

SFB 649 Discussion Paper 2008-042

# Gruppenvergleiche bei hypothetischen Konstrukten – Die Prüfung der Übereinstimmung von Messmodellen mit der Strukturgleichungsmethodik

Dirk Temme\*  
Lutz Hildebrandt\*



\* 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

# **Gruppenvergleiche bei hypothetischen Konstrukten – Die Prüfung der Übereinstimmung von Messmodellen mit der Strukturgleichungsmethodik<sup>1</sup>**

**– Eine Untersuchung am Beispiel der Markenforschung –**

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

## **Abstract**

Comparing groups with respect to hypothetical constructs requires that the measurement models are equal across groups. Otherwise conclusions drawn from the observed indicators regarding differences at the latent level (mean differences, differences in the structural relations) might be severely distorted. This article provides a state of the art on how to apply multi-group confirmatory factor analysis to assess measurement invariance. The required steps in the analysis of the observed indicator means and variances/covariances are described, placing special emphasis on how to identify noninvariant indicators. The procedure is demonstrated considering the construct brand strength (“Brand Potential Index”, BPI<sup>®</sup>) introduced by GfK Market Research as an example.

**Keywords:** Measurement invariance; Partial metric invariance; Multi-group confirmatory factor analysis; Brand strength

**JEL-Codes:** C31, C51, C81, M31

---

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

# **Gruppenvergleiche bei hypothetischen Konstrukten – Die Prüfung der Übereinstimmung von Messmodellen mit der Strukturgleichungsmethodik – Eine Untersuchung am Beispiel der Markenforschung –**

## **1 Einleitung**

Ein sinnvoller Vergleich von Merkmalen unterschiedlicher Gruppen (z. B. Länder, Kundensegmente) setzt einen gemeinsamen Maßstab voraus. Leicht zu erhebende Eigenschaften wie etwa Mengen oder monetäre Größen werfen hier relativ geringe Probleme auf. Schwieriger wird die Situation allerdings, wenn der Vergleich sich auf hypothetische Konstrukte bezieht, die i. d. R. in Befragungen durch Multi-Item-Skalen gemessen werden. Hier ist gefordert, dass die Messbeziehungen zwischen den beobachteten Indikatoren und ihren zugrunde liegenden latenten Variablen in den Gruppen gleich sind. Nur so ist sichergestellt, dass z. B. bei gleicher Ausprägung einer latenten Variablen in allen Gruppen auch für die einzelnen Indikatoren die jeweils gleichen Werte in den Gruppen zu erwarten sind. Gruppenspezifische Eigenschaften (z. B. der kulturelle Hintergrund) können aber die Reaktion der Befragten auf Fragebogenstatements und damit die Messbeziehungen so beeinflussen, dass die beobachteten Indikatoren die „wahren“ Gruppendifferenzen auf der Konstruktebene, insbesondere Mittelwertdifferenzen und Unterschiede in den strukturellen Beziehungen zwischen den latenten Variablen, nur verzerrt wiedergeben. Derartige Verzerrungen bleiben aber i. d. R. unentdeckt, wenn die Messäquivalenz der Skalen in den Gruppen nicht explizit mit einem geeigneten Ansatz überprüft wird. Erforderlich ist hierfür die simultane Schätzung eines Faktormodells für alle Gruppen, wobei sowohl die Varianzen und Kovarianzen als auch die Mittelwerte der Indikatoren einzubeziehen sind.

In diesem Beitrag wird gezeigt, welche Konsequenzen aus einer fehlenden Messäquivalenz resultieren können und wie Messäquivalenz im Einzelnen geprüft werden sollte. Basis für diesen State-of-the-Art ist eine detaillierte Auswertung der umfangreichen methodischen Literatur zur Messinvarianzproblematik. Besonderes Augenmerk wird dabei dem Fall gewidmet, dass sich die Hypothese gleicher Messbeziehungen in den Gruppen nicht bestätigt. Um die hierfür verantwortlichen Indikatoren eines Konstrukts zu identifizieren, werden zum Teil Ansätze vorgeschlagen und von den Anwendern aufgegriffen, die zu Fehlschlüssen

führen können<sup>2</sup>. Vorgestellt wird deshalb ein Verfahren, mit dem diese Gefahr vermieden werden kann.

Der Beitrag gliedert sich im Weiteren wie folgt: Der *Abschnitt 2* soll den Leser zunächst durch eine konkrete Illustration der möglichen Konsequenzen fehlender Messäquivalenz für die Problematik sensibilisieren. Der *Abschnitt 3* stellt dann einen umfassenden Ansatz zur mehrstufigen Überprüfung der Messäquivalenz mit Hilfe der Mehrgruppenanalyse konfirmatorischer Faktormodelle vor, der auch die Mittelwertstruktur der Indikatoren einbezieht. Im *Abschnitt 4* wird das Testverfahren auf eine empirische Studie zur Messäquivalenz ausgewählter Indikatoren der Markenstärke bei unterschiedlich loyalen Konsumenten angewendet und verdeutlicht, welche Implikationen die festgestellten Nichtinvarianzen haben. Der Beitrag endet mit einer Zusammenfassung sowie der Ableitung weiteren Forschungsbedarfs.

## **2 Die Nichtäquivalenz von Messmodellen und ihre Auswirkungen**

Mit Fragen der Generalisierbarkeit von Konstrukten sowie der Vergleichbarkeit von Messungen hat sich insbesondere die interkulturelle Forschung beschäftigt<sup>3</sup>. Konstrukte, die typischerweise in einem bestimmten Kulturkreis eingeführt werden, sollen auch über kulturelle Grenzen hinweg die gleiche Interpretation erfahren und durch die gleichen Indikatoren operationalisiert werden. Ist dies gewährleistet, so ist darüber hinaus zu klären, ob die nur anhand von Indikatoren zu beobachtenden Unterschiede zwischen Ländern oder Kulturen die tatsächlichen Unterschiede im Hinblick auf die interessierenden Konstrukte korrekt wiedergeben. Messäquivalenz spielt aber nicht nur im interkulturellen Vergleich eine Rolle, sondern ist immer dann von Relevanz, wenn unterschiedliche Kontextbedingungen (z. B. Alter, Geschlecht, Zeitraum, Erhebungsform, Beurteiler, Experimentalsituation) zugrunde liegen, die potenziell die Beziehungen zwischen den latenten Variablen und ihren Indikatoren moderieren können<sup>4</sup>. Da hypothetische Konstrukte nicht direkt beobachtbar sind, muss die Messäquivalenz bzw. -invarianz<sup>5</sup> unter Zugrundelegung eines bestimmten Modelltyps für die

---

<sup>2</sup> Vgl. z. B. *Krafft/Litfin* 2002; *Steenkamp/Baumgartner* (1998).

<sup>3</sup> Z. B. *van de Vijver/Leung* (1997). Beispielhaft seien etwa die zahlreichen Studien genannt, die die Übertragbarkeit der in den USA entwickelten CETSCALE (*Shimp/Sharma* 1987) zur Messung der Ethnozentrierung von Konsumenten auf andere Länder untersucht haben (z. B. *Douglas/Nijssen* 2003; *Netemeyer/Durvasula/Lichtenstein* 1991; *Steenkamp/Baumgartner* 1998).

<sup>4</sup> So weist z. B. *Yoo* (2002) für US-amerikanische Konsumenten nach, dass für die aus einer Kombination von Geschlecht und Altersklassen gebildeten Gruppen bestimmte Indikatoren der CETSCALE nicht die für einen Mittelwertvergleich erforderliche Äquivalenz besitzen.

<sup>5</sup> Die Begriffe Messäquivalenz und Messinvarianz werden im Folgenden synonym verwendet.

Beziehungen zwischen den latenten Variablen und ihren Indikatoren überprüft werden. Als wichtigste Methodik hat sich hier die Mehrgruppenanalyse konfirmatorischer Faktormodelle etabliert<sup>6</sup> – sie soll daher auch im Fokus dieses Beitrags stehen.

Die Konsequenzen, die sich aus einer fehlenden Messäquivalenz ergeben können, werden nachfolgend anhand zweier Fragestellungen erörtert: (1) In welchem Ausmaß unterscheiden sich die Gruppen in der durchschnittlichen Ausprägung eines latenten Merkmals? (2) Inwieweit differieren die Gruppen in den Strukturbeziehungen der latenten Variablen?

### *2.1 Die Auswirkungen fehlender Messäquivalenz auf Mittelwertvergleiche*

Die möglichen Auswirkungen eines nichtäquivalenten Messmodells verdeutlicht *Abbildung 1*. Die Grafik gibt die Messrelationen zwischen einer latenten Variablen  $\eta$  und einem ihrer Indikatoren  $y$  in den drei Gruppen A, B, und C inklusive der Mittelwertunterschiede wieder. Unterstellt wird, dass es sich bei dem Indikator um ein positives Fragebogenstatement handelt, für das die Befragten den Grad ihrer Zustimmung auf einer Ratingskala mit 5 Antwortkategorien (z. B. 1 = „stimme überhaupt nicht zu“ bis 5 = „stimme voll und ganz zu“) angeben können<sup>7</sup>. Die konfirmatorische Faktorenanalyse erklärt die beobachteten Varianzen/Kovarianzen und Mittelwerte über einen linearen Zusammenhang zwischen den latenten Variablen und ihren Indikatoren. Dieser wird – analog zu den Absolutgliedern und Steigungsparametern in der Regressionsanalyse – durch Konstanten ( $\nu$ ) und Faktorladungen ( $\lambda$ ) beschrieben. Als Faktormodell für den Indikator  $y$  einschließlich seines Mittelwerts  $\mu$  resultiert damit:

$$y = \nu + \lambda\eta + \varepsilon, \quad (1)$$

$$\mu = \nu + \lambda\alpha, \quad (2)$$

wobei  $\varepsilon$  den zufälligen Messfehler und  $\alpha$  den latenten Mittelwert repräsentieren. Zu Illustrationszwecken wird angenommen, dass die latente Variable in allen drei Gruppen den gleichen Mittelwert aufweist. Folglich müssten bei Messinvarianz die beobachteten Mittelwerte des Indikators in den Gruppen (d. h.  $\mu_A$ ,  $\mu_B$ , und  $\mu_C$ ) ebenfalls gleich sein. Wie zu erkennen ist, unterscheiden sich diese Mittelwerte aber aufgrund gruppenspezifischer Konstanten und/oder Faktorladungen erheblich voneinander.

---

<sup>6</sup> Vgl. Meade/Lautenschlager (2004a); Steenkamp/Baumgartner (1998).

<sup>7</sup> Auf die mögliche Problematik eines ordinalen Skalenniveaus soll an dieser Stelle nicht eingegangen werden. Zum Einsatz eines Faktormodells für ordinale Daten bei der Messinvarianzüberprüfung siehe z. B. Temme (2006a) sowie den *Abschnitt 3.1*.

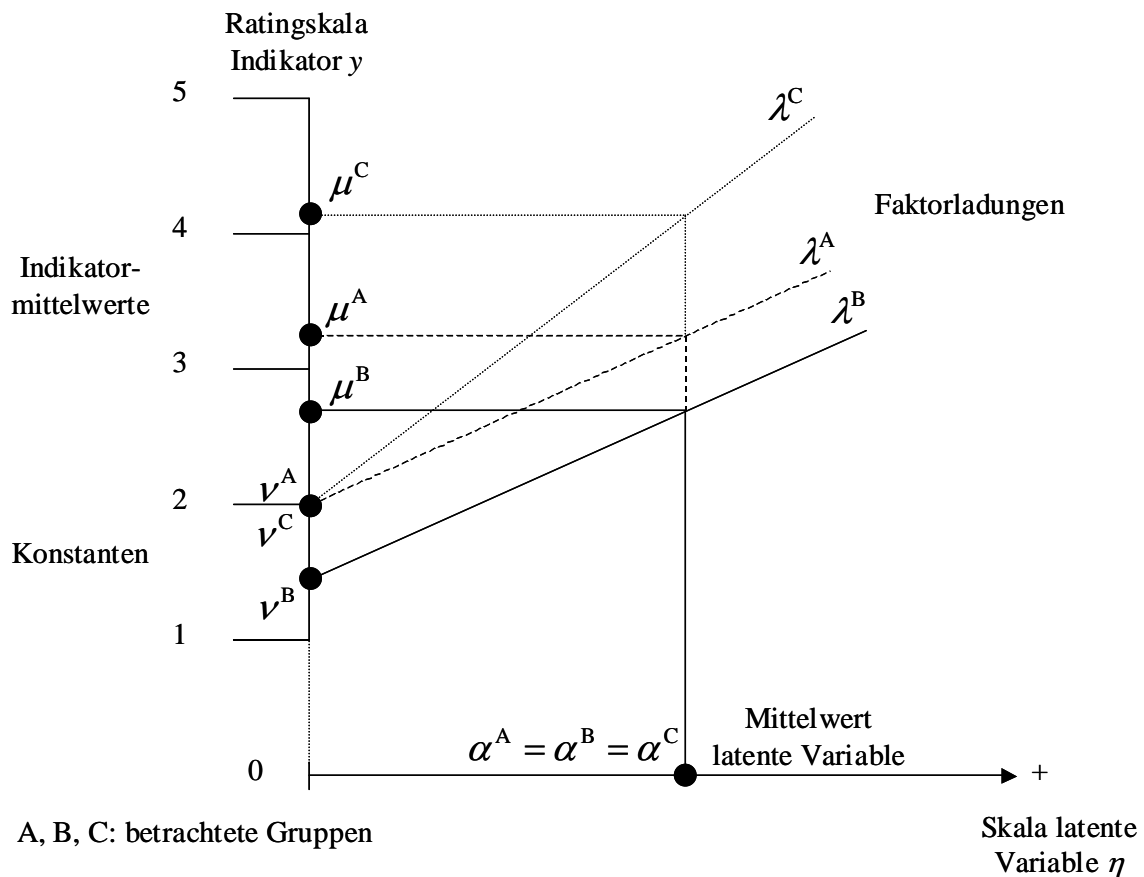


Abbildung 1: Auswirkungen fehlender Messinvarianz auf die beobachteten Mittelwerte

Nimmt man die Gruppe A als Ausgangspunkt, so wird deutlich, dass die Faktorladung (d. h. die Steigung der Geraden) in der Gruppe B zwar gleich ist, die Konstante (d. h. der Schnittpunkt der Geraden mit der Ratingskala) aber geringer ausfällt. Inhaltlich bedeutet dies, dass sich bei *gleicher* Ausprägung der latenten Variablen die Befragten in der Gruppe B immer weniger positiv in Bezug auf das verwendete Statement äußern als die Befragten in der Gruppe A. Bei Mittelwertvergleichen zwischen den beiden Gruppen A und B kommt es somit zu einer konstanten Verzerrung der Differenz.

Im Gegensatz zur Gruppe B weist die Messbeziehung in der Gruppe C die gleiche Konstante wie in Gruppe A auf; dafür ist jedoch in der Gruppe C die Faktorladung größer. Erhöht sich der Wert der latenten Variablen um einen bestimmten Betrag, so steigt der beobachtete Indikatorwert in der Gruppe C stärker an als in der Gruppe A (d. h. es kommt in der Gruppe C zu einem stärkeren Anstieg der Zustimmung). Mit zunehmendem Abstand vom Nullpunkt der latenten Skala wird daher die Differenz in den beobachteten Mittelwerten (bei gleichen Mittelwerten für die latente Variable) immer größer. Zusätzlich zu den in *Abbildung 1*

dargestellten Fällen fehlender Messäquivalenz sind auch Kombinationen aus differierenden Konstanten und Faktorladungen möglich, wobei sich z. B. für den Vergleich der beiden Gruppen A und D im positiven Wertebereich der latenten Skala schneidende (z. B.  $\nu^D > \nu^A$  und  $\lambda^D < \lambda^A$ ) und nicht-schneidende Geraden (z. B.  $\nu^D > \nu^A$  und  $\lambda^D > \lambda^A$ ) ergeben können.

Bisher wurde nur ein einzelner Indikator hinsichtlich der Auswirkungen fehlender Messinvarianz auf Mittelwertvergleiche betrachtet. In der Praxis erfolgt ein Gruppenvergleich dagegen häufig auf Basis eines Durchschnitts der Indikatormittelwerte (oder alternativ ihrer Summe). Bei einem solchen Index können sich Unterschiede in den Messbeziehungen der verschiedenen Indikatoren noch gegenseitig verstärken (z. B. wenn noch weitere Indikatoren in der Gruppe B eine geringere Konstante als in der Gruppe A aufweisen), im günstigsten Fall aber auch nivellieren (z. B. wenn in der Gruppe B sowohl Indikatoren mit geringerer als auch mit höherer Konstante als in der Gruppe A auftreten). Ohne einen expliziten Test auf Gleichheit der Konstanten und Faktorladungen in den Gruppen können hierüber jedoch keine Aussagen gemacht werden.

## 2.2 Die Auswirkung fehlender Messäquivalenz auf den Vergleich von Strukturbeziehungen

Verletzungen der Messäquivalenzbedingung können nicht nur Vergleiche von Mittelwerten, sondern auch von Strukturbeziehungen verzerren. Dieses Problem illustriert die *Abbildung 2* am Beispiel zweier latenter Variablen  $\xi$  und  $\eta$ , die zur Vereinfachung nur durch jeweils einen Indikator –  $x$  bzw.  $y$  – gemessen werden. Da die absolute Höhe der Mittelwerte für die Kovarianzen zwischen Variablen ohne Bedeutung ist, wird unterstellt, dass sowohl die beobachteten als auch die latenten Variablen als Abweichung von ihrem jeweiligen gruppenspezifischen Mittelwert gemessen werden. Differenzen in den Konstanten haben damit keinen Einfluss auf mögliche Unterschiede in den linearen strukturellen Beziehungen<sup>8</sup>, so dass nur die Faktorladungen berücksichtigt werden müssen. Angenommen wird, dass die Stärke des Einflusses der exogenen Variable  $\xi$  auf die endogene Variable  $\eta$ , also der Regressionsparameter  $\beta_{\eta\xi}$ , in den beiden Gruppen A und B gleich ist<sup>9</sup>. In diesem Fall sollten die in den beiden Gruppen zu beobachtenden Regressionsbeziehungen zwischen den Indikatoren  $x$  und  $y$  (d. h.  $\beta_{yx}^A$  und  $\beta_{yx}^B$ ) die Übereinstimmung auf der latenten Ebene korrekt

---

<sup>8</sup> Die Konstanten werden durch die Zentrierung der Variablen eliminiert.

<sup>9</sup> Regressionskonstanten brauchen aufgrund der Variablenzentrierung ebenfalls nicht betrachtet zu werden, so dass alle Geraden im Ursprung des Koordinatensystems beginnen.

widerspiegeln. Wie die beiden unterschiedlichen Regressionsgeraden im rechten oberen Quadranten erkennen lassen, ist diese Forderung im vorliegenden Beispiel aber nicht erfüllt.

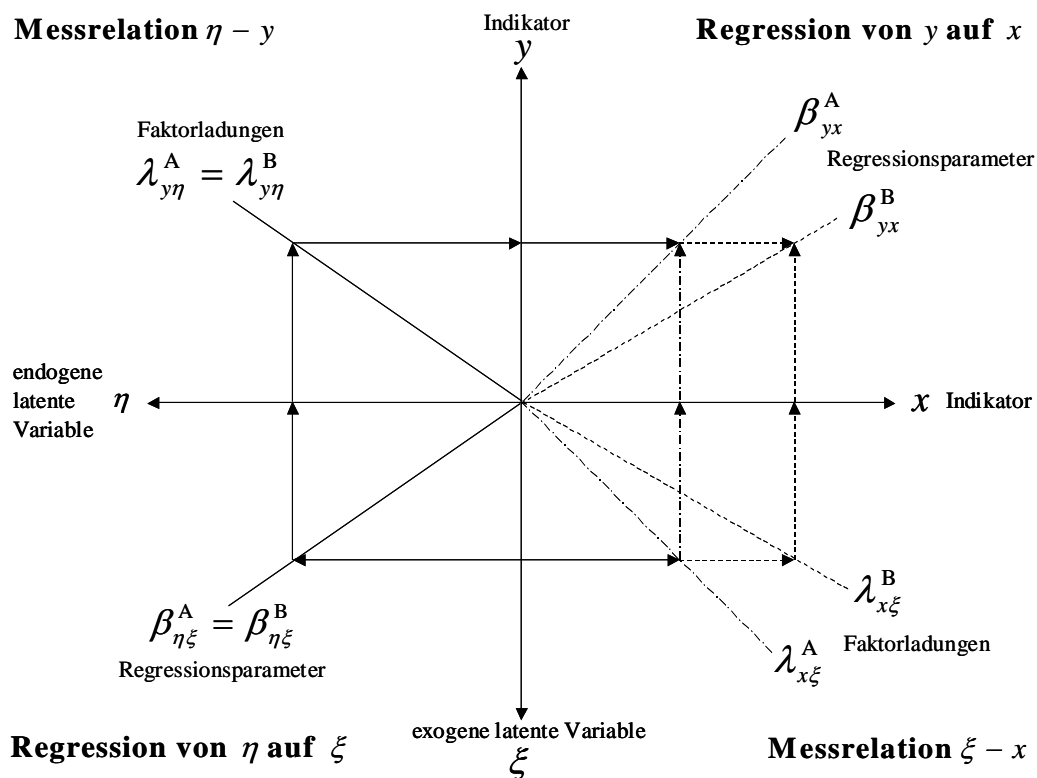


Abbildung 2: Auswirkung fehlender Messinvarianz von Faktorladungen auf die geschätzten Beziehungen zwischen latenten Variablen im Ein-Indikator-Fall

Worauf lassen sich diese Unterschiede nun zurückführen? Aufgrund des in den Gruppen A und B identischen linearen Zusammenhangs zwischen  $\xi$  und  $\eta$  wird ein bestimmter  $\xi$ -Wert jeweils in den gleichen  $\eta$ -Wert  $\beta_{\eta\xi} \cdot \xi$  transformiert (siehe die Pfeile im linken unteren Quadranten in *Abbildung 2*). Da für die Messbeziehung zwischen  $\eta$  und  $y$  von gleichen Faktorladungen in beiden Gruppen ausgegangen wird, resultiert für jeden  $\eta$ -Wert jeweils der gleiche  $y$ -Wert  $\lambda_{y\eta} \cdot \eta$  in den Gruppen (siehe die Pfeile im linken oberen Quadranten). Im Gegensatz dazu lädt der Indikator  $x$  in der Gruppe B höher auf den Faktor  $\xi$  als in der Gruppe A (d. h.  $\lambda_{x\xi}^B > \lambda_{x\xi}^A$ ). Dementsprechend wird ein bestimmter  $\xi$ -Wert in der Gruppe B in einen  $x$ -Wert  $\lambda_{x\xi}^B \cdot \xi$  transformiert, der höher ist als der entsprechende  $x$ -Wert  $\lambda_{x\xi}^A \cdot \xi$  in Gruppe A (siehe die Pfeile im rechten unteren Quadranten). Da diesen unterschiedlichen  $x$ -Werten aber in beiden Gruppen die gleichen  $y$ -Werte gegenüberstehen (siehe die Pfeile im rechten oberen Quadranten), resultiert in der Gruppe B eine gegenüber der Gruppe A geringere Steigung der beobachteten Regressionsgeraden (d. h.  $\beta_{yx}^B < \beta_{yx}^A$ ). Im Gegensatz zur tatsächlichen



Beziehung auf der latenten Ebene würde somit davon ausgegangen, dass die Variable  $\xi$  in der Gruppe A einen stärkeren Effekt auf die Variable  $\eta$  hat als in der Gruppe B.

