G. Rohwer, April 2010

Seminar: Statistische Mehrebenenmodelle

Leitfaden für die Diskussion

Gliederung

1. Konzeptionelle und statistische Mehrebenenmodelle.

Man kann *konzeptionelle* und *statistische* Mehrebenenmodelle unterscheiden (vgl. Hummell 1972; Eeden und Hüttner 1982). Konzeptionelle Mehrebenenmodelle bemühen sich um einen begrifflichen Rahmen, mit dessen Hilfe sinnvoll von mehreren Ebenen und ihren Beziehungen gesprochen werden kann (vgl. z.B. Baumgartner et al. 1976; Huinink 1989, 1995). Im Unterschied zu statistischen Mehrebenenmodellen implizieren sie nicht unmittelbar eine bestimmte Methode der statistischen Datenverarbeitung und Modellbildung. Statistische Mehrebenenmodelle können demgegenüber als Versionen von Regressionsmodellen betrachtet werden, wie sie in der empirischen Sozialforschung bereits seit langem verwendet werden. Im Folgenden beschränken wir uns auf eine Diskussion statistischer Mehrebenenmodelle.

Die Grundidee, von der statistische Mehrebenenmodelle fast immer ausgehen, wurde von Jan de Leeuw (1992, S. xiii) so formuliert:

> „In the social sciences, data structures are often hierarchical in the following sense: We have variables describing individuals, but the individuals also are grouped into larger units, each unit consisting of a number of individuals. We also have variables describing these higher order units.“

Hiervon ausgehend kann man unter statistischen Mehrebenenmodellen Regressionsmodelle verstehen, bei denen sich die Regressorvariablen auf zwei oder mehr unterschiedliche Ebenen beziehen.

2. Ebenenbildung durch Partitionen

Es ist wichtig, zu verstehen, wie in diesem Zusammenhang von „Ebenen“ gesprochen wird. Ebenen entstehen dadurch, dass eine Referenzmenge (Population) in Teilgesamtheiten partitioniert wird. Die Elemente der Referenzmenge werden Elemente der ersten Ebene (Individuen) genannt, die daraus gebildeten Teilgesamtheiten sind die Elemente der zweiten Ebene. Dieser Prozess der Ebenenbildung kann natürlich fortgesetzt werden.

Bei der Bildung von Teilgesamtheiten kann man sich an beliebigen Merkmalen der Individuen bzw. ihrer Umgebungen orientieren. Tatsächlich impliziert jede statistische Variable eine Ebenenbildung im eben definierten Sinn. Zur Erläuterung sei zunächst kurz erklärt, wie im Folgenden von statistischen Variablen gesprochen wird. Gemeint sind Funktionen der Form

$$X : \Omega \longrightarrow \tilde{\mathcal{X}}$$

Hierbei ist Ω eine endliche Referenzmenge (Population oder Stichprobe) und $\tilde{\mathcal{X}}$ ein Merkmalsraum, d.h. irgendeine Zusammenfassung möglicher Attribute für die Elemente von Ω (die Individuen). Die statistische Variable X ist somit formal eine Funktion, die jedem Element ω der Referenzmenge Ω ein durch $X(\omega)$ bezeichnetes Element des Merkmalsraums (einen

Merkmalswert) zuordnet.

Offenbar induziert nun jede statistische Variable auch eine Partition ihrer Referenzmenge, wobei die Teilgesamtheiten jeweils Individuen zusammenfassen, die bei dieser Variablen den gleichen Wert aufweisen. Bezieht man sich z.B. auf das Geschlecht, induziert diese Variable eine Partition in Männer und Frauen.

3. Einfache lineare Regressionsmodelle

Wir beginnen mit einem einfachen linearen Regressionsmodell:

$$Y = \alpha + X\beta + \epsilon \tag{1}$$

Y ist die abhängige Variable, X ist eine (möglicherweise mehrdimensionale) unabhängige Variable. Hat man Daten (x_i, y_i) für $i = 1, \dots, n$, können Schätzwerte $\hat{\alpha}$ und $\hat{\beta}$ der Modellparameter berechnet werden.

Jetzt nehmen wir an, dass sich die Individuen, auf die sich das Modell (1) bezieht, in unterschiedlichen Kontexten befinden. Die Kontexte repräsentieren wir formal durch $\mathcal{K}_1, \dots, \mathcal{K}_m$. Um diese Kontexte im Modell zu berücksichtigen, gibt es zwei Möglichkeiten:

a) Man kann Dummy-Variablen $D_1, \dots, D_m$ verwenden, durch die für jedes Individuum erfasst wird, welchem Kontext es angehört.

b) Man kann die Kontexte durch weitere Variablen charakterisieren und dann diese Variablen zusätzlich in das Modell aufnehmen. Handelt es sich bei den Kontexten z.B. um Schulklassen, könnte man beispielsweise Variablen für die Klassengröße oder für bestimmte Eigenschaften des Lehrpersonals oder des Unterrichts verwenden.

Wir beginnen mit der Variante (b) und nehmen an, dass zur Charakterisierung der Kontexte eine (ggf. wiederum mehrdimensionale) Variable Z verwendet wird.

4. Interaktionseffekte und separate Modelle

Da man nicht voraussetzen kann, dass X und Z voneinander unabhängige Bedingungen für die abhängige Variable Y sind, ist es sinnvoll, mit der Möglichkeit von Interaktionseffekten zu rechnen. Der neue Modellansatz sieht dann folgendermaßen aus:

$$Y = \alpha_0 + X\beta_x + Z\beta_z + XZ\beta_{xz} + \eta \tag{2}$$

Daraus gewinnt man auch sofort für jeden möglichen Wert z der Variablen Z ein separates Regressionsmodell

$$Y = (\alpha_0 + z\beta_z) + X(\beta_x + z\beta_{xz}) + \eta_z \tag{3}$$

Hat man hinreichend viele Daten, um die Parameter des Modells (2) zu schätzen, können daraus sogleich auch Schätzwerte der separaten Modelle

berechnet werden. (Ggf. können auch separate Modelle separat geschätzt werden; man kann jedoch i.a. nicht davon ausgehen, zu den gleichen Schätzwerten für die Parameter zu gelangen.)

5. Abhängigkeit der Parameter vom Kontext

Als ein wichtiger Grund für die Verwendung von Mehrebenenmodellen wird in der Literatur oft das Argument angeführt, dass man durch sie Einsichten gewinnt, wie Modellparameter von einem Kontext (von Werten der Kontextvariablen) abhängen.[1] Bei Mason, Wong und Entwisle (1984, S. 74f.) findet man folgende Formulierung:

> „Our fundamental assumption is that the micro values of the response variable in some way depend on context and that the *effects* of the micro determinants *may* vary systematically as a function of context.“

Ähnlich heißt es bei Blien, Wiedenbeck und Arminger (1994, S. 270):

> „In the analysis of a two level structure we are mainly interested in the variation of the regression coefficients across groups.“

Bezieht man sich auf das Modell (2), richtet sich die Frage darauf, wie der Zusammenhang zwischen X und Y seinerseits vom durch Z gegebenen Kontext abhängig ist. Zur Präzisierung der Frage kann man eine Funktion

$$x, z \longrightarrow \mathrm{E}(Y|X = x, Z = z)$$

betrachten, die jeweils bestimmten Werten der Variablen X und Z einen durch sie bedingten Mittel- bzw. Erwartungswert der abhängigen Variablen Y zuordnet. Die Frage bezieht sich dann darauf, wie die partielle Ableitung $\partial\mathrm{E}(Y|X = x, Z = z)/\partial x$ von z abhängt. Für das Modell (2) findet man

$$\partial\mathrm{E}(Y|X = x, Z = z)/\partial x \;=\; \beta_x + z\beta_{xz} \tag{4}$$

also eine lineare Abhängigkeit von den Werten von Z.

6. Ein Beispiel mit fiktiven Daten

Um die Modellansätze zu illustrieren, betrachten wir ein einfaches Beispiel mit fiktiven Daten. Es gibt 20 abhängig Beschäftigte, die in vier Unternehmen $(\mathcal{K}_1, \mathcal{K}_2, \mathcal{K}_3, \mathcal{K}_4)$ beschäftigt sind. X erfasst ihr Alter, Y ihren monatlichen Verdienst, Z die Größe des Unternehmens, in dem die Person beschäftigt ist. Box 1 zeigt die fiktiven Daten, Box 2 zeigt Parameterschätzwerte für die Modelle (1) und (2) sowie für zwei weitere Modellvarianten, die mit dem OLS-Verfahren berechnet wurden.

[1] Man vgl. zum Beispiel Bryk und Raudenbush 1992, S. 6; Goldstein 1995, S. 17; Engel 1998, S. 73.

Box 1 Fiktive Daten für das Beispiel 1.

Y	X	K	Z	Y	X	K	Z	Y	X	K	Z	Y	X	K	Z
3000	25	1	20	2500	30	2	50	2400	25	3	100	2000	20	4	120
3200	30	1	20	2800	30	2	50	3000	30	3	100	3000	30	4	120
3300	40	1	20	2600	35	2	50	3200	30	3	100	4000	40	4	120
3500	45	1	20	3000	40	2	50	3500	40	3	100	3500	40	4	120
3700	50	1	20	3500	45	2	50	4000	45	3	100	4000	45	4	120

Box 2 OLS-Parameterschätzungen mit den Daten aus Box 1.

Modell	(1)	(*)	(*)	(2)
Interzept	1233.21	3125.50	961.58	2258.31
X	54.60		56.73	21.67
Z		0.82	2.70	-14.88
XZ				0.48

Mithilfe der Formulierung (3) kann man aus den Angaben in Box 2 kontextspezifische Parameter für den Zusammenhang zwischen Alter und Höhe des Erwerbseinkommens berechnen. Man findet:

z	$\beta_x + z\beta_{xz}$
20	31.36
50	45.90
100	70.13
120	79.82

In diesem Beispiel hängt also der Zusammenhang zwischen Alter und Verdiensthöhe auch vom Kontext (der Unternehmensgröße) ab.

