Skalenbereinigung steht deshalb im Vordergrund einer Vielzahl von Studien, die eine kausalanalytische Validierung dokumentieren. In den Leitfadens-Aufsätzen vorgeschlagene Bereinigungsverfahren über die Reliabilitätsindizes einzelner Items bzw. Validitätsindizes der Messmodelle werden so lange eingesetzt, bis ein im statistischen Sinne akzeptables Modell vorliegt. Wenn allein die Methodik ausschlaggebend für die Elimination von Indikatoren war, kann ein inhaltlich invalides, aber statistisch „valides“ Testergebnis entstehen. Bei welchen Prozeduren der Validierung dies auftreten kann und welche Lösungsmöglichkeiten es gibt, wird im folgenden aufgezeigt.

3. Kernprobleme im Prozess der Validierung von Skalen

3.1 Grundlegende Probleme

Die vorliegenden Schemata zur Entwicklung von Messkonzeptionen für theoretische Konstrukte liefern einen Rahmen, an dem sich ein Forscher orientieren kann, lösen aber nicht die Probleme, die mit einem Theoriekonstruktionsprozess verbunden sind. Das Problem der konzeptuellen Definition und der Abgrenzung eines Konstruktes verlangt eine theoretische Begründung sowie Aussagen darüber, in welcher Beziehungsstruktur ein Konstrukt zu ähnlichen oder vor- bzw. nachgelagerten Konstrukten steht [28]. Der Grad der inhaltlichen Überschneidung wird sich in der Itemstruktur widerspiegeln, aber u. U. erst beim Theorietest einer komplexen Theorie offensichtlich. Deshalb ist der Sicherung der Inhaltsvalidität die größte Aufmerksamkeit zu schenken.

Das nächste Problem entsteht bei der Nutzung statistischer Methoden, mit denen die Erfüllung von Kriterien wie Eindimensionalität, Konsistenz und Reliabilität sowie grundlegender Validitätskriterien nachgewiesen wird. Die Verfahrensvorschläge zur Validierung von Skalen gehen von einer stufenweisen Skalenbereinigung unter Verwendung von exploratorischer und confirmatorischer Faktorenanalyse aus, wobei sowohl die Wahl der Methodik als auch die Entscheidungskriterien zur Eliminierung von Items einen Einfluss auf die Struktur des resultierenden Messmodells haben. Das Ergebnis einer Skalenkonstruktion ist somit letztlich auch methodenabhängig.

Das dritte Problem tritt beim Nachweis der Konstruktvalidität mit dem Test von üblichen Konvergenz- und Diskriminanzvalidität auf. Hier wird z. B. meist der Fit der Modelle als Grad der Erreichung eines Kriteriums verwendet und zur Verbesserung des Fit die Methode der Eliminierung gering ladender Items vorgeschlagen [29]. Prüfmechanismen zur Existenz anderer Effekte, wie z. B. Methodeneffekte oder Antworttendenzen, werden vernachlässigt. Die Möglichkeit, dass Befragte ein Konstrukt anders bewerten als theoretisch erwartet, wird nicht bei der Validierung mit einbezogen.

3.2 Das Problem der Inhaltsvalidität

Am Beginn der Konstruktion eines Messmodells stellt sich für den Untersucher zunächst die Frage, welches Konstrukt er messen will, welche Merkmale (Attribute) oder Dimensionen das Konstrukt beinhaltet und wie das „Universum“ der relevanten Items aussieht, das das zugrundeliegende Konstrukt erfasst. Aus der theoretischen Begründung und über die

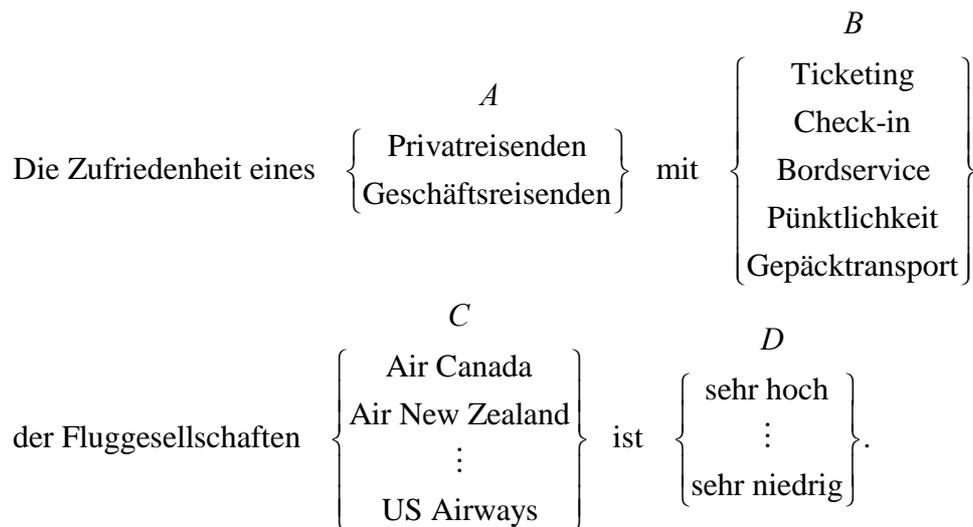
Merkmale in Form verbal spezifizierter Indikatoren sowie der Vorgabe von Antwortkategorien sollte sich dann ergeben, wie die inhaltliche Validität zu prüfen ist und wann sie vorhanden ist.

Die Frage, wie das Universum der Items aussieht, ist dabei eher eine philosophische Frage und wird üblicherweise durch Deduktion festgelegt [30]. Die Auswahl der Items kann systematisch erfolgen, z. B. um die unterschiedliche Bedeutung von Attributen zu berücksichtigen oder aber durch Bildung eines repräsentatives Samples, das die inhaltlichen Facetten, die das Domain eines Konstrukts bilden, abdeckt. Das Problem der inhaltlichen Validität ist nicht, das Konstrukt vollständig zu erfassen, sondern die Bereiche des Konstrukts in das Messmodell mit aufzunehmen, die letztendlich für die Forschungsaufgabe relevant sind. Dabei ist darauf zu achten, dass nicht bestimmte Aspekte in den Items überrepräsentiert sind oder wichtige Aspekte unberücksichtigt bleiben.

Insofern ist inhaltliche Validität immer daran geknüpft, ob die Forschungsfrage mit einem Messmodell adäquat beantwortet wird. Hier spielen theoretische Überlegungen eine größere Rolle als das Ergebnis einer statistischen Analyse. Daher finden sich auch Arbeiten, die statistische Verfahren z. T. ganz ablehnen oder ihre Bedeutung eher als gering ansehen [31] und eine systematische Prozedur zur Itementwicklung und Auswahl aufgrund von semantischen Kriterien vorschlagen, die an den Anfang einer Untersuchung zu stellen ist. Für die Konzeptualisierung kann z. B. der Facettenansatz genutzt werden [32], der gleichermaßen für reflektive und formative Konstrukte geeignet ist, da er die konzeptionelle Struktur der Abbildung eines Konstrukts festlegt, so dass inhaltliche Probleme schon hier offensichtlich werden. Formal lässt sich das Schema für die Konstruktion von Items als Abbildung („Mapping“) des kartesischen Produktes aus der Population der Subjekte P und der Stimuli S (mit den Facetten A, B, C, \dots) auf die Menge der zulässigen Antworten R darstellen:

$$P \times S \rightarrow R.$$

Aus dem formalen Modell kann über eine sogenannte „Mapping sentence“ die Itemkonstruktion für die inhaltliche Domain des Konstruktes erfolgen. Für die Messung der Zufriedenheit mit den Leistungen von Fluggesellschaften könnte eine solche „Mapping Sentence“ z. B. folgendes Aussehen haben:



Hier kann eine statistische Analyse mit Verfahren der Multidimensionalen Skalierung schon bei kleinen Stichproben zum einen dem Untersucher die Inhaltsvalidität indizieren und zum anderen den Nutzern des Messmodells Informationen liefern, zu welchem Grad die Items eines Messmodells die Facetten des Konstrukts abbilden oder methodenspezifische Varianzanteile in den Messungen enthalten sind [33].

3.2.1 Das Problem der Eindimensionalität der Skalen

Das Kriterium der Eindimensionalität einer Skala verlangt, dass die verwendeten Indikatoren zur Messung eines Konstruktes tatsächlich nur ein gemeinsames Merkmal erfassen [34]. Für die Überprüfung der Eindimensionalität sind verschiedene Verfahren vorgeschlagen worden. Das in vielen Studien verwendete Cronbachs α ist hierfür jedoch kein adäquates Maß, sondern setzt Eindimensionalität vielmehr voraus, um sinnvoll interpretiert werden zu können [35]. Als geeignete Verfahren zur Feststellung der Dimensionalität der zur Messung verwendeten Indikatoren stehen die exploratorische und die konfirmatorische Faktorenanalyse zur Verfügung, die häufig sukzessive eingesetzt werden. Von Eindimensionalität wird üblicherweise ausgegangen, wenn die Itemkorrelationen hinreichend gut durch einen einzelnen zugrundeliegenden Faktor erklärt werden können, was auch die Abwesenheit korrelierter Messfehler einschließt [36].

Die empirische Überprüfung der Dimensionalität von Indikatoren mehrerer Konstrukte mit Hilfe der exploratorischen Faktorenanalyse ist mit einer Reihe von Fallstricken für den Anwender verbunden [37]. Neben der Festlegung der zu analysierenden Indikatoren ist die Wahl (a) des Modells (Hauptkomponentenanalyse versus Faktorenanalyse), (b) der Kriterien

zur Bestimmung der Faktorenanzahl sowie (c) des Rotationsverfahrens für die Lösung von Bedeutung. Sowohl in der psychologischen Forschung [38] als auch in der Marketingforschung [39] bildet die Hauptkomponentenanalyse in Verbindung mit dem Eigenwertkriterium und der orthogonalen Varimax-Rotation das am häufigsten verwendete Design. Um die Faktorenstruktur zu identifizieren, die den Indikatoren zugrunde liegt, erweist sich aber gerade dieses Design als besonders ungeeignet [40]: Die Hauptkomponentenanalyse als reines Datenreduktionsverfahren berücksichtigt weder die Messfehlervarianz noch die spezifische Varianz der Indikatoren, was häufig zu einer Überschätzung der Faktorladungen führt; die Extraktion von Indikatoren ausschließlich anhand des Kriteriums „Eigenwert größer als 1“ ermöglicht nur in seltenen Fällen die korrekte Bestimmung der Faktorenanzahl; die Varimax-Rotation mit ihrer Annahme unkorrelierter Faktoren führt angesichts der Tatsache, dass die Faktoren in empirischen Studien häufig doch eine zumindest moderate Korrelation aufweisen, zu einer Einfachstruktur [41], die i.d.R. ein weniger klares Bild über die Zuordnung der Indikatoren zeichnet, als bei Verwendung eines obliquen Rotationsverfahrens. Stattdessen empfiehlt sich die Schätzung eines faktoranalytischen Modells mit dem ML-Schätzverfahren. Die Anzahl der Faktoren sollte dabei anhand verschiedener Kriterien [42] bestimmt werden. Bei Verletzungen der Normalverteilungsannahme kann auf robuste Schätzverfahren [43] zurückgegriffen werden, die korrigierte Standardfehler und χ^2 -Statistiken liefern. Darüber hinaus sollte für die Rotation der Faktorenlösung ein schiefwinkliges Verfahren (z. B. Promax) genutzt werden, da dieses auch bei tatsächlich unkorrelierten Faktoren korrekte Ergebnisse liefert. Resultieren aus einer derartigen Faktorenanalyse Indikatoren, die auf keinem der Faktoren eine substantielle Ladung aufweisen [44] oder auf mehrere Faktoren laden, so müssen diese u. U. eliminiert oder modifiziert werden. Dabei müsste auch berücksichtigt werden, wie sich die Korrelationen zwischen den Faktoren durch die Elimination dieser Items verändern [45].

Bei hinreichend konkreten Vorstellungen über die Dimensionalität der Konstrukte kann die konfirmatorische Faktorenanalyse verwendet werden. Die Eindimensionalität gilt dann als gegeben, wenn die globalen Gütekriterien nicht zu einer Ablehnung des Modells führen und keine signifikanten Kovarianzen zwischen den Residuen der Indikatoren vorliegen. Bei korrelierten Residuen muss davon ausgegangen werden, dass zumindest ein Teil der Indikatorkorrelation auf weitere Faktoren oder Methodeneffekte zurückzuführen ist.

Führt die konfirmatorische Faktorenanalyse zu einer Ablehnung des Modells, stellt sich die Frage nach der weiteren Vorgehensweise. Häufig werden in diesem Fall, orientiert am Kriterium der internen Konsistenz, Indikatoren eliminiert. Eine solche Vorgehensweise greift aber u. U. zu kurz, da sie auch hier zu Lasten der Inhaltsvalidität gehen kann [46]. Wird die repräsentative Abdeckung der inhaltlichen Domäne des Konstruktes nicht mehr erreicht, sollte das Modell neu konzipiert werden.

