Friedrich Pukelsheim

Schätzen

von Mittelwert und Streuungsmatrix

in GAUSS-MARKOFF-Modelllen

*

Diplomarbeit

Freiburg im Breisgau

September 1974

Ich habe diese Arbeit selbständig angefertigt und die benutzten Quellen angegeben.

(Versicherung gemäß Abschnitt III, § 1, Satz 2 der Diplomprüfungsordnung für Studierende der Mathematik vom 1. 10. 1962, Mathematische Fakultät der Albert-Ludwigs-Universität, Freiburg im Breisgau.)

Friedrich Pukelsheim

Einleitender Überblick

Inhaltsverzeichnis

Symbolliste

Einleitender Überblick

GAUSS-MARKOFF-Modelle sind lineare Modelle linearer Regression:
- lineare Modelle, weil der Parameter aus einem linearen Raum kommt,
- linearer Regression, weil die Regressionsfunktion für diesen Parameter linear ist.

Gesucht sind lineare Schätzfunktionen mit geeigneten Optimalitätseigenschaften. Um diese Aufgabe zu lösen, werden Hilfsmittel aus der Theorie endlich-dimensionaler linearer Räume herangezogen (vgl. Teil I).

In Teil II wird die Mittelwert-Schätzung erörtert, genauer: Geschätzt werden lineare Funktionen des Regressionsparameters für den Mittelwert. Dafür werden drei Optimalitätsbegriffe diskutiert und die zugehörigen optimalen Schätzer angegeben:
- Minimum Norm - Minimum Bias - lineare Schätzfunktionen (Kap. 4),
- Minimum Varianz - Minimum Bias - lineare Schätzfunktionen (Kap. 5),
- Minimum mean square error - lineare Schätzfunktionen (Kap. 6).

Asymptotische Fragestellungen werden nicht untersucht.

Teil III beschäftigt sich mit Streuungs-Schätzungen. Dabei wird wesentlich auf die Ergebnisse der Schätzung von Mittelwerten zurückgegriffen: Zu einem gegebenen "ursprünglichen" GAUSS-MARKOFF-Modell wird ein neues "abgeleitetes" GAUSS-MARKOFF-Modell konstruiert, so daß der Regressionsparameter für die Streuung im ursprünglichen Modell erscheint als Regressionsparameter für den Mittelwert im abgeleiteten Modell. Mit dieser "Streuungs-Mittelwert-Korrespondenz" werden die hergeleiteten Ergebnisse für die Mittelwert-Schätzung in natürlicher Weise auf die Streuungs-Schätzung übertragen.

Damit erhält man auch für die Streuung eine Theorie der
- Minimum Norm - Minimum Bias - Schätzung (Kap. 7),
- Minimum Varianz - Minimum Bias - Schätzung (Kap. 8),
- Minimum mean square error - Schätzung (Kap. 9).

Ein viertes Schätzkonzept ist RAOs MInimum Norm Quadratic Unbiased Estimation. Von dieser MINQUE-Theorie wird gezeigt (Abschnitte 7.4; 8.1), daß sie in engem Bezug steht sowohl zur Minimum Norm - Minimum Bias - Schätzung des Kapitels 7 als auch zur Minimum Varianz - Minimum Bias - Schätzung des Kapitels 8 .

Auf grund der Streuungs-Mittelwert-Korrespondenz lassen sich auch die optimalen Schätzer für die Streuung leicht explizit darstellen und bereits bekannte Ergebnisse aus der Literatur bequem einordnen (45:84; 48:91; 50:94; 52:98; 54:101)[1].

*

Die Literatur zur Mittelwert-Schätzung zerfällt in
- die klassische "Maximalrang-Theorie" und
- die neuere "ranglose Theorie".

In der Maximalrang-Theorie wird gefordert, daß alle auftretenden Matrizen maximalen Rang haben. Die beiden Hauptergebnisse dieser klassischen Theorie sind der Satz von GAUSS-MARKOFF und der von AITKEN (vgl. Abschnitt 5.2). Zwar kann man auch Modelle mit entarteten Matrizen behandeln, doch die Rechnung ist umständlich und führt zu unübersichtlichen Ergebnissen: Mit einer geeigneten Transformation wechselt man zu einem Maximalrang-Modell, berechnet hier die Schätzer und transformiert zurück (vgl. RAO 1971c: 376f)[2].

In der ranglosen Theorie wird auf Voraussetzungen über die Ränge verzichtet. "The unified theory appears to be somewhat simpler as it lays down a common numerical procedure for all situations and avoids the complicated algebraic notations generally used in the discussion of the general GAUSS-MARKOFF model." (RAO 1971c: 378). Möglich wurde dieser vereinheitlichende Ansatz durch die Theorie verallgemeinerter Matrixinverser, wie sie seit 1955 von R. PENROSE, C.R. RAO und anderen entwickelt wurde (vgl. Bibliographie in RAO & MITRA 1971). Grundlagen und Ergebnisse der neueren Schätztheorie diskutieren zum Beispiel die Lehrbücher von ALBERT (1972) und von RAO & MITRA (1971).

---

[1] Verweis auf Satz 45, Seite 84 dieser Arbeit, etc.
[2] Verweis auf Seite 376, 377 der unter RAO 1971c angegebenen Literatur.

In der Literatur zur Schätzung von Streuungsparametern gibt bisher nur RAO mit seiner MINQUE-Theorie (1970; 1971a; 1972) ein geschlossenes Schätzkonzept an. Bei den meisten anderen Verfahren, die in der Literatur empfohlen werden, wird Optimalität - nach Kenntnis des Verfassers - fast nur unter Normalverteilungsannahmen gezeigt.

Einige Kriterien ohne solch starke Verteilungsannahmen geben HSU (1938), RAO (1971b) und DRYGAS (1972). LAMOTTE (1973) berechnet Klassen optimaler Schätzer und gibt teilweise explizite Darstellungen.

*

Wir werden zum Teil anders Vorgehen als in der Literatur üblich: Statt Linearformen werden gleich lineare Funktionen geschätzt; alle Plausibilitätsüberlegungen werden auf die eine Matrizenungleichung $\|AB\| \leq \|A\|\cdot\|B\|$ gestützt; an die Stelle der Erwartungstreue tritt von vorneherein die Minimierung des Bias; alle Schätzer werden explizit berechnet. Im einzelnen führt dies zu den folgenden neuen Ergebnissen:

Minimum Norm - Minimum Bias - Schätzung des Mittelwerts liefert eine weitere Rechtfertigung der Methode der kleinsten Quadrate.

Bei der Minimum Varianz - Minimum Bias - Schätzung ist die von RAO & MITRA (1971) geforderte hinreichende Bedingung zur AITKEN-Schätzung auch notwendig (33.i:62). Die von RAO & MITRA (1971) und ALBERT (1972) angegebenen BLUEs sind gleich (36:68).

Die Schätzer des hier eingeführten mean square Risikos sind nicht in einem Punkt, sondern auf einer Sphäre optimal (Abschnitt 6.1); sie lassen sich aus einem Minimum Varianz - Minimum Bias - Schätzer berechnen, und umgekehrt (38:73).

Bei der Streuungs-Schätzung ergibt - wie schon oben erläutert - die Streuungs-Mittelwert-Korrespondenz (Abschnitt 7.2) eine abgerundete Schätztheorie für die Streuung (Teil III); RAOs MINQUE-Theorie wird dabei neu begründet und verallgemeinert (Abschnitt 8.1).

Inhaltsverzeichnis

Symbolliste

Vektorräume

$\mathbb{R}^n$	n-dimensionaler reeller Raum von Spaltenvektoren
$\mathbb{E}^n$	n-dimensionaler EUKLIDischer Raum, S. 13
e_ν^n	$= (0,\dots,0,1,0,\dots,0)^T$ ν-ter EUKLIDischer Basisvektor
dim	Dimension
span	linearer Spann
$\mathfrak{L}^\perp$	orthogonales Komplement eines Teilraums $\mathfrak{L}$
$\mathrm{proj}(\cdot\vert\mathfrak{L})$	Projektion(smatrix) auf einen Teilraum $\mathfrak{L}$, S. 16
$\mathfrak{L}+\mathfrak{K}$	$= \{\ l+k \mid l\in\mathfrak{L},\ k\in\mathfrak{K}\ \}$, für Vektormengen $\mathfrak{L}$, $\mathfrak{K}$
$\mathfrak{L}\oplus\mathfrak{K}$	$\mathfrak{L}+\mathfrak{K}$, falls $\mathfrak{L}\perp\mathfrak{K}$
L(V,W)	Raum der linearen Abbildungen von einem Vektorraum V in einen Vektorraum W
$L^2(V\times W,U)$	Raum der bilinearen Abbildungen von $V\times W$ in U

Matrizen

M(n,m)	Raum der reellen (n,m)-Matrizen
M(n)	Raum der (quadratischen) (n,n)-Matrizen
Sym(n)	Raum der symmetrischen (n,n)-Matrizen
Diag(n)	Raum der diagonalen (n,n)-Matrizen
Proj(n)	Menge der (n,n)-Projektionsmatrizen , S. 13
GL(n)	generelle lineare Gruppe in M(n)
O(n)	orthogonale Gruppe in M(n)
$E_{\nu\mu}^{(n,m)}$	$= e_\nu^n (e_\mu^m)^T$: (ν,μ)-te EUKLIDische Basismatrix in M(n,m), S. 13
I_n	$= \sum E_{\nu\nu}^{(n,n)}$: (n,n)-Einheitsmatrix
p.s.d.	positiv semidefinit
p.d.	positiv definit
S^2	p.s.d. Matrix mit p.s.d. Wurzel S, S. 17
DY	Streuungsmatrix (Varianz-Kovarianz-Matrix) eines Zufallsvektors Y
EY	Erwartungswert eines Zufallsvektors Y
Var Y	Varianz einer reellwertigen Zufallsgröße Y

Matrizenfunktionen

[ ]	eckige Klammern : Blockmatrix bzw. Blockvektor
$\otimes$	KRONECKERprodukt
Diag	S. 14
vec	S. 27
tr	Spur
$\Vert\cdot\Vert$	EUKLIDische Vektor- und Matrixnorm
$V(\cdot,\cdot)$	quasi-inneres Produkt, S. 43
A^T	transponierte Matrix von A
A^+	(MOORE-PENROSE-)pseudoinverse Matrix von A
rk A	Rang von A
im A	Bild von A
ker A	Kern von A
Lsg(Ax=B)	affiner Lösungsraum der Matrizengleichung Ax = B, S. 41
Lsg($\Vert Ax-B\Vert$ =inf)	affiner Lösungsraum der Minimierungsaufgabe $\Vert Ax - B\Vert = \inf$, S. 42

GAUSS-MARKOFF-Modelle

Übersichtstabelle: GAUSS-MARKOFF-Modelle		
GM-Modell	Bausteine	Struktur
(Y,Xb,n,p)	Y $\mathbb{R}^n$-wertiger Zufallsvektor $X \in M(n,p)$	$EY \in \text{im } X$
(Y,Xb,n,p,S^2)	$S^2 \in \text{Sym}(n)$, p.s.d.	$EY \in \text{im } X$ $DY = S^2$
$(Y,Xb,n,p,\sum t_\kappa D_\kappa)$	$(D_\kappa)_{\kappa=1,\dots,k} \in \text{Sym}(n)^k$	$EY \in \text{im } X$ $DY \in \text{span } D_\kappa$
$(Y,Xb,n,p,\sum t_\kappa D_\kappa,F^2)$	$F^2 \in \text{Sym}(n^2)$	$EY \in \text{im } X$ $DY \in \text{span } D_\kappa$ $D\ (Y-EY)\otimes(Y-EY) = F^2$

(lin.Mod.)	S. 48
(lin.Regr.)	S. 49
(GMM1,2,3)	S. 49
(GMM4)	S. 50
(GMM5,6)	S. 77
(GMM7)	S. 77

Begriffstabelle: Namen und gebundene Variable in GM-Modellen	
Y	Beobachtungsvektor
n	Anzahl der Beobachtungen
X	Mittelwert-Design
s	$:= \mathrm{rk}\, X$
b	Mittelwert-Regressionsparameter
p	Anzahl der Regressoren für den Mittelwert
$S_p(\beta)$	$:= \{ b \in \mathbb{R}^p : \lVert b \rVert = \beta \}$: p-dimensionale Sphäre
M	$:= I_n - XX^+ = \mathrm{proj}(\cdot \mid \mathrm{im}^{\perp} X) \in \mathrm{Proj}(n)$, $\mathrm{rk}\, M = n-s$
MY	streuungsrelevanter Anteil des Beobachtungsvektors Y
$(D_\kappa)_k$	Streuung-Design
t	Streuung-Regressionsparameter
$S_k(\tau)$	$:= \{ t \in \mathbb{R}^k : \lVert t \rVert = \tau \}$: k-dimensionale Sphäre
D	$:= [\mathrm{vec}\, D_\kappa]_{\kappa=1,\dots,k} \in M(n^2,k)$
N	$:= M \otimes M - M \otimes M \cdot D (M \otimes M \cdot D)^+ \in \mathrm{Proj}(n^2)$
V_κ^2	$= D_\kappa$, falls D_κ p.s.d.
V^2	$:= \sum V_\kappa^2$
M_V	$:= I_n - V^+ X (V^+ X)^+$

F-transformiertes GM-Modell	S. 50
AITKEN-Modell	S. 50
ursprüngliches Modell	S. 50
abgeleitetes Modell	S. 79
MINQUE-Modell	S. 86
erweitertes Modell	S. 88
HSU-Modell	S. 94

<u>Schätzfunktionen</u>

lin. Schätzfkt. } (Mittelwert-)Schätzer }	lineare Schätzfunktion, S. 51
inv. quadr. Schätzfkt. } (Streuungs-)Schätzer }	invariante quadratische Schätzfunktion, S. 80
Min.Norm-Min.Bias-Schätzer	S. 55; S. 82
Min.RAOnorm-Min.Bias-Schätzer	S. 88
Min.Var.-Min.Bias-Schätzer	S. 58
$S_p(\beta)$-lokale GAUSS-Schätzer	S. 70
$S_k(\tau)$-lokale GAUSS-Schätzer	S. 100

* * *

Teil I

Lineare Algebra

Die Schätztheorie für GAUSS-MARKOFF-Modelle beruht auf zwei Stützen aus der linearen Algebra: dem Tensorenkalkül und der Pseudoinversion von Matrizen.

Das natürliche Tensorprodukt zweier EUKLIDischer Räume ist das KRONECKERprodukt $x \otimes y$. Mit der Spur tr AB^T als innerem Produkt auf Matrizenräumen ist die Funktion vec A, die aus einer Matrix einen Vektor macht, ein isometrischer tensorprodukt-treuer Vektorraum-Isomorphismus (Kap. 1).

Die Pseudoinversion beliebiger Matrizen ist eine Verallgemeinerung der Inversion regulärer Matrizen (Kap. 2). Mit Pseudoinversen A^+ kann man die Lösungen linearer Matrizengleichungen einfach und überschaubar darstellen (Kap. 3).

Kap. 0: Bezeichnungen

0.1 EUKLIDische Räume

Der n-dimensionale EUKLIDische Raum $\mathbb{E}^n$ ist der $\mathbb{R}^n$, versehen mit seiner natürlichen Basis und dem Standardskalarprodukt $x^T y$. Der ν-te Basisvektor werde mit e_ν^n bezeichnet: $e_\nu^n = (0,\dots,0,\underset{\substack{\uparrow \\ \nu\text{-te Stelle}}}{1},0,\dots,0)^T$.

Für Teilmengen $\mathfrak{M}$ von Vektoren sei $\operatorname{span}\mathfrak{M}$ die lineare Hülle. $\mathfrak{L}^\perp$ sei das orthogonale Komplement eines Teilraums $\mathfrak{L}$.

Da in EUKLIDischen Räumen die Basis bestimmt ist, fällt der Raum $L(\mathbb{E}^m, \mathbb{E}^n)$ aller linearen Abbildungen vom $\mathbb{E}^m$ in den $\mathbb{E}^n$ und der Raum $M(n,m)$ aller reellen (n,m)-Matrizen in natürlicher Weise zusammen: $L(\mathbb{E}^m, \mathbb{E}^n) = M(n,m)$.

0.2 Matrizen-Bezeichnungen

Matrizenmengen: Die Bezeichnungen für die benutzten Matrizenmengen sind zusammengestellt in der

Begriffstabelle Matrizen		
Name	Definition	
$M(n,m)$	alle reellen (n,m)-Matrizen	Matrizen
$M(n)$	$M(n,n)$	quadratische Matrizen
$Sym(n)$	$\{ S \in M(n) : S=S^T \}$	symmetrische Matrizen
$Diag(n)$	$\{ D \in M(n) : D = Diag\ D \}$	Diagonalmatrizen
$Proj(n)$	$\{ M \in M(n) : M=M^T=M^2 \}$	Projektionsmatrizen

I_n sei die Einheitsmatrix in der generellen linearen Gruppe $GL(n)$.

$E_{\nu\mu}^{(n,m)}$ sei die (ν,μ)-te Basismatrix des Raumes $M(n,m)$, d.h. sie hat eine 1 auf Platz (ν,μ) und sonst Nullen.

Es gilt[34] $E^{(n,m)}_{\nu\mu} = e^{n}_{\nu}(e^{m}_{\mu})^{T}$, $I_n = \sum E^{(n,n)}_{\nu\nu}$.