7. Unterschiedliche Arten von Variablen?

Offenbar ist (2) ein gewöhnliches Regressionsmodell. Man könnte zwar von einem „Mehrebenenmodell“ sprechen, weil es auf Kontextvariablen Bezug nimmt. Dabei sollte jedoch bedacht werden, dass eine Unterscheidung zwischen „individuell zurechenbaren“ und „Kontextvariablen“ nicht nur obskur ist, denn Werte der Kontextvariablen werden hier zur Charakterisierung von Individuen verwendet, sondern im Rahmen statistischer Begriffsbildungen gar nicht durchgeführt werden kann. In formaler Hinsicht unterscheiden sich die Variablen X und Z in keinerlei Weise.

8. Verwendung von Dummy-Variablen

Jetzt betrachten wir die Variante (a) aus Abschnitt 3, bei der unterschiedliche Kontexte durch Dummy-Variablen $D_1, \dots, D_m$ erfasst werden. Als

Box 3 OLS-Parameterschätzungen des Modells (5) mit den Daten aus Box 1.

Interzept	1165.93
X	57.25
D1	1210.82
D2	-212.99
D3	-288.15
D4	-709.68
XD1	-31.91
XD2	-3.72
XD3	11.63
XD4	24.00

eine zu (2) analoge Formulierung erhält man dann den Modellansatz

$$Y = \alpha_0 + X\beta_x + \sum\nolimits_j D_j\delta_j + \sum\nolimits_j XD_j\delta_{x,j} + \eta \tag{5}$$

Dabei wird, um einen einfachen Vergleich zu ermöglichen, angenommen, dass die Parameter für die Dummy-Variablen als Abweichungen vom globalen Mittelwert α_0 definiert sind; es ist somit nicht erforderlich, eine Referenzkategorie festzulegen und auszulassen.[2]

Wiederum kann man der Frage nachgehen, wie der Zusammenhang zwischen X und Y vom Kontext abhängt. Analog zu (4) erhält man jetzt

$$\partial \mathrm{E}(Y|X=x)/\partial x \;=\; \beta_x + \sum\nolimits_j D_j\delta_{x,j} \tag{6}$$

wodurch deutlich wird, wie der Zusammenhang zwischen X und Y von den Werten der den Kontext erfassenden Dummy-Variablen abhängt.

Die durch (6) definierten Größen entsprechen gerade den Parametern kontextspezifischer Regressionsmodelle. Aus (5) gewinnt man nämlich für jeden Kontext j ein spezifisches Regressionsmodell

$$Y_j = (\alpha_j + \delta_j) + X_j(\beta_x + \delta_{x,j}) + \eta_j \tag{7}$$

wobei X_j und Y_j die Einschränkungen von X und Y für den Kontext j, d.h. die ihm entsprechende Referenzmenge, bezeichnen.

[2] Es sei angemerkt, dass sich eine vollständige formale Analogie zwischen (5) und (2) herstellen lässt. Man kann nämlich Variablen D und D_x konstruieren, die die Werte der Parameter δ_j bzw. $\delta_{x,j}$ erfassen (gehört ein Individuum dem Kontext $\mathcal{K}_j$ an, hat D den Wert δ_j und D_x den Wert $\delta_{x,j}$). Unter Verwendung dieser Variablen kann dann das Modell (5) in der Form

$$Y = \alpha_0 + X\beta_x + D + XD_x + \eta$$

geschrieben werden.

9. Fortsetzung des Beispiels mit Dummy-Variablen

Zur Illustration verwenden wir wieder das Beispiel 1, also die fiktiven Daten aus Box 1. Verwendet man das OLS-Verfahren, um mit diesen Daten die Parameter des Modells (5) zu schätzen, findet man die in Box 3 angegebenen Werte.[3] Wie die Abhängigkeit der Verdiensthöhe vom Alter mit dem Kontext variiert, wird jetzt durch $\beta_x + \delta_{x,j}$ (für den Kontext $j = 1, \ldots, 4$) angegeben. Die Werte können aus den Angaben in Box 3 unmittelbar berechnet werden, nämlich:

Kontext	$\beta_x + \delta_{x,j}$
1	25.34
2	53.53
3	68.88
4	81.25

Anders als bei der Verwendung des Modells (2) sind die Schätzwerte in diesem Fall mit denjenigen identisch, die man bei einer separaten Modellschätzung innerhalb der einzelnen Kontexte erhalten würde.

10. Interpretation von Kontexteffekten

Anhand dieses Beispiels kann auf einen wichtigen Unterschied zwischen den beiden in Abschnitt 3 unterschiedenen Methoden zur Definition von Kontextvariablen hingewiesen werden. Verwendet man Dummy-Variablen, gelangt man zwar möglicherweise zu der Einsicht, dass bestimmte Zusammenhänge (wie in unserem Beispiel der Zusammenhang zwischen Alter und Verdiensthöhe) in den verschiedenen Kontexten unterschiedlich ausgeprägt sind oder, wie man auch sagen könnte, bestimmte Effekte kontextabhängig sind. Aber da die Dummy-Variablen nur eine Partition der Individuen repräsentieren, liefern sie für sich genommen keinerlei inhaltliche Hinweise, an denen sich eine Interpretation orientieren könnte.

Wenn man dagegen die Kontexte durch inhaltlich bedeutsame Variablen charakterisiert, können deren Bedeutungen auch zur Interpretation kontextabhängiger Effekte verwendet werden. In unserem Beispiel könnte man einen Zusammenhang zur Unternehmensgröße herstellen.

Dann kann auch deutlich werden, dass Kontextabhängigkeit nicht einfach aus einer Aufteilung der Individuen auf unterschiedliche Kontexte resultiert, sondern aus jeweils bestimmten Eigenschaften dieser Kontexte. Oder anders formuliert: Relevant ist nicht die Partitionierung der Individuen durch die Kontexte im Sinne von Teilmengen der Individuengesamtheit, sondern ihre Partitionierung durch die Werte der die Kontexte charakterisierenden substantiellen Variablen. (Man denke z.B. bei unserem Beispiel an ein fünftes Unternehmen mit derselben Größe wie das Unternehmen Nr. 1, bei dem auch der Zusammenhang von Alter und Verdiensthöhe ähnlich ist wie im Unternehmen Nr. 1.)

[3] Um jeweils alle Dummy-Variablen verwenden zu können, wurden die Bedingungen $\Sigma_j \delta_j = 0$ und $\Sigma_j \delta_{x,j} = 0$ als Constraints verwendet.

11. Vorgehensweise bei einer Vielzahl von Kontexten

In unserem Beispiel gibt es nur vier unterschiedliche Kontexte. Wenn ihre Anzahl sehr groß ist – wie beispielsweise dann, wenn man Haushalte als Kontexte betrachtet –, kann es offenbar problematisch werden, ein Modell der Form (5) zu schätzen, das für jeden Kontext eine Dummy-Variable enthält. Aber warum sollte man in solchen Fällen überhaupt ein Modell dieser Art schätzen wollen?

Dass diese Frage berechtigt ist, zeigt auch die Überlegung des vorangegangenen Abschnitts. Denn selbst wenn es mit einem hinreichend großen Datensatz möglich wäre, das Modell zu schätzen, könnte man aus der großen Anzahl der dann resultierenden Modellparameter bestenfalls lernen, dass bestimmte (andere) Effekte kontextabhängig sind, zum Beispiel davon abhängen, in welchem Haushalt eine Person lebt. Aber um zu einer potentiell interessanten Interpretation zu gelangen, müssten die Kontexte inhaltlich charakterisiert werden; und dann wäre es auch sogleich möglich, diese inhaltlichen Charakterisierungen anstelle der die Kontexte repräsentierenden Dummy-Variablen in die Modelle aufzunehmen.

Wenn also die Anzahl unterschiedlicher Kontexte (im Sinne von Teilgesamtheiten einer Partitionierung der Referenzmenge) sehr groß ist, ist es meistens nicht sinnvoll, Modelle zu konzipieren, die Effekte anderer Variablen *von diesen Kontexten* abhängig machen.

12. Identifizierte und anonyme Kontexte

Folgende Unterscheidung liefert eine mögliche Sinnvoraussetzung für die Verwendung von Kontextdefinitionen durch Dummy-Variablen:

a) Einerseits kann man sich Anwendungen vorstellen, bei denen es sinnvoll möglich erscheint, die unterschiedlichen Kontexte inhaltlich zu identifizieren. Zum Beispiel Länder bei internationalen Vergleichen oder Berufe oder Produktionszweige.

b) Andererseits kann die Anzahl unterschiedlicher Kontexte so gross sein, dass eine Identifizierung der jeweils individuellen Kontexte nicht mehr sinnvoll erscheint. Zum Beispiel: Haushalte und Unternehmen.

Ungeachtet einer möglicherweise großen Anzahl unterschiedlicher Kontexte erscheint es im ersten Fall sinnvoll möglich, sie durch Dummy-Variablen zu repräsentieren, weil sich inhaltliche Interpretationen an eine Identifizierung der Kontexte anschließen lassen. Man denke zum Beispiel an einen Vergleich von Ländern oder Berufen.

13. Stochastische Regressionsmodelle

Außer den bisher besprochenen Modellansätzen werden in der Literatur auch Mehrebenenmodelle mit zufälligen Parametern diskutiert. In der englischsprachigen Literatur wird von „multilevel random coefficient models“ gesprochen, man vgl. etwa Mason, Wong und Entwisle (1984) oder De Leeuw und Kreft (1986); im Deutschen könnte man sie „statistische Mehrebenenmodelle mit zufälligen Parametern“ nennen. Wir sprechen im Folgenden kurz von MLRC-Modellen.

Im Unterschied zu Regressionsmodellen, die sich auf der Grundlage statistischer Variablen entwickeln lassen, knüpfen MLRC-Modelle an die Idee stochastischer Regressionsmodelle an, die sich auf Zufallsvariablen bzw. Zufallsgeneratoren beziehen. Zur Erläuterung beziehen wir uns nochmal auf das Modell (1), also

$$Y = \alpha + X\beta + \epsilon$$

Werden X und Y als statistische Variablen aufgefasst, bezieht sich das Modell auf die gemeinsame Verteilung der durch diese Variablen gegebenen Daten. Insbesondere ist dann die Residualvariable ϵ keine Zufallsvariable, sondern *entsteht* durch die Modellkonstruktion: $\epsilon := Y - \alpha - X\beta$.

