

SEM 3: Definitions- und Strukturgleichungen, Identifikation & Parameterschätzung



We are happy to share our materials openly:

The content of these Open Educational Resources by Lehrstuhl für Psychologische Methodenlehre und Diagnostik, Ludwig-Maximilians-Universität München is licensed under CC BY-SA 4.0. The CC Attribution-ShareAlike 4.0 International license means that you can reuse or transform the content of our materials for any purpose as long as you cite our original materials and share your derivatives under the same license.

- In dieser Vorlesung geht es darum, wie die Parameter eines linearen Strukturgleichungsmodells konkret geschätzt werden.
- Die technischen Details sind mathematisch anspruchsvoll. Sie sind nicht im Fokus der Prüfung, tragen aber zum Gesamtverständnis bei.
- Warum ist ein grobes Verständnis der Parameterschätzung nützlich?
 - Vertieftes Verständnis des Modells und seinen Annahmen (siehe Definitionsgleichungen, Strukturgleichungen, Voraussetzungen der Schätzmethoden)
 - Wissen, wie man überprüft, ob ein gewünschtes Modell überhaupt schätzbar ist (siehe Identifizierbarkeit, Identifikationsgleichungen)
 - Auswahl zwischen verschiedenen Schätzalgorithmen in der Praxis (siehe Schätzmethoden)
 - Orientierung im Output in R und Hilfe zur Selbsthilfe bei Fehlermeldungen in der Praxis (siehe Unteridentifikation, Konvergenz)

Die Realität liefert Daten

	mimic	voice	hrtrt	att1	att2	att3	att4	errrs	siri
mimic	1.12								
voice	0.22	1.15							
heartrate	-0.05	0.05	1.19						
att1	0.01	0.01	0.15	0.94					
att2	0.04	0.03	0.02	0.13	1.09				
att3	0.05	-0.13	-0.03	-0.02	0.05	1.07			
att4	-0.06	0.11	-0.12	0.02	0.07	0.29	1.46		
errors	0.55	0.48	0.07	0.18	0.33	0.42	0.71	2.93	
siri	0.30	0.19	0.05	0.13	0.21	0.19	0.13	0.97	1.29

S empirische
Kovarianzmatrix



Das Modell sagt Daten vorher

	mimic	voice	hertrt	att1	att2	att3	att4	errors	siri
mimic	1.49								
voice	0.24	1.12							
heartrate	0.16	0.08	1.05						
att1	0.01	0.01	0.00	1.04					
att2	0.02	0.01	0.01	0.08	1.16				
att3	0.02	0.01	0.01	0.06	0.12	1.09			
att4	0.03	0.02	0.01	0.12	0.25	0.18	1.37		
errors	0.79	0.40	0.26	0.16	0.32	0.24	0.48	2.78	
siri	0.36	0.18	0.12	0.03	0.06	0.04	0.09	0.70	1.00

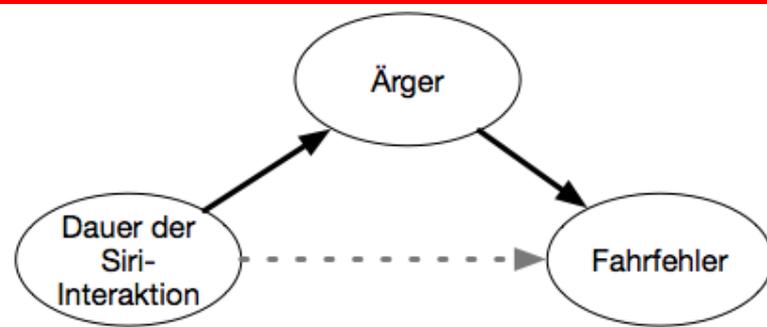
modell-implizierte
Kovarianzmatrix

$\hat{\Sigma}$



- Die modell-implizierte Kovarianzmatrix soll eigentlich mit der Populationskovarianzmatrix verglichen werden
- Da wir die Populationskovarianzmatrix nicht kennen, nehmen wir statt dessen die aus Stichprobendaten geschätzte Populationskovarianzmatrix (mit $\frac{1}{n-1}$ statt $\frac{1}{n}$; in R `cov(Daten)`)

Formales Kausalmodell eines Phänomens



Die Realität liefert Daten

	mimic	voice	hrtrt	att1	att2	att3	att4	errrs	siri
mimic	1.12								
voice	0.22	1.15							
heartrate	-0.05	0.05	1.19						
att1	0.01	0.01	0.15	0.94					
att2	0.04	0.03	0.02	0.13	1.09				
att3	0.05	-0.13	-0.03	-0.02	0.05	1.07			
att4	-0.06	0.11	-0.12	0.02	0.07	0.29	1.46		
errors	0.55	0.48	0.07	0.18	0.33	0.42	0.71	2.93	
siri	0.30	0.19	0.05	0.13	0.21	0.19	0.13	0.97	1.29

S empirische
Kovarianzmatrix

Das Modell sagt Daten vorher

	mimic	voice	hertrt	att1	att2	att3	att4	errors	siri
mimic	1.49								
voice	0.24	1.12							
heartrate	0.16	0.08	1.05						
att1	0.01	0.01	0.00	1.04					
att2	0.02	0.01	0.01	0.08	1.16				
att3	0.02	0.01	0.01	0.06	0.12	1.09			
att4	0.03	0.02	0.01	0.12	0.25	0.18	1.37		
errors	0.79	0.40	0.26	0.16	0.32	0.24	0.48	2.78	
siri	0.36	0.18	0.12	0.03	0.06	0.04	0.09	0.70	1.00

modell-implizierte
Kovarianzmatrix

$\hat{\Sigma}$

Wie kommt man vom graphischen Modell zur modell-implizierten Kovarianzmatrix?

	x1	x2	x3
x1	1.36	0.41	0.58
x2	0.41	1.39	0.45
x3	0.58	0.45	1.28

S empirische
Kovarianzmatrix

	x1	x2	x3
x1	$\sigma^2(x1)$	$\sigma(x1, x2)$	$\sigma(x1, x3)$
x2	-	$\sigma^2(x2)$	$\sigma(x2, x3)$
x3	-	-	$\sigma^2(x3)$

modell-implizierte
Kovarianzmatrix $\hat{\Sigma}$

Berechnung der maximalen **Freiheitsgrade einer Kovarianzmatrix (bekannte Parameter)***

Achtung: Bitte klar trennen von den **freien Parametern eines Modells** (welche auch „Freiheitsgrade“ genannt werden) und den Freiheitsgraden eines SEMs, welche auch „Freiheitsgrade“ genannt werden.

$$p = \frac{k \cdot (k + 1)}{2} = \frac{3 \cdot (3 + 1)}{2} = 6$$

- p = Anzahl der bekannten Parameter
- k = Anzahl der Items bzw. manifesten Variablen

Beispiel: 3 manifeste Variablen (x1, x2 und x3)

	x1	x2	x3
x1	$\sigma^2(x1)$	$\sigma(x1, x2)$	$\sigma(x1, x3)$
x2	-	$\sigma^2(x2)$	$\sigma(x2, x3)$
x3	-	-	$\sigma^2(x3)$

$p = 3 \cdot (3 + 1) / 2 = 6$

Diese 6 Parameter wollen wir mit dem Modell abbilden

- Im folgenden besprechen wir 3 Arten von Gleichungssystemen:
 - **Definitionsgleichungen:** Wie hängen die Variablen und Parameter im Modell zusammen? Jede Annahme zur Struktur des Modells explizit oder implizit enthalten. Vergleichbar mit Modellgleichung in der Regressionsanalyse.
 - **Strukturgleichungen:** Wie hängen die Modellparameter mit den Varianzen und Kovarianzen der manifesten Variablen zusammen? Zeigen, wie die modell-implizierte Kovarianzmatrix entsteht.
 - **Identifikationsgleichungen:** Hängen die Modellparameter nach der Berücksichtigung von Parameterfixierungen nur noch von bekannten Größen ab? Zeigen, ob das Modell identifiziert und damit grundsätzlich schätzbar ist.