<u>Matrizenfunktionen:</u> $\mathrm{tr}\ G = \sum G_{\nu\nu}$ bezeichne die Spur einer quadratischen Matrix $G \in M(n)$. Unter der Spur kommutieren Matrizenprodukte: $\mathrm{tr}\ AB = \mathrm{tr}\ BA$; ist nämlich A eine (n,m)-Matrix und B eine (m,n)-Matrix, so gilt
$\mathrm{tr}\ AB = \sum A_{\nu\mu}B_{\mu\nu} = \sum B_{\mu\nu}A_{\nu\mu} = \mathrm{tr}\ BA$.

Der Name Diag wird für zwei Funktionen verwendet; wir schreiben Diag mit einem großen D, weil beide Funktionen Matrizen als Werte annehmen.

Die erste Diag-Funktion ist für quadratische Matrizen G definiert. Diag G stanzt die Hauptdiagonale aus und setzt alle anderen Einträge auf Null.

Die zweite Diag-Funktion ist für Vektoren a definiert. Diag a schiebt den Vektor a auf die Hauptdiagonale und ergänzt mit Nullen.

<u>Matrizen als lineare Abbildungen:</u> Eine (n,m)-Matrix A kann als lineare Abbildung vom $\mathbb{E}^m$ in den $\mathbb{E}^n$ aufgefaßt werden.
Es bezeichne

$\ker A = A^{-1}\{0_{\mathbb{R}^n}\}$	den Kern (Nullraum, kernel) von A,
$\mathrm{im}\ A = A\ \mathbb{R}^m$	das Bild (Bildraum, image, ränge) von A,
$\mathrm{rk}\ A = \dim \mathrm{im}\ A$	den Rang (rank) von A.

ker A ist ein Unterraum im $\mathbb{E}^m$, im A ist ein Unterraum im $\mathbb{E}^n$.

Der folgende Satz i zeigt, wie (i) Kern und Bild zusammenhängen und (ii) sich mit GRAMschen Matrizen beschreiben lassen; Aussage iii ist ein Hilfssatz zu Satz 26.i:53[5].

---

[4] Wir halten uns an die folgende <u>Summationskonvention</u>:
Bei einem Summationszeichen ohne Summationsindizes wird über <u>alle</u> griechischen Indizes summiert, und zwar von 1 bis zu ihrem lateinischen Äquivalent.

[5] Aus den Beweisen ersieht man sofort, daß allgemeiner die Aussagen i, ii in HILBERT-Räumen und iii in Vektorräumen gelten.

<u>Satz 1: Kern, Bild und GRAMsche Matrizen; Urbild-Spann</u>

i (Kern und Bild)

Für jede (n,m)-Matrix A gilt:

$\ker^{\perp} A = \operatorname{im} A^T$.

ii (GRAMsche Matrizen)

Für jede (n,m)-Matrix A gelten die Gleichungen

$\ker A = \ker A^T A$,

$\ker A^T = \ker AA^T$,

$\operatorname{im} A^T = \operatorname{im} A^T A$,

$\operatorname{im} A = \operatorname{im} AA^T$

iii (Spann vom Urbild = Urbild vom Spann)

Sei $\mathfrak{M}$ eine nichtleere Menge im Bild einer (n,m)-Matrix A : $\emptyset \neq \mathfrak{M} \subset \operatorname{im} A$. Dann gilt:

$\operatorname{span} A^{-1}\mathfrak{M} = A^{-1} \operatorname{span} \mathfrak{M}$.

Bew.: i. Es wird gezeigt: $\ker A = \operatorname{im}^{\perp} A^T$.

$$
\begin{aligned}
x \in \ker A \quad &\Leftrightarrow \quad Ax = 0 \\
&\Leftrightarrow \quad \forall\ y \in \mathbb{R}^n : \quad y^T A x = 0 \\
&\Leftrightarrow \quad \forall\ y \in \mathbb{R}^n : \quad (A^T y)^T x = 0 \\
&\Leftrightarrow \quad x \perp \operatorname{im} A^T \\
&\Leftrightarrow \quad x \in \operatorname{im}^{\perp} A^T
\end{aligned}
$$

ii. I[6]. $\ker A \subset \ker A^T A$ klar. Für $x \in \ker A^T A$ verschwindet aber auch $x^T A^T A x = \|Ax\|^2$ und damit Ax .

II. Man setze in der ersten Gleichung von A^T an die Stelle von A .

III. $\operatorname{im} A^T = \ker^{\perp} A = \ker^{\perp} A^T A = \operatorname{im} (A^T A)^T = \operatorname{im} A^T A$.

IV. Man setze in III A^T statt A .

iii. $\subset$: $\operatorname{span} A^{-1}\mathfrak{M} \subset A^{-1}\operatorname{span} \mathfrak{M}$ folgt sofort aus $A^{-1}\mathfrak{M} \subset A^{-1}\operatorname{span} \mathfrak{M}$.

$\supset$: Für die Menge

$\mathfrak{L} := \{ y \in \operatorname{span} \mathfrak{M} : A^{-1}\{y\} \subset \operatorname{span} A^{-1}\mathfrak{M} \}$

gilt offenbar $\mathfrak{M} \subset \mathfrak{L} \subset \operatorname{span} \mathfrak{M}$; ist $\mathfrak{L}$ zudem Vektorraum, so folgt $\mathfrak{L} = \operatorname{span} \mathfrak{M}$:

jeder Punkt $y \in \operatorname{span} \mathfrak{M}$ liegt auch in $\mathfrak{L}$,

d.h. jedes $y \in \operatorname{span} \mathfrak{M}$ hat die Eigenschaft $A^{-1}\{y\} \subset \operatorname{span} A^{-1}\mathfrak{M}$,

d.h. $A^{-1} \operatorname{span} \mathfrak{M} \subset A^{-1}\mathfrak{M}$.

Bleibt also zu zeigen, daß $\mathfrak{L}$ Vektorraum ist:

I. Für $y \in \mathfrak{L}$, $\alpha \neq 0$ ist $\alpha y \in \mathfrak{L}$ wegen $A^{-1}\{\alpha y\} = \alpha A^{-1}|y\}$.

---

[6] Große römische Zahlen dienen zur Gliederung von Beweisen.

II. Für $0 \in \mathfrak{L}$ ist zu zeigen: $\ker A \subset \operatorname{span} A^T\mathfrak{M}$.

Wegen $\mathfrak{M} \neq \emptyset$ existiert ein $x \in A^{-1}\mathfrak{M}$, $y := Ax \in \mathfrak{M}$.

Für $x' \in \ker A$ ist dann $x+x' \in A^{-1}\mathfrak{M}$ wegen $A(x+x') = y$.

Somit $x' = x+x' - x \in \operatorname{span} A^{-1}\mathfrak{M}$.

III. Sei $y, y' \in \mathfrak{L}$; es folgt dann $y+y' \in \mathfrak{L}$ falls $x'' \in \operatorname{span} A^{-1}\mathfrak{M}$ für alle x'' mit $Ax'' = y+y'$.

Nach Voraussetzung ist $y, y' \in \operatorname{span} \mathfrak{M} \subset \operatorname{im} A$, also kann man x, x' wählen mit $Ax = y$, $Ax' = y'$; sei ferner $Ax'' = y+y'$. Dann ist $A(x+x'-x'') = 0$ und daher einerseits $x+x'-x'' \in \ker A \subset \operatorname{span} A^{-1}\mathfrak{M}$. Andererseits sind $y, y' \in \mathfrak{L}$ und deshalb $x, x' \in \operatorname{span} A^{-1}\mathfrak{M}$. Somit $x'' = x + x' - (x+x'-x'') \in \operatorname{span} A^{-1}\mathfrak{M}$. _/

<u>Projektionsmatrizen:</u> Die Projektionsmatrizen $\operatorname{Proj}(n)$ fallen zusammen mit den orthogonalen Projektionen des $\mathbb{E}^n$. Für "orthogonale Projektion" wird kurz "Projektion" geschrieben: In dieser Arbeit sind alle Projektionen EUKLIDisch orthogonal.

Für einen Unterraum $\mathfrak{L}$ des $\mathbb{E}^n$ wird die Projektion auf $\mathfrak{L}$ notiert mit $\operatorname{proj}(\cdot|\mathfrak{L})$ und - als lineare Abbildung – identifiziert mit der entsprechenden Projektionsmatrix. Für die Projektion auf das orthogonale Komplement $\mathfrak{L}^\perp$ von $\mathfrak{L}$ gilt: $\operatorname{proj}(\cdot|\mathfrak{L}^\perp) = I_n - \operatorname{proj}(\cdot|\mathfrak{L})$ . . Statt eines Matrizenprodukts $\operatorname{proj}(\cdot|\mathfrak{L})\cdot A$ schreiben wir kürzer $\operatorname{proj}(A|\mathfrak{L})$.

Jede Projektionsmatrix M projiziert genau auf ihren Bildraum: $M = \operatorname{proj}(\cdot|\operatorname{im} M)$. Mit unserer Schreibweise hat man dann $\operatorname{proj}(A|\operatorname{im} M) = \operatorname{proj}(\cdot|\operatorname{im} M)\cdot A = MA$.

Im folgenden Satz 2 wird der Rang $\operatorname{rk} M$ einer Projektionsmatrix M berechnet (i) und ein Kriterium gegeben (ii), wann das Produkt MN von Projektionen M,N wieder eine Projektion ist.

<u>Satz 2: Projektionen</u>

i (Rang einer Projektion)

Für jede Projektion M ist $\operatorname{rk} M = \operatorname{tr} M$.

ii (Produkt von Projektionen)

Für zwei (n,n)-Projektionsmatrizen M,N gilt:

$MN \in \operatorname{Proj}(n) \Leftrightarrow MN = NM \Leftrightarrow MN = \operatorname{proj}(\cdot|\operatorname{im} M \cap \operatorname{im} N)$.

Bew. : i. Projektionen haben die Eigenwerte 0 oder 1, denn $\lambda x = Mx = M\cdot Mx = \lambda^2 x$, $\lambda = \lambda^2$. Wird M mit der orthogonalen Matrix R diagonalisiert, so folgt:
$\operatorname{tr} M = \operatorname{tr} MR^TR = \operatorname{tr} RMR^T$ = Summe der Eigenwerte = Anzahl der nicht-verschwindenden Eigenwerte = $\dim \operatorname{im} M = \operatorname{rk} M$.
ii. I. Sei MN Projektion. Dann ist $MN = (MN)^T = NM$.
II. Bei $MN = NM$ ist MN symmetrisch: $(MN)^T = NM = MN$, und idempotent: $(MN)^2 = MN\cdot NM = MNM = MN$.
III. Es reicht zu zeigen: $\operatorname{im} MN = \operatorname{im} M \cap \operatorname{im} N$.
$\operatorname{im} MN = \operatorname{im} NM \subset \operatorname{im} M \cap \operatorname{im} N$ ist klar . Für $x \in \operatorname{im} M \cap \operatorname{im} N$ gilt wegen $x = Nx = Mx = MNx$ aber auch $x \in \operatorname{im} MN$. _/

Streuungsmatrizen: Die Streuungsmatrix (Varianz-Kovarianz-Matrix, dispersion matrix) $E\,(Y-EY)(Y-EY)^T$ eines Zufallsvektors Y werde mit DY bezeichnet. Streuungsmatrizen sind positiv semidefinit (p.s.d.): $DY \in \operatorname{Sym}(n)$ und für alle $x \in \mathbb{R}^n$:
$x^T\cdot DY\cdot x = E\, x^T(Y-EY)(Y-EY)^Tx = E(\,(Y-EY)^Tx\,)^2 \geq 0$.

Zu jeder p.s.d. Matrix $S \in \operatorname{Sym}(n)$ gibt es genau eine p.s.d. Matrix $\sqrt{S} \in \operatorname{Sym}(n)$ mit der Eigenschaft $S = (\sqrt{S})^2$ (ALBERT 1972:38). Um das Wurzelzeichen zu vermeiden, werden p.s.d. Matrizen gleich als Quadrat dieser ihrer p.s.d. Wurzel geschrieben: also S^2, S statt S, $\sqrt{S}$; und $DY = S^2$.

Ist S^2 eine p.s.d. Matrix und $\delta^2 > 0$, so ist $S^2 + \delta^2 I_n$ positiv definit (p.d.) .

* * *

Kap. 1: Tensorprodukte

Mit dem Tensorenkalkül wird die Diskussion von bi- und multilinearen Abbildungen reduziert auf die Theorie linearer Abbildungen.

Zunächst werden Tensorprodukte für Vektorräume (1.1) und lineare Abbildungen (1.2) eingeführt - wir folgen dabei der Darstellung von GREUB (1967) -, dann werden diese Ergebnisse auf EUKLIDische Räume angewendet (1.3, 1.4).

1.1 Tensorprodukte zweier Vektorräume

In diesem Abschnitt sei K ein (Skalaren-)Körper; Hinweise auf K , wie K-Vektorraum, K-dim etc., bleiben der Einfachheit halber fort. V und W seien zwei Vektorräume. $L(V,W)$ sei der Raum aller linearen Abbildungen von V in W ; $L^2(V\times W,U)$ sei der Raum aller bilinearen Abbildungen von $V\times W$ in einen weiteren Vektorraum U .

<u>Def. 1:</u> i (Tensorprodukte von Vektorräumen)

Ein Paar (T,φ) heißt "<u>Tensorprodukt für V und W</u>", falls

(TP1) T ein Vektorraum ist, und

(TP2) $\varphi\in L^2(V\times W,T)$ eine bilineare Abbildung von $V\times W$ in T ist mit der <u>Faktorisierungseigenschaft</u>:

(TP3) Für jeden Vektorraum U und jede bilineare Abbildung $f\in L^2(V\times W,U)$ von $V\times W$ in U existiert genau eine lineare Abbildung $f^* \in L(T,U)$ mit $f = f^*\circ\varphi$.

ii (Tensorprodukt von Vektoren)

Ist (T,φ) ein Tensorprodukt für V und W und sind $v\in V$ und $w\in W$ zwei Vektoren, so heißt $\varphi(v,w)$ das "<u>Tensorprodukt von v und w</u>". _/

Die Faktorisierungseigenschaft (TP3) bedeutet: Jedes Diagram

$$\begin{tikzcd} V\times W \arrow[rr, "f \text{ bil.}"] \arrow[dr, "\varphi"'] & & U \\ & T & \end{tikzcd}$$

läßt sich eindeutig zu einem kommutativen Diagram vervollständigen:

$$\begin{tikzcd} V\times W \arrow[rr, "f \text{ bil.}"] \arrow[dr, "\varphi"'] & & U \\ & T \arrow[ur, "f^* \text{ lin.}"'] & \end{tikzcd}$$

Die Zuordnung $*$ nach (TP3) liefert die schon angedeutete Reduktion von bilinearen auf lineare Abbildungen:

<u>Satz 3: Reduktionssatz</u>

(LANG 1966:224) Sei (T,φ) ein Tensorprodukt für V und W ; sei U ein Vektorraum.
Dann ist die über die Faktorisierungseigenschaft (TP3) gegebene Zuordnung $* : L^2(V\times W,U) \to L(T,U)$, $f \mapsto f^*$ ein Vektorraum-Isomorphismus.

<u>Bew.:</u> I. $*$ ist wohldefiniert, da nach (TP3) für jedes bilineare f <u>genau eine</u> lineare Abbildung f^* existiert.
II. $*$ ist surjektiv; denn für eine lineare Abbildung $h\in L(T,U)$ von T in U ist offenbar $f := h\circ\varphi$ bilinear und wegen der Eindeutigkeit von f^* gerade $h = f^*$.
III. $*$ ist injektiv: aus $f^* = g^*$ folgt $f = f^*\circ\varphi = g^*\circ\varphi = g$.
IV. $*$ ist linear: es gilt $(\alpha f+\beta g)^* = \alpha f^*+\beta g^*$ wegen $(\alpha f^*+\beta g^*)\circ\varphi = \alpha f^*\circ\varphi + \beta g^*\circ\varphi = \alpha f + \beta g$. _/

Weil in (TP3) <u>eindeutige</u> Faktorisierbarkeit gefordert wird, ist der Raum T minimal in dem Sinne, daß er die lineare Hülle über dem Bild von φ ist:

Satz 4: Minimaleigenschaft von T (GREUB 1967:6)

Sei (T,φ) ein Tensorprodukt für V und W.
Dann gilt $T = \operatorname{span} \varphi\, V\mathrm{x}W$.

Bew.: I. φ ist bilinear und (T,φ) ein Tensorprodukt, also existiert genau eine lineare Abbildung $\varphi^* \in L(T,T)$ mit $\varphi = \varphi^* \circ \varphi$. Offensichtlich gilt $\varphi^* = id_T$.
II. Sei $T' := \operatorname{span} \varphi\, V\mathrm{x}W$ der Spann des Bildes von φ. Definiert man $\varphi' : V\mathrm{x}W \to T'$, $(v,w) \mapsto \varphi(v,w)$, dann ist φ' bilinear und faktorisiert also gemäß (TP3): $\varphi' = f \circ \varphi$, $f \in L(T,T')$. Mit der Inklusionsabbildung $j : T' \hookrightarrow T$ gilt dann $\varphi = j \circ \varphi' = j \circ f \circ \varphi$.
III. Aus der Eindeutigkeit von φ^* folgt mit I, II: $id_T = j \circ f$. Also ist j surjektiv: $T' = T$. _/

Wegen dieses Satzes reicht es, bei linearen Abbildungen $h \in L(T,U)$ auf T das Verhalten auf dem Bild $\varphi\, V\mathrm{x}W$ von φ zu untersuchen. Zum Beispiel sind zwei lineare Abbildungen $g,h \in L(T,U)$ gleich, wenn die bilinearen Abbildungen $g \circ \varphi$, $h \circ \varphi$ gleich sind.