3.2.2 Prüfung der Reliabilität der Skalen

Bei der Prüfung der Reliabilität im Sinne des „True-score“-Modells hat sich als wichtigster Ansatz die interne Konsistenz, insbesondere unter Verwendung von Cronbachs α [47], etabliert [48]. Cronbachs α repräsentiert den Anteil der Gesamtvarianz einer Skala, der auf einen gemeinsamen Faktor zurückgeführt werden kann, und ist auf einem Intervall von 0 bis 1 definiert. Es setzt Eindimensionalität sowie die Eigenschaft der schwachen Tau-Äquivalenz [49] der Messungen voraus. Faktorenanalytisch entspricht dies der Annahme gleicher Faktorladungen für die Indikatoren eines Konstruktes, was selten erfüllt ist.

Cronbachs α ist in vielfältiger Weise kritisiert worden [50], so dass bei Schätzung eines konfirmatorischen Faktormodells auf geeignetere Maße übergegangen wird. In der Literatur werden hierzu mit der Indikatorreliabilität (ρ_{ii}), der Faktorreliabilität (ρ_c bzw. ω) [51] und der durchschnittlich erfassten Varianz (ρ_{DEV}) [52] drei Kriterien vorgeschlagen, die unterschiedlich strenge Anforderungen an das Messmodell stellen [53]:

$$\rho_{ii} = \frac{\lambda_{ij}^2 \phi_{jj}}{\lambda_{ij}^2 \phi_{jj} + \theta_{ii}}, \quad \rho_c = \frac{\left(\sum_{i=1}^q \lambda_{ij} \right)^2 \phi_{jj}}{\left(\sum_{i=1}^q \lambda_{ij} \right)^2 \phi_{jj} + \sum_{i=1}^q \theta_{ii}}, \quad \rho_{DEV} = \frac{\sum_{i=1}^q \lambda_{ij}^2 \phi_{jj}}{\sum_{i=1}^q \lambda_{ij}^2 \phi_{jj} + \sum_{i=1}^q \theta_{ii}},$$

mit λ_{ij} als Faktorladung, ϕ_{jj} als Varianz der latenten Variable und θ_{ii} als Varianz der Störgröße. Alle drei Reliabilitätskriterien sind auf dem Intervall 0 bis 1 definiert. Während die Indikatorreliabilität eine Schätzung für die Reliabilität eines einzelnen Indikators liefert, stellen die Faktorreliabilität und die durchschnittlich erfasste Varianz analog zu Cronbachs α Maße für die Reliabilität der Gesamtsumme über alle Indikatoren eines Faktors dar. Für die Reliabilitätsmaße werden als Zielgrößen bei der Bereinigung von Messmodellen mit Hilfe der konfirmatorischen Faktorenanalyse bestimmte Schwellenwerte angegeben, die in Tabelle 1 zusammengefasst sind. Allerdings sind diese Anforderungskriterien nur als grobe Orientierung zu verstehen, die eher generelle Anhaltspunkte liefern [54].

Reliabilitätsmaß	Anforderung	Quelle
Indikatorreliabilität ρ_{ii}	$\geq 0,5$	Bagozzi/Yi 1988
	$\leq 0,9$	Netemeyer et al. 2003, S. 153
Faktorreliabilität ρ_c	$\geq 0,6$	Bagozzi/Yi 1988
Durchschnittlich erfasste Varianz ρ_{DEV}	$\geq 0,5$	Fornell/Larcker 1981

Zu beachten ist, dass für $\rho_{DEV} = 0,5$ durchschnittliche (standardisierte) Faktorladungen von ungefähr 0,7 (dies entspricht $\rho_{ii} = 0,49$) erforderlich sind [55]. Bei einer in der Literatur häufig auch noch als akzeptabel angesehenen Indikatorreliabilität von 0,4 [56] müssen bei gleichzeitiger Aufrechterhaltung des oben angegebenen Schwellenwertes für die DEV einige Indikatoren u. U. eine deutlich höhere Faktorladung als 0,7 aufweisen.

Tab. 1: Gütemaße für die Beurteilung der Reliabilität von Messmodellen mit der konfirmatorischen Faktorenanalyse

Welche Reliabilitätswerte im Einzelfall akzeptiert werden können, hängt von einer Reihe von Einflussgrößen (z. B. inhaltliche Bandbreite eines Konstruktes, Anzahl der Indikatoren, Redundanz der Itemformulierungen; Stichprobe) ab, die darüber hinaus auch noch in einer komplexen Weise interagieren. Dieser Aspekt soll kurz skizziert werden. Mit steigender Anzahl an Indikatoren steigt z. B., ceteris paribus, Cronbachs α und damit auch die Faktorreliabilität [57]. Daher würde man bei längeren Skalen tendenziell höhere Anforderungen an diese beiden Kriterien stellen. Gleichzeitig ist aber davon auszugehen, dass die Anzahl der Indikatoren von der Bandbreite eines Konstruktes bzw. der einzelnen Dimensionen abhängig ist. Ist ein Konstrukt (bzw. eine Dimension) aber eher breit definiert, so dürfte i.d.R. auch die durchschnittliche Korrelation zwischen den Indikatoren geringer sein, was wiederum einen negativen Effekt auf die verwendeten Reliabilitätsmaße hätte. Empirisch zeigt sich daher auch ab einer Indikatorenanzahl von drei nur noch ein geringer systematischer Zusammenhang zwischen der Reliabilität und der Skalenlänge [58].

Welche Konsequenzen eine Skalenbereinigung allein auf Basis der internen Konsistenz (z. B. Indikatorreliabilität) haben kann, lässt sich an einem Beispiel nach Little et al. (1999) verdeutlichen. Die Beziehungen der Indikatoren zu ihren Faktoren sowie der Indikatoren untereinander werden dabei in einem mehrdimensionalen Raum durch Vektoren dargestellt [59]. Hier wird von einem dreidimensionalen Raum und zwei orthogonalen, also unkorrelierten Faktoren ausgegangen (vgl. Abbildung 2). Jeder der beiden Faktoren ξ_1 und ξ_2 wird durch jeweils sechs Indikatoren (x_1 - x_6 bzw. x_7 - x_{12}) gemessen.

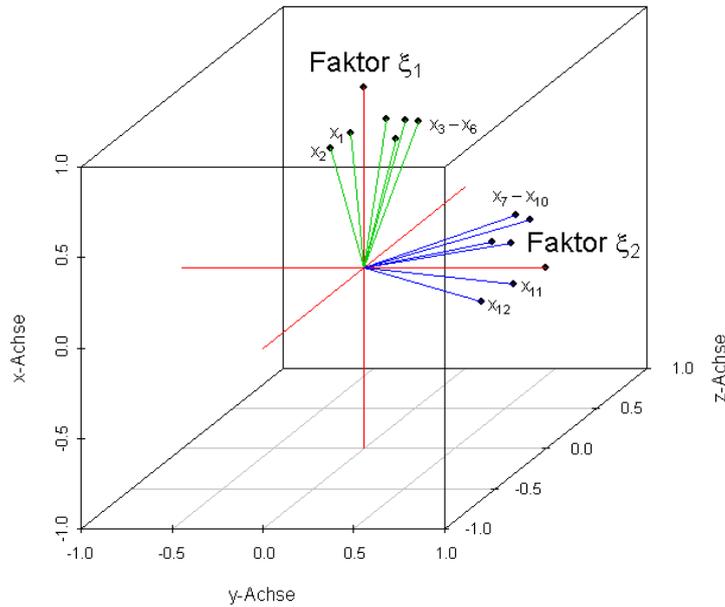


Abb. 2: Geometrische Darstellung des Zwei-Faktoren-Modells im standardisierten dreidimensionalen Raum

Bei einer standardisierten Metrik lässt sich der Raum als eine Einheitskugel auffassen, d. h. der Abstand vom Mittelpunkt der Kugel zu ihrer Oberfläche beträgt 1. Die Vektoren der Faktoren weisen in diesem Fall eine Länge von 1 auf, während die Indikator-Vektoren entsprechend ihrer Reliabilität eine Länge ≤ 1 besitzen (die Reliabilität entspricht dabei dem Quadrat der Vektorlänge). Entsprechend der „Domain sampling“-Perspektive stellen die Indikatoren eine Auswahl aller möglichen Indikatoren dar, durch die die beiden Konstrukte gemessen werden könnten. Diese gesamte inhaltliche Domäne eines Konstruktes kann man sich als einen Kegel um den Faktorvektor vorstellen, der mit seiner Spitze im Zentrum der Kugel steht. Ausgehend von einer bestimmten Konstellation der Vektoren im Vektorraum können die sich hieraus ergebenden Korrelationen zwischen den Indikatoren berechnet werden. Die Korrelation $r_{ii'}$ für zwei Indikatoren x_i und $x_{i'}$ ergibt sich dabei aus

$$r_{ii'} = l_i l_{i'} \cos(\alpha_{ii'}), \quad (3)$$

wobei l_i und $l_{i'}$ die Länge der Vektoren i bzw. i' sowie $\alpha_{ii'}$ den Winkel zwischen den beiden Vektoren angeben.

Die Konstellation der Indikator-Vektoren im vorliegenden Beispiel spiegelt eine Situation wider, in der jeweils vier Indikatoren relativ nahe beieinander liegen (Indikatoren x_3-x_6 bzw. x_7-x_{10}), die übrigen zwei Indikatoren (x_1, x_2 bzw. x_{11}, x_{12}) in Relation dazu aber sowohl

untereinander als auch von den zuvor genannten Indikatoren weiter entfernt sind (vgl. Abbildung 2). Die als vergleichsweise hoch angenommene Korrelation zwischen jeweils vier Indikatoren der beiden Konstrukte [60] kann z. B. auf einen sich stark überschneidenden semantischen Gehalt der Itemformulierungen, Methodeneffekte [61] oder u. U. auf die Existenz eigenständiger Subdimensionen zurückzuführen sein. Eine starke semantische Übereinstimmung weisen z. B. die in einer Arbeit von Stock (2004) verwendeten Indikatoren zur Messung des Konstrukts „Präsenz von Normen“ innerhalb eines Teams von Organisationsmitgliedern auf 62. Erfasst werden soll dabei die Bewertung der Teammitglieder, „inwieweit **verschiedene Aspekte** im Hinblick auf Normen in ihrem Team ausgeprägt sind“ [63]:

In unserem Team ...

- x_1 – ist klar festgelegt, welches Verhalten akzeptiert wird und welches nicht.
- x_2 – ist klar festgelegt, welches Verhalten von den einzelnen Teammitgliedern erwartet wird.
- x_3 – gibt es klare Vorstellungen bezüglich der gewünschten Verhaltensweisen der einzelnen Mitglieder.
- x_4 – ist den einzelnen klar, welche Verhaltensweisen von ihnen erwartet werden.
- x_5 – bestehen Erwartungen an die einzelnen Teammitglieder in Bezug auf gewünschte Verhaltensweisen.

Im Prinzip wird hier die Frage, ob es im Team klare Erwartungen an die Verhaltensweisen der einzelnen Mitglieder gibt, nur in unterschiedliche Satzkonstruktionen gekleidet (eine Ausnahme stellt bestenfalls das erste Item dar, da die Akzeptanz/Nichtakzeptanz von Verhalten über reine Erwartungen hinausgeht). Werden diese Indikatoren zusammen mit weiteren Statements analysiert, die tatsächlich verschiedene Aspekte von Normen abbilden (z. B. implizite Verhaltensregeln bei Konflikten oder der Entscheidungsfindung innerhalb der Gruppe), so wird sich mit hoher Wahrscheinlichkeit eine deutlich geringere Korrelation mit den oben angegebenen Indikatoren ergeben.

Auf Basis der aus der Vektorkonfiguration in Abbildung 2 abgeleiteten Korrelationsmatrix für die Indikatoren (siehe Tabelle A1 im Anhang) wird nun nach einem unter Reliabilitäts Gesichtspunkten optimierten Messmodell gesucht. Die exploratorische Phase (mit

Cronbachs α , Item-to-Total-Korrelation etc.) wird dabei übersprungen und gleich eine konfirmatorische Faktorenanalyse verwendet [64].

Die ermittelte Korrelationsmatrix wird als Stichprobenmatrix analysiert, wobei eine gängige Stichprobengröße von $N = 250$ [65] unterstellt wird [66]. Für alle nachfolgenden Modelle wird aus Gründen der Identifikation die Varianz der latenten Variablen auf 1 gesetzt. In einem ersten Schritt wird die Faktorstruktur jedes Faktors separat auf Eindimensionalität und interne Konsistenz hin überprüft. Für das Faktormodell der latenten Variable ξ_1 ergibt sich ein exzellenter Fit (vgl. Tabelle 2). Orientiert man sich nun strikt an dem Kriterium, dass die Indikatorreliabilität $\geq 0,5$ betragen soll (vgl. Tabelle 1), so sind die Indikatoren x_1 und x_2 aufgrund ihrer zu geringen Reliabilität von der weiteren Analyse auszuschließen. Für den Faktor ξ_2 zeigt die Schätzung ebenfalls einen guten Fit (vgl. Tabelle 2); hier würden aufgrund unzureichender Reliabilität die Indikatoren x_{11} und x_{12} eliminiert. Selbst wenn aufgrund der exzellenten Modellanpassung die jeweils zwei Indikatoren geringerer Reliabilität auf dieser Stufe noch nicht aus dem Itempool entfernt würden, so wäre dies spätestens dann der Fall, wenn ein gemeinsames konfirmatorisches Faktormodell mit allen Indikatoren (vgl. Abbildung 3) geschätzt wird.

