

*Dieter Holtmann*

## Multivariate Modellbildung für metrische Daten

Abstract: The application of multiple regression and path analysis is discussed in regard to the exclusive use of the beta coefficients. Beta is one of the possible ways of controlling for the effects of the remaining predictors, others are part and partial correlation, part and partial covariance. A typology is developed for the difference between total and controlled effect. With this instrument (for controlled effect = beta) it can be shown under which conditions the sum of the total explaining power ( $r^2_{y,x_i}$ ) and the sum of the additional explaining power (Part correlation<sup>2</sup>) do exceed or not Multiple R<sup>2</sup>. By the criteria of sign rule and of the relative advantages of the decomposition of total effect in controlled effect and rest effect it is argued that beta is still the best way (of those considered) of control. Next some problems of the otherwise recommended LISREL approach are discussed: Estimation, data-fitting versus model test, model-fitting, standardization. Finally some parallels between models for metric and models for non-metric data are specified.

In der folgenden Darstellung werden einige Probleme der Pfadanalyse, nicht-rekursiver Modelle und der Berücksichtigung von latenten Konstrukten (LISREL) diskutiert, es kann sich dabei natürlich nur um eine Auswahl von Problemen handeln, wobei ich mich auf einige Punkte beschränke, die in der Literatur noch nicht oder nicht hinreichend behandelt worden sind, weitere Probleme werden in der angeführten Literatur diskutiert.

### A) Pfadanalyse

Die Pfadanalyse ist eine Synthese aus Gleichungssystemen (mittels Regression gewonnen) und Kausalstruktur (graphisch dargestellt durch ein Pfeildiagramm).

Kausalmodelle bzw. Pfadmodelle bzw. structural equation models haben wohl deshalb eine so starke Verbreitung unter empirischen Soziologen gefunden, weil sie die Lücke zwischen Theorie

und Empirie zu schließen versprechen.

Dies ist insofern richtig, als hier kausale Annahmen aus der Theorie mit geschätzten Daten aus der Empirie "belegt" werden. (Oder: Kausalrelationen und Korrelationen werden in einem Modell vereint.) Allerdings läßt sich jede mögliche der Kausalordnungen in diesem technischen Sinne "belegen": Die Daten sind verträglich mit verschiedenen Modellen. Benötigt werden Kriterien, um zu entscheiden, welches Kausalmodell sich durch die Daten besser belegen läßt. Dazu wäre z.B. auszuwerten, wie hoch der Erklärungsanteil der verschiedenen Teilsysteme (Regressionen) ist. Da in den Modellen aber i.a. mehrere Regressionen auftreten, liegt die Gesamtbeurteilung nicht auf der Hand. Die Lösung könnte in den "full information" Schätzungen und Tests liegen (wie z.B. Jöreskogs LISREL-Ansatz, s.u.), da dann das Gesamtmodell beurteilt werden kann.

Die grundsätzlichsste Kritik der Pfadanalyse ist nach meiner Auffassung die Problematisierung der Zerlegung der Effekte in direkte und indirekte, die im folgenden genauer diskutiert werden soll. (Mein Haupteinwand gegen die Pfadanalyse bezieht sich auf eine derartige Verwendung, daß man allein die  $\beta$ -Koeffizienten der Regression berücksichtigt, die nur einen begrenzten Aussagewert haben: Es sind ja zunächst nur die besten Koeffizienten im technischen Sinne der Prädiktorgewichtung zur Vorhersage.)

In ihrem sehr guten Überblick über die Modellbildung in den Sozialwissenschaften argumentieren Kiiveri und Speed (1982, 279):

"We have said nothing about path analysis (...) because of its potential to mislead in modeling multivariate data. (...) Whether the path coefficients are useful in interpreting sociological models we leave sociologists to decide, but we do believe that the decomposition of a covariance or correlation into its direct and indirect components has no useful statistical meaning."

Der Kritik an der Pfadanalyse ist insofern zuzustimmen, als neuere Entwicklungen wie LISREL (Linear Structural

Relationships) einen Gesamttest des Modells erlauben und insofern vorzuziehen sind. Das Argumentieren mit direkten und indirekten Effekten ist aber auch in LISREL gebräuchlich, dort werden direkte und indirekte Effekte bereits vom Programm selbst berechnet. Es soll nun diskutiert werden, ob und wie man dieser Zerlegung einen statistischen Sinn verleihen kann.

### 1) Korrelierte Effekte (Multiple Regression)

Wegen der größeren Übersichtlichkeit der Formeln werden standardisierte Variablen,  $y, x_1, \dots, x_k$  vorausgesetzt, die Zurückrechnung auf nicht standardisierte Variablen ist ja immer möglich.

Das Problem der multiplen Regression besteht bekanntlich darin, eine abhängige Variable  $y$  aufgrund einer Linearkombination

$$\hat{y} = \sum_{i=1}^k \beta_i x_i$$

der Prädiktoren möglichst gut zu approximieren im Sinne der Minimierung der Fehler

$$\sum_{j=1}^n (y_j - \hat{y}_j)^2 \quad (\text{über die Einheiten } j).$$

Die Lösung der multiplen Regression lautet bekanntlich:

$\beta = C^{-1}r$ , wobei:

$$\beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} \text{ der Koeffizientenvektor ist, } r = \begin{pmatrix} r_{y,x_1} \\ \vdots \\ \bar{r}_{y,x_k} \end{pmatrix} \text{ der}$$

Vektor der Korrelationen der abhängigen Variablen mit den Prädiktoren und  $C$  die Korrelationsmatrix der Prädiktoren.

Also:  $C\beta = r$ , ausführlicher:

$$r_{y,x_i} = \sum_{j=1}^k r_{x_i,x_j} \beta_j$$

Mit Hilfe des Skalarprodukts  $\langle a, b \rangle := \sum_{i=1}^n a_i b_i$  für 2 beliebige Vektoren  $a=(a_1, \dots, a_n)$ ,  $b=(b_1, \dots, b_n)$  erhält man:

$$\begin{aligned} \langle y, x_i \rangle &= \sum_{j=1}^k \langle \beta_j x_j, x_i \rangle \\ &= \langle \hat{y}, x_i \rangle \end{aligned}$$

oder:  $\langle y - \hat{y}, x_i \rangle = 0$ , d.h. das Residuum  $y - \hat{y}$  ist orthogonal zu allen Prädiktoren  $x_i$ , korreliert also nicht mit ihnen.

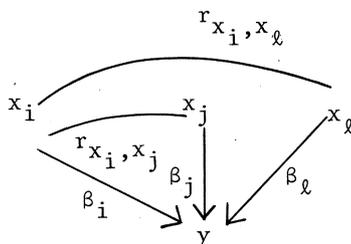
$$\text{Es folgt: } r_{y, x_i} = \beta_i + \sum_{\substack{j \\ j \neq i}} r_{x_i, x_j} \beta_j$$

Rein algebraisch ist also die Korrelation zerlegt in verschiedene Bestandteile. Dies belegt noch nicht, daß sich dies besonders gut interpretieren läßt, dazu kann man aber folgendes berücksichtigen:  $r_{y, x_i}$  läßt sich als Gesamteffekt eines Prädiktors  $x_i$  auf die zu erklärende Variable  $y$  betrachten, denn eine Änderung von  $x_i$  um eine Einheit ergibt eine Änderung in  $y$  um  $r_{y, x_i}$ , weil die einfache Regression lautet (standardisierte Variablen unterstellt):

$$y = r_{y, x_i} x_i + \varepsilon_i$$

$$\text{Also: } \Delta y = r_{y, x_i} \underbrace{\Delta x_i}_{1} \quad (x_i \text{ korreliert nicht mit dem Fehler } \varepsilon_i)$$

Die folgende graphische Darstellung illustriert die Unterscheidung in direkte und korrelierte Effekte:

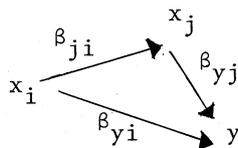


$$y = \beta_i x_i + \sum_{j \neq i} \beta_j x_j + \epsilon$$

Eine Änderung von  $x_i$  um eine Einheit bewirkt direkt eine Änderung in  $y$  um  $\beta_i$ , in diesem Sinne ist  $\beta_i$  der direkte Effekt. Da  $x_i$  mit den  $x_j$  ( $j \neq i$ ) korreliert, ergibt die Änderung in  $x_i$  eine Änderung in  $x_j$  ( $j \neq i$ ) und dadurch schließlich eine weitere Änderung in  $y$ . ( $x_i$  korreliert nicht mit dem Fehler  $\epsilon$ .) Alle Änderungen letzterer Art zusammen sollen als korrelierte Effekte verstanden werden.

$$r_{y, x_i} = \beta_i + \sum_{j \neq i} r_{x_i, x_j} \beta_j$$

II) Indirekte Effekte (Pfadanalyse)



Die Regressionsgleichungen eines solchen Pfadmodells lauten:

$$x_j = \beta_{ji} x_i + \epsilon_j$$

$$y = \beta_{yi} x_i + \beta_{yj} x_j + \epsilon$$

Eine Änderung in  $x_i$  um eine Einheit ergibt unmittelbar eine Änderung  $\beta_{yi}$  in  $y$ , dies soll wieder als direkter Effekt

verstanden werden. Die Änderung in  $x_i$  ergibt außerdem eine Änderung in  $x_j$  um  $\beta_{ji}$  und deshalb eine Änderung  $\beta_{yj} \beta_{ji}$  in  $y$ .

$$y = \beta_{yi} x_i + \beta_{yj} \beta_{ji} x_i + \beta_{yj} \varepsilon_j + \varepsilon$$

Die Residuen  $\varepsilon_j$  und  $\varepsilon$  korrelieren nicht mit  $x_i$ , also impliziert die Änderung in  $x_i$  keine weitere Änderung in  $y$ .

Die algebraische Zerlegung des Gesamteffekts in den direkten und indirekten Effekt ergibt sich aus:

$$r_{y, x_i} = s_{y, x_i} = s_{y, x_i}^{\hat{}} = \beta_{yi} + \beta_{yj} \beta_{ji}$$

(Die Verallgemeinerung auf mehr indirekte Pfade besteht dann darin, daß der indirekte Effekt die Summe der Koeffizientenprodukte entlang der Pfade ist.)

Insgesamt:

1) Ändert man  $x_i$  um eine Einheit, so ändert sich  $y$  um  $r_{y, x_i}$ .

2) Für  $\beta_i$  muß man (vgl. Holtmann 1980) dagegen wie folgt argumentieren:

$$\beta_i = \frac{s_{y, x_i - \hat{x}_i}}{s_{x_i - \hat{x}_i}}$$

Dies ist der Regressionskoeffizient in der einfachen Regression von  $y$  auf den bereinigten Prädiktor  $x_i - \hat{x}_i$ . Ändert man also  $x_i - \hat{x}_i$  um eine Einheit, so ändert sich  $y$  um  $\beta_i$ . Problematisch daran ist, daß  $x_i - \hat{x}_i$  nicht standardisiert ist, die Änderung um eine Einheit bei  $x_i - \hat{x}_i$  und  $x_j - \hat{x}_j$  ( $j \neq i$ ) ist also nicht ohne weiteres vergleichbar.