Bei der Verwendung von Multi-Item-Skalen zur Messung der Konstrukte ist die Wirkung nichtinvarianter Faktorladungen z. B. davon abhängig, ob für die Skalierung der latenten Variablen (durch Fixierung der Faktorladung eines sogenannten Referenzindikators in allen Gruppen auf den Wert 1) messäquivalente Indikatoren verwendet werden oder nicht; dabei können sich sowohl Verstärkungs- als auch Nivellierungseffekte ergeben. Weist z. B. im Fall einer Regression von  $\eta$  auf  $\xi$  der Referenzindikator für den Faktor  $\xi$  in der Gruppe B eine höhere Faktorladung als in der Gruppe A auf, so führt dies – analog zum vorangehenden Beispiel in *Abbildung 2* – im Vergleich zur Gruppe A zu einer Unterschätzung der Regressionsbeziehung in der Gruppe B. Diese Verzerrung wird noch verstärkt, wenn gleichzeitig die Ladung des Referenzindikators auf den Faktor  $\eta$  in der Gruppe B niedriger als in der Gruppe A ausfällt. Einen solchen Fall unterstellt das in *Abbildung 3* dargestellte Strukturgleichungsmodell.

In diesem Modell wird für den Einfluss der Variable  $\xi$  auf die Variable  $\eta$  ein in beiden Gruppen übereinstimmendes Regressionsgewicht in Höhe von 0,5 unterstellt. Schaut man sich dagegen die Faktorladungen an, so wird deutlich, dass diese sowohl für die Indikatoren  $x_1$ ,  $x_2$  und  $x_3$  als auch die Indikatoren  $y_1$ ,  $y_2$  und  $y_3$  unterschiedlich sind: In Gruppe B weisen die drei  $x$ -Indikatoren höhere, die drei  $y$ -Indikatoren dagegen niedrigere Faktorladungen als in Gruppe A auf. Für die Indikatoren  $x_4$ ,  $x_5$  und  $x_6$  sowie  $y_4$  und  $y_5$  wird von gleichen Faktorladungen ausgegangen. Da das Beispiel möglichst einfach gehalten werden soll, wird auch für die übrigen Parameter des Modells (Varianz für  $\xi$ , Störgrößenvarianz für  $\eta$ , Messfehlervarianzen der Indikatoren) Gleichheit über die Gruppen angenommen. Um den verzerrenden Einfluss der nichtinvarianten Faktorladungen auf die Schätzung der Regressionsbeziehung zwischen den beiden latenten Variablen  $\xi$  und  $\eta$  zu quantifizieren, werden auf Basis der Parameter in *Abbildung 3* die modellimplizierten gruppenspezifischen Populations-Kovarianzmatrizen abgeleitet<sup>10</sup>. Diese dienen als Dateninput für separate „Schätzungen“ des Strukturgleichungsmodells für die Gruppen A und B.

---

<sup>10</sup> Um zu eindeutigen Aussagen über die Verzerrung kommen zu können, werden hier bewusst keine Stichprobendaten simuliert, sondern die durch das Modell und seine Parameterwerte determinierten Varianzen und Kovarianzen in der Grundgesamtheit ermittelt (vgl. zur Berechnung *Bollen* 1989, S. 325; für den Indikator  $x_1$  ergibt sich hierdurch z. B. eine Varianz in Höhe von  $0,8^2 * 1,0 + 0,2 = 0,84$ ). Die Ergebnisse unterliegen somit keinem Stichprobenfehler.

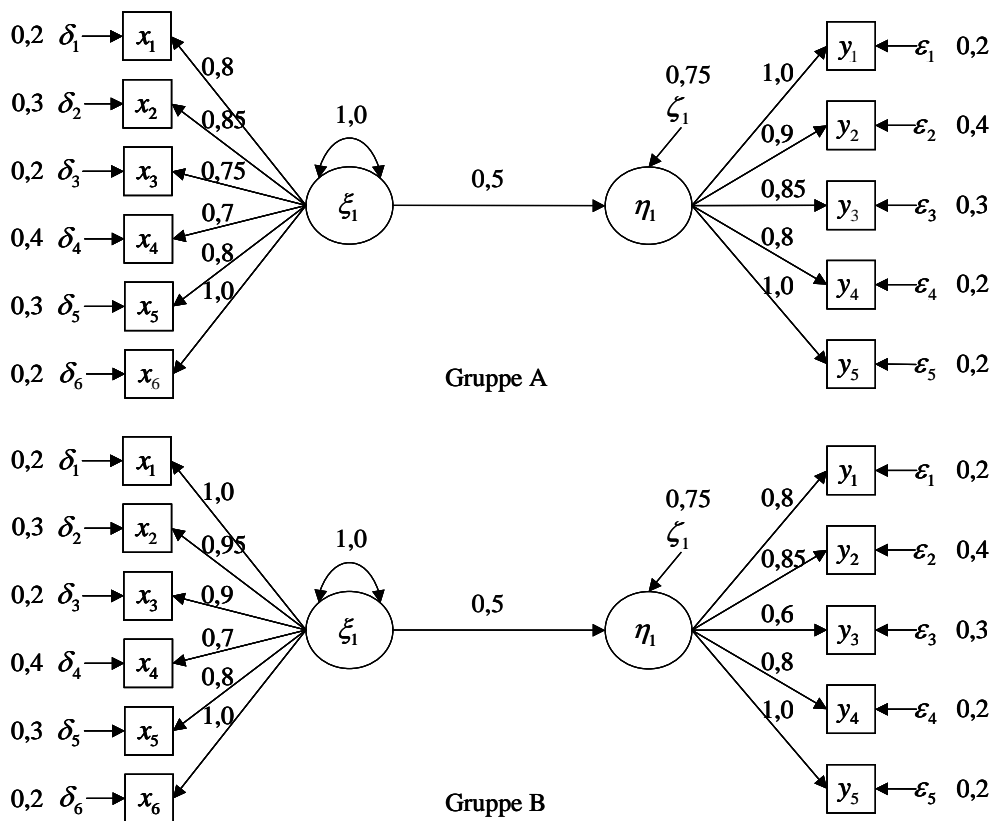


Abbildung 3: Auswirkung fehlender Messinvarianz von Faktorladungen auf die geschätzten Beziehungen zwischen latenten Variablen im Fall multipler Indikatoren – Strukturgleichungsmodell zur Erzeugung der Populationsdaten

Für die Skalierung der beiden latenten Variablen werden die nichtinvarianten Indikatoren  $x_1$  bzw.  $y_1$  als Referenzindikatoren ausgewählt (d. h. ihre Faktorladungen werden in beiden Gruppen auf den Wert 1 fixiert) – da in empirischen Studien ohne explizite Überprüfung der Messinvarianz a priori nicht bekannt ist, welche Indikatoren äquivalent sind oder nicht, ist dies ein durchaus realistischer Fall. Aufgrund der Verwendung von Populationsdaten führen die beiden „Schätzungen“ des Strukturgleichungsmodells jeweils zu einem perfekten Fit. Vergleicht man die geschätzten Modellparameter für die Gruppen A und B (vgl. *Abbildung 4*), so fallen zuerst die deutlichen Unterschiede in den Regressionskoeffizienten auf ( $\gamma_{11}^A = 0,625$  versus  $\gamma_{11}^B = 0,4$ ). Aufgrund dieses Ergebnisses liegt somit der (Fehl-)Schluss nahe, dass die Variable  $\xi$  in der Gruppe A einen deutlich stärkeren Einfluss auf die Variable  $\eta$  besitzt als in der Gruppe B.

Interessant ist darüber hinaus die Tatsache, dass für die *nichtinvarianten* Indikatoren  $x_1$  bis  $x_3$  sowie  $y_1$  bis  $y_3$  die Faktorladungen in den beiden Gruppen ähnlicher ausfallen als diejenigen

für die eigentlich *messäquivalenten* Indikatoren  $x_4$ ,  $x_5$  und  $x_6$  sowie  $y_4$  und  $y_5$ . Auf Basis eines Vergleichs der separat geschätzten Strukturgleichungsmodelle kommt es somit nicht nur zu einer Fehlbeurteilung bzgl. der Unterschiede in der Regressionsbeziehung auf der latenten Ebene, sondern es werden u. U. auch noch die falschen Indikatoren als messäquivalent bzw. nicht messäquivalent angesehen. Würden dagegen als invariant identifizierte Indikatoren (z. B.  $x_6$  und  $y_5$ ) als Referenzindikatoren verwendet<sup>11</sup>, so würde die Schätzung des Strukturgleichungsmodells in *Abbildung 3* in beiden Gruppen zum gleichen Regressionsparameter führen.

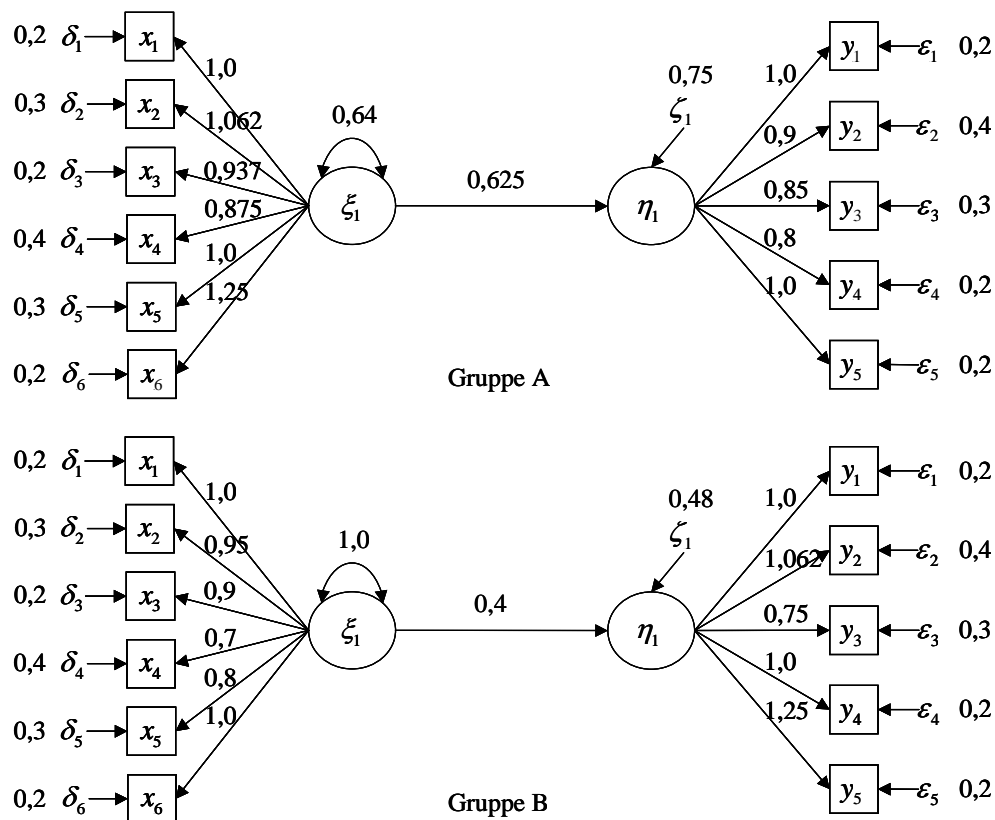


Abbildung 4: Auswirkung fehlender Messinvarianz von Faktorladungen auf die geschätzten Beziehungen zwischen latenten Variablen im Fall multipler Indikatoren – geschätztes Strukturgleichungsmodell mit nichtinvarianten Referenzindikatoren

### 2.3 Weitere Formen fehlender Messäquivalenz

Eine noch strengere Form von Messäquivalenz als bisher behandelt (gleiche Faktorladungen und gleiche Konstanten) stellt die Gleichheit der Messfehlervarianzen der Indikatoren in den einzelnen Gruppen dar. Eine Verletzung dieser Annahme kann zu einem dazu führen, dass

<sup>11</sup> Welche Probleme die Bestimmung invarianter Indikatoren aufwirft, wird in *Abschnitt 3.3* ausführlich behandelt.

die gleichen Indikatoren in den Gruppen unterschiedlich reliabel sind<sup>12</sup>. Zum anderen können hieraus bei Selektionsentscheidungen (z. B. der Auswahl von Kunden aus verschiedenen Segmenten) anhand der Ausprägung eines Konstrukts unterschiedliche Fehlerquoten (z. B. fälschlicherweise nicht ausgewählte Kunden) in den Gruppen resultieren<sup>13</sup>.

Bei der bisherigen Illustrierung der Messinvarianzproblematik wurde implizit unterstellt, dass die Indikatoren in den untersuchten Gruppen die gleichen Konzepte erfassen und die Messmodelle eine identische Struktur aufweisen. Laden aber z. B. bestimmte Indikatoren in einzelnen Gruppen nicht auf die ihnen zugeordnete latente Variable oder laden sie sogar auf andere Konstrukte, so liegen den Antworten der Befragten offensichtlich unterschiedliche Vorstellungen des Konzepts zugrunde. Damit ist ein Vergleich der Befunde aus verschiedenen Gruppen nicht möglich. Es empfiehlt sich daher für empirische Studien, zunächst diesen Aspekt der Messinvarianz zu überprüfen. *Tabelle 1* stellt im Überblick nochmals die zuvor erörterten Konsequenzen dar, die mit Verletzungen spezifischer Invarianzbedingungen im Rahmen konfirmatorischer Faktormodelle verbunden sein können.

*Tabelle 1: Verletzungen der Messäquivalenzbedingungen und ihre Konsequenzen*

<b>Unterschiede in den Faktormodellen</b>	<b>Verletzte Messäquivalenzbedingung</b>	<b>Mögliche Konsequenzen</b>
Indikatoren sind den Faktoren nicht bzw. unterschiedlich zugeordnet	konfigurale Invarianz	Statt gleicher werden unterschiedliche Konzepte verglichen
Messkonstanten unterschiedlich, Faktorladungen gleich	skalare Invarianz	Mittelwertunterschiede sind verzerrt
Messkonstanten gleich/unterschiedlich, Faktorladungen unterschiedlich	Metrische Invarianz	Mittelwertunterschiede sowie Unterschiede in den strukturellen Beziehungen sind verzerrt
Messfehlervarianzen unterschiedlich	Invarianz der Messfehlervarianzen	Reliabilität und Selektionsgenauigkeit sind unterschiedlich

<sup>12</sup> Für einen Vergleich der Indikatorreliabilitäten müssen neben den Differenzen in den Messfehlervarianzen auch die Unterschiede in der Varianz der zugehörigen latenten Variable berücksichtigt werden.

<sup>13</sup> Zur Beurteilung der Auswirkungen von Nichtinvarianz auf Selektionsprozesse vgl. *Millsap/Kwok* (2004).

### 3 Die Überprüfung der Äquivalenz von Messmodellen mit der Mittelwert- und Kovarianzstrukturanalyse

#### 3.1 Ansätze zur Überprüfung der Messäquivalenz

Die Mehrgruppenanalyse konfirmatorischer Faktormodelle stellt in den Sozialwissenschaften den dominierenden Ansatz zur Überprüfung der Messinvarianz dar. Daneben existiert mit der Item-Response-Theorie aber eine alternative Methode, deren Vor- und Nachteile im Vergleich zur konfirmatorischen Faktorenanalyse an dieser Stelle kurz diskutiert werden sollen.

Die Item-Response-Theorie (IRT) berücksichtigt im Gegensatz zum linearen Faktormodell explizit den ordinalen Charakter der in Fragebögen häufig verwendeten Ratingskalen<sup>14</sup>. Für jeden Indikator wird eine nichtlineare, monotone Funktion geschätzt, die die Wahrscheinlichkeit für die Wahl einer bestimmten Antwortkategorie in Abhängigkeit von der Ausprägung der latenten Variable angibt. Der Funktionsverlauf wird dabei durch zwei Arten von Parametern bestimmt, dem Diskriminierungsparameter<sup>15</sup> sowie den Lageparametern<sup>16</sup>. Ein gängiges Verfahren zur Überprüfung der Messinvarianz ist der Likelihood-Verhältnis-Test<sup>17</sup>. Für ein Basismodell, in dem bei allen Indikatoren die Diskriminierungs- und Lageparameter über die Gruppen gleichgesetzt sind, wird mittels ML-Schätzung die  $\chi^2$ -Teststatistik berechnet. Nachfolgend wird dann für jeden einzelnen Indikator separat überprüft, ob die simultane Aufhebung aller zugehörigen Gleichheitsrestriktionen zu einer signifikanten Verbesserung des Fit führt, d. h. keine Messäquivalenz vorliegt<sup>18</sup>.

Diejenigen Simulationsstudien, die die Validität der linearen Faktorenanalyse und des IRT-Ansatzes bei der Überprüfung der Messinvarianz vergleichen, kommen zu sehr unterschiedlichen Befunden<sup>19</sup>. Während etwa *Meade/Lautenschlager* (2004b) deutliche Vorteile für die IRT sehen, gelangen *Stark/Chernyshenko/Drasgow* (2006) zu dem Ergebnis,

---

<sup>14</sup> Wir beziehen uns hier auf das Graded Response-Modell von *Samejima* (1969). Für einen umfassenden Vergleich von linearer Faktorenanalyse und Item-Response-Theorie siehe z. B. *Singh* (2004). Zu einer vergleichenden Darstellung der Messäquivalenzüberprüfung mit den beiden Verfahren siehe *Raju/Laffitte/Byrne* (2002); *Reise/Widaman/Pugh* (1993).

<sup>15</sup> Dieser bestimmt die Steigung der Funktion und ist konzeptionell mit den Ladungen im linearen Faktormodell vergleichbar.

<sup>16</sup> Diese bestimmen die durchschnittliche Position eines Indikators auf der Skala der latenten Variable und geben damit Auskunft darüber, wie „leicht“ man einem Statement zustimmen kann. Hier gibt es bei aller Unterschiedlichkeit eine gewisse Analogie zu den Konstanten im linearen Faktormodell.

<sup>17</sup> Zu den verschiedenen Verfahren vgl. z. B. *Raju/Laffitte/Byrne* (2002), S. 522.

<sup>18</sup> Eine sehr vielversprechende Erweiterung des klassischen IRT-Ansatzes der Messinvarianzüberprüfung stellt das von *De Jong/Steenkamp/Fox* (2007) vorgeschlagene hierarchische IRT-Modell dar.

<sup>19</sup> *Flowers/Raju/Oshima* (2002); *González-Romá/Hernández/Gómez-Benito* (2006); *Meade/Lautenschlager* (2004b); *Stark/Chernyshenko/Drasgow* (2006).

dass beide Verfahren annähernd gleich valide sind; bei Stichprobengrößen von  $n = 500$  je Gruppe erweist sich das lineare Faktormodell sogar als vorteilhaft<sup>20</sup>. Ohne an dieser Stelle auf die vielfältigen Unterschiede im Design der Studien (z. B. Datengenerierung, Verteilung der Daten, Art und Ausmaß der Invarianz) eingehen zu können, lässt sich folgendes Fazit ziehen: Je stärker das datengenerierende Modell vom linearen Faktormodell abweicht, desto größere Vorteile zeigen sich i. d. R. für den IRT-Ansatz, da dieser die Nichtinvarianz von Indikatoren aufgrund der größeren Parameteranzahl in einer differenzierteren Weise erfassen kann<sup>21</sup>.

Vorteile für das lineare Faktormodell ergeben sich bei Untersuchungen mit mehreren hypothetischen Konstrukten. Da die gängigen IRT-Programme eindimensionale Konstrukte voraussetzen, müssten die Messmodelle jeweils einzeln auf ihre Äquivalenz überprüft werden<sup>22</sup>. Damit wird aber bei der Invarianzüberprüfung im Gegensatz zur Kovarianzstrukturanalyse die Korrelation der Faktoren nicht berücksichtigt. Ist man an der Schätzung von Konstruktbeziehungen interessiert, so müsste darüber hinaus wieder auf die Mehrgruppenanalyse konfirmatorischer Faktormodelle zurückgegriffen werden, bei der die eventuelle Nichtinvarianz bestimmter Parameter durch Freisetzung kontrolliert werden kann. Die theoretische Vorteilhaftigkeit des IRT-Ansatzes relativiert sich auch insofern, als Faktormodelle existieren, die ebenfalls dem ordinalen Charakter von Fragebogendaten Rechnung tragen<sup>23</sup>. Die Messinvarianzüberprüfung mit derartigen Modellen läuft dabei grundsätzlich analog zu der nachfolgend beschriebenen Vorgehensweise für das lineare Faktormodell ab, weist aber einige Besonderheiten auf<sup>24</sup>. Die Mehrgruppenanalyse mit ordinalen Faktormodellen kann mit weit verbreiteten SEM-Programmen wie z. B. Mplus<sup>25</sup> oder – mit gewissen Einschränkungen – LISREL durchgeführt werden<sup>26</sup>. Simulationsstudien, die die Validität des ordinalen Faktormodells bei der Messinvarianzüberprüfung mit dem IRT-Ansatz vergleichen, stehen derzeit noch aus<sup>27</sup>. Berücksichtigt man die weite Verbreitung der Strukturgleichungsmethodik, so steht hier ein Instrumentarium zur Verfügung, das auch

---

<sup>20</sup> In der Studie von *Stark/Chernyshenko/Drasgow* (2006) wurden Stichprobengrößen von  $n = 500$  und  $n = 1000$  untersucht.