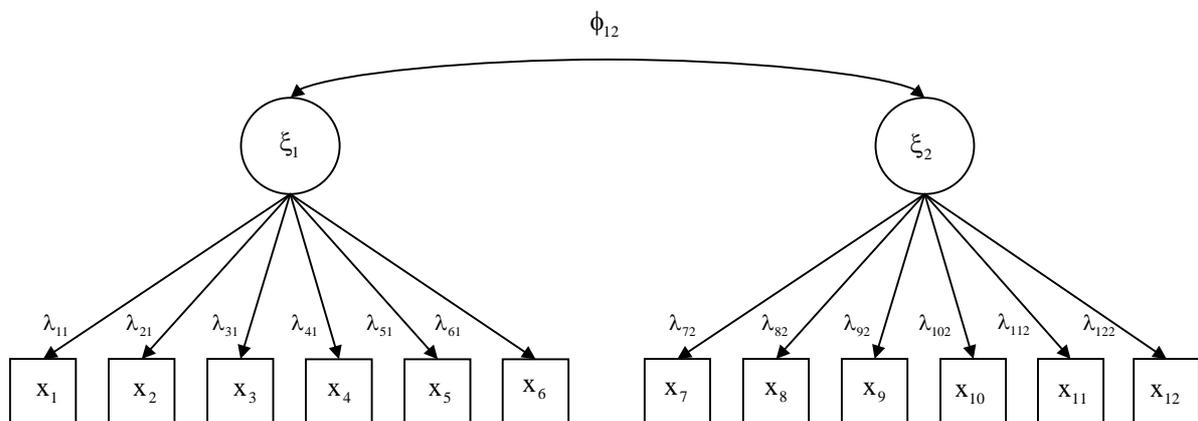


Abb. 3: Konfirmatorisches Zwei-Faktorenmodell mit allen Indikatoren

Beurteilungskriterien	Ein-Faktormodell ξ_1	Ein-Faktormodell ξ_2	Zwei-Faktorenmodell mit allen Indikat.	Zwei-Faktorenmodell nach Elimination
<i>Fit-Maße</i>				
χ^2	5,88	10,94	169,24	21,01
df	9	9	53	19
CFI	1,000	0,998	0,938	0,998
RMSEA	0,000	0,029	0,094	0,021
<i>Indikatorreliabilität ρ_{ii}</i>				
x_1	0,44^a	–	0,42	–
x_2	0,47	–	0,45	–
x_3	0,69	–	0,71	0,73
x_4	0,60	–	0,61	0,61
x_5	0,63	–	0,63	0,62
x_6	0,70	–	0,70	0,67
x_7	–	0,77	0,77	0,76
x_8	–	0,63	0,64	0,63
x_9	–	0,66	0,66	0,64
x_{10}	–	0,78	0,80	0,82
x_{11}	–	0,49	0,47	–
x_{12}	–	0,46	0,45	–

^aDie fett gesetzten Werte kennzeichnen Indikatoren mit zu geringer Indikatorreliabilität

Tabelle 2: Fit-Maße und Indikatorreliabilitäten für die geschätzten Modelle

Die Gütekriterien für die Schätzung dieses Modells (vgl. Tabelle 2) legen eindeutig eine Ablehnung nahe. Schätzt man dagegen ein Faktorenmodell ohne die Items x_1 und x_2 sowie x_{11} und x_{12} , so resultiert hieraus eine sehr gute Anpassung an die Korrelationsmatrix (vgl. Tabelle 2). Nach den gängigen Gütekriterien würde das Modell somit als empirisch bestens bewährt angesehen werden. Auch die übrigen Detailkriterien für Reliabilität sowie Konvergenz- und Diskriminanzvalidität werden ausnahmslos erfüllt (vgl. Tabellen 2 und 3). Damit kann die geschätzte Korrelation zwischen den beiden Konstrukten interpretiert werden – mit einem Wert von $r_{\xi_1\xi_2} = 0,506$ ist diese als hoch einzuschätzen 67, und dies obwohl die zugrundeliegenden Faktoren in Wirklichkeit unkorreliert sind. Trotz eines „optimierten“ Faktormodells repräsentieren die verbliebenen Indikatoren ihre jeweiligen Faktoren nur noch

schlecht, was dann letztlich auch zu einer stark verzerrten Schätzung der Beziehung zwischen den Faktoren führt.

Faktor	Faktor-reliabilität ρ_c	Durchschnittlich erfasste Varianz ρ_{DEV}	Quadrierte Faktor-korrelation $r_{\xi_1\xi_2}^2$
ξ_1	0,88	0,66	0,26
ξ_2	0,91	0,71	

Tabelle 3: Aggregierte Reliabilitäts- und Validitätsmaße für das optimierte Messmodell

Das Beispiel verdeutlicht, wie wichtig eine ausgewogene Repräsentation eines Konstruktes durch die anfänglich vorhandenen Indikatoren ist. Wird an dieser Stelle das Kriterium der Inhaltsvalidität verletzt, so kann mit Hilfe der gängigen Skalenbereinigungsprozeduren zwar ein nach Reliabilitäts Gesichtspunkten optimiertes Messmodell generiert werden, die bei der Festlegung des Itempools gemachten Fehler lassen sich aber hierdurch nicht korrigieren bzw. werden sogar noch verstärkt. Selbstverständlich bedeutet dies nicht, dass jede Skalenbereinigung mit der Strukturgleichungsmethodologie zwangsläufig ein Artefakt produziert. Indikatoren mit zu geringer Reliabilität können entfernt werden, solange die inhaltliche Domäne durch die Indikatoren weiter gut abgedeckt wird.

3.3 Probleme beim Nachweis von Konstruktvalidität

Wie im Abschnitt 2.3 ausgeführt, werden bei der kausalanalytischen Validitätsprüfung der Nachweis von Konvergenz- und Diskriminanzvalidität als wesentliche Indikatoren für das Vorhandensein von Konstruktvalidität gefordert. Eine zusätzliche Forderung ist nomologische Validität, d.h. die Bewährung der Messkonzepte in einem komplexen Kausalmodell [68]. Für den Grad der Schärfe der Modellprüfungen ist dabei entscheidend, ob für die Messungen der Konstrukte die gleichen oder aber unterschiedliche Methoden vorliegen. Im ersten Fall erfolgt die Konstruktvalidierung auf Basis eines klassischen konfirmatorischen Faktorenmodells, während im zweiten Fall komplexere Modellformen zur Zerlegung von Fehlerstrukturen auf Basis von MTMM-Matrizen vorliegen [69]. Tatsächlich verbanden Campbell und Fiske (1959) mit Konvergenzvalidität die Überlegung, dass valide Messungen mit maximal unterschiedlichen Methoden hoch miteinander korrelieren sollten [70]. Steenkamp und van Trijp (1991) sehen die Überprüfung der Inner-Methoden-Konvergenz deshalb nur als Vorstufe einer Analyse der Konvergenz von Messungen mit verschiedenen Methoden an. In

der Marketingforschung wird diese Anforderung jedoch üblicherweise aufgegeben und man begnügt sich mit der Konvergenz von multiplen Messungen mit praktisch identischer Methode (z. B. Fragebogen-Statements mit gleicher Antwortskala für die einzelnen Indikatoren). Da die Konvergenz im Rahmen eines konfirmatorischen Faktorenmodells aber auch das Ergebnis eines Methodenfaktors sein kann, liefert ein solcher Test immer vorsichtig zu beurteilende Ergebnisse. Wir werden uns aufgrund der großen forschungspraktischen Bedeutung auf die Konstruktmessung mit einer Methode konzentrieren und aufzeigen, wie mögliche Fehlereinflüsse aufgedeckt und kontrolliert werden können.

Eine empirische Bestätigung für Konvergenz liegt vor, wenn in einem fittenden konfirmatorischen Faktorenmodell die Ladungen signifikant von Null verschieden und hinreichend groß sind [71]. Zusätzlich werden die Faktorreliabilität und die durchschnittlich erfasste Varianz als Kriterien vorgeschlagen [72]. Ziel der Diskriminanzvalidierung ist es, sicherzustellen, dass die Messungen eines Konstruktes nicht zu hoch mit den Messungen eines anderen Konstruktes korrelieren [73]. Der Nachweis der Diskriminanzvalidität erfolgt über die Beurteilung der Korrelation zwischen den Faktoren in einem Mehrfaktorenmodell. Dabei stehen im Mono-Methodenfall drei Optionen zur Verfügung [74]. Zunächst kann für jedes Konstruktpaar (bzw. Paar aus Dimensionen eines Konstruktes) überprüft werden, ob der Wert 1 im 95 %-Konfidenzintervall der geschätzten Korrelation zwischen den beiden Faktoren enthalten ist. Ist dies der Fall, so ist dies ein Anzeichen dafür, dass die beiden Konstrukte inhaltlich nicht zu unterscheiden sind. Darüber hinaus kann mit Hilfe eines χ^2 -Differenztests getestet werden, ob die Annahme, die Indikatoren zweier Konstrukte würden das gleiche Konstrukt messen, zu einer signifikanten Verschlechterung des Modellfits führt [75]. Im einfachsten Fall nur zweier Konstrukte ist hierfür, ausgehend vom unrestringierten Modell, nur die Fixierung der Korrelation zwischen den Faktoren auf den Wert 1 erforderlich. Beinhaltet das gesamte konfirmatorische Faktorenmodell allerdings mehr als zwei Konstrukte, so muss zusätzlich die Korrelation der beiden Faktoren mit je einem weiteren Faktor auf Gleichheit restringiert werden [76]. Ein weiteres Kriterium zur Überprüfung der Diskriminanzvalidität haben Fornell und Larcker (1981) formuliert. Demnach muss für jedes Paar von Konstrukten die durchschnittlich erfasste Varianz ρ_{DEV} für jeden der beiden Faktoren größer als ihre quadrierte Korrelation sein.

Zusätzlich zur Konvergenz- und Diskriminanzvalidität wird für Konstruktvalidität gefordert, dass die theoretisch zu erwartenden Beziehungen eines Konstruktes mit anderen, vor- oder

nachgelagerten Konstrukten empirisch gestützt werden [77]. Für die Schätzung eines solchen nomologischen Netzwerkes als Strukturgleichungsmodell ist daher zunächst eine adäquate Anpassung an die Ausgangsdaten zu fordern. Im Detail sollten sich darüber hinaus die Hypothesen über die Beziehungen des im Fokus stehenden Konstruktes mit den anderen Konstrukten bestätigen. In den Sozialwissenschaften stellt sich hier allerdings das Problem, dass die entwickelten Hypothesen wenig restriktiv sind, da typischerweise nur die Richtung einer Beziehung (positiver oder negativer Einfluss) angegeben wird. Damit kann theoretisch auch eine Vielzahl weiterer Konstrukte mit den gefundenen Beziehungsstrukturen kompatibel sein [78]. Die Stringenz der nomologischen Validierung hängt somit stark davon ab, wie präzise die Hypothesen formuliert sind. Darüber hinaus werden mit der Akzeptanz des spezifizierten Modells häufig implizit auch alternative Modelle akzeptiert. Diese kovarianzäquivalenten Modelle führen zwar zum gleichen Fit, können aber zum Teil genau entgegengesetzte Kausalbeziehungen beinhalten [79]. Kausale Inferenzalgorithmen [80] können hier den Anwender in der Suche nach alternativen Modellen und der Identifikation invarianter Variablenbeziehungen (d. h. Beziehungen, die Bestandteil aller äquivalenten Modelle sind) unterstützen [81].

3.3.1 Die Kontrolle von Methodeneffekten bei der Konstruktvalidierung

Führt die Validierung einer theoretisch hergeleiteten Messkonzeption mit der konfirmatorischen Faktorenanalyse zu einer unzureichenden Approximation der empirischen Kovarianzstruktur, so sollte – statt das Problem einfach über die schrittweise Elimination von Indikatoren „zu lösen“ – überprüft werden, ob hierfür andere systematische Einflüsse verantwortlich sind. Um diesen Aspekt zu verdeutlichen, wird die in Gleichung (1) dargestellte „True-score“-Konzeption der klassischen Messtheorie um die in der Strukturgleichungsmethodologie vorzufindende Auffassung erweitert, nach der sich jede Messung eines Indikators aus dem Konstrukteinfluss, systematischen Fehlerquellen sowie zufälligen Fehlern zusammensetzt [82]:

$$x_i = \underbrace{\lambda_{ij}\xi_j}_{\text{Konstrukteinfluss}} + \underbrace{\sum_{k=1}^K \pi_{ik}\eta_k}_{\text{system. Fehler}} + \underbrace{\varepsilon_i}_{\text{zuf. Fehler}} = \underbrace{\lambda_{ij}\xi_j}_{\text{Konstrukteinfluss}} + \underbrace{\delta_i}_{\text{Residualgröße}} \quad (4)$$

Dabei sind x_i der beobachtete Wert für den i -ten Indikator des Konstruktes ξ_j , η_k eine systematische Fehlerquelle (z. B. unberücksichtigtes Konstrukt oder Methodeneinfluss) und ε_i der zufällige Messfehler. Die Parameter λ_{ij} und π_{ik} geben den jeweiligen Einfluss der latenten

Variablen ξ_j und η_k auf die Messung x_i an. Während im konventionellen konfirmatorischen Faktorenmodell die Residualgröße δ_i sowohl systematische Fehler als auch zufällige Messfehler umfasst [83], können systematische Einflüsse wie z. B. Methodeneffekte durch geeignete Erweiterungen dieses Modells isoliert werden [84].

