

2.7 Welche Verfahren sollten zur SE-Modellschätzung benutzt werden?

Die Berechnung der freien und auch der restringiert-freien³² Modellparameter eines SE-Modells erfolgt in der SEM-Analyse mittels aufwändiger statistischer Schätzverfahren. In diesen Verfahren werden diejenigen Werte als bestmögliche Schätzwerte für die freien Modellparameter ermittelt, mit denen die empirisch beobachteten Datenstrukturen eines beliebigen Datensatzes im Kontext des analysierten SE-Modells möglichst exakt reproduziert werden können.

Dies erfolgt in Form eines Datenabgleichs. Bei diesem Datenabgleich werden diejenigen Parameterwerte, die als Kandidaten für die letztendlich auszuwählenden, optimalen Schätzwerte anzusehen sind, dazu benutzt, eine Kovarianzmatrix (Σ) aller im jeweiligen Modell berücksichtigten manifesten Variablen zu berechnen. Wenn die Werte dieser, mittels Verwendung von Parameter-Schätzwerten berechneten Kovarianzmatrix (Σ) möglichst dicht an den empirisch beobachteten Werten der Kovarianzmatrix (S) der entsprechenden Variablen im gegebenen Datensatz liegen, dann können die Parameter-Schätzwerte als "gute" Schätzwerte gelten. Und die besten Schätzwerte sind diejenigen Werte, bei denen die Differenz zwischen Σ und S am geringsten ist.

Abbildung 2.10 verdeutlicht dieses Schätzprinzip an einem didaktisch vereinfachten, fiktiven Modellbeispiel:

In einem Datensatz lasse sich die in Abbildung 2.10a veranschaulichte Korrelation (=standardisierte Kovarianz) zwischen den beiden Variablen Y_1 und Y_2 von $r=0.70$ empirisch ermitteln. Mit dieser Korrelation sei das in Abbildung 2.10b gezeigte Modell zu schätzen. Nach diesem Modell beeinflusst die exogene Variable F die beiden endogenen Y -Variablen durch zwei direkte Pfadverbindungen und die Koeffizienten der beiden Pfade von F auf Y_1 und von F auf Y_2 seien statistisch zu schätzen (was durch Sternchen in Abb. 2.10b gekennzeichnet ist.).

In einem ersten Verfahrensschritt werden für die beiden Pfade die Startwerte "0.80" und "0.50" angenommen (vgl. Abb. 2.10c). Mit diesen beiden Werten lässt sich die Korrelation zwischen Y_1 und Y_2 berechnen. Denn nach einer Rechenregel der Pfadanalyse kann in einem rekursiven Pfadmodell (d.h. in einem Pfadmodell ohne Feedback-Schleifen) die Korrelation zwischen zwei Variablen ermittelt werden, indem das Produkt aus den Koeffizienten derjenigen Pfade gebildet wird, auf

32 Unter "restringiert-freien" Parametern werden solche Modellparameter verstanden, die zwar zu schätzen sind und in diesem Sinne "frei" sind, für die jedoch bestimmte Restriktionen festgelegt werden (auch "constraints" genannt). Eine typische Anwendung erfolgt in der Multigruppenanalyse (a.a.O.). Dort wird üblicherweise vorgegeben, dass die unstandardisierten Faktorladungen von jedem Faktor in allen Gruppen mit identischen Werten geschätzt werden.

denen entlang zu wandern ist, um von der einen Variablen zur anderen Variablen zu gelangen (wobei evtl. auch gegen die Pfeilrichtung "anzuschwimmen" ist). Im Beispiel beträgt dieses Produkt: $0.80 \times 0.50 = 0.40$ (vgl. Abb. 2.10c). Mit diesem Korrelationswert von 0.40 ist die Schätzung der beiden freien Parameter nicht besonders gut gelungen, denn die Korrelation von 0.40 wäre relativ weit von der empirisch beobachteten Korrelation von 0.70 entfernt.

Besser ist die in Abbildung 2.10d dargestellte Schätzung gelungen. Die beiden geschätzten Pfadkoeffizienten von 0.80 und 0.75 ergeben eine Korrelation von 0.60 und diese liegt wesentlich näher an der beobachteten Korrelation von 0.70. Deshalb ist die zweite Schätzung (Abb. 2.10d) deutlich besser als die erste Schätzung (Abb. 2.10c).