Grundsätzlich anders verhält es sich bei stochastischen Regressionsmodellen, bei denen ϵ als eine Zufallsvariable aufgefasst wird. Zufallsvariablen sind durch einen „zufälligen Mechanismus“ definiert, durch den ihre Werte entstehen können. Bei einer konstruktiven Deutung kann man sich auf Zufallsgeneratoren beziehen. Ein stochastisches Regressionsmodell bezieht sich dann auf einen Zufallsgenerator für die Zufallsvariable ϵ. Dabei kann irgendeine Verteilung angenommen werden, die mehr oder weniger weitgehend spezifiziert sein kann(in jedem Fall wird normalerweise vorausgesetzt, dass der Mittelwert von ϵ Null ist).

Somit kann man sich gedanklich auf einen Mechanismus beziehen, durch den beliebig viele Realisationen von ϵ entstehen können. Insbesondere kann man für jeden möglichen Wert der unabhängigen Variablen X beliebig viele Realisationen der abhängigen Variablen Y erzeugen, die bei stochastischen Regressionsmodellen natürlich auch als eine Zufallsvariable aufzufassen ist.

14. Reale und fiktive Zufallsgeneratoren

Reale Zufallsgeneratoren sind realisierbare Verfahren, durch die Werte einer Zufallsvariablen mit einer bestimmten Verteilung erzeugt werden können. Die Verteilung kann (qua Konstruktion) bekannt sein, wie z.B. bei der Verwendung von Würfeln und Urnen, aber das ist keine notwendige Voraussetzung; wichtig ist nur, dass für den Zufallsgenerator irgendeine bestimmte Verteilung, die sich bei wiederholten Realisationen nicht verändert, angenommen werden kann.

Dementsprechend kann es oft sinnvoll sein, auch technische Geräte und

experimentelle Prozeduren als Zufallsgeneratoren zu betrachten. An diese Betrachtungsweise können ihre Charakterisierungen durch stochastische Regressionsmodelle anschließen.

Anders verhält es sich bei der Analyse sozialstatistischer Daten. Sie entstehen nicht durch Mechanismen, die als reale Zufallsgeneratoren betrachtet werden können. Wenn man für ihre Analyse gleichwohl stochastische Regressionsmodelle verwendet, werden die Daten gewissermaßen als Resultate fiktiver Zufallsgeneratoren, die zum Zweck der Datenanalyse ausgedacht worden sind, betrachtet.[4]

Betrachten wir zur Illustration einige mögliche stochastische Regressionsmodelle für das Beispiel aus Abschnitt 6.

a) $Y = \alpha + X\beta + \epsilon$. In diesem Fall wird angenommen, dass für jedes Individuum i ein Wert y_i der abhängigen Variablen dadurch zustande kommt, dass zunächst durch den Zufallsgenerator für ϵ eine Zufallszahl ϵ_i erzeugt wird, woraus dann $y_i = \alpha + x_i\beta + \epsilon_i$ entsteht.

b) $Y_j = \alpha_j + X_j\beta_j + \epsilon$, wobei der Index j auf die Unternehmen verweisen soll, in denen die Individuen beschäftigt sein können. In diesem Fall wird angenommen, dass es für jedes Unternehmen unterschiedliche Modellparameter geben kann, dass jedoch Werte der Residualvariablen ϵ für alle Unternehmen mit demselben Zufallsgenerator erzeugt werden.

c) $Y_j = \alpha_j + X_j\beta_j + \epsilon_j$. In diesem Fall gibt es außerdem für jedes Unternehmen einen eigenen Zufallsgenerator zur Erzeugung von Werten der Residualvariablen. Somit besteht auch die Möglichkeit anzunehmen, dass sich ihre Verteilungen unterscheiden.

Offenbar kann man sich viele weitere Möglichkeiten für fiktive Zufallsgeneratoren und darauf basierende stochastische Regressionsmodelle ausdenken.

15. Modelle mit zufälligen Parametern

Jetzt kann versucht werden, MLRC-Modelle zu erläutern. Dazu knüpfen wir an den in Abschnitt 8 besprochenen Modellansatz an, bei dem für jeden Kontext unterschiedliche Modellparameter angenommen werden (vgl. (7)). Wie bereits besprochen wurde, kann es schwierig oder unmöglich sein, diese Parameter zu schätzen, wenn es sehr viele unterschiedliche Kontexte gibt.

Eine mögliche Konsequenz, die in Abschnitt 11 besprochen wurde, besteht darin, die Kontexte nicht durch Dummy-Variablen, sondern durch inhaltlich interpretierbare Variablen zu unterscheiden. MLRC-Modelle verdanken sich dagegen einer anderen Überlegung. Der Schwierigkeit, für jeden Kontext separate Parameter zu schätzen, wird mit der Überlegung begegnet, dass es möglich sein könnte, eine Verteilung dieser Parameter zu berechnen.

[4] Eine ausführliche Diskussion findet man bei Rohwer und Pötter 2002b.

Zu diesem Zweck wird angenommen, dass die Modellparameter zufällig zwischen den Kontexten variieren; oder anders formuliert: dass auch Unterschiede in den Modellparametern zwischen den Kontexten durch fiktive Zufallsgeneratoren entstehen. Die Modellformulierung beginnt mit dem einfachen Ansatz $Y = \alpha + X\beta + \epsilon$. Dann werden die Parameter dieses Modells von kontextspezifischen Zufallsvariablen abhängig gemacht:

$$\alpha = \alpha_0 + \epsilon_\alpha \tag{8}$$

$$\beta = \beta_0 + \epsilon_\beta \tag{9}$$

Durch Einsetzen entsteht dann das MLRC-Modell

$$Y = \alpha_0 + X\beta_0 + (\epsilon_\alpha + X\epsilon_\beta + \epsilon) \tag{10}$$

Man wird sich vielleicht wundern, dass bei dieser Modellformulierung überhaupt kein expliziter Verweis auf die unterschiedlichen Kontexte erfolgt; also zum Beispiel kein Index j auftritt, mit dem auf jeweils bestimmte Kontexte verwiesen wird. Aber solche Verweise würden voraussetzen, dass man sich auf eine bestimmte Menge unterschiedlicher Kontexte (wie sie beispielsweise in einem Datensatz gegeben sind) beziehen kann. MLRC-Modellen liegt jedoch die Idee zugrunde, dass unterschiedliche Kontexte erst *durch Realisationen von Zufallsvariablen entstehen.* Bei dem Modell (10) entstehen unterschiedliche Kontexte durch Realisationen der Zufallsvariablen ϵ_α und ϵ_β.

MLRC-Modelle setzen also voraus, dass man die Kontexte als anonym betrachtet und ggf. vorhandene Informationen über ihre Identität nicht verwendet.[5]

16. Eine einfache Modellvariante.

Bei dem Modell (10) wird angenommen, dass sowohl α als auch β zufällig zwischen den Kontexten variieren können. Oft wird eine einfachere Modellvariante verwendet, bei der nur der Interzept α zufällig variieren kann. Das Modell (oft als Varianz-Komponenten-Modell bezeichnet) sieht dann folgendermaßen aus:

$$Y = \alpha_0 + X\beta + (\epsilon_\alpha + \epsilon) \tag{11}$$

Der stochastische Modellteil besteht jetzt aus zwei Komponenten: Realisierungen von ϵ_α liefern zufällige Unterschiede zwischen den Kontexten, Realisierungen von ϵ liefern zufällige Unterschiede zwischen den Individuen, die zu den Kontextunterschieden dazukommen.

[5] Longford (1993, S. 11) hat das so gesagt: „In our setup the focus is on a *sample* of clusters [Kontexte]; the clusters can be thought of as anonymously labelled units, in the same way as can the elementary observations.“

17. Annahmen über Zufallsvariablen.

Bei statistischen Regressionsmodellen, die sich auf durch statistische Variablen gegebene Daten beziehen, entstehen Residualvariablen durch die Modellspezifikation (und das jeweils verwendete Schätzverfahren). Es ist infolgedessen sinnlos, Annahmen über Eigenschaften der Verteilungen zu machen. Anders verhält es sich bei stochastischen Regressionsmodellen, die sich auf reale oder fiktive Zufallsgeneratoren beziehen. Da dann angenommen wird, dass die Daten aus Zufallsgeneratoren resultieren, die in irgendeiner Form real oder fiktiv existieren, wird es grundsätzlich möglich, Annahmen über die Beschaffenheit dieser Zufallsgeneratoren zu treffen. Da solche Annahmen die für das Modell vorausgesetzten Zufallsgeneratoren spezifizieren, gehören sie als Bestandteile zur Modellspezifikation.

Für das Modell (11) wird meistens angenommen: $\mathrm{E}(\epsilon_\alpha) = \mathrm{E}(\epsilon) = 0$ sowie $\mathrm{Cov}(\epsilon_\alpha, \epsilon) = 0$. Die gesamte Residualvarianz ergibt sich somit additiv durch $\mathrm{V}(\epsilon_\alpha) + \mathrm{V}(\epsilon)$.

Für das Modell (10) werden meistens die gleichen Annahmen getroffen, außerdem: $\mathrm{E}(\epsilon_\beta) = 0$ und $\mathrm{Cov}(\epsilon_\beta, \epsilon) = 0$. Es wird jedoch zugelassen, dass ϵ_α und ϵ_β korreliert sein können.[6]

Oft werden auch weitere Verteilungsannahmen über die Zufallsvariablen gemacht. Verbreitet ist insbesondere die Annahme, dass es sich um normalverteilte Variablen handelt.[7]

18. Welche Erkenntnisse liefern MLRC-Modelle?

Als Ergebnis der Schätzung eines MLRC-Modells erhält man einerseits Schätzwerte für die mit Regressorvariablen assoziierten Modellparameter und andererseits Schätzwerte für die Varianzen (und ggf. Kovarianzen) der Zufallsvariablen, die den stochastischen Modellteil bilden. Besonders übersichtlich ist das einfache Varianz-Komponenten-Modell (11), bei dem man zusätzlich zu den Schätzwerten für α_0 und β_0 einen Schätzwert $\hat{\sigma}^2_{\epsilon_\alpha}$ für die Varianz von ϵ_α und einen Schätzwert $\hat{\sigma}^2_\epsilon$ für die Varianz von ϵ erhält. Da angenommen wird, dass ϵ_α und ϵ unabhängig sind, kann man auch schätzen, welche Anteile der gesamten Residualvarianz formal der Ebene der Kontexte bzw. der Individualebene zugerechnet werden können.