Derartige Zerlegungen sind z. B. durch die kausalanalytische Prüfung der Konstruktvalidität anhand einer Multitrait-Multimethod-Matrix (MTMM-Matrix) möglich. Das vorgeschlagene Design der MTMM-Matrix von Campbell und Fiske (1959) verlangt zur Identifizierung des entsprechenden Faktorenmodells bei Berücksichtigung aller Methodeneffekte allerdings mindestens drei unterschiedliche Messverfahren und mindestens drei Konstrukte [85].

Die Anforderung, Befragte in einer Studie drei verschiedene Skalen für ein Konstrukt beantworten zu lassen, werden in der Marketingforschung selten zu erfüllen sein. Üblicherweise wird deshalb ein Konstrukt nur über eine Multi-Item-Skala erfasst. In den mehrstufigen Konzeptionen für den Skalenentwicklungsprozess wird daher die explizite Prüfung von Methodeneffekten zum Teil nicht problematisiert [86], und selbst in angesehenen Zeitschriften wie dem *Journal of Consumer Research* kontrollieren nur wenige Studien derartige systematische Fehlereinflüsse [87]. Tatsächlich existieren mehrere Vorschläge, wie über spezielle Strukturen typische Methodeneffekte in Multi-Item-Messmodellen aufgefangen werden können, in denen nur eine Methode zur Messung eingesetzt wird. Auf drei ausgewählte Modelle zum Test verschiedener Methodeneffekte, nämlich

- unterschiedliche Antwortformate (Modell 1)
- Verwendung eines gemeinsamen Messinstruments (Modell 2)
- Reihenfolgeeffekte (Modell 3)

wird nachfolgend eingegangen.

Das erste Modell (vgl. Abbildung 4) entspricht in seiner Grundstruktur dem für die Analyse von MTMM-Matrizen vorgeschlagenen „Correlated uniqueness“ (CU)-Modell [88]. Ein solches Modell kann z. B. verwendet werden, wenn bei der Messung eines Konstruktes unterschiedliche Antwortformate (im dargestellten Fall z. B. drei) eingesetzt wurden. Dabei werden für Indikatoren, die das gleiche Antwortformat verwenden (z. B. die Indikatoren x_1 , x_4 und x_7) Korrelationen zwischen den Residuen eingeführt. Hier ist der Forscher in der Lage, sowohl Konvergenz- als auch Diskriminanzvalidität unter Berücksichtigung möglicher Methodeneffekte zu prüfen [89]. Neben dem CU-Modell, das ein festes Schema für die

Fehlerkorrelationen vorgibt, können darüber hinaus auch Modelle mit einem flexiblen Korrelationsmuster spezifiziert werden, bei dem Fehlerkorrelationen innerhalb eines Konstruktes und/oder über die Konstrukte hinweg zugelassen werden. Dies kann z. B. sinnvoll sein, wenn positive und negative Itemformulierungen verwendet werden [90].

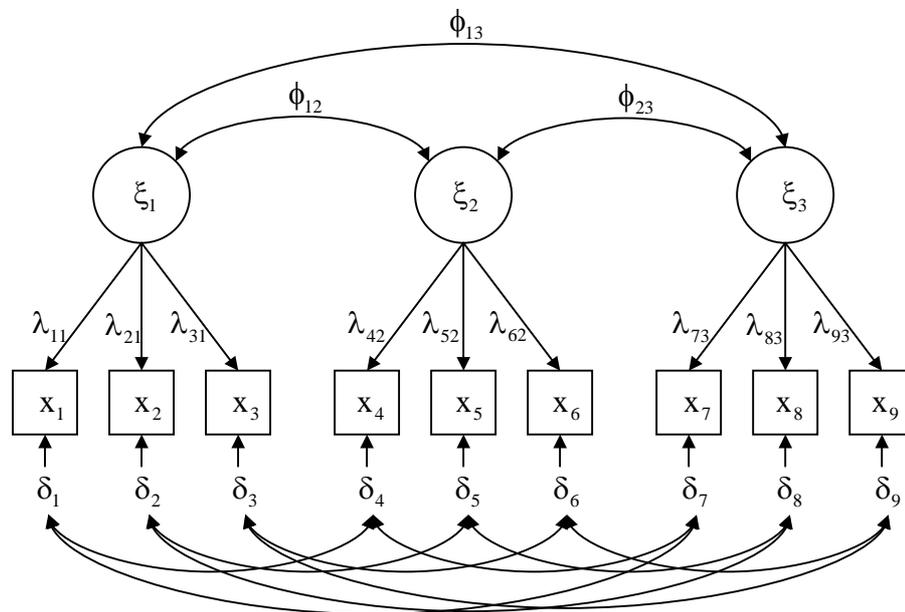


Abb. 4: „Correlated uniqueness“-Modell – Kontrolle von Methodeneinflüssen ohne explizite Einführung eines Methodenfaktors (Modell 1)

Im zweiten Modell (vgl. Abbildung 5) werden Kovariationen, die durch die Verwendung gleicher Skalen in Itembatterien oder sozial erwünschte Antworten hervorgerufen werden, durch die Einführung eines generellen Methodenfaktors in einem konfirmatorischen Faktorenmodell erfasst [91]. Die Stärke des Methodeneinflusses lässt sich dabei an der Höhe der Faktorladungen für den Methodenfaktor ablesen. Ein solches Modell ist besser geeignet, „Common method“-Varianz zu erfassen, als Ansätze, die auf der exploratorischen Faktorenanalyse beruhen [92]. Dort wird untersucht, ob sich die Kovarianzen zwischen den Indikatoren auf einen einzelnen Faktor zurückführen lassen bzw. ob ein sogenannter „Generalfaktor“ den größten Teil der Kovarianzen erklären kann [93]. Alternative Spezifikationen ergeben sich dadurch, dass ein Methodenfaktor nur ausgewählte Indikatoren beeinflusst [94] oder mehr als ein Methodenfaktor eingeführt wird [95]. Letzteres kann z. B. wieder bei der Verwendung positiv und negativ formulierter Statements sinnvoll sein. Da mit korrelierten Residuen sowie der Spezifikation einer unterschiedlichen Anzahl von Methodenfaktoren verschiedene Optionen zur Berücksichtigung methodischer Einflüsse bestehen, sollten entsprechende Modellvergleiche durchgeführt werden [96].

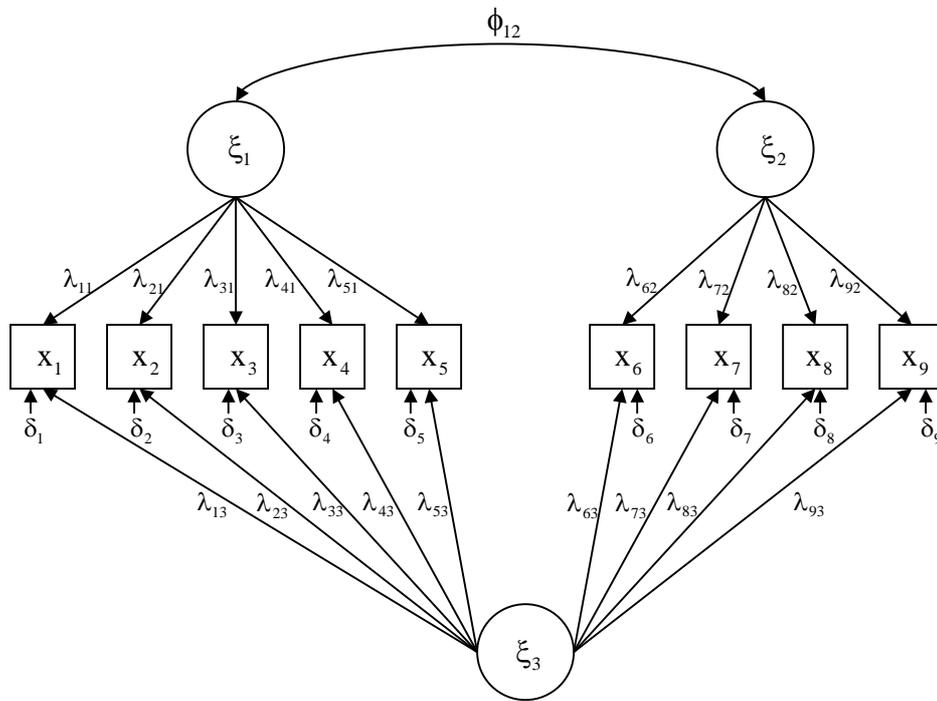


Abb. 5: Kontrolle von Methodeneinflüssen im Monomethodenfall (Modell 2)

Ein genereller Vorteil der expliziten Spezifikation von Methodenfaktoren gegenüber Modellen mit korrelierten Residuen besteht in erweiterten Testmöglichkeiten (z. B. auf Korrelation der Methodeneffekte) und der leichteren Ermittlung von Varianzanteilen, die auf Methodenfaktoren zurückzuführen sind [97]. Insgesamt sind allerdings der Möglichkeit, Methodenfaktoren einzuführen, aus Identifikationsgründen Grenzen gesetzt, insbesondere dann, wenn die Anzahl der Indikatoren pro Faktor relativ klein ist [98].

Das dritte Modell (vgl. Abbildung 6) geht davon aus, dass zwischen den Items insofern Reihenfolgeeffekte vorliegen, als die Antwort auf ein Item die Reaktion auf das nachfolgende Item beeinflusst [99]. Eine solche Antwortverzerrung kann z. B. durch das Motiv, konsistente Antworten geben zu wollen, hervorgerufen werden [100]. Neben der Modellierung eines einfachen autoregressiven Prozesses 1. Ordnung, der zu einer Simplex-Struktur der Korrelationsmatrix für die Residuen führt, können komplexere Modelle, z. B. durch die Einführung von Memory-Faktoren, spezifiziert und miteinander verglichen werden [101]. Lassen sich Konsistenzmotive durch die Schätzung solcher Modelle nachweisen, so muss der Anwender vor dem Hintergrund der Forschungsfrage entscheiden, ob er seine Messkonzeption weiter verwenden will. Zumindest erhält er wertvolle Informationen über das Ausmaß der Antwortverzerrung und kann so geeignete Maßnahmen (z. B. Änderung des Antwortformates) ergreifen, um diese Effekte für weitere Anwendungen seines Messmodells

zu eliminieren oder zumindest zu verringern. Bei jeder Verwendung von Multi-Itemskalen müsste standardmäßig ein Test auf Reihenfolgeeffekt durchgeführt werden.

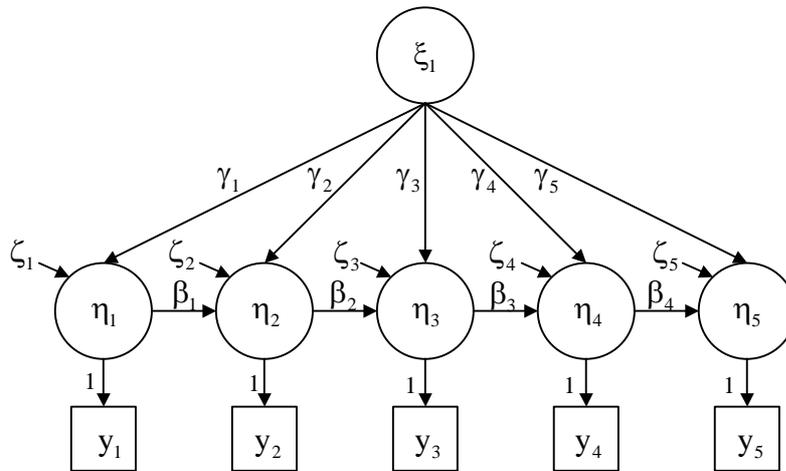


Abb. 6: Modell mit direkter autoregressiver Beziehung zwischen den Indikatoren (Modell 3)

Die Analyse in Abschnitt 3.2.2 demonstrierte die möglichen Folgen der Itemselektion nach vorgegebenen Reliabilitätsanforderungen. Die Strukturgleichungsmethodologie bietet hier, über die oben gezeigten Modelltypen, Alternativen zu den üblichen Validierungen mit konfirmatorischen Faktorenmodellen für Konvergenz- und Diskriminanzvalidität. Für unseren Datensatz wird deshalb – als eine mögliche Option – geprüft, ob die vergleichsweise hohe Korrelation der Indikatoren x_3 - x_6 bzw. x_7 - x_{10} auf einen Methodeneffekt zurückzuführen ist (z. B. Verwendung des gleichen Antwortformates wie „stimme zu – stimme nicht zu“). Hierzu wird das konfirmatorische Zwei-Faktorenmodell um einen Methodenfaktor erweitert, auf den die Indikatoren x_3 bis x_{10} laden (vgl. Abbildung 7). Dabei wird unterstellt, dass der Methodenfaktor mit den beiden Konstrukt Faktoren unkorreliert ist. Um das Modell möglichst sparsam zu parametrisieren, werden darüber hinaus die Faktorladungen des Methodenfaktors auf Gleichheit restringiert. Die Schätzung dieses Modells führt zu einer deutlichen Verbesserung gegenüber dem Ausgangsmodell mit allen Indikatoren ($\Delta\chi^2 = 52,29$, $df = 1$, $p = 0,000$) und einem insgesamt ausreichenden Fit ($\chi^2 = 116,32$, $df = 52$; CFI = 0,966; RMSEA = 0,070). Für den Methodenfaktor ergeben sich eine hochsignifikante Faktorladung und ein Erklärungsanteil an der Varianz der Indikatoren x_3 bis x_{10} , der zwischen 22 % und 25 % liegt. Durch die Kontrolle von Methodeneffekten ist die geschätzte Korrelation zwischen den beiden Konstrukt Faktoren nicht mehr signifikant ($r_{\xi_1, \xi_2} = -0,051$, $z = -0,639$). Vergleicht man diesen Wert mit den hochsignifikanten Korrelationen für die beiden Modelle ohne Methodenfaktoren ($r_{\xi_1, \xi_2} = 0,375$ für das Modell mit allen Indikatoren und $r_{\xi_1, \xi_2} = 0,506$ für das

Modell nach Itemelimination), so konnte offensichtlich der verzerrende Einfluss methodischer Effekte kontrolliert werden. Einschränkend muss angemerkt werden, dass die Kontrolle von Methodenfaktoren in Kausalmodellen kein Allheilmittel zur Lösung von Problemen darstellen, die in früheren Phasen der Skalenentwicklung entstanden sind. Vor ihrer Einführung in ein Modell sollten darüber hinaus begründete Vermutungen darüber existieren, welche möglichen methodischen Effekte (z. B. positive versus negative Itemformulierungen, Antwortskaleneffekte, Konsistenzmotive der Befragten) vorliegen können und welche substantiellen Einflüsse (z. B. nichtberücksichtigte Konstrukte/inhaltliche Subdimensionen) hierdurch u. U. auch aufgefangen werden. Darüber hinaus sind i.d.R. alternative Modellspezifikationen zu testen.

