

Ohlwein, Martin

Working Paper

Die Prüfung der globalen Güte eines Kausalmodells auf Stabilität mit Hilfe eines nichtparametrischen Bootstrap-Algorithmus

ISM Working Paper, No. 2

Provided in Cooperation with:

International School of Management (ISM), Dortmund

Suggested Citation: Ohlwein, Martin (2015) : Die Prüfung der globalen Güte eines Kausalmodells auf Stabilität mit Hilfe eines nichtparametrischen Bootstrap-Algorithmus, ISM Working Paper, No. 2, ISBN 978-3-95645-508-7, Verlagshaus Monsenstein und Vannerdat, Münster

This Version is available at:

<https://hdl.handle.net/10419/313077>

Standard-Nutzungsbedingungen:

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

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

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

Terms of use:

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

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

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

ISM Workingpaper No. 2

Martin Ohlwein

Die Prüfung der globalen
Güte eines Kausalmodells
auf Stabilität mit Hilfe
eines nichtparametrischen
Bootstrap-Algorithmus

Martin Ohlwein

**Die Prüfung der globalen Güte eines
Kausalmodells auf Stabilität mit Hilfe eines nicht-
parametrischen Bootstrap-Algorithmus**

Ohlwein, Martin: Die Prüfung der globalen Güte eines Kausalmodells auf Stabilität mit Hilfe eines nichtparametrischen Bootstrap-Algorithmus

© 2015 der vorliegenden Ausgabe, Verlagshaus Monsenstein und Vannerdat OHG Münster

www.mv-wissenschaft.de

© 2015 ISM Dortmund

Alle Rechte vorbehalten

Herstellung: MV-Verlag

ISBN 978-3-95645-508-7

ISM - International School of Management gGmbH

Otto-Hahn-Str. 19 · 44227 Dortmund

www.ism.de

Tel.: 0231.975139-0 · Fax: 0231.975139-39

ism.dortmund@ism.de

Ohlwein, Martin: Die Prüfung der globalen Güte eines Kausalmodells auf Stabilität mit Hilfe eines nichtparametrischen Bootstrap-Algorithmus, Dortmund und Münster, MV-Verl., 2015 (Working Paper ; 2)

ISBN 978-3-95645-508-7

Inhaltsverzeichnis

Abbildungsverzeichnis	IV
Tabellenverzeichnis.....	IV
1 Das Spannungsfeld zwischen wissenschaftstheoretischem Bezugsrahmen und Kovarianzstrukturanalyse.....	1
2 Das Resampling von Stichproben mit Hilfe der Bootstrap-Technik	3
2.1 Der Grundgedanke und der Algorithmus des Bootstrap-Ansatzes.....	3
2.2 Ein Schema zur Prüfung der Stabilität von Gütekriterien eines Kausalmodells	5
3 Problemfelder einer Verknüpfung von Bootstrap-Methode und Kovarianzstrukturanalyse.....	5
3.1 Der Bootstrap-Schätzungen innewohnende Bias	5
3.1.1 Die Ursache der Verzerrung.....	5
3.1.2 Die Modifikation des Resampling-Raumes als Lösungsansatz.....	6
3.1.3 Mit einer Transformation der Daten verbundene Probleme.....	7
3.2 Die Charakterisierung der empirischen Verteilung einer Bootstrap-Statistik.....	9
4 Das Erkenntnispotential eines nichtparametrischen Bootstrapping bei der Prüfung der Stabilität globaler Fit-Maße	11
4.1 Die der Analyse zugrundeliegende Untersuchung.....	11
4.1.1 Anliegen und Design der Studie	11
4.1.2 Ein Kausalmodell zur Erklärung der Einstellung zum Verkauf eines gebrauchten Fernsehapparates.....	11
4.2 Die Prüfung der Stabilität globaler Gütekriterien mittels des Bootstrap-Ansatzes	12
4.2.1 Das für die Analyse gewählte Vorgehen	12
4.2.2 Die zutage geförderten Befunde.....	14
4.2.3 Konsequenzen der empirischen Befunde	18
5 Ansatzpunkte für weiterführende Forschungsarbeiten.....	19
Literaturverzeichnis	20

Abbildungsverzeichnis

Abbildung 1:	Ein Kausalmodell zur Erklärung der Einstellung zum Verkauf eines gebrauchten Fernsehapparates.....	13
--------------	---	----

Tabellenverzeichnis

Tabelle 1:	Die Güte des empirisch entwickelten Modells widerspiegelnde Fit-Maße	12
Tabelle 2:	Kenngroßen der Verteilung der betrachteten Gütemaße	14

Abstract english

Scientific realism has become exceedingly popular as a frame of reference, especially within research drawing upon structural equation models. This concept allows for modification of a causal model derived from theoretical considerations based on preliminary empirical findings. This is significant as such an approach increases the risk of ending up with a model that misleadingly signals a high degree of validity and reliability. It is, therefore, imperative to interrogate whether the indices which reflect the fit of the model are consistent with alternative samples. This paper addresses the aforementioned problem. Additionally, the goal here is to examine to what extent the bootstrapping technique contributes to the evaluation of the stability of global fit indices.

1 Das Spannungsfeld zwischen wissenschaftstheoretischem Bezugsrahmen und Kovarianzstrukturanalyse

Jedes Forschungskonzept, dem eine anwendungsorientierte Fragestellung zugrunde liegt, bedarf einer wissenschaftstheoretischen Verankerung. Eines der populärsten Orientierungssysteme verkörpert der im Wesentlichen durch die Schriften Poppers geprägte Kritische Rationalismus (vgl. Popper 1963: 42 ff.; Popper 1984: 47 ff.; Popper 1991: 41 ff. und 65 ff.; Meyer 1979: 29 ff.; Raffée, Abel 1979: 3 ff.; Albert 1980: 12 ff.; Spinner 1982: 63 ff.; Pähler 1986: 56 ff.; Raffée 1993: 18 ff.).

Diesem Ansatz zufolge kann die Überprüfung eines auf der Grundlage theoretischer Überlegungen abgeleiteten Beziehungsgeflechts niemals in dessen Verifikation münden, sondern allenfalls zur Widerlegung des Modells führen (Falsifikationsprinzip). Lässt sich eine Hypothese nicht verwerfen, so gilt sie nicht als bestätigt, sondern lediglich als (vorläufig) bewährt. Der Erkenntnisfortschritt in der Wissenschaft vollzieht sich innerhalb dieses Rahmens als iterativer, über die deduktive Formulierung von Ursache-Wirkungs-Zusammenhängen gesteuerter Prozess. Als forschungsmethodologisch unzulässig erweist sich hingegen die induktive Schlussweise.

In den letzten Jahren verstärkt sich jedoch die Kritik am Kritischen Rationalismus. Seine Opponenten führen insbesondere an, dass das beschriebene Prinzip den faktischen Gegebenheiten der Sozialwissenschaften ungenügend Rechnung trage und folglich als wissenschaftstheoretische Leitlinie nur bedingt hilfreich sei¹. Insbesondere im Kontext kausalanalytischer Studien ging mit dieser Kritik eine

¹ Zur Kritik am Kritischen Rationalismus im Einzelnen siehe Kubicek 1975: 48 ff.; Witte, Grün, Bronner 1975: 797; Witte 1977: 272; Anderson 1983: 21; Deshpande 1983: 105 ff.; Arndt 1985: 14; Martin 1989: 19 ff. und 23; Homburg 2000: 62 ff.

Öffnung gegenüber dem Wissenschaftlichen Realismus² einher. Eine umfassende Darstellung dieses wissenschaftstheoretischen Ansatzes liefert Hunt 1984; 1991. Dieses Konzept führt deduktive und induktive Erkenntnisfindung zusammen, indem es die sukzessive Modifikation eines auf der Basis theoretischer Überlegungen abgeleiteten Hypothesengeflechts wissenschaftstheoretisch legitimiert. Gleichwohl setzt die Anforderung, „hypothesizing after the results are known“ (HARKing) zu vermeiden, enge wissenschaftliche Grenzen (vgl. Kerr 1998: 197 ff.; Diekmann 2011: 631 f.).

Die hierdurch gewonnene zusätzliche Gestaltungsfreiheit bei der Modellbildung erschwert es allerdings beträchtlich, die Güte eines Kausalmodells zu bewerten. Bei einer sukzessiven Überarbeitung eines Hypothesensystems droht der Forscher zu einem Beziehungsgeflecht zu gelangen, das zwar in Bezug auf das vorliegende Datenmaterial eine hohe Validität und Reliabilität signalisiert, aber im Hinblick auf die Grundgesamtheit nicht über solche verfügt (vgl. Temme, Hildebrandt 2009: 167). Infolgedessen ist nach Abschluss der Phase der Modellmodifikation die Stabilität jener Werte zu testen, die die Güte des Hypothesengeflechts widerspiegeln. Hierzu bedient man sich häufig der Methode der Known group-Validierung (vgl. zur Vorgehensweise bei einer Known group-Validierung Müller 1991: 218 ff.). Dieser liegt die Idee zugrunde, ein kausalanalytisches Modell auf der Basis einer Explorationsstichprobe zu entwickeln und in einem zweiten Schritt an einem Validierungssample zu überprüfen.

Ein solches Vorgehen ist jedoch nicht frei von Problemen. So führt eine Zerlegung der Gesamtstichprobe in zwei Teile unweigerlich zu einem Informationsverlust, sofern man sich bei der Schätzung der Parameter nur eines der beiden Teilsamples bedient (vgl. Berekoven, Eckert, Ellenrieder 2009: 57 ff.). An Brisanz gewinnt dieser Mangel im Zusammenhang mit Linearen Strukturgleichungsmodellen wegen der Konsistenz-eigenschaften der Schätzverfahren. Als Folge hiervon können die Befunde nachhaltig an Präzision verlieren, wenn die Berechnungen auf einer Stichprobe geringen Umfangs³ beruhen (vgl. Tanaka 1987: 136; Balderjahn 1986: 155). Da sozialwissenschaftlichen Untersuchungen häufig Datensätze zugrunde liegen, die vergleichsweise wenige Fälle umfassen, erweist sich der skizzierte Validierungsansatz somit oftmals als unangemessen.

Angesichts der thematisierten Probleme stellt sich die Frage nach einer methodischen Alternative. Eine solche verkörpert der Bootstrap-Ansatz⁴. An diesem

² Als Vertreter dieses Konzepts gelten unter anderem Sellars 1963; Causey 1979; MacKinnon 1979; Boyd 1984; Levin 1984; McMullin 1984; Harré 1986; Leplin 1986.