Struktur- und Messmodell lassen sich in Definitionsgleichungen darstellen:

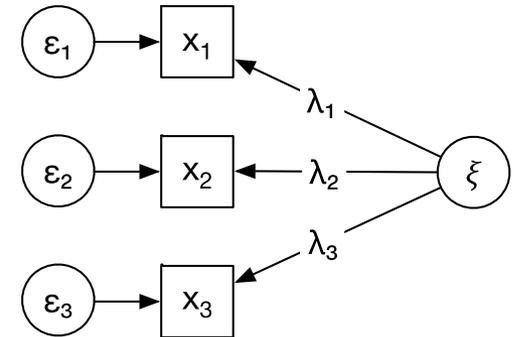
Definition der Ausprägungen endogener manifester Variablen:

$$X_1 = \lambda_1 \xi + \varepsilon_1$$

$$X_2 = \lambda_2 \xi + \varepsilon_2$$

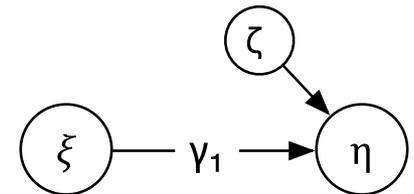
$$X_3 = \lambda_3 \xi + \varepsilon_3$$

Erinnerung: tau-
kongenerisches
Messmodell



Definition der Ausprägungen endogener latenter Variablen:

$$\eta = \gamma_1 \xi + \zeta$$



Implizite Annahmen/ Definition der Orthogonalitäten:

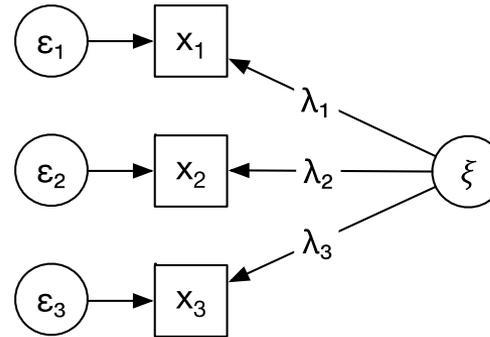
$\sigma(\varepsilon_i, \varepsilon_j) = 0$ mit $i \neq j$ → Die Fehlerterme der Indikatorvariablen sind untereinander unkorreliert.

$\sigma(\varepsilon_i, \xi) = 0$ → Die Fehlerterme der Indikatorvariablen kovariieren nicht mit den latenten Variablen

Generell: Nicht spezifizierte Pfade werden = 0 gesetzt

Die Definitionsgleichungen lassen sich nun in Strukturgleichungen überführen.
Diese sind nach den Elementen der Kovarianzmatrix aufgelöst, d.h., nach den (Ko-)Varianzen der manifesten Variablen des Modells!

$$\begin{aligned} X_1 &= \lambda_1 \xi + \varepsilon_1 \\ X_2 &= \lambda_2 \xi + \varepsilon_2 \\ X_3 &= \lambda_3 \xi + \varepsilon_3 \end{aligned}$$



$$\begin{aligned} \sigma(X_1, X_2) &= \sigma(\lambda_1 \xi + \varepsilon_1, \lambda_2 \xi + \varepsilon_2) \\ \sigma(X_1, X_3) &= \sigma(\lambda_1 \xi + \varepsilon_1, \lambda_3 \xi + \varepsilon_3) \\ \sigma(X_2, X_3) &= \sigma(\lambda_2 \xi + \varepsilon_2, \lambda_3 \xi + \varepsilon_3) \end{aligned}$$

Kovarianzen

sowie

$$\begin{aligned} \sigma^2(X_1) &= \sigma^2(\lambda_1 \xi + \varepsilon_1) \\ \sigma^2(X_2) &= \sigma^2(\lambda_2 \xi + \varepsilon_2) \\ \sigma^2(X_3) &= \sigma^2(\lambda_3 \xi + \varepsilon_3) \end{aligned}$$

Varianzen

	x1	x2	x3
x1	$\sigma^2(x1)$	$\sigma(x1, x2)$	$\sigma(x1, x3)$
x2	-	$\sigma^2(x2)$	$\sigma(x2, x3)$
x3	-	-	$\sigma^2(x3)$

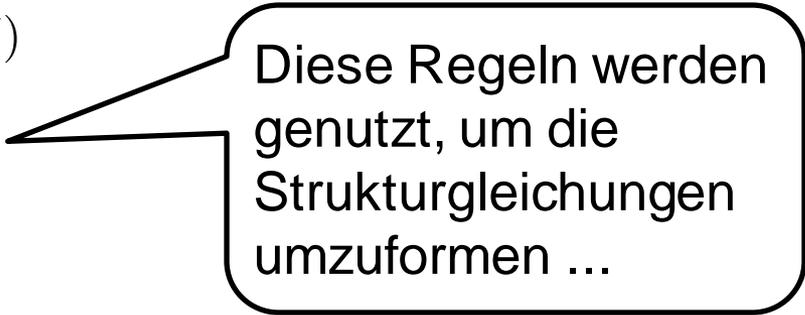
→ a und b sind Konstanten, X und Y Zufallsvariablen.

Einige Rechenregeln für Varianzen:

- $\sigma^2(a) = 0$
- $\sigma^2(a + X) = \sigma^2(X)$
- $\sigma^2(aX) = a^2 \sigma^2(X)$
- $\sigma^2(X + Y) = \sigma^2(X) + \sigma^2(Y) + 2\sigma(X, Y)$

Einige Rechenregeln für Kovarianzen:

- $\sigma(a, b) = 0$
- $\sigma(a, X) = 0$
- $\sigma(X, Y) = \sigma(Y, X)$
- $\sigma(X + a, Y + b) = \sigma(X, Y)$
- $\sigma(aX, bY) = ab\sigma(X, Y)$



Diese Regeln werden
genutzt, um die
Strukturgleichungen
umzuformen ...

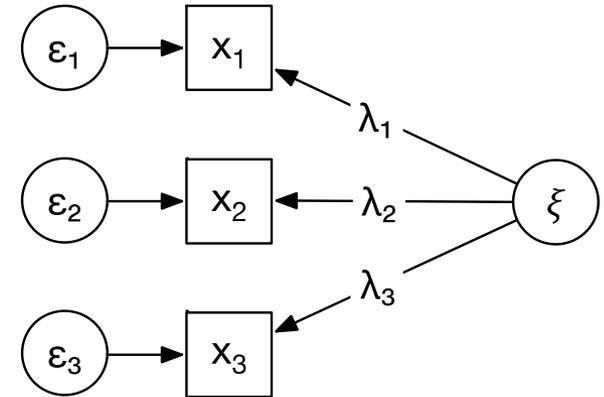
Umrechnung von **Kovarianzen**

am Beispiel X_1 und X_2 :

$$\sigma(X_1, X_2) = \sigma(\lambda_1 \xi + \varepsilon_1, \lambda_2 \xi + \varepsilon_2)$$

$$\sigma(X_1, X_2) = \lambda_1 \lambda_2 \sigma(\xi, \xi) + \lambda_1 \sigma(\xi, \varepsilon_2) + \lambda_2 \sigma(\xi, \varepsilon_1) + \sigma(\varepsilon_1, \varepsilon_2)$$

$$\sigma(X_1, X_2) = \lambda_1 \lambda_2 \sigma^2(\xi) \quad \underbrace{\hspace{1.5cm}}_{=0} \quad \underbrace{\hspace{1.5cm}}_{=0} \quad \underbrace{\hspace{1.5cm}}_{=0}$$



ξ , ε_1 und ε_2 sind
Zufallsvariablen,
 λ_1 und λ_2 sind Konstanten

Umrechnung von **Varianzen**

am Beispiel X_1 :

$$\sigma^2(X_1) = \sigma(\lambda_1 \xi + \varepsilon_1, \lambda_1 \xi + \varepsilon_1)$$

$$\sigma^2(X_1) = \lambda_1^2 \sigma^2(\xi) + \sigma^2(\varepsilon_1) + \underbrace{2\lambda_1 \sigma(\xi, \varepsilon_1)}_{=0}$$

$$\sigma^2(X_1) = \underbrace{\lambda_1^2 \sigma^2(\xi)}_{\text{= systematische Varianz}} + \sigma^2(\varepsilon_1)$$

