

Einführung

Lineare Strukturgleichungsmodelle sind aus der verhaltenswissenschaftlichen Forschung nicht mehr wegzudenken. Sie sind ein sehr nützliches Werkzeug, um Hypothesen über Beziehungen zwischen Variablen zu prüfen und – mehr noch – Implikationen kausaler Strukturen zu testen. Während in den ersten Jahrzehnten der Nutzung von „SEM“ (structural equation modeling)¹ der Enthusiasmus groß war und Parameterschätzungen vorschnell als wirkliche, kausale Effekte interpretiert wurden, hat sich in der letzten Zeit genau das Gegenteil eingestellt – nämlich, dass kausales Denken quasi aus SEM verbannt und die Parameter lediglich in Termini von „Beziehungen“ interpretiert werden. Dies stellt allerdings einen Rückschritt dar. SEM sind, wenn sie richtig angewendet werden, eine mächtige Methode, um bei Fehlen von randomisierten Experimenten (d.h. bei Feldstudien) kausale Hypothesen zu testen, ohne gleichermaßen in die Falle der naiven kausalen Interpretation der Koeffizienten in Modellen zu tappen.

Dieses Buch bietet eine Einführung in die Hintergründe, Prinzipien, Möglichkeiten und Grenzen von SEM. Unter Verwendung eines realen Modells wird eine Schritt-für-Schritt-Anleitung gegeben, um Einsteigern einen schnellen und praktikablen Start zu ermöglichen. Wegen der Kürze können die Themen nicht so detailliert behandelt werden, wie dies in gängigen Standardwerken der Fall ist. Daher sei eine weitere Vertiefung angeraten. Vieles in diesem Buch wurde beeinflusst von Hayduk (1987), Kline (2011), Shipley (2000) und Bollen (1989). Die Illustration erfolgt mit der freien Statistiksoftware R, die in den letzten Jahren eine rasante Verbreitung gefunden hat. Dabei sollen nicht nur die spezielle Modellierung mit dem R-Paket „lavaan“ behandelt werden (siehe dazu auch Beaujean, 2014), sondern auch wichtige Vorgehensweisen zur Vor- und Nachbereitung. Dies sind zum Beispiel das Einlesen von Daten aus anderen Programmen, oder das Überprüfen und Veranschaulichen der Daten im Rahmen der

¹ In diesem Buch soll die Abkürzung SEM sowohl für Strukturgleichungsmodelle als auch für den gesamten Ansatz der Modellierung benutzt werden, da diese Abkürzung auch von nicht-englischen Forschern und Forscherinnen üblicherweise verwendet wird.

Überprüfung zentraler Annahmen, die ein Modell beinhaltet. Vorkenntnisse von R werden nicht vorausgesetzt. Allerdings sind Vorkenntnisse in Varianzen und Kovarianzen von Variablen, der Regressionsanalyse und deskriptiver und Inferenzstatistik hilfreich. Der Fokus liegt dabei nicht auf methodischen und statistischen Aspekten von SEM (dies ist in den o.g. Büchern ausführlicher und auch fundierter behandelt), sondern vielmehr in der Verknüpfung von Theorie und Modell. Dabei werden diejenigen Aufgaben und Probleme diskutiert, die bei der Modellierung in der Regel auftauchen. Diese bestehen u.a. in der adäquaten Übersetzung von theoretischen Konstrukten in ein Modell, der Präzisierung der eigenen Vorstellungen, der Übersetzung dieser Vorstellungen in ein Modell und Reflektion der kausalen Implikationen dieser Vorstellungen. Schließlich soll der produktive Umgang mit „nicht-fittenden“ Modellen behandelt werden.