3) Betrachtet man deshalb die standardisierte bereinigte

Variable  $\frac{x_i - \hat{x}_i}{s_{x_i - \hat{x}_i}}$ , so ergibt eine Änderung dieser Variablen um eine Einheit in  $y$  die Änderung  $r_{y, x_i - \hat{x}_i}$ , dies ist gerade Part Correlation.

4) Die Gesamterklärungskraft eines Prädiktors ist:  $r_{y,x_i}^2$

Die zusätzliche Erklärungskraft eines Prädiktors ist:

$$r_{y,x_i-\hat{x}_i}^2 = R^2 - R_{(i)}^2. \quad \beta_i^2 \text{ ist kein Erklärungsanteil. Ferner}$$

gibt es keine standardisierte Variable, deren induzierte Änderung  $\beta_i$  ergäbe. D.h., daß die Interpretation von  $\beta_i$  als Effekt nicht so unproblematisch ist, wie der Begriff Effekt aus der orthogonalen Varianzanalyse bei experimentellem Design suggeriert.

5) Da die Prädiktoren korrelieren dürfen, läßt sich der Effekt von  $x_i$  nicht von dem Einfluß der übrigen Prädiktoren isolieren. Gerade als Reaktion darauf ließe sich die Unterscheidung in einen direkten und einen indirekten Effekt interpretieren: Eine Änderung von  $x_i$  um eine Einheit induziert nicht nur die Änderung  $\beta_i = \beta_i \underbrace{\Delta x_i}_1$  in  $y$ , sondern außerdem noch:

$$\sum_{\substack{j \\ j \neq i}} r_{x_i, x_j} \beta_j$$

### III) Problematisierung der Pfadanalyse

In der Pfadanalyse werden in einem Diagramm die direkten Effekte  $\beta_i$  eingetragen, um die relative Einflußstärke der Variablen auszudrücken. Genau so wichtig aber ist der Gesamteffekt  $r_{y,x_i}$  einer Variablen. Es soll nun im folgenden gezeigt werden, daß die Beschränkung der Interpretation auf die  $\beta_i$  völlig unangemessen ist, da die  $r_{y,x_i}$  in jeder Richtung von den  $\beta_i$  abweichen können. Eine Berücksichtigung beider Koeffizienten dagegen verhilft zur Klärung der Situation.

Im folgenden soll mit dem Konzept des Gesamteffekts ( $r_{y,x_i}$ ) des Prädiktors  $x_i$  und des bereinigten Effekts ( $\beta_i$ ), d.h. des Effekts des bereinigten Prädiktors  $x_i - \hat{x}_i$ , gearbeitet werden.

Der Effekt ist die induzierte Änderung in der abhängigen Variablen, wenn der Prädiktor um eine Einheit geändert wird.

Also:

Gesamt-Effekt von  $x_i$ :  $r_{y,x_i}^2$

Bereinigter Effekt von  $x_i$ ,  
d.h. Effekt des bereinigten  
Prädiktors  $x_i - \hat{x}_i$ :  $\beta_i$

Gesamt-Erklärungsbeitrag  
von  $x_i$ :  $r_{y,x_i}^2$

Bereinigter Erklärungs-  
beitrag von  $x_i$ , d.h. Bei-  
trag des bereinigten Prä-  
diktors  $x_i - \hat{x}_i$ :  $r_{y,x_i - \hat{x}_i}^2 = R^2 - R_{(i)}^2$

In Holtmann (1980) waren nun zwei Zerlegungen von Multiple  $R^2$  angegeben:

$$R^2 = \sum_{i=1}^k r_{y,x_i - \hat{x}_i}^2 + \Delta_1$$

und

$$R^2 = \sum_{i=1}^k r_{y,x_i}^2 - \Delta_2$$

Meine Bezeichnung dieser Zerlegungen als Zerlegung "von unten" und Zerlegung "von oben" ist irreführend, denn  $\Delta_1$  und  $\Delta_2$  müssen nicht positiv sein, sie können auch Null oder kleiner Null sein, solche Phänomene sollen nun genauer untersucht werden.

Die Ursache des Problems besteht darin, daß  $x_i - \hat{x}_i$  orthogonal zu allen  $x_j$  ( $j \neq i$ ) ist, nicht aber zu  $\hat{x}_j$ , da in  $\hat{x}_j$  wieder der Prädiktor  $x_i$  auftritt. Rechnet man aus  $x_i$  die übrigen Prädiktoren heraus, so korreliert  $x_i - \hat{x}_i$  nicht mehr mit  $x_j$ . Bereinigt

man auch  $x_j$  um die übrigen Prädiktoren, so kommt dadurch wieder der Einfluß von  $x_i$  herein:  $x_i - \hat{x}_i$  und  $x_j - \hat{x}_j$  ( $j \neq i$ ) müssen nicht orthogonal sein. Die einzige Möglichkeit, dieses Problem zu vermeiden, ist die Vorgabe einer Hierarchie der Prädiktoren: Rechnet man aus  $x_i$  alle Prädiktoren  $x_\ell, \ell < i$  heraus, und aus  $x_j$  ( $j > i$ ) alle Prädiktoren  $x_\ell, \ell < j$ , so tritt das Problem nicht mehr auf:  $x_j - \hat{x}_j$  ist orthogonal zu allen  $x_\ell, \ell < j$ , also auch zu  $x_i - \hat{x}_i$ . In  $x_i - \hat{x}_i$  tritt  $x_j$  nicht auf, weil  $\hat{x}_i$  die Regression von  $x_i$  auf  $x_1, \dots, x_{i-1}$  ist und  $j > i$ .

#### Vergleich von Gesamteffekt und bereinigtem Effekt: Typologie

Der Gesamteffekt (ges) läßt sich zerlegen in den bereinigten Effekt (ber) und einen Rest (res=ges-ber).

(Die Typologie ist so allgemein formuliert, daß sie sich auch z.B. für ber=Part Correlation verwenden ließe. Die Verwendung von ber=  $\beta_i$  scheint mir aber fruchtbarer, da dann res(= korrelierte/indirekte Effekte) besser zu analysieren ist, wie unten auch an einem Beispiel gezeigt wird.)

a) Ein Suppressor-Phänomen liegt vor, wenn der bereinigte Effekt von  $x_i$  (dies sei  $\beta_i$ ) größer ist als der Gesamteffekt ( $r_{y,x_i}$ ). D.h. die Störfaktoren supprimieren den "wahren" Einfluß  $\beta_i$  von  $x_i$ .

Formal:

$$\text{ges} > 0, \text{ so: } \text{ber} > \text{ges} > 0 \quad (\text{Es folgt: } \text{res} < 0)$$

$$\text{ges} < 0, \text{ so: } \text{ber} < \text{ges} < 0 \quad (\text{Es folgt: } \text{res} > 0)$$

Daraus folgt jeweils:  $|\text{ber}| > |\text{res}|$

$$\text{sign}(\text{ber}) \neq \text{sign}(\text{res}), \text{ sign}(\text{ges}) \neq \text{sign}(\text{res})$$

$$\text{sign}(\text{ges}) = \text{sign}(\text{ber})$$

┌

a1) Sei  $\text{ges} > 0$ .

$$\text{ber} > \text{ges} = \text{ber} + \text{res} \Rightarrow \text{res} < 0, -\text{res} > 0$$

$$0 < \text{ges} = \text{ber} + \text{res} \Rightarrow -\text{res} < \text{ber}$$

$$\text{Also: } |\text{res}| < |\text{ber}|$$

a2) Sei  $ges < 0$ .  
 $ber < ges = ber + res \Rightarrow res > 0$   
 $0 > ges = ber + res \Rightarrow res < -ber$   
 Also:  $|res| < |ber|$

b) Ein Distorter-Phänomen (Verzerrung) liegt vor, wenn der bereinigte Einfluß von  $x_i$  (dies sei  $\beta_i$ ) ein entgegengesetztes Vorzeichen zu dem Gesamteffekt ( $r_{y,x_i}$ ) hat. D.h. die Störfaktoren verzerren den "wahren" Einfluß von  $x_i$ .

Formal:  $ges > 0, so: ges > 0 > ber$  (Es folgt:  $res > 0$ )  
 $ges < 0, so: ges < 0 < ber$  (Es folgt:  $res < 0$ )  
 Daraus folgt jeweils:  $|ber| < |res|$   
 $sign(ber) \neq sign(res), sign(ges) = sign(res)$   
 $sign(ges) \neq sign(ber)$

b1) Sei  $ges > 0$ .  
 $ber + res = ges > ber \Rightarrow res > 0$   
 $0 < ges = ber + res \Rightarrow res > -ber$   
 Also:  $|res| > |ber|$

b2) Sei  $ges < 0$ .  
 $ber + res = ges < ber \Rightarrow res < 0$   
 $0 > ges = ber + res \Rightarrow ber < -res$   
 Also:  $|ber| < |res|$

Dies kann man, wie in der Tabellenanalyse bekannt, in anschaulicher Form durch folgendes Schema darstellen:

	Gesamteffekt	Bereinigter Effekt	Rest
Suppressor Phänomen	+	++	-
	-	--	+
Distorter Phänomen	+	-	++
	-	+	--

c) Ein "Schein-Effekt" ist ein abgeschwächtes Distorter Phänomen: Rechnet man die Störfaktoren heraus, so wechselt zwar nicht das Vorzeichen, aber der "wahre" Effekt (=  $ber$ ) erweist

sich als verschwindend.

Formal:  $ges > 0, so: ber = 0$  (Es folgt:  $res > 0$ )

$ges < 0, so: ber = 0$  (Es folgt:  $res < 0$ )

Die Subsumtion unter das Distorter-Phänomen zeigt sich daran,

daß:  $|ber| = 0 < |res|$ ,  $sign(ges) = sign(res)$

Ferner gilt:  $ber = 0, ges = res$

┌ c1)  $0 < ges = ber + res = res \Rightarrow res = ges > 0.$

Also:  $|res| > |ber|$

c2)  $0 > ges = ber + res = res \Rightarrow res = ges < 0.$

$-res > 0.$

└ Also:  $|res| > |ber|$

Schema:

	Gesamteffekt	Bereinigter Effekt	Rest
Schein-Effekt {	+	0	+
	-	0	-

d) Eine "Überschneidung" liegt vor, wenn der "wahre" Einfluß von  $x_1$  ( $:=ber$ ) als geringer sichtbar wird (aber nicht das Vorzeichen wechselt oder Null wird), wenn die anderen Prädiktoren als Störfaktoren herausgerechnet werden. D.h. es handelt sich um eine Abschwächung des "Schein-Effekts".