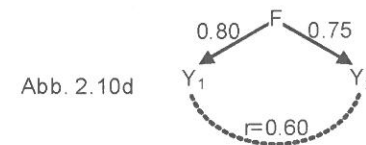
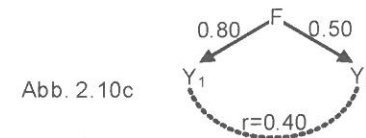


Abbildung 2.10: Schätzung von Pfadkoeffizienten (didaktisch vereinfachtes Beispiel)

Bei der SEM-Schätzung wird also eine geschätzte Kovarianzmatrix mit der empirisch beobachteten Kovarianzmatrix verglichen. Die Distanz zwischen beiden Matrizen wird durch den minimalen Wert der sogenannten "Fit-Funktion" angezeigt. Dieser Funktionswert liegt zwischen 0 und ∞ . Ist er null, dann sind die

beiden Matrizen identisch. Für eine gute Schätzung sollte er also möglichst nahe an null liegen.

Rein praktisch betrachtet arbeiten alle hier betrachteten Schätzverfahren bei überidentifizierten Modellen (a.a.O.: Identifikation) in iterativer Weise, d.h. sie suchen schrittweise nach Schätzwerten für die freien Parameter, um Σ möglichst dicht an S heranzubringen. Dazu beginnen sie mit bestimmten Startwerten und versuchen, so lange den Wert der Fit-Funktion zu optimieren, wie dies durch Veränderung der Schätzwerte noch möglich ist. Liegt die Veränderung unterhalb eines sehr geringen Schwellenwertes (häufig: 0.000001), so brechen sie die Iteration ab und konvergieren zu einer finalen Lösung. Kommt keine Konvergenz zustande, d.h. bewegt sich das Schätzverfahren in einer Endlosschleife, bei der die Veränderungen niemals so klein ausfallen, dass sie unterhalb des Abbruchkriteriums liegen, wird die Schätzung erfolglos beendet. Dafür kann es viele Gründe geben (vgl. dazu das folgende Kapitel 3).

Im Folgenden werden in aller Kürze einige Erläuterungen zu den zwei am häufigsten eingesetzten Schätzverfahren in der SEM-Analyse gegeben. Erläutert wird:

- (a) das "maximum likelihood" (ML/ML-robust)-Schätzverfahren, das zur Schätzung von Modellen mit metrischen und metrisch-definierten Variablen benutzt wird, und
- (b) das "weighted least squares" (WLS/WLSMV)-Schätzverfahren, das zur Schätzung von Modellen mit kategorialen Variablen zu verwenden ist.³³

Beide Schätzverfahren arbeiten nach dem Prinzip, dass sie diejenigen Schätzwerte suchen, die ihre Fit-Funktion minimieren. Sie liefern konsistente Schätzwerte, sodass mit größeren Fallzahlen (a.a.O.) auch eine größere Annäherung an die wahren Parameter der Population zu erwarten ist (sie werden deshalb auch als asymptotische Verfahren bezeichnet). Zudem sind ihre Schätzwerte bei großen Samples zumindest annäherungsweise normalverteilt, zeigen in ihrem Verteilungszentrum den wahren Wert der Population und sind zudem effizient, was bedeutet, dass die Verteilung der einzelnen Schätzwerte im Prinzip die kleinste mögliche Varianz erreichen kann, wodurch sie besonders präzise sein können (vgl. dazu auch die folgenden Unterkapitel).

³³ Im SEM-Programmpaket EQS wird noch eine weitere Möglichkeit zur Schätzung von SE-Modellen mit kategorialen Indikatoren angeboten, bei der einzelne Schritte aus ML- und WLS-Schätzung miteinander kombiniert werden (Byrne 2006: 163-176). Vgl. dazu die Erläuterungen im letzten Teil von Kap. 2.7.2.

2.7.1 Die ML/ML(robust)-Schätzung

Die Maximum-Likelihood-Schätzung (ML-Schätzung) wird in SEM-Analysen mit kontinuierlich verteilten bzw. metrisch skalierten endogenen Modellvariablen am häufigsten eingesetzt.³⁴

Bei Berechnung der Fit-Funktion im ML-Schätzverfahren (s.o.) wird eine Gewichtung der Summe der quadrierten Differenzen zwischen Σ und S (s.o.) benutzt. Im Detail ist die Fit-Funktion der ML-Schätzung kompliziert und wird hier nicht vorgestellt (vgl. dazu Bollen 1989: 107-111).

Die Bezeichnung "ML" kommt daher, dass im Schätzverfahren solche Schätzwerte gesucht werden, welche die Wahrscheinlichkeit (likelihood) dafür maximieren, dass bei Gültigkeit dieser Schätzwerte in der Population auch mit der jeweiligen Modellspezifikation die empirischen Kovarianzen gut zu beobachten sind.