Satz 4 legt auch nahe, wie man Tensorprodukte auf Teilräumen erhält:

Satz 5: Tensorprodukt für Teilräume (GREUB 1967:13)

Sei (T,φ) ein Tensorprodukt für V und W; seien Ω_V, Ω_W Teilräume in V bzw. W. Setze
$T_\Omega := \operatorname{span} \varphi\, \Omega_V \mathrm{x} \Omega_W$, und
$\varphi_\Omega : \Omega_V \mathrm{x} \Omega_W \to T_\Omega$, $(v,w) \mapsto \varphi(v,w)$.
Dann ist $(T_\Omega, \varphi_\Omega)$ ein Tensorprodukt für Ω_V und Ω_W.

Bew.: (TP1): T_Ω ist ein Vektorraum.
(TP2): $\varphi_\Omega \in L^2(\Omega_V \mathrm{x} \Omega_W, T_\Omega)$ ist bilinear von $\Omega_V \mathrm{x} \Omega_W$ in T_Ω.
(TP3): Sei U ein Vektorraum und $f \in L^2(\Omega_V \mathrm{x} \Omega_W, U)$ bilinear. Man wähle eine bilineare Fortsetzung $F \in L^2(V\mathrm{x}W,U)$ von f: $f = F/_{\Omega_V \mathrm{x} \Omega_W}$. Dann existiert $F^* \in L(T,U)$ mit $F = F^* \circ \varphi$. Es folgt $f = F/_{\Omega_V \mathrm{x} \Omega_W} = (F^* \circ \varphi)/_{\Omega_V \mathrm{x} \Omega_W} = F^*/_{T_\Omega} \circ \varphi_\Omega$: f faktorisiert.
Bei einer zweiten Faktorisierung $f = h \circ \varphi_\Omega$ ist $h \circ \varphi_\Omega = F^*/_{T_\Omega} \circ \varphi_\Omega$; wegen der Minimalität von T_Ω und der Linearität der Abbildungen folgt $h = F^*/_{T_\Omega}$: f faktorisiert eindeutig. _/

Zwei Tensorprodukte für V und W unterscheiden sich genau um einen tensorprodukt-treuen Isomorphismus:

<u>Satz 6: Eindeutigkeit von Tensorprodukten</u> (GREUB 1967:9)

Sei (T,φ) ein Tensorprodukt für V und W; sei $\bar{T}$ ein Vektorraum, $\bar{\varphi}\in L(V\times W,\bar{T})$ bilinear. $(\bar{T},\bar{\varphi})$ ist genau dann ein Tensorprodukt für V und W, wenn ein Vektorraum-Isomor phismus $i \in L(T,\bar{T})$ existiert, der tensorprodukt-treu ist: $i\circ\varphi = \bar{\varphi}$.

<u>Bew.:</u> $\Rightarrow$: I. Es ist $\bar{\varphi}$ bilinear und (T,φ) ein Tensorprodukt, also existiert ein lineare Abbildung $i \in L(T,\bar{T})$ mit $\bar{\varphi} = i\circ\varphi$.

II. Es ist φ bilinear und $(\bar{T},\bar{\varphi})$ ein Tensorprodukt, also existiert eine lineare Abbildung $h \in L(\bar{T},T)$ mit $\varphi = h\circ\bar{\varphi}$.

III. Somit gilt

$\varphi(v,w) = h\circ\bar{\varphi}(v,w) = h\circ i\circ\varphi(v,w)$, und

$\bar{\varphi}(v,w) = i\circ\varphi(v,w) = i\circ h\circ\bar{\varphi}(v,w)$.

Da nach Satz 4:20 über die Minimalität von T die Bilder von φ bzw. $\bar{\varphi}$ die Räume T bzw. $\bar{T}$ aufspannen, gilt $h\circ i = id_T$, und $i\circ h = id_{\bar{T}}$. Aus der ersten Gleichung folgt, daß i injektiv und h surjektiv ist, aus der zweiten, daß h injektiv und i surjektiv ist:

i ist ein Vektorraum-Isomorphismus von T auf $\bar{T}$ mit $i\circ\varphi = \bar{\varphi}$; h ist die Umkehrung von i.

$\Leftarrow$: Sei $f\in L^2(V\times W,U)$ bilinear mit Werten in einem Raum U. f faktorisiert bzgl. $\bar{\varphi}$ wegen $f = f^*\circ i^{-1}\circ i\circ\varphi =: \bar{f}^*\circ\bar{\varphi}$. Gilt ebenfalls $f = h\circ\bar{\varphi}$, so folgt $f = h\circ\bar{\varphi} = h\circ i\circ\varphi$, $f^* = h\circ i$, $h = f^*\circ i^{-1}$: f faktorisiert eindeutig. _/

Wir erhalten das folgende kommutative Diagram:

$$\begin{tikzcd}
V\times W \arrow[rr, "f \text{ bil.}"] \arrow[dr, "\varphi"] \arrow[ddr, "\bar{\varphi}"'] & & U \\
& T \arrow[ur, "f^* \text{ lin.}"] \arrow[d, "i"] & \\
& \bar{T} \arrow[uur, "\bar{f}^* \text{ lin.}"'] &
\end{tikzcd}$$

Beim folgenden Existenzsatz beschränken wir uns auf die uns interessierenden Räume von endlicher Dimension; in diesen Fällen können Tensorprodukte in naheliegender Weise konstruiert werden:

<u>Satz 7: Existenz von Tensorprodukten</u> (LANG 1966:222f)

V habe die Dimension n , W die Dimension m .
Dann existiert ein Tensorprodukt für V und W .

<u>Bew.:</u> Es seien $(e^V_\nu)_{\nu=1,\dots,n} \in V^n$, $(e^W_\mu)_{\mu=1,\dots,m} \in W^m$ geordneten Basen von V bzw. von W .

I. (TP1) Konstruktion von T :

Sei S die Menge der Paare von Basisvektoren:

$S := \left\{ (e^V_\nu, e^W_\mu) \in V \times W : \nu \in \{1,\dots,n\} \wedge \mu \in \{1,\dots,m\} \right\}$.

Dann ist die Menge

$T := K^S$ aller Abbildungen von S in den Skalarenkörper K in natürlicher Weise ein Vektorraum mit der Dimension $\dim T = \# S = mn$, und mit der "natürlichen" geordneten Basis $((t_{\nu\mu})_{\mu=1,\dots,m})_{\nu=1,\dots,n}$, wobei die Abbildung $t_{\nu\mu} \in K^S$ definiert ist durch

$$t_{\nu\mu}(e^V_{\bar\nu}, e^W_{\bar\mu}) = \delta_{\nu\bar\nu}\delta_{\mu\bar\mu} = \begin{cases} 1_K & \text{falls } \nu = \bar\nu \text{ , } \mu = \bar\mu \\ 0_K & \text{sonst} \end{cases} .$$

II. (TP2) Konstruktion von φ :

Wegen der Auszeichnung von Basen kann man Vektoren v, w identifizieren mit ihren Koordinaten-Tupeln:

$v = (v_\nu)_{\nu=1,\dots,n}$, $w = (w_\mu)_{\mu=1,\dots,m}$. Definiert man damit

$\varphi : V \times W \to T$, $(v,w) \mapsto \Sigma\, v_\nu w_\mu t_{\nu\mu}$, dann ist φ bilinear:

$$\varphi(\alpha v + \bar\alpha \bar v, w) = \Sigma\, (\alpha v_\nu + \bar\alpha \bar v_\nu) w_\mu t_{\nu\mu} =$$
$$= \alpha\, \Sigma\, v_\nu w_\mu t_{\nu\mu} + \bar\alpha\, \Sigma\, \bar v_\nu w_\mu t_{\nu\mu} = \alpha\, \varphi(v,w) + \bar\alpha\, \varphi(\bar v, w) ;$$

die Linearität in der zweiten Komponente folgt analog.

Insbesondere ist $\varphi(e^V_\nu, e^W_\mu) = t_{\nu\mu}$.

III. (TP3) Nachweis der Faktorisierungseigenschaft:

Sei $f \in L^2(V \times W, U)$ bilinear auf $V \times W$ mit Werten in einem Vektorraum U . $f^* \in L(T,U)$ sei diejenige lineare Abbildung von T in U mit den Basisbildern $f^*(t_{\nu\mu}) = f(e^V_\nu, e^W_\mu)$.

Dann faktorisiert f : $f(v,w) = \Sigma\, v_\nu w_\mu f^*(t_{\nu\mu}) = f^* \circ \varphi(v,w)$.

Ist zudem $f = h \circ \varphi$, so gilt $h(t_{\nu\mu}) = f(e^V_\nu, e^W_\mu) = f^*(t_{\nu\mu})$, also $h = f^*$: f faktorisiert eindeutig. $\lrcorner$

Der im Beweis konstruierte Raum T hat die Dimension nm. Mit Satz 6:2) über die Eindeutigkeit von Tensorprodukten folgt sofort der

Satz 8: Dimension von Tensorprodukten

V, W seien von endlicher Dimension, (T, φ) ein Tensorprodukt für V und W.

Dann gilt $\dim T = \dim V \cdot \dim W$. _/

1.2 Tensorprodukt linearer Abbildungen

In diesem Abschnitt seien V_1, V_2, W_1, W_2 endlich-dimensionale Vektorräume.

$A \in L(V_1, V_2)$ sei eine lineare Abbildung von V_1 in V_2,

$B \in L(W_1, W_2)$ sei eine lineare Abbildung von W_1 in W_2.

(T_1, φ_1) sei eine Tensorprodukt für V_1 und W_1,

(T_2, φ_2) sei eines für V_2 und W_2.

Die Abbildung $f : V_1 x W_1 \to T_2$, $(v,w) \mapsto \varphi_2(Av, Bw)$ ist bilinear, also existiert genau eine lineare Abbildung $\psi_{21}(A,B) \in L(T_1, T_2)$ mit $\varphi_2(Av, Bw) = \psi_{21}(A,B)\ \varphi_1(v,w)$. Dies rechtfertigt die

Def. 2: (Tensorprodukt linearer Abbildungen)

Eine lineare Abbildung $\psi_{21}(A,B) \in L(T_1, T_2)$ von T_1 T_2 heißt "Tensorprodukt von A und B (bzgl. (T_1, φ_1) und (T_2, φ_2))", falls

$\psi_{21}(A,B) \circ \varphi_1(v,w) = \varphi_2(Av, Bw)$ $\forall v \in V_1$ $\forall w \in W_1$. _/

Wie erwähnt existiert das Tensorprodukt von A und B und ist eindeutig.

Die Zuordnung $(A,B) \mapsto \psi_{21}(A,B)$ definiert eine bilineare Abbildung ψ_{21} von $L(V_1, V_2) \times L(W_1, W_2)$ in $L(T_1, T_2)$. Sind die Räume V_1 und W_1 von endlicher Dimension (wie hier), dann ist $(L(T_1, T_2), \psi_{21})$ ein Tensorprodukt für $L(V_1, V_2)$ und $L(W_1, W_2)$, allgemein ist das nicht der Fall (GREUB 1967:23f).

Bei endlich-dimensionalen Räumen braucht man also nicht zu unterscheiden zwischen dem Tensorprodukt von A und B als Vektoren (Def. 1.ii:18) und als linearen Abbildungen und kann $\psi_{21}(A,B)$ einfach das "Tensorprodukt von A und B " nennen .

<u>Satz 9: Rang eines Tensorprodukts</u> (GREUB 1967:26)

Es ist $\operatorname{rk} \psi_{21}(A,B) = \operatorname{rk} A \,.\, \operatorname{rk} B$.

<u>Bew.:</u> I. Setze $T_\Omega := \operatorname{span} \varphi_2 \operatorname{im} A \times \operatorname{im} B$, und $\varphi_\Omega : \operatorname{im} A \times \operatorname{im} B \to T_\Omega$, $(v,w) \mapsto \varphi_2(v,w)$.
Dann ist nach Satz 5:20 über Tensorprodukte für Teilräume das Paar $(T_\Omega, \varphi_\Omega)$ ein Tensorprodukt für $\operatorname{im} A$ und $\operatorname{im} B$.
Für die Dimension gilt mit Satz 8:23
$\dim T_\Omega = \dim \operatorname{im} A \,.\, \dim \operatorname{im} B = \operatorname{rk} A \,.\, \operatorname{rk} B$.
II. Da das Bild von φ_1 den Raum T_1 aufspannt (Satz 4:20), gilt:

$$\operatorname{im} \psi_{21}(A,B) = \psi_{21}(A,B)\, T_1 = \psi_{21}(A,B) \operatorname{span} \varphi_1 V\times W =$$
$$= \operatorname{span} \psi_{21}(A,B)\circ\varphi_1 V\times W = \operatorname{span} \varphi_2 \operatorname{im} A \times \operatorname{im} B = T_\Omega \;.$$

III. Zusammen ergibt das

$$\operatorname{rk} \psi_{21}(A,B) = \dim \operatorname{im} \psi_{21}(A,B) \underset{\mathrm{II}}{=} \dim T_\Omega \underset{\mathrm{I}}{=} \operatorname{rk} A \,.\, \operatorname{rk} B \;. \qquad \_/$$

<u>Satz 10: Tensor von Produkten</u> (GREUB 1967:24)

Seien V_3, W_3 weitere endlich-dimensionale Vektorräume und (T_3,φ_3) ein Tensorprodukt für V_3 und W_3 . Seien $\bar{A}\in L(V_2,V_3)$, $\bar{B}\in L(W_2,W_3)$ weitere lineare Abbildungen.
Dann gilt $\psi_{31}(\bar{A}A,\bar{B}B) = \psi_{32}(\bar{A},\bar{B})\circ\psi_{21}(A,B)$.

<u>Bew.:</u> Es ist zu zeigen, daß die rechte Seite die $\psi_{31}(\bar{A}A,\bar{B}B)$ definierende Gleichung löst:

$$\psi_{32}(\bar{A},\bar{B})\circ\psi_{21}(A,B)\circ\varphi_1(v,w) = \psi_{32}(\bar{A},\bar{B})\circ\varphi_2(Av,Bw) = \varphi_3(\bar{A}Av,\bar{B}Bw) \;. \qquad \_/$$

1.3 Das KRONECKERprodukt

Wir wenden nun die gerade hergeleiteten Ergebnisse auf EUKLIDische Räume an: Das KRONECKERprodukt $\otimes$, genauer: $(E^{nm},\otimes)$, ist ein Tensorprodukt für E^n und E^m .

In Satz 7:22 über die Existenz von Tensorprodukten haben wir auf "natürliche Weise" Tensorprodukte konstruiert. Diese Konstruktionen waren abhängig von Basen für V und W . Wir wiederholen nun dieses Vorgehen, indem wir für die Räume E^n und E^m die EUKLIDischen Basen $(e^n_\nu)_{\nu=1,\dots,n}$ und $(e^m_\mu)_{\mu=1,\dots,m}$ zugrundelegen:

$S = \left\{ (e^n_\nu, e^m_\mu) \in E^n x E^m : \nu \in \{1,\dots,n\} \wedge \mu \in \{1,\dots,m\} \right\}$, $T = R^S$,

$\varphi(x,y) = \Sigma\, x_\nu y_\mu t_{\nu\mu}$.

Der Raum T hat die Dimension nm , also existiert ein Isomorphismus in den E^{nm} . Ein solcher Isomorphismus i wird in natürlicher Weise gegeben, indem man die Abbildung $t_{\nu\mu}$ mit dem nm-Tupel ihrer Funktionswerte identifiziert:

$i(t_{\nu\mu}) = ((\, t_{\nu\mu}(e^n_{\bar\nu}, e^m_{\bar\mu})\,)_{\bar\mu=1,\dots,m})_{\bar\nu=1,\dots,n}^T \in R^{nm}$;

diesen nm-Vektor kann man als Blockvektor schreiben[7]:

$$i(t_{\nu\mu}) = \begin{bmatrix} (e^n_\nu)_1 e^m_\mu \\ \vdots \\ (e^n_\nu)_n e^m_\mu \end{bmatrix}$$

. Nach Satz 6:21 ist $(E^{nm}, i\circ\varphi)$ ein Tensorprodukt für E^n und E^m , es heißt das "<u>KRONECKERprodukt</u> von E^n und E^m " und wird notiert mit $(E^{nm},\otimes)$.

<u>Satz 11: Das KRONECKERprodukt EUKLIDischer Räume</u>

Das KRONECKERprodukt $\otimes : E^n \times E^m \to E^{nm}$,

$$(x,y) \mapsto x\otimes y = \begin{bmatrix} x_1 y \\ \vdots \\ x_n y \end{bmatrix}$$

, ist ein Tensorprodukt für E^n und E^m .

<u>Bew.:</u> Es bleibt zu zeigen, daß $x\otimes y$ die angegebene Darstellung hat:

---

7 Die eckigen Klammern von Vektoren bzw. Matrizen sollen daran erinnern, daß die Einträge selbst wieder Vektoren bzw. Matrizen sind.

$$x\otimes y = i\circ\varphi(x,y) = \Sigma\, x_\nu y_\mu . i(t_{\nu\mu}) = \Sigma\, x_\nu y_\mu \begin{bmatrix} (e^n_\nu)_1 e^m_\mu \\ \vdots \\ (e^n_\nu)_n e^m_\mu \end{bmatrix} =$$

$$= \begin{bmatrix} \Sigma\, x_\nu y_\mu . (e^n_\nu)_1 e^m_\mu \\ \vdots \\ \Sigma\, x_\nu y_\mu . (e^n_\nu)_n e^m_\mu \end{bmatrix} = \begin{bmatrix} x_1 y \\ \vdots \\ x_n y \end{bmatrix} .$$

Für Matrizen $A\in M(n,m)=L(E^m,E^n)$, $B\in M(k,l)=L(E^l,E^k)$ induziert das KRONECKERprodukt eine lineare Abbildung $\psi(A,B)$ von $T_1 = E^{ml}$ in $T_2 = E^{nk}$. Die Matrix dieser linearen Abbildung heißt das <u>KRONECKERprodukt</u> $A\otimes B$ der Matrizen A und B . $A\otimes B$ ist also die eindeutige Lösung der Matrizengleichung

$A\otimes B.x\otimes y = Ax\otimes By \quad \forall x\in E^m\ \forall y\in E^l$.