Formal:

$ges > 0, so: ges > ber > 0$  (Es folgt:  $res > 0$ )

$ges < 0, so: ges < ber < 0$  (Es folgt:  $res < 0$ )

Die Subsumtion unter die Distorter Richtung zeigt sich daran,

daß:  $sign(ges) \neq sign(res)$

Insgesamt:  $sign(ber) = sign(res) = sign(ges)$

┌ d1) Sei  $ges > 0.$

$ber < ges = ber + res \Rightarrow res > 0$

d2) Sei  $ges < 0.$

└  $ber > ges = ber + res \Rightarrow res < 0$

Schema:

	Gesamteffekt	Bereinigter Effekt	Rest
Überschneidung	++ --	+ -	+ -

e) Ein "scheinbarer no-effect" liegt vor, wenn die Existenz eines Einflusses erst sichtbar wird, wenn die übrigen Prädiktoren herausgerechnet werden. Dies ist ein Suppressor-Phänomen mit Gesamteffekt Null.

Formal:

ges = 0, ber > ges = 0 (Es folgt: res < 0)  
 oder:  
 ges = 0, ber < ges = 0 (Es folgt: res > 0)  
 Insgesamt: ber = -res

┌ 0 = ges = ber + res => ber = -res  
 e1) ber > 0 => res < 0  
 └ e2) ber < 0 => res > 0

Schema:

	Gesamteffekt	Bereinigter Effekt	Rest
Scheinbarer no-effect	0 0	+ -	- +

f) Schließlich können Gesamteffekt und bereinigter Effekt übereinstimmen, d.h. die anderen Prädiktoren wirken nicht als Störfaktoren.

Formal:

ges = ber (Es folgt: res = 0)

Schema:

	Gesamteffekt	Bereinigter Effekt	Rest
Übereinstimmung	+	+	0
	-	-	0
	0	0	0

g) Identifikation der Typen

- A) Falls  $ges = ber$  : Übereinstimmung
- B) Falls  $ges \neq ber$ 
  - B1) Falls  $ges \cdot ber < 0$ , so Distorter Phänomen
  - B2) Falls  $ges \cdot ber = 0$ 
    - B2a) Falls  $ges=0$  (wegen  $ber \neq ges$  gilt  $ber \neq 0$ ), so: Scheinbarer no-effect
    - B2b) Falls  $ber=0$  (wegen  $ges \neq ber$  gilt  $ges \neq 0$ ), so: Schein-Effekt
  - B3) Falls  $ges \cdot ber > 0$  (wegen  $ges \neq ber$  nur  $|ges| > |ber|$  oder  $|ges| < |ber|$  möglich)
    - B3a) Falls  $|ges| > |ber|$ , so: Überschneidung
    - B3b) Falls  $|ber| > |ges|$ , so: Suppressor-Phänomen

h) Wegen der Gleichung  $R^2 = \sum_{i=1}^k \beta_i r_{y,x_i}$  ließe sich  $R^2$  leicht

zerlegen in

$$\sum_{\beta_i \neq 0} r_{y,x_i}^2 : \text{Übereinstimmung}$$

$$\sum_{\beta_i r_{y,x_i} < 0} \beta_i r_{y,x_i} : \text{Distorter (negativer Summand)}$$

$$\sum_{\substack{\beta_i r_{y,x_i} > 0 \\ |\beta_i| < |r_{y,x_i}|}} \beta_i r_{y,x_i} : \text{Überschneidung}$$

$$\frac{\beta_i r_{y,x_i}}{\beta_i r_{y,x_i} > 0} : \text{Suppressor}$$

$$|\beta_i| > |r_{y,x_i}|$$

Solange es jedoch keine Rechtfertigung für eine Interpretation von  $\beta_i r_{y,x_i}$  gibt, ist diese Zerlegung nur von eingeschränkter Bedeutung. Sie zeigt rein rechnerisch, wie Multiple  $R^2$  zustande kommt.

#### Beispiel 1

Die dargestellten Argumente sollen nun auf ein Beispiel angewendet werden: Die Konzepte der "gewerkschaftlichen Orientierung" (repräsentiert durch die Jusos) und der "Alternativ-Orientierung" (repräsentiert durch die Basisgruppen (BG)) sollten durch verschiedene Indikatoren charakterisiert werden. Mit den Indikatoren, die sich im t-Test als diskriminierend erwiesen, als Prädiktoren wurde eine multiple Regression durchgeführt, wobei die Juso-BG-Wahldichotomie die abhängige Variable ist. (Die Problematik einer abhängigen Dichotomie soll hier ausgeklammert bleiben.)

Bei der Fragestellung, die (nominale) Wahl auf (metrische) Indikatoren zurückzuführen, wird man gleich an die Diskriminanzanalyse denken. Im Fall von zwei Gruppen sind die Koeffizienten  $a_i$  der (einzigen, weil  $\min \{k-1, p\} = 1$  für  $k=2$ ) Diskriminanzfunktion  $d = a_1 x_1 + \dots + a_p x_p$  aber proportional zu den  $\beta_i$  der Regression der Dummy-Variablen der Wahlentscheidung auf die Prädiktoren:

$$d = \frac{\hat{y}}{s_{\hat{y}}}, \quad s_{\hat{y}} = \text{Multiple } R, \text{ also:}$$

$$a_i = \frac{\beta_i}{R}$$

Juso-Wahl (vs. BG-Wahl)

(= Abhängige Variable)

Prädiktoren	Gesamt- effekt r	Gesamt- erklärungs- kraft r <sup>2</sup>	Effekt von x <sub>i</sub> - $\hat{x}_i$ , d.h. Beta	F	Erklärungs- beitrag von x <sub>i</sub> - $\hat{x}_i$ , d.h. r <sup>2</sup> <sub>y, x<sub>i</sub>-<math>\hat{x}_i</math></sub> = R <sup>2</sup> -R <sup>2</sup> <sub>(i)</sub>	R <sup>2</sup> bei schritt- weisem Vorgehen	R <sup>2</sup>
Bunte Liste positiv	-0,56	31 %	-0,38	49,3	12 %	31 %	31 %
SPD positiv	0,48	23 %	0,31	37,4	9 %	47 %	16 %
Für Frauenbe- rufstätigkeit	0,24	6 %	0,37	53,3	13 %	57 %	10 %
Für Wohngemein- schaft	0,39	15 %	-0,21	15,1	4 %	65 %	8 %
Verdienst wich- tig bei Berufs- entscheidung	0,29	8 %	0,27	28,6	7 %	71 %	6 %
Vater gewerk- schaftsnah	0,28	8 %	0,23	22,3	5 %	77 %	6 %
Kommunistischer Bund positiv	-0,46	21 %	-0,22	17,2	4 %	82 %	5 %
Berufsaussich- ten positiv	0,40	16 %	0,20	14,7	4 %	85 %	3 %

$$r_{y, x_i - \hat{x}_i}^2 = \frac{r_{y, x_i - \hat{x}_i}^2}{(1 - R^2)} / (n - k - 1), \text{ berechnet man: } r_{y, x_i - \hat{x}_i}^2 = F \cdot \frac{1 - 0,85362}{61}$$

Deshalb reicht es, eine multiple Regression durchzuführen. Die Anzahl der Prädiktoren einer schrittweisen Regression läßt sich mit Hilfe einer graphischen Darstellung des Erklärungszuwachses (wie in der Faktorenanalyse gebräuchlich) bestimmen. (Eine Alternative wäre, so viele Prädiktoren zu berücksichtigen, bis im nächsten Schritt ein Effekt nicht mehr signifikant ist.) Nach 8 Schritten ließ sich  $R^2 = 85,4\%$  der Varianz erklären. Der Gesamt-F-Wert lag mit 44,5 weit über dem kritischen Wert  $F_{8,61} \approx 2,1$ . (Die Stichprobe von 324 Studenten wird dadurch drastisch reduziert, daß nur die Teilgruppen BG- und Juso-Wähler berücksichtigt werden.) Die einzelnen Variablen liefern alle einen nach dem F-Test (kritischer Wert  $F_{1,61} \approx 4$ ) signifikanten Erklärungszuwachs.

1) Bunte Liste positiv

Typ B3a, nämlich:  $ges \neq ber$ ,  $ges.ber > 0$ ,  $|ges| > |ber|$   
Also: Überschneidung

Es soll nun genauer dargestellt werden, warum der bereinigte Effekt kleiner ist.

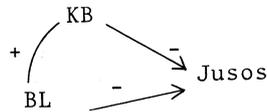
$r_{y,x_1}$

-0,56

$\beta_1$	$r_{x_1x_2}\beta_2 =$	$r_{x_1x_3}\beta_3 =$	$r_{x_1x_4}\beta_4 =$	$r_{x_1x_5}\beta_5 =$	$r_{x_1x_6}\beta_6 =$	$r_{x_1x_7}\beta_7 =$	$r_{x_1x_8}\beta_8 =$	$\Sigma =$
=-0,38	-.15.31 =-.05	=.15.37 =.06	.23(-.21) =-.05	-.08.27 =-.02	-.01.23 =-.002	.36(-.22) =-.08	-.17.20 =-.03	-0,56
BL	SPD	FR.	WG	VE.	VG.	KB	BE.	

Die positiven "Pfade" ergeben .06, die negativen -.25, weshalb aus dem bereinigter Effekt -0,38 ein Gesamteffekt von -0,56 ( $= -0,38 + 0,056 - 0,232$ ) wird. Am stärksten ins Gewicht fallen die negativen Pfade über KB, SPD, WG.

Graphisch z.B.:



Die KB-Variable ist ähnlich zu der BL-Variablen, so daß bei Kontrolle der KB-Variablen der Effekt der BL-Variablen geringer wird.

2) SPD positiv

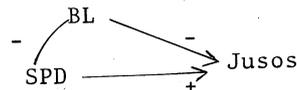
Typ B3a, also Überschneidung.

$r_{y,x_2}$

0,48

$\beta_2 =$	$r_{x_2x_1}\beta_1 =$	$r_{x_2x_3}\beta_3 =$	$r_{x_2x_4}\beta_4 =$	$r_{x_2x_5}\beta_5 =$	$r_{x_2x_6}\beta_6 =$	$r_{x_2x_7}\beta_7 =$	$r_{x_2x_8}\beta_8 =$	$\Sigma =$
0,31	-.15(-.38) =.06	.03.37 =.01	-.11(-.21) =.02	.02.27 =.005	.06.23 =.01	-.14(-.22) =.03	.17.20 =.03	0,48
SPD	BL	FR.	WG	VE.	VG.	KB	BE.	

Alle korrelierten Pfade sind positiv, sodaß aus dem bereinigten Effekt 0,31 ein Gesamteffekt von 0,48 wird (= 0,31 + 0,17). Am stärksten ins Gewicht fällt der Pfad über die BL.



Die bunte Liste hängt negativ mit SPD und Jusos zusammen, so daß ein Herausrechnen der BL den Effekt der SPD auf die Jusos reduziert.

