

Konfirmatorische Faktorenanalyse II

Regressionsmodelle für Politikwissenschaftler

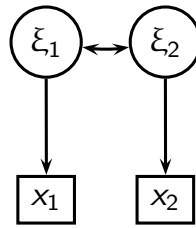
Wiederholung
Das vollständige Modell
Kausalität
Fazit

Wie war das mit der NPD?

„Ein Drittel der NPD=Wähler macht in einer Umfrage keine Angabe zu ihrem Wahlverhalten (1), ein weiteres Drittel gibt an nicht zur Wahl zu gehen oder die Union zu wählen (2). Liegt bei (1) beziehungsweise (2) ein missing data Mechanismus vor, und wenn ja, welcher?“

- ▶ (1) Hier fehlen Werte. Der Mechanismus ist NI, da die Wahrscheinlichkeit des Ausfalls von y vom Wert von y abhängt
- ▶ (2) Hier werden falsche Angaben gemacht – kein missing data, sondern ein systematischer Meßfehler

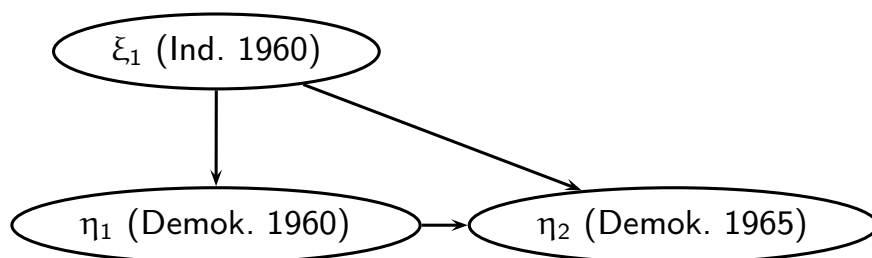
Wie war das mit der Reliabilität?



„Aus dem Grundstudium kennen Sie das Konzept der Reliabilität, aus dem Aufbaukurs die Attenuation-Formel. Wo zeigt sich Reliabilität bei der Faktorenanalyse? Kann diese die Attenuation-Formel ersetzen?“

- ▶ Reliabilität bezieht sich auf die Zuverlässigkeit der Messung, d. h. auf die Stärke des Zusammenhangs zwischen Indikator und latenter Variable → entspricht den λ im Modell
- ▶ Attenuation-Formel geht von Schätzungen für λ_{11} und λ_{22} aus, um ϕ_{12} zu schätzen
- ▶ *Nicht* zu ersetzen, wenn es jeweils nur einen Indikator gibt

Wie sieht das Beispiel aus?



- ▶ Industrialisierung 1960 ist *exogen*, weil sie innerhalb des Modells nicht erklärt wird → ξ
- ▶ Demokratie 1960 und 1965 sind *endogen*, weil sie innerhalb des Modells auf andere Variablen zurückgeführt werden → η
- ▶ η_1 hängt von ξ_1 und zufälligen Einflüssen ab: $\eta_1 = \gamma_{11}\xi_1 + \zeta_1$
- ▶ η_2 hängt von ξ_1 , η_1 und zufälligen Einflüssen ab:
 $\eta_2 = \beta_{21}\eta_1 + \gamma_{21}\xi_1 + \zeta_2$

Noch mehr Terminologie?

Matrix	Dimension	Kovarianz	Dimension	Bedeutung
ξ	$(n \times 1)$	Φ	$(n \times n)$	exogene latente Variablen
x	$(q \times 1)$	Σ	$(p + q) \times (p + q)$	Indikatoren für ξ
Λ_x	$(q \times n)$			Faktorladungen für x
δ	$(q \times 1)$	Θ_δ	$(q \times q)$	Meßfehler für x
η	$(m \times 1)$		$(m \times m)$	endogene latente Variablen
ζ	$(m \times 1)$	Ψ	$(m \times m)$	latente Fehler
y	$(p \times 1)$	Σ	$(p + q) \times (p + q)$	Indikatoren für η
Λ_y	$(p \times n)$			Faktorladungen für y
ϵ	$(p \times 1)$	Θ_ϵ	$(p \times p)$	Meßfehler für y
B	$(m \times m)$			Koeffizienten für endogene latente Variablen
Γ	$(m \times n)$			Koeffizienten für exogene latente Variablen

Wie sieht das vollständige Modell nun aus?