= systematische
Varianz

Diese Pfade sind im Modell nicht
enthalten und sind deshalb = 0
(vgl. Modelle der klassischen
Testtheorie)

Modellannahme in diesem Beispiel:

Alle Pfade λ_x sind auf 1 fixiert

Kovarianz von vorheriger Folie:

$$\sigma(X_1, X_2) = \lambda_1 \lambda_2 \sigma^2(\xi)$$

$$\sigma(X_1, X_2) = \sigma^2(\xi)$$

Varianz von vorheriger Folie:

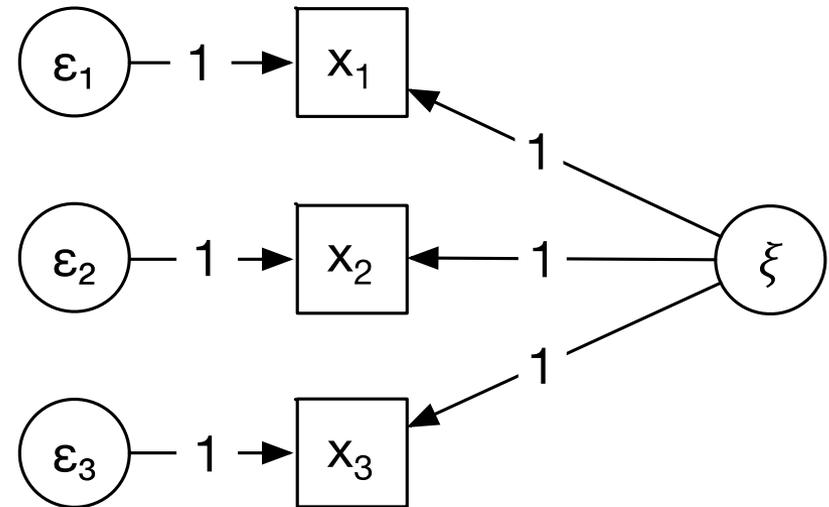
$$\sigma^2(X_1) = \lambda_1^2 \sigma^2(\xi) + \sigma^2(\varepsilon_1)$$

$$\sigma^2(X_1) = \sigma^2(\xi) + \sigma^2(\varepsilon_1)$$

Varianz der
beobachteten
Variable =

Systematische
Varianz der latenten
Variable +

Nicht erklärte
(Fehler)-Varianz,
„Uniqueness“



Annahme:
Alle Pfade auf 1 fixiert

$$\sigma(X_1, X_2) = \sigma^2(\xi)$$

$$\sigma(X_1, X_3) = \sigma^2(\xi)$$

$$\sigma(X_2, X_3) = \sigma^2(\xi)$$

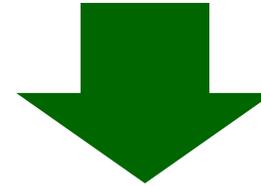
$$\sigma^2(X_1) = \sigma^2(\xi) + \sigma^2(\varepsilon_1)$$

$$\sigma^2(X_2) = \sigma^2(\xi) + \sigma^2(\varepsilon_2)$$

$$\sigma^2(X_3) = \sigma^2(\xi) + \sigma^2(\varepsilon_3)$$

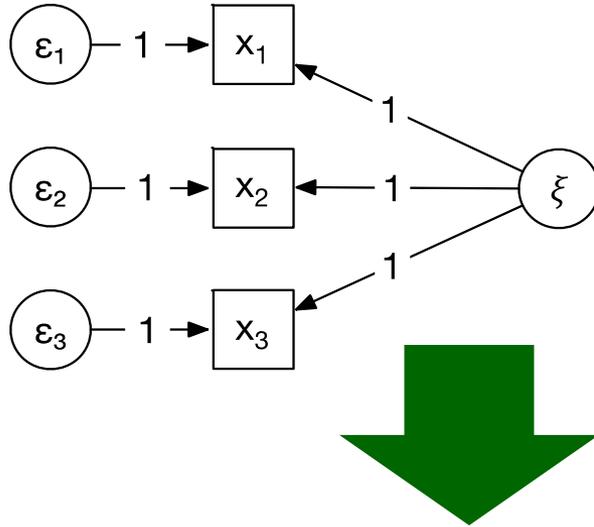
	x1	x2	x3
x1	$\sigma^2(x1)$	$\sigma(x1, x2)$	$\sigma(x1, x3)$
x2	-	$\sigma^2(x2)$	$\sigma(x2, x3)$
x3	-	-	$\sigma^2(x3)$

τ -Äquivalenz



	x1	x2	x3
x1	$\sigma^2(\xi) + \sigma^2(\varepsilon_1)$	$\sigma^2(\xi)$	$\sigma^2(\xi)$
x2	-	$\sigma^2(\xi) + \sigma^2(\varepsilon_2)$	$\sigma^2(\xi)$
x3	-	-	$\sigma^2(\xi) + \sigma^2(\varepsilon_3)$

Die modell-implizierten Kovarianzen



Dieses Strukturmodell
impliziert eine bestimmte
Struktur der Kovarianzmatrix

	x1	x2	x3
x1	$\sigma^2(\xi) + \sigma^2(\varepsilon_1)$	$\sigma^2(\xi)$	$\sigma^2(\xi)$
x2	-	$\sigma^2(\xi) + \sigma^2(\varepsilon_2)$	$\sigma^2(\xi)$
x3	-	-	$\sigma^2(\xi) + \sigma^2(\varepsilon_3)$

Folge der Modellannahmen:
Alle 3 Kovarianzen sind gleich!

Identifizierbarkeit

Identifikation

Um ein Modell zu identifizieren, sind grundsätzlich zwei Dinge notwendig

1. Es muss **mehr bekannte** als **unbekannte** (zu schätzende) Parameter geben
2. Jede **latente Variable** und jede **Fehlervariable** muss eine Metrik aufweisen
→ z.B. **Parameterfixierung eines Indikators auf Eins (ULI)**

Anmerkung: Dies ist eine vereinfachte Darstellung! Bei komplexen Modellen kann die Identifizierbarkeit von zusätzlichen Eigenschaften abhängen. Eine gute Beschreibung ist in Kline (2010) gegeben.

Die Empirische Kovarianzmatrix S stellt die **bekannten** Parameter:

Im Beispiel:

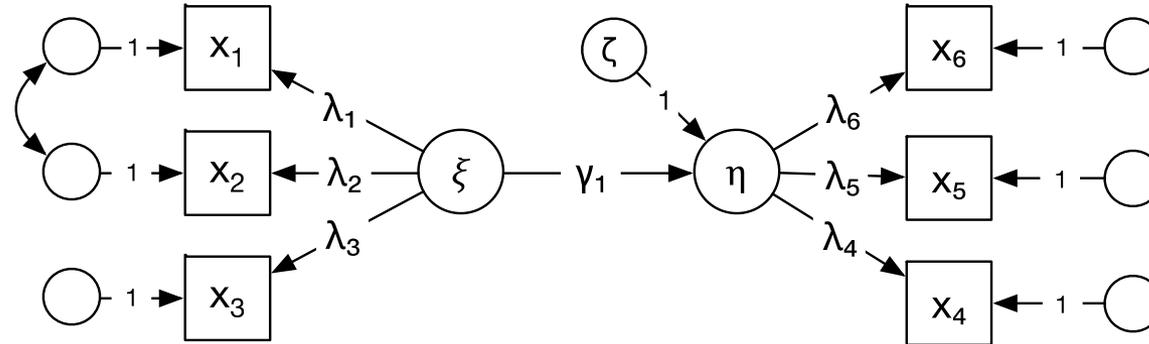
	x1	x2	x3
x1	1.36	0.41	0.58
x2	0.41	1.39	0.45
x3	0.58	0.45	1.28