3) Für Frauenberufstätigkeit

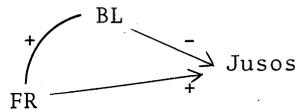
Typ B3b, nämlich: ges  $\neq$  ber, ges.ber > 0, |ber| > |ges|

Also: Suppressor-Phänomen

$r_{y,x_3}$								0,24
$\beta_3 =$	$r_{x_3x_1}\beta_1 =$	$r_{x_3x_2}\beta_2 =$	$r_{x_3x_4}\beta_4 =$	$r_{x_3x_5}\beta_5 =$	$r_{x_3x_6}\beta_6 =$	$r_{x_3x_7}\beta_7 =$	$r_{x_3x_8}\beta_8 =$	$\Sigma =$
0,37	.15(-.38) =-.06	.03.31 =.009	.18(-.21) =-.04	-.17.27 =-.05	.09.23 =.02	.11(-.22) =-.02	.0.20 =0	0,24
FR!	BL	SPD	WG	VE.	VG.	KB	BE.	

Die positiven Pfade ergeben nur 0,0300, die negativen dagegen -0,1649, sodaß sich aus dem bereinigten Effekt 0,37 ein geringerer Gesamteffekt 0,24 (= 0,37 + 0,0300 - 0,1649) ergibt. Am stärksten ins Gewicht fallen die negativen Pfade über BL, VE, WG.

Graphisch z.B.:



Anhänger der Bunten Liste befürworten die Frauenberufstätigkeit, lehnen die Jusos aber eher ab. Kontrolliert man die Einschätzung der BL, so wird ein stärkerer Zusammenhang zwischen der Befürwortung der Jusos und der Frauenberufstätigkeit sichtbar.

4) Alle anderen Effekte sind vom Typ der Überschneidung, so daß dies hier nicht genauer dargestellt werden soll.

Wie das Beispiel zeigt, dürften Überschneidungen am häufigsten sein, ferner Suppressor-Phänomene, schließlich Distorter-Phänomene, wenn sie auch in diesem Beispiel nicht auftraten. Perfekte Übereinstimmung oder perfektes Verschwinden von ges oder ber dürfte selten sein (Die Abweichung von den perfekten Typen wäre mit einem Signifikanztest zu behandeln.).

Da in dem Beispiel Überschneidungen überwiegen, muß gelten:

$$R^2 < \sum r_{y,x_i}^2$$

$$R^2 > \sum r_{y,x_i-\hat{x}_i}^2$$

	(a) $r_{y,x_i} \beta_i$	$a \lesseqgtr b$	(b) $r_{y,x_i}^2$	$a \lesseqgtr c$	(c) $r_{y,x_i-\hat{x}_i}^2$
Überschneidung	0,21	<	0,31	>	0,12
Überschneidung	0,15	<	0,23	>	0,09
Suppressor	0,09	>	0,06	<	0,13
Überschneidung	0,08	<	0,15	>	0,04
Überschneidung	0,078	<	0,084	>	0,07
Überschneidung	0,06	<	0,08	>	0,05
Überschneidung	0,10	<	0,21	>	0,04
Überschneidung	0,08	<	0,16	>	0,04
	$\Sigma = R^2$ = 0,85		$\Sigma = 1,28$		$\Sigma = 0,58$

$$\sum (r_{y,x_i}^2 - r_{y,x_i} \beta_i) = 0,46$$

Überschneidung

$$\sum (r_{y,x_i} \beta_i - r_{y,x_i}^2) = 0,03$$

Suppressor

Also: Multiple  $R^2 = 0,85 < \sum r_{y,x_i}^2 = 1,28 (=0,85 + 0,46 - 0,03)$

$$\sum (r_{y,x_i} \beta_i - r_{y,x_i-\hat{x}_i}^2) = 0,31$$

Überschneidung

$$\sum (r_{y,x_i-\hat{x}_i}^2 - r_{y,x_i} \beta_i) = 0,04$$

Suppressor

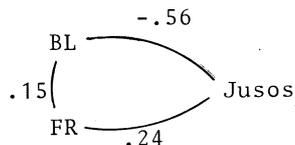
Also: Multiple  $R^2 = 0,85 > \sum r_{y,x_i-\hat{x}_i}^2 = 0,58 (=0,85 - 0,31 + 0,04)$

### Beispiel 2

Es soll nun gezeigt werden, daß:  $\sum r_{y,x_i}^2 < R^2 < \sum r_{y,x_i-\hat{x}_i}^2$  möglich ist.

Dazu beschränke ich mich einfach auf das Suppressor-Phänomen im vorigen Beispiel.

Korrelationen



Die Beta-Koeffizienten ergeben sich dann wegen

$$\beta_{yx} = \frac{r_{yx} - r_{xT}r_{yT}}{1 - r_{xT}^2}$$

als:  $\beta_1 = -0,61$        $\beta_2 = 0,33$

$$\text{Multiple } R^2 = \sum \beta_i r_{yx_i} = 0,42$$

$$\sum r_{y,x_i}^2 = 0,37$$

$$r_{y,x_1-\hat{x}_1}^2 = R^2 - r_{y,x_2}^2 = 0,36$$

$$r_{y,x_2-\hat{x}_2}^2 = R^2 - r_{y,x_1}^2 = 0,11$$

Also:  $\sum r_{y,x_i-\hat{x}_i}^2 = 0,47$

Insgesamt:  $\sum r_{y,x_i}^2 = 0,37 < R^2 = 0,42 < \sum r_{y,x_i-\hat{x}_i}^2 = 0,47$

(Ferner gilt im Fall von 3 Variablen (nicht aber allgemein, wie etwa Beispiel 1 zeigt):  $\Delta_1 = \Delta_2$

Denn:  $\sum \beta_i s_{\hat{y}(i),x_i}^2 (= \sum \beta_i s_{y,x_i}^2) = \beta_1 r_{yT} r_{xT} + \beta_2 r_{yx} r_{xT}$

(Zur 1. Gleichung:  $\beta_i r_{y,x_i}^2 = \beta_i s_{y,x_i-\hat{x}_i}^2 + \beta_i s_{y,\hat{x}_i}^2 =$   
 $r_{y,x_i-\hat{x}_i}^2 + \beta_i s_{y,\hat{x}_i}^2$ )

Und:  $\Delta_2 = r_{yx} r_{xT} \beta_2 + r_{yT} r_{xT} \beta_1$

Hier:  $\Delta_1 = \Delta_2 = 0,05$

### Beispiel 3

Es soll nun ein Beispiel dargestellt werden, bei dem Distorter Phänomene auftreten.

Die Einschätzung der Bunten Liste soll statistisch auf die Einschätzungen von Hochschulgruppen zurückgeführt werden.

Abhängige Variable Bunte Liste (Multiple  $R^2 = 0,22807$ )

Prädiktoren      Gesamteffekt  $r$       Effekt von  $x_i - \hat{x}_i$ , d.h. Beta      F

Prädiktoren	Gesamteffekt $r$	Effekt von $x_i - \hat{x}_i$ , d.h. Beta	F
SB	0,37	0,26	8,52
KB	0,36	0,23	7,95
BG	0,32	0,13	2,74
MSB	0,08	-0,15	2,96
RCDS	-0,21	-0,09	1,15
Jusos	0,08	-0,07	0,67
SHB	0,15	0,05	0,24
LHV	-0,01	0,03	0,17
SLH	-0,05	0,01	0,03

Daraus lassen sich berechnen:  $r_{y, x_i - \hat{x}_i}^2 = F \cdot \frac{1 - 0,22807}{171}$

( $n - k - 1 = 171$ )

(a)                      (b)                      (c)  
 $a \lesseqgtr b$                        $a \lesseqgtr c$   
 $r_{y, x_i} \beta_i$                        $r_{y, x_i}^2$                        $\bar{r}_{y, x_i - x_i}^2$

Überschneidung	0,10	<	0,14	>	0,04
Überschneidung	0,08	<	0,13	>	0,04
Überschneidung	0,04	<	0,10	>	0,01
Distorter	-0,01	<	0,01	<	0,01
Überschneidung	0,02	<	0,04	>	0,01
Distorter	-0,01	<	0,01	<	0,00
Überschneidung	0,01	<	0,02	>	0,00
Distorter	-0,00	<	0,00	<	0,00
Distorter	-0,00	<	0,00	<	0,00

$\Sigma = R^2$   
 $= 0,23$

$\Sigma = 0,45$

$\Sigma = 0,11$

$$\Sigma (r_{y,x_i}^2 - r_{y,x_i} \beta_i) = 0,18$$

Überschneidung

$$\Sigma (r_{y,x_i}^2 - r_{y,x_i} \beta_i) = 0,04$$

Distorter

$$\text{Also: Multiple } R^2 = 0,23 < \Sigma r_{y,x_i}^2 = 0,45 (= 0,23 + 0,18 + 0,04)$$

$$\Sigma (r_{y,x_i} \beta_i - r_{y,x_i}^{\hat{x}_i}) = 0,15$$

Überschneidung

$$\Sigma (r_{y,x_i}^{\hat{x}_i} - r_{y,x_i} \beta_i) = 0,03$$

Distorter

$$\text{Also: Multiple } R^2 = 0,23 > \Sigma r_{y,x_i}^{\hat{x}_i} = 0,11 (= 0,23 - 0,15 + 0,03)$$

Distorter-Phänomene reduzieren Multiple  $R^2$  sowohl verglichen mit  $\Sigma r_{y,x_i}^2$  als auch verglichen mit  $\Sigma r_{y,x_i}^{\hat{x}_i}$ , weil sie zu negativen Termen  $r_{y,x_j} \beta_i$  führen. In dem Beispiel dominieren aber die Überschneidungen.

i) Für Überschneidungen (d.h. o.B.d.A.  $r_{y,x_i} > \beta_i > 0$ ) gilt:

$$r_{y,x_i} \beta_i > r_{y,x_i}^{\hat{x}_i}$$

(Selbstverständlich gilt:  $r_{y,x_i}^2 > r_{y,x_i} \beta_i$ )

┌

$$\text{Denn: } s_{y,x_i} > \frac{s_{y,x_i}^{\hat{x}_i}}{2} \Rightarrow$$

$$s_{x_i}^{\hat{x}_i}$$

$$(*) \quad s_{y, x_i - \hat{x}_i} < s_{y, x_i} \quad s_{x_i - \hat{x}_i}^2 \leq s_{y, x_i}, \text{ weil:}$$

$$s_{x_i - \hat{x}_i}^2 = \frac{1}{n} \langle x_i - \hat{x}_i, x_i - \hat{x}_i \rangle = \frac{1}{n} \langle x_i - \hat{x}_i, x_i \rangle$$

$$= 1 - R_{x_i, x_j}^2 (j \neq i) \leq 1$$

$$\text{Also: } r_{y, x_i - \hat{x}_i}^2 = \frac{s_{y, x_i - \hat{x}_i}^2}{s_{x_i - \hat{x}_i}^2}$$

$$= \beta_i s_{y, x_i - \hat{x}_i}$$

$$< \beta_i s_{y, x_i} \quad (\text{nach } (*) \text{ und weil } \beta_i > 0)$$

└

Für Suppressor-Phänomene (o.B.d.A.  $\beta_i > r_{y, x_i} > 0$ ) gilt das Entsprechende nicht:

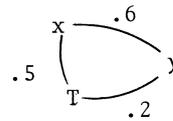
$$\text{I.a. nicht: } r_{y, x_i} \beta_i < r_{y, x_i - \hat{x}_i}^2$$

$$\text{Noch nicht einmal: } r_{y, x_i}^2 < r_{y, x_i - \hat{x}_i}^2$$

(Selbstverständlich gilt:  $r_{y, x_i}^2 < r_{y, x_i} \beta_i$ )

Beispiel:

Korrelationen



$$\beta_1 = \frac{.6 - .5 \cdot .2}{1 - .5^2} = 0,67$$

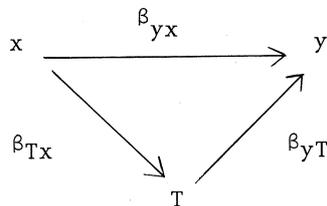
$$\beta_2 = \frac{.2 - .5 \cdot .6}{1 - .5^2} = -0,13$$

	r	$\beta$	(a) r $\beta$	$a \lesseqgtr b$	(b) $r^2$	$a \lesseqgtr c$	(c) $r^2_{y, x_i - \hat{x}_i}$
(Suppressor)	.6	.67	.402	>	.36	>	.336
(Distorter)	.2	-.13	-.026		.04		.016
			$\Sigma = R^2$ = .376				

Dies zeigt, daß Aussagen über Effekte und Aussagen über Erklärungskraft unabhängig voneinander variieren können:  
Obwohl ein Effekt bei Kontrolle größer wird, muß der Erklärungsanteil bei Kontrolle nicht wachsen.

#### j) Vorzeichenregel

Im Fall von 3 Variablen gibt es eine Vorzeichenregel analog zu Davis' Regel für die Tabellenanalyse:



Gesamteffekt von x auf y:  $r_{yx}$

Bereinigter Effekt von x auf y:  $\beta_{yx}$

Rest-Effekt:  $\beta_{yT} \beta_{Tx}$

Die Vorzeichenregel bezieht sich auf die Beziehung der Störvariablen T zu den Variablen x und y.

- a) Falls der Rest-Effekt kleiner Null ist, so haben die Beta-Koeffizienten von T zu X bzw. Y unterschiedliche Vorzeichen.

$$\text{Formal: } 0 > \beta_{YT} \beta_{TX} \Rightarrow \text{sign}(\beta_{YT}) \neq \text{sign}(\beta_{TX})$$

- b) Falls der Rest-Effekt größer Null ist, so haben die Beta-Koeffizienten von T zu X bzw. Y das gleiche Vorzeichen.

$$\text{Formal: } 0 < \text{res} = \beta_{YT} \beta_{TX} \Rightarrow \text{sign}(\beta_{YT}) = \text{sign}(\beta_{TX})$$

Bei der Darstellung der Typen ist jeweils angegeben worden, ob  $\text{res} \gtrless 0$ , damit die Information zur Vorzeichenregel in der Darstellung der Typen enthalten ist.

Im Fall von mehr als 3 Variablen hat man zunächst die zusammenfassende Information, ob der Rest-Effekt größer oder kleiner Null ist. Wie dieser Rest-Effekt zustande kommt, kann dann in der Art analysiert werden, wie oben an Beispiel 1 demonstriert.

Hinweis: Die Vorzeichenregel gilt nicht für die einfachen Korrelationskoeffizienten statt der Beta-Koeffizienten:

Das Beispiel in (i) liefert:

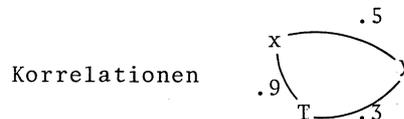
Die Relation (X,Y) ist Suppressor-Phänomen.

$$\text{Also: } \beta_{YT} \beta_{TX} < 0 \quad (\text{da } r_{YX} > 0)$$

$$\text{Aber: } r_{YT} r_{TX} > 0$$

Würde man eine Typologie auf der Basis von  $\text{ber} = \text{Part Correlation}$  oder  $\text{ber} = \text{Partial Correlation}$  erstellen, so würde man auch keine entsprechende Vorzeichenregel für die einfachen Korrelationskoeffizienten erhalten:

Beispiel:



$$r_{XY.T} = \frac{.5 - .9 \cdot .3}{\sqrt{1-.9^2} \sqrt{1-.3^2}} = 0,55$$

$$r_{Y, X-\hat{X}(T)} = \frac{.5 - .9 \cdot .3}{\sqrt{1-.9^2}} = .53$$

Unter beiden Voraussetzungen handelt es sich um ein Suppressor-Phänomen mit  $res > 0$ . Also  $res < 0$ .

Aber:  $r_{XT} \cdot r_{YT} > 0$

$$\begin{aligned} \text{(Dies liegt daran, daß: } r_{XY} &= \sqrt{1-r_{XT}^2} \sqrt{1-r_{YT}^2} r_{XY.T} + r_{XT} r_{YT}, \\ r_{XY} &= \sqrt{1-r_{XT}^2} r_{Y, X-\hat{X}(T)} + r_{XT} r_{YT}, \end{aligned}$$

d.h. die Gewichtung von  $ber$  ist jeweils zu berücksichtigen.)  
D.h. von den naheliegenden Koeffizienten liefert keiner eine Vorzeichenregel für die einfachen Korrelationskoeffizienten, wie sie der Intuition in Analogie zur Tabellenanalyse entspricht. Die auf der Basis von Beta erstellte Typologie leistet dies für die Beta.

Die Kovarianz liefert eine Vorzeichenregel für die einfachen Korrelationskoeffizienten:

$$\begin{aligned} r_{XY} = s_{XY} = s_{x-\hat{x}(T), y-\hat{y}(T)} + s_{\hat{x}(T), \hat{y}(T)} &= s_{XY.T} + s_{XT} s_{YT} = \\ &= s_{XY.T} + r_{XT} r_{YT} \end{aligned}$$

Oder:

$$r_{XY} = s_{y, x-\hat{x}(T)} + r_{XT} r_{YT}$$

$s_{XY.T}$  soll als Partial Covariance,  $s_{y, x-\hat{x}(T)}$  als Part Covariance bezeichnet werden. Wählt man als bereinigten Effekt Partial oder Part Covariance, so folgt unmittelbar:

$$(res > 0) \iff (r_{XT} r_{YT} > 0)$$

Es soll nun untersucht werden, ob diese Konzepte für mehr als zwei Prädiktoren günstige Darstellungen liefern.

Es gilt wegen der Orthogonalität der Residuen zu den Prädiktoren:

$$\langle y, x_i \rangle - \langle y, \hat{x}_i \rangle = \langle y, x_i - \hat{x}_i \rangle = \langle y - \hat{y}_{(i)}, x_i - \hat{x}_i \rangle =$$

$$\langle y - \hat{y}_{(i)}, x_i \rangle = \langle y, x_i \rangle - \langle \hat{y}_{(i)}, x_i \rangle$$

D.h. daß Partial und Part Covariance die gleichen Zerlegungen liefern.

$$1. \text{ Zerlegung: } r_{y, x_i} = s_{y, x_i - \hat{x}_i} + s_{y, \hat{x}_i}$$

$$= s_{y, x_i - \hat{x}_i} + s_y \sum_{j \neq i} \beta_{x_i \cdot x_j} x_j$$

$$= s_{y, x_i - \hat{x}_i} + \sum_{j \neq i} \beta_{x_i \cdot x_j} r_{y, x_j}$$

Diese Zerlegung ist deshalb nicht günstig, weil hier der statistische Einfluß von  $y$  auf  $x_i$  formuliert wird, man sich aber für die andere Richtung interessiert.

$$2. \text{ Zerlegung: } \langle x_i, x_j \rangle = \langle \hat{x}_i, x_j \rangle \text{ für } j \neq i, \text{ also:}$$

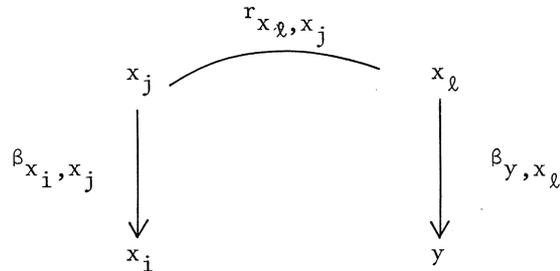
$$\langle x_i, \hat{y}_{(i)} \rangle = \langle \hat{x}_i, \hat{y}_{(i)} \rangle$$

Deshalb:

$$r_{y, x_i} = s_{y, x_i - \hat{x}_i} + s_{\hat{x}_i, \hat{y}_{(i)}}$$

$$= s_{y, x_i - \hat{x}_i} + s \sum_{j \neq i} \beta_{x_i \cdot x_j} x_j, \sum_{\ell \neq i} \beta_{y, x_\ell} x_\ell$$

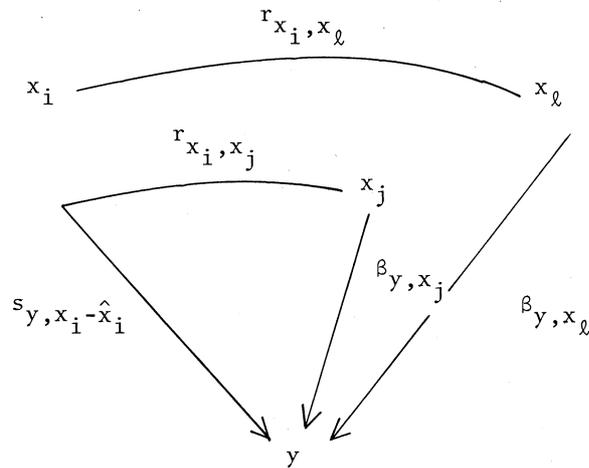
$$= s_{y, x_i - \hat{x}_i} + \sum_{\substack{j, \ell \\ j, \ell \neq i}} \beta_{y, x_\ell} r_{x_\ell, x_j} \beta_{x_i \cdot x_j}$$



Diese Zerlegung ist zwar anschaulich, benötigt aber zwei Regressionen.

$$\begin{aligned}
 3. \text{ Zerlegung: } r_{y, x_i} &= s_{y, x_i - \hat{x}_i} + s_{\hat{y}(i), x_i} \\
 &= s_{y, x_i - \hat{x}_i} + s \sum_{\substack{j \\ j \neq i}} \beta_{y, x_j} x_{j, x_i} \\
 &= s_{y, x_i - \hat{x}_i} + \sum_{\substack{j \\ j \neq i}} \beta_{y, x_j} r_{x_j, x_i}
 \end{aligned}$$

Graphisch:



Verglichen mit der Zerlegung für die Beta-Koeffizienten

$$(r_{y,x_i} = \beta_{y \cdot x_i} + \sum_{\substack{j \\ j \neq i}} r_{x_i, x_j} \beta_{y \cdot x_j})$$

hat man hierbei die Nachteile:

1) Die Zerlegung ist zwar anschaulich, benötigt aber 3 Konzepte: Kovarianz, Beta, Korrelation.