Die Logik des ML-Verfahrens lässt sich an folgendem kleinen Beispiel verdeutlichen (nach Andreß et al. 1997: 40-45):

Gegeben seien zwei Populationen von jeweils 10 Personen. In der Population A sind 3 Personen als Postmaterialisten zu bezeichnen, in der Population B sind es 8 Personen. Bei einer Stichprobe von $N=3$ Personen sind alle 3 Personen als Postmaterialisten zu bezeichnen. Aus welcher Population stammt diese Stichprobe mit der höchsten Wahrscheinlichkeit? Dies lässt sich mit Hilfe der Binomialverteilung leicht berechnen (Urban/Mayerl 2011: 327). Die darin einzusetzenden Werte sind: $n=3$, $s=3$ (s = Personenanzahl mit der interessierenden Eigenschaft), $\pi=0.3$ (bei A), $\pi=0.8$ (bei B). Es ergeben sich Wahrscheinlichkeiten von 0.03 für A und 0.51 für B. Demnach bestätigt die Stichprobe die Gültigkeit von $\pi=0.8$ mit der maximal erreichbaren Wahrscheinlichkeit.

Die maximal erreichbare Wahrscheinlichkeit wird im ML-Schätzverfahren als "likelihood" bezeichnet, weil es sich dabei streng genommen nach den Axiomen der Wahrscheinlichkeitstheorie nicht um eine Wahrscheinlichkeit handelt (denn die Summe der likelihood-Werte kann größer 1.00 werden). Diese Unterscheidung ist aber für ein adäquates Verständnis des Schätzverfahrens nicht wichtig.³⁵

³⁴ In SEM-Analysen mit ML-Schätzung dürfen allerdings die exogenen Prädiktoren auch dichotom skaliert sein (z.B. die Variable "Geschlecht").

³⁵ In der Praxis benutzen die meisten Iterationsverfahren den negativen Log-Likelihoodwert "-LL" als Annäherungskriterium, so dass das Maximum der Schätzung dort erreicht wird, wo der absolute Wert von -LL am geringsten ist (d.h., dass bei Vergleich der beiden Log-Likelihoodwerte "-7.986" und "-10.658" der Wert von -7.986 einen besseren Fit indiziert).

Voraussetzung für eine möglichst unverzerrte ML-Schätzung ist die Gültigkeit zweier Annahmen:

Zum einen sollten die Werte aller exogenen Variablen (soweit sie metrisch verteilt sind) und die Werte aller endogenen Variablen möglichst gut multivariat normalverteilt sein (allerdings sind geringe bis mittlere Abweichungen vom Modell der Normalverteilung bei Einsatz einer Modifikation der Schätzung, s.u., tolerierbar).

Zum anderen sollten die exogenen Variablen nicht mit den Schätzfehlern (Residuen) der endogenen Variablen korrelieren (was eine ML-Schätzung von non-rekursiven Modellen, d.h. von Modellen mit Feedback-Schleifen, schwierig macht).

Ein spezielles Maximum-Likelihood-Schätzverfahren (hier als "ML(robust)"-Schätzung bezeichnet (a.a.O.)) ist u.a. in EQS und MPLUS implementiert und sollte bei leichten bis mittelgradigen Abweichungen der empirischen Variablenverteilungen vom Normalverteilungsmodell eingesetzt werden. Es liefert verteilungsrobust korrigierte Schätzwerte der Standardfehler (a.a.O.), die korrigierte Satorra-Bentler-SCALED- χ^2 -Statistik (a.a.O.), den robusten Anpassungsindex CFI (a.a.O.) sowie den robusten RMSEA-Wert (a.a.O.).

Vorteile der ML-Schätzung sind:

- (-) Sie ist weniger abhängig von der Fallzahl (a.a.O.) des zu analysierenden Datensatzes und von der Kurtosis der Modellvariablen, und ist deshalb auch stabiler und von höherer Präzision als andere Schätzverfahren (vgl. Olsson et al. 2000; Boomsma/Hoogland 2001).
- (-) Sie liefert bei Abwesenheit von Spezifikationsfehlern und bei Existenz von (annäherungsweise) normalverteilten Daten solche χ^2 -Testergebnisse (a.a.O.), die auch bei kleineren Stichproben ($N \leq 200$) unverzerrt sind (Curran et al. 1996).
- (-) ML(robust) (a.a.O.) liefert mit Satorra-Bentler-Korrektur (a.a.O.) auch bei nicht normalverteilten Variablen (und auch bei kleinen Fallzahlen von ca. $N=200$) relativ unverzerrte χ^2 -Testwerte (Curran et al. 1996).