3 Varianzen
+ 3 Kovarianzen
= 6 Bekannte Parameter

Allgemein:

$$p = \frac{k \cdot (k + 1)}{2} = \frac{3 \cdot (3 + 1)}{2} = 6$$

- p = Anzahl der bekannten Parameter
- k = Anzahl der Items bzw. manifesten Variablen



- Wie viele **freie** (und damit **unbekannte**) **Parameter** gibt es in einem SEM? Geschätzt werden:
 - Die freien (d.h., nicht-fixierten) gerichteten Pfade (im Bsp: **7**)
 - Die Kovarianzen (im Bsp: **1**)
 - Die nicht-fixierten Varianzen der exogenen Variablen; d.h. aller ξ , der Fehlervariablen von manifesten (ε) und latenten (ζ) Variablen (im Bsp. **8**)
- Diese freien Parameter werden dann die **Modellparameter** genannt; die Gesamtheit aller Modellparameter eines Modells wird als **Parametervektor θ** (Theta) bezeichnet: $\theta = [\sigma^2(\varepsilon_1), \sigma^2(\varepsilon_2), \sigma^2(\varepsilon_3), \dots, \sigma^2(\xi), \sigma^2(\zeta), \dots, \lambda_1, \lambda_2, \dots]$
- Equality constraints werden von der Zahl freier Parameter abgezogen
 - z.B. $\lambda_1 = \lambda_2 \rightarrow$ es wird effektiv nur ein Parameter geschätzt

- Bei SEMs berechnet man die Zahl der Freiheitsgrade (df) über die Differenz der bekannten und unbekanntem Parameter:

$$df = \text{Anzahl}_{\text{bekannte P.}} - \text{Anzahl}_{\text{unbekannte P.}}$$

- $\text{Anzahl}_{\text{bekannte P.}}$ ist die Anzahl der Varianzen und Kovarianzen der Kovarianzmatrix S .
- $\text{Anzahl}_{\text{unbekannte P.}}$ ist die Anzahl der zu schätzenden Parameter θ

- (1) Bekannte Parameter = Freiheitsgrade der Kovarianzmatrix der manifesten Variablen: Wie viele Elemente der Kovarianzmatrix können empirisch frei variieren?
 - (2) Zu schätzende Parameter = Anzahl der freien Parameter in einem Modell: Wie viele Modellparameter können frei variieren?
 - (3) Freiheitsgrade für SEM Modelltests = Wie viele Elemente der empirischen Kovarianzmatrix können noch frei variieren, nachdem man die freien Modellparameter aus den Daten geschätzt hat? (d.h., Anzahl bekannte Parameter minus Anzahl zu schätzende Parameter)
- **Im Zweifelsfall meint man mit „Freiheitsgraden“ den Fall 3**

Beispiel eines Identifikationsproblems

$$\sigma(X_1, X_2) = \lambda_1 \lambda_2 \sigma^2(\xi)$$

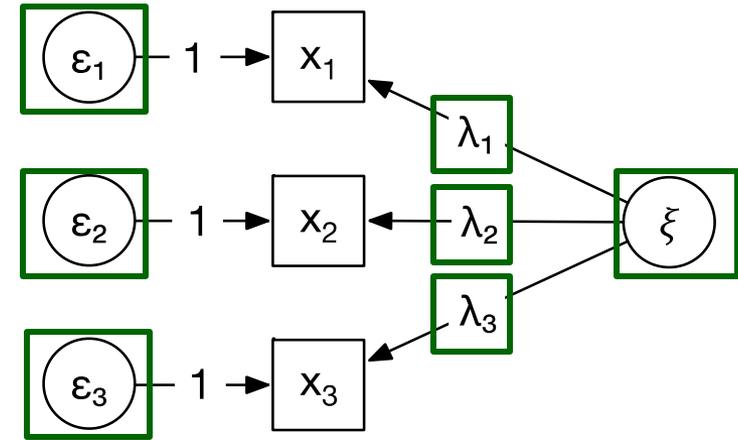
$$\sigma(X_1, X_3) = \lambda_1 \lambda_3 \sigma^2(\xi)$$

$$\sigma(X_2, X_3) = \lambda_2 \lambda_3 \sigma^2(\xi)$$

$$\sigma^2(X_1) = \lambda_1^2 \sigma^2(\xi) + \sigma^2(\varepsilon_1)$$

$$\sigma^2(X_2) = \lambda_2^2 \sigma^2(\xi) + \sigma^2(\varepsilon_2)$$

$$\sigma^2(X_3) = \lambda_3^2 \sigma^2(\xi) + \sigma^2(\varepsilon_3)$$



6 bekannte Parameter



7 unbekannte Parameter

- Auflösen der Gleichungen nach den Unbekannten (d.h., nach den Modellparametern):

$$\sigma^2(\xi) = \frac{\sigma(X_1, X_2)\sigma(X_1, X_3)}{\lambda_1^2 \sigma(X_2, X_3)}$$

$$\lambda_2 = \lambda_1 \frac{\sigma(X_2, X_3)}{\sigma(X_1, X_3)}$$

$$\lambda_3 = \lambda_1 \frac{\sigma(X_2, X_3)}{\sigma(X_1, X_2)}$$

$$\sigma^2(\epsilon_1) = \sigma^2(X_1) - \frac{\sigma(X_1, X_3)\sigma(X_1, X_2)}{\sigma(X_2, X_3)}$$

$$\sigma^2(\epsilon_2) = \sigma^2(X_2) - \frac{\sigma(X_2, X_3)\sigma(X_1, X_2)}{\sigma(X_1, X_3)}$$

$$\sigma^2(\epsilon_3) = \sigma^2(X_3) - \frac{\sigma(X_1, X_3)\sigma(X_2, X_3)}{\sigma(X_1, X_2)}$$

- Alle (Ko)Varianzen sind bekannt: Dafür kann man konkrete empirische Schätzwerte einsetzen.
- Das Gleichungssystem ist nicht lösbar, weil auf der rechten Seite immer ein unbekannter Parameter (im Bsp.: λ_1) steht.

$$\sigma(X_1, X_2) = 1 \lambda_2 \sigma^2(\xi)$$

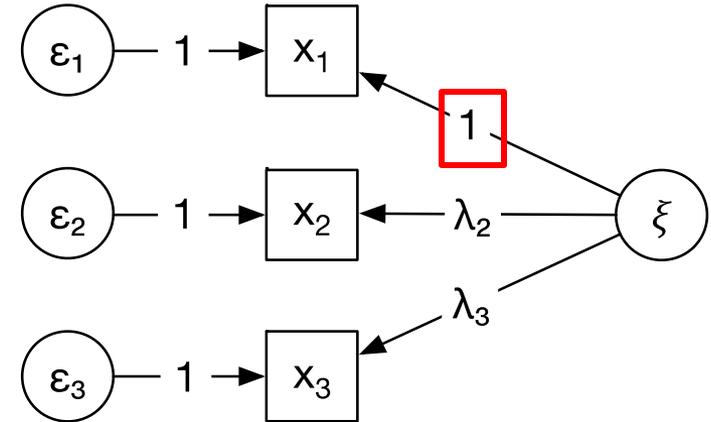
$$\sigma(X_1, X_3) = 1 \lambda_3 \sigma^2(\xi)$$

$$\sigma(X_2, X_3) = \lambda_2 \lambda_3 \sigma^2(\xi)$$

$$\sigma^2(X_1) = 1 \sigma^2(\xi) + \sigma^2(\varepsilon_1)$$

$$\sigma^2(X_2) = \lambda_2^2 \sigma^2(\xi) + \sigma^2(\varepsilon_2)$$

$$\sigma^2(X_3) = \lambda_3^2 \sigma^2(\xi) + \sigma^2(\varepsilon_3)$$



➤ Gleichungssystem ist nun lösbar!