(Bei der Verwendung von Beta braucht man nur 2 Konzepte: Beta und die Korrelation.)

Ferner sind die Beta-Koeffizienten nicht die Koeffizienten der Ausgangsregression ( $y$  auf  $x_1, \dots, x_k$ ), sondern der Regression von  $y$  auf  $\{x_1, \dots, x_k\} - \{x_i\}$ . Wegen letzterem handelt es sich bei der Betrachtung verschiedener Prädiktoren auch um verschiedene Regressionen.

#### Schlußfolgerung:

Die Verwendung von Part (oder Partial) Correlation als Bereinigung liefert nicht die Vorzeichenregel für die einfachen Korrelationskoeffizienten. Die Verwendung von Part (oder Partial) Covariance liefert zwar die Vorzeichenregel für die einfachen Korrelationskoeffizienten, aber im Falle von mehr als zwei Prädiktoren ist die Zerlegung in bereinigten und Rest-Effekt weniger günstig als bei den Beta-Koeffizienten. Verwendet man als bereinigten Effekt die Beta-Koeffizienten, so erhält man die Vorzeichenregel für die Beta-Koeffizienten, was ja im Hinblick auf Konsistenz durchaus überzeugend ist, auch wenn in der Tabellenanalyse üblicherweise die Vorzeichenregel für die einfachen Korrelationskoeffizienten formuliert wird, was - wie gezeigt - bei Verwendung der Kovarianz möglich ist. Von den betrachteten Möglichkeiten scheint mir die pfadanalytische Interpretation der Zerlegung des Gesamteffektes in einen bereinigten/direkten und korrelierte/indirekte Effekte noch der günstigste Lösungsversuch zur Erhellung der Korrelationsstruktur bei multivariaten Problemstellungen.

B) Nicht-rekursive Ansätze

Die Pfadanalyse ist ein rekursiver Ansatz: Sie setzt eine schwache kausale Ordnung ( $x_1 \leq \dots \leq x_k$ ) voraus derart, daß ein kausaler Einfluß auf  $x_i$  nur von  $x_j$  ( $j < i$ ) ausgehen darf. Es sind also keine Wechselwirkungen möglich. Dies leisten nicht-rekursive Ansätze, wobei zu unterscheiden ist, ob jede Gleichung einzeln geschätzt wird ("limited information", z.B. durch two stage least squares, wobei Instrumentalvariablen verwendet werden) oder alle Gleichungen gemeinsam ("full information", z.B. durch three stage least squares).

C) Nicht-rekursive Ansätze mit latenten Konstrukten (LISREL)

Die ML-Schätzung ist die asymptotisch beste Schätzung, unter dem Aspekt der Schätzung ist also ein Ansatz überlegen, wenn er mit der ML-Schätzung arbeitet. Dies ist bei Jöreskogs Ansatz LISREL (Linear structural relationships) der Fall, mit dem man nicht-rekursive Modelle behandeln kann. Auf die anderen nicht-rekursiven Ansätze soll deshalb gar nicht weiter eingegangen werden. Ferner ist Jöreskogs Ansatz ein Angebot an die allgemeinen Soziologen, da die Unterscheidung in theoretische Konstrukte und Operationalisierung durch Indikatoren Bestandteil des Modells ist.

$$\text{Meßmodell: } \begin{cases} y = \Lambda_y \eta + \varepsilon; \eta = \text{abhängige theoretische Konstrukte} \\ x = \Lambda_x \xi + \delta; \xi = \text{unabhängige theoretische Konstrukte} \end{cases}$$

(y,x sind die Indikatoren)

$$\text{Strukturmodell für die latenten Konstrukte } \begin{cases} \eta = B \eta + \Gamma \xi + \zeta \end{cases}$$

Jöreskog leistet also eine Synthese aus den faktoranalytischen Meßmodellen der Psychometrie und den Strukturgleichungsmodellen der Ökonometrie. (Die Lösung besteht letztlich in der Zurückführung des Problems auf eine Faktorenanalyse mit Restriktionen.)

Anmerkung: In der Berechnung der ML-Schätzung von LISREL ist ein kleiner Fehler enthalten, der sich bisher in allen Darstellungen fortpflanzt:

Die Likelihood-Funktion lautet, wenn  $S =$

$$\frac{1}{N-1} (Z'Z - N\bar{Z}\bar{Z}') \text{ (wie in LISREL V, S. I.25):}$$

$$\ln L(\mu, \Sigma) = -\frac{n}{2} \ln |\Sigma| - \frac{n-1}{2} \text{Spur}(\Sigma^{-1}) + \text{constans}$$

Den Faktor  $\frac{n}{2}$  ignorieren, wie Jöreskog dies macht, darf man nur, wenn man mit der ML-Schätzung  $\tilde{S} = \frac{1}{N} (Z'Z - N\bar{Z}\bar{Z}')$  arbeitet (die nicht erwartungstreu (unbiased) ist wie  $S$ ).

Jöreskog minimiert:  $\ln |\Sigma| + \text{Spur}(\Sigma^{-1}) + \text{constans}$ . Dies ist nur korrekt mit  $\tilde{S}$  statt  $S$ .

(Oder: Man benutzt  $S$  und die oben angegebene Gewichtung mit  $n$  bzw.  $n - 1$ .)

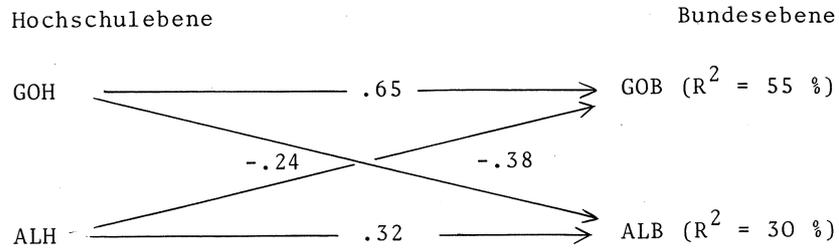
#### Data-Fitting versus Modellüberprüfung

Ein Hauptvorteil von Jöreskogs Ansatz besteht darin, daß die Faktorenanalyse nicht nur exploratorisch, sondern konfirmatorisch verwendet wird. Dadurch ist ein Modelltest möglich. Dieser wesentliche Unterschied soll nun kurz an einem Beispiel dargestellt werden.

a) Pfadanalyse mit Indizes:

GO = Gewerkschaftliche Orientierung

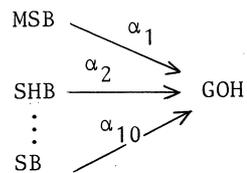
AL = Alternativ - Orientierung



$F_{2,178;0,05} = 3,05$ , also beide Regressionen signifikant.

$F_{1,78;0,05} = 3,9$ , also alle 4 Effekte signifikant.

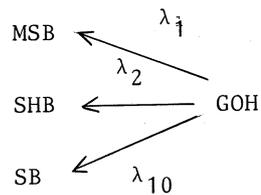
Ein Index ist eine Definition:



$$GOH = \alpha_1 MSB + \alpha_2 SHB + \dots + \alpha_{10} SB$$

Eine (Nominal-) Definition ist natürlich nicht wahr oder falsch, sondern mehr oder weniger zweckmäßig.

b) Ein (konfirmatorisches) faktorenanalytisches Modell dagegen ist falsifizierbar:



$$MSB = \lambda_1 GOH + \delta_1$$

$$SHB = \lambda_2 GOH + \delta_2 \quad \text{etc.}$$



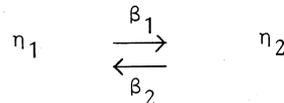
$$R_{YY} = \frac{1}{N} Y Y' = \Lambda \left( \frac{1}{N} F F' \right) \Lambda' = \Lambda \Lambda'$$

Also wird bei LISREL die empirische Kovarianz/Korrelationsmatrix mit der unter den Modellannahmen zu erwartenden Kovarianz/Korrelationsmatrix verglichen.

Anmerkung: Wie Hoppe (1981) zurecht bemerkt, läßt sich die Existenz von latenten Variablen nicht beweisen - dies müßte die Struktur eines Gottesbeweises haben. Aber diese Behauptung braucht man auch gar nicht: Zwar sind verschiedene Kausalmodelle mit vorliegenden Daten verträglich, aber es gibt Kriterien, um zu beurteilen, welches Modell besonders gut mit den Daten verträglich ist, in LISREL ist insbesondere jeweils ein Gesamttest für die Abweichung der Daten vom Modell möglich.

#### Nicht-rekursive Effekte

In LISREL sind nicht-rekursive Beziehungen zwischen theoretischen Konstrukten  $\eta_i$  modellierbar.



Nach dem Vorschlag von Graff und Schmidt, der in LISREL aufgenommen wurde, kann man von folgender Vorstellung ausgehen: Eine Änderung von  $\eta_1$  um eine Einheit induziert eine Änderung von  $\eta_2$  um  $\beta_1$  und deshalb eine Änderung von  $\eta_1$  um  $\beta_1 \beta_2$ . Im nächsten Zyklus ist die Änderung  $(\beta_1 \beta_2)^2$  etc. Also ist der Gesamteffekt:

$$\sum_{i=1}^{\infty} (\beta_1 \beta_2)^i = \frac{1}{1 - (\beta_1 \beta_2)} = \frac{\beta_1 \beta_2}{1 - \beta_1 \beta_2} ,$$

falls  $|\beta_1 \ \beta_2| < 1$ , denn sonst gilt die erste Gleichung nicht. Da Beta-Koeffizienten größer als 1 sein dürfen, muß also Vorsorge getroffen werden, daß im Fall  $|\beta_1 \ \beta_2| \geq 1$  die Gesamteffekte nicht derart berechnet werden.

(Die Verallgemeinerung in LISREL besteht in:

$$\eta = B\eta + \Gamma\xi + \zeta$$

Gesamteffekt von  $\eta$  auf sich selbst ist dann:

$$\sum_{i=1}^{\infty} B^i = (E - B)^{-1} - E, \text{ wobei}$$

E die Einheitsmatrix ist. Diese Gleichung gilt nur, wenn

$$\lim_{i \rightarrow \infty} B^i = 0.)$$

Bei LISREL (S. III 39) wird auch ein direkter Effekt von  $\xi$  auf  $y$  angegeben, nämlich:  $\Lambda_y \Gamma$

$$\xi \xrightarrow{\Gamma} \eta \xrightarrow{\Lambda_y} y$$

Das Pfeildiagramm zeigt, daß die Bezeichnung direkter Effekt hier nicht angemessen ist.