Insgesamt betrachtet ist die ML-Schätzung bei Abwägung aller Störeinflüsse und deren Kombinationen (geringe Fallzahlen, nicht-normalverteilte Daten, Spezifikationsfehler) im Vergleich zu anderen Schätzverfahren die beste Wahl (Olsson et al. 2000; Boomsma/Hoogland 2001). Allerdings muss stets berücksichtigt werden, dass auch für das ML-Schätzverfahren die Schätzung von Standardfehlern (a.a.O.) und die Berechnung von Chi-Quadrat-Statistiken (a.a.O.) zum Test des Modellfits (a.a.O.) problematisch ist, wenn die Fallzahlen eher klein sind (Daumenregel: N sollte oberhalb von 200 liegen, a.a.O.: Fallzahlen), oder wenn wenige Indikatoren

pro Faktor benutzt werden (Daumenregel: pro Faktor sollten mehr als 3 Indikatoren gegeben sein, a.a.O.: Fallzahlen), oder wenn die Daten nicht (oder zumindest nicht annäherungsweise) multivariat normalverteilt sind (Boomsma/ Hoogland 2001).

Wenn bei einer SEM-Analyse zwar keine metrisch skalierten jedoch kategorial geordnete Indikatoren mit mindestens 5 Kategorien pro Variable zur Verfügung stehen, und wenn die Variablen höchstens moderat von der Normalverteilungsform abweichen (a.a.O.), wird von einigen SEM-Autoren empfohlen (u.a. von Bentler 2006), auch unter diesen Bedingungen die ML-Schätzung bzw. die ML(robust)-Schätzung einzusetzen. So könne den Problemen verzerrter Standardfehler-Schätzungen, die oftmals bei Schätzungen nach der WLS-Methode auftreten (vgl. die nachfolgenden Ausführungen) aus dem Weg gegangen werden.

Weitere ausführliche Informationen zur ML-Schätzung mit ordinal verteilten Variablen, die als metrisch definiert werden können, finden sich in Kap. 4.4 "Müssen die empirischen Variablenwerte immer metrisch und normalverteilt sein?"

2.7.2 Die WLS/WLSMV-Schätzung³⁶

Die WLS-Schätzmethode wird in der SEM-Analyse vor allem dann eingesetzt, wenn bei der Modellschätzung kategoriale Daten (d.h. Variablen, die auf nominalem oder ordinalem Messniveau liegen) als endogene Indikatorvariablen zu berücksichtigen sind. In den folgenden Erläuterungen konzentrieren wir uns auf eine spezielle Variante der WLS-Schätzmethode, die im SEM-Softwarepaket "Mplus" implementiert ist, und die von uns deshalb auch als "Mplus-Strategie zur WLS-Schätzung" bezeichnet wird.^{37, 38}

Die WLS-Schätzmethode ist recht kompliziert und in ihren mathematischen Algorithmen für den sozialwissenschaftlichen SEM-Anwender nur schwer zu verstehen. Im Folgenden sollen dennoch einige Hinweise zur WLS-Schätzmethode gegeben werden, um so zumindest eine erste Vorstellung von der Logik des Verfahrens zu vermitteln.

36 Die folgenden Ausführungen zum WLS-Schätzverfahren entsprechen weitgehend den Erläuterungen des Verfahrens in: Urban 2004: 18-26.

37 Vgl. dazu Muthén 1983, 1984, 1993; Muthén/Satorra 1995; Xie 1989; Muthén/Muthén 2001: "Appendix 1: Regression with a categorical dependent variable" (S. 339-343) und "Appendix 2: The general modeling framework" (S. 345-352).

38 In anderen Software-Paketen ist die WLS-Schätzung unter anderen Bezeichnungen implementiert z.B. als ADF in Amos und als AGLS in EQS.

Die Schätzmethode basiert auf einem WLS (weighted least squares)-Verfahren, das als Input empirische Informationen über Verteilungen und Zusammenhänge von kategorialen Variablen (z.B. polychorische Korrelationen oder Probitkoeffizienten, dazu mehr im Folgenden) sowie eine Schätzung von deren asymptotischer Kovarianzmatrix nutzt. Auf diese Weise kann das WLS-Verfahren auch solche Modellschätzwerte erzielen, für welche die Voraussetzung multivariater Normalverteilung nicht erfüllt ist.

Für das WLS-Verfahren ist vorauszusetzen, dass es für jede kategoriale Indikatorvariable (Y) eine dieser zugrunde liegende, kontinuierlich- und normalverteilte, latente Indikatorvariable (Y^*) gibt. Hinsichtlich dieser zugrunde liegenden, kontinuierlich verteilten Variablen (Y^*) wird angenommen, dass sie aus prinzipiellen oder auch aus praktischen Gründen nur unvollkommen gemessen werden kann und dass deshalb ihre jeweiligen Ausprägungen in der empirischen Forschung allein in kategorialer Form zu beobachten sind (mehr dazu im Folgenden).