- ▶ Das bereits bekannte Modell für die konfirmatorische Faktorenanalyse wird ergänzt. . .
- ▶ Durch *endogene* latente Variablen, die auf exogene latente Variablen und
- ▶ zufällige Einflüsse zurückgeführt werden
- ▶ Für die endogenen latenten Variablen wird ebenfalls ein Meßmodell benötigt
- ▶ Hauptdiagonale von B immer gleich null
- ▶ Standardannahmen für die zusätzlichen zufälligen Einflüsse

Was passiert weiter?

- ▶ Modell in bekannter Weise graphisch darstellbar
- ▶ Zerlegung der Kovarianzen ermöglicht Schätzung
- ▶ Pfadmodell gestattet Zerlegung in direkte und indirekte Effekte (Beispiel auf Seite 37)
- ▶ Gesamteffekt = Direkter Effekt + Indirekte Effekte
- ▶ Welchen Effekt hat Industrialisierung 1960 auf Demokratie 1965?
 $\gamma_{21} + \gamma_{11}\beta_{21}$
- ▶ Welchen Effekt hat Industrialisierung 1960 auf den Indikator y_8 ?
 $0 + \gamma_{21}\lambda_{11} + \gamma_{11}\beta_{21} + \lambda_{11}$

Was ist Kausalität?

- ▶ Lange und kontroverse Debatten
- ▶ Weitverbreitete Arbeitsdefinition: y_1 ist von allen Einflüssen außer x_1 isoliert, eine Veränderung von x_1 geht mit einer Veränderung von y_1 einher → x_1 ist eine Ursache von y_1
- ▶ Drei wesentliche Aspekte
 1. Isolation
 2. Assoziation
 3. Richtung
- ▶ In der Realität ist Isolation nicht möglich; $y_1 = \gamma_{11}x_1 + \zeta_1$; keine Kovarianz zwischen x_1 ζ_1 → in der Praxis nicht zu prüfen → „Pseudo-Isolation“
- ▶ Etwas besser: Mehrere unabhängige Variablen $x_2, x_3 \dots$

Was passiert ohne Pseudo-Isolation?

- ▶ Intervenierende Variable wird ausgelassen (siehe S. 46)
 - ▶ Schätzung für γ_{21} konvergiert zum *totalen* Effekt von x_1
 - ▶ Problematisch, wenn direkte und indirekte Effekte verschiedenes Vorzeichen haben (Beispiel Intelligenz, Langweile, Fehler am Fließband)
- ▶ Meistens schlimmer: Gemeinsame Hintergrundvariable nicht modelliert
 - ▶ „Scheinkorrelation“ (spurious correlation)
 - ▶ „scheinbare Non-Korrelation“ (Suppressor)
- ▶ Non-rekursive Zusammenhänge (wechselseitige Verursachung)
 - ▶ Exogene Variablen nicht wirklich exogen; Assoziation zwischen systematischen und zufälligen Einflüssen
 - ▶ Ohne besondere Vorkehrungen: verzerrte Schätzungen
- ▶ Zusammenhänge zwischen zufälligen Einflüssen (S. 56) oder Fehlspezifikationen können zum gleichen Problem führen

Was kann beim Zusammenhang schiefgehen?

- ▶ Bivariater Zusammenhang weder notwendig noch hinreichend für Kausalität → Kontrolle anderer Einflüsse wichtig
- ▶ *De facto* keine deterministischen Zusammenhänge
 - ▶ Standardfehler wichtig
 - ▶ Autokorrelation und Heteroskedastizität
 - ▶ Multikollinearität
- ▶ Replikation, Replikation, Replikation

Was kann bei der Richtung der Kausalität schiefgehen?

- ▶ Seit Hume wird Bedeutung der zeitlichen Reihenfolge hervorgehoben
- ▶ In der Praxis kaum prüfbar/zu restriktiv
- ▶ Häufig Verwendung von lagged variables → kann Richtung der Kausalität *nicht* absichern, wenn x autokorreliert (S. 62-63)
- ▶ Etwas besser funktioniert es, wenn y_t auf y_{t-1} und auf x_{t-1} regrediert wird; auch hier Fallen (Autokorrelation der Fehler)
- ▶ Indikatoren für latente Variablen häufig eher Ursache als Wirkung

Was sind die wichtigsten Ergebnisse?

- ▶ Komplexere Theorien mit Hilfe von Strukturgleichungsmodellen angemessen modellierbar
- ▶ Berücksichtigung von Meß(fehler)theorien
- ▶ Aber: Kein Wundermittel, Standardprobleme bei der Modellierung und bei der Prüfung von Kausalität