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. 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, 2005), 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, 2010) wird sowohl auf die statistische Modellierung der Wachstumskurven als Mehrebenenmodell als auch als Kovarianzstrukturgleichungsmodell eingegangen. Auf die letztere der beiden Möglichkeiten wird sich dieses Lehrbuch konzentrieren. In der angelsächsischen Literatur hat sich hierfür die Bezeichnung *latent growth curve model* durchgesetzt.

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 des Kovarianzstrukturgleichungsmodells thematisiert werden, sollten Leser mit diesem Modellansatz vertraut sein. Gute und leicht verständliche Einführungen bieten Schumacker und Lomax (2010), Kline (2010) und die jeweils auf verschiedene Strukturgleichungsprogramme basierenden Bücher von Byrne (2006, 2009, 2011). Erfahrene Leser können auch auf Bollen (1989), Kaplan (2009) oder Mulaik (2009) zurückgreifen. Im deutschsprachigen Raum sind die Werke von Reinecke (2005), Weiber und Mühlhaus (2009), Geiser (2011) sowie Christ und Schlüter (2012) zu nennen. Die beiden zuletzt genannten Werke legen ihre Schwerpunkte auf das auch hier verwendete Programm M*plus* (L. K. Muthén & Muthén, 1998-2010).

Die vorgenommene Gliederung des Buches ist denkbar einfach: Kapitel 2 erörtert unterschiedliche Spezifikationen von Wachstumsmodellen in theoretischer und formaler Hinsicht. Kapitel 3 zeigt entsprechende Anwendungsbeispiele. Zunächst werden in Kapitel 2 ein-, zwei- und mehrfaktorielle Wachstumsmodelle vorgestellt. Konditionale und parallele Modellierungen erweitern das Spektrum. Eine höhere Komplexität wird durch autoregressive Wachstumsmodelle, durch Modelle mit latenten Differenzen, durch Faktoren höherer Ordnung sowie durch Mischverteilungsmodelle erreicht. Die Skalierungsmöglichkeiten der Zeit, die Annahmen und Schätzfunktionen sowie die statistischen Kriterien zur Modellbeurteilung schließen sich an. Abgeschlossen wird das Kapitel durch einen Abschnitt, der sich mit den Möglichkeiten der Modellierung von Datenausfällen befaßt. Kapitel 3 stellt zunächst das verwendete Datenmaterial vor. Die Reihenfolge der berechneten Beispiele orientiert sich an der Abfolge der theoretischen und formalen Erörterungen in Kapitel 2. Obwohl fast alle Modelle mit verbreiteten Strukturgleichungsprogrammen (z. B. EQS, LISREL, AMOS) berechnet werden können, ist hier aus Kapazitätsgründen nur das Programm M*plus* (L. K. Muthén & Muthén, 1998-2010) verwendet worden, in dem ein sehr breites Spektrum von Modellierungsmöglichkeiten implementiert wurde. Zudem können in den Abschnitten 2.8 und 3.12 vorgestellten und angewendeten Mischverteilungsmodelle derzeit nur mit dem Programm M*plus* berechnet werden.

Da die Komplexität der vorgestellten Modelle in einem Einführungsbuch begrenzt sein sollte, 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, Ram & Hamagami, 2011; Long & Ryoo, 2010), Modelle mit Interaktionsvariablen (vgl. Li, Duncan & Acock, 2000; Preacher, Curran & Bauer, 2006) sowie Modelle, die nicht intervallskaliertes Meßniveau bei den Variablen annehmen (vgl. Mehta, Neale & Flay, 2004).

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