<sup>21</sup> Dies kann z. B. von Bedeutung sein, wenn Gruppen in Bezug auf die Verwendung der Extremkategorien (z. B. 1 und 7 auf einer 7er-Ratingskala) miteinander verglichen werden sollen.

<sup>22</sup> Vgl. z. B. *González-Romá/Hernández/Gómez-Benito* (2006); *Stark/Chernyshenko/Drasgow* (2006).

<sup>23</sup> Vgl. *Flora/Curran* (2004); *Muthén* (1983; 1984).

<sup>24</sup> Vgl. *Millsap/Yun-Tein* (2004); *Muthén/Asparohou* (2002); Zu einer Diskussion divergierender Vorschläge zum Ablauf der Invarianzüberprüfung sowie einer Anwendung siehe *Temme* (2006a).

<sup>25</sup> *Muthén/Muthén* (1998-2007).

<sup>26</sup> Vgl. hierzu *Millsap/Yun-Tein* (2004).

<sup>27</sup> *Carle/Millsap/Cole* (2008) kommen bei einem empirischen Vergleich dieser beiden Ansätze zu weitgehend übereinstimmenden Befunden..

methodisch wenig vorgebildeten Anwendern eine Überprüfung der Messinvarianz erlaubt. Die dafür erforderlichen Schritte werden im Folgenden dargestellt.

### 3.2 Konfirmatorisches Faktormodell für den Mehrgruppenfall

Die Basis für die Überprüfung der Messäquivalenz bildet das Grundmodell der konfirmatorischen Faktorenanalyse für Kovarianz- und Mittelwertstrukturen im Mehrgruppenfall<sup>28</sup>. In diesem Modell stellt der beobachtete Wert  $y_{ijg}$  eines Individuums  $i$  ( $i = 1, K, I$ ) aus der Gruppe  $g$  ( $g = 1, K, G$ ) für den Indikator  $j$  ( $j = 1, K, J$ ) eine lineare Funktion der Faktorwerte  $\eta_{im}$  der  $M$  latenten Variablen ( $m = 1, K, M$ ) sowie der Residualgröße  $\varepsilon_{ijg}$  dar<sup>29</sup>:

$$y_{ijg} = \nu_{jg} + \sum_{m=1}^M \lambda_{jmg} \eta_{im} + \varepsilon_{ijg}. \quad (3)$$

Dabei bezeichnen  $\nu_{jg}$  die Konstante und  $\lambda_{jmg}$  die Ladung des Indikators  $y_j$  auf den Faktor  $\eta_m$ . In Matrixschreibweise lässt sich das Faktormodell kompakt formulieren als

$$\mathbf{y}_g = \boldsymbol{\nu}_g + \boldsymbol{\Lambda}_g \boldsymbol{\eta}_g + \boldsymbol{\varepsilon}_g. \quad (4)$$

Unter den Standardannahmen der konfirmatorischen Faktorenanalyse<sup>31</sup> ergibt sich hieraus für die Mittelwerte  $\boldsymbol{\mu}_g$  und die Varianzen/Kovarianzen  $\boldsymbol{\Sigma}_g$  der beobachteten Indikatoren

$$\boldsymbol{\mu}_g = \boldsymbol{\nu}_g + \boldsymbol{\Lambda}_g \boldsymbol{\alpha}_g, \quad (5)$$

$$\boldsymbol{\Sigma}_g = \boldsymbol{\Lambda}_g \boldsymbol{\Psi}_g \boldsymbol{\Lambda}_g' + \boldsymbol{\Theta}_g, \quad (6)$$

wobei  $\boldsymbol{\alpha}_g$  ein  $M \times 1$ -Vektor mit latenten Mittelwerten,  $\boldsymbol{\Psi}_g$  eine  $M \times M$ -Kovarianzmatrix der latenten Variablen und  $\boldsymbol{\Theta}_g$  eine  $J \times J$ -Kovarianzmatrix der Residualgrößen sind.

<sup>28</sup> Dieses Grundmodell basiert auf den Arbeiten von Jöreskog (1971) und Sörbom (1974).

<sup>29</sup> Im folgenden wird unterstellt, dass es sich bei den beobachteten Variablen um sogenannte reflektive Indikatoren handelt, die die nicht direkt beobachtbare Ausprägung einer zugrunde liegenden latenten Variable bei einem Individuum widerspiegeln. Eine Anwendung des hier vorgestellten Instrumentariums auf formative Messmodelle ist nicht sinnvoll, obwohl die Kovarianzstrukturanalyse die Schätzung derartiger Modelle prinzipiell ermöglicht (vgl. z. B. Temme 2006b). Zu einer Diskussion reflektiver vs. formativer Indikatoren in der Marketingforschung siehe Albers/Hildebrandt (2006); Diamantopoulos/Winklhofer (2001); Jarvis/MacKenzie/Podsakoff (2003); Rossiter (2002).

<sup>30</sup> Bei einem Faktormodell mit Einfachstruktur, bei dem jeder Indikator nur genau einem Faktor zugeordnet ist, sind bis auf die entsprechende Faktorladung alle anderen Ladungen gleich Null.

<sup>31</sup> Vgl. z. B. Bollen (1989), S. 233.

### 3.3 Test auf faktorielle Invarianz

Die Messinvarianz einer Skala kann dann als gesichert gelten, wenn sich durch eine simultane Analyse der gruppenspezifischen Kovarianzen und Mittelwerte die *strikte faktorielle Invarianz*<sup>32</sup> (d. h. die Invarianz von Konstanten, Faktorladungen und Residualvarianzen) des entsprechenden Faktormodells (vgl. die Gleichungen (5) und (6) im *Abschnitt 3.2*) nachweisen lässt. Für den empirischen Nachweis empfiehlt sich ein Prozedere, das sukzessive die Äquivalenz der Faktorstruktur (konfigurale Invarianz), der Faktorladungen (metrische Invarianz), der Konstanten (skalare Invarianz) sowie der Residualvarianzen in den untersuchten Gruppen überprüft<sup>33</sup>; umfassende Testhierarchien liegen z. B. von *Steenkamp/Baumgartner* (1998) sowie *Vandenberg/Lance* (2000) vor. Allerdings finden sich in der Literatur auch Vorschläge, die die Analyse von Mittelwertstrukturen ausklammern und damit nur die Überprüfung von Teilaspekten der Messinvarianz erlauben<sup>34</sup>. *Baumgartner/Steenkamp* (2006, S. 431) fordern dagegen, dass „... marketing researchers should make the means of scale items an explicit component of measurement analysis and should test for the invariance of item loadings and intercepts as a prerequisite for meaningful comparisons of scale means across samples and over time.“

Nachfolgend werden daher die einzelnen Stufen eines vollständigen Invarianztests auf Basis der Kovarianz- und Mittelwertinformationen erläutert (siehe den Überblick in *Abbildung 5*). Anschließend wird dann das Konzept partieller Invarianz behandelt und auf die Identifikation derjenigen Indikatoren eingegangen, die auf der jeweiligen Teststufe für eine Verletzung der Invarianzbedingung verantwortlich sind.

---

<sup>32</sup> Die Messinvarianz stellt gegenüber der strikten faktoriellen Invarianz insofern das allgemeinere Konzept dar, als sie unabhängig von einem bestimmten statistischen Modell (z. B. dem Faktormodell) als erfüllt angesehen wird, wenn bei gegebenen Werten für eine latente Variable die zugehörige bedingte Verteilung der beobachteten Indikatorwerte unabhängig von der Gruppenzugehörigkeit der Individuen ist. Vgl. *Meredith* (1993).

<sup>33</sup> Häufig wird vorgeschlagen, in einem ersten Schritt zunächst die Kovarianzmatrizen sowie Mittelwerte auf Gleichheit zu testen. Siehe hierzu z. B. *Steenkamp/Baumgartner* (1998) oder *Vandenberg/Lance* (2000). Der praktische Nutzen dieser Tests ist allerdings eher gering. Sind z. B. aufgrund theoretischer Überlegungen deutliche Gruppenunterschiede in den Faktorvarianzen oder den latenten Mittelwerten zu erwarten, so lassen diese Testergebnisse keine Schlussfolgerungen im Hinblick auf die Messinvarianz der Indikatoren zu. Vgl. *Cheung/Rensvold* (1999); *Vandenberg/Lance* (2000).

<sup>34</sup> Vgl. z. B. den von *Krafft/Litfin* (2002) in Anlehnung an *Jöreskog* (1971) in dieser Zeitschrift vorgeschlagenen Ansatz der „simultanen Faktorenanalyse“.



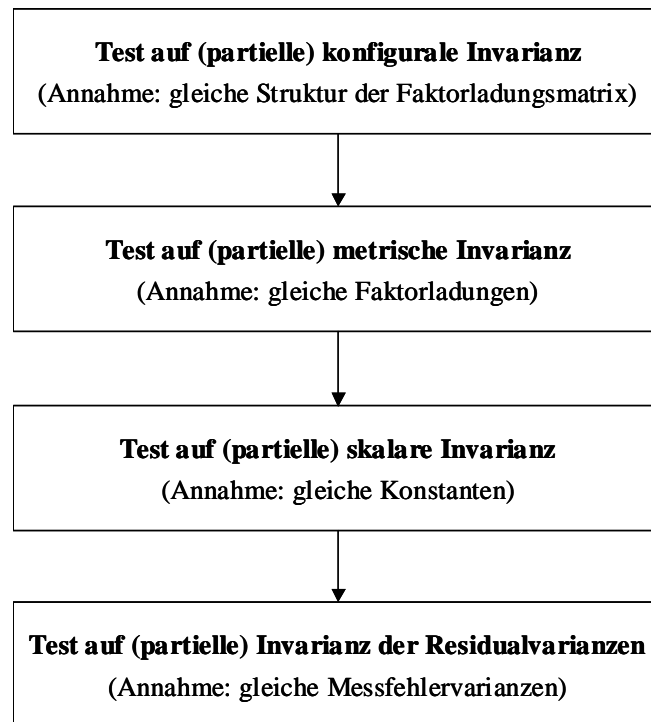


Abbildung 5: Schematischer Ablauf der Messinvarianzüberprüfung

### 3.3.1 Konfigurale Invarianz

Der Test auf konfigurale Invarianz (bzw. Forminvarianz) prüft zunächst, ob die Befragten in den verschiedenen Gruppen den gleichen gedanklichen Bezugsrahmen bei ihrer Reaktion auf die Fragen bzw. Statements eines Messinstrumentes haben. Hiervon kann ausgegangen werden, wenn die Konstrukte die gleiche Dimensionalität aufweisen und eine identische Zuordnung der Indikatoren zu den Konstrukten vorliegt. Für den empirischen Nachweis wird daher für alle Gruppen das gleiche Muster an fixierten (i. d. R. 0 bzw. 1) und freien (gruppenspezifischen) Parametern in den Faktorladungsmatrizen  $\Lambda_g$  spezifiziert<sup>35</sup>. Obwohl sich die Forminvarianz alleine auf Basis von Varianz/Kovarianz-Informationen überprüfen lässt, kann es bereits in diesem Stadium der Messinvarianzüberprüfung sinnvoll sein, simultan ein Modell für die beobachteten Mittelwerte der Indikatoren zu schätzen<sup>36</sup>. Um den Test nicht zu verzerren, wird für die Mittelwerte ein sogenanntes saturiertes Modell spezifiziert<sup>37</sup>. Dies

<sup>35</sup> Vgl. Horn/McArdle (1992).

<sup>36</sup> Wicherts/Dolan (2004) weisen z. B. darauf hin, dass Informationskriterien (z. B. AIC, BIC) sowie der ECVI bei einem Vergleich von Modellen ohne (z. B. Test auf konfigurale Invarianz) und mit Mittelwertstruktur (z. B. Test auf skalare Invarianz) zu einer falschen Rangfolge des Modellfits führen können. Deshalb empfehlen sie ein saturiertes Mittelwertmodell auch dann mitzuschätzen, wenn der Invarianztest dies eigentlich nicht erfordert.

<sup>37</sup> Ein solches Modell besitzt keine Freiheitsgrade und liefert somit eine exakte Anpassung an die Daten.

lässt sich dadurch erreichen, dass die Mittelwerte der latenten Variablen  $\alpha_g$  in allen Gruppen auf Null fixiert, die Konstanten dagegen frei geschätzt werden. Letztere entsprechen daher exakt den beobachteten Mittelwerten der Indikatoren in den einzelnen Gruppen. Damit resultiert als Modell konfiguraler Invarianz:

$$\boldsymbol{\mu}_g = \boldsymbol{v}_g, \quad (7)$$

$$\boldsymbol{\Sigma}_g = \boldsymbol{\Lambda}_g \boldsymbol{\Psi}_g \boldsymbol{\Lambda}_g' + \boldsymbol{\Theta}_g, \quad (8)$$

wobei  $Struktur(\boldsymbol{\Lambda}_1) = Struktur(\boldsymbol{\Lambda}_2) = \dots = Struktur(\boldsymbol{\Lambda}_G)$ . Zeigen die gängigen Gütemaße eine akzeptable Modellanpassung, und sind die geschätzten Faktorladungen signifikant sowie auch substantiell von Null verschieden<sup>38</sup>, so ist davon auszugehen, dass den Antworten in den untersuchten Gruppen zumindest eine identische Faktorstruktur zugrunde liegt<sup>39</sup>. Dieses Modell kann damit als Basismodell für die Überprüfung restriktiverer Invarianzbedingungen herangezogen werden. Lässt sich die Annahme konfiguraler Invarianz dagegen nicht empirisch stützen, so ist zu vermuten, dass die Items unterschiedliche hypothetische Konstrukte in den untersuchten Gruppen messen. In der Literatur herrscht weitgehend die Auffassung vor, dass in einer solchen Situation weitere Invarianztests nicht sinnvoll und Vergleiche zwischen den Gruppen unzulässig sind<sup>40</sup>.

### 3.3.2 Metrische Invarianz

Den empirischen Nachweis der konfiguralen Invarianz des Messmodells vorausgesetzt, wird im nächsten Schritt untersucht, ob die Stärke des Zusammenhangs zwischen den Faktoren und ihren zugehörigen Indikatoren in den betrachteten Gruppen identisch ist. Hierzu werden die zuvor frei geschätzten Faktorladungen in der Matrix  $\boldsymbol{\Lambda}_g$  je Indikator auf Gleichheit in den Gruppen restringiert; die Faktorladungsmatrizen stimmen somit in allen Gruppen überein. Für die Kovarianzstruktur wird damit folgendes Modell geprüft<sup>41</sup>:

$$\boldsymbol{\Sigma}_g = \boldsymbol{\Lambda} \boldsymbol{\Psi}_g \boldsymbol{\Lambda}' + \boldsymbol{\Theta}_g. \quad (9)$$

Kann die Hypothese gleicher Faktorladungen nicht abgelehnt werden (dies ist z. B. dann der Fall, wenn sich durch die Gleichheitsrestriktionen für die Faktorladungen der Modell-Fit im

<sup>38</sup> Vgl. Steenkamp/Baumgartner (1998).

<sup>39</sup> Streng genommen muss darüber hinaus auch das Kriterium der Diskriminanzvalidität in allen Gruppen erfüllt sein. Vgl. Steenkamp/Baumgartner (1998).

<sup>40</sup> Vgl. z. B. Vandenberg/Lance (2000). Baumgartner/Steenkamp (1998) schlagen einen verallgemeinerten Ansatz vor, in dem unterschiedliche Mengen von Indikatoren und latenten Variablen berücksichtigt werden können. Zu der besonderen Problematik einer solchen Invarianzüberprüfung siehe aber Meredith/Horn (2001).

<sup>41</sup> Die Mittelwertstruktur bleibt in diesem Modell von den Gleichheitsrestriktionen der Faktorladungen unberührt, da alle latenten Mittelwerte wiederum auf Null fixiert sind und nur die Konstanten geschätzt werden (siehe Gleichung (7)).

Vergleich zum vorangehenden Modell mit gruppenspezifischen Ladungen nicht signifikant verschlechtert), so besitzen die Indikatoren in den einzelnen Gruppen jeweils die gleichen Skaleneinheiten<sup>42</sup>. Bei einer in allen Gruppen identischen Veränderung einer latenten Variablen ist somit zu erwarten, dass auch die Veränderungen der zugehörigen Indikatoren jeweils in allen Gruppen gleich sind. Wie *Abschnitt 2* gezeigt hat, sind dagegen bei unterschiedlichen Faktorladungen die Mittelwertunterschiede der betroffenen Indikatoren i. d. R. verzerrt. Darüber hinaus können die Strukturparameter für die Beziehungen zwischen den latenten Variablen nicht mehr ohne weiteres über die Gruppen verglichen werden (vgl. nochmals das geschätzte Strukturgleichungsmodell in *Abbildung 4*).

### 3.3.3 Skalare Invarianz

Unter der Prämisse metrischer Invarianz des Messmodells wird in einem weiteren Schritt überprüft, ob auch die Konstanten der Indikatoren in den Gruppen jeweils identisch sind. Hierzu wird getestet, ob das entsprechend restringierte Faktormodell in der Lage ist, nicht nur die gruppenspezifischen Varianz-Kovarianzmatrizen, sondern zusätzlich auch die Indikatormittelwerte adäquat zu reproduzieren<sup>43</sup>. Da die Konstanten der Indikatoren und die Mittelwerte der latenten Variablen nicht unabhängig voneinander identifiziert sind<sup>44</sup>, werden hierzu die latenten Mittelwerte in einer Referenzgruppe (z. B. der ersten Gruppe) auf Null fixiert und die Konstanten der selben Indikatoren jeweils auf Gleichheit in allen Gruppen restringiert<sup>45</sup>. Die latenten Mittelwerte in den übrigen Gruppen werden dagegen frei geschätzt, so dass unterschiedliche Ausprägungen des Konstrukts in den Gruppen (dies wäre z. B. bei einem Vergleich der Einstellung loyaler und nichtloyaler Kunden zu einer Marke zu erwarten) bei diesem Test berücksichtigt werden<sup>46</sup>. Für  $\alpha_1 = 0$  resultiert damit folgendes Mittelwertmodell<sup>47</sup>:

---

<sup>42</sup> Für diesen Nachweis ist es wichtig, dass tatsächlich Kovarianzen untersucht werden und nicht, wie z. B. bei *Krafft/Litfin* (2002, S. 75), Korrelationen. In diesem Fall ist es nämlich nicht mehr möglich, beim Test auf gleiche Faktorladungen für unterschiedliche Varianzen der latenten Variablen in den Gruppen zu kontrollieren, so dass die Ergebnisse des Invarianztests verzerrt sein können. Vgl. auch *Steenkamp/Baumgartner* (1998).

<sup>43</sup> Dieser Test ist in dem von *Krafft/Litfin* (2002) vorgeschlagenen Ansatz nicht enthalten.

<sup>44</sup> Bei einer Faktorladung von z. B. 1 kann jede *Reduktion* der Konstante um einen Betrag  $\tau$  durch eine *Erhöhung* des latenten Mittelwertes um den Betrag  $\tau$  kompensiert werden, ohne dass sich der Mittelwert des Indikators hierdurch verändert.

<sup>45</sup> In der Literatur wird alternativ eine Spezifikation der Mittelwertstruktur vorgeschlagen, bei der die Konstanten eines Referenz-Items in allen Gruppen auf Null fixiert werden. Vgl. z. B. *Steenkamp/Baumgartner* (1998). Diese Identifikationsrestriktion ist jedoch problematisch, da die Schätzung der latenten Mittelwerte sowie der zugehörigen Standardfehler von der Wahl des Referenz-Items abhängig ist. Vgl. *Little* (2000) sowie *Meredith/Horn* (2001).

<sup>46</sup> In der aktuellen Literatur kursieren insofern Fehlinterpretationen dieses Tests, als unterstellt wird, ungleiche Konstanten könnten *auch* auf Unterschiede in den latenten Mittelwerten zurückzuführen sein. Siehe z. B. *Ployhart/Oswald* (2004), S. 35 ff.; *Raju/Laffitte/Byrne* (2002), S. 520; *Rensvold* (2002), S. 8 f.; *Vandenberg/Lance* (2000), S. 38. So schlussfolgert etwa *Rensvold* (2002): „If one wishes to test the equality of

$$\boldsymbol{\mu}_1 = \boldsymbol{\nu}, \quad (10)$$

$$\boldsymbol{\mu}_g = \boldsymbol{\nu} + \boldsymbol{\Lambda} \boldsymbol{\alpha}_g, \quad g = 2, \dots, K \text{ } G. \quad (11)$$

Wird die Annahme skalarer Invarianz nicht abgelehnt, so liegt zumindest eine sogenannte *starke Form der faktoriellen Invarianz* (gleiche Faktorladungen und gleiche Konstanten) vor<sup>48</sup>. In diesem Fall können zusätzlich zu den Beziehungen zwischen den latenten Variablen auch die Mittelwerte der Indikatoren in den einzelnen Gruppen miteinander verglichen werden. Darüber hinaus ist eine Gegenüberstellung der Gruppen anhand von Indizes (d. h. Durchschnitts- oder Summenscores) sinnvoll. Mittelwertvergleiche (inkl. Signifikanztests) können aber auch direkt auf der Ebene der latenten Variablen vorgenommen werden, da die frei geschätzten latenten Mittelwerte (d. h.  $\alpha_g$ , mit  $g \neq$  Referenzgruppe) auf einer gemeinsamen Skala gemessene Differenzen in Relation zur Referenzgruppe darstellen. Ist die Bedingung skalarer Invarianz dagegen verletzt, so sind sowohl die Mittelwertunterschiede der Indikatoren (siehe nochmals *Abbildung 1*) als auch die latenten Mittelwerte verzerrt.