Allerdings hat das "klassische" WLS-Schätzverfahren, das z.B. auch im Programmpaket LISREL installiert ist, einen entscheidenden Pferdefuß. Es werden sehr große Datensätze mit mehreren tausend Fällen (ca. 2000 bis 5000 Fälle)³⁹ benötigt, um die dafür erforderliche asymptotische Kovarianzmatrix zu schätzen (a.a.O.). Ohne solch große Fallzahlen kann das Verfahren zwar stabile Parameter-Schätzwerte, aber keine unverzerrten Schätzwerte für Standardfehler (a.a.O.) und Chi-Quadrat-Wert (a.a.O.) liefern. Letztere sind dann häufig so unzuverlässig, dass z.B. der für den üblichen Chi-Quadrat-Anpassungstest (a.a.O.) und für diverse, darauf beruhende Fit-Indizes (a.a.O.) benötigte Chi-Quadrat-Wert weit überhöht und deshalb unbrauchbar wird.

Somit ergibt sich das Dilemma, dass mit der WLS-Methode zwar ein Verfahren zur Schätzung von Strukturmodellen mit kategorialen Daten vorhanden ist, dafür jedoch Fallzahlen verlangt werden, die in sozialwissenschaftlichen Studien kaum zu erreichen sind. Um mit kategorialen Messwerten und kleineren Fallzahlen (z.B. schon mit $N=150$) dennoch zuverlässige und aussagekräftige SEM-Analysen durchführen zu können, muss zur Modellschätzung eine modifizierte Version des oben angesprochenen WLS-Verfahrens benutzt werden: die von uns sogenannte Mplus-Strategie der WLS-Schätzung.

Die Mplus-WLS-Schätzung benutzt einen Schätzalgorithmus, der auch noch kategoriale SE-Modelle mit 150 bis 200 Fällen relativ stabil schätzen kann.⁴⁰ Denn er verwendet eine diagonale Weight-Matrix, die nicht die Stabilitätsproble-

me des traditionellen WLS-Ansatzes (s.o.) kennt.⁴¹ Zudem gibt es eine Variante der Mplus-WLS-Schätzung, die für die praktische Sozialforschung höchst interessant ist. Sie wird als WLSMV-Schätzung bezeichnet (WLSMV= "weighted least squares estimator with standard errors and mean- and variance adjusted chi-square test statistic"). Diese WLS-Variante liefert mittelwert- und varianzjustierte robuste Schätzwerte für Standardfehler und Chi-Quadrat-Wert (vergleichbar mit der ML-Variante "ML(robust)", a.a.O.). Deshalb bleiben WLSMV-Schätzungen auch bei schief verteilten kategorialen Indikatorwerten noch relativ unverzerrt und können besonders gut für substanziell gehaltvolle inferenzstatistische Analysen genutzt werden.⁴²

Die WLS/WLSMV-Schätzung basiert im Wesentlichen auf einer Annahme und auf drei Schritten (S1 bis S3):⁴³

Die Annahme betrifft die bereits oben angesprochene Existenz einer kontinuierlich normalverteilten, latenten Indikatorvariablen (Y^*), auf die jede beobachtete geordnet-kategoriale Indikatorvariable (Y) zurückzuführen ist. Hinter jeder kategorialen Ausprägung von Y (also z.B. hinter den Zustimmungskategorien: 0=nein, 1=vielleicht, 2=ja eines trichotomen Items Y_1) stehen demnach ganze Wertebereiche der kontinuierlichen Variablen Y^* , und im ersten Schritt der WLS-Schätzung (S1) müssen die Grenzwerte, die sogenannten "Schwellenwerte" (thresholds) der Y^* -Variablen ermittelt werden, welche die einzelnen Wertebereiche für jeden Wert von Y voneinander abgrenzen. Was ist damit gemeint?

Wie bereits erwähnt, wird die Annahme getroffen, dass jeder Messwert einer kategorialen Y -Variablen stellvertretend für die sehr vielen Werte eines bestimmten Wertebereichs einer kontinuierlich normalverteilten Y^* -Variablen steht. Die Abbildung 2.11 verdeutlicht dies am Beispiel der kategorialen Variablen Y_1 , die prinzipiell nur einen von drei Variablenwerten annehmen kann (0, 1 oder 2). Zu jedem dieser drei Werte korrespondiert ein bestimmter Wertebereich der Y_1^* -Variablen. Für den kategorialen Wert "0" ist dies der gesamte Y_1^* -Bereich, der in Abbildung 2.11 links des Wertes von τ_1 liegt (unter Einschluss von τ_1 selbst). Für den kategorialen Wert "1" ist das der Y_1^* -Skalenabschnitt zwischen τ_1 und

41 Nach: B.O. Muthén in SEMNET vom 18.5.1999. Vgl. auch: Brown 2006: 388f; Kaplan 2000: 85-87.

42 Vgl. Muthén 1993; Kaplan 2000: 85-87. Die mittelwertjustierte Schätzung wird dabei ähnlich des "Satorra-Bentler-Verfahrens" zur Skalierung korrigierter Chi-Quadrat-Werte bei kontinuierlichen Indikatoren durchgeführt (vgl. Hu/Bentler/Kano 1992; Satorra/Bentler 1994).