<u>Satz 12: Das KRONECKERprodukt von Matrizen</u>

i (Darstellung mit Blockmatrizen)

Für jede (n,m)-Matrix A und jede (k,l)-Matrix B

gilt: $A\otimes B = \begin{bmatrix} A_{11}B & A_{12}B & \dots & A_{1m}B \\ \vdots & \vdots & & \vdots \\ A_{n1}B & A_{n2}B & \dots & A_{nm}B \end{bmatrix} \in M(nk,ml)$.

ii (Rang eines KRONECKERprodukts)

Für jede (n,m)-Matrix A und jede (k,l)-Matrix B

gilt: $\mathrm{rk}\, A\otimes B = \mathrm{rk}\, A \,.\, \mathrm{rk}\, B$.

iii (Produkt von KRONECKERprodukten[8])

Für (n_3,n_2)-Matrizen $\bar{A}$, (n_2,n_1)-Matrizen A ,

für (m_3,m_2)-Matrizen $\bar{B}$, (m_2,m_1)-Matrizen B gilt:

$\bar{A}\otimes\bar{B} \,.\, A\otimes B = \bar{A}A\otimes\bar{B}B$.

iv (Transposition eines KRONECKERprodukts[9])

Für jede (n,m)-Matrix A und jede (k,l)-Matrix B

gilt: $(A\otimes B)^T = A^T\otimes B^T$.

v (KRONECKERprodukt von Projektionen)

Für Projektionsmatrizen $M\in\mathrm{Proj}(m)$, $N\in\mathrm{Proj}(n)$

gilt: $M\otimes N \in \mathrm{Proj}(mn)$.

<u>Bew.:</u> i. Es ist zu zeigen, daß die rechte Seite der Gleichung die $A\otimes B$ definierende Gleichung erfüllt:

---

8 Wir benutzen die folgenden Regeln zur Klammerersparnis:
$A\otimes B.C := (A\otimes B)C$; $C.A\otimes B := C(A\otimes B)$;
$A\otimes BC := A\otimes (BC)$; $CA\otimes B := (CA)\otimes B$.

9 Entsprechende Aussagen gelten allgemein für Tensorprodukte auf Inneren-Produkt-Räumen (GREUB 1967:34) .

$$\begin{bmatrix} A_{11}B & \dots & A_{1m}B \\ \vdots & & \vdots \\ A_{n1}B & \dots & A_{nm}B \end{bmatrix} \begin{bmatrix} x_1 y \\ \vdots \\ x_m y \end{bmatrix} = \begin{bmatrix} \Sigma\ A_{1\mu} x_\mu By \\ \vdots \\ \Sigma\ A_{n\mu} x_\mu By \end{bmatrix} = \begin{bmatrix} (Ax)_1 By \\ \vdots \\ (Ax)_n By \end{bmatrix} = Ax \otimes By \ .$$

ii folgt aus Satz 9:24 über den Rang eines Tensorprodukts.
iii ist Satz 10:24 über das Tensorprodukt von Produkten.
iv ersieht man aus der in i gegebenen Darstellung.
v. $(M \otimes N)^T = M \otimes N$ nach iv , $(M \otimes N)^2 = M \otimes N$ nach iii . _/

Das KRONECKERprodukt von Matrizen ist eine konsistente Erweiterung des KRONECKERprodukts von Vektoren (Satz 11:25; Satz 12.i:26).

1.4 Die Funktion vec

Der Matrizenraum $M(n,m)$ hat wie der E^{nm} die Dimension nm und läßt sich ebenfalls zu einem Tensorprodukt für E^n und E^m machen.

Als Isomorphismus gemäß Satz 7:22 wähle man etwa j , definiert durch $j(t_{\nu\mu}) = e^n_\nu (e^m_\mu)^T$; dann ist $j \circ \varphi(x,y) = xy^T$:
$(\ M(n,m), xy^T\ )$ ist ein Tensorprodukt für E^n und E^m .

Nach Satz 6:21 existiert ein tensorprodukt-treuer Isomorphismus von $M(n,m)$ auf E^{nm} . Dieser Isomorphismus macht aus einer Matrix einen Vektor und bekommt deshalb den Namen "vec". Aus der Tensorprodukt-Treue $\operatorname{vec} xy^T = x \otimes y$ berechnet man für (n,m)-Matrizen A :

$$\begin{aligned} \operatorname{vec} A &= \Sigma\ A_{\nu\mu} \operatorname{vec} E^{(n,m)}_{\nu\mu} = \Sigma\ A_{\nu\mu} \operatorname{vec} e^n_\nu (e^m_\mu)^T = \Sigma\ A_{\nu\mu} . e^n_\nu \otimes e^m_\mu = \\ &= ((A_{\nu\mu})_{\mu=1,\dots,m})^T_{\nu=1,\dots,n} = \\ &= \text{Zeile an Zeile, dann Transposition .} \end{aligned}$$

Es gilt

$$\operatorname{vec} Axy^T C^T = \operatorname{vec} Ax(Cy)^T = Ax \otimes Cy = A \otimes C . x \otimes y = A \otimes C . \operatorname{vec} xy^T \ .$$

Daraus folgt für das Produkt dreier Matrizen A, B, C :

$$\begin{aligned} \operatorname{vec} ABC &= \Sigma\ B_{\mu\varkappa} . \operatorname{vec} AE^{(m,k)}_{\mu\varkappa} C = \Sigma\ B_{\mu\varkappa} . A \otimes C^T . \operatorname{vec} E^{(m,k)}_{\mu\varkappa} = \\ &= A \otimes C^T . \operatorname{vec} B \ . \end{aligned}$$

Das von vec auf dem Raum $M(n,m)$ induzierte innere Produkt ist

$$(\operatorname{vec} A)^T \operatorname{vec} B = \Sigma\ A_{\nu\mu} B_{\nu\mu} = \Sigma\ A_{\nu\mu} (B^T)_{\mu\nu} = \Sigma\ (AB^T)_{\nu\nu} = \operatorname{tr} AB^T \ .$$

vec ist also eine Isometrie von $(M(n,m), \mathrm{tr}\, AB^T)$ auf E^{nm}.

Die zum inneren Produkt $\mathrm{tr}\, AB^T$ gehörige Norm
$\|A\| = (\mathrm{tr}\, AA^T)^{\frac{1}{2}} = (\Sigma\, A_{\nu\mu}^{\;2})^{\frac{1}{2}}$ ist eine konsistente Erweiterung der EUKLIDischen Vektornorm und heißt "EUKLIDische Matrixnorm".

Für (n,m)-Matrizen A und (m,k)-Matrizen B gilt bei Verwendung der CAUCHY-SCHWARZschen Ungleichung:

$$\|AB\|^2 = \sum_{\nu=1}^{n} \sum_{\varkappa=1}^{k} \Big(\sum_{\mu=1}^{m} A_{\nu\mu} B_{\mu\varkappa} \Big)^2 \leqq$$
$$\leqq \sum_{\nu=1}^{n} \sum_{\varkappa=1}^{k} \Big(\sum_{\mu=1}^{m} A_{\nu\mu}^{\;2} \Big) \Big(\sum_{\mu=1}^{m} B_{\mu\varkappa}^{\;2} \Big) = \|A\|^2 . \|B\|^2 .$$

(Diese Abschätzung werden wir regelmäßig benutzen, um die Gütekriterien für Schätzfunktionen zu rechtfertigen.)

Fassen wir die obigen Ergebnisse in einem Satz zusammen:

<u>Satz 13: Die Funktion vec</u> [10]

i (vec tensorprodukt-treu)
$(M(n,m), xy^T)$ ist ein Tensorprodukt für E^n und E^m.
vec ist tensorprodukt-treuer Isomorphismus von $(M(n,m), xy^T)$ auf $(E^{nm}, \otimes)$:
$\mathrm{vec}\; xy^T = x \otimes y$.

ii (vec von Produkten)
Für drei Matrizen $A \in M(n,m)$, $B \in M(m,k)$, $C \in M(k,l)$ gilt:
$\mathrm{vec}\; ABC = A \otimes C^T . \mathrm{vec}\; B$.

iii (vec isometrisch)
$(M(n,m), \mathrm{tr}\, AB^T)$ ist ein Innerer-Produkt-Raum.
vec ist innerer- produkt-treuer Isomorphismus von $(M(n,m), \mathrm{tr}\, AB^T)$ auf $(E^{nm}, x^T y)$:
$\mathrm{tr}\, AB^T = (\mathrm{vec}\; A)^T \mathrm{vec}\; B$.

iv (Standardabschätzung)
Für jede (n,m)-Matrix A und (m,k)-Matrix B gilt:
$\|AB\| \leqq \|A\| . \|B\|$. _/

In der abschließenden Übersichtstabelle sind diejenigen Ergebnisse zusammengestellt, von denen wir im statistischen Teil der Arbeit Gebrauch machen werden.

---

10 Entsprechende Aussagen gelten allgemein für Tensorprodukte auf endlich-dimensionalen Inneren-Produkt-Räumen (GREUB 1967:35-39).

Übersichtstabelle: Tensorprodukte für E^n und E^m	
$(E^{nm}, \otimes)$ $(M(n,m), xy^T)$	Tensorprodukte für E^n und E^m
$\mathrm{vec}\, A =$ = Zeile an Zeile, dann Transposition	Vektorraum-Isomorphismus tensorprodukt-treu: $\mathrm{vec}\, xy^T = x \otimes y$ isometrisch: $\mathrm{tr}\, AB^T = (\mathrm{vec}\, A)^T \mathrm{vec}\, B$ $\mathrm{vec}\, ABC = A \otimes C^T . \mathrm{vec}\, B$
$\lVert A \rVert = (\Sigma\, A_{\nu\mu}^{2})^{\frac{1}{2}}$	$\lVert AB \rVert \leqq \lVert A \rVert . \lVert B \rVert$

$$
\begin{tikzcd}
E^n \times E^m \arrow[rr, "f \text{ bilinear}"] \arrow[dr, "xy^T"] \arrow[ddr, "\otimes"'] & & U \\
& M(n,m) \arrow[ur, "f^* \circ \mathrm{vec}"] \arrow[d, "\mathrm{vec}"] & \\
& E^{nm} \arrow[uur, "f^* \text{ linear}"'] &
\end{tikzcd}
$$

Dieses Diagram kommutiert.

* * *

Kap. 2: Pseudoinversion

In diesem Kapitel leiten wir die elementaren Eigenschaften von Pseudoinversen A^+ her: Existenz und Eindeutigkeit, Darstellungen in Spezialfällen und Rechenregeln (2.1).
Im zweiten Abschnitt (2.2) entwickeln wir eine Formel für $X^T(S^2+XX^T)^+$.

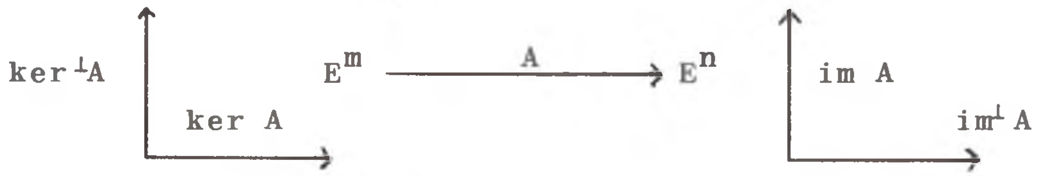
2.1 MOORE-PENROSE-Pseudoinversion

MOORE (1920) und PENROSE (1955) verallgemeinerten die Inversion regulärer Matrizen und führten für beliebige (n,m)-Matrizen A die "pseudoinverse" (m,n)-Matrix A^+ ein[11]. Dieser (MOORE-PENROSE-)Pseudoinversion entspricht die folgende geometrische Anschauung:

Interpretieren wir die (n,m)-Matrix A als lineare Abbildung vom E^m in den E^n . Kern und Bild von A zerlegen den Definitions- bzw. den Zielraum:

$E^m = \ker A \oplus \ker^{\perp} A$,

$E^n = \operatorname{im} A \oplus \operatorname{im}^{\perp} A$.



Bei regulärer Inversion ergibt AA^{-1} die Identität des Zielraumes; für beliebige Matrizen A wird man "höchstens" fordern können, daß AA^+ zur "Identität" auf dem Bildraum von A führt:

$AA^+ = \operatorname{proj}(./\operatorname{im} A)$.

Dann muß notwendig gelten[12]:

(PK1) AA^+ symmetrisch, und

(PK3) $AA^+A = A$.

---

11 Zur Vereinbarkeit beider Definitionen siehe RADO (1956).- Die heute übliche Formulierung, elegant, einprägsam und handlich, stammt von R. PENROSE.

12 PK = PENROSE-Kriterium

Dies ist aber auch hinreichend für $AA^+ = \text{proj}(./\text{im } A)$.
Denn mit (PK3) ist AA^+ idempotent: $(AA^+)^2 = AA^+A.A^+ = AA^+$.
Da AA^+ symmetrisch ist (PK1), ist AA^+ eine Projektion.
Für deren Bild gilt:

$$\text{im } AA^+ \subset \text{im } A \underset{PK3}{=} \text{im } AA^+A \subset \text{im } AA^+ \text{ , im } AA^+ = \text{im } A \text{ .}$$

Zusammen gibt das: $AA^+ = \text{proj}(./\text{im } AA^+) = \text{proj}(./\text{im } A)$.

Bei regulärer Inversion ergibt $A^{-1}A$ die Identität des Definitionsraumes von A , d.h. des Zielraumes von A^{-1} ; für beliebige Matrizen A wird man "höchstens" fordern können, daß A^+A zur "Identität" auf dem Bildraum von A^+ führt:

$A^+A = \text{proj}(./\text{im } A^+)$;

dabei sollte das Bild von A^+ zusammenfallen mit dem orthogonalen Komplement des Kernes von A :

(*) $\text{im } A^+ = \ker^{\perp} A$.

Die Forderung: $A^+A = \text{proj}(./\text{im } A^+)$ ist wie oben äquivalent zu

(PK2) A^+A symmetrisch, und
(PK4) $A^+AA^+ = A^+$.

Die Eigenschaft (*) wird von den PENROSE-Kriterien schon impliziert. Wegen Satz 1.i:15 ist zu zeigen: $\text{im } A^+ = \text{im } A^T$:

$$\text{im } A^+ \underset{PK4}{=} \text{im } A^+AA^+ \underset{PK2}{=} \text{im } (A^+A)^TA^+ = \text{im } A^TA^{+T}A^+ \subset \text{im } A^T =$$

$$\underset{PK3}{=} \text{im } (AA^+A)^T = \text{im } (A^+A)^TA^T \underset{PK2}{=} \text{im } A^+AA^T \subset \text{im } A^+ \text{ .}$$

Zu jeder Matrix A haben die PENROSE-Kriterien (PK1,2,3,4) genau eine Lösung A^+ : die "Pseudoinverse von A ". Wir zeigen zunächst, daß je drei der PENROSE-Kriterien keine eindeutige Lösung bestimmen.

Fall (PK1,2,3): $A = 0 \in \mathbb{R}$, $A^+ = \alpha \in \mathbb{R}$.
Dann gilt $AA^+ = 0 = A^+A$, und $AA^+A = 0 = A$.

Fall (PK1,2,4): $A = 1 \in \mathbb{R}$, $A^+ = \alpha \in \{0,1\}$.
Dann gilt $AA^+ = \alpha = A^+A$, und $A^+AA^+ = \alpha^2 = A^+$.

Fall (PK1,3,4): $A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, $A^+ = \begin{pmatrix} 1 & 0 \\ \alpha & 0 \end{pmatrix}$, $\alpha \in \mathbb{R}$.
Dann gilt $AA^+ = A$ symmetrisch, $AA^+A = A^2 = A$, und $A^+AA^+ = A^+$.

Fall (PK2,3,4): $A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, $A^+ = \begin{pmatrix} 1 & \alpha \\ 0 & 0 \end{pmatrix}$, $\alpha \in R$.

Dann gilt $A^+A = A$ symmetrisch, $AA^+A = A^2 = A$, und $A^+AA^+ = A^+$.

Wir beweisen nun die eindeutige Existenz von A^+ [13]:

Satz 14: Die PENROSE-Kriterien

(PENROSE 1955:406) Zu jeder (n,m)-Matrix A existiert genau eine (m,n)-Matrix A^+ , die die vier PENROSE-Kriterien erfüllt:

(PK1) $AA^+ \in \mathrm{Sym}(n)$,

(PK2) $A^+A \in \mathrm{Sym}(m)$,

(PK3) $AA^+A = A$,

(PK4) $A^+AA^+ = A^+$.

