Sozialwissenschaftliche Forschungsmethoden 3

Jost Reinecke

Wachstumsmodelle

2. Auflage

Nomos

Sozialwissenschaftliche Forschungsmethoden

Band 3

Jost Reinecke

Wachstumsmodelle

2., durchgesehene und erweiterte Auflage

Nomos

Onlineversion
Nomos eLibrary

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.d-nb.de abrufbar.

ISBN 978-3-98542-101-5 (Print)

ISBN 978-3-95710-502-8 (ePDF)

2., durchgesehene und erweiterte Auflage 2024

Vorwort zur ersten Auflage

Die statistische Modellierung von Entwicklungsprozessen hat sich in den letzten Jahren nicht zuletzt durch die Zunahme von Studien, die Längsschnittdaten produzieren, deutlich verstärkt. Innerhalb des Strukturgleichungsansatzes haben sich sowohl einfache als auch komplexere Wachstumsmodelle in vielen sozialwissenschaftlichen Bereichen durchgesetzt. Dieses Lehrbuch soll Einsteigern mit Kenntnissen in der multivariaten Statistik und Datenanalyse helfen, sich diesen Modellierungsmöglichkeiten zu nähern. In vielen einführenden Lehrbüchern zu Strukturgleichungsmodellen werden Wachstumsmodelle zwar auch beispielhaft erörtert, eine Systematisierung der vielfältigen Modellierungsmöglichkeiten fehlt für den deutschsprachigen Raum aber bisher.

Durch die Längsschnittstudie „Kriminalität in der modernen Stadt“, die ich seit dem Jahre 2002 mit dem Kollegen Klaus Boers von der Universtät Münster leite, steht ein Fundus an Datenmaterial zur Verfügung, mit dem ich systematisch die Modellierung von Entwicklungsprozessen nicht nur formal, sondern auch anwendungsorientiert verdeutlichen kann. Für die Aufbereitung dieses Datenmaterials standen mir Andreas Daniel, Daniela Pollich, Andreas Pöge und Jochen Wittenberg hilfreich zur Seite. Ergänzend konnte ich Daten aus dem Längsschnittprojekt „Gruppenbezogene Menschenfeindlichkeit“ des Instituts für Konflikt- und Gewaltforschung (IKG) der Universität Bielefeld nutzen, wofür ich dem Direktor, Wilhelm Heitmeyer, zu Dank verpflichtet bin.

Unterstützung bei Fragen zu Ausfallprozessen und den damit verbundenen fehlenden Werten erhielt ich von Kristian Kleinke, der mir auch bei der Durchsicht einzelner Kapitel tatkräfig zur Seite stand. Bei der Erstellung der Abbildungen, der Systematisierung der Literatur, der technischen Unsetzung des Autorenregister und Sachindex und bei der abschließenden Formatierung der einzelnen Textteile stand mir Jakob Guzy hilfreich zur Seite. Bei ihm möchte ich mich ausdrücklich für seine sorgfältige Arbeit bedanken.

Bielefeld im März 2012 Jost Reinecke

Vorwort zur zweiten Auflage

Für die zweite Auflage ist das Lehrbuch neu strukturiert und um aktuelle Entwicklungen zur Modellierung von Wachstumsmodellen erweitert worden. Insbesondere sind hier Modelle unter Berücksichtigung der Bayesianischen Statistik und zeitkontinuierliche Ansätze hervorzuheben. Wie auch in der ersten Auflage praktiziert, stehen die an empirischen Beispielen vorgenommenen Erörterungen im Vordergrund. Bei den Bayesianischen Modellen bin ich von Jasper Bendler und bei den zeitkontinuierlichen Modellen von Manuel Voelkle beraten und unterstützt worden. Hierfür möchte ich mich ausdrücklich bedanken.