³ Simulationsstudien zufolge liefert beispielsweise das häufig verwendete Maximum Likelihood-Schätzverfahren für Stichproben, die dessen Verteilungsannahmen genügen, ab einem Umfang von $n = 200$ stabile Werte.

⁴ Einen Einblick in die Vorgehensweise und die Leistungsfähigkeit des Bootstrap-Verfahrens vermittelt Gutsche 1994.

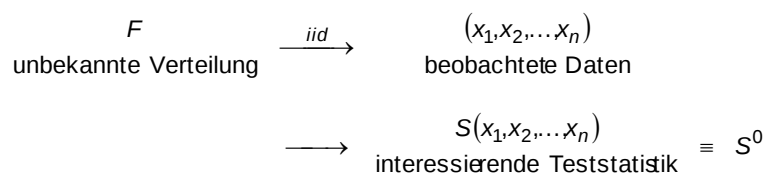
Punkt knüpft die vorliegende Arbeit an. Ihr liegt das Ziel zugrunde zu überprüfen, inwieweit die Technik des Bootstrapping die Stabilität von Kriterien zu evaluieren erlaubt, die die globale Güte eines Kausalmodells widerspiegeln.

Ausgehend von einer Darstellung des Bootstrap-Algorithmus in seiner allgemeinen Form ist dieser zunächst auf die im Mittelpunkt stehende Fragestellung zu übertragen. In einem nächsten Schritt sind jene Unwägbarkeiten zu diskutieren, die aus einer Zusammenführung des Bootstrap-Ansatzes und der Kovarianzstrukturanalyse resultieren. Anhand eines empirischen Beispiels soll dann das Erkenntnispotential untersucht werden, über das ein nichtparametrisches Bootstrapping verfügt, wenn es die Stabilität der globalen Güte eines LISREL-Modells zu prüfen gilt. Eine Zusammenfassung der Befunde sowie eine Identifikation von Ansatzpunkten für weiterführende Forschungsaktivitäten runden die Ausführungen ab.

2 Das Resampling von Stichproben mit Hilfe der Bootstrap-Technik

2.1 Der Grundgedanke und der Algorithmus des Bootstrap-Ansatzes

Das Anliegen des Bootstrap-Verfahrens besteht darin, die unbekannte Verteilung einer Teststatistik mit Hilfe von Stichprobendaten zu ermitteln. Den Ausgangspunkt markiert eine, ebenfalls unbekannte, Wahrscheinlichkeitsverteilung F , aus der per Zufallsauswahl ein Sample vom Umfang n gezogen wird. Dieses Vorgehen definiert eine n -dimensionale Zufallsstichprobe $X = (X_1, X_2, \dots, X_n)$ mit unabhängig identisch verteilten Stichprobenvariablen X_i (vgl. Efron, Tibshirani 1986: 55). Verkörpern $X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$ die beobachteten Realisationen, so kann man mit diesen den Wert S^0 der Teststatistik $S(x_1, x_2, \dots, x_n)$ berechnen. Das skizzierte Prinzip spiegelt sich im folgenden Schema wider (vgl. Efron 1990: 79).



Des auf diese Art gewonnenen Datenmaterials bedient sich die Bootstrap-Methode, deren Simulationsalgorithmus aus den fünf folgenden Schritten besteht (vgl. Efron, Gong 1983: 36 f.; Efron, Tibshirani 1986: 56; Efron 1990: 79 f.; Gutsche 1995: 169).

- (1) Zunächst wird jeder Realisation x_i , mit $i = 1, 2, \dots, n$, die Wahrscheinlichkeit n^{-1} zugeordnet (Laplace-Raum) und hierdurch die Experimentalpopulation $\hat{F}$ definiert.
- (2) Aus der Verteilung $\hat{F}$ zieht man m -mal mit Zurücklegen und gelangt dadurch zu einem m -dimensionalen Bootstrap-Sample $\hat{x}^j = (\hat{x}_1^j, \hat{x}_2^j, \dots, \hat{x}_m^j)$ (j verkörpert

einen Index zur Kennzeichnung der einzelnen Bootstrap-Replikationen. Vgl. hierzu auch Schritt 4). Bedingt durch die Struktur des Zufallsgenerators kann jede in $\hat{F}$ enthaltene Ausprägung x_i kein einziges Mal, genau einmal oder mehrfach in der Bootstrap-Stichprobe vorkommen (Aus der methodischen Sicht sind alle Größenrelationen zwischen m und n zulässig. Für $m = n$ spricht, dass es in den Fällen $m < n$ und $m > n$ zu einem Konvergenzproblem kommen kann. Vgl. Bickel, Freedman 1981: 1197; Efron, Tibshirani 1986: 56. Dessen ungeachtet erweist es sich bei einigen Fragestellungen als vorteilhaft, m kleiner als n zu wählen. Vgl. Bickel, Götze, Zwet 1994; Yung, Bentler 1996: 198. Bei einzelnen Kenngrößen, die die Güte eines Kausalmodells widerspiegeln, gilt es zu berücksichtigen, dass sie mit dem Stichprobenumfang variieren. Folglich liegt es nahe, die Größe des Bootstrap-Samples so festzulegen, dass sie dem Umfang der aus der Grundgesamtheit gezogenen Stichprobe gleicht.).

- (3) Für die Bootstrap-Stichprobe $\hat{x}^j$ wird der Wert der interessierenden Teststatistik als $S(\hat{x}_1^j, \hat{x}_2^j, \dots, \hat{x}_m^j) = \hat{S}_j$ berechnet. Da $\hat{x}^j$ nicht aus der Grundgesamtheit F , sondern aus der Experimentalpopulation $\hat{F}$ stammt, bezeichnet man $S(\hat{x}^j)$ als Bootstrap-Teststatistik.
- (4) Wiederholt man die Schritte 2 und 3 ($b - 1$)-mal, gelangt man zu b statistisch unabhängigen Bootstrap-Stichproben $\hat{x}^1, \hat{x}^2, \dots, \hat{x}^b$ und damit auch zu b Realisationen der Bootstrap-Teststatistik $\hat{S}_j = S(\hat{x}^j)$ mit $j = 1, 2, \dots, b$. (Dabei wählt man b aus forschungsökonomischen Gründen in der Regel so, dass die Anzahl der Bootstrap-Replikationen die Zahl theoretisch denkbarer Kombinationen von n Objekten, n^n , beträchtlich unterschreitet, vgl. Yung, Bentler 1996: 197 f.). Eine solche approximative Ermittlung der Bootstrap-Verteilung hat sich als eine praktikable Vorgehensweise bewährt. Sie wirft jedoch unmittelbar die Frage auf, welcher Wert von b als angemessen gelten kann. Hierfür existieren bislang lediglich Faustregeln, wobei Konsens darüber besteht, dass b unter Berücksichtigung des jeweiligen statistischen Anliegens zu bestimmen ist. Vgl. Efron, Tibshirani 1986: 72; Efron 1987: 180 f.; Efron 1990: 79. Da für den hier gegebenen Kontext bislang kein Anhaltspunkt hinsichtlich einer für b adäquaten Größenordnung verfügbar ist, verkörpert die Klärung dieser Frage ein Teilziel der vorliegenden Untersuchung.
- (5) Anhand der b verfügbaren Ausprägungen von $\hat{S}_j$ wird die Bootstrap-Verteilung $\hat{G}$ dieser Teststatistik ermittelt. Hierzu ordnet man jedem Wert $\hat{S}_j$ die Wahrscheinlichkeit b^{-1} zu. Die derart definierte Verteilung $\hat{G}$ dient als Approximation für die unbekannte Verteilung G von S .

2.2 Ein Schema zur Prüfung der Stabilität von Gütekriterien eines Kausalmodells

Um die Stabilität von Maßen, die den Fit eines Linearen Strukturgleichungsmodells widerspiegeln, mit Hilfe der Bootstrap-Methodik überprüfen zu können, ist der im vorangegangenen Abschnitt vorgestellte allgemeine Algorithmus geringfügig zu modifizieren. Aus der um Datensätze mit fehlenden Werten bereinigten Experimentalpopulation $\hat{F}$ wird durch m -maliges Ziehen mit Zurücklegen ein m Fälle umfassendes Bootstrap-Sample $\hat{X} = (\hat{X}_1, \hat{X}_2, \dots, \hat{X}_m)$ konstruiert (Schritte 1 und 2) (Das ex ante Aussondern der Datensätze mit fehlenden Werten stellt sicher, dass in Schritt 3 jede Modellschätzung auf einer Stichprobe der Größe m beruht.). Hierbei verkörpert $\hat{X}_i$ einen $(p \times 1)$ -Vektor von Realisationen der p im Modell berücksichtigten Indikatoren. Für diese Stichprobe schätzt man das anhand empirischer Befunde modifizierte Kausalmodell und berechnet jene Maße, die dessen Güte widerspiegeln (Schritt 3). Wiederholt man diesen Vorgang $(b - 1)$ -mal, ergeben sich b statistisch voneinander unabhängige Stichproben und dementsprechend für jeden Modellparameter und jede Teststatistik b Realisationen (Schritt 4) (vgl. zur Vorgehensweise Yung, Bentler 1996: 197). Ordnet man jeder Ausprägung eines interessierenden Fit-Kriteriums die Wahrscheinlichkeitsmasse b^{-1} zu, verfügt man über dessen empirische Verteilung $\hat{G}$ beziehungsweise eine Approximation der unbekannten Verteilung G (Schritt 5).

3 Problemfelder einer Verknüpfung von Bootstrap-Methode und Kovarianzstrukturanalyse

3.1 Der Bootstrap-Schätzungen innewohnende Bias

3.1.1 Die Ursache der Verzerrung

Das Konzept des Bootstrapping beruht auf der Annahme, dass das Verteilungsverhalten einer Teststatistik nicht davon abhängt, ob man eine Stichprobe aus der Experimentalpopulation $\hat{F}$ oder aus der Grundgesamtheit F zieht (Bollen, Stine 1993: 113). Für Maße, die die Güte eines Linearen Strukturgleichungsmodells widerspiegeln, gilt diese Indifferenz jedoch nicht (Diesen Sachverhalt belegen Bollen, Stine 1993: 113 ff. und 127 ff., anhand von Beispielen.). Die zu beobachtende Divergenz ist Folge der mathematischen Struktur der Gütekriterien, die zum Ausdruck bringen, in welchem Umfang die vom postulierten Modell reproduzierte Varianz-Kovarianzmatrix $\Sigma(\hat{\theta})$ von der Varianz-Kovarianzmatrix Σ abweicht, die der Schätzung zugrunde lag. Ein Blick auf das Prinzip, auf dem die Bewertung des Fit eines Kausalmodells beruht, verdeutlicht diesen Sachverhalt.