### 3.3.4 Invarianz der Residualvarianzen

An den Nachweis skalarer Invarianz kann sich als letzter Schritt ein Test anschließen, bei dem geprüft wird, ob die Indikatoren jeweils gleiche Residual- bzw. Messfehlervarianzen in den einzelnen Gruppen aufweisen<sup>49</sup>. Werden entsprechende Gleichheitsrestriktionen auf die Varianzen in der Matrix  $\Theta_g$  gelegt, so ergibt sich folgendes Kovarianzstrukturmodell:

$$\boldsymbol{\Sigma}_g = \boldsymbol{\Lambda} \boldsymbol{\Psi}_g \boldsymbol{\Lambda}' + \boldsymbol{\Theta}_g. \quad (12)$$

Wird dieses Modell empirisch gestützt, so erfüllt das Messinstrument alle Bedingungen für *strikte faktorielle Invarianz*. Wie erläutert, impliziert dies unter der Annahme multivariat normalverteilter Daten eine vollständige Messinvarianz der Indikatoren.

### 3.4 Partielle Invarianz

Die zuvor dargestellte Testprozedur führt im günstigsten Fall zum Nachweis der vollständigen Messinvarianz aller Indikatoren. In empirischen Studien dürfte eine solche

---

intercepts, one must assume the equality of latent means, and vice versa" (S. 8). Tatsächlich lässt sich aber sowohl theoretisch als auch anhand von Simulationsstudien (vgl. *Stark/Chernyshenko/Drasgow* 2006) zeigen, dass die Invarianz der Konstanten auch bei unterschiedlichen latenten Mittelwerten in den Gruppen valide überprüft werden kann.

<sup>47</sup> Das Kovarianzstrukturmodell entspricht demjenigen für metrische Invarianz.

<sup>48</sup> Vgl. *Meredith* (1993).

<sup>49</sup> Bei der Auswahl von Individuen (z. B. Kunden) auf Basis der Gesamt- oder Summenscores implizieren Differenzen in den Residualvarianzen eine unterschiedliche Selektionsgenauigkeit. Vgl. *DeShon* (2004); *Lubke et al.* (2003); *Millsap/Kwok* (2004). Zu kontroversen Auffassungen über die Stellung dieses Tests innerhalb der Hierarchie der Modelltests siehe *Little* (2000) und *Lubke/Dolan* (2003).

Konstellation allerdings eher die Ausnahme darstellen. Bei Ablehnung einer bestimmten Äquivalenzhypothese (z. B. gleiche Faktorladungen) stellt sich somit die Frage nach dem weiteren Vorgehen. Aufgrund der Nichtinvarianz auf einen Vergleich der Gruppen zu verzichten, ist unter forschungspraktischen Gesichtspunkten sicherlich unbefriedigend. Insofern besitzt das Konzept partieller Invarianz<sup>50</sup> große Relevanz für die empirische Forschung. Die Grundidee lässt sich wie folgt skizzieren: Ist nur ein kleiner Teil der einem Faktor zugeordneten Indikatoren nichtinvariant, so werden diese z. B. bei weiterführenden Invarianztests oder Tests auf Unterschiede in den Mittelwerten der latenten Variablen zwar weiter einbezogen, durch die Schätzung gruppenspezifischer Parameter (d. h. Faktorladungen, Konstanten und Residualvarianzen) für diese Indikatoren wird aber die fehlende Messinvarianz kontrolliert<sup>51</sup>. Der entscheidende erste Schritt bei der Anwendung dieses Konzeptes besteht somit darin, die invarianzverletzenden Indikatoren valide zu identifizieren. Die damit verbundene Problematik wird hier anhand partieller metrischer Invarianz illustriert.

Idealerweise sollte die Identifikation derjenigen Indikatoren, die zu einer Ablehnung der Annahme metrischer Invarianz für das gesamte Messmodell eines Konstrukts führen<sup>52</sup>, auf theoretischen Überlegungen basieren<sup>53</sup>. In empirischen Studien lässt sich aber der Anspruch einer streng konfirmatorischen Herangehensweise i. d. R. nicht erfüllen, so dass letztlich exploratorisch unter Zuhilfenahme statistischer Kriterien vorgegangen wird. Hierzu sind unterschiedliche Ansätze vorgeschlagen worden<sup>54</sup>. Bei denjenigen Verfahren, bei denen die Skalierung der latenten Variable über die Fixierung der Faktorladung eines Referenzindikators erfolgt, besteht das wesentliche Unterscheidungsmerkmal darin, ob der Referenzindikator systematisch variiert wird oder nicht<sup>55</sup>.

---

<sup>50</sup> Vgl. *Byrne/Shavelson/Muthén* (1989).

<sup>51</sup> Vgl. z. B. *Steenkamp/Baumgartner* (1998) zum Konzept partieller konfiguraler Invarianz.

<sup>52</sup> Vereinfachend wird hier von einem Modell mit einem Faktor und zwei Gruppen ausgegangen. In einem Modell mit mehreren Faktoren werden zunächst diejenigen Faktoren identifiziert, für deren Indikatoren eine Invarianzverletzung vorliegt. Die nichtinvarianten Indikatoren werden dann in der für den Ein-Faktor-Fall beschriebenen Weise ermittelt. Vgl. z. B. *Rensvold/Cheung* (1998). Zur Durchführung von Invarianztests bei mehr als zwei Gruppen siehe insbesondere *Rensvold* (2002).

<sup>53</sup> Vgl. *Vandenberg/Lance* (2000).

<sup>54</sup> Vgl. hierzu ausführlich *Cheung/Rensvold* (1999).

<sup>55</sup> Die alternative Form der Skalierung einer latenten Variable durch Standardisierung (d. h. Fixierung der Varianz auf den Wert 1) bei Messinvarianztests findet nur selten Anwendung. Da die Standardisierung nur für eine Gruppe vorgenommen werden darf, ist allerdings für die Modellidentifikation auch hier wieder die Auswahl eines als invariant angenommenen Indikators notwendig (vgl. *Reise/Widaman/Pugh* 1993). Aus diesem Grund beschränken wir uns hier auf die Referenzindikatorstrategie. Vgl. aber *Yoon/Millsap* (2007) zu ersten Befunden über die Leistungsfähigkeit des alternativen Skalierungsansatzes bei Verwendung von Modifikationsindizes.

Zu letzteren Ansätzen zählt z. B. das von *Krafft/Litfin* (2002) vorgeschlagene Verfahren, bei dem von einem Modell ausgegangen wird, in dem alle Ladungen für die Indikatoren eines Faktors jeweils auf Gleichheit in den Gruppen restringiert sind. Danach wird separat für jeden Indikator (mit Ausnahme des Referenzindikators) ein Modell geschätzt, in dem dessen Faktorladung über die Gruppen frei variieren kann, alle anderen Faktorladungen aber weiterhin einer Gleichheitsrestriktion unterliegen. Als nichtinvariant werden dann diejenigen Indikatoren angesehen, für die sich aufgrund des Chi-Quadrat-Differenztests eine signifikante Verbesserung des Fit im Vergleich zum Ausgangsmodell ergibt<sup>56</sup>. Problematisch wird eine solche, auf den ersten Blick durchaus sinnvoll erscheinende Vorgehensweise durch die Beibehaltung des gleichen Referenzindikators in allen Tests. Für diesen Indikator wird nämlich durch die Fixierung der Faktorladung auf den Wert 1 in allen Gruppen implizit metrische Invarianz unterstellt. Trifft diese Annahme aber nicht zu, so sind die einzelnen Tests auf Invarianz verzerrt und können zu falschen Schlussfolgerungen bezüglich invarianter bzw. nichtinvarianter Items führen<sup>57</sup>.

Da die zuvor angerissene Problematik in der betriebswirtschaftlichen Forschung weitgehend vernachlässigt wird<sup>58</sup>, wird zur Illustration nochmals auf das Modell in *Abbildung 3* zurückgegriffen. Auf Basis der Parameter des Messmodells der latenten exogenen Variablen  $\xi$  wurden für die  $x$ -Indikatoren Stichprobendaten für die beiden Gruppen A und B simuliert (mit  $n = 500$  in jeder Gruppe). Als Referenzindikator diente dabei wiederum die Variable  $x_1$ . Die Schätzung eines Modells mit gleichen Faktorladungen in den beiden Gruppen führt im Vergleich zu einem Modell ohne Gleichheitsrestriktionen zu einer signifikanten Verschlechterung, d. h. die Annahme metrischer Invarianz wird korrekterweise abgelehnt. Wird nun das oben beschriebene Verfahren von *Krafft/Litfin* (2002) angewendet, um die Ursache der Nichtinvarianz zu lokalisieren, so ergeben sich nur für die beiden Indikatoren  $x_5$  und  $x_6$  signifikante Verbesserungen des Modell-Fit. Damit werden also gerade zwei tatsächlich invariante Indikatoren (vgl. *Abbildung 3*) als nichtinvariant gekennzeichnet. Für diejenigen Indikatoren, die im wahren Modell unterschiedliche Faktorladungen in den Gruppen aufweisen ( $x_1$ ,  $x_2$  und  $x_3$ ), würde dagegen von metrischer Invarianz ausgegangen. Insgesamt wird damit also ein verzerrtes Bild über die Messäquivalenz der untersuchten

---

<sup>56</sup> Alternativ kann auch das Modell konfiguraler Invarianz als Ausgangspunkt gewählt werden. In diesem Fall werden Modelle geschätzt, bei denen neben dem Referenzindikator nur die Faktorladung je eines Indikators auf Gleichheit restringiert ist. Vgl. *Byrne/Shavelson/Muthén* (1989).

<sup>57</sup> Vgl. *Cheung/Rensvold* (1999).

<sup>58</sup> Dies zeigt bereits ein cursorischer Blick in die Literatur. Vgl. z. B. die an *Steenkamp/Baumgartner* (1998) angelehnte Überprüfung auf partielle metrische Invarianz bei *De Jong/Steenkamp/Fox* (2007), *Erdem/Swait/Valenzuela* (2006) oder *Netemeyer/Brashear-Alejandro/Boles* (2004).

Indikatoren gezeichnet<sup>59</sup>. Selbst wenn für die vermeintlich nichtinvarianten Indikatoren gruppenspezifische Faktorladungen zugelassen werden, ergeben sich wiederum verzerrte Schätzer für die strukturellen Beziehungen zwischen den latenten Variablen.

Die zentrale Schwäche der zuvor beschriebenen Gruppe von Ansätzen umgehen *Cheung/Rensvold* durch eine systematische Variation des Referenzindikators<sup>60</sup>. Getestet wird dabei jeweils die metrische Invarianz von Indikatorpaaren bestehend aus einem Referenz- und einem Argumentindikator. Für die übrigen Indikatoren werden die Faktorladungen frei geschätzt. Ist ein bestimmter Referenzindikator (z. B.  $x_1$ ) mit allen anderen Indikatoren ( $x_2$  bis  $x_6$ ) kombiniert worden, wird der nächste Indikator ( $x_2$ ) als Referenzindikator ausgewählt und jeweils zusammen mit den verbleibenden Indikatoren ( $x_3$  bis  $x_6$ ) auf metrische Invarianz getestet. Dieses Prozedere wird solange wiederholt, bis alle Indikatoren (bis auf  $x_6$ ) einmal als Referenzindikator fungiert haben. Über Chi-Quadrat-Differenztests kann dann wieder bestimmt werden, für welche Indikatorpaare die Annahme gleicher Faktorladungen in den Gruppen zu einer signifikanten Verschlechterung des Fit gegenüber dem Modell ohne Gleichheitsrestriktionen führt<sup>61</sup>. Sind auf diesem Weg alle invarianten und nichtinvarianten Indikatorpaare ermittelt worden, so lässt sich hieraus, z. B. mit dem Programm *Stepwise.exe*<sup>62</sup>, die Menge invarianter Indikatoren extrahieren<sup>63</sup>. Dabei kann es passieren, dass verschiedene Sets von Indikatoren resultieren, die mit der Annahme metrischer Invarianz kompatibel sind. Die Entscheidung darüber, welches Set letztlich als invariant angesehen wird, sollte in diesem Fall auf theoretische Überlegungen rekurrieren. Führt die Schätzung des Modells mit Gleichheitsrestriktionen für die zuvor als invariant gekennzeichneten Faktorladungen im Vergleich zum Modell konfiguraler Invarianz nicht zu einer signifikanten Verschlechterung, so kann von partieller metrischer Invarianz ausgegangen werden.

Wird das oben beschriebene Verfahren von *Cheung/Rensvold* auf den simulierten Datensatz angewendet, so führen die paarweisen Chi-Quadrat-Differenztests zu den in *Tabelle 2* zusammengefassten Ergebnissen.

---

<sup>59</sup> Bei der von *Steenkamp/Baumgartner* (1998) vorgeschlagenen Orientierung an Modifikationsindizes werden ebenfalls die Indikatoren  $x_5$  und  $x_6$  als nichtinvariant identifiziert.

<sup>60</sup> Vgl. *Cheung/Rensvold* (1999); *Rensvold/Cheung* (1998, 2001).

<sup>61</sup> Alternativ können hier auch inkrementelle Fit-Indizes wie z. B. der CFI genutzt werden. Vgl. *Cheung/Rensvold* (2002).

<sup>62</sup> Diese von *Rensvold* entwickelte Software kann auf Anfrage vom Autor dieses Beitrags angefordert werden.

<sup>63</sup> Vgl. zu alternativen Vorgehensweisen *Rensvold/Cheung* (2001); *French/Finch* (2008).

Tabelle 2: Ergebnisse der Chi-Quadrat-Differenztests nach Cheung und Rensvold für simulierte Daten

		Argument-Item				
		$x_2$	$x_3$	$x_4$	$x_5$	$x_6$
Referenz-Item	$x_1$	4,40	4,94	7,38	28,27*	50,29*
	$x_2$	—	0,001	1,21	9,93*	17,57*
	$x_3$	—	—	1,34	11,08*	20,63*
	$x_4$	—	—	—	2,04	3,58
	$x_5$	—	—	—	—	0,15

Die mit \* gekennzeichneten Teststatistiken sind bei einem Freiheitsgrad signifikant für  $\alpha = 0,0033$ <sup>64</sup>.

Das vorliegende Muster aus signifikanten und nichtsignifikanten Teststatistiken ist mit der Existenz zweier verschiedener Mengen von „invarianten“ Indikatoren kompatibel:

- (1)  $x_1, x_2, x_3, x_4$
- (2)  $x_4, x_5, x_6$

Dass sich hierunter auch das Set (1) mit ausschließlich *nichtinvarianten* Indikatoren befindet, hängt damit zusammen, dass mit dem Invarianztest nur bestimmte Proportionalitätsrestriktionen getestet werden können<sup>65</sup>. Der Vorteil des Verfahrens von Cheung/Rensvold im Vergleich zu den Ansätzen mit fixiertem Referenzindikator besteht also darin, dass er ein vollständiges Bild über alternative Invarianzsets liefert. Daher empfehlen wir dieses Verfahren für empirische Überprüfungen der Messinvarianz.

Im Anschluss an die Identifikation der Indikatoren mit invarianten Faktorladungen wird in einem weiteren Schritt die partielle skalare Invarianz des Messmodells überprüft. Im Ausgangsmodell für den Test auf partielle skalare Invarianz bleiben die Konstanten der Indikatoren mit nichtinvarianten Faktorladungen unbeschränkt<sup>66</sup>. Alle übrigen Konstanten werden auf Gleichheit in den Gruppen restringiert. Wird dieses Modell abgelehnt, so erfolgt analog zum Fall partieller metrischer Invarianz die Suche nach den nichtinvarianten

<sup>64</sup> Durch die mehrfache Durchführung von Chi-Quadrat-Differenztests ergibt sich die Problematik, wie das nominale Testniveau von z. B. 5 % für alle Tests zusammen gewährleistet werden kann. Eine Möglichkeit ist die Bonferroni-Korrektur, durch die im vorliegenden Fall das Signifikanzniveau für die individuellen Tests entsprechend der Formel  $\alpha_{\text{kor}} = 0,05/[\text{Anzahl der Indikatoren} * (\text{Anzahl der Indikatoren} - 1)/2]$  angepasst wird.

<sup>65</sup> Vgl. Zu den Details Cheung/Rensvold (1999).

<sup>66</sup> Vgl. Byrne/Shavelson/Muthén (1989); Steenkamp/Baumgartner (1998).



Indikatoren<sup>67</sup>. Auch in dieser Situation können die notwendigen Identifikationsrestriktionen in einem Modell mit minimalen Beschränkungen (d. h. zwei auf Gleichheit in den Gruppen restringierte Konstanten) die Testergebnisse verzerren<sup>68</sup>. Obwohl nicht von *Rensvold* und *Cheung* diskutiert, kann ihr Ansatz auf die Identifikation nichtinvarianter Konstanten übertragen werden<sup>69</sup>. Sind hinreichend viele Konstanten (formal mindestens zwei pro Faktor) über die Gruppen gleich, so liegt partielle skalare Invarianz vor. Damit sind auch bei unvollständiger Invarianz Gruppenvergleiche der latenten Mittelwerte möglich<sup>70</sup>. Abschließend kann für diejenigen Indikatoren, die sowohl metrisch als auch skalar invariant sind, die Äquivalenz der Residualvarianzen überprüft werden.

#### **4 Eine empirische Überprüfung der Messäquivalenz am Beispiel der Markenstärke**

Ein kontinuierliches Monitoring der wahrgenommenen Attraktivität einer Marke bzw. der psychologischen Markenstärke ist für eine zielgerichtete Markenführung eine wichtige Informationsbasis. Dabei ist es sinnvoll, nicht nur die aggregierte Entwicklung für alle Konsumenten in einer bestimmten Produktkategorie zu verfolgen, sondern auch segmentspezifische Verläufe zu analysieren. Ein Vergleich loyaler und nichtloyaler Konsumenten zeigt z. B., ob sich die Differenz in der wahrgenommenen Markenattraktivität im Zeitverlauf eher verringert, konstant bleibt oder sogar noch vergrößert<sup>71</sup>. Ein valider Vergleich der beiden Gruppen setzt aber die Invarianz des verwendeten Messmodells der Markenstärke voraus. Hier setzt die empirische Studie an: Anhand der in *Abschnitt 3* dargestellten Mehrgruppenanalyse konfirmatorischer Faktormodelle wird die Messäquivalenz kognitiver, affektiver und verhaltensorientierter Indikatoren der Markenstärke bei unterschiedlich loyalen Konsumenten überprüft.

##### *4.1 Die Operationalisierung der Markenstärke durch den BPI<sup>®</sup>*

Für die Operationalisierung des Markenstärkekonstrukts wird auf den „Brand Potential Index“ (BPI<sup>®</sup>) der GfK Marktforschung zurückgegriffen. Dieser wird in der Praxis z. B. im Rahmen

---

<sup>67</sup> Sind im Ausgangsmodell nur die Minimalrestriktionen für partielle metrische Invarianz erfüllt, d. h. besitzen nur zwei Indikatoren gleiche Faktorladungen in den Gruppen, so lassen sich bei einer Ablehnung allerdings keine weiterführenden Tests auf partielle Invarianz mehr durchführen.

<sup>68</sup> Für den Test auf invariante Messkonstanten eines Argument-Items muss die Konstante eines Referenz-Items auf Gleichheit in den Gruppen restringiert werden. Hiermit ist wiederum eine implizite Invarianzannahme für diesen Indikator verbunden.

<sup>69</sup> Zu einem alternativen mehrstufigen Verfahren siehe *Meredith/Horn* (2001).

<sup>70</sup> Formal sind zwei Indikatoren mit invarianten Faktorladungen und Konstanten pro Faktor ausreichend, um im Rahmen eines Modells partieller starker faktorieller Invarianz die latenten Mittelwerte miteinander vergleichen zu können. Vgl. z. B. *Steenkamp/Baumgartner* (1998). Die meisten Autoren vertreten aber die Auffassung, dass nur eine Minderheit an Indikatoren nichtinvariant sein sollte. Vgl. z. B. *Vandenberg/Lance* (2000).

<sup>71</sup> Vgl. z. B. *Huber/Herrmann/Peter* (2002).

von Trackingstudien, als Bestandteil eines umfassenden finanziellen Markenbewertungssystems<sup>72</sup> oder zur Erhebung von Markenrankings<sup>73</sup> in vielfältiger Weise genutzt. In der aktuell gültigen Konzeption<sup>74</sup> umfasst das Messmodell des BPI® insgesamt 10 Variablen, die einen großen Teil der in der Markenliteratur diskutierten Indikatoren der Markenstärke abbilden<sup>75</sup>. Trotz der Bezeichnung als Index sind die Indikatoren mit einer Ausnahme als reflektiv anzusehen, da sich im Antwortverhalten auf die entsprechenden Statements die von einer Person empfundene Attraktivität einer etablierten Marke (also ein bereits vor der Befragung existenter interner Zustand) manifestiert<sup>76</sup>. Insgesamt ist die Konstruktoperationalisierung dennoch nicht unproblematisch, weshalb einige kritische Aspekte eingehender diskutiert werden sollen.

Zunächst ist zu konstatieren, dass der Indikator Markenbekanntheit nicht die Bekanntheit beim befragten Konsumenten selbst erfasst<sup>77</sup>. Stattdessen soll der/die Befragte angeben, in welchem Ausmaß er/sie glaubt, dass die Marke „bekannter als andere Marken“ ist. Dieses Statement misst somit eher die empfundene Popularität einer Marke<sup>78</sup>. Damit handelt es sich jedoch nicht um einen reflektiven Indikator der wahrgenommenen Attraktivität einer Marke<sup>79</sup>. So ist es z. B. denkbar, dass eine Marke als populär angesehen wird, ohne dass der/die Befragte dieser Marke selber eine hohe Wertschätzung entgegenbringt<sup>80</sup>. Mit diesen Überlegungen korrespondieren auch die Ergebnisse einer exploratorischen Faktorenanalyse. Es zeigt sich, dass die Markenbekanntheit zwar mit den übrigen BPI®-Indikatoren moderat positiv korreliert ist, dennoch aber einen separaten Faktor bildet. Für die vorliegende Analyse wurde daher der Indikator Markenbekanntheit sowohl aus konzeptionellen als auch statistischen Gründen ausgeschlossen.

Kritisch zu sehen ist darüber hinaus die Einbindung der verhaltensorientierten Variablen (Kaufabsicht, Mehrpreisakzeptanz und Bereitschaft zur Weiterempfehlung) in das

---

<sup>72</sup> Vgl. *Sattler/Högl/Hupp* (2002).

<sup>73</sup> Siehe z. B. [www.bestbrands.de](http://www.bestbrands.de).