Man erhält jedoch aus der Literatur über MLRC-Modelle kaum Aufschluss darüber, welche Schlussfolgerungen aus einer solchen Aufteilung der gesamten Residualvarianz auf die beiden Ebenen gezogen werden können. Andererseits werden oft Erkenntnisansprüche formuliert, die m.E. problematisch sind.

a) Wie in Abschnitt 5 zitiert wurde, wird ein wichtiges Ziel von Mehrebenenmodellen oft darin gesehen, dass man erkennt, wie Zusammenhänge

[6] Man vgl. etwa Goldstein 1995, S. 17.

[7] Man vgl. etwa Goldstein 1995, S. 22; Blien, Wiedenbeck und Arminger 1994, S. 270.

zwischen unabhängigen und abhängigen Variablen wiederum von Kontexten abhängen. Offenbar kann dies mit MLRC-Modellen nicht ermittelt werden.

b) Gelegentlich wird gesagt, dass zumindest eine Verteilung (für die Variation von Modellparametern zwischen den Kontexten, also der Zufallsvariablen ϵ_α) geschätzt werden könne (Goldstein 1995, S. 17). Hier muss man jedoch aufpassen: Bestenfalls kann die Varianz von ϵ_α geschätzt werden; über die Form der Verteilung erhält man nicht nur keine Informationen, sondern sie wird in Gestalt einer Normalverteilung vorausgesetzt. Aber für diese Annahme gibt es tatsächlich keinerlei Begründung.

c) MLRC-Modelle liefern auch keinen Beitrag zur Erklärung der Werte einer abhängigen Variablen durch Rückgriff auf substantielle Eigenschaften der unterschiedlichen Kontexte. Natürlich ist es keineswegs ausgeschlossen, solche Kontexteigenschaften (wie beispielsweise die Betriebsgröße in dem Beispiel in Abschnitt 6) durch zusätzliche Regressorvariablen auch in MLRC-Modellen zu berücksichtigen. Das Spezifikum von MLRC-Modellen liegt jedoch gerade darin, *stattdessen* zufällige Unterschiede zwischen den Kontexten anzunehmen. Wenn man es positiv formulieren will, dienen sie gewissermaßen dem Versuch, unbeobachtete Unterschiede zwischen den Kontexten zu erfassen.

d) Mithilfe von MLRC-Modellen kann man nichts über identifizierbare Kontexte aussagen. Sie eignen sich deshalb nicht für Anwendungen wie beispielsweise Vergleiche bestimmter Länder, bei denen es möglich und erforderlich ist, unterschiedliche Kontexte zu identifizieren.[8]

e) Aus dem gleichen Grund ist auch eine Verwendung von MLRC-Modellen zur Skalierung bzw. Bewertung unterschiedlicher Kontexte (z.B. Schulen) zumindest fragwürdig. Diese Verwendung wird z.B. von Goldstein (1995, S. 17) folgendermaßen angedeutet:

> „An important class of situations arises when we wish primarily to have information about each individual school in a sample, but where we have a large number of schools so that (2.2) [entspricht dem Modellansatz (5) ohne die Interaktionsterme] would involve estimating a very large number of parameters. Furthermore, some schools may have rather small numbers of students and application of (2.2) would result in imprecise estimates. In such cases, if we regard the schools as members of a population and then use our population estimates of the mean and between-school variation, we can utilize this information to obtain more precise estimates for each individual school.“

[8] Insofern ist es merkwürdig, dass Blien et al. (1994) bei der Interpretation des von ihnen geschätzten MLRC-Modells einige Kontexte (die Städte Wolfsburg und Pirmasens) identifizieren.

Offenbar setzt dieser Gedanke nicht nur eine Identifizierbarkeit von Kontexten voraus, die durch das MLRC-Modell eigentlich ausgeschlossen wird. Denkt man z.B. daran, etwas über eine bestimmte Schule herauszufinden, erscheint es auch ziemlich fragwürdig, zu diesem Zweck nicht nur Daten von dieser, sondern auch von anderen Schulen zu verwenden.

19. Goldsteins Beispiel mit Schulen

Zur Verdeutlichung betrachten wir ein Beispiel, das von Goldstein (1995, S. 15ff.) besprochen wird. Die Daten beziehen sich auf 728 Schüler in 48 Grundschulen in London. Es gibt folgende Variablen:

$$\begin{aligned} Y &\equiv \text{11-year math score} \\ X_1 &\equiv \text{8-year math score} \\ X_2 &\equiv \text{gender (boys - girls)} \\ X_3 &\equiv \text{social class (non-manual - manual)} \end{aligned}$$

Goldstein verwendet ein einfaches Varianz-Komponenten-Modell der Form

$$Y = \alpha_0 + X_1\beta_{01} + X_2\beta_{02} + X_3\beta_{03} + (\epsilon_\alpha + \epsilon) \tag{12}$$

Goldstein schätzt das Modell mit einem IGLS-Verfahren (iterative generalized least squares), das er in seinem Buch bespricht, und findet folgende Werte (jeweils geschätzte Standardfehler in Klammern):

$$\begin{aligned} &\hat{\alpha}_0 = 14.9,\ \hat{\beta}_{01} = 0.64\,(0.025),\ \hat{\beta}_{02} = -0.36\,(0.34),\ \hat{\beta}_{03} = 0.72\,(0.39) \\ &\hat{\sigma}^2_{\epsilon_\alpha} = 3.21\,(1.0),\ \hat{\sigma}^2_\epsilon = 19.6\,(1.1) \end{aligned}$$

Was kann dieses Modell im Unterschied zu einem Modell, das nur einen einfachen Residualterm aufweist und mit OLS geschätzt werden könnte, zeigen? Es kann zeigen, dass vermutlich ein Teil der „unerklärten Varianz“ auf Unterschiede zwischen den Schulen zurückgeführt werden könnte, d.h. mithilfe von Kontextvariablen erklärt werden *könnte* (denn Kontextvariablen werden in dem Modell gar nicht verwendet). Wie groß dieser Varianzanteil ist, kann natürlich mit dem Modell nicht verlässlich geschätzt werden, da bereits die Fragestellung von einer Unterscheidung zwischen Individual- und Kontextvariablen (sowie von einem Modell) abhängt.

Es ist auch nützlich, sich zu überlegen, wie man das Modell zur Erzeugung (Simulation) von Daten – z.B. für 48 oder für beliebig viele Schulen – verwenden könnte. Könnte man dann die durch das Modell erzeugten mit den für seine Berechnung verwendeten Daten sinnvoll vergleichen?

20. Korrelationen zwischen Individuen?

Bei der Diskussion von Mehrebenenmodellen wird in der Literatur oftmals die Vorstellung geäußert, dass die Individuen innerhalb der Kontexte in irgendeinem Sinn „ähnlicher“ sind als zwischen den Kontexten. Gelegentlich findet man auch den Versuch, diese Vorstellung gleichsam statistisch zu formulieren, indem von einer „Korrelation zwischen Individuen“ gesprochen wird (Goldstein 1995, S. 19) oder von einer Korrelation zwischen individuellen Fehlertermen, wie in folgendem Zitat von Blien, Wiedenbeck und Arminger (1994, S. 268f.):

> „The assumption of stochastic independence may not be fulfilled when the individuals are clustered within groups, such as classes or schools in education or regions in the labor market. For the individuals that belong to the same cluster, the errors may be correlated.“

Offenbar sind Formulierungen dieser Art obskur, denn der Begriff einer Korrelation bezieht sich auf statistische oder Zufallsvariablen, kann aber nicht sinnvoll auf Individuen oder ihre Fehlerterme angewendet werden.

Tatsächlich lässt sich der „assumption of stochastic independence“ bei Regressionsmodellen, die sich auf statistische Variablen beziehen, überhaupt keine bestimmte Bedeutung geben; und auch bei stochastischen Regressionsmodellen ist dies nicht ohne weiteres möglich und kann – in gewisser Weise – nur durch einen spekulativen Trick erreicht werden.

21. Modellformulierungen für Individuen

Der Trick besteht darin, für jedes Individuum, auf das man sich mithilfe eines Datensatzes beziehen kann, ein eigenes stochastisches Regressionsmodell anzunehmen. Werden Individuen durch $i = 1, \dots, n$ indiziert, würde man z.B. das Modell (1) auf folgende Weise in Gestalt von n Modellen schreiben:

$$Y_i = \alpha + X_i\beta + \epsilon_i \tag{13}$$

X_i und Y_i sind, wie X und Y im Modell (1), Variablen; aber im Unterschied zu X und Y, für die es jeweils n Werte gibt, gibt es für die Variablen X_i und Y_i nur jeweils eine Beobachtung (die vom Individuum i stammt). Für ϵ_i liegt natürlich überhaupt keine Beobachtung vor. Bei einer stochastischen Modellinterpretation wird aber ϵ_i als eine Zufallsvariable aufgefasst, für die man sich beliebig viele mögliche Realisationen denken kann. So kann man sich vorstellen, dass die beim Individuum i ermittelten Daten irgendeiner bestimmten Realisation der Zufallsvariablen ϵ_i entsprechen, dass man sich aber theoretisch auf den Zufallsgenerator beziehen kann, mit dem beliebig viele Realisationen der Zufallsvariablen erzeugt werden könnten.

Als Ergebnis hat man für jedes Individuum i eine Zufallsvariable ϵ_i; und da es sich nun um Variablen handelt, kann man Annahmen über mögliche Korrelationen zwischen diesen Variablen machen.

22. MLRC-Modellformulierungen für Individuen

In der Literatur, die sich mit stochastischen Regressionsmodellen beschäftigt, ist es sehr verbreitet, Modelle mit individuell indizierten Variablen zu formulieren. Dies gilt insbesondere für die Literatur zu MLRC-Modellen. Es ist nützlich, sich dies kurz zu vergegenwärtigen.