In der Regel signalisieren die Gütemaße eines Kausalmodells eine Divergenz zwischen der vom Modell reproduzierten Varianz-Kovarianzmatrix $\Sigma(\hat{\theta})$ und der aus den

erhobenen Daten $(x_1, x_2, \dots, x_n)$ errechneten Varianz-Kovarianzmatrix Σ_F . Dem Forscher obliegt die Aufgabe abzuschätzen, ob es die Ausprägungen der Fit-Maße erlauben, das Hypothesensystem als empirisch bewährt zu klassifizieren. Bei einigen Gütekriterien, wie dem Quotienten aus dem χ^2_{df} -Wert und der Anzahl an Freiheitsgraden, kann man sich hierzu eines Signifikanztests bedienen. Lässt sich die beobachtete Abweichung zwischen $\Sigma(\hat{\theta})$ und Σ_F allein mit einem zufälligen Einfluss erklären, hat sich die Nullhypothese, dass die empirisch ermittelte Varianz-Kovarianzmatrix aus einer Verteilung stammt, deren unbekannte Varianz-Kovarianzmatrix der modelltheoretisch reproduzierten gleicht, bewährt. Bei Fit-Kriterien, für die ein solcher Test nicht zur Verfügung steht, lässt sich die Fragestellung lediglich anhand mehr oder minder etablierter Konventionen beantworten.

Schätzt man ein Kausalmodell mit einer Stichprobe, die aus der Experimentalpopulation $\hat{F}$ stammt, weichen die Gegebenheiten in einem zentralen Punkt von den beschriebenen ab. Sowohl die vom Hypothesengeflecht reproduzierte Varianz-Kovarianzmatrix $\Sigma(\hat{\theta})$ als auch die für das Sample aus der Experimentalpopulation $\hat{F}$ beziehungsweise die Stichprobe aus der Grundgesamtheit F ermittelten Varianz-Kovarianzmatrizen $\Sigma_{\hat{F}}$ und Σ_F sind bekannt. Folglich handelt es sich bei der Differenz zwischen $\Sigma(\hat{\theta})$ und Σ_F um eine solche deterministischer Natur, so dass sich eine inferenzstatistische Betrachtung des Unterschieds zwischen $\Sigma(\hat{\theta})$ und Σ_F erübrigt. Häufig weichen Σ_F und $\Sigma(\hat{\theta})$ jedoch selbst dann voneinander ab, wenn sich die den Fit-Maßen zugrundeliegende Nullhypothese bewährt hat, dass $\Sigma(\hat{\theta})$ und die unbekannte Varianz-Kovarianzmatrix, die sich für die Grundgesamtheit ergibt, identisch sind. Dies führt dazu, dass der Erwartungswert eines Gütekriteriums, das für ein aus der Experimentalpopulation $\hat{F}$ gezogenes Sample berechnet wird, im Mittel stärker von der optimalen Ausprägung des Fit-Maßes entfernt liegt als jener, der sich für eine Stichprobe aus der Grundgesamtheit F ergibt.⁵ Infolgedessen weisen die für Bootstrap-Stichproben errechneten globalen Gütekriterien eine Verzerrung auf.⁶

3.1.2 Die Modifikation des Resampling-Raumes als Lösungsansatz

Die beschriebene Unzulänglichkeit wirft die Frage auf, wie ihr entgegnet werden sollte. Einschlägige Forschungsbemühungen werden derzeit von der Strategie dominiert, die beobachteten Realisationen x_i so zu modifizieren, dass die von diesen implizierte der vom Modell reproduzierten Varianz-Kovarianzmatrix entspricht. Der Resampling-Raum wird rekonfiguriert.

⁵ Vgl. Bollen, Stine 1993: 118 ff.

⁶ Vgl. Bollen, Stine 1993: 116 und 119.

Beran/Srivastava schlagen vor, die Realisationen $(x_1, x_2, \dots, x_n)$, die per Zufallsgenerator aus der Grundgesamtheit mit der unbekannten Verteilung F gezogen wurden, mit Hilfe der Transformationsvorschrift

$$R := \left\{ y_i = \Sigma(\hat{\theta})^{\frac{1}{2}} \cdot \Sigma_F^{-\frac{1}{2}} \cdot x_i, \text{ mit } i = 1, 2, \dots, n \right\},$$

mit

$\Sigma(\hat{\theta}) :=$ durch das Modell reproduzierte Varianz-Kovarianzmatrix und

$\Sigma_F :=$ symmetrische $(p \times p)$ -Matrix der Varianzen und Kovarianzen der p Indikatoren, in n Datensätze y_i zu überführen.⁷ Die auf diesem Wege ermittelte Bootstrap-Verteilung konvergiert, wie die Autoren analytisch beweisen, asymptotisch gegen die Verteilung G .

Einen ähnlichen Ansatz verfolgen Bollen/Stine, die die Realisationen x_i mit Hilfe der Formel

$$R := \left\{ y_i = \Sigma(\hat{\theta})^{\frac{1}{2}} \cdot \Sigma_F^{-\frac{1}{2}} \cdot (x_i - \bar{x}), \text{ mit } i = 1, 2, \dots, n \right\}$$

mit

$\Sigma(\hat{\theta}) :=$ durch das Modell reproduzierte Varianz-Kovarianzmatrix und

$\Sigma_F :=$ symmetrische $(p \times p)$ -Matrix der Varianzen und Kovarianzen der p Indikatoren, in n Datensätze y_i überführen und den Bootstrap-Algorithmus auf diesen Resampling-Raum anwenden.⁸ Die Autoren belegen die Gültigkeit des postulierten Ansatzes anhand von Beispielen und einer Simulation. Dass man auf dem von Bollen/Stine beschrittenen Weg zu einer angemessenen Schätzung für G gelangt, bestätigt darüber hinaus eine Simulationsstudie von Ichikawa/Konishi.⁹

3.1.3 Mit einer Transformation der Daten verbundene Probleme

Die beiden referierten Ansätze verfügen ohne Frage über methodische Eleganz. Diese darf jedoch nicht den Blick für jene Probleme verdecken, die mit einer Modifikation des Resampling-Raumes einhergehen. Erst nach einer Abwägung zwischen den sich eröffnenden Chancen und den verfahrensspezifischen Risiken lässt sich abschätzen, ob die propagierten Transformationsvorschriften als angemessen gelten können.

Als ein zentraler Vorteil der Bootstrap-Technik gilt, dass es sich grundsätzlich sowohl um einen verteilungs- als auch um einen modellfreien Ansatz handelt. Da Beran/Srivastava und Bollen/Stine bei der Transformation der x_i zu y_i auf die vom Hypothesensystem reproduzierte Varianz-Kovarianzmatrix zurückgreifen, kann die

⁷ Vgl. Beran, Srivastava 1985.

⁸ Vgl. Bollen, Stine 1993: 120 ff.

⁹ Vgl. Ichikawa, Konishi 1995.

Methodik zwar noch Verteilungs-, nicht aber Modellfreiheit beanspruchen.¹⁰ Das der nichtparametrischen Statistik innewohnende Vorteilspotential lässt sich folglich nach einer derartigen Modifikation der Daten nicht mehr beanspruchen.¹¹

Die unzureichende theoretische Verankerung einer Transformation des Resampling-Raumes verkörpert einen weiteren fundamentalen Mangel. Eine derartige Modifikation der Daten neutralisiert zwar die augenfälligen Symptome, behindert allerdings eine Auseinandersetzung mit dem grundsätzlichen Problem eher als dass sie diese fördert. Dies ist umso bedenklicher, als dem Ansatz, die x_i so in y_i zu überführen, dass sich die Varianz-Kovarianzmatrix der y_i und die vom Modell reproduzierte gleichen,¹² substantielle Unwägbarkeiten innewohnen.

Lässt sich die Diskrepanz zwischen $\Sigma(\hat{\theta})$ und Σ_F mit dem Zufall erklären, konzentriert sich eine Diskussion, ob das beschriebene Vorgehen angemessen ist, weitestgehend auf wissenschaftstheoretische Grundpositionen. In der sozialwissenschaftlichen Forschung stellt ein Modell, für das die vom Hypothesensystem erzeugte nicht statistisch signifikant von der der Schätzung zugrundeliegenden Varianz-Kovarianzmatrix abweicht, jedoch eher eine Ausnahme denn die Regel dar. Bei der Evaluation eines Kausalmodells kommt es deshalb vielfach nicht vornehmlich darauf an, dessen Fehlerfreiheit zu belegen, sondern abzuschätzen, ob die Ungenauigkeiten gravierend oder vernachlässigbar sind.¹³ Transformiert man bei einem Modell, das einen statistisch signifikanten Fehler aufweist, das Datenmaterial in der beschriebenen Art, ergibt sich für den Resampling-Raum eine Identität der für die y_i errechneten und der modelltheoretisch reproduzierten Varianz-Kovarianzmatrix, obwohl eine solche Beziehung für $\Sigma(\hat{\theta})$ und die unbekannte Matrix der Varianzen und Kovarianzen der p Indikatoren in der Grundgesamtheit nicht gilt. Tatsächlich vorhandene Unzulänglichkeiten drohen verdeckt zu werden, so dass der Bootstrap-Algorithmus zu einer zu positiven Einschätzung der Modellgüte führt.

Einer Rekonstruktion des Datenraumes in der beschriebenen Art liegt zudem die Annahme zugrunde, dass das identifizierte Beziehungsgeflecht richtig ist. Auch in einem solchen Fall kann es von Interesse sein, die Verteilung der geschätzten Parameter zu bestimmen. Hinfällig ist es hingegen zu prüfen, ob sich die von den Fit-Maßen attestierte Modellgüte als stabil erweist. Sofern eine entsprechende Transformation des Resampling-Raumes Berechtigung besitzt, beantwortet sich diese Frage von selbst, und es erübrigt sich, das Bootstrap-Verfahren einzusetzen.

¹⁰ Vgl. Yung, Bentler 1996: 216.

¹¹ Als nichtparametrische Statistik gelten im Rahmen dieser Arbeit Ansätze, die sowohl verteilungs- als auch modellfrei sind. Eine Bootstrap-Methode, die auf ein bestimmtes Hypothesensystem rekurriert, wird hingegen als semi-nichtparametrisch bezeichnet. Vgl. Yung, Bentler 1996: 197.

¹² Vgl. Bollen, Stine 1993: 120.

¹³ Vgl. Bentler, Bonett 1980: 588 ff.; Jöreskog, Sörbom 1993: 26; Homburg, Baumgartner 1995: 166.