<sup>74</sup> Die Anzahl der Indikatoren des BPI® hat sich im Zeitverlauf erhöht. Vgl. z. B. *Grimm/Högl/Hupp* (2000, S. 8); *Högl/Twardawa/Hupp* (2001, S. 39); *Hupp/Xu* (2003, S. 225).

<sup>75</sup> Vgl. zu einer Synopse von Markenstärkeindikatoren z. B. *Morschett* (2002).

<sup>76</sup> Vgl. z. B. *Rossiter* (2002).

<sup>77</sup> Die gestützte Markenbekanntheit wird bereits vorab abgefragt und ist Voraussetzung für die vollständige Durchführung eines Interviews.

<sup>78</sup> Vgl. *Aaker* (1996).

<sup>79</sup> Bezeichnenderweise ist das Statement zur Markenbekanntheit das einzige, das nicht das Personalpronomen „ich“ enthält.

<sup>80</sup> Als Beispiel können hier etwa eingefleischte UNIX-Anwender angeführt werden, die nach der Popularität der Marke WINDOWS gefragt werden.

Messmodell für den BPI<sup>®</sup>. Als Alternative wäre eine Modellierung denkbar, in der ein ausschließlich über affektive und kognitive Indikatoren gemessener BPI<sup>®</sup> die drei Verhaltensvariablen als nachgelagerte Größen beeinflusst. Interessanterweise ist ein solches Strukturmodell aber statistisch äquivalent zum hier unterstellten Messmodell<sup>81</sup>. Sollte sich also z. B. herausstellen, dass die Faktorladungen der verhaltensorientierten Indikatoren über die Gruppen variieren, so wäre dies gleichbedeutend mit der Tatsache, dass im äquivalenten Strukturmodell die Steigungsparameter für die Regression der Verhaltensvariablen auf den BPI<sup>®</sup> in den Gruppen jeweils verschieden sind. Die Interpretation dieser Differenzen ist in diesem Fall analog zur derjenigen für das reflektive Messmodell: Bei gleichem Anstieg der wahrgenommenen Markenattraktivität würde für die einzelnen Verhaltensindikatoren die Zustimmung in den Gruppen jeweils unterschiedlich stark zunehmen. Obwohl also die Modellierung der Verhaltensvariablen als abhängige Größen eine sinnvolle Alternative darstellt, wird im Folgenden die ursprüngliche Konzeption des BPI<sup>®</sup>-Messmodells beibehalten.

Für die Kauf- und Weiterempfehlungsabsicht ist die Annahme plausibel, dass sich diese Indikatoren mehr oder weniger im Gleichklang mit den affektiven und kognitiven Items verändern, wenn die wahrgenommene Attraktivität der Marke variiert. Für die Mehrpreisakzeptanz gilt dies u. U. aber nur eingeschränkt; so muss sich bei einer Discountmarke (z. B. Aldi oder Ryanair) die Steigerung der wahrgenommenen Markenattraktivität nicht zwangsläufig auch in einer Bereitschaft zur Zahlung höherer Preise niederschlagen<sup>82</sup>. Während also für bestimmte Marken u. U. nur eine schwache oder gar keine Korrelation mit den übrigen Markenstärkeindikatoren zu erwarten ist, kann bei der hier untersuchten führenden Marke in ihrer Produktkategorie<sup>83</sup> allerdings davon ausgegangen werden, dass sich die zunehmende Attraktivität der Marke für eine Person auch in einer entsprechenden Mehrpreisakzeptanz widerspiegeln würde. Diese Einschätzung wird durch die empirischen Ergebnisse gestützt, auch wenn der Indikator Mehrpreisakzeptanz im Vergleich zu den anderen Indikatoren die geringste Indikatorreliabilität in beiden Konsumentensegmenten aufweist<sup>84</sup>.

---

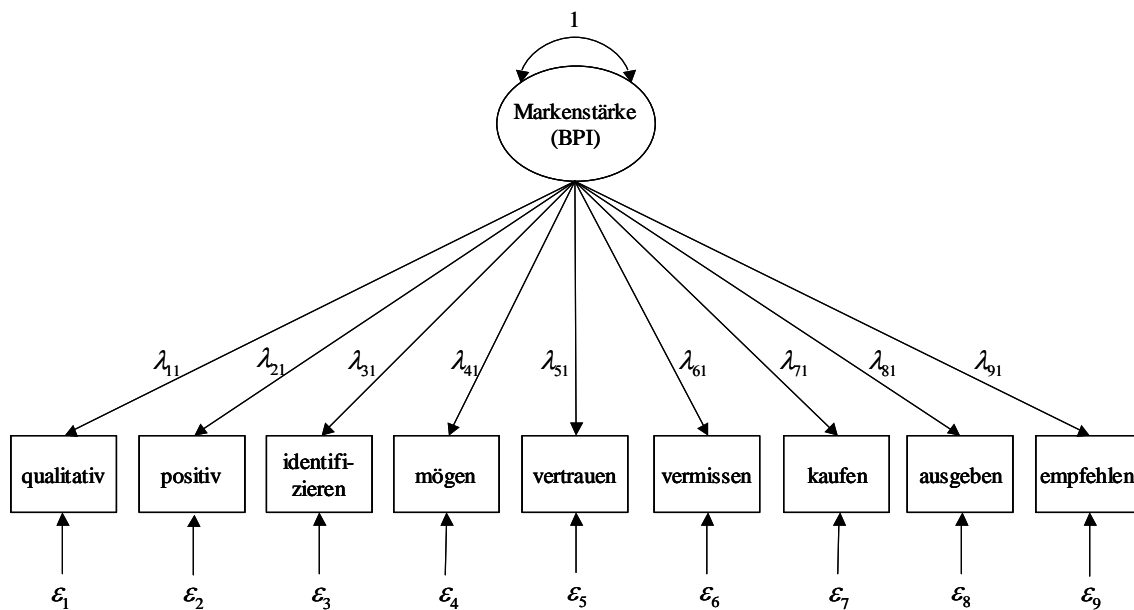
<sup>81</sup> Vgl. *DeShon* (2004).

<sup>82</sup> Wir danken einem anonymen Gutachter für diesen Hinweis.

<sup>83</sup> Zu näheren Informationen zur Marke und der zugehörigen Produktkategorie siehe *Abschnitt 4.2*.

<sup>84</sup> Vgl. *Tabelle 4* in *Abschnitt 4.3.1*.

Aus der kritischen Diskussion der BPI<sup>®</sup>-Operationalisierung lässt sich abschließend folgendes Fazit ziehen: Die Markenstärke wird – nach Ausschluss der Variable Markenbekanntheit – in dieser Studie über 9 reflektive Indikatoren gemessen, die kognitive, affektive und verhaltensorientierte Facetten abdecken (vgl. zum Faktormodell 1. Ordnung und seinen Indikatoren *Abbildung 6* sowie *Tabelle A1* im Anhang). Der Grad ihrer Zustimmung zu den einzelnen Markenstärke-Statements wird von den befragten Personen auf siebenstufigen Ratingskalen angegeben, deren Endpunkte mit „trifft überhaupt nicht zu“ (1) und „trifft voll und ganz zu“ (7) bezeichnet sind.



*Abbildung 6: Messung der Markenstärke durch den Brand Potential Index (in Anlehnung an Hupp/Xu 2003, S. 225)*

#### 4.2 Daten

Die Daten der empirischen Studie entstammen einer Befragung von 1048 haushaltsführenden Personen durch die GfK Marktforschung aus dem Jahre 2000. Im Rahmen persönlicher Interviews wurden die befragten Konsumenten gebeten, konkurrierende Marken aus dem Segment der Kartoffelhalbfertigprodukte anhand der BPI<sup>®</sup>-Indikatoren sowie genereller bzw. produktspezifischer Eigenschaften zu beurteilen. Darüber hinaus wurden u. a. Informationen zum vergangenen Kaufverhalten sowie zu sozio-demografischen Kriterien erhoben. Für die vorliegende Analyse wurden nur Daten für die im untersuchten Segment führende Herstellermarke verwendet, da für die übrigen Marken keine ausreichende Fallzahl für eine Mehrgruppenanalyse erreicht wurde und eine simultane Analyse mehrerer Marken die

Annahme unabhängig verteilter Beobachtungen verletzt hätte<sup>85</sup>. Nach einer Eliminierung von Fällen mit inkonsistentem Antwortverhalten stand eine effektive Stichprobe von 1037 Individuen zur Verfügung.

Die Aufteilung der Gesamtstichprobe in zwei unterschiedliche Loyalitätssegmente basiert auf den vergangenheitsbezogenen Angaben der Befragten zu ihrer Markenwahl. Die Gruppe 1 ( $n = 638$ ) setzt sich aus denjenigen Personen zusammen, die angaben, in den letzten 3 bis 6 Monaten überwiegend die im Fokus der Untersuchung stehende Marke in einer der beiden wesentlichen Produktkategorien gekauft zu haben, in der diese Marke angeboten wird. Alle übrigen Befragten wurden der Gruppe 2 ( $n = 399$ ) zugeordnet. Nachfolgend werden die Befragten in Gruppe 1 und 2 auch als loyale bzw. nichtloyale Konsumenten bezeichnet. Die Kovarianzmatrizen sowie die Mittelwerte für die beiden Gruppen gibt die *Tabelle A2* im *Anhang* wieder. Ein vorgeschaltetes Daten-Screening mit PRELIS 2.80 zeigt, dass die Markenstärkeindikatoren zwar statistisch signifikant von der Normalverteilung abweichen, nach den in der Literatur vorgeschlagenen Daumenregeln aber höchstens für die Indikatoren „kaufen“ und „vermissen“ ein nennenswertes Ausmaß an Nichtnormalverteilung vorliegt<sup>86</sup>.

#### 4.3 Ergebnisse der Messinvarianztests

Der Testhierarchie in *Abschnitt 3* folgend, wird zunächst die Annahme überprüft, dass den Indikatoren des BPI<sup>®</sup>-Messmodells bei loyalen und nichtloyalen Konsumenten die gleiche eindimensionale Faktorstruktur zugrunde liegt<sup>87</sup>. Hierzu sowie auch für alle weiteren Analysen wird der in Mplus 5.0<sup>88</sup> implementierte MLM-Schätzer verwendet. Dieser gibt die *Satorra-Bentler (SB)* Chi-Quadrat-Statistik  $\chi_{SB}^2$ <sup>89</sup> sowie robuste Standardfehler aus, die jeweils für Abweichungen von der Normalverteilungsannahme korrigieren. Der Maximum-

---

<sup>85</sup> Vgl. z. B. *Bentler/Chou* (1987).

<sup>86</sup> Diese Schlussfolgerung bezieht sich auf *Muthén/Kaplan* (1985), die in ihrer relativ konservativen Einschätzung davon ausgehen, dass eine beträchtliche Abweichung von der Normalverteilung erst ab einer Schiefe bzw. Wölbung von größer |1| vorliegt. Andere Autoren legen hier weit großzügigere Maßstäbe an. *West/Finch/Curran* (1995) sprechen z. B. erst ab Werten von Schiefe = |2| und Wölbung = |7| von einer substantiellen Abweichung. Diese Größenordnungen werden in den beiden Gruppen für die Indikatoren „kaufen“ (stand. Schiefe = |1,27| und stand. Wölbung = |1,17| in Gruppe 1) und „vermissen“ (stand. Wölbung = |1,02| in Gruppe 2) deutlich unterschritten. Für alle anderen Indikatoren liegen die Schiefe- bzw. Wölbungswerte unter |0,86|.

<sup>87</sup> Für den konzeptionell als dreidimensionales Konstrukt angelegten BPI<sup>®</sup> wurde ein alternatives konfirmatorisches Faktorenmodell mit einem kognitiven, einem affektiven und einem konativen Faktor spezifiziert. Die Schätzung führte zu einer unzulässigen Lösung, da für die Gruppe der loyalen Konsumenten Faktorkorrelationen  $> 1$  resultierten. Dies kann als klares Indiz für fehlende Diskriminanzvalidität gewertet werden. In einer separaten Analyse anhand des *Fornell-Larcker-Kriteriums* (*Fornell/Larcker* 1981) musste auch für die Gruppe der nichtloyalen Konsumenten die Annahme der Diskriminanzvalidität abgelehnt werden.

<sup>88</sup> *Muthén/Muthén* (1998-2007).

<sup>89</sup> Vgl. *Satorra/Bentler* (1988); im folgenden wird i. d. R. auf das Subskript „SB“ verzichtet.

Likelihood-Schätzer für die Parameter ist angesichts der eher geringen Verletzung der Normalverteilungsannahme hinreichend robust, so dass von validen Schätzwerten auszugehen ist<sup>90</sup>. Die Fit-Maße aller getesteten Modelle sowie die Ergebnisse der Modellvergleiche auf Basis von Likelihood-Verhältnis-Tests (bzw. Chi-Quadrat-Differenztests) sind in *Tabelle 3* dokumentiert.

#### 4.3.1 Test auf konfigurale Invarianz

Für die Schätzung des Modells konfigurale Invarianz (M1.1) resultieren die folgenden Anpassungsmaße<sup>91</sup>:  $\chi^2_{(54)} = 172,60$ ,  $p = 0,000$ , CFI = 0,981, RMSEA = 0,065, SRMR = 0,023 und WRMR = 0,631. Damit ergibt sich ein recht uneinheitliches Bild der Modellgüte<sup>92</sup>. Zwar erfüllen CFI, SRMR und WRMR die erforderlichen Kriterien für einen guten Fit, nicht aber RMSEA sowie das Verhältnis der Chi-Quadrat-Statistik zu den Freiheitsgraden des Modells ( $\chi^2/df = 3,20$ ). Für die Kovarianzmatrix der Messresiduen zeigen darüber hinaus die erwarteten Parameterveränderungen sowie die Modifikationsindizes, dass eine Reihe korrelierter Residualgrößen vorliegt.

Während eine a posteriori Freisetzung von Parametern allein auf der Basis statistischer Kriterien grundsätzlich Probleme aufwirft<sup>93</sup>, tritt im Rahmen der Invarianzüberprüfung bei einer Kovariation von Residualgrößen ein weiteres Problem hinzu: Wird die Korrelation der Residuen zweier Indikatoren als Resultat einer gemeinsamen Beeinflussung durch einen zusätzlichen unbeobachteten Faktor angesehen, so können die zugehörigen Faktorladungen nicht getrennt von der Varianz dieses Faktors in den Gruppen auf Invarianz überprüft werden<sup>94</sup>. Der Invarianzstatus bleibt damit letztlich ungeklärt. Dennoch sprechen gewichtige Argumente für eine Freisetzung signifikanter Kovarianzen zwischen den Residualgrößen: Zum einen kann die Nichtspezifikation vorhandener Kovarianzen einen verzerrenden Einfluss auf die Invari-

---

<sup>90</sup> Vgl. *Boomsma/Hoogland* (2001).

<sup>91</sup> Für die Identifikation des Modells wurden die Faktorvarianz des BPI® in der Gruppe der nichtloyalen Konsumenten auf den Wert 1 fixiert sowie eine Gleichheitsrestriktion für die Faktorladung des Indikators „positiv“ spezifiziert.

<sup>92</sup> In Anlehnung an die Simulationsstudie von *Hu/Bentler* (1999) werden für die Modellbeurteilung die folgenden Schwellenwerte für eine gute Anpassung an die Ausgangsdaten zugrunde gelegt: CFI  $\geq 0,95$ , RMSEA  $\leq 0,06$ , SRMR  $\leq 0,08$ . Aufgrund der nichtnormalverteilten Daten wird zusätzlich der Weighted Root-Mean-Square Residual (WRMR) berichtet (*Muthén/Muthén* 1998, S. 361 f.), für den eine Akzeptanzschwelle von WRMR  $\leq 0,95$  verwendet wird (*Yu* 2002). Ergänzend werden für das Verhältnis  $\chi^2/df$  im Anschluss an *Bollen* (1989) Werte  $\leq 2,5$  gefordert. Zu einer umfassenden Diskussion von Fit-Maßen für Strukturgleichungsmodelle siehe z. B. *Homburg/Baumgartner* (1998).

<sup>93</sup> Vgl. z. B. *MacCallum/Roznowski/Necowitz* (1992).

<sup>94</sup> Vgl. *Lubke et al.* (2003).

anztests ausüben<sup>95</sup>. Zum anderen wird bei Ablehnung einer globalen Invarianzhypothese für einen bestimmten Parametertyp (z. B. Faktorladungen) ebenfalls auf statistische Kriterien zurückgegriffen, um nichtinvariante Indikatoren zu identifizieren. Eine Aufhebung von Invarianzrestriktionen im Rahmen eines Modells partieller Invarianz aufgrund derartiger Kriterien lässt sich aber nur schwer rechtfertigen, wenn gleichzeitig hoch signifikante Fehlerkorrelationen toleriert werden<sup>96</sup>.

Tabelle 3: Ergebnisse der Tests auf faktorielle Invarianz

Mo- dell	Hypothesen	$\chi^2_{SB}$	df	p	Refe- renz	$\Delta\chi^2_{SB}$	df	p	RMSEA	CFI	SRMR	WRMR
<b>Konfigurale Invarianz</b>												
M1.1	$\Lambda_1^{Form} = \Lambda_2^{Form}$	172,60	54	0,000	-	-	-	-	0,065	0,981	0,023	0,631
M1.2	M1.1 + korr. Resid.	80,95	45	0,001	M1.1	86,51	9	0,000	0,039	0,994	0,015	0,421
<b>Metrische Invarianz</b>												
M2.1	M1.2 + $\Lambda_1 = \Lambda_2$	101,38	53	0,000	M1.2	22,83	8	0,004	0,042	0,992	0,037	0,989
M2.2	M2.1 + $\lambda_1^{empfehlen} \neq \lambda_2^{empfehlen}$	94,16	52	0,000	M2.1	10,41	1	0,001	0,040	0,993	0,033	0,876
M2.3	M2.2 + $\lambda_1^{kaufen} \neq \lambda_2^{kaufen}$	87,23	51	0,001	M2.2	6,42	1	0,010	0,037	0,994	0,020	0,548
					M1.2	3,95	6	0,683				
<b>Skalare Invarianz</b>												
M3	M2.3 + $\tau_1 = \tau_2,$ $\tau_1^{empfehlen} \neq \tau_2^{empfehlen},$ $\tau_1^{kaufen} \neq \tau_2^{kaufen}$	96,31	57	0,001	M2.3	8,86	6	0,182	0,036	0,994	0,024	0,636
					M1.2	13,27	12	0,350				
<b>Invarianz der Residualvarianzen</b>												
M4.1	M3 + $\Theta_1 = \Theta_2,$ $\theta_1^{2(empf.)} \neq \theta_2^{2(empf.)},$ $\theta_1^{2(kaufen)} \neq \theta_2^{2(kaufen)}$	153,02	64	0,000	M3	47,23	7	0,000	0,052	0,985	0,035	0,910
M4.2	M3 + $\theta_1^{2(ident.)} = \theta_2^{2(ident.)},$ $\theta_1^{2(symp.)} = \theta_2^{2(symp.)}$	95,84	59	0,002	M3	0,82	2	0,664	0,035	0,994	0,024	0,635
					M1.2	12,99	14	0,527				

<sup>95</sup> Vgl. Byrne/Shavelson/Muthén (1989).

<sup>96</sup> Vgl. Steenkamp/Baumgartner (1998).

Das Ausgangsmodell M1.1 wird daher durch die Einführung gruppenspezifischer Kovarianzen zwischen mehreren Residualgrößen modifiziert<sup>97</sup>. Die Schätzung dieses Modells (M1.2) führt zu einem deutlich verbesserten Fit:  $\chi^2_{(45)} = 80,95$ ,  $p = 0,001$ , CFI = 0,994, RMSEA = 0,039, SRMR = 0,015 und WRMR = 0,421. Dementsprechend ist die *SB*-skalierte Chi-Quadrat-Differenz mit  $\Delta\chi^2_{SB} = 86,51$  ( $df = 9$ ,  $p = 0,000$ ) hoch signifikant<sup>98</sup>. Da nun alle deskriptiven Fit-Maße die entsprechenden Anforderungskriterien erfüllen, kann die globale Anpassungsgüte dieses modifizierten Faktormodells als sehr gut bezeichnet werden.

Bevor auf die weiteren Tests auf Messinvarianz eingegangen wird, soll zunächst die Reliabilität bzw. Validität der Operationalisierung der wahrgenommenen Markenattraktivität durch den BPI<sup>®</sup> beurteilt werden. Ausgehend von der akzeptablen Gesamtanpassung des modifizierten Modells konfiguraler Invarianz (M1.2) werden hierzu lokale Gütekriterien herangezogen (vgl. *Tabelle 4*). Alle Faktorladungen sind hochgradig signifikant ( $z$ -Werte  $\geq 13,03$  für Gruppe 1 und  $\geq 17,82$  für Gruppe 2) und substanziell von Null verschieden. Wird für den Anteil der Indikatorvarianz, der durch die empfundene Markenattraktivität mindestens erklärt werden muss (Indikatorreliabilität), ein Grenzwert von 0,4 zugrunde gelegt<sup>99</sup>, so erfüllen alle Indikatoren dieses Kriterium. Mit Ausnahme der beiden verhaltensorientierten Variablen „kaufen“ und „ausgeben“ liegt die Indikatorreliabilität sogar bei über 0,65. Die weiteren Kriterien (Faktorreliabilität und durchschnittlich erfasste Varianz) können als Gütemaße für Konvergenzvalidität interpretiert werden. Sie geben an, wie gut der BPI<sup>®</sup> von der Gesamtheit seiner Indikatoren erfasst wird. Als Schwellenwerte werden hier 0,5 für die Faktorreliabilität und 0,6 für die durchschnittlich erfasste Varianz herangezogen<sup>100</sup>. Auch diese Kriterien werden durch das Messmodell des BPI<sup>®</sup> in beiden Gruppen erfüllt. Als Fazit lässt sich festhalten, dass die Operationalisierung der wahrgenommenen Attraktivität einer Marke durch den BPI<sup>®</sup> alle gängigen Reliabilitäts- und Validitätskriterien erfüllt. Abgesehen von der unterschiedlichen Struktur der Residualkovarianzen erweist sich damit die eindimensionale Faktorstruktur sowohl für loyale als auch nichtloyale Konsumenten als

---

<sup>97</sup> Insgesamt wurden 9 Kovarianzen zwischen Residualgrößen zugelassen, wobei ein kritischer Wert von 7,88 ( $\alpha = 0,005$ ) für die Modifikationsindizes zugrundegelegt wurde. Dabei entfielen 6 Kovarianzen auf das Segment der loyalen und 3 auf das Segment der nichtloyalen Konsumenten. Eine strukturell übereinstimmende Korrelation von Residualgrößen in beiden Gruppen liegen nur für die beiden Indikatoren „qualitativ“ und „positiv“ vor, die beide die kognitive Facette des BPI<sup>®</sup> abdecken.