➤ Es gibt genau eine Lösung:

➤ 6 bekannte \leftrightarrow 6 unbekannte (zu schätzende) Parameter

➤ $df = 0$

- Identifikationsgleichungen:

$$\sigma^2(\xi) = \frac{\sigma(X_1, X_2)\sigma(X_1, X_3)}{1\sigma(X_2, X_3)}$$
$$\lambda_2 = 1 \frac{\sigma(X_2, X_3)}{\sigma(X_1, X_3)}$$
$$\lambda_3 = 1 \frac{\sigma(X_2, X_3)}{\sigma(X_1, X_2)}$$

$$\sigma^2(\epsilon_1) = \sigma^2(X_1) - \frac{\sigma(X_1, X_3)\sigma(X_1, X_2)}{\sigma(X_2, X_3)}$$
$$\sigma^2(\epsilon_2) = \sigma^2(X_2) - \frac{\sigma(X_2, X_3)\sigma(X_1, X_2)}{\sigma(X_1, X_3)}$$
$$\sigma^2(\epsilon_3) = \sigma^2(X_3) - \frac{\sigma(X_1, X_3)\sigma(X_2, X_3)}{\sigma(X_1, X_2)}$$

Jetzt stehen ausschließlich bekannte Größen auf der rechten Seite – d.h., die unbekannt Modellparameter können einfach berechnet werden, indem man die (Ko)Varianzen aus der empirischen Kovarianzmatrix einsetzt.

Die Identifikationsgleichungen erlauben eine **Diagnostik der Identifizierbarkeit**, wobei man drei Arten unterscheidet:

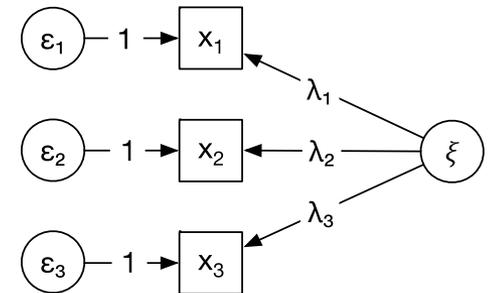
1. **Modell ist unteridentifiziert**: Für wenigstens einen Modellparameter lässt sich keine Identifikationsgleichung angeben ($df < 0$)
2. **Modell ist gerade/genau identifiziert**: Für jeden Modellparameter lässt sich genau eine Identifikationsgleichung angeben ($df = 0$)
3. **Modell ist überidentifiziert**: Für wenigstens einen Modellparameter lassen sich mehr als eine Identifikationsgleichung angeben ($df > 0$)

Unteridentifiziert (*underidentified*): Für mindestens einen Parameter lässt sich keine Identifikationsgleichung angeben

- Mehr zu schätzende Parameter als bekannte Parameter
→ viele mögliche Lösungen für die Parameterschätzungen; Parameter können nicht festgelegt werden

Beispiel: $a + b = 6$

- „a“ und „b“ zu schätzende Parameter
- „6“ ist die beobachtete Variable
- Das Gleichungssystem ist nicht eindeutig lösbar
- Viele Lösungen können den Wert 6 erzeugt haben
 - $2 + 4 = 6$; $3 + 3 = 6$, usw.



Um das Modell eindeutig zu identifizieren und eine Parameterschätzung vorzunehmen, muss eine weitere beobachtete Variable bekannt sein, oder ein freier Parameter des Modells fixiert werden.

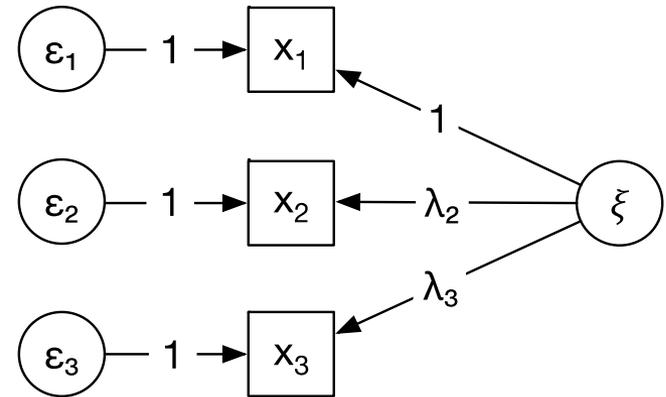
```
Warning message:  
In lav_model_vcov(lavmodel = lavmodel, lavsamplestats = lavsamplestats,  
lavaan WARNING: could not compute standard errors!  
lavaan NOTE: this may be a symptom that the model is not identified.
```

Gerade identifiziert (*just identified / saturiertes Modell*):

Für jeden Parameter lässt sich genau eine Identifikationsgleichung angeben

→ So viele zu schätzende wie bekannte Parameter

- Es gibt eine einzige Lösung
- Modell hat daher null Freiheitsgrade
→ daher kein Modelltest möglich!



Beispiel: $a + b = 6$ und $2a + b = 10$

- Für beide Unbekannte gibt es eine eindeutige Lösung: $a = 4$ und $b = 2$
- Modell ist gerade identifiziert

Überidentifiziert (*overidentified*): Für wenigstens einen Parameter lassen sich mehr als eine Identifikationsgleichung angeben

→ Mehr bekannte als zu schätzende Parameter

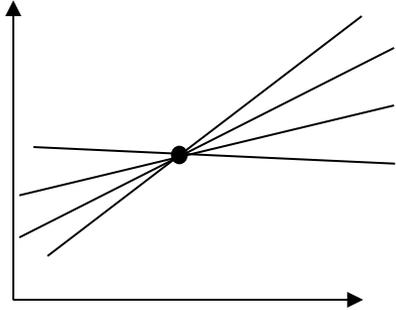
- Resultierendes Gleichungssystem näherungsweise bestimmbar
- Stellt den Regelfall im Rahmen von Strukturgleichungsmodellen dar
- Ein Modelltest ist möglich

Beispiel: drei Gleichungen

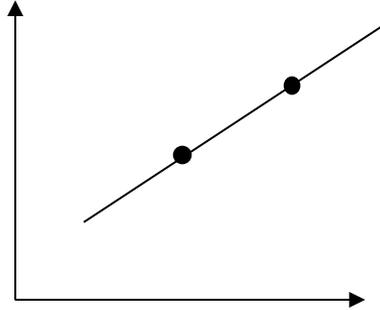
		Exemplarische Werte für die Unbekannten	Abw. 1	Abw. 2	Abw. 3	$\Sigma(\text{Abw.}^2)$
Gleichung 1	$a + b = 6$	$a = 3; b = 3$	0	1	0	1
Gleichung 2	$2 a + b = 10$	$a = 3; b = 4$	1	0	1	2
Gleichung 3	$3 a + b = 12$	$a = 2; b = 6$	2	0	0	4
		$a = 3; b = 3.3$	0.3	0.7	0.3	0.67

- Widersprüchliche Werte! → welche Werte für a und b sind nun die „Richtigen“?
- Diejenigen, die am ehesten passen (z.B. die geringste Abweichung haben):
 $a = 3$ und $b = 3.3$

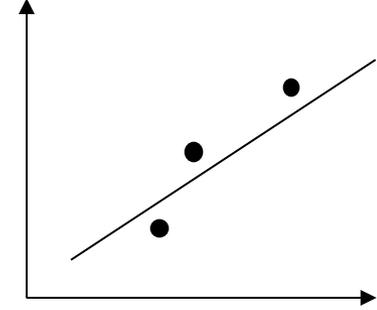
Modell mit 2 freien Parametern: $y = b_0 + b_1 \cdot x + e$