Ob ein nichtparametrisches Bootstrapping zu einem adäquaten Ergebnis führt, d. h. die „richtige“ Antwort gibt, hängt darüber hinaus vor allem von der Frage ab, die beantwortet werden soll. Diese scheinbar triviale Feststellung lenkt den Blick auf einen Aspekt, dem nur selten Aufmerksamkeit zuteilwird. Gilt es, die unbekannte Verteilung G einer interessierenden Statistik S in der Grundgesamtheit zu schätzen, erweist es sich als problematisch, die Bootstrap-Samples aus der Stichprobe $(x_1, x_2, \dots, x_n)$ zu ziehen.¹⁴ Einen Rückschluss auf die Stabilität der Güte eines Kausalmodells erlaubt jedoch auch die Verteilung der auf der Grundlage der Experimentalpopulation $\hat{F}$ ermittelten Statistik S^0 . Jene lässt sich allerdings nicht auf Basis der y_i , sondern nur mit Hilfe der x_i näherungsweise bestimmen. Liegt der für S^0 berechnete Wert in einer angemessenen Größenordnung und erweist sich diese Klassifikation für aus $\hat{F}$ gezogene Samples als stabil, so besitzt das Beziehungsgeflecht nicht nur in Bezug auf die Experimentalpopulation $\hat{F}$, sondern im allgemeinen einen hinreichenden Fit.

Angesichts der Probleme, die mit einer Modifikation des Resampling-Raumes einhergehen, erscheint es angemessen, auf eine Transformation der Daten zu verzichten und stattdessen bei der Bewertung der zutage geförderten Befunde die Grenzen eines nichtparametrischen Bootstrapping ins Kalkül zu ziehen. So führt die dem Test innewohnende Verzerrung, deren Richtung, nicht aber genaues Ausmaß man kennt, dazu, dass die Hypothese einer Identität der modelltheoretisch reproduzierten und der für die p Indikatoren in der Grundgesamtheit geltenden Varianz-Kovarianzmatrix zu häufig verworfen wird. Dieses Defizit beunruhigt jedoch nicht grundsätzlich, da die Grenzen, ab denen der Wert eines Gütekriteriums einen hinreichenden Fit signalisiert, bei der Mehrzahl verfügbarer Maße lediglich Konventionen verkörpern und somit nicht inferenzstatistisch fundiert sind. Dies erleichtert es zugleich, die induzierte Erhöhung des Anforderungsniveaus dadurch abzufangen, dass man den vorgegebenen Mindestwert an die veränderten Rahmenbedingungen anpasst.

3.2 Die Charakterisierung der empirischen Verteilung einer Bootstrap-Statistik

Soll die Stabilität eines Kriteriums, in dem die globale Güte eines Kausalmodells zum Ausdruck kommt, überprüft werden, liegt es nahe, nicht auf der in Abschnitt 2.1 als Schritt 5 gekennzeichneten Erkenntnisstufe zu verharren. Vielmehr ist in einem weiteren Analyseschritt die Form der empirischen Verteilung $\hat{G}$ näher zu untersuchen und ein Konfidenzintervall für das betrachtete Fit-Maß zu ermitteln. Bewegen sich für ein vorgegebenes Signifikanzniveau beide Grenzen dieses Vertrauensbereiches in

¹⁴ Vgl. Bollen, Stine 1993: 118.

einer angemessenen Größenordnung, kann der hinreichende Fit des Beziehungsgeflechts als gegenüber alternativen Datensätzen robust gelten.

Zur Konstruktion eines Bootstrap-Konfidenzintervalls eignen sich mehrere Methoden.¹⁵ So lässt sich mittels eines Tests auf Schiefe und Exzess untersuchen, ob die für ein globales Gütemaß ermittelten Realisationen $\hat{S}_j$ normalverteilt sind.¹⁶ Trifft dies zu, kann man unmittelbar den sogenannten kritischen Wert bestimmen, für den die Irrtumswahrscheinlichkeit kleiner oder gleich α ist.¹⁷ Bewegt sich jener auf einem angemessenen Niveau, spricht dies für die Stabilität der Modellgüte.

Weicht die ermittelte Verteilung statistisch signifikant von einer Normalverteilung ab, gilt es, das Konfidenzintervall und den kritischen Wert mit Hilfe eines alternativen Ansatzes zu ermitteln. Von den in der wissenschaftlichen Literatur diskutierten Methoden hat sich in der Forschungspraxis neben der Percentile- vor allem die Bias-Corrected-Percentile-Methode bewährt. Diese verfügt gegenüber dem zuerst genannten Verfahren über den Vorteil, auch bei einer asymmetrischen Bootstrap-Verteilung zu einem unverzerrten Vertrauensbereich zu führen.¹⁸ Bei der Korrektur eines Bias recurriert diese Methode allerdings auf die Verteilungsfunktion der Bootstrap-Statistik $\hat{S}$ und damit indirekt auf die Struktur des postulierten Hypothesensystems, so dass es sich um kein modellfreies Vorgehen handelt. Infolgedessen erhält im Rahmen dieser Studie die Percentile-Methode den Vorzug.¹⁹

¹⁵ Vgl. Efron 1981; Efron 1982; Efron 1985; DiCiccio, Tibshirani 1987; Efron 1987; Beran 1988; Beran 1990; Franklin, Wasserman 1992. Einen Überblick vermittelt Gutsche 1995: 170 ff.

¹⁶ Vgl. zur Struktur dieses Tests Mardia 1970; Mardia, Foster 1983; Mardia 1985; D'Agostino 1986; Bollen 1989: 420 ff.; Jöreskog, Sörbom 1996: 168 f.

¹⁷ Vgl. Bortz, Schuster 2010: 102 ff.; Hedderich, Sachs 2012: 387 ff.

¹⁸ Mit dem Problem der Ermittlung eines unverzerrten Bootstrap-Vertrauensintervalls setzt sich detailliert Efron 1987 auseinander.

¹⁹ Die Entscheidung zugunsten dieses Verfahrens legt es nahe, die Verteilung des betrachteten Fit-Maßes auf Symmetrie hin zu überprüfen. Hierzu eignet sich der in PRELIS enthaltene Test auf Schiefe sowie der ebenfalls von diesem Programm erzeugte Verteilungsplott. Weist dieser keine substantielle Asymmetrie aus, erscheint es unproblematisch, den Bootstrap-Vertrauensbereich mit Hilfe der Percentile-Methode zu bestimmen. Der Verzicht, den Bias zu korrigieren, kann auch insofern als vertretbar gelten, als man hiermit vor allem das Ziel verfolgt, die Fehlerwahrscheinlichkeiten an den beiden Grenzen des Konfidenzintervalls zu vereinheitlichen. Vgl. Efron, Tibshirani 1986: 69. Dieses Vorhaben besitzt im Rahmen der hier betrachteten Fragestellung jedoch nur eine untergeordnete Bedeutung, da lediglich einer der beiden Grenzwerte für die Entscheidung, ob die Güte eines Kausalmodells als stabil gelten kann, relevant ist. Beim RMSEA sowie dem Quotienten aus dem χ^2_{df} -Wert und der Zahl der Freiheitsgrade handelt es sich dabei um die obere, beim GFI, IFI, CFI und TLI um die untere Schranke.

4 Das Erkenntnispotential eines nichtparametrischen Bootstrapping bei der Prüfung der Stabilität globaler Fit-Maße

4.1 Die der Analyse zugrundeliegende Untersuchung

4.1.1 Anliegen und Design der Studie

Um empirisch überprüfen zu können, inwieweit sich ein nichtparametrisches Bootstrapping zur Evaluation der Güte eines Kausalmodells eignet, war auf eine Studie zu rekurrieren, die sich der Kovarianzstrukturanalyse bedient. Die Wahl fiel auf eine Untersuchung, die der Identifikation von Determinanten der Entstehung eines Marktes für gebrauchte Güter am Beispiel von Fernsehgeräten aus zweiter Hand gewidmet war. In deren Rahmen wurden 1 508 private Haushalte schriftlich befragt. Diese Stichprobe verkörperte eine zufällige Auswahl aus der Grundgesamtheit jener bundesdeutschen Privathaushalte, die über einen Telefonanschluss verfügen. Von den kontaktierten Personen antworteten 379 (Brutto-Rücklaufquote = 25,1 %). Infolge partieller Antwortverweigerung mussten 26 Fragebogen ausgesondert werden, so dass die Befunde letztendlich auf 353 Datensätzen (Netto-Rücklaufquote = 23,4 %) beruhen (vgl. Ohlwein 1999: 212 ff.).

4.1.2 Ein Kausalmodell zur Erklärung der Einstellung zum Verkauf eines gebrauchten Fernsehapparates

Ein Modul des Forschungsprogramms, das der angesprochenen Untersuchung zugrunde lag, verkörperte das Bemühen, Größen zu identifizieren, die die Neigung eines Individuums beeinflussen, ein ausrangiertes Erzeugnis auf dem Sekundärmarkt zu veräußern. Hierzu wurde auf der Basis theoretischer Überlegungen ein Kausalmodell formuliert. In diesem fanden neben dem Zielkonstrukt „Die Einstellung zum Verkauf eines gebrauchten Fernsehapparates“ sechs latente Variablen Berücksichtigung. Bei diesen handelte es sich um die wahrgenommene Vorteilhaftigkeit des Verkaufens eines Fernsehapparates aus zweiter Hand, die Bindung an das ausrangierte Fernsehgerät, den Erlebniswert des Veräußerns eines Second hand-Fernsehempfängers, die Haltung der sozialen Bezugsgruppe zur entgeltlichen Weitergabe eines ausrangierten Fernsehers, die wahrgenommene Transparenz des Marktes für Fernsehapparate aus zweiter Hand sowie die antizipierte altruistische Komponente des Verkaufens eines derartigen Gerätes.

Tabelle 1: Die Güte des empirisch entwickelten Modells widerspiegelnde Fit-Maße

Gütekriterium	Wert des Gütekriteriums
RMSEA	0,032
GFI	0,971
χ^2_{df} - Wert/df	1,359
IFI	0,988
CFI	0,988
TLI	0,984

Die empirische Überprüfung des theoretisch abgeleiteten Beziehungsgeflechts offenbarte die Notwendigkeit, drei der postulierten Hypothesen mangels Signifikanz aus dem Modell zu entfernen. Diese Modifikation zog die Elimination von drei der sieben theoretischen Phänomene und eines weiteren Parameters des Strukturmodells nach sich und führte zu dem in Abbildung 1 dargestellten Hypothesensystem. Dieses zeichnete sich, wie die in Tabelle 1 wiedergegebenen Fit-Maße belegen, durch eine hohe globale Güte aus. Da darüber hinaus alle ermittelten Koeffizienten statistisch signifikant waren, ergab sich kein Bedarf nach einer weiteren Überarbeitung.