<sup>98</sup> Die Berechnung der  $\Delta\chi^2_{SB}$ -Statistik weicht von derjenigen für die konventionelle ML-Schätzung ab, da die Differenzen zwischen *SB*-skalierten Chi-Quadrat-Statistiken nicht  $\chi^2$ -verteilt sind. Zur genauen Berechnung siehe *Satorra/Bentler* (2001).

<sup>99</sup> Vgl. *Bagozzi/Baumgartner* (1994).

<sup>100</sup> Vgl. *Bagozzi/Baumgartner* (1994).



empirisch haltbar. Das modifizierte Faktormodell konfiguraler Invarianz (Modell M1.2) fungiert deshalb für die nachfolgenden Invarianztests als Basismodell.

Tabelle 4: Reliabilität und Validität des Modells konfiguraler Invarianz

Indikator	Indikatorreliabilität		Faktorreliabilität		Durchschnittlich erfasste Varianz	
	Gruppe 1	Gruppe 2	Gruppe 1	Gruppe 2	Gruppe 1	Gruppe 2
qualitativ	0,73	0,71				
positiv	0,71	0,67				
identifizieren	0,69	0,70				
mögen	0,77	0,79				
vertrauen	0,81	0,75	0,96	0,95	0,68	0,69
vermissen	0,73	0,69				
kaufen	0,59	0,57				
ausgeben	0,44	0,55				
empfehlen	0,80	0,80				

Gruppe 1 = loyale Konsumenten, Gruppe 2 = nichtloyale Konsumenten

#### 4.3.2 Test auf (partielle) metrische Invarianz

Für den Test auf metrische Invarianz werden alle Faktorladungen je Indikator auf Gleichheit in den beiden untersuchten Gruppen restringiert (Modell M2.1). Aus der Schätzung dieses Modells resultiert eine  $\chi^2$ -Statistik in Höhe von 101,38 ( $df = 53$ ). Dies impliziert gegenüber dem Basismodell M1.2 eine signifikante Verschlechterung des Modellfit ( $\Delta\chi_{SB}^2 = 22,83$ ,  $df = 8$ ,  $p = 0,004$ ). Während der WRMR den Schwellenwert von 0,95 nun knapp überschreitet, erhöhen sich der RMSEA und der CFI allerdings nur wenig.

Obwohl die deskriptiven Fit-Maße nach den gängigen Kriterien letztlich keine eindeutige Ablehnung des Modells indizieren, scheint die Annahme globaler metrischer Invarianz dennoch zu strikt. Dies zeigen auch die relativ hohen erwarteten Parameterveränderungen sowie Modifikationsindizes für die Gleichheitsrestriktionen der Faktorladungen der beiden Indikatoren „empfehlen“ und „kaufen“. Werden sukzessive gruppenspezifische Faktorladungen zunächst für den Indikator „empfehlen“ (Modell M2.2) sowie anschließend für den Indikator „kaufen“ (Modell M2.3) zugelassen, so resultieren hieraus jeweils signifikante

Verbesserungen des Modellfit ( $\Delta\chi_{SB}^2 = 10,41, p = 0,001$  bzw.  $\Delta\chi_{SB}^2 = 6,42, p = 0,010$ )<sup>101</sup>. Die Gleichheitsrestriktionen für die Faktorladungen der übrigen sieben Indikatoren führen dagegen nicht zu einer signifikanten Verschlechterung der Anpassung des Modells an die Ausgangsdaten, wie ein Vergleich mit dem Basismodell (M1.2) zeigt ( $\Delta\chi_{SB}^2 = 3,95, df = 6, p = 0,683$ ). Alle Fit-Maße indizieren darüber hinaus eine sehr gute Anpassung des Modells partieller metrischer Invarianz (M2.3) an die Daten:  $\chi_{(51)}^2 = 87,23, p = 0,001, CFI = 0,994, RMSEA = 0,037, SRMR = 0,020$  und  $WRMR = 0,548$ .

#### 4.3.3 Test auf (partielle) skalare Invarianz

Für den Test auf partielle skalare Invarianz (Modell M3) werden zusätzlich Gleichheitsrestriktionen auf die Konstanten derjenigen Indikatoren gelegt, für die metrische Invarianz nachgewiesen werden konnte. Der latente Mittelwert des BPI<sup>®</sup> wird in der Gruppe der nicht-loyalen Konsumenten auf den Wert Null fixiert, so dass für die loyalen Konsumenten mit einem positiven Mittelwert (d. h. einer höheren wahrgenommenen Markenattraktivität) zu rechnen ist. Aus der Schätzung dieses Modells resultieren die folgenden Fit-Maße:  $\chi_{(57)}^2 = 96,31, p = 0,001, CFI = 0,994, RMSEA = 0,036, SRMR = 0,024$  und  $WRMR = 0,636$ . Sowohl im Vergleich zum Modell partieller metrischer Invarianz ( $\Delta\chi_{SB}^2 = 8,86, df = 6, p = 0,182$ ) als auch zum modifizierten Modell konfiguraler Invarianz ( $\Delta\chi_{SB}^2 = 13,27, df = 12, p = 0,350$ ) verschlechtert sich der Fit durch die Annahme identischer Konstanten in den beiden Gruppen nicht signifikant. Der RMSEA weist das Modell M3 sogar als bestes unter allen bisher getesteten Modellen aus.

#### 4.3.4 Test auf (partielle) Invarianz der Residualvarianzen

Als letztes wird überprüft, ob die Varianzen der Residualgrößen in beiden Gruppen identisch sind (Modell M4.1)<sup>102</sup>. Aufgrund der merklichen Verschlechterung des Gesamtfit ( $\Delta\chi_{SB}^2 = 47,23, df = 7, p = 0,000$ ) wird die Hypothese gleicher Residualvarianzen eindeutig zurückgewiesen. Separate Tests für die einzelnen Indikatoren zeigen, dass die Annahme gleicher Residualvarianzen nur für die Indikatoren „identifizieren“ und „sympathisch“ haltbar ist; das Modell mit Gleichheitsrestriktionen für die beiden Residualvarianzen (M4.2) zeigt einen sehr

<sup>101</sup> Zusätzlich wurden systematisch paarweise Invarianztests durchgeführt (zum Verfahren von *Cheung/Rensvold* siehe nochmals den *Abschnitt 3.3*), die in diesem Fall zu den gleichen Befunden führten.

<sup>102</sup> Von diesem Test wurden wiederum die Indikatoren „empfehlen“ und „kaufen“ ausgeschlossen.

guten Fit:  $\chi^2_{(59)} = 95,84^{103}$ ,  $p = 0,002$ , CFI = 0,994, RMSEA = 0,035, SRMR = 0,024 und WRMR = 0,635. Darüber hinaus hat sich der Chi<sup>2</sup>-Wert im Vergleich zum Basismodell M1.2 nicht signifikant erhöht ( $\Delta\chi^2_{SB} = 12,99$ ,  $df = 14$ ,  $p = 0,527$ ). Die geschätzten Parameter des abschließenden Faktormodells M4.2 sind in *Tabelle 5* wiedergegeben.

*Tabelle 5: Parameterschätzer für das Modell M4.2 mit partieller faktorieller Invarianz (unstandardisierte Lösung)*

Indikator	Konstanten		Faktorladungen		Residualvarianzen		Reliabilitäten	
	Gruppe 1	Gruppe 2	Gruppe 1	Gruppe 2	Gruppe 1	Gruppe 2	Gruppe 1	Gruppe 2
qualitativ	4,10		1,27		0,49	0,75	0,73	0,68
positiv	4,16		1,24		0,49	0,78	0,72	0,66
identifizieren	3,46		1,45		0,82		0,68	0,72
mögen	3,81		1,42		0,51		0,77	0,80
vertrauen	3,83		1,41		0,41	0,65	0,80	0,76
vermissen	3,66		1,53		0,71	1,05	0,73	0,69
kaufen	4,71	4,24	1,13	1,37	0,73	1,42	0,59	0,57
ausgeben	3,10		1,36		1,92	1,36	0,44	0,58
empfehlen	4,01	3,97	1,36	1,55	0,39	0,62	0,80	0,79
			Gruppe 1		Gruppe 2			
<b>Latenter Mittelwert BPI</b>			1,07		0,00 <sup>a</sup>			
<b>Varianz BPI</b>			0,83		1,00 <sup>a</sup>			

Gruppe 1 = loyale Konsumenten, Gruppe 2 = nichtloyale Konsumenten; <sup>a</sup> fixiert

#### 4.4 Ausmaß und praktische Bedeutung der diagnostizierten Nichtinvarianz

Da von insgesamt 9 betrachteten BPI<sup>®</sup>-Indikatoren nur die verhaltensorientierten Indikatoren „kaufen“ und „empfehlen“ in den beiden Loyalitätssegmenten durch unterschiedliche Zusammenhänge mit der latenten Markenstärke gekennzeichnet sind, liegt für das BPI<sup>®</sup>-Messmodell insgesamt eine starke partielle faktorielle Invarianz bei loyalen und nichtloyalen Konsumenten vor. Die latenten Mittelwerte für diese beiden Gruppen können somit sinnvoll miteinander verglichen werden. Bei einer Standardisierung auf Basis der Faktorvarianz in der Gruppe 1 ergibt sich demnach eine hochsignifikante Differenz in der wahrgenommenen Markenstärke in Höhe von 1,18. Wie erwartet, wird die betrachtete Marke von den loyalen Konsumenten im Durchschnitt deutlich attraktiver wahrgenommen als von den nichtloyalen Konsumenten. Auch ein Vergleich der Beziehungen zwischen der Markenstärke und anderen

<sup>103</sup> Die Ursache dafür, dass im Vergleich zum Modell ohne Gleichheitsrestriktionen für die Residualvarianzen (M3) das restriktivere Modell M4.2 eine geringere SB-skalierte Chi-Quadrat-Statistik aufweist, liegt an den leicht unterschiedlichen Korrekturfaktoren für nichtnormalverteilte Daten.

latent Variablen wäre im Rahmen eines Mehrgruppenmodells für loyale und nichtloyale Konsumenten bei entsprechender Berücksichtigung der gruppenspezifischen Parameter des Messmodells möglich<sup>104</sup>.

Im folgenden wird nun näher auf die Interpretation und Signifikanz der festgestellten Unterschiede in den Konstanten und Faktorladungen der beiden verhaltensorientierten Indikatoren eingegangen. Die Konstanten sind bei beiden Indikatoren in der Gruppe der loyalen Konsumenten größer als in der Gruppe der nichtloyalen Konsumenten (vgl. *Tabelle 5*). Beim Indikator „empfehlen“ ist die Differenz aber statistisch nicht signifikant und das zugehörige 95%-Konfidenzintervall legt sich fast symmetrisch um den Nullpunkt (vgl. *Tabelle 6*)<sup>105</sup>. Im Gegensatz dazu ist der Unterschied für den Indikator „kaufen“ signifikant ( $p = 0,0006$ ) und vergleichsweise groß; das Konfidenzintervall schließt in diesem Fall den Wert Null nicht ein. Wird für beide Gruppen ein hypothetischer Wert von Null für die Markenstärke unterstellt (beide Gruppen würden in diesem Fall die Marke im Durchschnitt als gleich attraktiv ansehen), würde die Zustimmung der loyalen Befragten zum Statement für die Kaufintention im Durchschnitt um ca. 0,5 Skalenpunkte höher ausfallen als bei den nichtloyalen Konsumenten. Für die Weiterempfehlungsabsicht ist dagegen kein Unterschied zu erwarten.

Im Vergleich zu den Konstanten ergibt sich für die Faktorladungen ein genau entgegengesetztes Bild. Hier sind bei den Indikatoren „kaufen“ und „empfehlen“ die Faktorladungen in der Gruppe der loyalen Konsumenten geringer als in der Gruppe der nichtloyalen Konsumenten (vgl. *Tabelle 5*). Die Unterschiede sind jeweils statistisch signifikant ( $p = 0,016$  für „kaufen“ und  $p = 0,017$  für „empfehlen“) und die zugehörigen Konfidenzintervalle enthalten den Wert Null nicht (vgl. *Tabelle 6*). Fällt also z. B. die wahrgenommene Attraktivität der untersuchten Marke, so nimmt bei den loyalen Konsumenten die Zustimmung zu den entsprechenden Statements nicht so stark ab wie bei den nichtloyalen Konsumenten. Allgemeiner formuliert bedeutet dies, dass Veränderungen der wahrgenommenen Attraktivität der Marke bei den loyalen Konsumenten weniger stark auf ihre Kaufintention und Weiterempfehlungsabsicht durchschlagen.

---

<sup>104</sup> Selbstverständlich erfordert dies die zumindest partielle Äquivalenz der Messmodelle für die weiteren Konstrukte.

<sup>105</sup> Der zur Berechnung der z-Statistik und des Konfidenzintervalls benötigte Standardfehler ( $SF$ ) für die Differenz zwischen den geschätzten Konstanten bzw. Faktorladungen der beiden Gruppen lässt sich über die Formel  $SF(\hat{\nu}_1 - \hat{\nu}_2) = \sqrt{SF(\hat{\nu}_1)^2 + SF(\hat{\nu}_2)^2}$  bzw.  $SF(\hat{\lambda}_1 - \hat{\lambda}_2) = \sqrt{SF(\hat{\lambda}_1)^2 + SF(\hat{\lambda}_2)^2}$  ermitteln. Vgl. Meade/Bauer (2007).

Tabelle 6: Gruppenunterschiede in den Konstanten und Faktorladungen der Indikatoren „kaufen“ und „empfehlen“ bei loyalen und nichtloyalen Konsumenten

Indikator	Differenz Gruppe 1 – Gruppe 2		z-Wert		95%-Konfidenzintervall	
	Konstante	Faktorladung	Konstante	Faktorladung	Konstante	Faktorladung
kaufen	0,47	-0,24	3,44	-2,42	[0,20; 0,74]	[-0,44; -0,05]
empfehlen	0,04	-0,19	0,32	-2,39	[-0,20; 0,28]	[-0,34; -0,03]

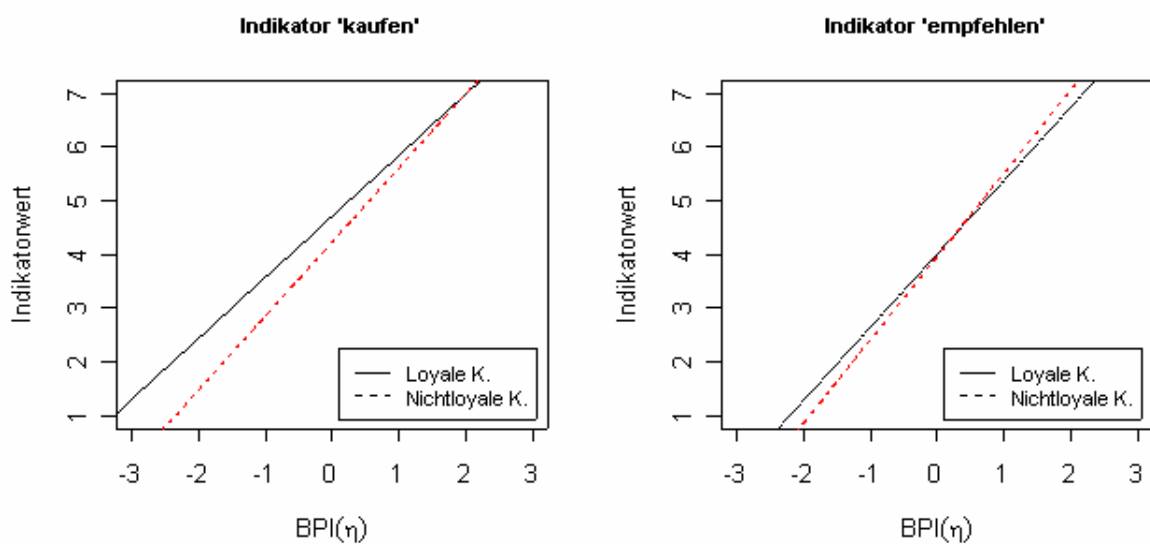
Gruppe 1 = loyale Konsumenten, Gruppe 2 = nichtloyale Konsumenten

Aufgrund der höheren Konstanten in der Gruppe der loyalen Konsumenten und der höheren Faktorladungen in der Gruppe der nichtloyalen Konsumenten tritt der in *Abschnitt 2* bereits kurz erörterte Fall ein, dass sich die gruppenspezifischen Messgeraden schneiden (siehe *Abbildung 7*). In diesem Fall ist zwar eine Ermittlung der „wahren“ Differenz der Itemmittelwerte auf Basis der beobachteten Mittelwerte der beiden Gruppen nicht möglich, die Differenzen können aber zumindest in eine gemeinsame Metrik transformiert werden<sup>106</sup>. Da die Ergebnisse von der Wahl der Bezugsgruppe abhängen, wird diese Berechnung für beide Gruppen vorgenommen (vgl. *Tabelle 7*)<sup>107</sup>. Verglichen mit der entsprechend der Metrik in Gruppe 1 skalierten Differenz ist der beobachtete Unterschied zwischen den Mittelwerten des Items „kaufen“ ca. 39 % größer; auf Basis der Metrik in Gruppe 2 beträgt die Überschätzung immerhin noch ca. 14 %. Ein Vergleich der loyalen und der nichtloyalen Konsumenten auf Basis der in dieser Studie beobachteten Differenz der Mittelwerte für den Indikator „kaufen“ überzeichnet folglich den tatsächlichen Unterschied in der durchschnittlichen Markenstärke. Wie *Abbildung 7* deutlich macht, fällt die Verzerrung dabei umso größer aus, je geringer die wahrgenommene Markenstärke ist. Im Gegensatz dazu lässt sich für den Indikator „empfehlen“ keine eindeutige Schlussfolgerung ziehen. Wird die Skala in Gruppe 1 herangezogen, so besteht praktisch kein Unterschied zwischen der beobachteten und der transformierten Differenz; auf Basis der Skala in Gruppe 2 fällt der beobachtete

<sup>106</sup> Für einen metrisch invarianten Indikator  $y_j$  mit gruppenspezifischen Konstanten kann die korrigierte, „wahre“ Mittelwertdifferenz zwischen den Gruppen 1 und 2 gemäß  $\mu_{j1} - \mu_{j2} = \lambda_j(\alpha_1 - \alpha_2)$  berechnet werden. Vgl. *Scholderer/Grunert/Brunso* (2005). Variieren dagegen auch die Faktorladungen, so lässt sich mit Hilfe der Formel die Differenz jeweils in der Metrik der Gruppe 1 und der Gruppe 2 ermitteln, wobei  $\lambda_{j1}$  bzw.  $\lambda_{j2}$  verwendet wird.

<sup>107</sup> Die zugrunde liegende Idee lässt sich anhand der Ermittlung von Temperaturunterschieden verdeutlichen. Hierzu folgendes Beispiel: In einem Land A wird eine Durchschnittstemperatur von 68° Fahrenheit (F) gemessen, während die mittlere Temperatur im Land B 15° Celsius (C) beträgt. Wird nicht berücksichtigt, dass die Temperaturen auf unterschiedlichen Skalen gemessen wurden, so ergibt sich eine Differenz in den Durchschnittswerten von 53°. Um zu ermitteln, ob und in welchem Umfang diese Differenz den „wahren“ Unterschied verzerrt, können die Durchschnittstemperaturen einheitlich entweder in °F oder in °C berechnet werden. Als Temperaturdifferenzen resultieren hieraus 9 °F (= 68 °F – 59 °F) oder 5 °C (= 20 °C – 15 °C), womit deutlich wird, dass die ursprüngliche Differenz den tatsächlichen Unterschied deutlich überzeichnet.

Mittelwertunterschied dagegen 10 % geringer aus. Neben diesen relativ geringen Abweichungen zeigen auch die linearen Responsefunktionen in beiden Gruppen mit Ausnahme der äußeren Skalenbereiche einen ähnlichen Verlauf (vgl. *Abbildung 7*).



*Abbildung 7: Zusammenhang zwischen BPI<sup>®</sup> und den Indikatoren „kaufen“ und „empfehlen“*

*Tabelle 7: Beobachtete und transformierte Mittelwertdifferenzen für die Indikatoren „kaufen“ und „empfehlen“*

<b>Indikator</b>	<b>Differenz</b>	<b>Über-/ Unterschätzung</b>
<b>kaufen</b>		
- beobachtet	1,68	–
- Skala Gruppe 1	1,21	+38,8 %
- Skala Gruppe 2	1,47	+14,3 %
<b>empfehlen</b>		
- beobachtet	1,49	–
- Skala Gruppe 1	1,46	+0,2 %
- Skala Gruppe 2	1,66	-10,2 %

Zusammenfassend lässt sich feststellen, dass für den praktischen Einsatz der BPI<sup>®</sup>-Skala ein Vergleich von loyalen und nichtloyalen Konsumenten auf Basis der Mittelwerte von 8 der insgesamt 9 untersuchten Indikatoren sinnvoll ist. Beim Indikator „kaufen“ ist dagegen davon auszugehen, dass die beobachteten Mittelwertdifferenzen den „wahren“ Unterschied in der

zugrunde liegenden Wertschätzung für die Marke überzeichnen. Für die loyalen Konsumenten zeigt sich, dass diese im Vergleich zu den nichtloyalen Konsumenten weniger sensitiv auf Veränderungen der wahrgenommenen Markenattraktivität reagieren, also z. B. ihre Kauf- und Weiterempfehlungsabsicht weniger stark reduzieren.

#### *4.5 Mögliche Ursachen der diagnostizierten Nichtinvarianz*

Abschließend stellt sich die Frage, welche Ursachen für die Nichtinvarianz der beiden Indikatoren „kaufen“ und „empfehlen“ ausgemacht werden können. Zunächst liegt es dabei nahe, nach statistischen Erklärungen für die mangelnde Messäquivalenz zu suchen. Mögliche Ursachen könnten die Nichtnormalverteilung sowie die Ordinalität der Daten (aufgrund der Verwendung von Ratingskalen) sein<sup>108</sup>.