1.1 Was ist ein Kausalmodell?

Das Ziel der Anwendung von SEM besteht darin, ein Kausalmodell zu spezifizieren, das die theoretisch erwarteten kausalen Effekte von **Variablen** repräsentiert. Eine Variable ist eine Dimension, auf der sich Fälle (z.B. Personen) graduell unterscheiden. Dies können beispielsweise Unterschiede in der Häufigkeit eines Verhaltens, im Ausmaß der Zustimmung zu bestimmten Aussagen, in der Anzahl von zählbaren Ereignissen oder in der Evaluation eines Zustandes (z.B. Zufriedenheit) sein. Die Variablen des Modells repräsentieren theoretische oder **hypothetische Konstrukte**, wie sie in wissenschaftlichen Theorien enthalten sind, oder – genauer – diejenigen **empirischen Phänomene**, die diese Konstrukte zu beschreiben versuchen (vgl. Kap. 3). Das Modell beschreibt daher einen Ausschnitt der Realität, da in ihm lediglich eine Auswahl von Variablen enthalten ist. Ob und wann dies problematisch ist, soll später erörtert werden.

Weiterhin beschreibt ein Modell eine mehr oder weniger ausgeprägte *Struktur von Beziehungen* zwischen den Variablen, die sich sehr gut in **Pfaddiagrammen** (siehe nächster Abschnitt) als Struktur von Pfeilen illustrieren lässt. Auch wenn in solchen Diagrammen meist die expliziten Beziehungen ins Auge fallen, sei hier bereits angemerkt, dass die Struktur erst durch die Nicht-Beziehungen („Restriktionen“) innerhalb des Modells vollständig definiert ist. So hat in einem Modell nicht jede Variable auf jede andere einen Effekt. Es sind diese nicht-sichtbaren sowie impliziten Restriktionen, die sowohl einen **Test** des Modells ermöglichen, als auch den Grund für einen **misfit** (d.h. eine inakzeptable Vorhersage der Daten) ausmachen. Das Modell lässt sich schließlich als ein System von

linearen **Strukturgleichungen** interpretieren, in dem abhängige Variablen als die Folge des gewichteten Einflusses anderer Variablen beschrieben werden .

1.2 Pfaddiagramme als Illustrationen von Modellen

Wie oben erwähnt, sind **Pfaddiagramme** eine sehr hilfreiche Methode, um ein Modell darzustellen. Ein Pfaddiagramm enthält alle postulierten Effekte der Variablen, sowie bivariate Beziehungen (Kovarianzen), für die keine spezifische kausale Hypothese besteht. Insbesondere in der frühen Phase der Konzeptualisierung eines Forschungsprozesses ist der Detailgrad eines postulierten Forschungsmodells oft sehr gering und hypothetisierte Beziehungen zwischen theoretischen Konstrukten eher abstrakt und vage. Ein Pfaddiagramm erfordert einen höheren Präzisionsgrad in der Konzeptualisierung der Variablen und ihrer Effekte, der über das theoretisch oft vage „Konstrukt A ist assoziiert mit Konstrukt B“ hinaus geht. Dies ist v.a. der Fall wenn Konstrukte multidimensional sind. Ein Pfaddiagramm ist daher enorm hilfreich, um eine eher abstrakte Theorie in ein spezifisches Modell zu übersetzen.