#### Einige Probleme von LISREL

Ich halte Jörgeskogs Ansatz für den im Augenblick besten Versuch zur Modellierung metrischer Daten. D.h. natürlich nicht, daß nicht einiges verbesserungswürdig ist, weshalb nun einige Probleme diskutiert werden sollen.

1) Ein großer Fortschritt von LISREL gegenüber der herkömmlichen Pfadanalyse besteht darin, daß ein Test des Gesamtmodells verfügbar ist, so daß im Prinzip das Problem der Modellüberprüfung durch LISREL gelöst ist. Dabei gibt es

jedoch mehrere Probleme, die im folgenden diskutiert werden sollen.

a) Problematik des  $\chi^2$  - Test bei der Modellüberprüfung

Sowohl bei Modellen für metrische als auch bei Modellen für nicht-metrische Daten verwendet man einen  $\chi^2$  - Test zur Überprüfung des Gesamtmodells. Während man aber bei einem klassischen Test wie z.B. dem  $\chi^2$ -Unabhängigkeitstest in Kontingenztafeln versucht, die Nullhypothese der Unabhängigkeit zu falsifizieren (mit i. a. 95 % Sicherheit, d. h. 5 % Irrtumswahrscheinlichkeit), so daß man die Nullhypothese nur bei 5 % der Stichproben fälschlicherweise ablehnt (Fehler 1. Art), versucht man bei dem Modelltest eher, das Modell zu "verifizieren": Die Nullhypothese besteht darin, daß das Modell zutrifft, die Alternativhypothese bei LISREL darin, daß keine Restriktion für die Kovarianzmatrix vorliegt. Das Modell wird beibehalten, so lange der  $\chi^2$  - Wert nicht in den kritischen 5 % - Bereich fällt. Damit hat man eine gute Chance, ein Modell beizubehalten, obwohl es falsch ist. Dieser Fehler 2. Art wird aber bisher allgemein ignoriert. Das Problem, daß man die Nullhypothese "verifizieren" will, würde z.B. vermieden durch einen Test, bei dem die Ablehnung der Nullhypothese der Annahme des Modells entspricht. Der Test, den Bentler und Bonett (1980) vorschlagen, hat diese Eigenschaft (die Nullhypothese lautet, es bestehe keine Abhängigkeit außer den fixierten Abhängigkeitswerten). Die andere Lösung besteht darin, daß der Test bei dem Modell, das einen Parameter zu wenig enthält, der klassischen Vorgehensweise entspricht: Die Ablehnung der üblichen Nullhypothese der Unabhängigkeit in Kontingenztafeln (mit zwei Merkmalen) bedeutet ja, daß man die Interaktion der beiden Merkmale berücksichtigen muß. Wird die Nullhypothese zurückgewiesen, so ist das korrekte Ergebnis dieses klassischen Tests, daß das saturierte Modell erforderlich ist (d. h. die Randverteilungseffekte plus der Interaktionseffekte). (Der Fehler 2. Art müßte außerdem berücksichtigt werden.)

- b) Es ergeben sich ferner Probleme bei dem  $\chi^2$  - Test, falls die Daten zu stark von dem Modell der Normalverteilung abweichen (vgl. Geweke und Singleton 1980 zu Abweichungen von den Regularitätsbedingungen). Ferner hat die Stichprobengröße einen starken Einfluß auf das Testergebnis (zu großen Stichproben vgl. Bentler und Bonett 1980 und Bagozzi 1981, zu kleinen Stichproben vgl. Geweke und Singleton 1980, Sprott 1980, Bearden et al. 1982 und Boomsma 1982). Jöreskog und Sörbom (1982) warnen selbst, daß der Test nicht angemessen ist für eine explorative Vorgehensweise, d. h. z.B. nicht für die übliche Modellsuche (model fitting). Bei dem Test wird die Hypothese geprüft, ob die Kovarianzmatrix  $\Sigma$  von der Form ist, die das Modell impliziert, gegen die Alternativthese, daß die Kovarianzmatrix keiner Bedingung unterliegt. Wenn eine solche These nicht vorliegt, ist der Test nicht adäquat. Stattdessen sollte man dann  $\chi^2/df$  als Maßzahl für das Gelingen des model-fitting verwenden. Ein Vorzug des Tests besteht darin, daß Vergleiche möglich sind zwischen Modellen, deren Parametermenge in einer Teilmengenrelation zueinander stehen. Dies läßt sich testen mit  $\chi^2$ , wobei die Differenz der Freiheitsgrade der beiden Modelle den neuen Freiheitsgrad ergibt, dies ist die Differenz der Parameteranzahl. Speziell mit  $\chi^2_1$  läßt sich also die Signifikanz eines einzelnen Parameters testen. - Bei dem Modellvergleich geht die Anzahl der beobachteten Kovarianzen nicht mehr in die Freiheitsgrade ein.

c) Model-Fitting

Model-Fitting ist eine Synthese von klassischem Hypothesentest und reiner Deskription. Während beim klassischen Hypothesentest nur einmal entschieden wird, ob Daten mit einem Modell verträglich sind, und bei einer Deskription i. a. nur eine Version formuliert wird, besteht Model-Fitting in dem schrittweise Ändern des Modells, um die Verträglichkeit des resultierenden Modells mit den Daten zu erhöhen. Ich halte die Vorgehensweise für sehr sinn-

voll, wobei man aber nicht vergessen darf, daß dies ein ad hoc-Verfahren ist: An den Daten, an denen man das Modell gewinnt, kann man es nicht falsifizieren, denn das Modell ist ja schrittweise so geändert worden, daß es möglichst gut zu den Daten paßt. (Ein Ausweg wäre z. B. split half der Daten, an einer Hälfte läßt sich das Modell generieren, an der anderen überprüfen.)

In LISREL V wird durch modification indices das Modell-Fitting erleichtert, weil diese Indizes angeben, in welchem Ausmaß  $\chi^2$  sinkt, wenn man einen Parameter frei setzt (statt wie zuvor fixiert). Dies ließe sich leicht als Einladung zum Mißbrauch kritisieren, etwa in der Art von Mc Phersons (1976) Kritik des "Theory-Trimming". Mir scheint die Verwendung der Modellsuche jedoch fruchtbarer als die Kritik, wenn man den ad hoc-Charakter stets in Erinnerung behält. Empirische Wissenschaft besteht ja nicht nur darin, aus vorliegenden Behauptungen aufgrund der Empirie die richtigen herauszufiltern, sondern nicht zuletzt darin, fruchtbare Thesen/Modelle aufzustellen.

Bei einem schrittweise Vorgehen führt man mehrere Tests durch, wovon das Signifikanzniveau nicht unberührt bleibt: Bei jeweiliger Verwendung eines 5 %-Signifikanzniveaus kann man nicht davon ausgehen, daß insgesamt ein 5 %-Signifikanzniveau resultiert. Die Anzahl der Tests läßt sich aber wie folgt zur Korrektur des Signifikanzniveaus verwenden: Wenn man k Tests durchführt mit den Fehlern 1. Art gleich  $\alpha_i$ , so gilt für den Fehler 1. Art  $\alpha_G$  der Gruppe von Tests, eine der Hypothesen abzulehnen, obwohl sie richtig ist, die Ungleichung (mit A = Ablehnungsbereich, t = Testwert):

$$\alpha_G = P\left(\bigcup_{i=1}^k \{t_i \in A_{\alpha_i}\}\right) \leq \sum_{i=1}^k P\{t_i \in A_{\alpha_i}\} = \sum_{i=1}^k \alpha_i$$

Will man also 5 Tests durchführen und insgesamt nur einen Fehler 1. Art von 5 % tolerieren, so kann man die 5 Tests jeweils auf dem 1 %-Niveau durchführen, um dies zu erreichen (vgl. z. B. Bielby und Kluegel 1977). (Ferner müßte auch noch der Fehler 2. Art berücksichtigt werden.)

## 2) Standardisierung:

Ein weiteres Problem ist die Fixierung der Skalen in LISREL. Gerbing et al. (1982) argumentieren, daß die Fixierung der Skala einer latenten unabhängigen Variablen durch Festsetzung einer Ladung auf 1 eine willkürliche Metrik ist, von der dann die meisten anderen Schätzungen abhängen. Dieses Problem läßt sich vermeiden, indem man die Varianzen der latenten unabhängigen Variablen auf 1 fixiert (um die Vergleichbarkeit der Strukturparameter zu erhöhen). Gerbing et al. finden dann auch die "standardisierte Lösung" von LISREL besonders vorteilhaft, die darin besteht, daß auch die latenten abhängigen Variablen noch standardisiert werden. Was nach meinem Wissen in der Literatur noch nicht diskutiert wird, ist der Punkt, daß dies keine echte standardisierte Lösung ist: Zunächst sind die  $\eta_i$  abhängige Variable, deren Varianz nicht einfach wie bei den unabhängigen  $\xi_j$  auf 1 fixiert werden kann (Die Varianzen der  $\eta_i$  sind keine Modellparameter, die man fixieren kann.) In der multiplen Regression erreicht man durch Standardisierung von  $x_j$  auch nicht, daß  $\hat{y}$  standardisiert ist. Sondern:  $s_{\hat{y}} = \text{Multiple } R$ . Zur Standardisierung von  $\hat{y}$  braucht man eine Schätzung unter der Nebenbedingung  $s_{\hat{y}} = 1$ . Wenn man die Likelihood-Funktion in Abhängigkeit u. a. von  $\eta_i$  maximiert und für die Lösung anschließend eine Standardisierung vornimmt, so ist dies i. a. nicht identisch mit dem Maximum der Likelihood-Funktion unter der Nebenbedingung, daß die  $\eta_i$  standardisiert sind. Da eine Maximierung unter Nebenbedingungen aufwendiger ist, hat Jöreskog wohl eine "praktische Annäherung" an die richtige Vorgehensweise gewählt.

Jöreskogs Ansatz als spezielle Faktorenanalyse 2. Ordnung

Jöreskogs Ansatz läßt sich darstellen als spezielle Faktorenanalyse 2. Ordnung:

$$\begin{pmatrix} Y \\ X \end{pmatrix} = \Lambda \begin{pmatrix} \eta \\ \xi \end{pmatrix} + \begin{pmatrix} \varepsilon \\ \delta \end{pmatrix}, \text{ wobei: } \Lambda = \begin{pmatrix} \Lambda_Y & 0 \\ 0 & \Lambda_X \end{pmatrix}$$

Also sind  $\begin{pmatrix} \eta \\ \xi \end{pmatrix}$  die Faktoren 1. Ordnung.