Blickt man zunächst auf die Verteilung der beobachteten Indikatorwerte, so zeigt sich, dass diese in der Gruppe der loyalen Konsumenten linksschief ist<sup>109</sup>. Die Nichtnormalverteilung der Daten als Erklärung für die Nichtinvarianz würde allerdings nur für den Indikator „kaufen“ greifen, der mit Abstand die größte Schiefe und Wölbung aufweist. Beim Indikator „empfehlen“ liegen diese Werte dagegen deutlich unter den als kritisch angesehenen Schwellenwerten<sup>110</sup>. Darüber hinaus sind für andere, messinvariante Indikatoren sogar größere Abweichungen von der Normalverteilung zu verzeichnen (z. B. für die Indikatoren „positiv“ oder „mögen“). Eine eng mit dem hier vorliegenden Verteilungsmuster verknüpfte Überlegung wäre, dass durch die Verwendung einer Ratingskala mit einer begrenzten Anzahl an Antwortkategorien ein Deckelungseffekt („ceiling effect“) in der Gruppe der loyalen Konsumenten auftritt. In der Tat wählt beim Indikator „kaufen“ ein großer Teil der Befragten die höchste Antwortkategorie, so dass hieraus u. U. eine geringere Faktorladung resultiert als bei den nichtloyalen Konsumenten. Auch hier zeigt sich wieder, dass diese Erklärung beim Indikator „empfehlen“ nicht greift. Das Antwortmuster für diesen Indikator ist vergleichbar mit dem anderer Indikatoren (z. B. „vertraue“), für die keine fehlende Invarianz der Faktorladung festgestellt wurde.

---

<sup>108</sup> Vgl. hierzu generell z. B. *DiStefano* (2002) sowie zur speziellen Problematik in Mehrgruppenanalysen *Lubke/Muthén* (2004).

<sup>109</sup> In der Gruppe der nichtloyalen Konsumenten ist die Verteilung ebenfalls linksschief, wenn auch in geringerem Ausmaß.

<sup>110</sup> Wie in *Abschnitt 4.2* beschrieben, liegen die univariaten Werte für Schiefe und Wölbung beim Indikator „empfehlen“ deutlich unter den als kritisch angesehenen Schwellenwerten.

Um Nichtnormalität und Ordinalität der Daten als alleinige Ursachen der Nichtinvarianz noch sicherer ausschließen zu können, wurde zusätzlich eine Mehrgruppenanalyse durchgeführt, bei der ein Faktormodell für ordinale Indikatoren unterstellt wurde<sup>111</sup>. Die Ergebnisse bestätigen fast vollständig die Befunde bei Annahme kontinuierlicher Indikatoren. Mangelnde Messinvarianz wird wie zuvor nur für die beiden verhaltensorientierten Indikatoren angezeigt. Für die Faktorladungen, die wiederum in Gruppe 2 höher als in Gruppe 1 ausfallen, ergeben sich signifikante Unterschiede. Analog zu den höheren Konstanten in der Gruppe der loyalen Konsumenten im linearen Faktormodell zeigen die unterschiedlichen Schwellenwerte für den Indikator „kaufen“, dass bei den loyalen Konsumenten in Relation zu den nichtloyalen Konsumenten durchgehend geringere Ausprägungen der latenten Markenstärke ausreichen, damit diese die nächsthöhere Antwortkategorie auf der Ratingskala wählen. Für den Indikator „empfehlen“ ist dagegen die Hypothese gleicher Schwellenwerte empirisch haltbar. Dies korrespondiert mit der Nichtsignifikanz der Gruppenunterschiede bei den Konstanten dieses Indikators. Zusammenfassend lässt sich somit das Fazit ziehen, dass die Ergebnisse der hier dokumentierten Messinvarianzüberprüfung nicht als statistische Artefakte anzusehen sind. Damit verlagert sich die Suche nach den Ursachen der diagnostizierten Nichtinvarianz auf die inhaltliche Ebene.

Wie *Abbildung 7* zeigt, äußern die loyalen Konsumenten praktisch über den gesamten Bereich der beobachteten Ratingskala im Vergleich zu den nichtloyalen Konsumenten bei gleicher wahrgenommener Markenattraktivität ein höheres Maß an Bereitschaft, die Marke zukünftig zu kaufen. Gleichzeitig geht die Kaufbereitschaft der loyalen Konsumenten bei abnehmender Markenattraktivität weniger stark zurück als bei den nichtloyalen Konsumenten<sup>112</sup>. Verringert sich die Bereitschaft zum Kauf der Marke, so ist dies bei bisher loyalen Konsumenten gleichbedeutend mit einer zunehmenden Neigung, andere Marken zu kaufen. Eine mögliche Erklärung für die Nichtäquivalenz des Items „kaufen“ sollte daher beleuchten, welche Gründe gegen einen Markenwechsel loyaler Konsumenten sprechen.

Da an dieser Stelle keine umfassende Diskussion aller möglichen Ursachen erfolgen kann, fokussieren wir im Folgenden auf psychologisch-soziale und/oder ökonomische

---

<sup>111</sup> Vgl. nochmals *Abschnitt 3.1*. Da dieser Ansatz nur für die jedem beobachteten Indikator zugeordnete latente Responsevariable Normalverteilung unterstellt, wird auch die Nichtnormalverteilung der beobachteten Indikatorwerte bei der Parameterschätzung berücksichtigt.

<sup>112</sup> Diese Interpretation ergibt sich aus der geringeren Faktorladung in der Gruppe der loyalen Konsumenten.



Wechselbarrieren als Erklärung<sup>113</sup>. Während derartige Wechselbarrieren bei Dienstleistungsunternehmen (z. B. aufgrund persönlicher Beziehungen zu einem Versicherungsvertreter) oder bestimmten Gebrauchsgütern (z. B. aufgrund spezifischer Kenntnisse bei der Anwendung einer Software) relativ offensichtlich sind, erscheint eine Übertragung dieses Konzepts auf Konsumgüter des täglichen Bedarfs zunächst fragwürdig. Eine genauere Betrachtung lässt allerdings erkennen, dass auch in diesen Produktbereichen Wechselbarrieren für die Konsumenten entstehen können. Diese können zum einen aus akkumuliertem markenspezifischem Know-how und Wissen resultieren, dass für die Effizienz und Effektivität regelmäßig anfallender Tätigkeiten im Haushalt von Bedeutung ist<sup>114</sup>. Das Ausmaß dieses Know-hows sowie dessen Auswirkungen auf das Kaufverhalten hängen dabei vom Komplexitätsgrad der Produktnutzung ab. Die hier untersuchte Produktkategorie umfasst Halbfertigprodukte, deren Zubereitung sowohl eine Reihe von Aktivitäten seitens des Konsumenten als auch die Zugabe weiterer Komponenten erfordert. Durch die wiederholte Verwendung einer bestimmten Marke wird der Konsument in die Lage versetzt, das Produkt besser (z. B. bessere Konsistenz oder besserer Geschmack) und/oder effizienter (z. B. schneller) zuzubereiten. Da beim Kauf einer anderen Marke das erworbene markenspezifische Know-how wertlos würde und neues aufgebaut werden müsste, besteht hier ein gewisser Anreiz, der Marke treu zu bleiben<sup>115</sup>. Empirische Befunde, die diese These zumindest indirekt stützen, können aus der Literatur zur Beziehung zwischen der Zufriedenheit und dem Wiederkauf abgeleitet werden<sup>116</sup>. *Mittal/Kamakura* (2001) weisen in ihrer Studie nach, dass ältere Kunden bei gleicher Zufriedenheit in höherem Maße die gleiche Marke wiederkaufen als jüngere Kunden. Das Alter kann dabei als Proxy-Variable für das Ausmaß an Produkterfahrung bzw. akkumuliertem markenspezifischem Know-how angesehen werden. In dem hier analysierten Datensatz ist der Anteil älterer Personen (50-65 Jahre) an den loyalen Konsumenten in der Tat leicht höher als in der nichtloyalen Gruppe. Umgekehrt sind die Jüngeren (18-29 Jahre) stärker in der Gruppe der nichtloyalen Konsumenten vertreten. Einschränkend ist allerdings festzustellen, dass diese Unterschiede statistisch nicht signifikant sind.

---

<sup>113</sup> Zu einer ausführlichen Darstellung verschiedener Formen von Wechselbarrieren vgl. *Burnham/Frels/Mahajan* (2006); *Peter* (2001).

<sup>114</sup> Vgl. *Ratchford* 2001; *Wernerfelt* (1985).

<sup>115</sup> Einschränkend sei angemerkt, dass Konsumenten in einer bestimmten Produktkategorie i. d. R. über ein Portfolio an Marken (z. B. nach *Hauser/Wernerfelt* (1990) durchschnittlich vier) verfügt, aus dem die Markenwahl getroffen wird. Daher ist davon auszugehen, dass die Konsumenten über markenspezifisches Wissen mehrerer konkurrierender Marken verfügen.

<sup>116</sup> Die Zufriedenheit mit einer Marke kann dabei als Antezedenz der Markenstärke angesehen werden.

Weitere Wechselbarrieren können im psychologischen Bereich angesiedelt sein. Konsumenten betrachten den Kauf von Gütern des täglichen Bedarfs häufig nicht als reine Transaktion, sondern bilden über die Jahre zu bestimmten Marken Beziehungen aus, die zumindest partiell denen zu Personen ähneln können<sup>117</sup>. Solche Markenbeziehungen befördern Verhaltensweisen wie höhere Toleranz gegenüber unerfüllten Erwartungen oder Abwertung von Produktalternativen, die beziehungsstabilisierend wirken und zum Wiederkauf der gleichen Marke führen.

Auch wenn die Bereitschaft zur Weiterempfehlung ein anderes Verhalten adressiert als die (Wiederkauf-)Intention<sup>118</sup>, so kann Erstere dennoch als Bestandteil der vorhandenen Einstellungsloyalität<sup>119</sup> des Konsumenten angesehen werden. Daher gehen wir davon aus, dass die zuvor dargestellten Ursachen für die Nichtinvarianz der Kaufintention in analoger Weise auch auf die Weiterempfehlungsabsicht zutrifft.

## **5 Zusammenfassung und Schlussfolgerungen**

Werden Gruppen (z. B. Länder, Kundensegmente) anhand hypothetischer Konstrukte miteinander verglichen, so setzt dies voraus, dass die Beziehungen zwischen den latenten Variablen und ihren Indikatoren bzw. die verwendeten Messmodelle in den betrachteten Gruppen gleich sind. Ist diese Äquivalenzbedingung verletzt, so werden die auf der Konstruktebene vorliegenden Unterschiede zwischen den Gruppen durch die beobachteten Variablen verzerrt wiedergegeben. Dies hat zur Konsequenz, dass sich keine validen Aussagen über Mittelwertunterschiede oder Differenzen in den geschätzten Beziehungen zwischen den latenten Variablen mehr treffen lassen. Die Problematik dabei ist, dass sich Unterschiede in den Messmodellen bei einer separaten Schätzung konfirmatorischer Faktormodelle oder umfassender Strukturgleichungsmodelle nicht aufdecken lassen. Bei Gruppenvergleichen sollte daher immer eine Invarianzüberprüfung mit einem geeigneten Verfahren vorgenommen werden.

Der vorliegende Beitrag zeigt als State-of-the-Art, wie mit der Erweiterung der konfirmatorischen Faktorenanalyse auf den Mehrgruppenfall (und unter Einbeziehung von Mittelwertstrukturen) vier spezifische Formen von Messinvarianz (konfigurale Invarianz, metrische Invarianz, skalare Invarianz, Invarianz der Messfehlervarianzen) überprüft werden

---

<sup>117</sup> Vgl. *Fournier* (1998).

<sup>118</sup> Vgl. z. B. *Söderlund* (2006).

<sup>119</sup> Vgl. *Jacoby/Chestnut* (1978).

können. Insbesondere wird dabei auf den für praktische Anwendungen hochrelevanten Fall eingegangen, dass nicht alle Indikatoren eines Konstruktes die erforderliche Messinvarianz aufweisen. Für die Identifikation dieser Items wird dabei häufig auf Ansätze zurückgegriffen, die – wie ein Beispiel belegt – aufgrund der Verwendung eines fixen Referenzindikators zu falschen Schlussfolgerungen führen können. Eine adäquate Vorgehensweise beim Test auf partielle Invarianz basiert dagegen auf der hier dargestellten systematischen Variation des Referenzindikators. Wird nur eine geringe Anzahl von nichtinvarianten Indikatoren identifiziert, so lassen sich Gruppenunterschiede in den latenten Mittelwerten oder den Strukturbeziehungen durch Freisetzung gruppenspezifischer Parameter in den Strukturgleichungsmodellen valide bestimmen.

Die einzelnen Testschritte werden anhand einer empirischen Studie zur Messäquivalenz einer in der Praxis vielfach genutzten Operationalisierung der Markenstärke bzw. wahrgenommenen Markenattraktivität (Brand Potential Index, BPI<sup>®</sup>) demonstriert. Konkret geht es dabei um die Vergleichbarkeit von Markenstärkemessungen bei unterschiedlich loyalen Konsumenten eines Markenartikels für „Fast moving consumer goods“. Von den insgesamt neun untersuchten Indikatoren der Markenstärke ergeben sich nur für die verhaltensorientierten Variablen Kaufintention und Weiterempfehlungsabsicht gruppenspezifische Faktorladungen; für die Kaufabsicht liegen darüber hinaus auch unterschiedliche Konstanten vor. Damit weist der BPI<sup>®</sup> zumindest eine partielle faktorielle Invarianz auf, die Vergleiche zwischen loyalen und nichtloyalen Konsumenten auf der Ebene der latenten Mittelwerte aber auch im Hinblick auf Strukturbeziehungen mit anderen, messinvarianten latenten Variablen rechtfertigt. Voraussetzung hierfür ist allerdings, dass bei der Mehrgruppenanalyse für die nichtinvarianten Indikatoren gruppenspezifische Parameter zugelassen werden. Darüber hinaus sind direkte Mittelwertvergleiche für die invarianten Indikatoren zulässig; unter praktischen Gesichtspunkten scheint dies, zumindest innerhalb einer bestimmten Bandbreite, auch für die Weiterempfehlungsabsicht sinnvoll. Die beobachteten Differenzen für die Kaufabsicht überschätzen dagegen die „wahren“ Unterschiede der empfundenen Attraktivität der Marke in den beiden Segmenten.

Die Diskussion möglicher Ursachen der fehlenden Messinvarianz zeigt, dass rein statistische Erklärungen wie Nichtnormalverteilung und das ordinale Skalenniveau der Indikatorwerte ausgeschlossen werden können. So bestätigen etwa die Ergebnisse der zusätzlichen Invarianzüberprüfung mit einem Faktormodell für ordinale Indikatoren die Befunde auf Basis des linearen Faktormodells. Da loyale Konsumenten im Falle abnehmender

Markenattraktivität ihre Kaufintention und Weiterempfehlungsabsicht in geringerem Maße als weniger loyale Konsumenten reduzieren, wird als inhaltlich motivierte Ursache die Existenz psychologischer oder ökonomischer Wechselbarrieren vorgeschlagen. Wie allen a posteriori Erklärungsversuchen haftet auch diesen Überlegungen etwas Spekulatives an. Wünschenswert wären daher weitergehende Studien, die die Bedeutung der postulierten Wechselbarrieren für Güter des täglichen Bedarfs detaillierter untersuchen. Darüber hinaus sollte die Invarianz des BPI<sup>®</sup> auch für weitere Marken sowie andere Produktkategorien überprüft werden, um die Generalisierbarkeit unserer empirischen Befunde zu zeigen.

Abschließend ist festzuhalten, dass die empirische Analyse auf einer a priori Segmentierung der befragten Konsumenten anhand ihres Loyalitätsstatus beruht. Trotz einer theoretisch fundierten Gruppierung ist jedoch nicht auszuschließen, dass andere, unbeobachtete Faktoren eine Nichtinvarianz der Markenstärkeindikatoren induzieren können. Um eine solche, unbeobachtete Heterogenität zu berücksichtigen, wären daher als Ergänzung zur hier vorgenommenen Mehrgruppenanalyse Untersuchungen mit dem Ansatz finiter Mischungen von konfirmatorischen Faktormodellen von Interesse<sup>120</sup>.

---

<sup>120</sup> Vgl. z. B. *Lubke/Muthén* (2007); *Yung* (1997).

## Literaturverzeichnis

*Aaker, David A.* (1996), Measuring Brand Equity Across Products and Markets, in: *California Management Review*, Vol. 38, No. 3, S. 102-120.

*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*, Jg. 58, H. 1, S. 2-33.

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

*Baumgartner, Hans/Steenkamp, Jan-Benedict E.M.* (1998), Multi-Group Latent Variable Models for Varying Numbers of Items and Factors with Cross-National and Longitudinal Applications, in: *Marketing Letters*, Vol. 9, No. 1, S. 21-35.

*Bentler, Peter M./Chou, Chih-Ping* (1987), Practical Issues in Structural Equation Modeling, in: *Sociological Methods and Research*, Vol. 16, No. 1, S. 78-116.

*Bollen, Kenneth A.* (1989), *Structural Equations with Latent Variables*.

*Boomsma, Anne/Hoogland, Jeffrey J.* (2001), The Robustness of LISREL Modeling Revisited, in: *Cudeck, Robert/du Toit, Stephen/Sörbom, Dag* (Eds.), *Structural Equation Modeling: Present and Future*, S. 139-168.

*Burnham, Thomas A./Frels, Judy K./Mahajan, Vijay* (2006), Consumer Switching Costs: A Typology, Antecedents, and Consequences, *Journal of the Academy of Marketing Science*, Vol. 31, No. 2, S. 109-126.

*Byrne, Barbara M./Shavelson, Richard J./Muthén, Bengt* (1989), Testing for the Equivalence of Factor Covariance and Mean Structures: The Issue of Partial Measurement Invariance, in: *Psychological Bulletin*, Vol. 105, No. 3, S. 456-466.

*Carle, Adam C./Millsap, Roger E./Cole, David A.* (2008), Measurement Bias Across Gender on the Children's Depression Inventory: Evidence for Invariance From Two Latent Variable Models, in: *Educational and Psychological Measurement*, Vol. 68, No. 2, S. 281-303.

*Cheung, Gordon W./Rensvold, Roger B.* (1999), Testing Factorial Invariance Across Groups: A Reconceptualization and Proposed New Method, in: *Journal of Management*, Vol. 25, No. 1, S. 1-27.

*Cheung, Gordon W./Rensvold, Roger B.* (2002), Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance, in: *Structural Equation Modeling*, Vol. 9, No. 2, S. 233-255.

*De Jong, Martijn G./Steenkamp, Jan-Benedict E.M./Fox, Jean-Paul* (2007), Relaxing Measurement Invariance in Cross-National Consumer Research Using a Hierarchical IRT Model, in: *Journal of Consumer Research*, Vol. 34, August, S. 260-278.

*DeShon, Richard P.* (2004), Measures are Not Invariant Across Groups Without Error Variance Homogeneity, in: *Psychology Science*, Vol. 46, No. 1, S. 137-149.

*DiStefano, Christine* (2002), The Impact of Categorization with Confirmatory Factor Analysis, in: Structural Equation Modeling, Vol. 9, No. 3, S. 327-346.

*Diamantopoulos, Adamantios/Winklhofer, Heidi M.* (2001), Index Construction with Formative Indicators: An Alternative to Scale Development, in: Journal of Marketing Research, Vol. 38, May, S. 269-277.

*Douglas, Susan P./Nijssen, Edwin J.* (2003), On the Use of “Borrowed” Scales in Cross-national Research: A Cautionary Note, in: International Marketing Review, Vol. 20, No. 6, S. 621-642.

*Erdem, Tülin/Swait, Joffre/Valenzuela, Ana* (2006), Brands as Signals: A Cross-Country Validation Study, in: Journal of Marketing, Vol. 70, January, S. 34-49.

*Flora, David B./Curran, Patrick J.* (2004), An Empirical Evaluation of Alternative Methods of Estimation for Confirmatory Factor Analysis With Ordinal Data, in: Psychological Methods, Vol. 9, No. 4, S. 466-491.

*Flowers, Claudia P./Raju, Nambury S./Oshima, T. Chris* (2002), A Comparison of Measurement Equivalence Methods Based on Confirmatory Factor Analysis and Item Response Theory, Paper presented at the Annual Meeting of the National Council on Measurement in Education, New Orleans, LA.

*Fornell, Claes/Larcker, David F.* (1981), Evaluating Structural Equation Models with Unobservable Variables and Measurement Error, in: Journal of Marketing Research, Vol. 18, February, S. 39-50.

*Fournier, Susan* (1998), Consumers and Their Brands: Developing Relationship Theory in Consumer Research, in: Journal of Consumer Research, Vol. 24, March, S. 343-373.

*French, Brian F./Finch, W. Holmes* (2008), Multigroup Confirmatory Factor Analysis: Locating the Invariant Referent Sets, in: Structural Equation Modeling, Vol. 15, No. 1, S. 96-113.

*González-Romá, Vicente/Hernández, Ana/Gómez-Benito, Juana* (2006), Power and Type I Error of the Mean and Covariance Structure Analysis Model for Detecting Differential Item Functioning in Graded Response Items, in: Multivariate Behavioral Research, Vol. 41, No. 1, S. 29-53.

*Grimm, Manuela/Högl, Siegfried/Hupp, Oliver* (2000), Target<sup>®</sup> Positioning – Ein bewährtes Tool zur Unterstützung des strategischen Markenmanagements, in: Jahrbuch der Absatz- und Verbrauchsforschung, 46. Jg., H. 1, S. 4-18.

*Hauser, John R./Wernerfelt, Birger* (1990), An Evaluation Cost Model of Consideration Sets, in: Journal of Consumer Research, Vol. 16, March, S. 393-408.

*Högl, Siegfried/Twardawa, Wolfgang/Hupp, Oliver* (2001), Key Driver starker Marken, in: GWA (Hrsg.), Key Driver starker Marken – Gibt es Regeln für erfolgreiche Marken?, S. 15-59.

*Homburg, Christian/Baumgartner, Hans* (1998), Beurteilung von Kausalmodellen: Bestandsaufnahme und Anwendungsempfehlungen, in: *Hildebrandt, Lutz/Homburg, Christian* (Hrsg.), *Kausalanalyse*, S. 343-369.