Bielefeld im Dezember 2023 Jost Reinecke

Inhaltsverzeichnis

Variablen- und Matrizenbezeichnungen

Buchstabe	Name	Allgemeine Bedeutung (Strukturgleichungsmodelle)
ξ	Ksi	unabhängige (exogene) latente Variable
η	Eta	abhängige (endogene) latente Variable
β	beta	Regressionskoeffizient ($\eta \to \eta$)
B		Matrix der Regressionskoeffizienten β
γ	gamma	Regressionskoeffizient ($\xi \to \eta$)
Γ	Gamma	Matrix der Regressionskoeffizienten γ
ϕ	phi	Varianz, Kovarianz der latenten Variablen ξ
Φ	Phi	Kovarianzmatrix der latenten Variablen ξ
ζ	Zeta	Residuum der latenten Variablen η
ψ	psi	Varianz, (Residual)kovarianz der latenten Variablen η
Ψ	Psi	Kovarianzmatrix der Residuen ζ
λ_x	lambda	Faktorenladung der manifesten Variablen x
λ_y	lambda	Faktorenladung der manifesten Variablen y
Λ_x	Lambda	Faktorenladungsmatrix für die manifesten Variablen x
Λ_y	Lambda	Faktorenladungsmatrix für die manifesten Variablen y
δ	delta	Meßfehler der manifesten Variablen x
ϵ	epsilon	Meßfehler der manifesten Variablen y
θ_δ	thetadelta	Kovarianzmatrix der Meßfehler δ
θ_ϵ	thetaepsilon	Kovarianzmatrix der Meßfehler ϵ
τ_x	tau	Mittelwert (intercept) der manifesten Variablen x
τ_y	tau	Mittelwert (intercept) der manifesten Variablen y
κ	kappa	Mittelwert der latenten Variablen ξ
α	alpha	Mittelwert der latenten Variablen η
$\bar{x}$,$\bar{y}$		Vektor der empirischen Mittelwerte der manifesten Variablen
μ	mü	Vektor der geschätzten Mittelwerte der manifesten Variablen
S		empirische Kovarianzmatrix der manifesten Variablen
Σ	Sigma	geschätzte Kovarianzmatrix der manifesten Variablen

Buchstabe	Name	Spezifische Bedeutung in diesem Buch
β	beta	Autoregressiver Regressionskoeffizient (Abschnitt 3.3)
Δ	Delta	latente Differenz (Abschnitt 3.4)
$\hat{\lambda}_L$	lambda	geschätzte untere Grenze des Konfidenzintervalls (Abschnitt 3.7)
$\hat{\lambda}_U$	lambda	geschätzte obere Grenze des Konfidenzintervalls (Abschnitt 3.7)
ν	nü	Erwartungswert der Poisson-Verteilung bzw. negativen Binomialverteilung (Abschnitt 3.7)
Θ	Theta	Parametervektor im Bayesianischen Modell (Abschnitt 3.6)
Θ	Theta	Parametervektor für die kompletten Daten (Abschnitt 3.8)
Ψ	Psi	Parametervektor für die unvollständigen Daten (Abschnitt 3.8)
γ	gamma	logistischer Regressionskoeffizient (Abschnitt 3.8)
Γ	Gamma	Matrix der logistischen Regressionskoeffizienten (Abschnitt 3.8)

1 Einführung

Längsschnittstudien mit wiederholten Messungen sind notwendig, um Stabilität und Wandel von interessierenden Sachverhalten bei Individuen und Gruppen zu untersuchen. Paneldaten eignen sich einerseits zur Analyse intraindividueller Entwicklungen substanzieller Variablen über einen gegebenen Zeitraum als auch andererseits zur Betrachtung interindividueller Dispositionen.

Die statistische Modellierung von Entwicklungsprozessen hat mit den so genannten Wachstumsmodellen (*growth curve models*) in verschiedenen Forschungsbereichen wie Biometrie (z. B. Rao, 1958; Liang & Zeger, 1986), Bildungsstatistik (z. B. Goldstein, 1987; Bryk & Raudenbush, 1992) und Psychometrie (z. B. Tucker, 1958; McArdle & Epstein, 1987) eine starke Verbreitung gefunden. Zuvor hat Wishart (1938) die mathematische Grundlage für die Wachstumsmodelle durch seine Längsschnittanalyse über die Gewichtsveränderung von Schweinen gelegt.