43 Vgl. dazu: Kaplan 2000: 83f; Muthén/Muthén 2001: Appendix 1 ("Regression with a categorical dependent variable"), Appendix 2 ("The general modeling framework") und Appendix 4 ("Estimators").

39 Vgl. Muthén/Kaplan 1985; Yuan/Bentler 1994.

40 Nach Brown 2006: 389; Flora/Curran 2004.

τ_2 (unter Einschluss von τ_2). Und für den kategorialen Wert "2" ist dies der Y_1^* -Skalenabschnitt, der oberhalb von τ_2 liegt.

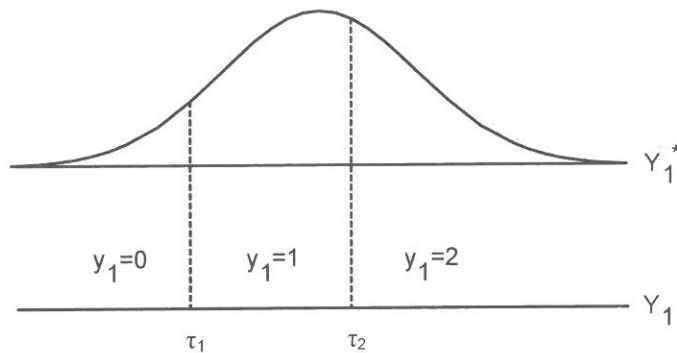


Abbildung 2.11: Illustration des Zusammenhangs zwischen der kategorialen manifesten Variablen Y_1 und der ihr zugrunde liegenden, kontinuierlich normalverteilten, latenten Variablen Y_1^*

Die Zuordnung von Y -Werten zu Y^* -Abschnitten erfordert also die Festlegung von τ -Werten, die auch als "Schwellenwerte" bezeichnet werden. Ihre Anzahl ist immer um eins kleiner als die Anzahl aller Kategorien (k) der Y -Variablen ($k-1$, hier also: $3-1 = 2$ Schwellenwerte: τ_1 und τ_2).

Diese Schwellenwerte können in der WLS/WLSMV-Schätzstrategie von Mplus auf zwei verschiedenen Wegen ermittelt werden:

- Wenn es für einen Faktor mehrere kategoriale Indikatoren und gleichzeitig noch mindestens einen Prädiktor für diesen Faktor oder zumindest für einen seiner Indikatoren gibt, wird ein multivariates Probit-Regressionsverfahren zur Schätzung der Schwellenwerte eingesetzt.⁴⁴
- Wenn es mehrere kategoriale Indikatoren ohne zusätzliche Prädiktoren (X -Variablen) gibt, sind die Schwellenwerte nichts anderes als "Z-Werte". Diese werden berechnet, indem der kumulative Prozentanteil von Fällen für jede Y -Kategorie ermittelt wird und mit diesem Prozentwert dann der dazu korrespondierende Wert der Wahrscheinlichkeitsverteilung einer Standardnormalverteilung gesucht wird (die entsprechenden Verteilungswerte gibt es in vielen Statistikbüchern in Tabellenform). Haben z.B. etwa 28% aller Beob-

⁴⁴ Dieses Verfahren verlangt noch nicht einmal die umfassende multivariate Normalverteilung aller Y^* -Indikatoren, sondern begnügt sich mit der weniger restriktiven Annahme einer konditionalen Normalverteilung von Y^* bei Vorliegen bestimmter Werte der entsprechenden Kovariaten (vgl. Muthén/Muthén 2001: 342).

achtungsfälle einen Wert von 0, so ergibt dafür ein Schwellenwert von 0.58. Und haben 88% der Beobachtungsfälle einen Wert von 0 oder 1, so ergibt sich dafür ein Schwellenwert von 1.17.