Dafür knüpfen wir an das Modell (10) in Abschnitt 15 an. Es wird angenommen, dass es m Kontexte $j = 1, \dots, m$ gibt und dass n_j die Anzahl der Individuen im Kontext j ist (so dass es insgesamt $n = \Sigma_j n_j$ Individuen gibt). Jedes Individuum wird somit durch ein Indexpaar ij identifiziert, wobei sich j auf den Kontext und i auf die Nummer des Individuums innerhalb des Kontextes bezieht.[9] Das Modell (10) erscheint dann in der Form

$$Y_{ij} = \alpha_0 + X_{ij}\beta_0 + (\epsilon_{\alpha,j} + X_{ij}\epsilon_{\beta,j} + \epsilon_{ij}) \tag{14}$$

Bemerkenswert ist, dass die Zufallsvariablen ϵ_α und ϵ_β nur mit j indiziert werden. Tatsächlich ist das keineswegs selbstverständlich, denn man könnte auch diese Variablen doppelt indizieren und dann hinzufügen, dass sich alle Variablen mit dem gleichen j-Index auf den gleichen Zufallsgenerator beziehen sollen. Das wäre aber ein anderer Modellansatz, nämlich: Generiere für jedes Individuum zufällig einen Kontext und einen Wert für ϵ_{ij}. Dagegen unterstellt das Modell in der Schreibweise (14) eine andere Regel: Generiere zunächst zufällig einen Kontext und dann für jedes Individuum in diesem Kontext einen Wert für ϵ_{ij}.

Offenbar gibt es diese beiden Möglichkeiten auch beim Varianz-Komponenten-Modell (11). In der Literatur wird fast immer die Formulierung

$$Y_{ij} = \alpha_0 + X_{ij}\beta_0 + (\epsilon_{\alpha_j} + \epsilon_{ij}) \tag{15}$$

verwendet, die der zweiten Regel entspricht.

23. Annahmen über die individuellen Zufallsvariablen

Hat man nun ein MLRC-Modell für Individuen formuliert, kann man weitere Annahmen für die individuellen – d.h. den Individuen eines Datensatzes fiktiv zugerechneten – Zufallsvariablen machen. Verwendet man sinngemäß die Annahmen aus Abschnitt 17, gelangt man für das Modell (15) zu den Annahmen

$$\mathrm{E}(\epsilon_{\alpha,j}) = 0, \quad \mathrm{E}(\epsilon_{ij}) = 0, \quad \mathrm{Cov}(\epsilon_{ij}, \epsilon_{\alpha,j'}) = 0 \;\; \text{(für alle } i, j, j'\text{)}$$

[9] Werden die Individuen durch $k = 1, \dots, n$ nummeriert, nehmen wir an, dass $j(k)$ den entsprechenden Kontext und $i(k)$ die Nummer des Individuums innerhalb des Kontextes liefert.

Hinzu kommen jetzt die Annahmen

$$\begin{aligned} &\mathrm{Cov}(\epsilon_{\alpha,j},\epsilon_{\alpha,j'}) = 0 \quad (\text{für } j \neq j') \\ &\mathrm{Cov}(\epsilon_{ij},\epsilon_{i'j'}) = 0 \quad (\text{für } i \neq i' \text{ oder } j \neq j') \end{aligned}$$

Außerdem werden meistens konstante Varianzen angenommen, also

$$\mathrm{V}(\epsilon_{ij}) = \sigma_\epsilon^2 \;\text{ und }\; \mathrm{V}(\epsilon_{\alpha,j}) = \sigma_\alpha^2 \quad (\text{für alle } i,j)$$

Für das Modell (14) kommen dann sinngemäß einige weitere Annahmen hinzu.

Schließlich kann man sich zur Aufklärung der in Abschnitt 20 erwähnten obskuren Formulierung, dass Individuen korreliert sein können, auf die individuellen Zufallsvariablen beziehen. Offenbar findet man für das einfache Varianz-Komponenten-Modell aufgrund der eben genannten Annahmen:

$$\mathrm{Cov}(\epsilon_{\alpha,j}+\epsilon_{ij},\epsilon_{\alpha,j'}+\epsilon_{i'j'}) = \begin{cases} \sigma_\alpha^2+\sigma_\epsilon^2 & \text{wenn } j=j',\, i=i' \\ \sigma_\alpha^2 & \text{wenn } j=j',\, i\neq i' \\ 0 & \text{andernfalls} \end{cases}$$

24. Verwendung der Matrix-Schreibweise

Für die weiteren Überlegungen ist es nützlich, die Modelle in der Matrix-Schreibweise zu formulieren.[10] Wir beginnen mit einem gewöhnlichen Regressionsmodell der Form

$$Y = X_1\beta_1 + \cdots + X_p\beta_p + e \tag{16}$$

wobei angenommen wird, dass $X_1 = 1$ ist, so dass β_1 der Interzept ist. Bezieht man sich auf Individuen $i = 1,\ldots,n$, kann man definieren:

$$\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{np} \end{bmatrix} \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} \quad \mathbf{e} = \begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix}$$

Damit kann man den Modellansatz auch in der Form

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \tag{17}$$

schreiben. Diese Formulierung kann sowohl für statistische als auch für stochastische Regressionsmodelle verwendet werden. Bei statistischen Regressionsmodellen sind die Komponenten von $\mathbf{y}$ und $\mathbf{X}$ die den Individuen zurechenbaren Daten; bei stochastischen Modellkonzeptionen sind die Komponenten von $\mathbf{e}$ und somit auch von $\mathbf{y}$ Zufallsvariablen.

[10] Eine kurze Einführung und Erläuterung der hier verwendeten Notationen findet man bei Rohwer und Pötter 2002a, Anhang A.

25. Ein allgemeines MLRC-Modell

Um zu einer allgemeinen Formulierung für MLRC-Modelle zu gelangen, gehen wir von (14) aus, erlauben aber sowohl im fixen als auch im stochastischen Teil des Modells mehrere Regressorvariablen, die auch nicht identisch zu sein brauchen. Eine erste Formulierung ist

$$Y_{ij} = X_{1,ij}\beta_1 + \cdots + X_{p,ij}\beta_p + u_{ij} + e_{ij} \tag{18}$$

Im fixen Teil des Modells gibt es die Regressorvariablen $X_1,\ldots,X_p$ mit den zugehörigen Parametern $\beta_1,\ldots,\beta_p$. Dabei nehmen wir an, dass $X_1 = 1$ ist, so dass β_1 einen Interzept liefert. Der stochastische Teil besteht aus den beiden Zufallsvariablen u_{ij} und e_{ij}.

e_{ij} entspricht der bisher ϵ_{ij} genannten Zufallsvariablen. Jeder dieser Zufallsvariablen korrespondiert ein eigener Zufallsgenerator, der von allen anderen unabhängig ist. Die Annahmen sind:

$$\mathrm{E}(e_{ij}) = 0, \quad \mathrm{Cov}(e_{ij},e_{i'j'}) = \begin{cases} \sigma_e^2 & \text{wenn } i=i' \text{ und } j=j' \\ 0 & \text{andernfalls} \end{cases} \tag{19}$$

Die Zufallsvariable u_{ij} liefert im einfachsten Fall für jeden Kontext j einen Wert und ist dann nicht von i abhängig (das entspricht dem einfachen Varianz-Komponenten-Modell). Im allgemeinen können die Werte jedoch auch von Regressorvariablen abhängen, so dass sich folgende Formulierung anbietet:

$$u_{ij} = Z_{1,ij}v_{1,j} + \cdots + Z_{q,ij}v_{q,j} \tag{20}$$

$Z_1,\ldots,Z_q$ sind Regressorvariablen, $v_{1,j},\ldots,v_{q,j}$ sind Zufallsvariablen. Die Regressorvariablen können teilweise oder vollständig mit den Regressorvariablen im fixen Modellteil übereinstimmen. Insbesondere kann $Z_1 = 1$ sein, so dass, wenn $q = 1$ ist, das einfache Varianz-Komponenten-Modell entsteht.

Für die Zufallsvariablen $v_{k,j}$ ($k = 1,\ldots,q$, $j = 1,\ldots,m$) wird angenommen, dass $\mathrm{E}(v_{k,j}) = 0$ ist und dass sie mit den Zufallsvariablen e_{ij} nicht korreliert sind, also $\mathrm{Cov}(e_{ij},v_{k,j'}) = 0$ für alle i,j und k,j'. Für ihre Varianzen wird $\mathrm{V}(v_{k,j}) = \sigma_{v_k}^2$ angenommen, so dass sie unabhängig vom Kontext j sind. Außerdem wird angenommen, dass diese Zufallsvariablen zwischen den Kontexten nicht korreliert sind, also $\mathrm{Cov}(v_{k,j},v_{k',j'}) = 0$, wenn $j \neq j'$ ist. Sie können aber innerhalb jedes Kontextes miteinander korreliert sein, so dass mit einer Kovarianz-Matrix für $v_{1,j},\ldots,v_{q,j}$ gerechnet werden muss (die jedoch wiederum unabhängig vom Kontext ist).