*Horn, John L./McArdle, J. Jack* (1992), A Practical and Theoretical Guide to Measurement Invariance in Aging Research, in: *Experimental Aging Research*, Vol. 18, No. 3, S. 117-144.

*Hu, Li-tze/Bentler, Peter M.* (1999), Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria Versus New Alternatives, in: *Structural Equation Modeling*, Vol. 6, No. 1, S. 1-55.

*Huber, Frank/Herrmann, Andreas/Peter, Sibylle* (2003), Ein Ansatz zur Steuerung der Markenstärke. Grundidee, Methodik und Implikationen, in: *Zeitschrift für Betriebswirtschaft*, 73. Jg., H. 4, S. 345-370.

*Hupp, Oliver/Xu, Zhen* (2003), Brand Performance Measurement mit dem Brand Assessment System (BASS), in: *Jahrbuch der Absatz- und Verbrauchsforschung*, 49. Jg., H. 3, S. 220-246.

*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, S. 199-218.

*Jöreskog, Karl G.* (1971), Simultaneous Factor Analysis in Several Populations, in: *Psychometrika*, Vol. 36, No. 4, S. 409-426.

*Krafft, Manfred/Litfin, Thorsten* (2002), Adoption innovativer Telekommunikationsdienste: Validierung der Rogers-Kriterien bei Vorliegen potenziell heterogener Gruppen, in: *Zeitschrift für betriebswirtschaftliche Forschung*, Jg. 54, H. 2, S. 64-83.

*Jacoby, Jacob/Chestnut, Robert W.* (1978), *Brand Loyalty: Measurement and Management*.

*Little, Todd D.* (2000), On the Comparability of Constructs in Cross-Cultural Research - A Critique of Cheung and Rensvold, in: *Journal of Cross-Cultural Psychology*, Vol. 31, No. 2, S. 213-219.

*Lubke, Gitta/Dolan, Conor V.* (2003), Can Unequal Residual Variances Across Groups Mask Differences in Residual Means in the Common Factor Model?, in: *Structural Equation Modeling*, Vol. 10, No. 2, S. 175-192.

*Lubke, Gitta/Dolan, Conor V./Kelderman, Henk/Mellenbergh, Gideon J.* (2003), On the Relationship Between Sources of Within- and Between-Group Differences and Measurement Invariance in the Common Factor Model, in: *Intelligence*, Vol. 31, No. 6, S. 543-566.

*Lubke, Gitta/Muthén, Bengt O.* (2004), Applying Multigroup Confirmatory Factor Models for Continuous Outcomes to Likert Scale Data Complicates Meaningful Group Comparisons, in: *Structural Equation Modeling*, Vol. 11, No. 4, S. 514-534.

*Lubke, Gitta/Muthén, Bengt O.* (2007), Performance of Factor Mixture Models as a Function of Model Size, Covariate Effects, and Class-Specific Parameters, in: *Structural Equation Modeling*, Vol. 14, No. 1, S. 26-47.

*MacCallum, Robert. C./Roznowski, Mary/Necowitz, Lawrence B. (1992), Model Modifications in Covariance Structure Analysis: The Problem of Capitalization on Chance, in: Psychological Bulletin, Vol. 111, No. 3, S. 490-504.*

*Meade, Adam W./Bauer, Daniel J. (2007), Power and Precision in Confirmatory Factor Analytic Tests of Measurement Invariance, in: Structural Equation Modeling, Vol. 14, No. 4, S. 611-635.*

*Meade, Adam W./Lautenschlager, Gary J. (2004a), A Monte-Carlo Study of Confirmatory Factor Analytic Tests of Measurement Equivalence/Invariance, in: Structural Equation Modeling, Vol. 11, No. 1, S. 60-72.*

*Meade, Adam W./Lautenschlager, Gary J. (2004b), A Comparison of Item Response Theory and Confirmatory Factor Analytic Methodologies for Establishing Measurement Equivalence/Invariance, in: Organizational Research Methods, Vol. 7, No. 4, S. 361-388.*

*Meredith, William (1993), Measurement Invariance, Factor Analysis and Factorial Invariance, in: Psychometrika, Vol. 58, No. 4, S. 525-543.*

*Meredith, William/Horn, John (2001), The Role of Factorial Invariance in Modeling Growth and Change, in: Collins, Linda M./Sayer, Aline G. (Eds.), New Methods for the Analysis of Change, S. 203-240.*

*Millsap, Roger E./Kwok, Oi-Man (2004), Evaluating the Impact of Partial Factorial Invariance on Selection in Two Populations, in: Psychological Methods, Vol. 9, No. 1, S. 93-115.*

*Millsap, Roger E./Yun-Tein, Jenn (2004), Assessing Factorial Invariance in Ordered-Categorical Measures, in: Multivariate Behavioral Research, Vol. 39, No. 3, S. 479-515.*

*Mittal, Vikas/Kamakura, Wagner A. (2001), Satisfaction, Repurchase Intent, and Repurchase Behavior: Investigating the Moderating Effect of Customer Characteristics, in: Journal of Marketing Research, Vol. 38, February, S. 131-142.*

*Morschett, Dirk (2002), Konzeptualisierung und Operationalisierung des Markenwerts von Einkaufsstätten, in: Marketing ZFP, 24. Jg., H. 4, S. 277-292.*

*Muthén, Bengt O. (1983), Latent Variable Structural Equation Modeling With Categorical Data, in: Journal of Econometrics, Vol. 22, No. 1-2, S. 48-65.*

*Muthén, Bengt O. (1984), A General Structural Equation Model With Dichotomous, Ordered Categorical, and Continuous Latent Variable Indicators, in: Psychometrika, Vol. 49, No. 1, S. 115-132.*

*Muthén, Bengt O./Asparouhov, Tihomir (2002), Latent Variable Analysis With Categorical Outcomes: Multi-Group and Growth Modeling in Mplus, Mplus Web Note # 4, Los Angeles, CA: <http://www.statmodel.com/mplus/examples/webnotes/CatMGLong.pdf>.*

*Muthén, Linda K./Muthén, Bengt O. (1998), Mplus User's Guide.*

*Muthén, Linda K./Muthén, Bengt O. (1998-2007), Mplus Version 5.*



*Muthén, Bengt O./du Toit, Stephen H.C./Spisic, Damir (1997), Robust Inference Using Weighted Least Squares and Quadratic Estimating Equations in Latent Variable Modeling With Categorical and Continuous Outcomes, unveröffentlichtes Manuskript.*

*Netemeyer, Richard G./Durvasula, Srinivas/Lichtenstein, Donald R. (1991), A Cross-National Assessment of the Reliability and Validity of the CETSCALE, in: Journal of Marketing Research, Vol. 28, August, S. 320-327.*

*Netemeyer, Richard G./Brashear-Alejandro, Thomas/Boles, James S. (2004), A Cross-National Model of Job-Related Outcomes of Work Role and Family Role Variables: A Retail Sales Context, in: Journal of the Academy of Marketing Science, Vol. 32, No. 1, S. 49-60.*

*Peter, Sibylle I. (2001), Kundenbindung als Marketingziel, 2. Aufl.*

*Ployhart, Robert E./Oswald, Frederick L. (2004), Applications of Mean and Covariance Structure Analysis: Integrating Correlational and Experimental Approaches, in: Organizational Research Methods, Vol. 7, No. 1, S. 27-65.*

*Raju, Nambury S./Laffitte, Larry J./Byrne, Barbara M. (2002), Measurement Equivalence: A Comparison of Methods Based on Confirmatory Factor Analysis and Item Response Theory, in: Journal of Applied Psychology, Vol. 87, No. 3, S. 517-529.*

*Ratchford, Brian T. (2001), The Economics of Consumer Knowledge, in: Journal of Consumer Research, Vol. 27, March, S. 397-411.*

*Reise, Steven P./Widaman, Keith F./Pugh, Robin H. (1993), Confirmatory Factor Analysis and Item Response Theory: Two Approaches for Exploring Measurement Invariance, in: Psychological Bulletin, Vol. 114, No. 3, S. 552-566.*

*Rensvold, Roger B. (2002), Metric Equivalence/Invariance Across Multiple Groups: Comparing Apples with Apples with Apples etc., Working Paper, City University of Hong Kong.*

*Rensvold, Roger B./Cheung, Gordon W. (1998), Testing Measurement Models for Factorial Invariance: A Systematic Approach, in: Educational and Psychological Measurement, Vol. 58, No. 6, S. 1017-1034.*

*Rensvold, Roger B./Cheung, Gordon W. (2001), Testing for Metric Invariance Using Structural Equation Models: Solving the Standardization Problem, in: Schriesheim, Chester A./Neider, Linda, L. (Eds.), Research in Management: Equivalence in Measurement, Vol. 1, S. 25-50.*

*Rossiter, John R. (2002), The C-OAR-SE Procedure for Scale Development in Marketing, in: International Journal of Research in Marketing, Vol. 19, No. 4, S. 305-335.*

*Samejima, F. (1969), Estimation of Latent Ability Using a Response Pattern of Graded Scores, in: Psychometrika Monograph, Vol. 17.*

*Sattler, Henrik/Högl, Siegfried/Hupp, Oliver (2002), Evaluation of the Financial Value of Brands, Proceedings of the 54. Annual Conference of the European Society for Opinion and Marketing Research (ESOMAR), Barcelona, Spanien, 481-502.*

*Satorra, Albert/Bentler, Peter M. (1988), Scaling Corrections for Chi-Square Statistics in Covariance Structure Analysis, in: ASA 1988 Proceedings of the Business and Economic Statistics Section, S. 308-313.*

*Satorra, Albert/Bentler, Peter M. (2001), A Scaled Difference Chi-Square Test Statistic for Moment Structure Analysis, in: Psychometrika, Vol. 66, No. 4, S. 507-514.*

*Scholderer, Joachim/Grunert, Klaus G./Brunsø, Karen (2005), A Procedure for Eliminating Additive Bias from Cross-Cultural Survey Data, in: Journal of Business Research, Vol. 58, No. 1, S. 72-78.*

*Shimp, Terence A./Sharma, Subhash (1987), Consumer Ethnocentrism: Construction and Validation of the CETSCALE, in: Journal of Marketing Research, Vol. 24, August, S. 280-289.*

*Singh, Jagdip (2004), Tackling Measurement Problems with Item Response Theory: Principles, Characteristics, and Assessment, with an Illustrative Example, in: Journal of Business Research, Vol. 57, No. 2, S. 184-208.*

*Söderlund, Magnus (2006), Measuring Customer Loyalty with Multi-Item Scales: A Case for Caution, in: International Journal of Service Industry Management, Vol. 17, No. 1, S. 76-98.*

*Sörbom, Dag (1974), A General Method for Studying Differences in Factor Means and Factor Structure Between Groups, in: British Journal of Statistical Psychology, Vol. 27, S. 137-163.*

*Stark, Steven/Chernyshenko, Oleksandr S./Drasgow, Fritz (2006), Detecting Differential Item Functioning with Confirmatory Factor Analysis and Item Response Theory? Toward a Unified Strategy, in: Journal of Applied Psychology, Vol 91, No. 6, S. 1292-1306.*

*Steenkamp, Jan-Benedict E.M./Baumgartner, Hans (1998), Assessing Measurement Invariance in Cross-National Consumer Research, in: Journal of Consumer Research, Vol. 25, June, S. 78-90.*

*Temme, Dirk (2006a), Assessing Measurement Invariance of Ordinal Indicators in Cross-National Research, in: Diehl, Sandra/Terlutter, Ralf (Eds.), International Advertising and Communication – Current Insights and Empirical Findings, S. 455-472.*

*Temme, Dirk (2006b), Die Spezifikation und Identifikation formativer Messmodelle der Marketingforschung in Kovarianzstrukturanalysen, in: Marketing ZFP, Jg. 28, H. 3, S. 183-196.*

*Vandenberg, Robert J./Lance, Charles E. (2000), A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research, in: Organizational Research Methods, Vol. 3, No. 1, S. 4-70.*

*Van de Vijver, Fons/Leung, Kwok (1997), Methods and Data Analysis for Cross-Cultural Research.*

*Wernerfelt, Birger (1985), Brand Loyalty and User Skills, in: Journal of Economic Behavior and Organization, Vol. 6, S. 381-385.*

*West, Stephen G./Finch, John F./Curran, Patrick J. (1995), Structural Equation Models With Nonnormal Variables: Problems and Remedies, in: Hoyle, Rick H. (Ed.), Structural Equation Modeling: Concepts, Issues, and Applications, S. 56-75.*

*Wicherts, Jelte M./Dolan, Conor V. (2004), A Cautionary Note on the Use of Information Fit Indexes in Covariance Structure Modeling With Means, in: Structural Equation Modeling, Vol. 11, No. 1, S. 45-50.*

*Yoo, Boonghee (2002), Cross-Group Comparisons: A Cautionary Note, in: Psychology & Marketing, Vol. 19, No. 4, S. 357-368.*

*Yoon, Myeongsun/Millsap, Roger E. (2007), Detecting Violations of Factorial Invariance Using Data-Based Specification Searches: A Monte Carlo Study, in: Structural Equation Modeling, Vol. 14, No. 3, S. 435-463.*

*Yu, Ching-Yun (2002), Evaluating Cutoff Criteria of Model Fit Indices for Latent Variable Models with Binary and Continuous Outcomes, Dissertation, University of California at Los Angeles (UCLA).*

*Yung, Yiu-Fai (1997), Finite Mixtures in Confirmatory Factor-Analysis Models, in: Psychometrika, Vol. 62, No. 3, S. 297-330.*

## Anhang: Informationen zur empirischen Studie

Tabelle A1: Indikatoren des BPI<sup>®</sup> nach GfK und Variablenbezeichnungen in der Studie<sup>121</sup>

Indikatoren des BPI <sup>®</sup>	Variablenbezeichnung
<b>Kognitive Facette</b>	
Qualität	qualitativ
Uniqueness	positiv
<b>Affektive Facette</b>	
Markenidentifikation	identifizieren
Markensympathie	mögen
Markenvertrauen	vertrauen
Markenloyalität	vermissen
<b>Konative Facette</b>	
Kaufabsicht	kaufen
Mehrpreisakzeptanz	ausgeben
Bereitschaft zur Weiterempfehlung	empfehlen

<sup>121</sup> Aus Gründen der Vertraulichkeit können die exakten Formulierungen an dieser Stelle nicht wiedergegeben werden.

*Tabelle A2: Kovarianzmatrizen und Mittelwerte für die loyalen und nichtloyalen Konsumenten*

<b>Variablen</b>	quali- tativ	positiv	identi- fizieren	mögen	ver- trauen	ver- missen	kaufen	ausge- ben	emp- fehlen
<b>Gruppe 1</b>									
<i>Kovarianzen</i>									
qualitativ	1,845								
positiv	1,379	1,729							
identifizieren	1,531	1,491	2,579						
mögen	1,515	1,415	1,711	2,122					
vertrauen	1,550	1,445	1,714	1,694	2,130				
vermissen	1,556	1,513	1,881	1,724	1,782	2,609			
kaufen	1,168	1,150	1,212	1,302	1,169	1,423	1,765		
ausgeben	1,426	1,387	1,734	1,508	1,620	1,751	1,030	3,431	
empfehlen	1,510	1,378	1,653	1,645	1,649	1,722	1,231	1,455	1,928
<i>Mittelwerte</i>	5,437	5,483	5,024	5,332	5,329	5,315	5,917	4,611	5,459
<b>Gruppe 2</b>									
<i>Kovarianzen</i>									
qualitativ	2,495								
positiv	1,864	2,363							
identifizieren	1,740	1,702	2,845						
mögen	1,944	1,857	2,082	2,595					
vertrauen	1,923	1,763	2,002	2,003	2,606				
vermissen	1,839	1,943	2,298	2,127	2,044	3,385			
kaufen	1,711	1,699	1,864	1,987	1,870	2,226	3,311		
ausgeben	1,546	1,586	2,053	1,817	1,747	2,159	1,848	3,012	
empfehlen	2,133	1,940	2,147	2,167	2,204	2,309	2,199	2,008	3,017
<i>Mittelwerte</i>	4,158	4,148	3,436	3,812	3,870	3,607	4,236	3,028	3,967

Gruppe 1 = loyale Konsumenten, Gruppe 2 = nichtloyale Konsumenten

## SFB 649 Discussion Paper Series 2008

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

- 001 "Testing Monotonicity of Pricing Kernels" by Yuri Golubev, Wolfgang Härdle and Roman Timonfeev, January 2008.
- 002 "Adaptive pointwise estimation in time-inhomogeneous time-series models" by Pavel Cizek, Wolfgang Härdle and Vladimir Spokoiny, January 2008.
- 003 "The Bayesian Additive Classification Tree Applied to Credit Risk Modelling" by Junni L. Zhang and Wolfgang Härdle, January 2008.
- 004 "Independent Component Analysis Via Copula Techniques" by Ray-Bing Chen, Meihui Guo, Wolfgang Härdle and Shih-Feng Huang, January 2008.
- 005 "The Default Risk of Firms Examined with Smooth Support Vector Machines" by Wolfgang Härdle, Yuh-Jye Lee, Dorothea Schäfer and Yi-Ren Yeh, January 2008.
- 006 "Value-at-Risk and Expected Shortfall when there is long range dependence" by Wolfgang Härdle and Julius Mungo, January 2008.
- 007 "A Consistent Nonparametric Test for Causality in Quantile" by Kiho Jeong and Wolfgang Härdle, January 2008.
- 008 "Do Legal Standards Affect Ethical Concerns of Consumers?" by Dirk Engelmann and Dorothea Kübler, January 2008.
- 009 "Recursive Portfolio Selection with Decision Trees" by Anton Andriyashin, Wolfgang Härdle and Roman Timofeev, January 2008.
- 010 "Do Public Banks have a Competitive Advantage?" by Astrid Matthey, January 2008.
- 011 "Don't aim too high: the potential costs of high aspirations" by Astrid Matthey and Nadja Dwenger, January 2008.
- 012 "Visualizing exploratory factor analysis models" by Sigbert Klinke and Cornelia Wagner, January 2008.
- 013 "House Prices and Replacement Cost: A Micro-Level Analysis" by Rainer Schulz and Axel Werwatz, January 2008.
- 014 "Support Vector Regression Based GARCH Model with Application to Forecasting Volatility of Financial Returns" by Shiyi Chen, Kiho Jeong and Wolfgang Härdle, January 2008.
- 015 "Structural Constant Conditional Correlation" by Enzo Weber, January 2008.
- 016 "Estimating Investment Equations in Imperfect Capital Markets" by Silke Hüttel, Oliver Mußhoff, Martin Odening and Nataliya Zynych, January 2008.
- 017 "Adaptive Forecasting of the EURIBOR Swap Term Structure" by Oliver Blaskowitz and Helmut Herwatz, January 2008.
- 018 "Solving, Estimating and Selecting Nonlinear Dynamic Models without the Curse of Dimensionality" by Viktor Winschel and Markus Krätzig, February 2008.
- 019 "The Accuracy of Long-term Real Estate Valuations" by Rainer Schulz, Markus Staiber, Martin Wersing and Axel Werwatz, February 2008.
- 020 "The Impact of International Outsourcing on Labour Market Dynamics in Germany" by Ronald Bachmann and Sebastian Braun, February 2008.
- 021 "Preferences for Collective versus Individualised Wage Setting" by Tito Boeri and Michael C. Burda, February 2008.

**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".



- 022 "Lumpy Labor Adjustment as a Propagation Mechanism of Business Cycles" by Fang Yao, February 2008.
- 023 "Family Management, Family Ownership and Downsizing: Evidence from S&P 500 Firms" by Jörn Hendrich Block, February 2008.
- 024 "Skill Specific Unemployment with Imperfect Substitution of Skills" by Runli Xie, March 2008.
- 025 "Price Adjustment to News with Uncertain Precision" by Nikolaus Hautsch, Dieter Hess and Christoph Müller, March 2008.
- 026 "Information and Beliefs in a Repeated Normal-form Game" by Dietmar Fehr, Dorothea Kübler and David Danz, March 2008.
- 027 "The Stochastic Fluctuation of the Quantile Regression Curve" by Wolfgang Härdle and Song Song, March 2008.
- 028 "Are stewardship and valuation usefulness compatible or alternative objectives of financial accounting?" by Joachim Gassen, March 2008.
- 029 "Genetic Codes of Mergers, Post Merger Technology Evolution and Why Mergers Fail" by Alexander Cuntz, April 2008.
- 030 "Using R, LaTeX and Wiki for an Arabic e-learning platform" by Taleb Ahmad, Wolfgang Härdle, Sigbert Klinke and Shafeeqah Al Awadhi, April 2008.
- 031 "Beyond the business cycle – factors driving aggregate mortality rates" by Katja Hanewald, April 2008.
- 032 "Against All Odds? National Sentiment and Wagering on European Football" by Sebastian Braun and Michael Kvasnicka, April 2008.
- 033 "Are CEOs in Family Firms Paid Like Bureaucrats? Evidence from Bayesian and Frequentist Analyses" by Jörn Hendrich Block, April 2008.
- 034 "JBendge: An Object-Oriented System for Solving, Estimating and Selecting Nonlinear Dynamic Models" by Viktor Winschel and Markus Krätzig, April 2008.
- 035 "Stock Picking via Nonsymmetrically Pruned Binary Decision Trees" by Anton Andriyashin, May 2008.
- 036 "Expected Inflation, Expected Stock Returns, and Money Illusion: What can we learn from Survey Expectations?" by Maik Schmeling and Andreas Schrimpf, May 2008.
- 037 "The Impact of Individual Investment Behavior for Retirement Welfare: Evidence from the United States and Germany" by Thomas Post, Helmut Gründl, Joan T. Schmit and Anja Zimmer, May 2008.
- 038 "Dynamic Semiparametric Factor Models in Risk Neutral Density Estimation" by Enzo Giacomini, Wolfgang Härdle and Volker Krätschmer, May 2008.
- 039 "Can Education Save Europe From High Unemployment?" by Nicole Walter and Runli Xie, June 2008.
- 042 "Gruppenvergleiche bei hypothetischen Konstrukten – Die Prüfung der Übereinstimmung von Messmodellen mit der Strukturgleichungsmethodik" by Dirk Temme and Lutz Hildebrandt, June 2008.

**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".