Bew.: I. (ALBERT 1972:29) Einzigkeit:

Erfülle auch $B \in M(m,n)$ die PENROSE-Kriterien. Dann gilt:

$B = BAB$	
$= (ABB^T)^T$	weil AB sym.
$= (AA^+ABB^T)^T$	$A = AA^+A$
$= (AA^+(BAB)^T)^T$	AB sym.
$= (AA^+B^T)^T$	$B = BAB$
$= BAA^+$	AA^+ sym., Halbzeit
$= (BA)^TA^+$	BA sym.
$= (AA^+A)^TB^TA^+$	$A = AA^+A$
$= A^+AA^TB^TA^+$	A^+A sym.
$= A^+ABAA^+$	BA sym.
$= A^+AA^+$	$ABA = A$
$= A^+$	

II. PENROSE (1955:408) erwähnt den folgenden Existenzbeweis, der ganz im Rahmen der linearen Algebra bleibt . Wir halten die einzelnen Schritte gleich in einem Satz fest:

---

13 RAO & MITRA (1972) gehen anders vor. Sie stellen sich ein Problem und führen zu dessen Lösung "generalisierte Inverse" ein. Zum Beispiel erfüllt die allgemeinste Form A^- solcher generalisierter Inversen gerade (PK3): $AA^-A = A$, und dient zur Lösung linearer Gleichungen $Ax = b$ (20f). Ein solches Konzept kann minimalistisch genannt werden: es werden nicht mehr Eigenschaften gefordert als notwendig.
Doch drohen meines Erachtens auch Nachteile: Ergebnisse mit generalisierten Inversen lassen sich nicht im selben Maße mit geometrischen Plausibilitätserklärungen stützen wie solche mit MOORE-PENROSE-Pseudoinversen. Zudem bedarf ein minimalistisches Konzept eines umfangreichen Begriffsapparats, der schon von selbst einen Schwerpunkt setzt; Pseudoinversion ist aber häufig "nur" ein Hilfsmittel, und das sollte auch im Aufwand erkennbar sein.

<u>Satz 15: Spezialfälle</u> (PENROSE 1955:408)

i Für reelle Zahlen $\alpha \in R = M(1,1)$ gilt:

$$\alpha^+ = \begin{cases} \alpha^{-1} & \text{falls } \alpha \neq 0 \\ 0 & \text{falls } \alpha = 0 \end{cases}$$

ii Für Diagonalmatrizen $D \in Diag(n)$ gilt:

$D^+ = Diag\,(D_{\nu\nu}^{\ +})_{\nu=1,\dots,n} \in Diag(n)$

iii Für symmetrische Matrizen $S \in Sym(n)$ gilt:

$S^+ = R^T(RSR^T)^+R \in Sym(n)$ falls $RS^TR \in Diag(n)$, $R \in O(n)$

iv[14] Für jede (n,m)-Matrix A gilt:

$A^+ = A^T(AA^T)^+ = (A^TA)^+A^T$.

<u>Bew.:</u> i. Das angegebene α^+ erfüllt die PENROSE-Kriterien.
ii. Das angegebene D^+ erfüllt die PENROSE-Kriterien.
iii. Sei R eine orthogonale Matrix, die S diagonalisiert:
$D := RSR^T \in Diag(n)$. Dann ist $S = R^TDR$ und die Matrix
$S^+ = R^TD^+R \in Sym(n)$ erfüllt die PENROSE-Kriterien.
SS^+ ist symmetrisch und idempotent, außer dem gilt
$\mathrm{im}\, SS^+ \subset \mathrm{im}\, S = \mathrm{im}\, SS^+S \subset \mathrm{im}\, SS^+$; dies gibt
$SS^+ = \mathrm{proj}(./\mathrm{im}\, SS^+) = \mathrm{proj}(./\mathrm{im}\, S)$.
iv. $A^T(AA^T)^+$ erfüllt die PENROSE-Kriterien[15]:
(PK1) $A.A^T(AA^T)^+ \in Sym(n)$ nach iii,
(PK2) $A^T(AA^T)^+.A \in Sym(m)$ nach iii,
(PK3) $A.A^T(AA^T)^+.A = \mathrm{proj}(A/\mathrm{im}\, AA^T) = \mathrm{proj}(A/\mathrm{im}\, A) = A$,
(PK4) $A^T\underbrace{(AA^T)^+.A.A^T(AA^T)^+}_{} \overset{iii}{=} A^T.(AA^T)^+$ [i,ii:15] .
Genauso zeigt man, daß $(A^TA)^+A^T$ die PENROSE-Kriterien erfüllt, zum Beispiel
(PK3) $A.(A^TA)^+A^T.A = (\ A.(A^TA)^+A^T.A\ )^{TT} = (\ A^TA(A^TA)^+A^T\ )^T =$
$= (\ \mathrm{proj}(A^T/\mathrm{im}\, A^TA)\ )^T = (\ \mathrm{proj}(A^T/\mathrm{im}\, A^T\ )^T = A^{TT} = A$.

Damit ist auch Satz 14 bewiesen. _/

Die Formeln $A^+ = A^T(AA^T)^+ = (A^TA)^+A^T$ werden wir häufig benutzen, sie geben die Pseudoinverse A^+ an über die GRAMschen Matrizen AA^T, A^TA .

---

14 Man kann die Pseudoinverse A^+ definieren als Limes
$A^+ = \lim_{\delta \to 0} A^T(AA^T+\delta^2 I_n)^{-1} = \lim_{\delta \to 0} (A^TA+\delta^2 I_m)^{-1}A^T$ (ALBERT 1972:19f) .
Für funktionalanalytische Methoden siehe zum Beispiel (BEUTLER 1965:459f).
15 Wenn es der Klarheit dient, werden in Beweisen diejenigen Aussagenteile <u>unterstrichen</u>, die von einem Schritt zum nächsten gerade Gegenstand der Betrachtung sind.

Satz 16: Weitere Spezialfälle

i Für Vektoren $a \in R^n$ gilt:

$$a^+ = \begin{cases} \frac{1}{\|a\|^2} \cdot a^T & \text{falls } a \neq 0 \\ 0 & \text{falls } a = 0 \end{cases} .$$

ii Für Projektionsmatrizen M gilt:

$M^+ = M$.

iii Für reguläre Matrizen $G \in GL(n)$ gilt:

$G^+ = G^{-1}$.

iv Für jede (n,m)-Matrix A gilt:

$$A^+ = \begin{cases} A^T(AA^T)^{-1} & \text{falls } \operatorname{rk} A = n \\ (A^TA)^{-1}A^T & \text{falls } \operatorname{rk} A = m \end{cases} .$$

v Für jede (n,m)-Matrix A und jede (k,l)-Matrix B gilt:

$(A \otimes B)^+ = A^+ \otimes B^+$. _/

Aussage 16.iii zeigt, daß die Pseudoinversion eine Verallgemeinerung der Inversion regulärer Matrizen ist.

Satz 17: Rechenregeln (PENROSE 1955:408)

Für alle Matrizen $A \in M(n,m)$, Projektionsmatrizen $M \in Proj(n)$, symmetrische Matrizen $S \in Sym(n)$, Skalare $\alpha \in R$ gilt:

i ($T \circ + = + \circ T$)

$A^{T+} = A^{+T}$,

ii (Pseudoinverse einer Pseudoinversen)

$A^{++} = A$,

iii (Skalare Multiplikation)

$(\alpha A)^+ = \alpha^+ A^+$,

iv (Spezielle Produkte)

$(A^TA)^+ = A^+A^{T+}$, $S^{2+} = S^{+2}$,

v (Produkt mit einer Projektionsmatrix)

$(A^TM)^+ = M(A^TM)^+$, $(MA)^+ = (MA)^+M$,

vi (Spezielle Produkte)

$SS^+ = S^+S$.

Bew.: i - v: Verifizieren der PENROSE-Kriterien.

vi. $SS^+ = (SS^+)^T = S^{+T}S^T = S^{T+}S = S^+S$. _/

Die Formeln 17.v werden wir häufig benutzen: Bei der Pseudoinversion von Produkten kann eine Projektion M "auf der anderen Seite" herausgezogen bzw. vernachlässigt werden.

Wir beschließen diesen Abschnitt mit den eingangs erwähnten Projektionseigenschaften von Pseudoinversen:

Satz 18: Projektionseigenschaften, Kern und Bild

(vgl. PENROSE 1955:408)

Für jede (n,m)-Matrix A gilt:

i (Projektionseigenschaften)

$$\begin{aligned} AA^{+} &= \mathrm{proj}(./\mathrm{im}\ A) \quad , \\ I_n - AA^{+} &= \mathrm{proj}(./\ker A^{T}) \quad , \\ A^{+}A &= \mathrm{proj}(./\mathrm{im}\ A^{T}) \quad , \\ I_m - A^{+}A &= \mathrm{proj}(./\ker A) \quad ; \end{aligned}$$

ii (Bild und Kern)

$$\begin{aligned} \mathrm{im}\ A^{+} &= \mathrm{im}\ A^{T} \quad , \\ \mathrm{im}\ A &= \mathrm{im}\ A^{T+} \quad , \\ \ker A^{+} &= \ker A^{T} \quad , \\ \ker A &= \ker A^{+}A = \ker (A^{T}A)^{+} \quad . \end{aligned}$$

Bew.: i. Diese Aussagen wurden im Text auf Seite 31 bewiesen.

ii. I. Siehe (*) auf Seite 31 .

II. Setze in I A^T statt A .

III. $\ker A^{T} \underset{1.i:15}{=} \mathrm{im}^{\perp} A \underset{II}{=} \mathrm{im}^{\perp} A^{T+} \underset{17.i:34}{=} \mathrm{im}^{\perp} A^{+T} \underset{1.i:15}{=} \ker A^{+}$.

IV. Aus $A^{+}A = \mathrm{proj}(./\mathrm{im}\ A^{T})$ folgt $\mathrm{im}\ A^{+}A = \mathrm{im}\ A^{T}$, damit:

$\ker A = \mathrm{im}^{\perp} A^{T} = \mathrm{im}^{\perp} A^{+}A = \ker (A^{+}A)^{T} = \ker A^{+}A$.

Und mit $\mathrm{im}\ A^{T} = \mathrm{im}\ A^{T}A$ (1.ii:15): PK2

$\ker A = \mathrm{im}^{\perp} A^{T} = \mathrm{im}^{\perp} A^{T}A \underset{II}{=} \mathrm{im}^{\perp}(A^{T}A)^{+T} = \ker (A^{T}A)^{+}$. _/

2.2 Eine Darstellung für $X^{T}(S^{2}+XX^{T})^{+}$

Im statistischen Teil der Arbeit werden verschiedene Risiken betrachtet. Um das GAUSSsche Risiko (mean square error) einordnen zu können, muß $X^{T}(S^{2}+XX^{T})^{+}$ anders dargestellt werden. Dazu wird in drei Schritten vorgegangen:

- In Satz 19 werden Blockmatrizen $[U\ V]$ pseudoinvertiert,
- in Satz 20 werden Summen $UU^{T}+VV^{T}$ pseudoinvertiert und
- in Satz 21 die gewünschte Darstellung hergeleitet.

Es wird wiederholt benutzt, daß die Summe einer p.d. und einer p.s.d. Matrix wieder p.d. und damit invertierbar ist.

<u>Satz 19: Pseudoinversion von Blockmatrizen</u> (CLINE 1964:594-596)

Sei U eine (n,m)-Matrix, V eine (n,k)-Matrix.

Setze $C := (I_n-UU^+)V \in M(n,k)$, und

$K := (I_k + (I_k-C^+C)V^TU^{T+}U^+V(I_k-C^+C)\ )^{-1} \in GL(k)$.

Dann gilt

$$[U\ V]^+ = \begin{bmatrix} U^+(I_n-VC^+) - U^+V(I_k-C^+C)KV^T\ U^{T+}U^+(I_n-VC^+) \\ C^+ + (I_k-C^+C)KV^TU^{T+}U^+(I_n-VC^+) \end{bmatrix} .$$

<u>Bew.:</u> I. Für $L \in M(k,n)$ wird definiert

$$X_L := \begin{bmatrix} U^+(I_n-VC^+) - U^+V(I_k-C^+C)L \\ C^+ + (I_k-C^+C)L \end{bmatrix} \in M(m+k,n) \quad ;$$

setzt man $A := [U\ V]$, und $L_o := KV^TU^{T+}U^+(I_n-VC^+)$,

dann lautet die Behauptung:

$A^+ = X_{L_o}$.

Für beliebige (k,n)-Matrizen L ist AX_L symmetrisch (II) und $AX_LA = A$ (V) . Für gewisse L gilt $X_LAX_L = X_L$ (VI), L_o hat die geforderte Eigenschaft (VII). Für gewisse L wird X_LA symmetrisch (VIII), L_o erfüllt die drei in VIII geforderten Bedingungen (X, XI, XII). Damit sind die PENROSE-Kriterien verifiziert: X_{L_o} ist die Pseudoinverse von A .

Sei im folgenden L eine beliebige (k,n)-Matrix.

II. $AX_L = UU^++CC^+ \in Sym(n)$; denn:

$$AX_L = [U\ V]\begin{bmatrix} U^+(I_n-VC^+) - U^+V(I_k-C^+C)L \\ C^+ + (I_k-C^+C)L \end{bmatrix} =$$

$$= UU^+ \underline{-\ UU^+VC^+} - UU^+V(I_k-C^+C)L \underline{+\ VC^+} + V(I_k-C^+C)L =$$

$$= UU^+ + \underline{(I_n-UU^+)V}C^+ + \underline{(I_n-UU^+)V}(I_k-C^+C)L =$$

$$= UU^+ + CC^+ + \underline{C(I_k-C^+C)}L =$$

$$= UU^+ + CC^+ + 0 .$$

III. $C^+U = 0$, und $U^+C = 0$; denn:

$$C^+U = (\ \underline{(I_n-UU^+)V}\ )^+U \underset{17.v:34}{=} (\ (I_n-UU^+)V\ )^+\underline{(I_n-UU^+)U} = 0 ,$$

$$C^+U = 0 \Leftrightarrow \forall x \in \mathrm{im}\ U \underset{18.ii:35}{=} \mathrm{im}\ U^{+T} : C^+x = 0$$

$$\Leftrightarrow C^+U^{+T} = 0$$

$$\Leftrightarrow U^+\underline{C^{+T}} = 0^T = 0$$

$$\Leftrightarrow U^+C = 0 \quad .$$

IV. $C^+V = C^+C$; denn mit 17.v:34 (wie in III):

$$C^+V = ((\underline{I_n-UU^+})V\ )^+V = (\ (I_n-UU^+)V\ )^+(I_n-UU^+)V = C^+C\ .$$

V. $AX_LA = A$; denn:

$$\begin{aligned} AX_LA &\underset{II}{=} (UU^++CC^+)[U\ V] = [UU^+U+\underline{CC^+U} \quad UU^+V+C\underline{C^+V}] = \\ &\underset{III,IV}{=} [U \quad UU^+V+\underline{C}] = [U \quad UU^+V+(I_n-UU^+)V] = [U\ V] = A\ . \end{aligned}$$

VI. $L = L(UU^++CC^+) \Rightarrow X_LAX_L = X_L$; denn:

$$\begin{aligned} X_LAX_L &\underset{II}{=} \begin{bmatrix} U^+ - U^+VC^+ - U^+V(I_k-C^+C)L \\ C^+ + (I_k-C^+C)L \end{bmatrix} (UU^++CC^+) = \\ &\underset{III}{=} \begin{bmatrix} \underline{U^+} +0-0 \underline{- U^+VC^+} - U^+V(I_k-C^+C)\underline{L(UU^++CC^+)} \\ C^+ + (I_k-C^+C)\underline{L(UU^++CC^+)} \end{bmatrix} \\ &\underset{Vor}{=} \begin{bmatrix} U^+(I_n-VC^+) - U^+V(I_k-C^+C)L \\ C^+ + (I_k-C^+C)L \end{bmatrix} = X_L\ . \end{aligned}$$

VII. $L_o = L_o(UU^++CC^+)$; denn:

$$U^+(I_n-V\underline{C^+)UU^+} \underset{III}{=} U^+ -0 \ ,$$

$$\underline{U^+}(I_n-VC^+)\underline{C}C^+ \underset{III}{=} 0- U^+VC^+ \ ,$$

$$\begin{aligned} KV^TU^{T+}U^+(I_n-VC^+)(UU^++CC^+) &= KV^TU^{T+}(U^+-U^+VC^+) = \\ &= KV^TU^{T+}U^+(I_n-VC^+)\ . \end{aligned}$$

VIII. $\Big(\ (\ U^+V(I_k-C^+C)(I_k-LV)\ )^T = (I_k-C^+C)LU\ \wedge$

$\wedge\ U^+V(I_k-C^+C)LU \in Sym(m)\ \wedge\ (I_k-C^+C)LV \in Sym(k)\ \Big) \Rightarrow$

$\Rightarrow X_LA \in Sym(m+k)$; denn:

$$\begin{aligned} X_LA &= \begin{bmatrix} U^+(I_n-VC^+) - U^+V(I_k-C^+C)L \\ C^+ + (I_k-C^+C)L \end{bmatrix} [U\ V] = \\ &\underset{III,IV}{=} \begin{bmatrix} U^+U -0 - U^+V(I_k-C^+C)LU & U^+V - U^+V\underline{C^+C} - U^+V(I_k-C^+C)LV \\ 0+ (I_k-C^+C)LU & \underline{C^+C} + (I_k-C^+C)LV \end{bmatrix} \\ &= \begin{bmatrix} U^+U - U^+V(I_k-C^+C)LU & U^+V(I_k-C^+C)(I_k-LV) \\ (I_k-C^+C)LU & C^+C + (I_k-C^+C)LV \end{bmatrix} , \end{aligned}$$

mit den drei Voraussetzungen folgt $X_LA = (X_LA)^T$.

IX. $(I_k-C^+C)K = K(I_k-C^+C) \in Sym(k)$; denn:

$$\begin{aligned} \underline{K^{-1}}(I_k-C^+C) &= (\ I_k + (I_k-C^+C)V^TU^{T+}U^+V(I_k-C^+C)\ )(I_k-C^+C) = \\ &\underset{Proj.}{=} (I_k-C^+C)(\ I_k + (I_k-C^+C)V^TU^{T+}U^+V(I_k-C^+C)\ ) = (I_k-C^+C)K^{-1}\ , \end{aligned}$$

Rechts- und Linksmultiplikation mit K ergibt die Behauptung.