4.2 Die Prüfung der Stabilität globaler Gütekriterien mittels des Bootstrap-Ansatzes

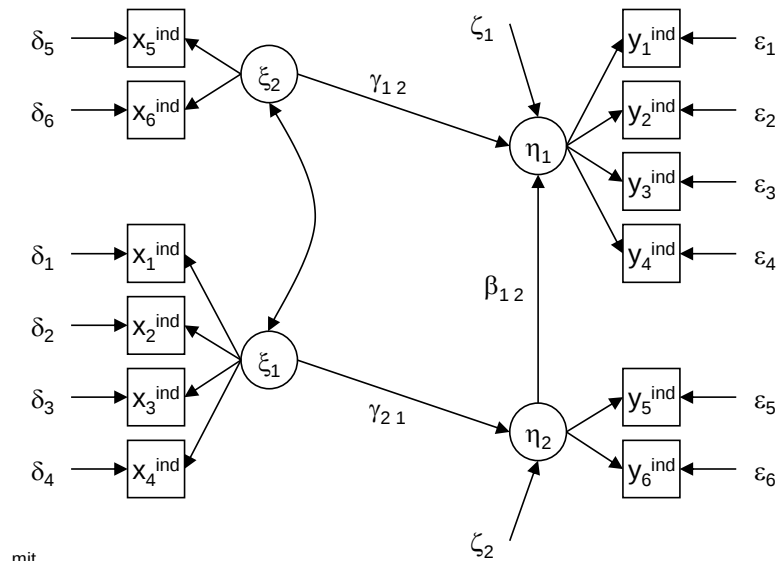
4.2.1 Das für die Analyse gewählte Vorgehen

Mit einer von empirischen Befunden geleiteten Modifikation eines Hypothesensystems geht, wie bereits problematisiert, die Gefahr einher, zu einem Modell zu gelangen, das zwar in Bezug auf die Experimentalpopulation, nicht jedoch im Hinblick auf die Grundgesamtheit über eine hohe Güte verfügt. Insofern ist mit Hilfe des Bootstrap-Ansatzes zu überprüfen, ob die für das empirisch entwickelte Beziehungsgeflecht ermittelte hohe globale Güte als stabil gelten kann.

Zu diesem Zweck wurde in Anlehnung an die in Abschnitt 2.1 skizzierte idealtypische Vorgehensweise aus den beobachteten Realisationen x_i , mit $i = 1, 2, \dots, 353$, ein Bootstrap-Sample $\hat{x}^j$ vom Umfang $m = n = 353$ gezogen. Diese Stichprobe diente in einem nächsten Schritt dazu, das postulierte Kausalmodell zu schätzen und Kriterien, die dessen globale Güte widerspiegeln, zu berechnen. Durchläuft man diesen Algorithmus b -mal, ergeben sich für jeden Parameter des Beziehungsgeflechts und jedes Globalmaß b Werte. Um einen Eindruck davon zu gewinnen, welcher Anzahl von Bootstrap-Replikationen es bedarf, um zu einer stabilen Schätzung zu gelangen, nahm b hierbei alternativ die Ausprägungen 100, 200, 300, 400, 500, 600, 800, 1 000, 1 500 und 2 000 an.

Die auf die beschriebene Art gewonnenen Werte definieren die empirische Verteilung eines Bootstrap-Gütekriteriums. Mit deren Hilfe ließen sich für das

Fit-Maß ein Konfidenzintervall und der kritische Wert berechnen. Hierzu galt es zunächst zu überprüfen, ob die empirische Verteilung die Form einer Normalverteilung besaß. Traf dies zu, konnte man mit der üblichen Formel unmittelbar ein 90 %-Vertrauensbereich und für ein Signifikanzniveau von $\alpha = 0,05$ die kritische Schwelle ermitteln, die zugleich die weiter vom Optimum des Fit-Maßes entfernt liegende Grenze des Konfidenzintervalls verkörpert. Unabhängig davon wurde mittels der Percentile-Methode ein 90 %-Vertrauensbereich und für $\alpha = 0,05$ der kritische Wert bestimmt.



mit

η_1 : Die Einstellung zum Verkauf eines gebrauchten Fernsehapparates

y_1^{ind} : Wenn ich ein noch funktionsfähiges Fernsehgerät durch ein neues ersetze, mache ich mir Gedanken, ob ich das alte verkaufen soll

y_2^{ind} : Ich finde es vernünftig, einen gebrauchten Fernsehapparat zu veräußern

y_3^{ind} : Ich kann mir vorstellen, ein gebrauchtes Fernsehgerät zu verkaufen

y_4^{ind} : Den Verkauf eines gebrauchten Fernsehapparates halte ich für sinnvoll

η_2 : Die wahrgenommene altruistische Komponente des Verkaufens eines Second hand-Fernsehgerätes

y_5^{ind} : Wenn ich ein gebrauchtes Fernsehgerät verkaufe, habe ich die Gewissheit, einer guten Sache gedient zu haben

y_6^{ind} : Ein Fernsehapparat, den ich durch einen neuen ersetze, könnte einem anderen noch gute Dienste leisten

ξ_1 : Die Haltung der sozialen Bezugsgruppe zum Verkauf eines ausrangierten Fernsehgerätes

x_1^{ind} : Wenn ich mein Fernsehgerät verkaufen würde, würde ich mich gegenüber meinen Bekannten schämen

x_2^{ind} : Der Verkauf eines gebrauchten Fernsehapparates stieße bei meinen Bekannten auf Unverständnis

x_3^{ind} : Wenn jemand ein Fernsehgerät gebraucht verkauft, erweckt dies den Eindruck, als ob er momentan knapp bei Kasse sei

x_4^{ind} : Wenn ich mein Fernsehgerät gebraucht verkaufen würde, wäre es mir wichtig, dass meine Bekannten das nicht erfahren

ξ_2 : Die wahrgenommene ökonomische Vorteilhaftigkeit des Verkaufens eines Fernsehapparates aus zweiter Hand

x_5^{ind} : Der Aufwand, der mit dem Verkauf eines gebrauchten Fernsehapparates einhergeht (Zeit, Wege etc.), ist im Verhältnis zum Erlös viel zu hoch

x_6^{ind} : Ein gebrauchtes Fernsehgerät ist so gut wie nichts mehr wert

Abbildung 1: Ein Kausalmmodell zur Erklärung der Einstellung zum Verkauf eines gebrauchten Fernsehapparates

4.2.2 Die zutage geförderten Befunde

Die im vorangegangenen Abschnitt skizzierten Analyseschritte führten zu den in Tabelle 2 zusammengestellten Ergebnissen. Anhand dieser lässt sich überprüfen, ob sich die Güte des theoretisch abgeleiteten und empirisch entwickelten Kausalmodells als stabil erweist. Darüber hinaus vermitteln die Befunde einen Eindruck von der Adäquanz eines nichtparametrischen Bootstrapping im Allgemeinen sowie der sechs ins Kalkül gezogenen Fit-Kriterien im Speziellen, und sie erlauben es, eine angemessene Größenordnung für die Anzahl von Bootstrap-Replikationen zu bestimmen.

Tabelle 2: Kenngrößen der Verteilung der betrachteten Gütemaße

Anzahl Boot-strap Replikationen (b)	Kleinsten beobachteter Wert	Mittelwert	Größter beobachteter Wert	Grenzen des 90%-Konfidenzintervalls für eine theoretische Normalverteilung		Grenzen des 90%-Konfidenzintervalls für die empirische Verteilung (Percentile-Methode)	
RMSEA							
100	0,043	0,067	0,088	0,065	0,069	0,052	0,085
200	0,042	0,067	0,094	0,066	0,068	0,052	0,082
300	0,034	0,067	0,094	0,066	0,068	0,051	0,082
400	0,034	0,066	0,094	0,065	0,067	0,051	0,082
500	0,034	0,066	0,096	0,065	0,067	0,051	0,082
600	0,034	0,066	0,096	0,065	0,067	0,051	0,082
800	0,033	0,066	0,096	0,065	0,067	0,051	0,082
1000	0,033	0,066	0,096	0,065	0,067	0,050	0,082
1500	0,033	0,066	0,101	0,065	0,067	0,051	0,081
2000	0,033	0,066	0,101	0,066	0,066	0,051	0,082
GFI							
100	0,923	0,946	0,964	0,944	0,948	0,931	0,959
200	0,920	0,946	0,965	0,945	0,947	0,933	0,959
300	0,920	0,946	0,970	0,945	0,947	0,932	0,959
400	0,920	0,947	0,970	0,946	0,948	0,933	0,959
500	0,918	0,947	0,970	-*	-*	0,933	0,960
600	0,918	0,947	0,970	-*	-*	0,932	0,960
800	0,918	0,947	0,971	-*	-*	0,933	0,960
1000	0,918	0,947	0,971	-*	-*	0,932	0,961
1500	0,906	0,947	0,971	-*	-*	0,933	0,960
2000	0,906	0,947	0,971	-*	-*	0,932	0,961

Anzahl Boot- strap Replika- tionen (b)	Kleinsten beobach- teter Wert	Mittelwert	Größter beobach- teter Wert	Grenzen des 90%- Konfidenzintervalls für eine theoretische Normalverteilung		Grenzen des 90%- Konfidenzintervalls für die empirische Verteilung (Percentile- Methode)	
χ^2_{df} -Wert/df							
100	1,647	2,598	3,727	2,512	2,683	1,932	3,573
200	1,622	2,601	4,113	2,539	2,663	1,996	3,372
300	1,399	2,608	4,113	2,556	2,659	1,905	3,367
400	1,399	2,585	4,113	2,541	2,628	1,905	3,372
500	1,399	2,571	4,277	-*	-*	1,903	3,372
600	1,377	2,579	4,277	-*	-*	1,898	3,376
800	1,377	2,558	4,277	-*	-*	1,884	3,347
1000	1,377	2,569	4,277	-*	-*	1,874	3,367
1500	1,377	2,573	4,560	-*	-*	1,881	3,347
2000	1,377	2,576	4,560	-*	-*	1,881	3,353
IFI							
100	0,914	0,951	0,981	-*	-*	0,928	0,970
200	0,907	0,951	0,983	-*	-*	0,928	0,973
300	0,907	0,951	0,989	0,949	0,953	0,928	0,973
400	0,907	0,952	0,989	0,951	0,953	0,929	0,974
500	0,907	0,952	0,989	-*	-*	0,929	0,974
600	0,907	0,952	0,989	-*	-*	0,928	0,974
800	0,907	0,952	0,989	-*	-*	0,929	0,973
1000	0,903	0,952	0,989	-*	-*	0,929	0,973
1500	0,893	0,952	0,989	-*	-*	0,928	0,973
2000	0,893	0,952	0,989	-*	-*	0,928	0,974
CFI							
100	0,913	0,950	0,980	-*	-*	0,927	0,970
200	0,906	0,951	0,982	-*	-*	0,927	0,973
300	0,906	0,951	0,989	0,949	0,953	0,927	0,973
400	0,906	0,951	0,989	0,950	0,952	0,928	0,973
500	0,906	0,952	0,989	-*	-*	0,928	0,973
600	0,906	0,951	0,989	-*	-*	0,927	0,973
800	0,906	0,952	0,989	-*	-*	0,928	0,973
1000	0,901	0,952	0,989	-*	-*	0,928	0,973
1500	0,892	0,951	0,989	-*	-*	0,927	0,973
2000	0,892	0,951	0,989	-*	-*	0,927	0,973