Sind die Schwellenwerte ermittelt, können im zweiten Schritt (S2) der WLS/WLSMV-Schätzung die latenten Korrelationen berechnet werden. Die Korrelationen zwischen direkt beobachteten Prädiktorvariablen X und den Y^* -Variablen werden als "polyserielle Korrelationen"⁴⁵ berechnet und die Korrelationen zwischen Variablen, die allein Y^* -Variablen sind, werden als "polychorische Korrelationen" ermittelt.⁴⁶ Dabei werden statistische Informationen benutzt, die sich auf jede mögliche Kombination von Werten der kategorial gemessenen Y -Variablen beziehen (so ergeben sich z.B. für zwei kategoriale Variablen mit den gemessenen Wertekategorien von 0/1/2 und 0/1 insgesamt sechs mögliche Wertekombinationen). Zu den benötigten Informationen gehören u.a. sowohl die Anzahl der Beobachtungsfälle, die für jede Wertekombination vorhanden sind (mehr dazu im Folgenden), als auch die Schwellenwerte, die dem jeweiligen Wertepaar zugeordnet werden können.⁴⁷

Somit steht für den anschließenden, dritten WLS/WLSMV-Schritt (S3) keine Kovarianzmatrix (wie in der traditionellen SEM-Schätzung mit kontinuierlichen Indikatorvariablen), sondern eine Korrelationsmatrix der Y^* -Variablen bzw. der Y^* - und X -Variablen zur Verfügung. Es wird dann ein Schätzer für die Kovarianzmatrix der latenten Korrelationen ermittelt und zur Kalkulation der WLS/WLSMV-Schätzung aller freien Parameter des SE-Modells benutzt. Für diese Schätzung wird im Falle einer robusten WLSMV-Schätzung die oben erwähnte

⁴⁵ Eine polyserielle Korrelation ist eine Korrelation "between an observed continuous variable and a latent continuous variable that underlies an ordinal variable, assuming that the observed and the latent continuous variables follow a bivariate normal distribution." (Xie 1989: 329).

⁴⁶ Eine polychorische Korrelation ist eine Korrelation "between two latent continuous variables that are assumed to be distributed as bivariate normal and to have generated the observed ordinal variables through thresholds." (Xie 1989: 329). Die ML-Schätzung der polychorischen Korrelation erfolgt nach den Gleichungen 9.105 und 9.106 in Bollen 1989: 442.

⁴⁷ Die Abhängigkeit polychorischer Korrelationskoeffizienten von den Fallzahlen einzelner Kategorienkombinationen kann erklären, warum die Schätzung der Koeffizienten instabil wird, wenn für viele Wertekombinationen keine oder nur sehr wenige Beobachtungsfälle vorhanden sind. Zudem ist eine polychorische Korrelationsschätzung unmöglich, wenn alle Fälle einer bestimmten Kategorie der Variablen Y_1 nur in Kombination mit einer einzigen Kategorie der Variablen Y_2 auftreten (wenn also die Fälle für jede Kategorie einer beliebigen Variablen Y_1 nicht über mehrere Kategorien einer beliebigen Variablen Y_2 streuen). Das entsprechende Risiko wird natürlich gemeinsam von der Kategorienanzahl der einzelnen kategorialen Indikatoren und von den vorhandenen Fallzahlen beeinflusst. Bei kleinen Fallzahlen sollten deshalb möglichst Kategorisierungen mit wenigen Werten benutzt werden.

diagonale Weight-Matrix benutzt. Sie liefert u.a. die Schätzwerte für die freien Pfadkoeffizienten und Faktorladungen des SE-Modells.

Die ermittelten Schätzwerte sind nicht einfach zu interpretieren, wenn als abhängige Variable die geordnet-kategoriale Variable Y betrachtet werden soll.⁴⁸ Deshalb wird in der kategorialen SEM-Analyse mit WLS/WLSMV-Schätzung fast immer nicht Y sondern Y^* als abhängige Variable betrachtet. Die Variable Y^* wird bei Analyse eines " $X \rightarrow Y$ "-Effektes stets als zwischengeschaltete, kontinuierliche Größe verstanden ($X \rightarrow Y^* \rightarrow Y$), die in einer nicht-linearen (S-kurvigen) Beziehung zur Wahrscheinlichkeit von Y steht. Und ist nicht Y sondern Y^* die abhängige Variable, so sind die Pfadkoeffizienten als kontinuierliche Veränderungsraten einer kontinuierlich skalierten abhängigen Variablen Y^* zu interpretieren (so wie bei der klassischen OLS-Regression).

Allerdings fehlt den Y^* -Variablen eine empirisch zu interpretierende Skalierung. Um diese zu erhalten, müsste auf das komplizierte, nicht-lineare Verhältnis zu Y zurückgegriffen werden. Wenn aber darauf, wie üblich, verzichtet wird, kann über die Beziehung zwischen Y^* und Y allein gesagt werden, dass wenn Y^* ansteigt (z.B. aufgrund eines positiven Effektes von X), auch die Wahrscheinlichkeit für einen höheren Wert bei Y ansteigt. Eine exakte, quantitative Kalkulation dieses Anstiegs erforderte jedoch eine recht aufwändige Berechnung.