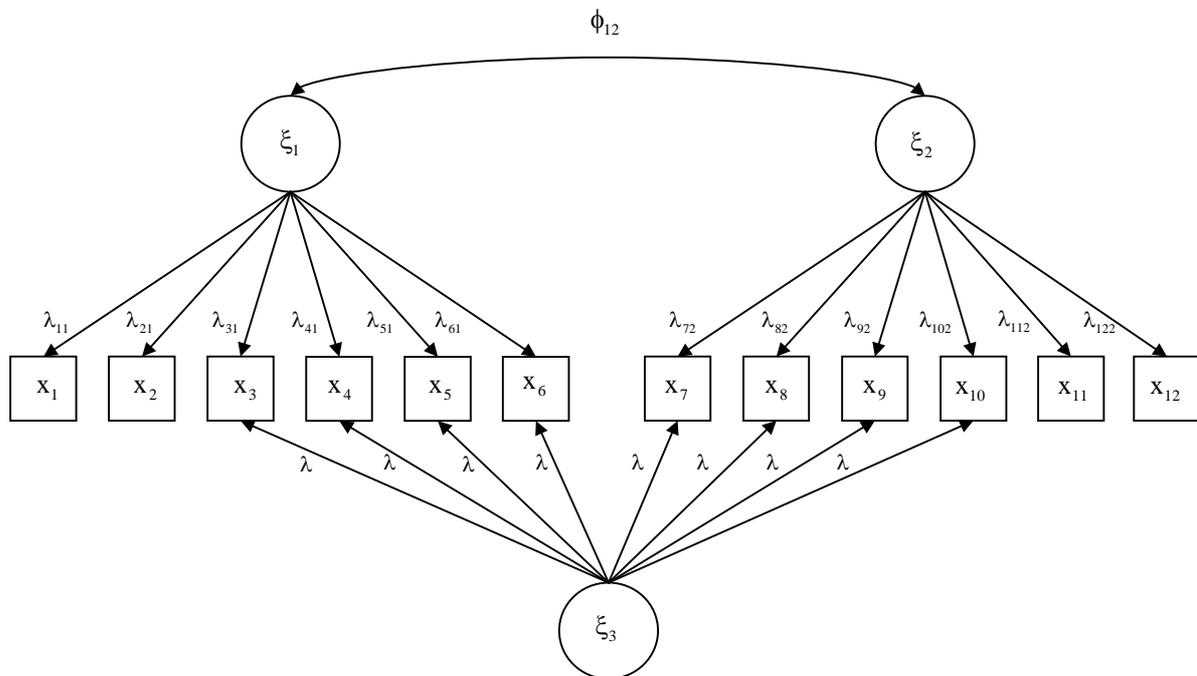


Abb. 7: Konfirmatorisches Zwei-Faktorenmodell mit allen Indikatoren und einem Methodenfaktor

4. Diskussion

Der Prozess der Skalenkonstruktion in der empirischen Marketingforschung beruht auf theoretischen Überlegungen und dem Einsatz statistischer Methoden bei der Validierung der entwickelten Messmodelle. Dabei nimmt die Validierung mit der Kausalanalyse in vielen Studien eine zentrale Rolle ein. Wir haben in diesem Beitrag versucht aufzuzeigen, dass mit der mechanistischen Verfolgung von Skalenbereinigerungsverfahren die inhaltliche Validität der Konstruktoperationalisierung gefährdet sein kann. Die ausschließlich an

Reliabilitätskriterien ausgerichtete Selektion von Indikatoren zur Erzeugung eines akzeptablen konfirmatorischen Faktormodells kann zum einen zu methodischen Artefakten führen. Zum anderen wird durch die Skalenbereinigung in komplexen Messmodellen u. U. die am Anfang einer Studie theoretisch abgeleitete Konzeption des Konstrukts soweit verändert, dass das zugrundeliegende Phänomen nicht oder nur noch teilweise über das am Ende entwickelte Messmodell erfasst wird.

Die Sicherung der Inhaltsvalidität sollte daher bei der kausalanalytischen Reliabilitäts- und Validitätsprüfung in den Vordergrund gestellt werden. Dies kann einerseits über die Anwendung eines systematischen facettentheoretischen Ansatzes bei der Itemkonstruktion erreicht werden, der die Beurteilung der Validität der Messkonzeption nach Itembereinigungen für den Leser nachvollziehbarer macht. Andererseits wird vorgeschlagen, über die übliche Form der Prüfung von Konvergenz- und Diskriminanzvalidität mit der konfirmatorischen Faktorenanalyse hinauszugehen, und explizit das Vorhandensein möglicher Methodeneffekte (z. B. „Common method bias“) zu testen. Die Strukturgleichungsmethodologie erlaubt es, durch eine Varianzzerlegung die Methodeneffekte in den Indikatoren von den zufälligen Fehleranteilen und den substantiellen Konstrukteinflüssen zu trennen und transparent zu machen. Die Prozedur ist zwar aufwändiger, da neben den substantiellen Überlegungen zusätzliche Hypothesen über die möglichen Ursachen von Methodeneinflüssen zu entwickeln sind. Der Vorteil besteht aber darin, dass diese Einflussgrößen – sofern vorhanden – explizit bei der Reliabilitäts- und Validitätsbeurteilung eines Messmodells einbezogen bzw. herausgerechnet werden können. Bei der Prüfung nomologischer Validität im Rahmen komplexerer Kausalstrukturen können zudem Methodeneffekte mit modelliert werden. Explizit in das Modell mit aufgenommene Methodenfaktoren würden auch eher der ganzheitlichen Theorietest-Konzeption von Bagozzi (1998) entsprechen als dem Ansatz, in der Vorstufe Items zu entfernen und damit u. U. wertvolle Informationen zu verlieren.

Anhang

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}
X_1	1,000											
X_2	0,505	1,000										
X_3	0,527	0,548	1,000									
X_4	0,488	0,534	0,662	1,000								
X_5	0,531	0,529	0,672	0,612	1,000							
X_6	0,562	0,586	0,696	0,647	0,661	1,000						
X_7	0,056	0,054	0,417	0,343	0,327	0,284	1,000					
X_8	0,027	0,083	0,374	0,331	0,274	0,258	0,686	1,000				
X_9	-0,010	0,007	0,329	0,272	0,239	0,203	0,710	0,641	1,000			
X_{10}	0,101	0,136	0,482	0,418	0,379	0,354	0,784	0,720	0,717	1,000		
X_{11}	-0,185	-0,204	0,097	0,045	0,028	-0,030	0,621	0,535	0,585	0,589	1,000	
X_{12}	-0,192	-0,143	0,103	0,081	0,016	-0,012	0,574	0,554	0,561	0,582	0,544	1,000

Tab. A1: Korrelationsmatrix für die Indikatoren des Zwei-Faktorenmodells

Verzeichnis der zitierten Literatur

Albers, Sönke/Hildebrandt, Lutz (2006): Methodische Probleme bei der Erfolgsfaktorenforschung – Messfehler, formative versus reflektive Indikatoren und die Wahl des Strukturgleichungs-Modells. In: Zeitschrift für betriebswirtschaftliche Forschung, 58. Jg. (2006), H. 1, S. 2–33.

Anderson, James C./Gerbing, David W. (1988): Structural Equation Modeling in Practice: A Review and Recommended Two-Step Approach. In: Psychological Bulletin, Vol. 103 (1988), No. 3, S. 411–423.

Bagozzi, Richard P. (1980): Causal Models in Marketing. New York 1980.

Bagozzi, Richard P. (1994): Measurement in Marketing Research: Basic Principles of Questionnaire Design. In: Bagozzi, Richard P. (Hrsg.): Principles of Marketing Research. Cambridge, MA, 1994, S. 1–49.

Bagozzi, Richard P. (1998): A Prospectus for Theory Construction in Marketing: Revisited and Revised. In: Hildebrandt, Lutz/Homburg, Christian (Hrsg.): Die Kausalanalyse. Instrument der empirischen betriebswirtschaftlichen Forschung. Stuttgart 1998, S. 44–81.

Bagozzi, Richard P./Phillips, Lynn W. (1982): Representing and Testing Organizational Theories: A Holistic Construal. In: Administrative Science Quarterly, Vol. 27 (1982), S. 459–489.

Bagozzi, Richard P./Baumgartner, Hans (1994): The Evaluation of Structural Equation Models and Hypothesis Testing. In: Bagozzi, Richard P. (Hrsg.): Principles of Marketing Research. Cambridge, MA, 1994, S. 386–482.

Bagozzi, Richard P./Yi, Youjae (1988): On the Evaluation of Structural Equation Models. In: Journal of the Academy of Marketing Science, Vol. 16 (1988), No. 1, S. 74–94.

Balderjahn, Ingo (2003): Validität – Konzept und Methoden. In: WiSt – Wirtschaftswissenschaftliches Studium, 32. Jg. (2003), H. 3, S. 130–135.

Balderjahn, Ingo (1998): Die Kreuzvalidierung von Kausalmodellen. In: Hildebrandt, Lutz/Homburg, Christian (Hrsg.): Die Kausalanalyse. Instrument der empirischen betriebswirtschaftlichen Forschung. Stuttgart 1998, S. 371–397.

- Bauer, Hans H./Grether, Mark/Borrmann, Ulrike (2001): Die Erklärung des Nutzerverhaltens in elektronischen Medien mit Hilfe der Flow-Theorie. In: *Marketing ZFP*, 23. Jg. (2001), Nr. 1, S. 17–30.
- Baumgartner, Hans/Steenkamp, Jan-Benedict E.M. (2001): Response Styles in Marketing Research: A Cross-National Investigation. In: *Journal of Marketing Research*, Vol. 38 (May 2001), S. 143–156.
- Bearden, William O./Netemeyer Richard G. (1998): *Handbook of Marketing Scales: Multi-Item Measures for Marketing and Consumer Behavior Research*. 2nd ed., Palo Alto, CA, 1998.
- Bollen, Kenneth A. (1989): *Structural Equations with Latent Variables*. New York 1989.
- Bollen, Kenneth/Lennox, Richard (1991): Conventional Wisdom on Measurement: A Structural Equation Perspective. In: *Psychological Bulletin*, Vol. 110 (1991), No. 2, S. 305–314.
- Bollen, Kenneth A. (2000): Modeling Strategies: In Search of the Holy Grail. In: *Structural Equation Modeling*, Vol. 7 (2000), No. 1, S. 74–81.
- Borg, Ingwer (1977): Some Basis Concepts of Facet Theory. In: Lingoës, James C. (Hrsg.): *Geometric Representations of Relational Data*. Ann Arbor, MI, 1977, S. 65–102.
- Borg, Ingwer/Shye, Samuel (1995): *Facet Theory: Form and Content*. Thousand Oaks CA, 1995.
- Borsboom, Denny/Mellenberg, Gideon J./van Heerden, Jaap (2004): The Concept of Validity. In: *Psychological Review*, Vol. 111 (2004), No. 4, S. 1061–1071.
- Bruner, Gordon C. (2003): Combating Scale Proliferation. In: *Journal of Targeting, Measurement and Analysis for Marketing*, Vol. 11 (2003), No. 4, S. 362–372.
- Bruner, Gordon C./Hensel, Paul J. (1997): *Marketing Scales Handbook: A Compilation of Multi-Item Measures*, 2nd ed., Chicago 1997.
- Campbell, Donald T./Fiske, Donald W. (1959): Convergent and Discriminat Validation by the Multitrait-Multimethod Matrix. In: *Psychological Bulletin*, Vol. 56 (1959), No. 2, S. 81–105.
- Campbell, Donald T. (1960): Recommendations for APA Test Standards Regarding Construct, Trait, or Discriminant Validity. In: *American Psychologist*, Vol. 15 (1960), S. 546–553.
- Cattell, Raymond B./Tsujioka, Bien (1964): The Importance of Factor-Trueness and Validity, versus Homogeneity and Orthogonality, in Test Scales. In: *Educational and Psychological Measurement*, Vol. 14 (1964), No. 1, S. 3–30.
- Churchill, Gilbert A. (1979): A Paradigm for Developing Better Measures of Marketing Constructs. In: *Journal of Marketing Research*, Vol. 16 (February 1979), S. 64–73.
- Clark, Lee A./Watson, David (1995): Constructing Validity: Basic Issues in Objective Scale Development. In: *Psychological Assessment*, Vol. 7 (1995), No. 3, S. 309–319.
- Cohen, Jacob (1988): *Statistical Power Analysis for the Behavioral Sciences*. New Jersey 1988.
- Conway, James M. (1998): Understanding Method Variance in Multitrait-Multirater Performance Appraisal Matrices: Examples Using General Impressions and Interpersonal Affect as Measured Method Factors. In: *Human Performance*, Vol. 11 (1989), No. 1, S. 29–55.