Ohlwein, Martin: Die Prüfung der globalen Güte eines Kausalmodells auf Stabilität mit Hilfe eines nichtparametrischen Bootstrap-Algorithmus

Anzahl Boot- strap Replika- tionen (b)	Kleinst- er beobach- teter Wert	Mittelwert	Größter beobach- teter Wert	Grenzen des 90%- Konfidenzintervalls für eine theoretische Normalverteilung		Grenzen des 90%- Konfidenzintervalls für Die empirische Verteilung (Percentile- Methode)	
TLI							
100	0,880	0,932	0,973	-*	-*	0,899	0,959
200	0,871	0,932	0,976	-*	-*	0,899	0,961
300	0,871	0,932	0,984	0,930	0,934	0,900	0,963
400	0,871	0,933	0,984	0,931	0,935	0,900	0,964
500	0,871	0,933	0,984	-*	-*	0,901	0,963
600	0,871	0,933	0,984	-*	-*	0,900	0,964
800	0,871	0,934	0,984	-*	-*	0,901	0,963
1000	0,864	0,933	0,984	-*	-*	0,900	0,963
1500	0,852	0,933	0,984	-*	-*	0,900	0,962
2000	0,852	0,933	0,984	-*	-*	0,900	0,963

–* Wert nicht ermittelbar, da die empirische Verteilung nicht die Form einer Normalverteilung besaß

Vergleicht man die in Tabelle 2 wiedergegebenen mit den Werten, die die Gütekriterien für die Experimentalpopulation annahmen, fällt ins Auge, dass keiner der betrachteten Fit-Maße im Mittel das sich in Tabelle 1 widerspiegelnde hohe Niveau erreichte. Bei zwei der sechs Kriterien (RMSEA, χ^2_{df} -Wert/df) verfehlte selbst die beste der 2 000 Ausprägungen der Bootstrap-Statistik den Vergleichswert, bei weiteren zwei (GFI, TLI) entsprachen sich beide Realisationen. Dieser Befund belegt empirisch, dass ein nichtparametrisches Bootstrapping tendenziell ein zu pessimistisches Bild von der Güte eines Kausalmodells vermittelt. Die Beobachtung, dass der diesem Ansatz innewohnende Unschärfenbereich zu einer Erhöhung des Anforderungsniveaus führt, stützt ex post einen der zentralen Gedanken, mit denen der Einsatz eines sowohl verteilungs- als auch modellfreien Bootstrap-Verfahrens rechtfertigt wurde.

Das Ausmaß, in dem die für die Experimentalpopulation beobachtete Realisation und der sich für die Bootstrap-Stichproben ergebende Mittelwert differieren, hängt dabei vom Gütekriterium ab. Für den RMSEA sowie den Quotienten aus dem χ^2_{df} -Wert und der Zahl der Freiheitsgrade, für die beide eine Ausprägung von 0 einen bestmöglichen Fit signalisiert, ergibt sich jeweils ungefähr eine Verdoppelung. Eine differenziertere Einschätzung erfordern GFI, IFI, CFI und TLI, die, mit Ausnahme des IFI, Ausprägungen zwischen 0 und 1 annehmen können, wobei ein höherer Wert einen besseren Fit signalisiert.²⁰ Die geringste Abweichung ist beim GFI zu

²⁰ Zwar spiegelt auch beim IFI der Wert 0 eine maximale Diskrepanz zwischen der empirisch ermittelten und der modelltheoretisch reproduzierten Varianz-Kovarianzmatrix wider, dieses Fit-Maß kann jedoch auch Werte größer als 1 annehmen.

beobachten, einen etwas größeren Abschlag haben IFI und CFI, die sich sehr ähnlich verhalten, hinzunehmen. Mit Abstand am sensitivsten reagiert der TLI.

Vergleicht man die Mittelwerte, die sich für die einzelnen Globalmaße bei unterschiedlichen b ergaben, fällt auf, dass sie sich bei allen sechs Fit-Kriterien gegenüber einer Variation der Anzahl von Bootstrap-Replikationen robust verhalten. Die maximale Differenz beträgt bei CFI und TLI 0,002, bei RMSEA, GFI und IFI gar nur 0,001. Eine Abweichung in dieser Größenordnung bedarf, wie auch der beim Quotienten aus dem χ^2_{df} -Wert und der Zahl der Freiheitsgrade beobachtete Unterschied, keiner Aufmerksamkeit. Mit der Wahl eines großen b geht folglich kein nennenswerter Erkenntnisgewinn einher.

Eine Variation der Anzahl von Bootstrap-Replikationen wirkt sich darüber hinaus nur geringfügig auf die Grenzen der beiden ermittelten 90 %-Vertrauensbereiche aus. Bei den für eine theoretische Normalverteilung berechneten Werten wird der Effekt zudem von zwei typischen Phänomenen überlagert. Zum einen verringert sich, bedingt durch eine nur marginal schwankende Standardabweichung, mit steigender Anzahl von Bootstrap-Replikationen die Breite des Konfidenzintervalls, das insgesamt als vergleichsweise eng gelten kann. Zum anderen folgt für $b > 500$ mit Ausnahme des RMSEA keiner der betrachteten Gütekriterien einer Normalverteilung, während dies für $b < 500$ zumindest bei einigen der Fit-Maße der Fall ist. Dies beruht auf der mit zunehmendem b steigenden Stärke des herangezogenen Tests, so dass selbst geringfügige Abweichungen von der Normalverteilung nicht mehr mit einem zufälligen Einfluss erklärt werden können.

Stets bestimmen lässt sich ein Vertrauensbereich mittels der Percentile-Methode. Jener fällt, wie ein Blick auf Tabelle 2 bestätigt, für jedes Gütekriterium beträchtlich breiter als das Konfidenzintervall aus, das auf der Basis einer theoretischen Normalverteilung ermittelt wurde. Die Spannbreite des mit der Percentile-Methode berechneten Vertrauensbereiches erweist sich hierbei für 300 oder mehr Bootstrap-Replikationen als weitestgehend konstant.

Für $\alpha = 0,05$ verkörpert die untere (GFI, IFI, CFI und TLI) beziehungsweise obere (RMSEA sowie Quotient aus dem χ^2_{df} -Wert und der Zahl der Freiheitsgrade) Grenze des 90 %-Konfidenzintervalls zugleich den kritischen Wert. Berücksichtigt man die bereits mehrfach problematisierte Verzerrung, die den Bootstrap-Schätzungen innewohnt, erscheinen die ermittelten Ausprägungen als grundsätzlich unbedenklich. Für den GFI, den IFI und den CFI liegen die Schwellenwerte ohne Ausnahme in einer akzeptablen Größenordnung.²¹ Beim TLI schwanken die mittels der Percentile-

²¹ Bei diesen Gütekriterien weist eine Ausprägung von größer oder gleich 0,90 auf einen guten Fit hin. Vgl. Jacobs 1992: 199 ff.; Homburg, Baumgartner 1995: 167 ff.; Hoyle, Panter 1995: 164; Korte 1995: 179.

Methode berechneten Werte um die als Hürde geltende Größe von 0,90,²² unterschreiten diese jedoch in der Mehrzahl der Fälle nicht. Die für den Quotienten aus dem χ^2_{df} -Wert und der Zahl der Freiheitsgrade auf der Basis einer Normalverteilung berechneten Grenzen überzeugen uneingeschränkt, und auch die mittels der Percentile-Methode bestimmten befriedigen, wenn auch nur knapp.²³ Beim RMSEA erachtet man einer Konvention zufolge Werte bis 0,05 als Hinweis für einen guten, Ausprägungen bis 0,08 als Zeichen für einen akzeptablen Fit.²⁴ Lediglich die mittels der Percentile-Methode berechneten Schranken liegen, lässt man jene für $b = 100$ außer Acht, geringfügig über der zuletzt genannten Hürde. Dies erscheint allerdings im Hinblick auf die den Werten innewohnende Verzerrung als duldbar. Somit kann die für die Experimentalpopulation beobachtete hinreichende Güte des Beziehungsgeflechts als gegenüber variierenden Datensätzen stabil gelten.

4.2.3 Konsequenzen der empirischen Befunde

Den Befunden dieser auf der explorativen Ebene angesiedelten Studie lassen sich in mehrfacher Hinsicht Hinweise für die Evaluation der Güte eines Kausalmodells mittels eines nichtparametrischen Bootstrap-Algorithmus entnehmen. Betrachtet man den Grad der Verzerrung, den die einzelnen Gütekriterien aufweisen, so ist zu konstatieren, dass jene beim GFI in relativ geringem, beim IFI und CFI in moderatem und beim TLI in vergleichsweise hohem Maße zu Tage tritt. Einem direkten Vergleich mit diesen vier Größen entziehen sich, bedingt durch ihre Struktur, der RMSEA sowie der Quotient aus dem χ^2_{df} -Wert und der Zahl der Freiheitsgrade. Von diesen beiden Fit-Maßen scheint das zuerst genannte etwas sensibler auf den durch den Prüfalgorithmus induzierten Bias zu reagieren. Die angesprochenen Relationen sollten allerdings nicht als Selektionskriterium für mehr oder minder geeignet erscheinende Gütemaße dienen, sondern die Interpretation von Bootstrap-Ergebnissen erleichtern.

Eine weitere Fragestellung, die es mit der Studie zu beantworten galt, war jene nach einer angemessenen Anzahl von Bootstrap-Replikationen. Da die geschätzten Werte bereits für Stichproben mit einem Umfang von kleiner 1 000 stabil sind, erscheint eine größere Zahl von Wiederholungen entbehrlich. Die Befunde zu den Grenzen der mit der Percentile-Methode berechneten Konfidenzintervalle legen es jedoch nahe, für b einen Wert von zumindest 300 vorzusehen. Forschungsökonomische Restriktionen sind darüber hinaus von lediglich untergeordneter Bedeutung, so dass für die Zahl der Bootstrap-Replikationen ein Wert zwischen 300 und 800 gewählt werden sollte. Beabsichtigt man, einen Vertrauensbereich unter Bezug auf eine

²² Vgl. Homburg, Baumgartner 1995: 168.