1 bekannt, 2 zu schätzen
→ $df = -1$
→ viele Lösungen, nicht identifiziert

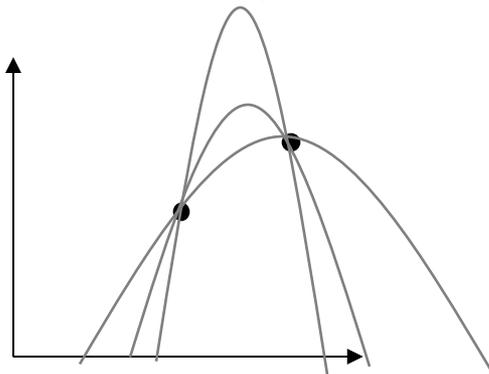


2 bekannt, 2 zu schätzen
→ $df = 0$
→ genau eine Lösung, gerade identifiziert

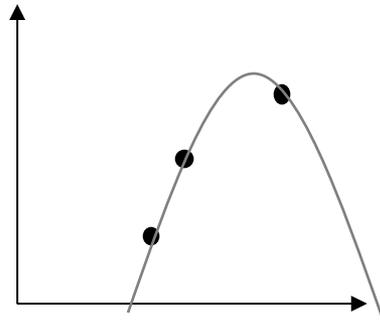


3 bekannt, 2 zu schätzen
→ $df = 1$
→ eine optimale Lösung, überidentifiziert

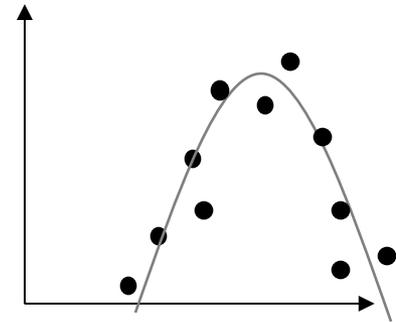
Polynomiales Modell mit 3 freien Parametern: $y = b_0 + b_1 \cdot x + b_2 \cdot x^2 + e$



2 bekannt, 3 zu schätzen
→ $df = -1$
→ viele Lösungen, nicht identifiziert



3 bekannt, 3 zu schätzen
→ $df = 0$
→ genau eine Lösung, gerade identifiziert



11 bekannt, 3 zu schätzen
→ $df = 8$
→ eine optimale Lösung, überidentifiziert

Zusammenfassung: Arten von Gleichungen

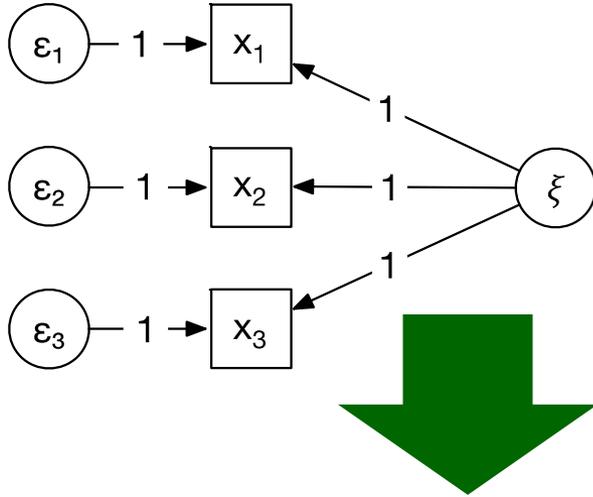
- Alle Begriffe beziehen sich auf das gleiche Gleichungssystem, das nach verschiedenen Variablen aufgelöst und umsortiert werden kann:
- **Definitionsgleichungen** (links stehen alle endogenen Variablen):
 - Jede endogene Variable wird durch eine Regressionsgleichung dargestellt (d.h., sie wird durch alle Variablen vorhergesagt, die einen gerichteten Pfeil auf sie haben). D.h., Gleichungen so aufstellen, dass links die endogenen Variable stehen (z.B. x_1 , oder η)
- **Strukturgleichungen** (links stehen die Elemente der Kovarianzmatrix):
 - Beschreiben die *Struktur* der modell-implizierten Kovarianzmatrix. Es geht nicht mehr um die manifesten Variablen an sich, sondern um die (Ko)Varianzen der Variablen → die Definitionsgleichungen so umstellen und als (Ko)Varianz ausdrücken, dass links die Elemente der modell-implizierten Kovarianzmatrix stehen, also z.B. $\sigma^2(x_1)$ oder $\sigma(x_1, x_2)$
- **Identifikationsgleichungen** (links stehen die Modellparameter):
 - So umstellen, dass links die *Modellparameter* stehen, also z.B. $\sigma^2(\xi)$, $\sigma^2(\varepsilon_1)$ oder λ_1 . Dabei muss dafür gesorgt werden, dass rechts nur bekannte Variablen (d.h., Varianzen und Kovarianzen der manifesten Variablen stehen). Hat man mehr unbekannte Modellparameter als bekannte Variablen, ist das nicht möglich, und das Modell ist unteridentifiziert.

Die Schätzung der Parameter mit Maximum Likelihood

- Strukturgleichungen implizieren eine bestimmtes **Größenverhältnis der Varianzen/Kovarianzen der manifesten Variablen**.
- Dieses **Größenverhältnis** wird als Kovarianz**struktur** bezeichnet.
- Die Kovarianzstruktur der Daten wird (bei überidentifizierten Modellen) stets von der modellimplizierten Struktur abweichen*.
 - z.B. selbst wenn in der Population die drei Kovarianzen exakt gleich sind (wie das Modell vorhersagt), wird in jeder konkreten Stichprobe eine gewisse Abweichung davon festzustellen sein.
 - D.h., man muss prüfen, ob diese Abweichung als zufällig oder systematisch angesehen werden sollte.

* Außer es ist ein riesiger Zufall und die realisierte Stichprobe bildet auf die 10. Nachkommastelle die wahre Kovarianzstruktur ab. Bei gerade identifizierten Modellen wird hingegen die empirische Kovarianzmatrix immer exakt von der modellimplizierten Struktur abgebildet.

Modell-implizierten Kovarianzmatrix vs. empirische Kovarianzmatrix



Das Modell
sagt vorher
(Hypothese):
Alle 3
Kovarianzen
sind gleich!

Empirisch: Die
3 Kovarianzen
sind ähnlich,
aber nicht
gleich!

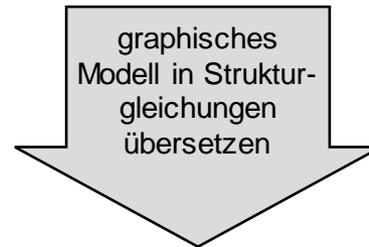
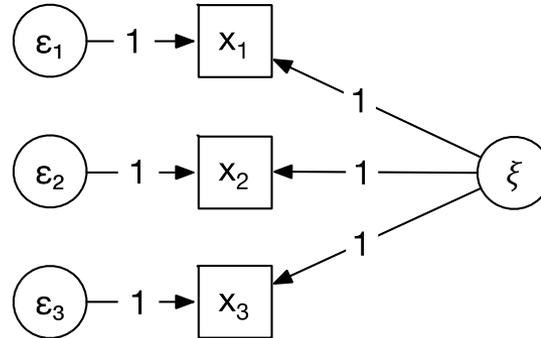
	x1	x2	x3
x1	$\sigma^2(\xi) + \sigma^2(\varepsilon_1)$	$\sigma^2(\xi)$	$\sigma^2(\xi)$
x2	-	$\sigma^2(\xi) + \sigma^2(\varepsilon_2)$	$\sigma^2(\xi)$
x3	-	-	$\sigma^2(\xi) + \sigma^2(\varepsilon_3)$

modell-implizierte Struktur der Kovarianzmatrix

	x1	x2	x3
x1	1.36	0.41	0.58
x2	-	1.39	0.45
x3	-	-	1.28

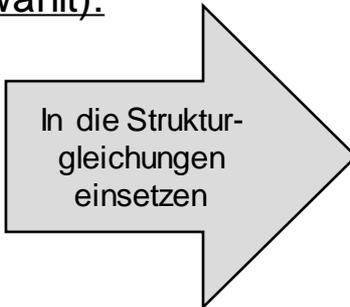
empirische Stichproben-Kovarianzmatrix

- Wie kommt man von der vom Modell implizierten Kovarianz**struktur** zur implizierten Kovarianz**matrix**, die Aussagen über die absolute Größe der Varianzen/Kovarianzen der manifesten Variablen macht?
- Werden die Modellparameter zunächst irgendwie numerisch festgelegt und dann in die Strukturgleichungen eingesetzt, so resultiert eine konkrete Realisierung der Kovarianzmatrix.
- Setzt man andere Startwerte für die Modellparameter ein, resultiert daraus eine andere Kovarianzmatrix
- Aber: Da jedes Set an Modellparametern die selben Strukturgleichungen „durchläuft“, folgt jede realisierte Matrix zwangsläufig der allgemeinen Kovarianzstruktur dieses Modells.