- Cortina, Jose M. (1993): What Is Coefficient Alpha? An Examination of Theory and Applications. In: *Journal of Applied Psychology*, Vol. 78 (1993), No. 1, S. 98–104.
- Costello, Anna B./Osborne, Jason W. (2005): Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most From Your Analysis. In: *Practical Assessment Research & Evaluation (Electronic Journal)*, Vol. 10 (2005), No. 7, S. 1–9.
- Cronbach, Lee J. (1951): Coefficient Alpha and the Internal Structure of Tests. In: *Psychometrika*, Vol. 16 (1951), No. 3, S. 297–334.
- Cronbach, Lee J./Meehl, Paul E. (1955): Construct Validity in Psychological Tests. In: *Psychological Bulletin*, Vol. 52 (1955), No. 4, S. 281–302.
- Cudeck, Robert (1989): Analysis of Correlation Matrices Using Covariance Structure Models. In: *Psychological Bulletin*, Vol. 105 (1989), No. 2, S. 317–327.
- Cudeck, Robert/Browne, Michael W. (1983): Cross-Validation of Covariance Structures. In: *Multivariate Behavioral Research*, Vol. 18 (1983), No. 2, S. 147–167.
- Diamantopoulos, Adamantios/Winklhofer, H. M. (2001): Index Construction with Formative Indicators: An Alternative to Scale Development. In: *Journal of Marketing Research*, Vol. 38 (May 2001), S. 269–277.
- Diller, Hermann (2003): Editorial: Das süße Gift der Kausalanalyse. In: *Marketing ZFP*, 26. Jg. (2003), Nr. 3, S. 177.
- Edwards, Jeffrey R. (2001): Multidimensional Constructs in Organisational Behavior Research: An Integrative Analytical Framework. In: *Organisational Research Methods*, Vol. 4 (2001), No. 2, S. 144–192.
- Edwards, Jeffrey R. (2003): Construct Validation in Organizational Behavior Research, in: Greenberg (Hrsg.): *Organization Behavior: The State of the Science*, 2nd ed., New Jersey 2003, S. 327–372.
- Ewing, Michael T./Napoli, Julie (2005): Developing and Validating a Multidimensional Nonprofit Brand Orientation Scale. In: *Journal of Business Research*, Vol. 58 (2005), S. 841–853.
- Fabrigar, Leandre R./Wegener, Duane T./MacCallum, Robert C./Strahan, Erin J. (1999): Evaluating the Use of Exploratory Factor Analysis in Psychological Research. In: *Psychological Methods*, Vol. 4 (1999), No. 3, S. 272–299.
- Floyd, Frank J./Widaman, Keith F. (1995): Factor Analysis in the Development and Refinement of Clinical Assessment Instruments. In: *Psychological Assessment*, Vol. 7 (1995), No. 3, S. 286–299.
- Fornell, Claes/Larcker, David F. (1981): Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. In: *Journal of Marketing Research*, Vol. 18 (February 1981), S. 39–50.
- Gerbing, David W./Anderson, James C. (1988): An Updated Paradigm for Scale Development Incorporating Unidimensionality and Its Assessment. In: *Journal of Marketing Research*, Vol. 25 (May 1988), S. 186–192.
- Green, Samuel B./Hershberger, Scott L. (2000): Correlated Errors in True Score Models and Their Effect on Coefficient Alpha. In: *Structural Equation Modeling*, Vol. 7 (2000), No. 2, S. 251–270.
- Gulliksen, Harold (1950): The Reliability of Speeded Tests. In: *Psychometrika*, Vol. 15 (1950), No. 3, S. 259–269.

- Guttman, Louis (1971): Measurement as Structural Theory. In: *Psychometrika*, Vol. 36 (1971), No. 4, S. 329–347.
- Harmann, Harry H. (1976): *Modern Factor Analysis*. Chicago 1976.
- Hattie, John (1985): Methodology Review: Assessing Unidimensionality of Tests and Items. In: *Applied Psychological Measurement*, Vol. 9 (June 1985), S. 139–164.
- Hildebrandt, Lutz (1986): A Facet Theoretical Approach for Testing Measurement and Structural Theories. An Application of Confirmatory MDS. In: Lutz, Richard J. (Ed.): *Advances in Consumer Research*, Vol. 8, Provo.
- Hildebrandt, Lutz (1998): Kausalanalytische Validierung in der Marktforschung. In: Hildebrandt, Lutz/Homburg, Christian (Hrsg.): *Die Kausalanalyse. Instrument der empirischen betriebswirtschaftlichen Forschung*. Stuttgart 1998, S. 85–110.
- Hildebrandt, Lutz (2000): Hypothesenbildung und empirische Überprüfung, In: Hermann, Andreas/Homburg, Christian (Hrsg.): *Marktforschung*. Wiesbaden 2000, S. 34–57.
- Hildebrandt, Lutz/Temme, Dirk (2005): Strukturgleichungsmodelle in der betriebswirtschaftlichen Forschung. In: Steven, Marion/Sonntag, Susanne (Hrsg.): *Quantitative Unternehmensführung – Denken in Austauschraten*, Heidelberg 2005, S. 49–65.
- Homburg, Christian/Baumgartner, Hans (1998): Beurteilung von Kausalmodellen: Bestandsaufnahme und Anwendungsempfehlungen. In: Hildebrandt, Lutz/Homburg, Christian (Hrsg.): *Die Kausalanalyse. Instrument der empirischen betriebswirtschaftlichen Forschung*. Stuttgart 1998, S. 343–369.
- Homburg, Christian/Giering, Annette (1996): Konzeptualisierung und Operationalisierung komplexer Konstrukte. Ein Leitfaden für die Marketingforschung. In: *Marketing ZFP*, 18. Jg. (1996), Nr. 1, S. 5–24.
- Horan, Patrick M./DiStefano, Christine/Motl, Robert W. (2003): Wording Effects in Self-Esteem Scales: Methodological Artifact or Response Style? In: *Structural Equation Modeling*, Vol. 10 (2003), No. 3, S. 435–455.
- Jarvis, Cheryl B./MacKenzie, Scott B./Podsakoff, Philip M. (2003): A Critical Review of Construct Indicators and Measurement Model Misspecification in Marketing and Consumer Research. In: *Journal of Consumer Research*, Vol. 30 (September 2003), S. 199–218.
- Jöreskog, Karl G. (1971): Statistical Analysis of Sets of Congeneric Tests. In: *Psychometrika*, Vol. 36 (1971), No. 2, S. 109–133.
- Johns, Gary (1994): How Often Were You Absent? A Review of the Use of Self-Reported Absence Data. In: *Journal of Applied Psychology*, Vol. 79 (1994), No. 4, S. 574–591.
- Kenny, David A. (1979): *Correlation and Causality*. New York 1979.
- Krafft, Manfred/Götz, Oliver/Liehr-Gobbers, Kerstin (2005): Die Validierung von Strukturgleichungsmodellen mit Hilfe des Partial-Least-Squares (PLS)-Ansatzes. In: Bliemel, Friedhelm/Eggert, Andreas/ Fassott, Georg/Henseler, Jörg (Hrsg.): *Handbuch PLS-Pfadmodellierung*, Stuttgart 2005, S. 71–86.
- Lance, Charles E./Noble, Carrie L./Scullen, Steven E. (2002): A Critique of the Correlated Trait-Correlated Method and Correlated Uniqueness Models for Multitrait-Multimethod Data. In: *Psychological Methods*, Vol. 7 (2002), No. 2, S. 228–244.
- Lee, Soonmook/Hershberger, Scott (1990): A Simple Rule for Generating Equivalent Models in Covariance Structure Modeling. In: *Multivariate Behavioral Research*, Vol. 25 (1990), No. 3, S. 313–334

- Little, Todd D./Lindenberger, Ulman/Nesselroade, John R. (1999): On Selecting Indicators for Multivariate Measurement and Modeling With Latent Variables: When “Good” Indicators Are Bad and “Bad” Indicators are Good. In: *Psychological Methods*, Vol. 4 (1999), No. 2, S. 192–211.
- Lord, Frederick M./Novick, Melvin R. (1968): *Statistical Theories of Mental Test Scores*. Oxford 1986.
- MacKenzie, Scott B. (2001): Opportunities for Improving Consumer Research through Latent Variable Structural Equation Modeling, In: *Journal of Consumer Research*, Vol. 28 (June 2001), S. 159–166.
- Marsh, Herbert W. (1989): Confirmatory Factor Analyses of Multitrait-Multimethod Data: Many Problems and a Few Solutions. In: *Applied Psychological Measurement*, Vol. 13 (1989), No. 4, S. 335–361.
- Marsh, Herbert W./Bailey, Michael (1991): Confirmatory Factor Analyses of Multitrait-Multimethod Data: A Comparison of Alternative Models, In: *Applied Psychological Measurement*, Vol. 15 (1991), No. 1, S. 47–70.
- McDonald, Roderick P. (1970): Theoretical Foundations of Principal Factor Analysis, Canonical Factor Analysis, and Alpha Factor Analysis. In: *British Journal of Mathematical and Statistical Psychology*, Vol. 23 (1970), No. 1, S. 1–21.
- McDonald, Roderick P. (1981): The Dimensionality of Tests and Items. In: *British Journal of Mathematical and Statistical Psychology*, Vol. 34 (1981), No. 1, S. 100–117.
- Muthén, Linda K./Muthén, Bengt O. (1998-2004), *Mplus User’s Guide*. 3rd ed., Los Angeles, CA, 2004.
- Muthén, Linda K./Muthén, Bengt O. (1998-2005), *Mplus Version 3.0*.
- Netemeyer, Richard G./Bearden, William O./Sharma, Subhash (2003): *Scaling Procedures. Issues and Applications*. Thousand Oaks, CA, 2003.
- Novick, Melvin R./Louis, Charles (1967): Coefficient Alpha and the Reliability of Composite Measures. In: *Psychometrika*, Vol. 32 (1967), No. 1, S. 1–13.
- Nunnally, Jum C. (1978): *Psychometric Theory*. 2nd ed., New York 1978.
- Peterson, Robert A. (2000): A Meta-Analysis of Variance Accounted for and Factor Loadings in Exploratory Factor Analysis. In: *Marketing Letters*, Vol. 11 (2000), No. 3, S. 261–275.
- Peterson, Robert A. (1994): A Meta-Analysis of Cronbach’s Coefficient Alpha. In: *Journal of Consumer Research*, Vol. 21 (September 1994), S. 381–391.
- Podsakoff, Philip M./MacKenzie, Scott B./Lee, Jeong-Yeon/Podsakoff, Nathan P. (2003): Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies, In: *Journal of Applied Psychology*, Vol. 88 (2003), No. 5, S. 879–903.
- Podsakoff, Philip M./Organ, Dennis W. (1986): Self-Reports in Organizational Research: Problems and Prospects. In: *Journal of Management*, Vol. 12 (1986), No. 4, S. 69–82.
- Preacher, Kristopher J./MacCallum, Robert C. (2003): Repairing Tom Swift’s Electric Factor Analysis Machine. In: *Understanding Statistics*, Vol. 2 (2003), No. 1, S. 13–43.
- Rossiter, John R. (2002): The C-OAR-SE Procedure for Scale Development in Marketing. In: *International Journal of Research in Marketing*, Vol. 19 (2002), No. 4, S. 305–335.