X. Zur zweiten Bedingung in VIII:

$$U^+V(I_k-C^+C).KV^TU^{T+}\underline{U^+(I_n-VC^+).U} =$$
$$\underset{III}{=} U^+V(I_k-C^+C).KV^TU^{T+}\underline{U^+U} -0 \underset{PK2}{=} U^+V(I_k-C^+C)KV^T\underline{U^{T+}U^TU^{T+}} =$$
$$= U^+V.(I_k-C^+C)K.V^TU^{T+} \underset{IX}{\in} \mathrm{Sym}(m) \ .$$

XI. Zur dritten Bedingung in VIII:

$$(I_k-C^+C).KV^TU^{T+}U^+\underline{(I_n-VC^+).V} \underset{IV}{=} \underline{(I_k-C^+C)}KV^TU^{T+}U^+V(I_k-C^+C) =$$
$$\underset{Proj.}{=} (I_k-C^+C)K(I_k-C^+C)V^TU^{T+}U^+V(I_k-C^+C) +(I_k-C^+C)K-(I_k-C^+C)K =$$
$$\underset{IX}{=} (I_k-C^+C)\underline{K(\ I_k + (I_k-C^+C)V^TU^{T+}U^+V(I_k-C^+C)\ )} - (I_k-C^+C)K =$$
$$= (I_k-C^+C).I_k - (I_k-C^+C)K \underset{IX}{\in} \mathrm{Sym}(k) \ .$$

XII. Zur ersten Bedingung in VIII:

$$(\ U^+V(I_k-C^+C)(\ I_k- KV^TU^{T+}U^+(I_n-VC^+).V\ )\ )^T =$$
$$= (\ U^+V(\ (I_k-C^+C) - \underline{(I_k-C^+C)KV^TU^{T+}U^+(I_n-VC^+)V}\ )\ )^T =$$
$$\underset{XI}{=} (\ U^+V(\ (I_k-C^+C) - (I_k-C^+C) + (I_k-C^+C)K\ )\ )^T =$$
$$= (\ U^+V\underline{(I_k-C^+C)K}\ )^T =$$
$$\underset{IX}{=} (I_k-C^+C)KV^T\underline{U^{T+}} =$$
$$\underset{X}{=} (I_k-C^+C).KV^TU^{T+}U^+(I_n-VC^+).U \ .$$

_/

<u>Satz 20: Pseudoinversion von Summen</u> (CLINE 1965:100f)

Unter denselben Voraussetzungen wie in Satz 19:36 gilt:

$$(UU^T+VV^T)^+ = C^{T+}C^+ +$$
$$+ (I_n-C^{+T}V^T)U^{T+}.(I_m - U^+V(I_k-C^+C)KV^TU^{T+}).U^+(I_n-VC^+) \ .$$

<u>Bew.:</u> I. Sei wieder $A = [U\ V]$, dann ist einerseits

$AA^T = [U\ V]\begin{bmatrix}U^T\\V^T\end{bmatrix} = UU^T+VV^T$, und andererseits nach 17.iv:34

$(AA^T)^+ = A^{+T}A^+$, wobei A^+ in Satz 19:36 angegeben wurde.

Der Rest ist Rechnung. Setzt man

$M := I_m- U^+V(I_k-C^+C)KV^TU^{T+} \underset{19.IX.37}{\in} \mathrm{Sym}(m)$,

$N := (I_k-C^+C)KV^TU^{T+}U^+(I_n-VC^+) \in M(k,n)$, so schreibt sich

$$A^+ = \begin{bmatrix} MU^+(I_n-VC^+) \\ C^+ + N \end{bmatrix} \ .$$

II. $N^TC^+ = 0$; denn $(I_k-C^+C)C^+ = 0$.

III. $(C^+ + N)^T(C^+ + N) \underset{II}{=} C^{+T}C^+ + 0 + 0 + N^TN = C^{+T}C^+ + N^TN$.

IV. $M^TM = M^2 =$

$$= I_m - 2U^+V(I_k-C^+C)KV^TU^{T+} + U^+V\underline{(I_k-C^+C)K}V^TU^{T+}.U^+V(I_k-C^+C)KV^TU^{T+} +$$

$$+ U^+V(I_k-C^+C)K.KV^TU^{T+} - U^+V(I_k-C^+C)K.KV^TU^{T+} =$$

$$\underset{19.IX:37}{=} I_m - 2U^+V(I_k-C^+C)KV^TU^{T+} + U^+V(I_k-C^+C)\underline{K.(\ I_k +}$$

$$\underline{+ (I_k-C^+C)V^TU^{T+}.U^+V(I_k-C^+C)\ )}.KV^TU^{T+} - U^+V(I_k-C^+C)K^2V^TU^{T+} =$$

$$= M - U^+V(I_k-C^+C)KV^TU^{T+} + U^+V(I_k-C^+C).I_k.KV^TU^{T+} - U^+V(I_k-C^+C)K^2V^TU^{T+} =$$

$$= M - U^+V(I_k-C^+C)K^2V^TU^{T+} \ .$$

V. $N^TN = (I_n-C^{T+}V^T)U^{T+}U^+VK(I_k-C^+C).(I_k-C^+C)KV^TU^{T+}U^+(I_n-VC^+) =$

$$\underset{19.IX:37}{=} (I_n-C^{T+}V^T)U^{T+}U^+V(I_k-C^+C)K^2.V^TU^{T+}U^+(I_n-VC^+) \ .$$

VI. Zusammen ergibt das:

$$A^{+T}A^+ = [\,(MU^+(I_n-VC^+))^T \quad (C^++N)^T] \begin{bmatrix} MU^+(I_n-VC^+) \\ C^+ + N \end{bmatrix} =$$

$$= (I_n-C^{T+}V^T)U^{T+}\underline{M^TM}U^+(I_n-VC^+) + \underline{(C^++N)^T(C^++N)} =$$

$$\underset{III}{=} (I_n-C^{T+}V^T)U^{T+}MU^+(I_n-VC^+) -$$

$$\underset{IV}{-} (I_N-C^{T+}V^T)U^{T+}.U^+V(I_k-C^+C)K^2V^TU^{T+}.U^+(I_n-VC^+) + N^TN + C^{T+}C^+ =$$

$$\underset{V}{=} (I_n-C^{T+}V^T)U^{T+}.(I_m - U^+V(I_k-C^+C)KV^TU^{T+}).U^+(I_n-VC^+) + C^{T+}C^+ \ . \ \_/$$

<u>Satz 21: Formel für $X^T(S^2+XX^T)^+$</u>

Sei X eine (n,p)-Matrix und $S\in Sym(n)$ symmetrisch.
Setze $M := I_n-XX^+ \ \in Proj(n)$, und
$B := X^+(I_n-S(MS)^+) \ \in M(p,n)$.
Dann gilt:

$$X^T(S^2+XX^T)^+ = (\ I_p + BS^2B^T\ )^{-1}B \ .$$

<u>Bew.:</u> I. Die Ergebnisse der Sätze 19, 20 werden übertragen

mit dem Lexikon:

U	X
m	p
V	S
k	n
C	MS
K	$(I_n+SB^TBS)^{-1}$

$$K = (I_n + (I_n-\underline{(MS)^+M}S)SX^{T+}X^+S(I_n-\underline{(MS)^+M}S)\ )^{-1} =$$

$$\underset{17.v:34}{=} (I_n + (X^+\underline{S(I_n-(MS)^+S)})^TX^+S(I_n-(MS)^+S)\ )^{-1} =$$

$$= (I_n + (X^+(I_n-S(MS)^+)S)^TX^+(I_n-S(MS)^+)S)\ )^{-1} =$$

$$= (I_n + SB^TBS)^{-1} \ .$$

II. $X^T\underline{(SM)^+} \underset{17.v:34}{=} X^TM(SM)^+ = (\underline{MX})^T(SM)^+ = 0$.

III. Mit Satz 20:38 und dem angegebenen Lexikon:

$X^T(S^2+XX^T)^+ =$

$= \underline{X^T}.(I_n-\underline{(SM)^+S})X^{+T}.\ (I_p - X^+S(I_n-(MS)^+S)KSX^{+T}).X^+(I_n-S(MS)^+) +$

$+ \underline{X^T.(SM)^+}(MS)^+ =$

$\underset{II}{=} \underline{(X^T} -0)\underline{X^{+T}}.(I_p - X^+(I_n-S(MS)^+)SKSX^{+T}).B + 0 =$

$= X^+X.(I_p - BSKSX^{+T}).B =$

$= \underline{X^+XB} - \underline{X^+XB}.SKSX^{+T}B = B - BSKSX^{+T}B = (I_p - BSKSX^{+T}).B$.

IV. $BSK = (I_p+BS^2B^T)^{-1}BS$; denn:

$(I_p+BS^2B^T)BS = BS(I_n+SB^TBS)$,und Multiplikation von rechts mit $K = (I_n+SB^TBS)^{-1}$ und von links mit $(I_p+BS^2B^T)^{-1}$ ergibt die Behauptung.

V. $BS^2\underline{(SM)^+} \underset{15.iv:33}{=} BS^2.\underline{(MS^2M)^+MS} \underset{17.v:34}{=} BS.\underline{SM(MS^2M)^+}.S =$

$= BS.(MS)^+.S = X^+\underline{(I_n-S(MS)^+)S}(MS)^+S = X^+S.\underline{(I_n-(MS)^+MS)(MS)^+S} = 0$.

VI. $I_p - \underline{BSK}SX^{T+} \underset{IV}{=} I_p - (I_p+BS^2B^T)^{-1}BS.SX^{T+} =$

$= I_p - (I_p+BS^2B^T)^{-1}BS^2(I_n \underline{-(SM)^+S+(SM)^+S})X^{T+} =$

$= I_p - (I_p+BS^2B^T)^{-1}BS^2((\underline{X^+(I_n-S(MS)^+)})^T+(SM)^+SX^{T+}) =$

$= I_p - (I_p+BS^2B^T)^{-1}(BS^2B^T + \underline{BS^2(SM)^+}SX^{T+}) =$

$\underset{V}{=} I_p - (I_p+BS^2B^T)^{-1}(BS^2B^T + 0 \underline{+I_p-I_p}) =$

$= I_p - \underline{(I_p+BS^2B^T)^{-1}(I_p+BS^2B^T)} + (I_p+BS^2B^T)^{-1} =$

$= 0 + (I_p+BS^2B^T)^{-1}$.

Damit in III:

$X^T(S^2+XX^T)^+ = (I_p + BS^2B^T)^{-1}B$. _/

* * *

Kap. 3: Lösungen linearer Matrizengleichungen

Optimale Schätzfunktionen in GAUSS-MARKOFF-Modellen ergeben sich als Lösungen von Matrizengleichungen $Ax = B$, solche Gleichungen werden zunächst behandelt (3.1). Zur Berechnung optimaler Schätzer bzgl. des GAUSS-Risikos (mean square error) müssen quasi-innere Produkte minimiert werden, diese Produkte diskutieren wir im zweiten Abschnitt (3.2).

3.1 Minimum-Norm-Lösungen linearer Matrizengleichungen

Sei A eine Matrix, b ein Vektor. Die Gleichung $Ax = b$ ist genau dann lösbar, wenn $b \in \operatorname{im} A$; denn genau dann existiert ein Urbild x_o: $b = Ax_o$. $b \in \operatorname{im} A$ ist äquivalent mit $b = \operatorname{proj}(b/\operatorname{im} A) = AA^+b$.

Mehrere Gleichungen $Ax = b_\mu$ mit rechten Seiten $b_1, \ldots, b_m$ können zu einer Matrixgleichung $Ax = B$ zusammengefügt werden, wobei die Matrix $B = [b_1, \ldots, b_m]$ aus den rechten Seiten b_μ spaltenweise zusammengesetzt wird. Die Theorie der Matrizengleichungen $Ax = B$ unterscheidet sich daher im wesentlichen nicht von der Theorie linearer Gleichungssysteme $Ax = b$.

Eine Matrizengleichung $Ax = B$ heiße "<u>lösbar</u>", falls eine Lösungsmatrix X existiert: $AX = B$. Die Menge aller Lösungen bezeichnen wir mit "<u>Lsg(Ax=B)</u>".

<u>Satz 22: Lösbarkeits- und Teilraumkriterium</u>

(PENROSE 1955:409) Für jede (n,k)-Matrix A und (n,m)-Matrix B gilt:

$$Ax = B \text{ lösbar} \Leftrightarrow \operatorname{im} B \subset \operatorname{im} A \Leftrightarrow AA^+B = B \quad .$$

<u>Bew.:</u> I. Sei $AX = B$ und z im Bild von B. Dann liegt z wegen $z = By = AXy$ auch im Bild von A.
II. AA^+ projiziert auf das Bild von A (18.i:35). Ist $\operatorname{im} B \subset \operatorname{im} A$, dann gilt offenbar $AA^+B = B$.
III. Bei $AA^+B = B$ ist A^+B sichtbar eine Lösung. _/

Wir berechnen nun den Lösungsraum lösbarer Gleichungen:

Satz 23: Lösungsraum einer Matrixgleichung

(PENROSE 1955:409) Sei A eine (n,k)-Matrix und B eine (n,m)-Matrix und die Gleichung $Ax = B$ lösbar. Dann gilt:

i (affiner Lösungsraum)

$$\begin{aligned} \mathrm{Lsg}(Ax=B) &= \{\, A^+B + (I_k - A^+A)Z \;:\; Z \in M(k,m) \,\} = \\ &= A^+B \oplus \mathrm{Lsg}(Ax=0)\ ; \end{aligned}$$

ii (Minimum-Norm-Eigenschaft von A^+B)

A^+B hat unter allen Lösungen minimale Norm.

Bew.: I. $\subset$: Sei X eine Lösung: $AX = B$. Mit $Z = X$ gilt:

$A^+B + (I_k - A^+A)X = A^+B + X - A^+\underline{AX} = X \underline{+A^+B - A^+B} = X$.

II. $\supset$: Für jede (k,m)-Matrix Z gilt:

$A(A^+B + (I_k - A^+A)Z) = \underline{AA^+B} + 0.Z = B$ nach 22:41 .

III. Mit $B = 0$ folgt aus der ersten Gleichung sofort

$\mathrm{Lsg}(Ax=B) = A^+B + \mathrm{Lsg}(Ax=0)$. Bleibt zu zeigen:

$A^+B \perp \mathrm{Lsg}(Ax=0)$:

$$\mathrm{tr}\, A^+B.(\, (I_k - A^+A)Z\,)^T = \mathrm{tr}\, A^+BZ^T(I_k - A^+A) \underset{\text{kom.}}{=} \mathrm{tr}\, BZ^T(I_k - A^+A)A^+ = 0\ .$$

ii. Nach dem Satz von PYTHAGORAS wird

$$\|A^+B + (I_k - A^+A)Z\|^2 = \|A^+B\|^2 + \|(I_k - A^+A)Z\|^2$$

minimiert bei $Z = 0$.

Bei Vektoren b schreibt man üblicherweise

$\mathrm{Lsg}(Ax=b) = A^+b \oplus \ker A$.

Dabei ist A^+b gerade diejenige inhomogene Lösung, die minimale Norm hat, d.h. die auf $\ker A$ senkrecht steht.

Bei nicht-lösbaren Gleichungen $Ax = B$ liegt es nahe, ersatzweise den Abstand $\|Ax-B\|$ zu minimieren. Der Lösungsraum dieser Minimierungsaufgabe - wir nennen ihn "$\underline{\mathrm{Lsg}(\|Ax-B\|=\inf)}$" - fällt nicht nur im Falle der Lösbarkeit, sondern immer zusammen mit $\{\, A^+B + (I_k - A^+A)Z \;:\; Z \in M(k,m) \,\}$.

Für Vektoren b wird $\|Ax-b\|$ minimiert von der Projektion $\mathrm{proj}(b/\mathrm{im}\, A) = AA^+b$; genauso bei Matrizen:

Satz 24: Lösungsraum bei Abstandsminimierung

(PENROSE 1956:17) Für jede (n,k)-Matrix A und (n,m)-Matrix B gilt:

i (Abstandsminimierung durch Projektion)

$$\inf \{ \|AX-B\| : X \in M(k,m) \} = \|AA^+B-B\| \quad ;$$

ii (affiner Lösungsraum)

$$\begin{aligned} Lsg(\|Ax-B\|=\inf) &= Lsg(Ax=AA^+B) \\ &= \{ A^+B + (I_k-A^+A)Z : Z \in M(k,m) \} \\ &= A^+B \oplus Lsg(Ax=0) \end{aligned}$$

iii (Minimum-Norm-Eigenschaft von A^+B)

A^+B hat unter allen Lösungen minimale Norm.

Bew.: i. Man setze

$\hat{B} := proj(B/im\ A) = AA^+B$, $\bar{B} := B - \hat{B} = (I_n-AA^+)B$.

Sei X eine beliebige (k,m)-Matrix. Wegen

$$\underbrace{tr\ (AX-\hat{B}).\bar{B}^T}_{=\,0} = tr\ A(X-A^+B).B^T(I_n-AA^+) \underset{kom.}{=} tr\ (X-A^+B)B^T.\underline{(I_n-AA^+)A}$$

stehen $AX-\hat{B}$ und $\bar{B}$ senkrecht aufeinander, und es gilt nach PYTHAGORAS:

(*) $\|AX-B\|^2 = \|AX-\hat{B}-\bar{B}\|^2 = \|AX-\hat{B}\|^2+\|\bar{B}\|^2 \geq \|\bar{B}\|^2 = \|AA^+B-B\|^2$.