Set 1 von Modellparametern
(beliebig gewählt):

$$\begin{aligned}\sigma^2(\xi) &= 0.5 \\ \sigma^2(\varepsilon_1) &= 0.3 \\ \sigma^2(\varepsilon_2) &= 0.4 \\ \sigma^2(\varepsilon_3) &= 0.6\end{aligned}$$



$$\begin{aligned}\sigma(X_1, X_2) &= \sigma^2(\xi) \\ \sigma(X_1, X_3) &= \sigma^2(\xi) \\ \sigma(X_2, X_3) &= \sigma^2(\xi) \\ \sigma^2(X_1) &= \sigma^2(\xi) + \sigma^2(\varepsilon_1) \\ \sigma^2(X_2) &= \sigma^2(\xi) + \sigma^2(\varepsilon_2) \\ \sigma^2(X_3) &= \sigma^2(\xi) + \sigma^2(\varepsilon_3)\end{aligned}$$



	x1	x2	x3
x1	0.8	0.5	0.5
x2	-	0.9	0.5
x3	-	-	1.1

Jedes beliebige Set an
Modellparametern folgt
der allgemeinen
Modellstruktur

Set 1 von Modellparametern:

$$\begin{aligned}\sigma^2(\xi) &= 0.5 \\ \sigma^2(\varepsilon_1) &= 0.3 \\ \sigma^2(\varepsilon_2) &= 0.4 \\ \sigma^2(\varepsilon_3) &= 0.6\end{aligned}$$

In die Struktur-
gleichungen
einsetzen

$$\begin{aligned}\sigma(X_1, X_2) &= \sigma^2(\xi) \\ \sigma(X_1, X_3) &= \sigma^2(\xi) \\ \sigma(X_2, X_3) &= \sigma^2(\xi) \\ \sigma^2(X_1) &= \sigma^2(\xi) + \sigma^2(\varepsilon_1) \\ \sigma^2(X_2) &= \sigma^2(\xi) + \sigma^2(\varepsilon_2) \\ \sigma^2(X_3) &= \sigma^2(\xi) + \sigma^2(\varepsilon_3)\end{aligned}$$

Resultierende
Kovarianzmatrix

	x1	x2	x3
x1	0.8	0.5	0.5
x2	-	0.9	0.5
x3	-	-	1.1

Set 2 von Modellparametern:

$$\begin{aligned}\sigma^2(\xi) &= 0.48 \\ \sigma^2(\varepsilon_1) &= 0.8 \\ \sigma^2(\varepsilon_2) &= 0.9 \\ \sigma^2(\varepsilon_3) &= 0.8\end{aligned}$$

In die Struktur-
gleichungen
einsetzen

$$\begin{aligned}\sigma(X_1, X_2) &= \sigma^2(\xi) \\ \sigma(X_1, X_3) &= \sigma^2(\xi) \\ \sigma(X_2, X_3) &= \sigma^2(\xi) \\ \sigma^2(X_1) &= \sigma^2(\xi) + \sigma^2(\varepsilon_1) \\ \sigma^2(X_2) &= \sigma^2(\xi) + \sigma^2(\varepsilon_2) \\ \sigma^2(X_3) &= \sigma^2(\xi) + \sigma^2(\varepsilon_3)\end{aligned}$$

Resultierende
Kovarianzmatrix

	x1	x2	x3
x1	1.3	0.48	0.48
x2	-	1.4	0.48
x3	-	-	1.3

Set 3 von Modellparametern:

...

- Welches Set an Parametern kommt so nahe wie möglich an die empirische Kovarianzmatrix heran? Wann hat man den besten *Fit*?
- Man benötigt ein Diskrepanzmaß, das angibt, „wie weit weg“ zwei Matrizen voneinander sind
- Bsp: Ordinary Least Squares = Summe der quadrierten Abweichungen
- Je kleiner die Diskrepanz ist, desto näher ist die vorhergesagte an der empirischen Matrix dran.
- Diskrepanz von 0 = perfekte Übereinstimmung von vorhergesagter mit empirischer Kovarianzmatrix

Implizierte Kovarianzmatrix von Modellparameter-Set 2:			
$\hat{\Sigma}(\theta_2)$	x1	x2	x3
x1	1.3	0.48	0.48
x2	-	1.4	0.48
x3	-	-	1.3

=

Empirische Kovarianzmatrix			
S	x1	x2	x3
x1	1.36	0.41	0.58
x2	-	1.39	0.45
x3	-	-	1.28

=

Differenz: $\hat{\Sigma}(\theta_2) - S$			
	x1	x2	x3
x1	-0.06	0.07	-0.10
x2	-	0.01	0.03
x3	-	-	0.02

Diskrepanzfunktionen

- Mehrere Möglichkeiten, die Diskrepanz zwischen zwei Matrizen zu berechnen
- Jedes Diskrepanzmaß ist eine Funktion von S und $\hat{\Sigma}(\Theta)$:

$$F[S, \hat{\Sigma}(\Theta)]$$

- Verschiedene Verrechnungsmethoden:

ML (Maximum Likelihood):

$$F_{ML} = tr(S \cdot S^{-1}) - p + \ln|S| - \ln|\hat{S}|$$

GLS (Generalized Least Squares):

$$F_{GLS} = .5 \cdot tr[(S - \hat{S}) \cdot S^{-1}]^2$$

ULS (Unweighted Least Squares; = OLS):

$$F_{ULS} = .5 \cdot tr[(S - \hat{S})]^2$$

WLSMV

(Weighted Least Squares Mean
and Variance Adjusted)

$$F_{WLSMV} = (S - \hat{S})' W^{-1} (S - \hat{S})$$

↑
Formeln = Exkurs

1. **Startwerte** werden für die Modellparameter θ eingesetzt
Es gibt Heuristiken, wie die Startwerte sinnvoll gewählt werden können
(so, dass sie vermutlich möglichst nahe an den optimalen Werten
starten)
 2. Diese Modellparameter werden in die **Strukturgleichungen** eingesetzt
und die implizierte Kovarianzmatrix berechnet.
 3. Die implizierte Kovarianzmatrix $\hat{\Sigma}(\theta)$ wird mit der empirischen (aus der
Stichprobe geschätzten) Kovarianzmatrix S verglichen: **Diskrepanz F**
zwischen beobachteter und implizierter Kovarianzmatrix wird ermittelt.
 4. Modellparameter werden etwas verändert \rightarrow wird die Diskrepanz
kleiner?
- Schritte 3 und 4 werden so lange wiederholt (\rightarrow **Iteration**; lat. iterare =
wiederholen) bis sich zwei aufeinanderfolgende Parameterschätzungen
kaum mehr, d.h. nach einem vorab bestimmten Kriterium, unterscheiden
(\rightarrow **Konvergenz**; lat. convergere = sich zuneigen).

Objective function =
Diskrepanzfunktion.
Kleiner Werte = weniger
Diskrepanz

```
> fit1 <- cfa(model = modell1, data =  
HolzingerSwineford1939[, c("x1", "x2", "x3")],  
verbose=TRUE, debug=FALSE)
```

Quasi-Newton steps using NLMINB:

```
Objective function = 0.4310868182753260  
Objective function = 0.0685167034222560  
Objective function = 0.0578306703906584  
Objective function = 0.0256211601368628  
Objective function = 0.0185081734725918  
Objective function = 0.0136902147986711  
Objective function = 0.0097227715131203  
Objective function = 0.0074841917429793  
Objective function = 0.0071239182926741  
Objective function = 0.0069038978752487  
Objective function = 0.0069017322951990
```

Ab hier sind die
geschätzten Parameter
fast am Optimum: weiteres
„Tuning“ der Parameter
bringt keine Verbesserung
in der Diskrepanz mehr.

```
Objective function = 0.0069017167113161  
Objective function = 0.0069017164105791  
Objective function = 0.0069017163208531  
Objective function = 0.0069017163074843  
Objective function = 0.0069017163073140  
Objective function = 0.0069017163073140  
convergence status (0=ok): 0  
nlminb message says: relative convergence (4)  
number of iterations: 15
```

Allgemeiner Algorithmus der Parameterschätzung

Startwerte für die Modellparameter $\sigma^2(\varepsilon_1)$, $\sigma^2(\varepsilon_2)$, $\sigma^2(\varepsilon_3)$, und $\text{Var}(\xi)$. Werden von lavaan automatisch gewählt, können aber auch von Hand gesetzt werden.