- Scheines, Richard/Spirtes, Peter/Glymour, Clark/Meek, Christopher (1994), TETRAD II: Tools for Causal Modeling, User's Manual. Hillsdale, 1994.
- Schmitt, Neal (1996): Uses and Abuses of Coefficient Alpha. In: Psychological Assessment, Vol. 8 (1996), No. 4, S. 350–353.
- Scullen, Steven E. (1999): Using Confirmatory Factor Analysis of Correlated Uniqueness to Estimate Method Variance in Multitrait-Multimethod Matrices. In: Organizational Research Methods, Vol. 2 (1999), No. 3, S. 275–292.
- Shye, Samuel (1998): Modern Facet Theory: Content Design and Measurement in Behavioral Research, In: European Journal of Psychological Assessment, Vol. 14 (1998), No. 2, S. 160–171.
- Steenkamp, Jan-Benedict E.M./van Trijp, Hans C.M. (1991): The Use of LISREL in Validating Marketing Constructs. In: International Journal of Research in Marketing, Vol. 8 (1991), No. 4, S. 283–299.
- Stelzl, Ingeborg (1986): Changing a Causal Hypothesis without Changing the Fit: Some Rules for Generating Equivalent Path Models. In: Multivariate Behavioral Research, Vol 21 (1986), No. 3, S. 309–331.
- Stock, Ruth (2004): Wirkungsweise von Normen in Organisationen. Theoretische Betrachtung und empirische Analyse am Beispiel von Teams. In: Zeitschrift für Betriebswirtschaft, 74. Jg. (2004), H. 8, S. 785–810.
- Streiner, David L. (2003): Starting at the Beginning: An Introduction to Coefficient Alpha and Internal Consistency. In: Journal of Personality Assessment, Vol. 80 (2002), No. 1, S. 99–103.
- Temme, Dirk (2006a): Constraint-Based Inference Algorithms for Structural Models with Latent Confounders – Empirical Application and Simulations. In: Computational Statistics, Vol. 21 (2006), S. 151–182.
- Temme, Dirk (2006b): Die Spezifikation und Identifikation formativer Messmodelle der Marketingforschung in Kovarianzstrukturanalysen. In: Marketing ZFP, 27. Jg. (2006), Nr. 2, in Druck.
- Thurstone, Louis L. (1947): Multiple Factor Analysis. Chicago.
- Tomás, José M./Oliver, Amparo (1999): Rosenberg's Self-Esteem Scale: Two Factors or Method Effects. In: Structural Equation Modeling, Vol. 6 (1999), No. 1, S. 84–98.
- van der Sluis, Sophie/Dolan, Conor V./Stoel, Reinoud D. (2005): A Note on Testing Perfect Correlations in SEM. In: Structural Equation Modeling, Vol. 12 (2005), No. 4, S. 551–577.
- Widaman, Keith F. (1985): Hierarchically Nested Covariance Structure Models for Multitrait-Multimethod Data. In: Applied Psychological Measurement, Vol. 9 (1985), No. 1, S. 1–26.
- Williams, Larry J./Ford, Lucy R./Nguyen, Nhung (2002): Basic and Advanced Measurement Models for Confirmatory Factor Analysis. In: Rogelberg, Steven G. (Hrsg.): Handbook of Research Methods in Industrial and Organizational Psychology, Malden MA, S. 366–389.
- Zimmermann, Donald W./Zumbo, Bruno D./Lalonde, Coralie (1993): Coefficient Alpha as an Estimate of Test Reliability under Violation of Two Assumptions. In: Educational and Psychological Measurement, Vol. 53 (1993), No. 1, S. 33–49.

Anmerkungen

-
- [1] Rossiter (2002, S. 315) legt z. B. eine solche Auffassung indirekt nahe, wenn er schreibt „[...] a formed attribut does not follow the domain *sampling* model. This means that items are not interchangeably, that is, items cannot be added or deleted from the scale.“
- [2] Vorgestellt wird die Entwicklung einer Skala („Brand orientation scale“), mit der gemessen werden soll, in welchem Ausmaß sich Non-Profit-Organisationen als Marke verstehen. Dass es sich dabei entgegen der Operationalisierung der Autoren um ein formatives Messmodell handelt, soll hier nicht weiter problematisiert werden.
- [3] Zur Diskussion formativer Messmodelle wird auf Albers/Hildebrandt 2006; Diamantopoulos/Winklhofer 2001; Jarvis et al. 2003; Rossiter 2002 verwiesen.
- [4] Vgl. z. B. Churchill 1979; Gerbing/Anderson 1988; Homburg/Giering 1996; Netemeyer et al. 2003.
- [5] Vgl. Rossiter 2002; Shye 1998.
- [6] Vgl. Bearden/Netemeyer 1998; Bruner/Hensel 1997.
- [7] Vgl. Bruner 2003; Diller 2004.
- [8] Vgl. Nunnally 1978.
- [9] Vgl. Churchill 1979.
- [10] Vgl. Balderjahn 1998; Cudeck/Browne 1983.
- [11] Vgl. Bollen/Lennox 1991.
- [12] Eine Ausnahme bilden Modelle, in denen die formativen „Indikatoren“ reflektiv gemessene Konstrukte darstellen; vgl. Edwards 2001.
- [13] Vgl. Temme 2006.
- [14] Vgl. Krafft/Götz/Liehr-Gobbers 2005.
- [15] Vgl. Churchill 1979; Nunnally 1978; Bagozzi 1998.
- [16] Vgl. Gulliksen 1950; Lord/Novick 1968.
- [17] Vgl. Hildebrandt 1998.
- [18] Vgl. z. B. Netemeyer et al. 2003, S. 11.
- [19] Vgl. z. B. Bagozzi 1980; Bagozzi/Baumgartner 1994.
- [20] Zu alternativen Vorstellungen der Validität im Sinne des Nachweises einer kausalen Beeinflussung der Messungen durch das zugrundeliegende Konstrukt siehe Borsboom et al. (2004). Fragen der Qualität einer Messung (z. B. Reliabilität) werden dabei zunächst bewusst von der Validitätskonzeption ausgeklammert.
- [21] Vgl. Bagozzi 1994; Netemeyer et al. 2003, S. 72 ff.
- [22] Vgl. Cronbach/Meehl 1955.
- [23] Vgl. Campbell/Fiske 1959.
- [24] Vgl. Bagozzi 1998; Balderjahn 2003; Hildebrandt 1998.
- [25] Vgl. z. B. Netemeyer et al. 2003, S. 82 f.
- [26] Vgl. z. B. Bagozzi 1994.
- [27] Vgl. Bagozzi 1980.
- [28] Vgl. Bagozzi 1998; Hildebrandt 2000.
- [29] Vgl. Homburg/Giering 1996.
- [30] Vgl. Cronbach/Meehl 1995; Guttman 1971.
- [31] Vgl. Borsboom et al. 2004; Rossiter 2002.
- [32] Vgl. Borg 1977; Hildebrandt 1985; Rossiter 2002.
- [33] Vgl. z. B. Borg/Shye 1995.
- [34] Vgl. z. B. Hattie 1985; Bei einem mehrdimensionalen Konstrukt ist dies für die jeweiligen Dimensionen zu überprüfen.
- [35] Vgl. Clark/Watson 1995; Cortina 1993.
- [36] Vgl. McDonald 1981.
- [37] Vgl. Fabrigar et al. 1999.
- [38] Vgl. z. B. Costello/Osborne 2005; Fabrigar et al. 1999; Floyd/Widaman 1995.

-
- [39] Vgl. Peterson 2000.
- [40] Vgl. Fabrigar et al. 1999; Preacher/MacCallum 2003.
- [41] Vgl. Thurstone 1947.
- [42] Z. B. Scree-Test, Parallel-Analyse, RMSEA; vgl. Fabrigar et al. 1999.
- [43] Z. B. der MLM-Schätzer in Mplus; vgl. Muthén/Muthén 1998-2004.
- [44] Z. B. Ladung $< 0,40$; vgl. Netemeyer et al. 2003, S. 125.
- [45] Vgl. Little et al. 1999.
- [46] Vgl. Cattell/Tsujioka 1964.
- [47] Vgl. Cronbach 1951.
- [48] Vgl. z. B. Netemeyer et al. 2003, S. 11.
- [49] „essential tau-equivalency“; vgl. Novick/Lewis 1967.
- [50] Vgl. Cortina 1993; Schmitt 1996; Streiner 2003; So hängt etwa die Höhe des Koeffizienten positiv von der Anzahl der Indikatoren ab (Cortina 1993). Darüber hinaus erweist sich die Prämisse der schwachen Tau-Äquivalenz in empirischen Studien in aller Regel als zu restriktiv. Unter der realistischeren Annahme gleichartiger Messungen („congeneric measures“; Jöreskog 1971), die unterschiedliche Faktorladungen zulässt, wird Cronbachs α häufig als Untergrenze für die Reliabilität angesehen. Allerdings können positiv korrelierte Residuen der Indikatoren dazu führen, dass die Reliabilität überschätzt wird (vgl. Zimmermann et al. 1993). Daher reicht streng genommen der Nachweis der Eindimensionalität mittels einer exploratorischen Faktorenanalyse für eine Interpretation dieses Koeffizienten nicht aus.
- [51] Vgl. McDonald 1971.
- [52] Vgl. Fornell/Larcker 1981.
- [53] Vgl. Bagozzi/Baumgartner 1994; Homburg/Giering 1996.
- [54] Vgl. zu weiteren Übersichten etwa Bagozzi/Baumgartner 1994; Homburg/Baumgartner 1998; Homburg/Giering 1996.
- [55] Vgl. Netemeyer et al. 2003, S. 153 f.
- [56] Vgl. Homburg/Giering 1996.
- [57] Vgl. Cortina 1993.
- [58] Vgl. Peterson 1994.
- [59] Vgl. z. B. Harman 1976.
- [60] Die durchschnittliche Korrelation für die Indikatoren x_3 - x_6 und x_7 - x_{10} beträgt $\bar{r} = 0,66$ bzw. $\bar{r} = 0,71$.
- [61] Ein solcher Methodeneffekt könnte z. B. daraus resultieren, dass für die je vier Indikatoren eine Likert-Skala mit Antwortkategorien zwischen den Extremen „stimme überhaupt nicht zu“ und „stimme voll und ganz zu“ verwendet wird (vgl. Baumgartner/Steenkamp 2001); für die übrigen je zwei Indikatoren wird dagegen die Intensität direkt durch abgestufte Antwortkategorien angegeben (vgl. zu einer Diskussion dieser unterschiedlichen Antwortformate z. B. Rossiter 2002).
- [62] Zu einem weiteren Beispiel siehe Bauer/Grether/Borrmann 2001
- [63] S. 797; Hervorhebung durch die Autoren.
- [64] Da es sich um ein illustrierendes Beispiel handelt, werden methodische Feinheiten vernachlässigt (z. B. bezüglich der Analyse einer Korrelations- statt einer Kovarianzmatrix; vgl. z. B. Cudeck 1989). Dies hat jedoch keinen Einfluss auf die grundlegenden Befunde.
- [65] Vgl. z. B. Hildebrandt/Temme 2005.
- [66] Für die Schätzung wurde der ML-Schätzer im Programm Mplus 3.0 (Muthén/Muthén 1998-2005) verwendet.
- [67] Vgl. Cohen 1988.
- [68] Vgl. Bagozzi 1998.
- [69] Vgl. z. B. Widaman 1985.

-
- [70] Vgl. auch Bagozzi/Phillips 1982; Hildebrandt 1986.
- [71] Vgl. Hildebrandt 1998.
- [72] Vgl. z. B. Homburg/Giering 1998.
- [73] Vgl. Campbell/Fiske 1959.
- [74] Vgl. z. B. Netemeyer et al. 2003, S. 154.
- [75] Vgl. Anderson/Gerbing 1988.
- [76] Vgl. van der Sluis et al. 2005.
- [77] Vgl. Campbell 1960; Cronbach/Meehl 1955.
- [78] Vgl. Borsboom/Mellenbergh/van Heerden 2004.
- [79] Vgl. Lee/Hershberger 1990; Stelzl 1986.
- [80] Z. B. TETRAD, Scheines et al. 1994.
- [81] Zu einer Anwendung siehe z. B. Temme 2006.
- [82] Vgl. Bagozzi 1998; Bollen 1989, S. 219.
- [83] Auf die Bedeutung, die dies für die Berechnung der Reliabilität hat, sei auf Bollen (1989, S. 20 f.) verwiesen.
- [84] Zu einer umfassenden Darstellung möglicher Ursachen eines „Common method bias“ siehe Podsakoff et al. (2003).
- [85] Vgl. Kenny 1979; Aufgrund der häufig auftretenden Schätzprobleme (z. B. Nichtkonvergenz und unzulässige Parameterlösungen) bei diesem Modelltyp sind alternative Modelle und Verfahrensweisen vorgeschlagen worden. Zu einer Diskussion dieser Modelle siehe z. B. Edwards (2003); Podsakoff. (2003).
- [86] Vgl. z. B. Homburg/Giering 1996.
- [87] Vgl. MacKenzie 2001.
- [88] Vgl. Kenny 1979; Marsh 1989; Marsh/Bailey 1991.
- [89] Zu einer Kritik des CU-Modells siehe Lance et al. (2002).
- [90] Vgl. Horan et al. 2003.
- [91] Vgl. Podsakoff et al. 2003.
- [92] Vgl. Podsakoff et al. 2003.
- [93] Vgl. Podsakoff/Organ 1986.
- [94] Vgl. z. B. Williams et al. 2002.
- [95] Vgl. Tomás/Oliver 1999.
- [96] Vgl. z. B. Tomás/Oliver 1999.
- [97] Zur Ermittlung von Varianzanteilen im CU-Modell siehe Conway (1998) sowie Scullen (1999).
- [98] Diese Beschränkung gilt selbstverständlich auch für korrelierte Residuen.
- [99] Vgl. z. B. Bollen 2000; Green/Hershberger 2000.
- [100] Vgl. z. B. Johns 1994; Podsakoff/Organ 1986.
- [101] Vgl. Green/Hershberger 2000.