Also löst A^+B die Minimierungsaufgabe.

ii. $\|Ax-B\|$ wird von X genau dann minimiert, wenn in (*) Gleichheit steht, d.h. $\|AX-\hat{B}\| = 0$, d.h. $AX = \hat{B} = AA^+B$, d.h. X löst $Ax = AA^+B$.

Die Gleichung $Ax = AA^+B$ ist immer lösbar (22:41): $AA^+.AA^+B = AA^+B$, und hat den Lösungsraum (23:42):

$$\begin{aligned} Lsg(Ax=AA^+B) &= \{ A^+.AA^+B + (I_n-A^+A)Z : Z \in M(k,m) \} = \\ &= \{ A^+B + (I_n-A^+A)Z : Z \in M(k,m) \} . \end{aligned}$$

Die restlichen Aussagen wurden schon in Satz 23:42 bewiesen. _/

3.2 Quasi-innere Produkte

Quasi-innere Produkte sind positiv semidefinite (Q3), symmetrische (Q2), "bi-affin-lineare" Abbildungen (Q1):

Def. 3: (quasi-inneres Produkt)

Sei M ein R-Vektorraum. Eine Abbildung $V : M \times M \to R$ heiße "quasi-inneres Produkt", falls

(Q1) $\forall B \in M$ $\quad V(A,B)-V(0,B)$ linear in A ,

(Q2) $\forall A,B \in M$ $\quad V(A,B) = V(B,A)$,

(Q3) $\forall A \in M$ $\quad V(A,A) \geq 0$. _/

Das folgende Minimierungskriterium werden wir im Beweis zu Satz 37:70 benutzen:

Satz 25: Minimierung von quasi-inneren Produkten (DRYGAS 1972:374f)
Sei M ein R-Vektorraum und $V : M \times M \to R$ ein quasi-inneres Produkt. Dann gilt für jeden Punkt $A \in M$: $V(A,A)$ ist minimal: $V(A,A) = \inf \{ V(B,B) : B \in M \}$ genau dann, wenn für alle $C \in M$ $V(A,C)-V(A,0) = o$.

Bew.: I. V induziert eine p.s.d. symmetrische Bilinearform $W(A,B) := V(A,B) - V(A,0) - V(0,B) + V(0,0)$. In II wird gezeigt, daß W tatsächlich eine symmetrische Bilinearform ist. Dann (III) wird eine binomische Formel bewiesen, aus der leicht folgt, daß W auch p.s.d. ist (IV). V und VI beweisen schließlich die Behauptung.

II. W ist symmetrisch ; denn:

$$W(A,B) = V(A,B) - V(A,0) - V(0,B) + V(0,0) =$$
$$\underset{Q2}{=} V(B,A) - V(0,A) - V(B,0) + V(0,0) = W(B,A) .$$

W ist bilinear; denn:

$$\begin{aligned} W(\alpha A+\gamma C,B) &= \underline{V(\alpha A+\gamma C,B)} - V(\alpha A+\gamma C,0) - \underline{V(0,B)} + V(0,0) = \\ &= \alpha(\ V(A,B)-V(0,B)\ ) + \gamma(\ V(C,B)-V(0,B)\ ) + \\ &\underset{Q1}{+} \alpha(-V(A,0)+V(0,0)\ ) + \gamma(-V(C,0)+V(0,0)\ ) = \\ &= \alpha\ W(A,B) + \gamma\ W(C,B) . \end{aligned}$$

III. (Binomische Formel) $\forall A,C \in M$:

$$V(A+C,A+C) = V(A,A) + W(C,C) + 2(\ V(A,C)-V(A,0)); \text{ denn:}$$

$$\begin{aligned} V(A+C,A+C) &= V(A+C,A+C) - V(A+C,0) - V(0,A+C) + V(0,0) + \\ &+ V(A+C,0) - V(0,0) + \\ &+ V(0,A+C) - V(0,0) + V(0,0) = \\ &= W(A+C,A+C) + \\ &\underset{Q1}{+} V(A,0) - V(0,0) + V(C,0) - V(0,0) + \\ &+ V(A,0) - V(0,0) + V(C,0) - V(0,0) + V(0,0) = \\ &= \underline{W(A.A)} + \underline{2\ W(A.C)} + W(C,C) + \\ &+ 2\ V(A,0) + 2\ V(C,0) - 3\ V(0,0) = \\ &= V(A,A) - 2\ V(A,0) + V(0,0) + \\ &+ 2\ V(A,C) - 2\ V(A,0) - 2\ V(0,C) + 2\ V(0,0) + \\ &+ W(C,C) + \\ &+ 2\ V(A,0) + 2\ V(C,0) - 3\ V(0,0) = \\ &= V(A,A) + W(C,C) + 2\ (\ V(A,C)-V(A,0)\ ) . \end{aligned}$$

IV. W ist p.s.d. ; denn:

$o \leqq V(A+\gamma C, A+\gamma C) = V(A,A) + \gamma^2 W(C,C) + 2\gamma\,(\,V(A,C)-V(A,0)\,)$;

Division durch γ^2 und Grenzübergang ergibt $\forall C \in M$

$$o \leqq \lim_{\gamma\to\infty} \gamma^{-2} V(A,A) + W(C,C) + 2\gamma^{-1}(\,V(A,C)-V(A,0)\,) = W(C,C) .$$

V. $\Rightarrow$: Sei $V(A,A)$ minimal. Sei $C \in M$, $\gamma > o$.

Mit III folgt

$$V(A,A) \leqq V(A+\gamma C, A+\gamma C) = V(A,A) + \gamma^2 W(C,C) + 2\gamma(V(A,C)-V(A,0)) ,$$

$$0 \leqq \lim_{\gamma\downarrow o} \gamma W(C,C) + 2(\,V(A,C)-V(A,0)\,) = 2(\,V(A,C)-V(A,0)\,) .$$

Also für alle C : $o \leqq V(A,C)-V(A,0)$;

insbesondere für $-C$ wegen der Linearität:

$o \leqq V(A,-C)-V(A,0) = -(V(A,C)-V(A,0))$;

zusammen gibt das: $o = V(A,C)-V(A,0)$.

VI. $\Leftarrow$: Sei $V(A,C)-V(A,0) = o$ für alle C . Sei $B \in M$,

$C := B-A$. Dann

$$V(B,B) = V(B-A+A, B-A+A) = V(A+C, A+C) \overset{\text{III}}{=}$$

$$= V(A,A) + W(C,C) + 2(\,V(A,C)-V(A,0)\,) \geqq$$

$$\geqq V(A,A) + o + o .$$

_/

* * *

Teil II

Schätzung des Mittelwerts in GAUSS-MARKOFF-Modellen

GAUSS-MARKOFF-Modelle sind lineare Modelle linearer Regression. Gesucht werden optimale Schätzfunktionen für lineare Funktionen des Mittelwert-Regressionsparameters.

Minimum Norm - Minimum Bias - lineare Schätzfunktionen sind optimal, indem zuerst der Fehler in den ersten Momenten (Bias) und dann der Fehler in den zweiten Momenten minimiert wird; letzteres geschieht gleichmäßig (Minimum Norm), da keine Annahme über die Streuung gemacht wird. Dieses Verfahren führt zur klassischen Methode der kleinsten Quadrate (Kap. 4).

Minimum Varianz - Minimum Bias - lineare Schätzfunktionen minimieren den Fehler in den zweiten Momenten mit Annahmen über die Streuung (Minimum Varianz). Dies verallgemeinert die klassische BLUE-Theorie (Kap. 5).

Bei Minimum mean square error - linearen Schätzfunktionen wird Bias und Varianz gleichzeitig minimiert (Kap. 6).

Alle Schätzer werden explizit angegeben, in Spezialfällen vereinfacht und untereinander in Beziehung gesetzt.

Kap. 4: Minimum Norm - Minimum Bias - lineare Schätzfunktionen

In diesem Kapitel stellen wir zunächst die typischen Überlegungen dar, die zur Annahme von GAUSS-MARKOFF-Modellen führen (4.1). Dann leiten wir ein erstes Schätzverfahren her (4.2 - 4.4) und vergleichen dieses Verfahren mit der Methode der kleinsten Quadrate (4.5).

4.1 GAUSS-MARKOFF-Modelle zur Mittelwert-Regression

"Given these data and their origin, what model do I use?" (SEARLE 1971:2). Welchen praktischen Fragestellungen ordnet man ein GAUSS-MARKOFF-Modell zu? Indem wir das Typische solcher Situationen beschreiben, entwickeln wir die charakterisierenden Modellannahmen. Drei Schritte führen zur Annahme eines GAUSS-MARKOFF-Modells (vgl. SEELY 1970:1726f; URQUHARDT, WEEKS & HENDERSEN 1973:310):

- Wahl des Mittelwerts als strukturbeschreibendem Parameter (Parameterwahl),
- Beschränkung dieses Parameters auf einen linearen Teilraum (Annahme eines linearen Modells), und
- Entscheidung für eine lineare Regression (Annahme eines Modells linearer Regression).

Grundannahme der Mathematischen Statistik (WITTING 1966:13): Das Beobachtungsmaterial - ein n-dimensionaler reeller Vektor Y - wird als Zufallsgröße von einem Meßraum $(\Omega,\mathfrak{A})$ in den BORELschen $(R^n,\mathfrak{B}_n)$ aufgefaßt. Über dem Meßraum $(\Omega,\mathfrak{A})$ ist eine Klasse von Wahrscheinlichkeitsmaßen $\mathfrak{P}$ gegeben. Bei vorliegenden Beobachtungen Y soll eine "wahre" Verteilung P^Y, $P\in\mathfrak{P}$, ausgewählt bzw. ein "wahrer" Parameterwert μP^Y bestimmt werden. Diese Entscheidung soll nicht willkürlich fallen, sondern wahrscheinlichkeitstheoretisch begründet sein.

Parameterwahl: Trivialerweise gilt die Zerlegung $Y = E_PY + (Y-E_PY)$. Das praktische Problem soll es gestatten, den Mittelwert E_PY als "struktur-beschreibend" zu betrachten und das "Residuum" $Y-E_PY$ als "Störgröße" einzustufen.

Definiert man als Parameter μ die Funktion

$\mu : \mathfrak{P}^Y \to R^n, \quad P^Y \mapsto \mu P^Y := E_P Y$,

dann möchte man also aus vorliegenden Beobachtungen Y auf den "wahren" Parameterwert μP^Y schließen.

<u>Annahme eines linearen Modells:</u> Die praktische Fragestellung soll die Annahme rechtfertigen, daß der Parameter μ nur Werte in einem Teilraum Ω des R^n annimmt; genauer:

$$\text{(lin.Mod.)} \qquad \operatorname{span} \mu\mathfrak{P}^Y = \Omega \quad .$$

In die Schätztheorie für lineare Modelle geht nur der Raum Ω und nicht die Menge $\mu\mathfrak{P}^Y$ der Parameterwerte ein. Es können daher Schätzungen in die Differenz $\Omega \setminus \mu\mathfrak{P}^Y$ auftreten: wir nennen dies den "<u>Fundamentaldefekt</u> linearer Modelle"; denn Schätzwerte in $\Omega \setminus \mu\mathfrak{P}^Y$ sind vom praktischen Standpunkt aus modellwidrig und schlimmstenfalls unrealisierbar.

Geben wir zwei Beispiele:

i. Das tatsächliche Experiment lasse alle Mittelwerte μP^Y zu, die in einem Teilraum Ω des R^n liegen und kürzer sind als β (>o): $\|\mu P^Y\| < \beta$. Dann gilt $\operatorname{span} \mu\mathfrak{P}^Y = \Omega$; Schätzwerte in $\Omega \setminus \mu\mathfrak{P}^Y$ sind von den praktischen Modellannahmen her modellwidrig, aber nicht unbedingt unrealisierbar.

ii. Bei einer reellwertigen Zufallsgröße Y soll der Erwartungswert von $\bar{Y} := (Y-EY)^2$, $E_P\bar{Y} = \operatorname{Var}_P Y$ geschätzt werden. Dann gilt $\operatorname{span} \mu\mathfrak{P}^{\bar{Y}} = R$, jedoch sind negative Schätzungen offenbar unrealisierbar.

Für den Fundamentaldefekt gibt es keine allgemeine Lösung. Er widerlegt zwar nicht die rein mathematische Behandlung dieser Modelle, stellt aber deren Sinn in Frage. Wenn man ihn in Kauf nimmt, dann mangels besserer Alternativen[16].

<u>Annahme eines Modells linearer Regression:</u> Das konkrete Problem möge es – häufig im Rahmen eines kausalen Erklärungsversuchs – erlauben, μP^Y als Wert einer Funktion $X : R^p \to R^n$ an der Stelle $b \in R^p$ anzusehen: $\mu P^Y = X(b)$.

---

16 Man könnte versucht sein, den Fundamentaldefekt "wegzudefinieren", indem man $\mu\mathfrak{P}^Y = \Omega$ fordert. Aber in vielen Fällen, in denen man ein lineares Modell annimmt und dessen Methoden einsetzt, gilt nur $\operatorname{span} \mu\mathfrak{P}^Y = \Omega$ und nicht $\mu\mathfrak{P}^Y = \Omega$; vergleiche die beiden Beispiele.

Zudem möge man mit Recht annehmen, daß die "Regressionsfunktion" X linear ist:

(lin.Regr.) $X \in M(n,p) \;\wedge\; {}_\mu\mathfrak{P}^Y \subset \operatorname{im} X$.

Das Interesse hat sich jetzt vom Parameterwert ${}_\mu P^Y$ auf den Wert b verschoben. Folglich sucht man nicht mehr nach Schätzfunktionen für ${}_\mu P^Y$, sondern nach solchen für b , oder allgemeiner: nach Schätzfunktionen für Lb , wobei $L \in M(1,p)$ eine lineare Funktion des Regressionsparameters b ist.

Das Schätzen linearer Funktionen Lb hat den Vorteil, daß lineare Modelle linearer Regression mehr liefern als lineare Modelle. Denn bei der speziellen Wahl L = X wird Xb geschätzt, d.i. ${}_\mu P^Y$, der Parameter des linearen Modells.

Ein <u>lineares Modell linearer Regression</u> beschreiben wir nun durch das 4-Tupel (Y,Xb,n,p) und nennen es GAUSS-MARKOFF-Modell:

<u>Def. 4:</u> (GM-Modell (Y,Xb,n,p))

Ein 4-Tupel (Y,Xb,n,p) heiße "<u>GAUSS-MARKOFF-Modell</u>", "<u>GM-Modell</u>" - kurz: GM-Modell (Y,Xb,n,p) -, falls

(GMM1) Y ein R^n-wertiger Zufallsvektor ist,

(GMM2) X eine (n,p)-Matrix ist, und

(GMM3) der Erwartungswert von Y im Bild von X liegt: $EY \in \operatorname{im} X$. _/

Die Eigenschaft (GMM3) ist eine kurze Schreibweise für $\operatorname{span} E_\mathfrak{P} Y = \operatorname{im} X$ und faßt damit die Annahmen

(lin.Mod.) $\operatorname{span} E_\mathfrak{P} Y = \Omega$ und

(lin.Regr.) $E_\mathfrak{P} Y \subset \operatorname{im} X$

zusammen. Üblicherweise wird auf die Mitführung von $\mathfrak{P}$ verzichtet[17]. -

17 Genauer könnte man ein GAUSS-MARKOFF-Modell angeben durch das 4-Tupel $(\mathfrak{P}^Y, Xb, n, p)$ mit den Modellannahmen

(1) $\mathfrak{P}^Y$ ist eine Klasse von Verteilungen über dem BORELschen $(R^n, \mathfrak{B}_n)$,

(2) X ist eine (n,p)-Matrix, und

(3) $\operatorname{span} E_\mathfrak{P} Y = \operatorname{im} X$.

Sei $S^2 \in Sym(n)$ eine Streuungsmatrix. Dann kann man sich auf diejenigen Wahrscheinlichkeitsmaße $P \in \mathfrak{P}$ beschränken, unter denen Y die Streuung S^2 hat:

Def. 5: (GM-Modell (Y,Xb,n,p,S^2))
Ein 5-Tupel (Y,Xb,n,p,S^2) heiße "GAUSS-MARKOFF-Modell mit bekannter Streuung S^2 " - kurz:
GM-Modell (Y,Xb,n,p,S^2) ⌐ falls
(Y,Xb,n,p) ein GM-Modell ist und
(GMM4) Y die Streuungsmatrix S^2 hat:

$$DY = S^2 .$$

_/

RAO (1971c:372) und RAO & MITRA (1971:136) beschreiben (mit unseren Bezeichnungen) ein GM-Modell durch das Tupel (Y,Xb,S^2) ; MILLER (1973:710) wählt $\mathfrak{M}_{np}[Y,b,X,S^2]$.-

In einem GM-Modell (Y,Xb,n,p) werden die folgenden Namen und gebundenen Variablen verwendet:
Y heißt "Beobachtungsvektor", n "Anzahl der Beobachtungen".
X wird "(Mittelwert-)Design" genannt, denn X "beschreibt" den Versuchsaufbau; b heißt "(Mittelwert-)Regressionsparameter", p "Anzahl der Regressoren".
$M := I_n - XX^+ \underset{18.i:35}{=} proj(./im^{\perp}X)$ ist die Projektion auf das orthogonale Komplement des Bildes von X . MY heißt "streuungsrelevanter Anteil von Y ", denn da man mit $\mathfrak{L} = im\, X$ statt mit $E_{\mathfrak{P}}Y$ rechnet (Fundamentaldefekt), kann $XX^+Y =$ $= proj(Y/im\, X)$ ganz durch die Mittelwert-Regression erklärt werden.