²³ Beim Quotienten aus dem χ^2_{df} -Wert und der Zahl der Freiheitsgrade gilt ein Wert von kleiner oder gleich 3 als Zeichen für einen guten Fit, in der Forschungspraxis werden aber auch Ausprägungen von bis zu 10 akzeptiert. Vgl. hierzu Fritz 1995: 125; Homburg 2000: 93 ff.

²⁴ Vgl. Browne, Cudeck 1993: 144; Albers, Hildebrandt 2006: 22.

theoretische Normalverteilung zu ermitteln, kann das Ziehen von 500 Bootstrap-Samples als ein guter Kompromiss gelten.

Eher um komplementäre denn um konfliktäre Methoden handelt es sich bei den beiden zur Bestimmung eines Konfidenzintervalls herangezogenen Techniken. Vor dem Hintergrund der Befunde erscheinen grundsätzlich beide Ansätze als geeignet. Welchem Verfahren in stärkerem Maße Aufmerksamkeit gebührt, ist somit vor dem Hintergrund des jeweiligen Forschungsanliegens zu entscheiden. Hierbei gilt es zu beachten, dass der sich auf eine theoretische Normalverteilung beziehende Ansatz nur dann zur Verfügung steht, wenn die Verteilung der begutachteten Statistik nicht signifikant von dieser abweicht. Um Vergleiche zwischen Studien ziehen zu können, liegt es deshalb nahe, stets auch das 5. und das 95. Perzentil der empirischen Verteilung zu bestimmen.

5 Ansatzpunkte für weiterführende Forschungsarbeiten

Die vorgestellte Untersuchung ist Beleg dafür, dass ein nichtparametrischer Bootstrap-Algorithmus die Evaluation der Güte eines Kausalmodells unterstützen kann. Aufgabe von Replikationsstudien wird es sein, diese Einschätzung kritisch zu überprüfen. Die grundsätzliche Eignung des betrachteten Ansatzes lenkt darüber hinaus den Blick auf nach wie vor ungeklärte Fragestellungen, die weiterführenden Forschungsarbeiten als Anknüpfungspunkte dienen können. So liegt es nahe, den Kreis begutachteter Gütekriterien sukzessive zu erweitern. Des Weiteren sollte die vorgestellte Technik auch bei solchen Kausalmodellen angewendet werden, die ein beträchtlich höheres Maß an Komplexität aufweisen.

Ein besonders fruchtbares Forschungsfeld eröffnet sich vor dem Hintergrund der theoretisch diskutierten und empirisch manifestierten Verzerrung der Bootstrap-Schätzung. Es bleibt zu vermuten, dass sich in jener zwei Sachverhalte widerspiegeln. Zum einen handelt es sich hierbei um die Güte, die das Modell im Hinblick auf die Experimentalpopulation aufweist, zum anderen um das Ausmaß, indem das theoretisch abgeleitete Beziehungsgeflecht im Rahmen seiner von empirischen Befunden geleiteten Modifikation an das Experimentalsample angepasst wurde. Ließen sich beide Effekte trennen, wäre dies ein wesentlicher Schritt in Richtung eines besseren Verständnisses des komplexen Phänomens Güte eines Kausalmodells.

Literaturverzeichnis

- ALBERS, S./HILDEBRANDT, L. (2006): *Methodische Probleme bei der Erfolgsfaktorenforschung*. In: zfbf 58 (2006), Februar, S. 2–33
- ALBERT, H. (1980): *Traktat über kritische Vernunft*. 4., verb. Aufl. Tübingen : Mohr (Die Einheit der Gesellschaftswissenschaften Bd. 9)
- ANDERSON, P. F. (1983): *Marketing, scientific progress, and scientific method*. In: Journal of Marketing 47 (1983), Nr. 4, S. 18–31
- ARNDT, J. (1985): *On making marketing science more scientific*. In: Journal of Marketing 49 (1985), Nr. 3, S. 11–23
- BALDERJAHN, I. (1986): *Das umweltbewusste Konsumentenverhalten*. Berlin : Duncker & Humblot (Betriebswirtschaftliche Schriften Heft 123)
- BENTLER, P. M./BONETT, D. G. (1980): *Significance tests and goodness of fit in the analysis of covariance structures*. In: Psychological Bulletin 88 (1980), Nr. 3, S. 588–606
- BERAN, R. (1988): *Prepivoting test statistics*. In: Journal of the American Statistical Association 83 (1988), Nr. 403, S. 687–697
- BERAN, R. (1990): *Refining bootstrap simultaneous confidence sets*. In: Journal of the American Statistical Association 85 (1990), Nr. 410, S. 417–426
- BERAN, R./SRIVASTAVA, M. S. (1985): *Bootstrap tests and confidence regions for functions of a covariance matrix*. In: The Annals of Statistics 13 (1985), Nr. 1, S. 95–115
- BEREKOVEN, L./ECKERT, W./ELLENRIEDER, P. (2009): *Marktforschung*. 12., überarb. und erw. Aufl. Wiesbaden : Gabler (Lehrbuch)
- BICKEL, P. J./FREEDMAN, D. A. (1981): *Some asymptotic theory for the bootstrap*. In: The Annals of Statistics 9 (1981), Nr. 6, S. 1196–1217
- BICKEL, P. J./GÖTZE, F./ZWET, W. R. VAN (1994): *Resampling fewer than n observations*. Bielefeld : Sonderforschungsbereich 343 (Arbeitspapier der Universität Bielefeld 94-084)

- BOLLEN, K. A./STINE, R. A. (1993): *Bootstrapping goodness-of-fit measures in structural equation models*. In: Bollen; Long (Hrsg.): Testing structural equation models. Newbury Park, Calif. [u. a.] : Sage Publ., 1993 (Sage focus editions, 154), S. 111–135
- BORTZ, J./SCHUSTER, C. (2010): *Statistik für Human- und Sozialwissenschaftler*. 7., vollst. überarb. u. erw. Aufl. Berlin : Springer (Springer-Lehrbuch)
- BOYD, R. N. (1984): *The current status of scientific realism*. In: Leplin (Hrsg.): Scientific realism. Berkeley : University of California Press, 1984 (Campus, 323), S. 41–82
- BROWNE, M. W./CUDECK, R. (1993): *Alternative ways of assessing model fit*. In: Bollen; Long (Hrsg.): Testing structural equation models. Newbury Park, Calif. [u. a.] : Sage Publ., 1993 (Sage focus editions, 154), S. 136–162
- CAUSEY, R. L. (1979): *Theory and observation*. In: Asquith; Kyburg (Hrsg.): Current research in philosophy of science. East Lansing - Mich : Philosophy of science Association, 1979, S. 187–206
- D'AGOSTINO, R. B. (1986): *Tests for the normal distribution*. In: D'Agostino; Stephens (Hrsg.): Goodness-of-fit techniques. New York : M. Dekker, 1986 (Statistics, textbooks and monographs, vol. 68), S. 367–419
- DESHPANDE, R. (1983): *"Paradigms Lost"*. In: Journal of Marketing 47 (1983), Nr. 4, S. 101–110
- DICICCIO, T./TIBSHIRANI, R. J. (1987): *Bootstrap confidence intervals and bootstrap approximations*. In: Journal of the American Statistical Association 82 (1987), Nr. 397, S. 163–170
- DIEKMANN, A. (2011): *Are most published research findings false?* In: Jahrbücher für Nationalökonomie und Statistik 231 (2011), 5/6, S. 628–635
- EFRON, B. (1981): *Nonparametric standard errors and confidence intervals*. In: The Canadian Journal of Statistics 9 (1981), Nr. 2, S. 139–172
- EFRON, B. (1982): *The Jackknife, the bootstrap and other resampling plans*. Philadelphia - Pa : Society for industrial and applied mathematics (CBMS-NSF regional conference series in applied mathematics 38)
- Ohlwein, Martin: Die Prüfung der globalen Güte eines Kausalmodells auf Stabilität mit Hilfe eines nichtparametrischen Bootstrap-Algorithmus

- EFRON, B. (1985): *Bootstrap confidence intervals for a class of parametric problems*. In: Biometrika 72 (1985), Nr. 1, S. 45–58
- EFRON, B. (1987): *Better bootstrap confidence intervals*. In: Journal of the American Statistical Association 82 (1987), Nr. 397, S. 171–185
- EFRON, B. (1990): *More efficient bootstrap computations*. In: Journal of the American Statistical Association 85 (1990), Nr. 409, S. 79–89
- EFRON, B./GONG, G. (1983): *A leisure look at the bootstrap, the jackknife, and cross-validation*. In: The American Statistician 37 (1983), Nr. 1, S. 36–48
- EFRON, B./TIBSHIRANI, R. J. (1986): *Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy*. In: Statistical Science 1 (1986), Nr. 1, S. 54–75
- FRANKLIN, L./WASSERMAN, G. (1992): *Bootstrap lower confidence limits for capability indices*. In: Journal of Quality Technology 24 (1992), Nr. 4, S. 196–210
- FRITZ, W. (1995): *Marketing-Management und Unternehmenserfolg*. 2., überarb. und erg. Aufl. Stuttgart : Schäffer-Poeschel (Betriebswirtschaftliche Abhandlungen n. F., Bd. 90)
- GUTSCHE, J. (1994): *Bootstrapping*. Mannheim : Inst. für Marketing (Arbeitspapier / Institut für Marketing, Universität Mannheim Nr. 103)
- GUTSCHE, J. (1995): *Produktpräferenzanalyse*. Berlin : Duncker & Humblot (Schriften zum Marketing 40)
- HARRÉ, R. (1986): *Varieties of realism*. Oxford : Basil Blackwell
- HEDDERICH, J./SACHS, L. (2012): *Angewandte Statistik*. 14., überarb. und erg. Aufl. Berlin [u. a.] : Springer
- HOMBURG, C. (2000): *Kundennähe von Industriegüterunternehmen*. 3. Aufl. Wiesbaden : Gabler (Neue betriebswirtschaftliche Forschung 147)
- HOMBURG, C./BAUMGARTNER, H. (1995): *Beurteilung von Kausalmodellen*. In: Marketing ZFP 17 (1995), Nr. 3, S. 162–176
- HOYLE, R. H./PANTER, A. T. (1995): *Writing about structural equation models*. In: Hoyle (Hrsg.): *Structural equation modeling*. Thousand Oaks, Calif. [u. a.] : Sage Publications, 1995, S. 158–176