SFB 649 Discussion Paper Series 2006

For a complete list of Discussion Papers published by the SFB 649, please visit <http://sfb649.wiwi.hu-berlin.de>.

- 001 "Calibration Risk for Exotic Options" by Kai Detlefsen and Wolfgang K. Härdle, January 2006.
- 002 "Calibration Design of Implied Volatility Surfaces" by Kai Detlefsen and Wolfgang K. Härdle, January 2006.
- 003 "On the Appropriateness of Inappropriate VaR Models" by Wolfgang Härdle, Zdeněk Hlávka and Gerhard Stahl, January 2006.
- 004 "Regional Labor Markets, Network Externalities and Migration: The Case of German Reunification" by Harald Uhlig, January/February 2006.
- 005 "British Interest Rate Convergence between the US and Europe: A Recursive Cointegration Analysis" by Enzo Weber, January 2006.
- 006 "A Combined Approach for Segment-Specific Analysis of Market Basket Data" by Yasemin Boztuğ and Thomas Reutterer, January 2006.
- 007 "Robust utility maximization in a stochastic factor model" by Daniel Hernández-Hernández and Alexander Schied, January 2006.
- 008 "Economic Growth of Agglomerations and Geographic Concentration of Industries - Evidence for Germany" by Kurt Geppert, Martin Gornig and Axel Werwatz, January 2006.
- 009 "Institutions, Bargaining Power and Labor Shares" by Benjamin Bental and Dominique Demougin, January 2006.
- 010 "Common Functional Principal Components" by Michal Benko, Wolfgang Härdle and Alois Kneip, January 2006.
- 011 "VAR Modeling for Dynamic Semiparametric Factors of Volatility Strings" by Ralf Brüggemann, Wolfgang Härdle, Julius Mungo and Carsten Trenkler, February 2006.
- 012 "Bootstrapping Systems Cointegration Tests with a Prior Adjustment for Deterministic Terms" by Carsten Trenkler, February 2006.
- 013 "Penalties and Optimality in Financial Contracts: Taking Stock" by Michel A. Robe, Eva-Maria Steiger and Pierre-Armand Michel, February 2006.
- 014 "Core Labour Standards and FDI: Friends or Foes? The Case of Child Labour" by Sebastian Braun, February 2006.
- 015 "Graphical Data Representation in Bankruptcy Analysis" by Wolfgang Härdle, Rouslan Moro and Dorothea Schäfer, February 2006.
- 016 "Fiscal Policy Effects in the European Union" by Andreas Thams, February 2006.
- 017 "Estimation with the Nested Logit Model: Specifications and Software Particularities" by Nadja Silberhorn, Yasemin Boztuğ and Lutz Hildebrandt, March 2006.
- 018 "The Bologna Process: How student mobility affects multi-cultural skills and educational quality" by Lydia Mechtenberg and Roland Strausz, March 2006.
- 019 "Cheap Talk in the Classroom" by Lydia Mechtenberg, March 2006.
- 020 "Time Dependent Relative Risk Aversion" by Enzo Giacomini, Michael Handel and Wolfgang Härdle, March 2006.
- 021 "Finite Sample Properties of Impulse Response Intervals in SVECMs with Long-Run Identifying Restrictions" by Ralf Brüggemann, March 2006.
- 022 "Barrier Option Hedging under Constraints: A Viscosity Approach" by Imen Bentahar and Bruno Bouchard, March 2006.

SFB 649, Spandauer Straße 1, D-10178 Berlin
<http://sfb649.wiwi.hu-berlin.de>

This research was supported by the Deutsche
Forschungsgemeinschaft through the SFB 649 "Economic Risk".



- 023 "How Far Are We From The Slippery Slope? The Laffer Curve Revisited" by Mathias Trabandt and Harald Uhlig, April 2006.
- 024 "e-Learning Statistics – A Selective Review" by Wolfgang Härdle, Sigbert Klinke and Uwe Ziegenhagen, April 2006.
- 025 "Macroeconomic Regime Switches and Speculative Attacks" by Bartosz Maćkowiak, April 2006.
- 026 "External Shocks, U.S. Monetary Policy and Macroeconomic Fluctuations in Emerging Markets" by Bartosz Maćkowiak, April 2006.
- 027 "Institutional Competition, Political Process and Holdup" by Bruno Deffains and Dominique Demougin, April 2006.
- 028 "Technological Choice under Organizational Diseconomies of Scale" by Dominique Demougin and Anja Schöttner, April 2006.
- 029 "Tail Conditional Expectation for vector-valued Risks" by Imen Bentahar, April 2006.
- 030 "Approximate Solutions to Dynamic Models – Linear Methods" by Harald Uhlig, April 2006.
- 031 "Exploratory Graphics of a Financial Dataset" by Antony Unwin, Martin Theus and Wolfgang Härdle, April 2006.
- 032 "When did the 2001 recession *really* start?" by Jörg Polzehl, Vladimir Spokoiny and Cătălin Stărică, April 2006.
- 033 "Varying coefficient GARCH versus local constant volatility modeling. Comparison of the predictive power" by Jörg Polzehl and Vladimir Spokoiny, April 2006.
- 034 "Spectral calibration of exponential Lévy Models [1]" by Denis Belomestny and Markus Reiß, April 2006.
- 035 "Spectral calibration of exponential Lévy Models [2]" by Denis Belomestny and Markus Reiß, April 2006.
- 036 "Spatial aggregation of local likelihood estimates with applications to classification" by Denis Belomestny and Vladimir Spokoiny, April 2006.
- 037 "A jump-diffusion Libor model and its robust calibration" by Denis Belomestny and John Schoenmakers, April 2006.
- 038 "Adaptive Simulation Algorithms for Pricing American and Bermudan Options by Local Analysis of Financial Market" by Denis Belomestny and Grigori N. Milstein, April 2006.
- 039 "Macroeconomic Integration in Asia Pacific: Common Stochastic Trends and Business Cycle Coherence" by Enzo Weber, May 2006.
- 040 "In Search of Non-Gaussian Components of a High-Dimensional Distribution" by Gilles Blanchard, Motoaki Kawanabe, Masashi Sugiyama, Vladimir Spokoiny and Klaus-Robert Müller, May 2006.
- 041 "Forward and reverse representations for Markov chains" by Grigori N. Milstein, John G. M. Schoenmakers and Vladimir Spokoiny, May 2006.
- 042 "Discussion of 'The Source of Historical Economic Fluctuations: An Analysis using Long-Run Restrictions' by Neville Francis and Valerie A. Ramey" by Harald Uhlig, May 2006.
- 043 "An Iteration Procedure for Solving Integral Equations Related to Optimal Stopping Problems" by Denis Belomestny and Pavel V. Gapeev, May 2006.
- 044 "East Germany's Wage Gap: A non-parametric decomposition based on establishment characteristics" by Bernd Görzig, Martin Gornig and Axel Werwatz, May 2006.
- 045 "Firm Specific Wage Spread in Germany - Decomposition of regional differences in inter firm wage dispersion" by Bernd Görzig, Martin Gornig and Axel Werwatz, May 2006.

SFB 649, Spandauer Straße 1, D-10178 Berlin
<http://sfb649.wiwi.hu-berlin.de>

This research was supported by the Deutsche
Forschungsgemeinschaft through the SFB 649 "Economic Risk".



- 046 "Produktdiversifizierung: Haben die ostdeutschen Unternehmen den Anschluss an den Westen geschafft? – Eine vergleichende Analyse mit Mikrodaten der amtlichen Statistik" by Bernd Görzig, Martin Gornig and Axel Werwatz, May 2006.
- 047 "The Division of Ownership in New Ventures" by Dominique Demougin and Oliver Fabel, June 2006.
- 048 "The Anglo-German Industrial Productivity Paradox, 1895-1938: A Restatement and a Possible Resolution" by Albrecht Ritschl, May 2006.
- 049 "The Influence of Information Costs on the Integration of Financial Markets: Northern Europe, 1350-1560" by Oliver Volckart, May 2006.
- 050 "Robust Econometrics" by Pavel Čížek and Wolfgang Härdle, June 2006.
- 051 "Regression methods in pricing American and Bermudan options using consumption processes" by Denis Belomestny, Grigori N. Milstein and Vladimir Spokoiny, July 2006.
- 052 "Forecasting the Term Structure of Variance Swaps" by Kai Detlefsen and Wolfgang Härdle, July 2006.
- 053 "Governance: Who Controls Matters" by Bruno Deffains and Dominique Demougin, July 2006.
- 054 "On the Coexistence of Banks and Markets" by Hans Gersbach and Harald Uhlig, August 2006.
- 055 "Reassessing Intergenerational Mobility in Germany and the United States: The Impact of Differences in Lifecycle Earnings Patterns" by Thorsten Vogel, September 2006.
- 056 "The Euro and the Transatlantic Capital Market Leadership: A Recursive Cointegration Analysis" by Enzo Weber, September 2006.
- 057 "Discounted Optimal Stopping for Maxima in Diffusion Models with Finite Horizon" by Pavel V. Gapeev, September 2006.
- 058 "Perpetual Barrier Options in Jump-Diffusion Models" by Pavel V. Gapeev, September 2006.
- 059 "Discounted Optimal Stopping for Maxima of some Jump-Diffusion Processes" by Pavel V. Gapeev, September 2006.
- 060 "On Maximal Inequalities for some Jump Processes" by Pavel V. Gapeev, September 2006.
- 061 "A Control Approach to Robust Utility Maximization with Logarithmic Utility and Time-Consistent Penalties" by Daniel Hernández-Hernández and Alexander Schied, September 2006.
- 062 "On the Difficulty to Design Arabic E-learning System in Statistics" by Taleb Ahmad, Wolfgang Härdle and Julius Mungo, September 2006.
- 063 "Robust Optimization of Consumption with Random Endowment" by Wiebke Wittmüß, September 2006.
- 064 "Common and Uncommon Sources of Growth in Asia Pacific" by Enzo Weber, September 2006.
- 065 "Forecasting Euro-Area Variables with German Pre-EMU Data" by Ralf Brüggemann, Helmut Lütkepohl and Massimiliano Marcellino, September 2006.
- 066 "Pension Systems and the Allocation of Macroeconomic Risk" by Lans Bovenberg and Harald Uhlig, September 2006.
- 067 "Testing for the Cointegrating Rank of a VAR Process with Level Shift and Trend Break" by Carsten Trenkler, Pentti Saikkonen and Helmut Lütkepohl, September 2006.
- 068 "Integral Options in Models with Jumps" by Pavel V. Gapeev, September 2006.
- 069 "Constrained General Regression in Pseudo-Sobolev Spaces with Application to Option Pricing" by Zdeněk Hlávka and Michal Pešta, September 2006.

SFB 649, Spandauer Straße 1, D-10178 Berlin
<http://sfb649.wiwi.hu-berlin.de>

This research was supported by the Deutsche
 Forschungsgemeinschaft through the SFB 649 "Economic Risk".



- 070 "The Welfare Enhancing Effects of a Selfish Government in the Presence of Uninsurable, Idiosyncratic Risk" by R. Anton Braun and Harald Uhlig, September 2006.
- 071 "Color Harmonization in Car Manufacturing Process" by Anton Andriyashin, Michal Benko, Wolfgang Härdle, Roman Timofeev and Uwe Ziegenhagen, October 2006.
- 072 "Optimal Interest Rate Stabilization in a Basic Sticky-Price Model" by Matthias Paustian and Christian Stoltenberg, October 2006.
- 073 "Real Balance Effects, Timing and Equilibrium Determination" by Christian Stoltenberg, October 2006.
- 074 "Multiple Disorder Problems for Wiener and Compound Poisson Processes With Exponential Jumps" by Pavel V. Gapeev, October 2006.
- 075 "Inhomogeneous Dependency Modelling with Time Varying Copulae" by Enzo Giacomini, Wolfgang K. Härdle, Ekaterina Ignatieva and Vladimir Spokoiny, November 2006.
- 076 "Convenience Yields for CO₂ Emission Allowance Futures Contracts" by Szymon Borak, Wolfgang Härdle, Stefan Trück and Rafal Weron, November 2006.
- 077 "Estimation of Default Probabilities with Support Vector Machines" by Shiyi Chen, Wolfgang Härdle and Rouslan Moro, November 2006.
- 078 "GHICA - Risk Analysis with GH Distributions and Independent Components" by Ying Chen, Wolfgang Härdle and Vladimir Spokoiny, November 2006.
- 079 "Do Individuals Recognize Cascade Behavior of Others? - An Experimental Study -" by Tim Grebe, Julia Schmid and Andreas Stiehler, November 2006.
- 080 "The Uniqueness of Extremum Estimation" by Volker Krätschmer, December 2006.
- 081 "Compactness in Spaces of Inner Regular Measures and a General Portmanteau Lemma" by Volker Krätschmer, December 2006.
- 082 "Probleme der Validierung mit Strukturgleichungsmodellen" by Lutz Hildebrandt and Dirk Temme, December 2006.

SFB 649, Spandauer Straße 1, D-10178 Berlin
<http://sfb649.wiwi.hu-berlin.de>

This research was supported by the Deutsche
Forschungsgemeinschaft through the SFB 649 "Economic Risk".