26. Matrix-Formulierung des allgemeinen MLRC-Modells

Jetzt formulieren wir den Modellansatz des vorangegangenen Abschnitts mit der Matrix-Schreibweise. Dafür nehmen wir an, dass Werte der Variablen Y_{ij} und X_{ij} durch y_{ij} bzw. x_{ij} bezeichnet werden. Die weiteren Definitionen erfolgen schrittweise, beginnend mit Definitionen für die durch j indizierten Kontexte ($j = 1, \ldots, m$):

$$\mathbf{y}_j = \begin{bmatrix} y_{1j} \\ \vdots \\ y_{n_j j} \end{bmatrix} \quad \mathbf{u}_j = \begin{bmatrix} u_{1j} \\ \vdots \\ u_{n_j j} \end{bmatrix} \quad \mathbf{e}_j = \begin{bmatrix} e_{1j} \\ \vdots \\ e_{n_j j} \end{bmatrix} \quad \mathbf{v}_j = \begin{bmatrix} v_{1,j} \\ \vdots \\ v_{q,j} \end{bmatrix}$$

$$\mathbf{X}_j = \begin{bmatrix} x_{1,1j} & \cdots & x_{p,1j} \\ \vdots & & \vdots \\ x_{1,n_j j} & \cdots & x_{p,n_j j} \end{bmatrix} \quad \mathbf{Z}_j = \begin{bmatrix} z_{1,1j} & \cdots & z_{q,1j} \\ \vdots & & \vdots \\ z_{1,n_j j} & \cdots & z_{q,n_j j} \end{bmatrix} \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}$$

Offenbar ist $\mathbf{u}_j = \mathbf{Z}_j \mathbf{v}_j$, und man kann für jeden Kontext j schreiben:

$$\mathbf{y}_j = \mathbf{X}_j \boldsymbol{\beta} + \mathbf{Z}_j \mathbf{v}_j + \mathbf{e}_j \tag{21}$$

Um schließlich ein Gesamtmodell zu formulieren, werden die kontextspezifischen Ausdrücke zusammengefasst:

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_m \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_m \end{bmatrix} \quad \mathbf{Z} = \begin{bmatrix} \mathbf{Z}_1 \\ \vdots \\ \mathbf{Z}_m \end{bmatrix} \quad \mathbf{u} = \begin{bmatrix} \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_m \end{bmatrix} \quad \mathbf{e} = \begin{bmatrix} \mathbf{e}_1 \\ \vdots \\ \mathbf{e}_m \end{bmatrix}$$

Somit kann man das Gesamtmodell in der Form

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} + \mathbf{e} \tag{22}$$

schreiben. Ergänzend können die im vorangegangenen Abschnitt getroffenen Annahmen über den stochastischen Modellteil folgendermaßen ausgedrückt werden:

a) $\mathrm{E}(\mathbf{u}) = 0, \;\; \mathrm{E}(\mathbf{e}) = 0, \;\; \mathrm{E}(\mathbf{v}_j) = 0 \quad (j = 1, \ldots, m)$

b) $\mathrm{V}(\mathbf{e}) = \mathrm{E}(\mathbf{e}\mathbf{e}') = \sigma_e^2\, \mathbf{I}_n$, wobei $\mathbf{I}_n$ eine Einheitsmatrix der Ordnung n ist.

c) $\mathrm{Cov}(\mathbf{u}, \mathbf{e}) = 0$, so dass $\mathrm{V}(\mathbf{u} + \mathbf{e}) = \mathrm{V}(\mathbf{u}) + \mathrm{V}(\mathbf{e})$ ist.

d) $\mathrm{V}(\mathbf{u}) = \mathrm{E}(\mathbf{u}\mathbf{u}')$ ist die Kovarianz-Matrix der Zufallsvariablen $\mathbf{u}$, die die Form einer Blockdiagonalmatrix hat, da die Variablen zwischen den Kontexten nicht korreliert sind. Der j.te Block ist

$$\mathrm{V}(\mathbf{Z}_j \mathbf{v}_j) = \mathrm{E}(\mathbf{Z}_j \mathbf{v}_j \mathbf{v}_j' \mathbf{Z}_j') = \mathbf{Z}_j \mathrm{E}(\mathbf{v}_j \mathbf{v}_j') \mathbf{Z}_j'$$

$\mathrm{E}(\mathbf{v}_j \mathbf{v}_j')$ ist die Kovarianz-Matrix der Zufallsvariablen $\mathbf{v}_j$. Da angenommen wird, dass sie nicht vom Kontext abhängt, kann sie durch eine kontextunabhngige Matrix $\mathbf{S}$ der Ordnung q repräsentiert werden, wobei in der Hauptdiagonalen die Varianzen $\sigma_{v_k}^2$ und außerhalb die Kovarianzen stehen.

27. Das OLS-Schätzverfahren

Bevor wir mit der Besprechung von MLRC-Modellen fortfahren, erinnern wir kurz an einige Aspekte der OLS- und GLS-Verfahren zur Parameterschätzung von Regressionsmodellen.

Beim OLS-Schätzverfahren wird die Summe der quadrierten Residuen minimiert. Geht man von (17) aus, wird also $\mathbf{e}'\mathbf{e}$ minimiert. Als Lösung findet man den Parametervektor

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \tag{23}$$

Daraus folgt auch sogleich, dass die Residuen $\hat{\mathbf{e}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$ und die Regressorvariablen unkorreliert sind. Denn aus (23) folgt $\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$, und daraus folgt

$$0 = \mathbf{X}'\mathbf{y} - \mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \mathbf{X}'\hat{\mathbf{e}}$$

Wenn man (17) als ein stochastisches Regressionsmodell auffasst, ist $\mathbf{e}$ ein Vektor, dessen Komponenten Zufallsvariablen sind, und es wird möglich, Annahmen über deren gemeinsame Verteilung zu machen. Eine besonders einfache Annahme ist

$$\mathrm{E}(\mathbf{e}) = 0 \;\text{ und }\; \mathrm{V}(\mathbf{e}) = \mathrm{E}(\mathbf{e}\mathbf{e}') = \sigma^2\, \mathbf{I} \tag{24}$$

wobei $\mathbf{I}$ eine Einheitsmatrix passender Ordnung bezeichnet.

Aus (23) folgt, dass dann auch die Komponenten von $\hat{\boldsymbol{\beta}}$ Zufallsvariablen sind, und man kann den Erwartungswert berechnen. Zunächst findet man:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \mathbf{e}) = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{e} \tag{25}$$

Somit folgt, wegen der Annahme $\mathrm{E}(\mathbf{e}) = 0$, dass $\mathrm{E}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$ ist, d.h. das OLS-Verfahren ist in diesem Fall erwartungstreu. Also kann man die Kovarianzmatrix von $\hat{\boldsymbol{\beta}}$ in der Form $\mathrm{V}(\hat{\boldsymbol{\beta}}) = \mathrm{E}((\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})')$ schreiben. Und da aus (25) $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{e}$ folgt, findet man

$$\mathrm{E}((\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})') = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\,\mathrm{E}(\mathbf{e}\mathbf{e}')\,\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$$

Somit folgt schließlich aus der in (24) getroffenen Annahme über $\mathrm{E}(\mathbf{e}\mathbf{e}')$, dass $\mathrm{V}(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ ist.[11]

[11] Darauf gründen sich die verbreiteten Signifikanztests, wobei zur Schätzung von σ^2 meistens $(\mathbf{y} - \hat{\mathbf{y}})'(\mathbf{y} - \hat{\mathbf{y}})/(n - p)$ verwendet wird, wobei $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$ ist.

Man kann sich fragen, ob das OLS-Verfahren an irgendwelche Voraussetzungen gebunden ist, wie manchmal behauptet wird. Das hängt in gewisser Weise davon ab, ob man sich auf ein statistisches oder ein stochastisches Regressionsmodell bezieht. Bezieht man sich auf ein statistisches Regressionsmodell, gibt es keine besonderen Voraussetzungen. Außerdem impliziert das OLS-Verfahren, dass Residuen und Regressorvariablen unkorreliert sind. Das gilt natürlich auch, wenn man das OLS-Verfahren zur Schätzung eines stochastischen Regressionsmodells verwendet. In diesem Fall kann aber ein Widerspruch auftreten, wenn die Modellspezifikation die Vorstellung nahelegt, dass die Zufallsvariablen im stochastischen Modellteil mit den Regressorvariablen korreliert sind.[12]

28. Das GLS-Schätzverfahren

Wie bereits erwähnt wurde, erlauben stochastische Regressionsmodelle, wenn sie für Individuen formuliert werden, Annahmen darüber, ob und ggf. wie die individuellen Zufallsvariablen ϵ_i korreliert sind. Zwar setzt das OLS-Verfahren nicht voraus, dass diese Zufallsvariablen nicht korreliert sind. Wie jedoch im vorangegangenen Abschnitt erwähnt wurde, wird dies für die übliche Schätzung einer Kovarianzmatrix $\mathrm{V}(\boldsymbol{\beta})$ für die Modellparameter angenommen.

Somit kann man auch fragen, wie man vielleicht zu einer solchen Kovarianzmatrix gelangt, wenn man die Annahme der Unkorreliertheit nicht treffen möchte. In einigen Fällen kann man dann das GLS-Verfahren (generalized least squares) verwenden. Dies Verfahren geht davon aus, dass die Kovarianzmatrix des Residualvektors $\mathbf{e}$ die Form

$$\mathrm{V}(\mathbf{e}) = \mathrm{E}(\mathbf{e}\mathbf{e}') = \sigma^2\,\mathbf{W}$$

hat, wobei σ^2 ein unbekannter (zu schätzender) Skalar und $\mathbf{W}$ eine bekannte positiv definite Matrix ist. Es folgt dann nämlich, dass $\mathbf{W}$ in der Form $\mathbf{W} = \mathbf{A}\mathbf{A}'$ geschrieben werden kann, wobei gilt:

$$\mathbf{A}^{-1}\mathbf{W}\mathbf{A}^{-1'} = \mathbf{I} \quad \text{und} \quad \mathbf{A}^{-1'}\mathbf{A}^{-1} = \mathbf{W}^{-1}$$

Definiert man nun $\mathbf{y}^* = \mathbf{A}^{-1}\mathbf{y}$, $\mathbf{X}^* = \mathbf{A}^{-1}\mathbf{X}$ und $\mathbf{e}^* = \mathbf{A}^{-1}\mathbf{e}$, kann man das Regressionsmodell

$$\mathbf{y}^* = \mathbf{X}^*\boldsymbol{\beta} + \mathbf{e}^* \tag{26}$$

betrachten, bei dem die Komponenten des Residualvektors unkorreliert sind, denn offenbar gilt:

$$\mathrm{E}(\mathbf{e}^*\mathbf{e}^{*'}) = \mathrm{E}(\mathbf{A}^{-1}\mathbf{e}\mathbf{e}'\mathbf{A}^{-1'}) = \mathbf{A}^{-1}\mathrm{E}(\mathbf{e}\mathbf{e}')\mathbf{A}^{-1'} = \sigma^2\,\mathbf{I}$$

[12] Als Beispiel kann man das allgemeine MLRC-Modell betrachten, wenn eine oder mehrere der X-Variablen im stochastischen Modellteil verwendet werden. Die als fragwürdig empfundene Korrelation ist natürlich in jedem Fall eine Folge der Modellspezifikation und kein Aspekt der Realität, auf die sich das Modell bezieht.