Die statistische Modellierung von latenten Wachstumsmodellen kann mit der exploratorischen Faktorenanalyse oder der Hauptkomponentenanalyse erfolgen. Die Faktoren des latenten Wachstumsmodells werden hierbei als Veränderungsfaktoren oder -komponenten interpretiert. Die Faktorenladungen repräsentieren den Grad der Abhängigkeit dieser Messungen von den genannten Veränderungsfaktoren (Rao, 1958; Tucker, 1958). Eine der wesentlichen Probleme dieser Ansätze ist das Rotationsproblem. Es existiert kein eindeutiges Rotationskriterium, daß zu einer Faktorenladungsmatrix führt, aus der die substantiellen Veränderungen in den Faktoren interpretierbar wären. Daher haben sich Modellierungen innerhalb des konfirmatorische Faktorenmodells weitgehend durchgesetzt, da sich hier das Rotationsproblem nicht stellt. Da das konfirmatorische Faktorenmodell ein Submodell des allgemeinen Kovarianzstrukturgleichungsmodell ist (vgl. Reinecke, 2014), lassen sich latente Wachstumsmodelle formal als Strukturgleichungsmodelle darstellen. Meredith und Tisak (1990) verwenden hierzu den Begriff *latent curve analysis*.

Alternative Bezeichnungen von latenten Wachstumsmodellen sind in verschiedenen Disziplinen mit statistischen Forschungsentwicklungen üblich. Beispielsweise werden die Bezeichnungen *random-effects analysis of variance* und *random coefficient modeling* in der Biometrie verwendet. In der empirischen Bildungsforschung sind die Bezeichnungen *hierarchical linear modeling* und *multilevel modeling* geläufig.

Latente Wachstumsmodelle können auch als Mehrebenenlängsschnittmodelle bezeichnet werden, da durch die wiederholte Datenerhebung im Paneldesign ein hierarchischer Datensatz erzeugt wird. In einigen Lehrbüchern (z. B. in Hox, Moerbeek & van de Schoot, 2017) wird sowohl auf die statistische Modellierung der Wachstumskurven als Mehrebenenmodell als auch als Kovarianzstrukturgleichungsmodell eingegangen. Die durch ein Paneldesign erzeugten Daten können in einem langen (*long-format*) oder einem weiten (*wide-format*) Format vorliegen. Die wiederholt gemessenen Variablen besetzen beim langen Format eine Datenspalte (z. B. als y bezeichnet), wobei zur Identifikation der Meßzeitpunkte $t_1, t_2, \ldots T$ eine weitere Variable (z. B. als *time* bezeichnet) notwendig wird. Der Paneldatensatz im langen Format hat die Länge $n \times T$. Diese Format wird verwendet, wenn Entwicklungsprozesse auf der Grundlage eines Mehrebenenmodells untersucht werden sollen.

Beim weiten Format wird für jeden Meßzeitpunkt $t_1, t_2, \ldots T$ eine Datenspalte für die Variable y angelegt. Die Anzahl der Spalten der Variablen y entspricht demnach der Anzahl der Meßzeitpunkte. Die Länge des Datensatzes kann variieren, je nachdem ob alle Untersuchungseinheiten unabhängig von der Häufigkeit ihrer Teilnahme an der Erhebung berücksichtigt werden oder ob nur diejenigen im Datensatz aufgenommen werden, die an allen Meßzeitpunkten beteiligt waren. Das weite Datenformat wird verwendet, wenn Entwicklungsprozesse auf der Grundlage eines Strukturgleichungsmodells untersucht werden sollen (vgl. auch Grimm, Ram & Estabrook, 2017). Auf die letztere der beiden Möglichkeiten wird sich dieses Lehrbuch konzentrieren.

Die Vorteile von Wachstumsmodellen beziehen sich nicht nur auf die Möglichkeiten, Veränderungen und Entwicklungen über die Zeit zu modellieren, sondern auch auf Techniken, diese Veränderungen und Entwicklungen als latente Variablen in komplexere Strukturgleichungsmodelle zu berücksichtigen. Die Datengrundlage muß auf wiederholte Messungen gleicher Untersuchungseinheiten basieren, da *intra*individuelle Veränderungen und Entwicklungen (innerhalb gleicher Untersuchungseinheiten) und *inter*individuelle Veränderungen und Entwicklungen (innerhalb verschiedener Untersuchungseinheiten) gleichzeitig untersucht werden können. Der integrative Charakter von Wachstumsmodellen zur statistischen Modellierung von Veränderungsprozessen wird in der Literatur immer wieder hervorgehoben (z. B. bei Voelkle, 2007).