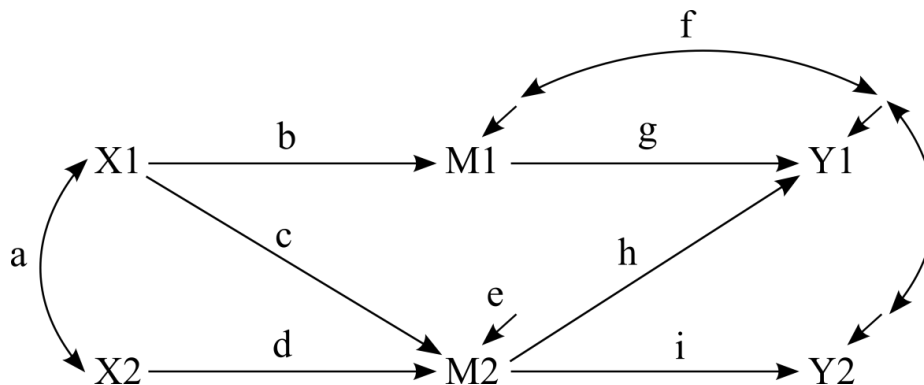


Abb. 1: Pfaddiagramm

Abb. 1 enthält ein exemplarisches Pfaddiagramm. In dem entsprechenden Modell gibt es sechs Variablen (X1, X2, M1, M2, Y1, Y2). Die Variablen X1 und X2 nennt man **exogene Variablen**; sie entsprechen den unabhängigen Variablen in herkömmlichen Forschungsdesigns, da von ihnen nur Effekte ausgehen, aber keine empfangen werden. Wir werden den Buchstaben „X“ in der Regel für solche exogenen Variablen verwenden. Die Variablen M1 – Y2 sind **endogene Variablen**, weil sie Effekte empfangen; sie entsprechen den abhängigen Variablen in herkömmlichen Designs.

Weiterhin enthält das Pfaddiagramm vier verschiedene Arten von Pfeilen:

1) unidirektionale Pfeile zwischen zwei Variablen (z.B. b),

- 2) unidirektionale Pfeile ohne Ursprung (z.B. e),
- 3) bidirektionale Pfeile zwischen Variablen (z.B. a) und
- 4) bidirektionale Pfeile zwischen Pfeilen (z.B. f).

Die unidirektionalen Pfeile zwischen zwei Variablen symbolisieren die **direkten Effekte** einer Variablen auf die andere. Zum Beispiel sieht das Modell vor, dass X1 direkt M1 und M2 beeinflusst. Hintereinander geschaltete Pfeile, die zwei Variablen verbinden, definieren einen **Pfad** – unabhängig davon, ob die Pfeile in dieselbe Richtung weisen. So ist cd ein Pfad, der X1 und X2 verbindet. Besteht ein Pfad aus Pfeilen, die in dieselbe Richtung weisen, spricht man von einem **indirekten Effekt**. So beeinflusst X1 die endogene Variable Y1 indirekt oder *vermittelt* über M1. Damit übernimmt M1 die Rolle eines **Mediators** (MacKinnon, Fairchild, & Fritz, 2007). Der indirekte Effekt von X1 (über M1) auf Y1 ergibt sich aus dem Produkt der direkten Effekte b und g. Schließlich gibt der **totale Effekt** einer Variablen auf eine andere den Gesamteffekt an, den diese Variable über alle direkten und indirekten Pfade hat. Zum Beispiel ist der totale Effekt von X1 auf Y1 die Summe der beiden indirekten Effekte bg und ch.

Die unidirektionalen Pfeile ohne Ursprung kennzeichnen den **Fehlerterm** der Variable. Er repräsentiert alle nicht im Modell befindlichen weiteren Einflussvariablen der Variable sowie Messfehler. Die bidirektionalen Pfeile zwischen Variablen kennzeichnen **Kovarianzen zwischen exogenen Variablen**, deren Grund nicht im Modell spezifiziert ist. So könnte der Grund für die Kovarianz zwischen X1 und X2 ein Einfluss jeweils einer dieser Variablen auf die andere sein (z.B. $X1 \rightarrow X2$), oder aber die Konsequenz einer oder mehrerer gemeinsamer Ursachen (vgl. 2.1). Kovarianzen zwischen endogenen Variablen gibt es in einem Modell nicht, da sie eine Funktion der Varianzen und Effekte vorangehender Variablen sind. So ist z.B. die Kovarianz zwischen Y1 und Y2 kein schätzbarer Modellparameter, sondern eine direkte Folge der Varianz von M2 und der beiden Effekte h und i. Stattdessen sind die bivariaten Pfeile zwischen den Fehlertermen **Kovarianzen zwischen den Fehlertermen** dieser Variablen und nicht zwischen den Variablen selbst. Sie repräsentieren *weitere*, nicht im Modell adressierte, gemeinsame Ursachen der betreffenden Variablen. Wie wir später noch sehen werden, macht es Sinn, diese Kovarianzen zu schätzen, außer man vertritt die Hypothese, dass alle gemeinsamen Ursachen im Modell enthalten sind. Auf diese Implikationen eines Modells für Varianzen und Kovarianzen von Variablen werden wir im nächsten Kapitel genauer eingehen.