Also kann man den Parametervektor $\boldsymbol{\beta}$ im Modell (26) mit dem OLS-Verfahren schätzen und erhält $\mathrm{V}(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})^{-1}$.

29. Iterierte GLS-Schätzung des MLRC-Modells

Um das in den Abschnitten 25 und 26 definierte allgemeine MLRC-Modell zu schätzen, sind unterschiedliche Verfahren vorgeschlagen worden. Wir besprechen hier das IGLS-Verfahren, wie es von Goldstein (1995, S. 38ff.) angegeben wurde.

Es sei $\mathbf{w} = \mathbf{u} + \mathbf{e}$ die Summe der Zufallsvektoren des Modells (22), und $\mathbf{W}$ sei die Kovarianzmatrix von $\mathbf{w}$, also $\mathbf{W} = \mathrm{Cov}(\mathbf{w}) = \mathrm{E}(\mathbf{w}\mathbf{w}')$. Es handelt sich um eine Blockdiagonalmatrix, die aus m Blöcken $\mathbf{W}_1, \ldots, \mathbf{W}_m$ besteht. Für den j.ten Block gilt aufgrund der Annahmen über den stochastischen Modellteil: $\mathbf{W}_j = \mathbf{Z}_j\mathbf{S}\mathbf{Z}_j' + \sigma_e^2\,\mathbf{I}_{n_j}$, wobei $\mathbf{I}_{n_j}$ eine Einheitsmatrix der Ordnung n_j ist.

Wäre $\mathbf{W}$ bekannt, könnten Schätzwerte für $\boldsymbol{\beta}$ mit dem GLS-Verfahren berechnet werden, nämlich durch

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}^{-1}\mathbf{y} \tag{27}$$

Da jedoch $\mathbf{W}$ zunächst nicht bekannt ist, wird versucht, mithilfe eines Iterationsverfahrens immer bessere Schätzungen für $\mathbf{W}$ und somit auch für $\boldsymbol{\beta}$ zu finden.

Um das Verfahren zu erklären, wird zunächst angenommen, dass $\boldsymbol{\beta}$ bekannt ist. Dann ist auch $\mathbf{w} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$ bekannt, und man kann die (n,n)-Matrix $\mathbf{w}\mathbf{w}'$ bilden. Diese Matrix wird nun als ein Vektor geschrieben, indem ihre Spalten untereinander angeordnet werden. Dafür wird die Notation $\mathrm{vec}(\mathbf{w}\mathbf{w}')$ verwendet;[13] es handelt sich um einen Spaltenvektor mit n^2 Elementen. Offenbar gilt: $\mathrm{vec}(\mathbf{W}) = \mathrm{vec}(\mathrm{E}(\mathbf{w}\mathbf{w}')) = \mathrm{E}(\mathrm{vec}(\mathbf{w}\mathbf{w}'))$.[14]

Wichtig ist nun eine Methode, um sich geordnet auf die Elemente der mit dem vec-Operator gebildeten Vektoren beziehen zu können. Um die Methode zu erläutern, beziehen wir uns auf den Vektor $\mathrm{vec}(\mathbf{W})$, der der (n,n)-Matrix $\mathbf{W}$ entspricht. Sei nun l $(1 \leq l \leq n^2)$ die Nummer irgendeines Elements von $\mathrm{vec}(\mathbf{W})$. Innerhalb der Matrix $\mathbf{W}$ befindet sich dieses Element in der Zeile

$$r(l) = (l-1)\,\%\,n + 1$$

und in der Spalte

$$c(l) = [(l-1)/n] + 1$$

[13] Diese Notation kann offenbar für beliebige Matrizen verwendet werden.

[14] Da $\mathbf{W}$ eine Blockdiagonalmatrix ist, sind viele Elemente dieses Vektors Null und könnten ignoriert werden. Es erleichtert jedoch die Notation, wenn mit den vollständigen Vektoren operiert wird.

(Hierbei ist % der Modulus-Operator und $[x]$ ist die größte ganze Zahl, die nicht größer als x ist.) Somit können auch die beteiligten Individuen identifiziert werden: Das erste Individuum hat die Nummer $i = i(r(l))$ im Kontext $j = j(r(l))$, das zweite Individuum hat die Nummer $i' = i(c(l))$ im Kontext $j' = j(c(l))$.[15] In der Hauptdiagonalen verweisen natürlich beide Indexpaare auf dasselbe Individuum.

Jetzt definieren wir einen Vektor, der die Parameter des stochastischen Modellteils, also die Elemente der Matrix $\mathbf{S}$ und außerdem σ_e^2, zusammenfasst:

$$\boldsymbol{\theta} = \begin{bmatrix} \text{vec}(\mathbf{S}) \\ \sigma_e^2 \end{bmatrix}$$

Die $q^2 + 1$ Elemente dieses Vektors sind die Parameter des stochastischen Modellteils, die zusätzlich zu den Elementen des Parametervektors $\boldsymbol{\beta}$ geschätzt werden müssen.[16]

Schließlich wird noch eine Design-Matrix $\mathbf{D}$ definiert. Sie hat n^2 Zeilen (entsprechend der Länge von $\text{vec}(\mathbf{W})$) und q^2+1 Spalten (entsprechend der Länge von $\boldsymbol{\theta}$). Sei nun l $(1 \le l \le n^2)$ ein Index für die Zeilen von $\mathbf{D}$. Wie oben erklärt wurde, gehören dazu der Zeilenindex $r(l)$, der Spaltenindex $c(l)$ sowie die Indizes (i, j) und (i', j') für die zugehörigen Individuen. Es wird nun definiert:

a) Wenn $j \neq j'$ ist, enthält die l.te Zeile von $\mathbf{D}$ nur Nullen.

b) Wenn $j = j'$ ist, wird die l.te Zeile von $\mathbf{D}$ durch den Zeilenvektor

$$(\text{vec}(\mathbf{z}'_{(r(l))}\mathbf{z}_{(c(l))})', \delta_{ii'})$$

definiert. Dabei ist $\delta_{ii'} = 1$, wenn $i = i'$, und andernfalls $= 0$. Außerdem wird $\mathbf{z}_{(r(l))}$ als Bezeichnung für die $r(l)$.te und $\mathbf{z}_{(c(l))}$ als Bezeichnung für die $c(l)$.te Zeile von $\mathbf{Z}$ verwendet.

Diese leider etwas umständlichen Definitionen führen zu einem einfachen Resultat:

$$\mathbf{D}\boldsymbol{\theta} = \text{vec}(\mathbf{W}) = \text{E}(\text{vec}(\mathbf{w}\mathbf{w}')) \tag{28}$$

Man kann also versuchen, Werte für $\boldsymbol{\theta}$ aus einem Regressionsansatz der Form

$$\text{vec}(\mathbf{w}\mathbf{w}') = \mathbf{D}\boldsymbol{\theta} + \mathbf{r} \tag{29}$$

zu schätzen (wobei $\mathbf{r}$ den Residualvektor bezeichnet). Wie von Goldstein gezeigt wird, sollte hierfür wiederum ein GLS-Verfahren verwendet werden, nämlich

$$\hat{\boldsymbol{\theta}} = (\mathbf{D}'\mathbf{U}^{-1}\mathbf{D})^{-1}\mathbf{D}'\mathbf{U}^{-1}\text{vec}(\mathbf{w}\mathbf{w}') \tag{30}$$

[15] Vgl. die Definition der hier verwendeten Operatoren in Abschnitt 22.

[16] Wegen der Symmetrie von $\mathbf{S}$ sind es nur $q(q+1)/2+1$ unterschiedliche Parameter.

Box 4 Das IGLS-Verfahren für das MLRC-Modell.

(1) $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$

(2) $\hat{\mathbf{w}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$

(3) $\hat{\boldsymbol{\theta}} = (\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'\text{vec}(\hat{\mathbf{w}}\hat{\mathbf{w}}')$

(4) $\hat{\mathbf{W}} = \text{ivec}(\mathbf{D}\hat{\boldsymbol{\theta}})$

(5) $\hat{\boldsymbol{\beta}}^* = (\mathbf{X}'\hat{\mathbf{W}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{W}}^{-1}\mathbf{y}$

(6) $\hat{\mathbf{w}}^* = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}^*$

(7) $\hat{\mathbf{U}} = \hat{\mathbf{W}} \otimes \hat{\mathbf{W}}$

(8) $\hat{\boldsymbol{\theta}}^* = (\mathbf{D}'\hat{\mathbf{U}}^{-1}\mathbf{D})^{-1}\mathbf{D}'\hat{\mathbf{U}}^{-1}\text{vec}(\hat{\mathbf{w}}^*\hat{\mathbf{w}}^{*\prime})$

(9) Ende der Iterationen, wenn $\hat{\boldsymbol{\beta}} \approx \hat{\boldsymbol{\beta}}^*$ und $\hat{\boldsymbol{\theta}} \approx \hat{\boldsymbol{\theta}}^*$
andernfalls $\hat{\boldsymbol{\beta}} \leftarrow \hat{\boldsymbol{\beta}}^*$, $\hat{\boldsymbol{\theta}} \leftarrow \hat{\boldsymbol{\theta}}^*$ und Fortsetzung bei (4).

wobei die Kovarianzmatrix $\mathbf{U} = \mathbf{W} \otimes \mathbf{W}$ verwendet werden sollte.[17] Mithilfe von $\hat{\boldsymbol{\theta}}$ kann also eine Schätzung der Kovarianz-Matrix $\mathbf{W}$ gewonnen werden, die dann wiederum für eine neue Schätzung von $\boldsymbol{\beta}$ verwendet werden kann.

Zusammenfassend ergibt sich ein iteratives Verfahren, dessen Grundzüge in Box 4 angegeben sind.

30. Ein Varianz-Komponenten-Modell für Beispiel 1

Um das Verfahren zu illustrieren, verwenden wir es zur Schätzung einiger Modelle für die Beispieldaten aus Abschnitt 6 (Box 1). Wir beginnen mit einem einfachen Varianz-Komponenten-Modell der Form (11) bzw. (15).