Die erste implizierte Kovarianzmatrix, die sich aus den Startwerten ergibt

Die Optimierung der Modellparameter über die 15 Iterationen hinweg.

Die finale (optimale) Schätzung: Näher kommt man mit dieser vorgegebenen Modellstruktur nicht an die empirische Kovarianzmatrix heran.

```
start.x = 0.6791849 0.6908919 0.6374324 0.05
```

```
      [,1]      [,2]      [,3]  
[1,] 0.7291849 0.0500000 0.0500000  
[2,] 0.0500000 0.7408919 0.0500000  
[3,] 0.0500000 0.0500000 0.6874324
```

```
Current free parameter values =
```

```
[1] 0.6791849 0.6908919 0.6374324 0.0500000  
[1] 0.8244866 0.8396709 0.7872044 1.0166026  
[1] 0.8349243 0.9019111 0.7650264 0.9667664  
[1] 0.8560510 1.1107821 0.6669687 0.7266743  
[1] 0.8612694 1.2323573 0.6285142 0.5062924  
[1] 0.8519810 1.1610864 0.6834877 0.5563524  
[1] 0.8402434 1.0909742 0.7394249 0.5334313  
[1] 0.8317424 0.9978216 0.7810783 0.4930666  
[1] 0.8370244 0.9873017 0.7661032 0.4901942  
[1] 0.8464091 0.9769520 0.7363039 0.4907334  
[1] 0.8455828 0.9765197 0.7393528 0.4904694  
[1] 0.8456546 0.9767305 0.7391665 0.4904783  
[1] 0.8456707 0.9767024 0.7391652 0.4904775  
[1] 0.8456849 0.9766974 0.7391703 0.4904755  
[1] 0.8456911 0.9767001 0.7391729 0.4904745  
[1] 0.8456913 0.9767009 0.7391730 0.4904745  
[1] 0.8456913 0.9767009 0.7391730 0.4904745
```

lavaan (0.5-20) converged normally after 15 iterations

Konvergenz der
Parameterschätzung nach 15
Iterationen

Number of observations 301

ML Schätzfunktion als default

Estimator ML
Minimum Function Test Statistic 4.155
Degrees of freedom 2
P-value (Chi-square) 0.125

df = 2: Es gibt zwei bekannte
Parameter mehr als
unbekannte Parameter

Parameter Estimates

Information Standard Errors Expected Standard

Latent Variables:

	Estimate	Std.Err	Z-value	P(> z)
visual =~				
x1	1.000			
x2	1.000			
x3	1.000			

Hier findet man die finalen
Parameterschätzungen von der
letzten Iteration

Variances:

	Estimate	Std.Err	Z-value	P(> z)
x1	0.846	0.089	9.538	0.000
x2	0.977	0.098	9.922	0.000
x3	0.739	0.081	9.121	0.000
visual	0.490	0.065	7.545	0.000

Die Wahl der geeigneten Schätzmethode hängt unter anderem ab von folgenden Faktoren:

1. Skalenniveau:
ordinal- oder intervallskaliert
2. Art der Verteilung der endogenen Variablen:
normalverteilt oder schief verteilt
3. Stichprobengröße

Maximum-Likelihood-Methode: Beantwortet die Frage, „*Welche Parameter in der Population sind am plausibelsten bei dem beobachteten Ereignis?*“

→ optimale Schätzung der Populationsparameter

GLS- und ULS-Methode: minimieren die quadrierten Abweichungen zwischen beobachteter und implizierter Kovarianzmatrix

ADF-Methode: Basiert auf einer speziellen Kovarianzmatrix und einer GLS- (Generalized Least Squares) Schätzung

WLSMV-Methode: Basiert auf einer gewichteten Schätzung der kleinsten quadratischen Abweichung, welche für arithmetisches Mittel und Varianz adjustiert ist

ML- und GLS Methode:

- multivariate Normalverteilung der endogenen Variablen (die exogenen können auch dichotom, schief, etc. verteilt sein)
- Intervalldatenniveau

Falls keine multivariate Normalverteilung der endogenen Variablen:

- Schätzungen des χ^2 -Wertes fallen überhöht aus (siehe nächste Sitzung).
 - Führt dazu, dass ein passendes Modell durch den Modelltest zu oft abgelehnt wird.
- Moderate bis schwerwiegende Unterschätzungen der Standardfehler der Parameter
 - Parameter: Fehlervarianzen, Korrelation, Ladungen
 - Der Standardfehler wird zur Ermittlung der Signifikanz der Parameter verwendet

ADF-Methode benötigt große Stichproben ($N > 3000$ empfohlen)
und ist nur bei wenig komplexen Modellen gut

- ADF-Methode hat allerdings den Vorteil, dass sie keine Verteilungsannahmen macht.

WLSMV-Methode ermöglicht eine robuste Schätzung von Modellen mit
dichotomen Variablen

- Muthén und Muthén empfehlen eine Stichprobengröße von $N > 200$

- Die **ML-Methode** ist robust gegenüber Verletzungen der Normalverteilungsannahme, was die Parameterschätzungen anbetrifft, nicht was die Signifikanztests (Modell und Parameter) anbetrifft.
 - Im Falle einer Verletzung der multivariaten Normalverteilung wendet man die **Bollen-Stine Bootstrap-Methode** an, die einen korrigierten p -Wert für das Gesamtmodell ergibt.
 - Diverse Korrekturen für robuste Standardfehler sind möglich.

- In der Praxis wird oft viel über die „richtige“ Schätzmethode diskutiert (z.B. ML vs. WLSMV, Bootstrap vs. Korrektur der Standardfehler, ...)
- Nach unserer Einschätzung hat eine „nicht optimale“ Wahl der Schätzmethode in der Praxis viel geringere Auswirkungen, als die folgenden Punkte:
 - Modellfehlspezifikation, z.B.
 - Falsche Anzahl von latenten Variablen (Stichwort Dimensionalität)
 - Falsche Struktur zwischen latenten und manifesten Variablen (Stichwort Einfachstruktur, Nebenladungen)
 - Wichtige Variablen, die bei der Schätzung der interessierenden kausalen Effekte berücksichtigt werden müssten, wurden nicht erhoben bzw. nicht ins Modell aufgenommen (Stichwort Confounder, siehe Vorlesung zur kausalen Inferenz).
 - Schlechte Messinstrumente (Stichwort Inhaltsvalidität)
 - Kleine Stichproben

- Grundlage der Parameterschätzung in SEMs ist ein Vergleich der empirischen Kovarianzmatrix S mit der modell-implizierten Kovarianzmatrix $\hat{\Sigma}$.
- Wie kommt man vom graphischen Modell zu $\hat{\Sigma}$?
- **Definitionsgleichungen** aufstellen → **Strukturgleichungen** aufstellen
→ nach **Modellparametern** auflösen
- **Identifikation** des Modells prüfen – lassen sich die Parameter überhaupt schätzen?
 - Unter-, gerade, oder überidentifizierte Modelle
- Die optimalen Werte für die **Modellparameter** finden, so dass $\hat{\Sigma}$ so nahe wie möglich an S herankommt
- Wie findet man die optimalen Werte? **Diskrepanzfunktion** minimieren (z.B. Maximum Likelihood, oder WLSMV)