Ferner war:  $B_\eta = \Gamma\xi + \zeta$ , also:  $\eta = B^{-1}\Gamma\xi + B^{-1}\Gamma$

$$\begin{pmatrix} \eta \\ \xi \end{pmatrix} = \kappa \begin{pmatrix} \zeta \\ \xi \end{pmatrix}, \text{ wobei: } \kappa = \begin{pmatrix} B^{-1} & B^{-1}\Gamma \\ 0 & E \end{pmatrix}$$

$\begin{pmatrix} \zeta \\ \xi \end{pmatrix}$  sind also spezielle Faktoren zweiter Ordnung (wobei es keinen Fehlerterm gibt und  $\xi$  nicht weiter zurückgeführt wird auf Faktoren 2. Ordnung).

Für die Kovarianzmatrix von  $z := \begin{pmatrix} Y \\ X \end{pmatrix}$  gilt dann:

$$\begin{aligned} S_{zz} &= E(z z') \\ &= \Lambda E \left( \begin{pmatrix} \eta \\ \xi \end{pmatrix} \begin{pmatrix} \eta \\ \xi \end{pmatrix}' \right) \Lambda' + \begin{pmatrix} \theta_\varepsilon & 0 \\ 0 & \theta_\delta \end{pmatrix} \\ &= \Lambda K \begin{pmatrix} \psi & 0 \\ 0 & \phi \end{pmatrix} K' \Lambda' + \begin{pmatrix} \theta_\varepsilon & 0 \\ 0 & \theta_\delta \end{pmatrix} \end{aligned}$$

Bentler (1976), Bentler und Weeks (1979, 1980) und Mc Donald (1978, 1979) entwickeln allgemeinere Modelle, die Faktoren beliebiger Ordnung zulassen. Dies scheint mir jedoch eine bloß formale Verallgemeinerung, während Jöreskogs Ansatz eine klare wissenschaftstheoretische Konzeption hat.

D) Einige Parallelen zwischen Modellen für metrische und Modellen für nicht-metrische Daten

- 1) Das Konzept der bedingten statistischen Unabhängigkeit liefert einen gemeinsamen Rahmen für metrische und nicht-metrische Daten.

Beispiel: Intervenierende Variable

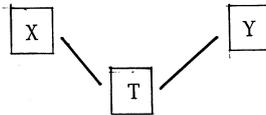
(z.B. X = Familienstand von Frauen, T = Umfang der Hausarbeit, Y = Fernbleiben vom Betrieb)

X und Y sind unabhängig unter Bedingung T. Im metrischen Fall heißt dies, daß der partielle Korrelationskoeffizient verschwindet:  $r_{XY.T} = 0$

Im nicht-metrischen Fall heißt dies, daß der Zusammenhang von X und Y in den Teilgruppen auf der Basis von T verschwindet:

$$[XY: T_1] = [XY: T_2] = 0$$

Verwendet man statt der klassischen Teilgruppenanalyse (Tabellenanalyse) die Parametrisierung durch ein log-lineares Modell, so heißt dies, daß nur die "Cliques" (X,T) (Y,T) des ungerichteten Graphen zu berücksichtigen sind:



X und Y hängen nur über T zusammen.

Eine ausführliche Diskussion der Parallelen von metrischen und nicht-metrischen Modellen auf der Basis des Konzepts der bedingten statistischen Unabhängigkeit findet sich bei Wermuth (1976 a,b, 1978, 1980) und bei Kiiveri und Speed (1982).

- 2) Als Verteilungsannahme geht man von der multivariaten Normalverteilung und der Multinomialverteilung aus, die beide zur multivariaten exponentiellen Familie gehören (zur Verwendung von letzterem vgl. Dempster 1971, 1972, Dempster et al. 1977).

- 3) Als Schätzverfahren verwendet man in beiden Fällen i.a. das bewährte Maximum Likelihood Prinzip.
- 4) In beiden Fällen gibt es einen Gesamttest für das Modell mit Hilfe von  $\chi^2_{n-p}$ , wobei  $n$  die Anzahl der Beobachtungswerte ist (im metrischen Fall die Anzahl der beobachteten Kovarianzen, im nicht-metrischen Fall die Anzahl der Subpopulationen, multipliziert mit  $k - 1$  (wobei  $k =$  Anzahl der Kategorien der abhängigen Variablen), also die Anzahl der - nicht redundanten - beobachteten Anteilswerte) und  $p$  die Anzahl der zu schätzenden Parameter.  
Die Probleme des Modelltests ("Verifizieren der Nullhypothese", Signifikanzniveau bei mehreren Tests) wurden oben bereits dargestellt, die Ausführungen zu LISREL gelten gleichlautend für den Modelltest bei nicht-metrischen Daten. Nur die Verteilungsannahme ist bei Modellen für nicht-metrische Daten unproblematischer.
- 5) Modellsuche: Ein Modell soll eine knappe Darstellung für die Daten liefern und nicht signifikant von den Daten abweichen (was natürlich gegenläufige Bedingungen sind). Im nicht-metrischen Fall hat man mit dem saturierten Modell, das nur eine Umformung der Daten ist, einen natürlichen Ausgangspunkt bei der Modellsuche. Es werden schrittweise Effekte weggelassen, bis das Modell nicht mehr zu den Daten paßt. Das letzte Modell vorher paßt noch zu den Daten und ist besonders sparsam (weil besonders wenige Effekte zugelassen werden). - Im metrischen Fall gibt es keinen solchen natürlichen Ausgangspunkt für die Modellsuche. Aber nachdem man von einem speziellen Modell ausgegangen ist, kann man die Darstellung knapper machen, indem man weitere Restriktionen einführt (Parameter auf einen bestimmten Wert fixieren (z.B. Effekte = 0, Varianzen = 1) oder die Gleichheit von Parametern festlegen), wobei die Anpassung an die Daten natürlich nur geringer werden kann. Umgekehrt kann man das Modell durch Aufnahme weiterer Parameter (frei setzen statt wie zuvor zu fixieren) abschwächen, um eine bessere Anpassung an die Daten zu erreichen.

Bibliographie

- Aigner, D.H. / A.S. Goldberger (Hrsg.) (1977), Latent variables in socio-economic models. Amsterdam, North-Holland Publishing Co.
- Bagozzi, R.P. (1981), Evaluating Structural Equation Models with Unobservable Variables and Measurement Error: A Comment. In: Journal of Marketing Research 18, 375-381
- Bearden, W.O. et al. (1982), Sample Size Effects on Chi Square and Other Statistics Used in Evaluating Causal Models. In: Journal of Marketing Research XIX, 425-430
- Bentler, P.M. (1976), Multistrukture statistical model applied to factor analysis. In: Multivariate Behavioral Research 11, 3-25
- (1978), The interdependence of theory, methodology, and empirical data: Causal modeling as an approach to construct validation. In: D.B. Kandel (Hrsg.): Longitudinal Research on drug use. New York: Wiley
  - (1980), Multivariate analysis with latent variables: Causal models. In: Annual Review of Psychology 31, 419-456
  - / D.G. Bonett (1980), Significance Tests and Goodness of Fit in the Analysis of Covariance Structures. In: Psychological Bulletin 88, 588-606
  - / D.G. Weeks (1979), Interrelations among models for the analysis of moment structures. In: Multivariate Behavioral Research 14, 169-186
  - / D.G. Weeks (1980), Linear Structural Equations with latent Variables. In: Psychometrika 45, 289-308
- Bielby, W.T. / J.R. Kluegel (1977), Statistical Inference and statistical power in applications of the general linear model. In: D.R. Heise (Hrsg.): Sociological Methodology. San Francisco: Jossey-Bass
- Boomsma, A. (1982), The robustness of LISREL against small sample sizes in factor analysis models. In: Jöreskog/Wold
- Dempster, A.P. (1971), An overview of multivariate analysis. In: Journal of Multivariate Analysis 1, 316-346
- (1972), Covariance selection. In: Biometrics 28, 157-175
  - et al. (1977), Maximum Likelihood from incomplete data via the EM algorithm. In: Journal of the Royal Statistical Society, Series B, 39, 1-21

- Gerbing, D.W. et al. (1982), The Metric of the latent variables in a LISREL IV Analysis. In: Educational and Psychological Measurement 42, 423-427
- Geweke, J.F. et al. (1980), Interpreting the Likelihood Ratio Statistic in Factor Models when Sample Size is Small. In: Journal of the American Statistical Association 75, 133-137
- Goldberger, A.S. / O.D. Duncan (Hrsg.) (1973), Structural equation models in the social sciences. New York: Seminar Press
- Graff, J. / P. Schmidt (1982), A general model for decomposition of effects. In: Jöreskog/Wold
- Holtmann, D. (1980), Interpretationsprobleme und Verwendung der Residuen in der multiplen Regression. Zeitschrift für Soziologie 9, 384-389
- Hoppe, H.H. (1981), Über die Verwendung ungemessener Variablen in Kausalmodellen. Eine epistemologische Kritik. In: Zeitschrift für Soziologie 10, 307-318
- Hummell, H.J. / R. Ziegler (1976), Korrelation und Kausalität. 2 Bde. Stuttgart: Enke
- Jöreskog, K.G. (1973), A general method for estimating a linear structural equation system.. In: Goldberger/Duncan
- / D. Sörbom (1979), Advances in factor analysis and structural equation models. Cambridge, Mass.: Abt Books
  - / D. Sörbom (1981), Analysis of linear structural relationships by maximum likelihood and least squares methods. LISREL V. University of Uppsala
  - / D. Sörbom (1982), Recent developments in structural equation modelling. In: Journal of Marketing Research XIX, 404-416
  - / H. Wold (1982), Systems under indirect observation: Causality and prediction. Amsterdam: North Holland
- Kenny, D.A. (1979), Correlation and Causality. New York: Wiley
- Kiiveri, H. / T.P. Speed (1982), Structural analysis of multivariate data: A review. In: S. Leinhardt (Hrsg.), Sociological Methodology. San Francisco: Jossey-Bass, 209-289
- McDonald, R.P. (1978), A simple comprehensive model for the analysis of covariance structures. In: Br. J. Math.Stat. Psychol. 31, 59-72
- (1979), The structural analysis of multivariate data: A sketch of a general theory. In: Multivariate Behavioral Research 14. 21-38

- McPherson. J.M. (1976). Theory trimming. In: Social Science Research 5, 21-38
- Opp. K.-D. / P. Schmidt (1976), Einführung in die Mehrvariablenanalyse. Hamburg: Rowohlt
- Sprott, D.A. (1980), Maximum Likelihood in Small Samples: Estimation in the Presence of Nuisance Parameters. In: Biometrika 67, 515-523
- Wermuth, N. (1976a), Analogies between multiplicative models in contingency tables and covariance selection. In: Biometrics 32, 95-108
- (1976b), Model search among multiplicative models. In: Biometrika 32, 253-263
  - (1978), Zusammenhangsanalysen medizinischer Daten. Berlin: Springer
  - (1980), Linear recursive equations, covariance selection and path analysis. In: Journal of the American Statistical Society, Series C (Applied Statistics), 26, 963-972
- Wright, S. (1934), The method of path coefficients. In: Ann. Math. Stat. 5, 161-215
- (1960), Path coefficients and path regressions: alternative or complementary concepts? In: Biometrics 16, 189-202