- HUNT, S. D. (1984): *Should marketing adopt relativism?* In: Anderson; Ryan (Hrsg.): Scientific method in marketing. Chicago, IL : American Marketing Association, 1984 (Proceedings series / American Marketing Association), S. 30–34
- HUNT, S. D. (1991): *Modern marketing theory*. Cincinnati, OH : South-Western Pub. Co.
- ICHIKAWA, M./KONISHI, S. (1995): *Application of the bootstrap methods in factor analysis*. In: Psychometrika 60 (1995), Nr. 1, S. 77–93
- JACOBS, S. (1992): *Strategische Erfolgsfaktoren der Diversifikation*. Wiesbaden : Gabler (Neue betriebswirtschaftliche Forschung 88)
- JÖRESKOG, K. G./SÖRBOM, D. (1993): *LISREL 8: structural equation modeling with the SIMPLIS command language*. Chicago, Ill., [u. a.] : Scientific Software International [u. a.]
- JÖRESKOG, K. G./SÖRBOM, D. (1996): *PRELIS 2 user's reference guide*. 3rd ed. Chicago, IL : Scientific Software International
- KERR, N. L. (1998): *HARKing: hypothesizing after the results are known*. In: Personality and social psychology review 2 (1998), Nr. 3, S. 196–217
- KORTE, C. (1995): *Customer satisfaction measurement*. New York [u. a.] : P. Lang (Schriften zu Distribution und Handel Bd. 15)
- KUBICEK, H. (1975): *Empirische Organisationsforschung*. Stuttgart : Poeschel (Sammlung Poeschel 78)
- LEPLIN, J. (1986): *Methodological realism and scientific rationality*. In: Philosophy of Science 53 (1986), Nr. 1, S. 31–51
- LEVIN, M. E. (1984): *What kind of explanations is truth?* In: Leplin (Hrsg.): Scientific realism. Berkeley : University of California Press, 1984 (Campus, 323), S. 124–139
- MACKINNON, E. (1979): *Scientific Realism*. In: Philosophy of Science 46 (1979), Nr. 4, S. 501–532
- MARDIA, K. V. (1970): *Measures of multivariate skewness and kurtosis with applications*. In: Biometrika 57 (1970), Nr. 3, S. 519–530
- Ohlwein, Martin: Die Prüfung der globalen Güte eines Kausalmodells auf Stabilität mit Hilfe eines nichtparametrischen Bootstrap-Algorithmus

- MARDIA, K. V. (1985): *Mardia's test of multinormality*, Bd. 5. In: Kotz; Johnson (Hrsg.): Encyclopedia of statistical sciences. New York : Wiley, 1985 (Encyclopedia of statistical sciences / Editors in chief: Samuel Kotz, 5), S. 217–221
- MARDIA, K. V./FOSTER, K. (1983): *Omnibus tests of multinormality based on skewness and kurtosis*. In: Communication in Statistic - Theory and Methods 12 (1983), Nr. 2, S. 207–221
- MARTIN, A. (1989): *Die empirische Forschung in der Betriebswirtschaftslehre*. Stuttgart : Poeschel (Betriebswirtschaftliche Abhandlungen. Neue Folge Band 73)
- MCMULLIN, E. (1984): *A case for scientific realism*. In: Leplin (Hrsg.): Scientific realism. Berkeley : University of California Press, 1984 (Campus, 323), S. 8–40
- MEYER, W. (1979): *Die Methodologie des Kritischen Rationalismus*. In: Raffée; Abel (Hrsg.): Wissenschaftstheoretische Grundfragen der Wirtschaftswissenschaften. München : Vahlen, 1979 (WiSt-Taschenbücher), S. 28–43
- MÜLLER, S. (1991): *Die Psyche des Managers als Determinante des Exporterfolges*. Stuttgart : M & P
- OHLWEIN, M. (1999): *Märkte für gebrauchte Güter*. Wiesbaden : DUV
- PÄHLER, K. (1986): *Qualitätsmerkmale wissenschaftlicher Theorien*. Tübingen : Mohr (Die Einheit der Gesellschaftswissenschaften Bd. 43)
- POPPER, K. R. (1963): *Conjectures and refutations*. London [u. a.] : Routledge and Kegan Paul
- POPPER, K. R. (1984): *Logik der Forschung*. 8., weiter verb. und verm. Aufl. Tübingen : J.C.B. Mohr (Paul Siebeck) (Die Einheit der Gesellschaftswissenschaften Bd. 4)
- POPPER, K. R. (1991): *Auf der Suche nach einer besseren Welt*. 6. Aufl. München : Piper
- RAFFÉE, H. (1993): *Gegenstand, Methoden und Konzepte der Betriebswirtschaftslehre*, Bd. 1. In: Baetge; Bitz (Hrsg.): Vahlens Kompendium der Betriebswirtschaftslehre. 3., überarb. und erw. Aufl. München : Vahlen, 1993, S. 1–46
- RAFFÉE, H./ABEL, B. (1979): *Aufgaben und aktuelle Tendenzen der Wissenschaftstheorie in den Wirtschaftswissenschaften*. In: Raffée; Abel (Hrsg.): Wissenschaftstheoretische Grundfragen der Wirtschaftswissenschaften. München : Vahlen, 1979 (WiSt-Taschenbücher), S. 1–10

- SELLARS, W. (1963): *Science, perception and reality*. London : Routledge and Kegan Paul
- SPINNER, H. F. (1982): *Ist der Kritische Realismus am Ende?* Weinheim [u. a.] : Beltz
- TANAKA, J. S. (1987): *How big is big enough?* In: Child Development 58 (1987), Nr. 1, S. 134–146
- TEMME, D./HILDEBRANDT, L. (2009): *Gruppenvergleiche bei hypothetischen Konstrukten*. In: zfbf 61 (2009), Nr. 2, S. 138–185
- WITTE, E. (1977): *Lehrgeld für empirische Forschung*. In: Köhler (Hrsg.): Empirische und handlungstheoretische Forschungskonzeption in der Betriebswirtschaftslehre. Stuttgart : C. E. Poeschel, 1977, S. 269–281
- WITTE, E./GRÜN, O./BRONNER, R. (1975): *Pluralismus in der betriebswirtschaftlichen Forschung*. In: zfbf 27 (1975), Nr. 12, S. 796–800
- YUNG, Y.-F./BENTLER, P. M. (1996): *Bootstrapping techniques in analysis of mean and covariance structures*. In: Marcoulides; Schumacker (Hrsg.): Advanced structural equation modeling. Mahwah, NJ : L. Erlbaum Associates, 1996, S. 195–226

Der Autor

Prof. Dr. Martin Ohlwein studierte Betriebswirtschaftslehre mit den Schwerpunkten Marketing, Internationales Management und Statistik an der Universität Mannheim. Nach seinem Diplomabschluss arbeitete er an dieser Hochschule als wissenschaftlicher Mitarbeiter am Lehrstuhl für Allgemeine Betriebswirtschaftslehre und Marketing I und promovierte an der dortigen Fakultät für Betriebswirtschaftslehre. Seit 1999 war er als Unternehmensberater tätig, u. a. im Frankfurter Büro von The Boston Consulting Group. In dieser Zeit entwickelte er für Klienten aus verschiedenen Branchen u. a. internationale Produktbereichs- und Markteintrittsstrategien. Für eine Karriere im Marketing entschied er sich, da er einen direkten Beitrag leisten wollte, Unternehmen marktorientiert zu steuern und auf diesem Weg neue Erfolgspotenziale zu erschließen sowie einen Wettbewerbsvorteil zu erzielen. Es folgten seit 2005 Positionen als u. a. für die Konzernmarktforschung verantwortlicher Vice President im Brand Management der Hilti AG (Fürstentum Liechtenstein) sowie als Mitglied der Geschäftsführung von Hilti Emirates LLC (Vereinigte Arabische Emirate), in der er für Marketing und die Vertriebskanäle verantwortlich zeichnete. Seit 2011 lehrt Ohlwein Marketing, Marktforschung, Strategisches Management und Businessplanung am Frankfurter Campus der International School of Management. Seit seinem Ausscheiden aus der Industrie arbeitet er zudem als selbstständiger Berater und Coach und ist Mitglied verschiedener Beiräte diverser Organisationen.



International School of Management

Die International School of Management (ISM) – eine staatlich anerkannte, private Hochschule – bildet seit 1990 in Dortmund, Frankfurt/Main, München, Hamburg und Köln Nachwuchsführungskräfte für die internationale Wirtschaft aus. Das Studienprogramm umfasst sieben Bachelor-Studiengänge (International Management, Tourism & Event Management, Communications & Marketing, Psychology & Management, Finance & Management, Global Brand & Fashion Management, Online-Marketing & E-Commerce), sechs konsekutive Master-Studiengänge (International Management, Strategic Marketing Management, Finance, International Logistics & Supply Chain Management, Psychology & Management, Luxury, Fashion & Sales Management) und vier berufsbegleitende Programme (B.A. Business Administration, M.A. Management, MBA General Management, MBA Energy Management). Alle Studiengänge der ISM zeichnen sich durch ihre Internationalität und Praxisorientierung aus. Diese Erfolgsfaktoren garantiert die ISM durch enge Kooperationen mit Unternehmen, Projekte in Kleingruppen sowie integrierte Auslandssemester und -module

an weltweit über 170 Partneruniversitäten. Die Qualität der Ausbildung bestätigen Studierende und Ehemalige ebenso wie Personaler in zahlreichen Hochschulrankings. Die ISM belegt dort seit Jahren konstant vorderste Plätze.

Mit dem ISM Working Paper wird die bisherige ISM-Reihe „Discussion Papers“ ausgeweitet, das heißt, es werden Ergebnisse von Arbeiten präsentiert, wie z. B. Thesen, Ergebnisse aus Workshops oder aus eigenen Forschungsarbeiten. Ähnlich wie beim ISM Research Journal, das ebenfalls zu den neuen ISM Publikationsreihen gehört, werden die Beiträge im ISM Working Paper einem fachlichen Bewertungsverfahren (Peer Review) unterzogen.

In der Reihe „Working Paper“ bisher erschienen:

- No. 1** Brock, Stephan; Antretter, Torben: Kapitalkostenermittlung als Grauzone wertorientierter Unternehmensführung, 2014
- No. 2** Ohlwein, Martin: Die Prüfung der globalen Güte eines Kausalmodells auf Stabilität mit Hilfe eines nichtparametrischen Bootstrap-Algorithmus, 2015

Diese Titel können u. a. über www.mv-buchhandel.de bestellt werden.