Alternativ zur hier erläuterten WLS/WLSMV-Schätzmethode kann zur Schätzung von SE-Modellen mit kategorialen Variablen auch das im EQS-Programmpaket implementierte dreistufige Schätzverfahren eingesetzt werden, in dem einzelne Schritte aus WLS- und ML-Schätzung miteinander kombiniert werden (vgl. dazu Byrne 2006: 163-176):

- Schritt 1: Schätzung von "Schwellenwerten" (s.o.) unter Annahme von normalverteilten, kontinuierlichen Variablen Y^* , die den kategorialen Messwerten zugrunde liegen (s.o.);
- Schritt 2: Berechnung von polychorischen/polyseriellen Korrelationen (s.o.) zwischen den metrisch rekodierten, ursprünglich kategorial gemessenen Modellvariablen (s.o.);
- Schritt 3: Schätzung der freien Modellparameter auf der Basis der berechneten polychorischen/polyseriellen Korrelationen (s.o.) unter Verwendung des ML(robust)-Schätzverfahrens (a.a.O.).

Allerdings fehlen zu dieser Methode noch systematische Evaluationen, die insbesondere die Stabilität und Unverzerrtheit des Verfahrens in Abhängigkeit von der

48 Vgl. dazu Muthén/Muthén 2001: Appendix 1 ("Regression with a categorical dependent variable"); Agresti 1990; Hosmer/Lemeshow 1989; Maddala 1983.

modellspezifischen Messkomplexität (Anzahl der Schwellenwerte pro Modellvariable) und den jeweils zur Verfügung stehenden Fallzahlen untersuchen.

2.8 Wann entsteht ein "Identifikationsproblem" bei der Konstruktion von SE-Modellen?

Damit die freien (und auch die restringiert-freien, a.a.O.) Parameter eines Strukturgleichungsmodells eindeutig geschätzt werden können, muss jedes SE-Modell "identifiziert" sein. Identifiziertsein heißt, dass genügend empirische Information zur Verfügung steht, um alle zu schätzenden Modellparameter auch unzweifelhaft ermitteln zu können. Denn genau so, wie arithmetische Gleichungssysteme mit mehr Unbekannten als Gleichungen nicht zu lösen sind,⁴⁹ sind auch die Gleichungssysteme von SE-Modellen, in denen eine möglichst weitgehende Annäherung von Σ und S erreicht werden soll (a.a.O.: Modellschätzung), immer dann nicht zu schätzen, wenn dazu die zur Verfügung stehende empirische Information nicht ausreicht.

Für eine erfolgreiche SEM-Schätzung wird also ein empirisch identifiziertes SE-Modell benötigt. Wenn allerdings in einer SEM-Analyse auch noch statistische Tests durchgeführt werden sollen, wenn z.B. der Modellfit (a.a.O.) des Gesamtmodells getestet werden soll, oder wenn überprüft werden soll, ob bestimmte Einflussverbindungen sinnvoll sind oder durch alternative Pfade ersetzt werden sollen, muss das zu schätzende SE-Modell überidentifiziert sein.

SE-Modelle sind dann überidentifiziert, wenn zu ihrer Schätzung mehr empirische Information zur Verfügung steht, als eigentlich zur Ermittlung der Parameter-Schätzwerte des betreffenden Modells erforderlich wäre. Da solche statistischen Tests üblicherweise in jeder SEM-Analyse durchgeführt werden, sollten die zu analysierenden SE-Modelle in aller Regel auch überidentifiziert sein (s.u.).

Grundsätzlich sollte der Identifikationsgrad von SE-Modellen für drei verschiedene Modellebenen festgestellt werden:

- (a) für jedes einzelne Messmodell,
- (b) für das Strukturmodell,
- (c) für das Gesamtmodell (bestehend aus Messmodell/en und Strukturmodell).

49 Die Gleichung: " $x+2y=7$ " ist unteridentifiziert, weil es eine unendliche Anzahl von Lösungen für x und y gibt (z.B. $x=5$, $y=1$ oder $x=6$, $y=0,5$). Es gibt in der Gleichung mehr Unbekannte als Bekannte. Demgegenüber ist das Gleichungssystem, das aus den zwei Gleichungen besteht: " $x+2y=7$ " und " $3x-y=7$ ", genau identifiziert, weil es dafür nur eine einzige Lösung gibt ($x=3$, $y=2$).