Box 5 zeigt das TDA-Skript für den in Box 4 beschriebenen Algorithmus, angepasst für die Daten in Box 1 (die sich in dem Datenfile `ml1.dat` befinden). Die abhängige Varible wird mit 10^{-3} multipliziert, um eine bessere Konditionierung der Daten herzustellen. Insgesamt werden fünf Iterationen vollzogen (eine Konvergenzprüfung entfällt).

Box 6 zeigt die Rechenergebnisse für fünf Iterationen. Man erkennt, dass der Prozess sehr schnell konvergiert.[18]

Die Schätzungen für die β-Parameter unterscheiden sich offenbar kaum von denjenigen, die aus der OLS-Schätzung eines einfachen Regressionsmodells resultieren. Die Besonderheit des Modells besteht nur in einer Zerlegung der Residualvarianz. Als Schätzungen erhält man

$$\hat{\sigma}^2_{\epsilon_\alpha} = 0.0218 \quad \text{und} \quad \hat{\sigma}^2_{\epsilon} = 0.0717$$

[17] $\otimes$ bezeichnet das Kronecker-Produkt.

[18] Natürlich hängt das Konvergenzverhalten des Algorithmus auch von den Daten ab.

Box 5 TDA-Skript zur Schätzung des Varianz-Komponenten-Modells

```
mdeff(DAT) = ml1.dat;

mscol(DAT,<1>,Y);
mscol(DAT,<2>,X);
mscol(DAT,<3>,G);
mexpr(Y / 1000,Y);

mdefc(20,1,1,Z);
mldes(Z,G,D);
mcath(Z,X,X);

mtransp(X,XT);
mcross(X,XX);
mginv(XX,XXI);
mmul(XXI,XT,Y,Beta);
mpr(Beta);

mmul(X,Beta,YE);
mexpr(Y - YE,W);
mtransp(W,WT);
mmul(W,WT,WW);
mcvec(WW,WWVec);
mtransp(D,DT);
mcross(D,DD);
mginv(DD,DDI);
mmul(DDI,DT,WWVec,Theta);
mpr(Theta);

repeat(n = 5);
    mmul(D,Theta,DTheta);
    mivec(DTheta,20,V);
    mginv(V,VI);
    mmul(XT,VI,X,XX);
    mginv(XX,XXI);
    mmul(XXI,XT,VI,Y,Beta1);
    mpr(Beta1);

    mmul(X,Beta1,YE);
    mexpr(Y - YE,W);
    mtransp(W,WT);
    mmul(W,WT,WW);
    mcvec(WW,WWVec);

    mkp(V,V,U);
    mginv(U,UI);
    mmul(DT,UI,D,DD);
    mginv(DD,DDI);
    mmul(DDI,DT,UI,WWVec,Theta1);
    mpr(Theta1);

    mexpr(Beta1,Beta);
    mexpr(Theta1,Theta);
endrepeat;
```


Box 6 Ergebnisses des IGLS-Verfahrens für 5 Iterationen

Iteration	Beta1	Beta2	Theta1	Theta2
0	1.2332	0.0546	0.0216	0.0719
1	1.2038	0.0554	0.0218	0.0717
2	1.2037	0.0554	0.0218	0.0717
3	1.2037	0.0554	0.0218	0.0717
4	1.2037	0.0554	0.0218	0.0717
5	1.2037	0.0554	0.0218	0.0717

Ergebnisse dieser Art werden oft so interpretiert, dass sich ein Teil der „unbeobachteten Heterogenität“ Unterschieden zwischen den Kontexten zurechnen lässt. Aber ist diese Interpretation gerechtfertigt?

31. Modifikationen durch weitere Regressorvariablen

Gesichtspunkte für Überlegungen zu dieser Frage erhält man, indem man weitere Regressorvariablen in das Modell aufnimmt. Als erstes verwenden wir zusätzlich die Variable Z aus Box 1, die sich auf die Betriebsgröße bezieht und deren Werte nur zwischen den Kontexten (Unternehmen) variieren. Das Modell sieht dann so aus:

$$Y = \alpha + X\beta_x + Z\beta_z + (\epsilon_\alpha + \epsilon_e) \tag{31}$$

Als Schätzwerte für die Modellparameter findet man mit dem IGLS-Algorithmus:

$$\begin{aligned} \hat{\alpha} &= 0.9736 & \hat{\sigma}^2_{\epsilon_\alpha} &= 0.0106 \\ \hat{\beta}_x &= 0.0564 & \hat{\sigma}^2_{\epsilon} &= 0.0717 \\ \hat{\beta}_z &= 0.0027 \end{aligned}$$

Die zusätzliche Aufnahme der Kontextvariablen Z hat den Varianzanteil, der formal den Kontexten zugerechnet wird, verringert; und dies scheint die übliche Interpretation zu stützen.

Aber hier ist noch einmal an die Bemerkungen in Abschnitt 7 zu erinnern, durch die die Unterscheidung zwischen Individual- und Kontextvariablen problematisiert wurde. Um das für unser gegenwärtiges Beispiel zu illustrieren, verwenden wir für die Variable Z anstelle von Angaben zur Betriebsgröße Angaben zum Geschlecht. Box 7 zeigt die modifizierten Daten (0 = Männer, 1 = Frauen).[19] Schätzt man nun (31) mit diesen

[19] Die Werte wurden so zugewiesen, dass Frauen eher niedrige, Männer eher höhere Verdienste haben.

Box 7 Modifizierte Daten für das Beispiel 1.

Y	X	K	Z	Y	X	K	Z	Y	X	K	Z	Y	X	K	Z
3000	25	1	1	2500	30	2	1	2400	25	3	1	2000	20	4	1
3200	30	1	1	2800	30	2	1	3000	30	3	1	3000	30	4	0
3300	40	1	0	2600	35	2	1	3200	30	3	0	4000	40	4	0
3500	45	1	0	3000	40	2	0	3500	40	3	0	3500	40	4	0
3700	50	1	0	3500	45	2	0	4000	45	3	0	4000	45	4	0

modifizierten Daten, findet man:

$$\hat{\alpha} = 1.74166 \qquad \hat{\sigma}^2_{\epsilon_\alpha} = 0.0090$$
$$\hat{\beta}_x = 0.0432 \qquad \hat{\sigma}^2_{\epsilon} = 0.0744$$
$$\hat{\beta}_z = -0.2518$$

Die „Individualvariable“ Geschlecht hat also die den Kontexten zugerechnete „unbeobachtete Heterogenität“ mehr verringert als die „Kontextvariable“ Betriebsgröße.

32. Weitere Modellvarianten

Als eine weitere Variante schätzen wir das MLRC-Modell (10), bei dem sowohl α als auch β als zufällige Parameter aufgefasst werden. Das Modell hat insgesamt 6 Parameter. Mit dem IGLS-Algorithmus findet man die Schätzwerte:

$$\hat{\alpha}_0 = 1.1909 \qquad \hat{\sigma}^2_{\epsilon_\alpha} = 0.4105$$
$$\hat{\beta}_0 = 0.0564 \qquad \hat{\sigma}^2_{\epsilon_\beta} = 0.0004$$
$$\hat{\sigma}_{\epsilon_\alpha,\epsilon_\beta} = -0.0119$$
$$\hat{\sigma}^2_{\epsilon} = 0.0365$$

Hierbei bezeichnet $\hat{\sigma}_{\epsilon_\alpha,\epsilon_\beta}$ den Schätzwert für die Kovarianz $\mathrm{Cov}(\epsilon_\alpha, \epsilon_\beta)$. Könnte dieses Ergebnis so interpretiert werden, dass eine „unbeobachtete Heterogenität“ hauptsächlich zwischen den Kontexten existiert?

Literatur

Baumgartner, T., Burns, T. R., Meeker, L. D., Wild, B. 1976. Open Systems and Multi-Level Processes: Implications for Social Research. International Journal of General Systems 3, 25–42.

Blien, U., Wiedenbeck, M., Arminger, G. 1994. Reconciling Macro and Micro Perspectives by Multilevel Models: An Application to Regional Wage Differences. In: I. Borg, P. P. Mohler (eds.), in: Trends and Perspectives in Empirical Social Research, 266–282. Berlin: de Gruyter.

Bryk, A. S., Raudenbush, S. W. 1992. Hierarchical Linear Models: Applications and Data Analysis Methods. Newbury Park: Sage.

DeLeeuw, J. 1992. Series Editor's Introduction to Hierarchical Linar Models. In: A. S. Bryk,S. W. Raudenbush, Hierarchical Linear Models: Applications and Data Analysis Methods, xii-xvi. Newbury Park: Sage.

DeLeeuw, J., Kreft, I. 1986. Random Coefficient Models for Multilevel Analysis. Journal of Educational Statistics 11, 57–85.

Eeden, P. van den, Hüttner, H. J. M. 1982. Multi-Level Research (= Current Sociology, Vol. 30, No. 3). London: Sage.

Engel, U. 1998. Einführung in die Mehrebenenanalyse. Opladen: Westdeutscher Verlag.

Goldstein, H. 1995. Multilevel Statistical Models. 2nd ed. London: Edward Arnold.

Huinink, J. 1989. Mehrebenensystem-Modelle in den Sozialwissenschaften. Wiesbaden: Deutscher Universitätsverlag.

Huinink, J. 1995. Warum noch Familie? Frankfurt: Campus.

Hummell, H. J. 1972. Probleme der Mehrebenenanalyse. Stuttgart: Teubner.

Longford, N. T. 1993. Random Coefficient Models. Oxford: Clarendon.

Mason, W. M., Wong, G. Y., Entwisle, B. 1984. Contextual Analysis Through the Multilevel Linear Model. In: S. Leinhardt (ed.), Sociological Methodology 1983-84, 72–103. San Francisco: Jossey-Bass.

Rohwer, G., Pötter, U. 2002a. Methoden sozialwissenschaftlicher Datenkonstruktion. Weinheim: Juventa.

Rohwer, G., Pötter, U. 2002b. Wahrscheinlichkeit. Begriff und Rhetorik in der Sozialforschung. Weinheim: Juventa.