Im folgenden werden einige, typische Fragestellungen aufgelistet, die mit Wachstumsmodellen untersucht werden können:

1. Wie verläuft der Entwicklungsprozeß einer mehrfach untersuchten Variable und mit welcher Einfachheit bzw. Komplexität kann der durchschnittliche Verlauf in Form einer Trajektorie geschätzt werden?
2. Gibt es bedeutsame erklärende Variablen für den Entwicklungsprozeß?

3. Wie groß ist die Heterogenität der Entwicklungsverläufe und welche Variablen stehen zur Verfügung, um das Ausmaß beobachteter Heterogenität abzuschätzen?
4. Welchen Stellenwert hat unbeobachtete Heterogenität im Entwicklungsverlauf und lassen sich unterschiedliche „Entwicklungstypen" identifizieren?
5. Steht der untersuchte Entwicklungsprozeß mit anderen Entwicklungen im Zusammenhang und können beispielsweise parallele Verläufe angenommen werden?

Da im vorliegenden Buch Modellierungen von latenten Wachstumskurven auf der Basis von Strukturgleichungen thematisiert werden, sollte man mit der Grundkonzeption dieses Gleichungssystems vertraut sein (vgl. auch die Übersicht der Kapitel im Handbuch von Hoyle, 2023). Gute und leicht verständliche Einführungen bieten Schumacker und Lomax (2016), Kline (2016) und Geiser (2013). Für mehr an statistische Grundlagen der Modellbildung mit Strukturgleichungen interessierte Personen stehen Bollen (1989), Kaplan (2009) oder Mulaik (2009) zur Verfügung. Im deutschsprachigen Raum sind die Werke von Reinecke (2014), Steinmetz (2015), Weiber und Sarstedt (2021), Geiser (2011) sowie Kleinke, Schlüter und Christ (2017) zu nennen. Die beiden zuletzt genannten Bücher legen ihre Schwerpunkte auf das auch hier verwendete Programm M*plus* (L. K. Muthén & Muthén, 1998-2017).

Das Buch ist folgendermaßen gegliedert: Die Kapitel 2 und 3 erörtern unterschiedliche Spezifikationen von Wachstumsmodellen in theoretischer und formaler Hinsicht, wobei zunächst einfache Modellspezifikationen im Vordergrund stehen und – darauf aufbauend – komplexere Modellierungsansätze erörtert werden.

Die Skalierungsmöglichkeiten der Zeit, die Annahmen und Schätzfunktionen sowie die statistischen Kriterien zur Modellbeurteilung schließen sich in Kapitel 4 an. In Kapitel 5 wird das verwendete Datenmaterial für die berechneten Wachstumsmodelle beschrieben.

In den Kapitel 6 und 7 werden analog zu den theoretischen und formalen Erörterungen der Modelle in den Kapiteln 2 und 3 entsprechende Anwendungsbeispiele vorgestelllt und diskutiert.

Die technische Umsetzung der Anwendungsbeispiele erfolgte – bis auf eine Ausnahme – mit dem Programm M*plus* (L. K. Muthén & Muthén, 1998-2017), in dem ein sehr breites Spektrum von Modellierungsmöglichkeiten implementiert ist. Die zeitkontinuierlichen Wachstumsmodelle in Abschnitt 3.9 wurden mit dem R-Paket `ctsem` (Driver & Voelkle, 2018) berechnet.

Da die Komplexität der vorgestellten Modelle in einem Einführungsbuch begrenzt ist, wird auf bestimmte spezielle Modellierungsmöglichkeiten mit Wachstumskurven im weiteren nicht eingegangen. Hierzu gehören Modelle, die verschiedene Formen exponentieller und anderer nicht linearer Funktionen zu Grunde legen (vgl. Grimm & Ram, 2009; Grimm, Ram & Estabrook, 2010; Grimm et al., 2017; Long & Ryoo, 2010; Grimm, Ram & Hamagami, 2011; Boedeker, 2021), Modelle mit Interaktionsvariablen (vgl. Li, Duncan & Acock, 2000; Preacher, Curran & Bauer, 2006) sowie Modelle, die kein intervallskaliertes Meßniveau annehmen (vgl. Mehta, Neale & Flay, 2004; Grimm et al., 2017).

Alle verwendeten Daten und Programmbeispiele können unter `http://www.sozialwissenschaftliche-forschungsmethoden.de/` heruntergeladen und für eigene Übungen verwendet werden.