Ist F eine meßbare Funktion vom R^n in den R^m und (Y,Xb,n,p) ein GM-Modell, so kann man in manchen Fällen für FY wieder ein GM-Modell angeben, das dann das "F-transformierte GM-Modell" heißt. Im Gegensatz dazu nennen wir (Y,Xb,n,p) das "ursprüngliche GM-Modell".

Geben wir ein Beispiel: Sei $F = A$ eine (m,n)-Matrix. Dann ist (AY,AXb,m,p,AS^2A^T) das A-transformierte GM-Modell zum ursprünglichen GM-Modell (Y,Xb,n,p,S^2) .

Das zum ursprünglichen GM-Modell (Y,Xb,n,p,S^2) gehörige S^+-transformierte GM-Modell (S^+Y,S^+Xb,n,p,S^+S) (17.vi:34) heiße das "AITKEN-Modell von (Y,Xb,n,p,S^2)" .-

An die Ränge von X und S^2 werden <u>keine</u> Bedingungen gestellt. Diese Allgemeinheit ist nicht nur theoretisch befriedigend, sondern auch praktisch nützlich. Geben wir ein Beispiel: Um in einem GM-Modell (Y,Xb,n,p,S^2) die lineare Nebenbedingung $y_o = X_o b$, $X_o \in M(m,p)$, $y_o \in R^m$, zu berücksichtigen, braucht man nur überzugehen zum GM-Modell $\left(\begin{bmatrix} Y \\ y_o \end{bmatrix}, \begin{bmatrix} X \\ X_o \end{bmatrix} b, n+m, p, \begin{bmatrix} S^2 & 0 \\ 0 & 0 \end{bmatrix}\right)$ [18].-

Im folgenden bestimmen wir optimale Verfahren zur Schätzung einer linearen Funktion Lb des Regressionsparameters b , $L \in M(1,p)$. Um überhaupt ein Entscheidungsverfahren gegen alle anderen vergleichen zu können, darf die Menge der zugelassenen Verfahren nicht "zu groß" sein: In GM-Modellen beschränkt man sich auf lineare Schätzfunktionen $\hat{L}Y$, $\hat{L} \in M(1,n)$. Solche Funktionen nennen wir "(Mittelwert-) <u>Schätzer</u>".

In ursprünglichen GM-Modellen (Y,Xb,n,p) soll nun gelten $\text{span } Y\Omega = R^n$. Weiß man nämlich, daß Y schon in einen echten Teilraum $\text{im } A \subset R^n$ abbildet, so kann man ($AA^+ = \text{proj}(./\text{im } A)$ nach 18.1:35) zum AA^+-transformierten GM-Modell übergehen.

Die linearen Schätzfunktionen $\hat{L}Y$ in ursprünglichen GM-Modellen (Y,Xb,n,p) bilden daher den Raum $L(R^n,R^1) = M(1,n)$ aller linearen Abbildungen vom R^n in den R^1 . Er hat die Dimension $\dim L(R^n,R^1) = n1$. Zwei Schätzer $\hat{L}Y$, $\check{L}Y$ sind gleich, wenn die Matrizen $\hat{L}$, $\check{L}$ gleich sind.

Die linearen Schätzfunktionen $\hat{L}AY$ in A-transformierten GM-Modellen $(A \in M(m,n))$ bilden den Raum $L(\text{im } A,R^1) = M(1,m).A$ aller linearen Abbildungen von $\text{im } A$ in den R^1 . Er hat die Dimension $\dim L(\text{im } A,R^1) = \text{rk } A \cdot 1 \leq n1$. Zwei Schätzer $\hat{L}AY$, $\check{L}AY$ können gleich sein: $\hat{L}A = \check{L}A$, ohne daß die Matrizen $\hat{L}$, $\check{L}$ gleich sind.-

"Given this model and its definitions, what consequences can I deduce?" (SEARLE 1971:2).

---

18 Die Streuungsmatrix muß offenbar die angegebene Form haben. Eine andere Wahl ergäbe ein "inkonsistentes" Modell (vgl. RAO 1971c:378; 1973:278). Fragen der Konsistenz spielen bei der mathematischen Behandlung aber keine Rolle und werden deshalb nicht weiter verfolgt.

4.2 Erwartungstreue lineare Schätzfunktionen

Gegeben sei
- ein GM-Modell (Y,Xb,n,p),
- eine lineare Funktion Lb des Regressionsparameters b : $L \in M(1,p)$, und
- eine lineare Schätzfunktion $\hat{L}Y$ für Lb : $\hat{L} \in M(1,n)$.

In diesem Abschnitt diskutieren wir die Erwartungstreue von $\hat{L}Y$.

Mit Erwartungstreue meint man in einem Regressionsmodell, daß der Regressionsparameter im Mittel treu ist: Wenn für den Parameterwert E_PY die Regression Xb gilt, dann soll $\hat{L}Y$ im Mittel gerade den Wert Lb annehmen:

Def. 6: i (erw.-treue Schätzer für Lb)
$\hat{L}Y$ heiße "erwartungstreu", "erw.-treu", "erwartungstreue lineare Schätzfunktion für Lb ", "erw.-treuer Schätzer für Lb ", falls
$\forall b \in R^p$: $EY = Xb \Rightarrow E\,\hat{L}Y = Lb$.
ii (erw.-treue Schätzbarkeit von Lb)
Lb heiße "erwartungstreu schätzbar", "erw.-treu schätzbar", falls
ein erw.-treuer Schätzer für Lb existiert. _/

Die Implikation $EY = Xb \Rightarrow E\,\hat{L}Y = Lb$ kann gelesen werden: Bei zugrundeliegendem wahren Regressionsparameter b hat $\hat{L}Y$ den Erwartungswert Lb . In Anlehnung an diejenigen statistischen Verfahren, in denen der zu schätzende Parameter eine Verteilung festlegt, wird an Stelle der Implikation $EY = Xb \Rightarrow E\,\hat{L}Y = Lb$ oft kürzer geschrieben: $E_b\hat{L}Y = Lb$.

Sind b, $\bar{b}$ zwei Regressionsparameter für denselben Parameterwert E_PY : $E_PY = Xb = X\bar{b}$ und ist $\hat{L}Y$ erw.-treu, dann gilt per Definition notwendig $E\,\hat{L}Y = Lb = L\bar{b}$. Bei erw.-treuer Schätzbarkeit von Lb müssen also die Konstanzbereiche von L über denen von X liegen. Nun sind aber X und L linear, und deshalb reicht es, die Konstanzbereiche der Null, d.h. die Kerne, zu vergleichen. Für erw.-treue Schätzbarkeit muß also notwendig $\ker L \supset \ker X$ gelten. Dies ist auch hinreichend[19]:

19 Eine Erwartungstreue "im weiteren Sinne", bei der auch eine vorliegende Streuungsannahme berücksichtigt wird, behandelt RAO (1973). $\ker L \supset \ker X$ ist dabei hinreichend, aber nicht notwendig für Erwartungstreue.

Satz 26: Erw.-treue Schätzbarkeit

i (Kriterium für Erw.-treue)

$\hat{L}Y$ ist genau dann erw.-treu, wenn $\hat{L}X = L$.

ii (Erw.-treue Schätzbarkeit)

Lb ist genau dann erw.-treu schätzbar,
wenn $LX^+X = L$, d.h. $\ker L \supset \ker X$.

iii (Spezialfall $L = I_p$)

b ist genau dann erw.-treu schätzbar, wenn $\operatorname{rk} X = p$.

Bew.: i. Aus $E_p Y = Xb$ folgt $E_p \hat{L}Y = \hat{L}Xb$, bei Erw.-treue folgt zudem $E_p \hat{L}Y = Lb$. Die Gleichung $\hat{L}Xb = Lb$ gilt also zunächst für alle $b \in X^{-1}E_{\mathfrak{P}}Y$, und wegen der Linearität dann weiter für alle $b \in \operatorname{span} X^{-1}E_{\mathfrak{P}}Y \underset{\text{GMM3}}{=} X^{-1}\operatorname{span} E_{\mathfrak{P}}Y = X^{-1}\operatorname{im} X = R^p$; dies ist $\hat{L}X \underset{1.iii:15}{=} L$.

ii. I. Ein erw.-treuer Schätzer $\hat{L}Y$ existiert genau dann, wenn $\hat{L}X = L$ lösbar ist in $\hat{L}$, d.h. $X^T x = L^T$ lösbar, d.h. (22:41) $X^T X^{T+} L^T = L^T$, $L = LX^+X$.

II. Nach dem Teilraumkriterium 22:41 ist $X^T X^{T+} L^T = L^T$ äquivalent mit $\operatorname{im} L^T \subset \operatorname{im} X^T$, d.h. (1.i:15) $\ker L = \operatorname{im}^{\perp} L^T \supset \operatorname{im}^{\perp} X^T = \ker X$.

iii. b ist nach ii genau dann erw.-treu schätzbar, wenn $I_p = I_p X^+ X \underset{15.iv:33}{=} (X^T X)^+ X^T X$; die GRAMsche Matrix $X^T X$ ist genau dann regulär, wenn $\operatorname{rk} X = p$. _/

Für Linearformen $L \in M(1,p)$ wird $\operatorname{im} L^T = \operatorname{span} L^T \subset \operatorname{im} X^T$ häufig angegeben als: L^T liegt im Bild von X^T .

Xb ist immer erw.-treu schätzbar: $XX^+X = X$.

$\operatorname{rk} X = p$ impliziert $n \geq p$: notwendig zur erw.-treuen Schätzbarkeit von b ist, daß die Anzahl n der Beobachtungen nicht kleiner ist als die Anzahl p der Regressoren.

4.3 Minimum Bias - lineare Schätzfunktionen

Die Minimierung des Bias verallgemeinert die Erw.-treue und ersetzt sie bei nicht erw.-treuen Schätzfunktionen.
Die Voraussetzungen seien dieselben wie in Abschnitt 4.2 .

Ist $\hat{L}Y$ nicht erw.-treu, so erscheint es im Blick auf 26.i:53 sinnvoll, hilfsweise den Abstand $\|\hat{L}X-L\|$ zu minimieren.

Dafür gibt es noch eine zweite Begründung. Man möchte den Schätzfehler in den ersten Momenten klein halten, d.h. den Bias des Schätzers $\hat{L}Y$ in b minimieren:
$\forall b \in R^p : \quad \|E_b\hat{L}Y - Lb\| = \|(\hat{L}X-L)b\| = \inf$.
Die Minimierung soll <u>gleichmäßig in b</u> geschehen, damit die Lösung unabhängig vom unbekannten Regressionsparameter b wird. Da aber $\|(\hat{L}X-L)b\| = \inf$ zu Lösungen führt, die von b abhängen, geht man mit der Standardabschätzung 13.iv:28 $\|(\hat{L}X-L).b\| \leq \|\hat{L}X-L\|.\|b\|$ – diese Abschätzung ist scharf – über zu der neuen Minimierungsaufgabe $\|\hat{L}X-L\| = \inf$ (1. Stufe des Risikos):

<u>Def. 7:</u> (Min.Bias-Schätzer für Lb)
$\hat{L}Y$ heiße "<u>minimal verzerrt</u>", "<u>Minimum Bias - lineare Schätzfunktion für Lb</u>", "<u>Min.Bias-Schätzer für Lb</u>", falls
$\|\hat{L}X-L\| = \inf \{ \|\check{L}X-L\| : \check{L} \in M(1,n) \}$. _/

Min.Bias-Schätzer für Lb existieren offenbar immer. Sie fallen mit den erw.-treuen genau dann zusammen, wenn Lb erw.-treu schätzbar ist.

Die Klasse aller Min.Bias-Schätzer für Lb ist per Definition der Lösungsraum einer Minimierungsaufgabe und kann daher mit den in 24:43 bereitgestellten Mitteln bestimmt werden:

<u>Satz 27: Alle Min.Bias-Schätzer für Lb</u>
Die Menge aller Min.Bias-Schätzer für Lb ist
$\{ LX^+Y + ZMY : Z \in M(1,n) \}$, $M = I_n - XX^+$.

<u>Bew.:</u> $\hat{L}Y$ ist Min.Bias-Schätzer für Lb
$\Leftrightarrow \quad \|\hat{L}X-L\| = \inf$
$\Leftrightarrow \quad \|X^T\hat{L}^T - L^T\| = \inf$
$\Leftrightarrow \quad \hat{L}^T \in Lsg(\|X^Tx-L^T\|=\inf)$
$\Leftrightarrow \quad \hat{L}^T \in \{ X^{T+}L^T + (I_n - X^{T+}X^T)Z : Z \in M(n,1) \}$
24.ii:43
$\Leftrightarrow \quad \hat{L} \in \{ LX^+ + Z(I_n - XX^+) : Z \in M(I,n) \}$ _/

4.4 Minimum Norm - Minimum Bias - lineare Schätzfunktionen

Die Voraussetzungen seien dieselben wie in Abschnitt 4.2 .

Unter allen Min.Bias-Schätzern für Lb "sollte" der (oder die) Schätzer ausgewählt werden, der (komponentenweise) minimale Varianz hat. Wir werden in Satz 30:59 zeigen, daß komponentenweise Minimierung der Varianz dasselbe ist wie die Minimierung der Spur $\operatorname{tr} D\hat{L}Y$ der Streuungsmatrix von $\hat{L}Y$. Unter der Voraussetzung, daß die Streuung von Y bekannt ist: $DY = S^2$, gilt
$\operatorname{tr} D\hat{L}Y = \operatorname{tr} \hat{L}.DY.\hat{L}^T = \operatorname{tr} \hat{L}S^2\hat{L}^T = \operatorname{tr} \hat{L}S(\hat{L}S)^T = \|\hat{L}S\|^2$.
Da nun aber die Streuung S^2 <u>nicht</u> bekannt ist - es ist ein GM-Modell (Y,Xb,n,p) vorausgesetzt -, erreichen wir mit der Standardabschätzung 13.iv:28 - ähnlich wie in 4.3 - Unabhängigkeit vom unbekannten S^2 : $\|\hat{L}S\|^2 \leqq \|\hat{L}\|^2.\|S\|^2$. Im GM-Modell (Y,Xb,n,p) erscheint es also sinnvoll, die Norm $\|\hat{L}\|$ zu minimieren (2. Stufe des Risikos):

<u>Def. 8:</u> (Min.Norm-Min.Bias-Schätzer für Lb)
$\overset{*}{L}Y$ heiße "<u>Minimum Norm - Minimum Bias - lineare Schätzfunktion für Lb</u> ", "<u>Min.Norm-Min.Bias-Schätzer für Lb</u> ",falls
$\overset{*}{L}Y$ Min.Bias-Schätzer für Lb ist, und
$\hat{L}$ minimale Norm unter allen Min.Bias-Schätzern hat:
$\|\hat{L}\| = \inf \{ \|\check{L}\| : \check{L}Y \text{ Min.Bias-Schätzer für } Lb \}$. _/

In Satz 28 berechnen wir den Min.Norm-Min.Bias-Schätzer und geben dann in Satz 29 Beispiele:

<u>Satz 28: Hauptsatz über Min.Norm-Min.Bias-Schätzer für Lb</u>
In einem GM-Modell (Y,Xb,n,p) sei Lb eine lineare Funktion des Regressionsparameters b : $L \in M(1,p)$.
Dann ist LX^+Y der einzige Min.Norm-Min.Bias-Schätzer für Lb .

<u>Bew.:</u> Wegen $\|A\| = \|A^T\|$ kann man gleichwertig nach Min.Norm-Elementen $\hat{L}^T$ in (27:54): $\{ LX^+ + ZM : Z \in M(1,n) \}^T$ suchen; nach 23.ii:42 ist dies $\hat{L}^T = (LX^+)^T$, $\hat{L} = LX^+$. _/

Satz 29: Beispiele für Min.Norm-Min.Bias-Schätzer

i (Schätzung von b : $L = I_p$)

X^+Y ist der Min.Norm-Min.Bias-Schätzer für b , er ist genau dann erw.-treu, wenn X Spaltenrang p hat.

ii (Schätzung von Xb : $L = X$)

$XX^+Y = \mathrm{proj}(Y/\mathrm{im}\, X)$ ist der Min.Norm-Min.Bias-Schätzer für Xb , er ist erw.-treu; er ist invariant unter allen Designs X mit demselben Bild im X .

iii (Gleiche Mittelwerte μ : $X = 1_n$)

Wenn die Komponenten von Y denselben Mittelwert μ haben, dann ist das Stichprobenmittel $\bar{Y} = \frac{1}{n} \Sigma Y_\nu$ der Min.Norm-Min.Bias-Schätzer für μ , er ist erw.-treu.

Bew.: i und ii mit Hauptsatz 28:55 und Satz 26:53.

iii. Für das GM-Modell $(Y, 1_n\mu, n, 1)$, $1_n := (1, \dots, 1)^T$ gilt (16.i:34)

$(1_n)^+Y = \frac{1}{n}(1, \dots, 1)Y = \frac{1}{n} \Sigma Y_\nu = \bar{Y}$. _/

Min.Norm-Min.Bias-Schätzer sind nicht robust: Wegen $X^{++} = X$ gehört X^+Y einzig und allein zum GM-Modell (Y, Xb, n, p) .

4.5 Die Methode der kleinsten Quadrate

Der Min.Norm-Min.Bias-Schätzer X^+Y für b fällt zusammen mit dem Kleinste-Quadrate-Schätzer. Zur Methode der kleinsten Quadrate gibt es drei Zugänge: über die Minimierung der Residuen, die Maximum-Likelihood-Methode und den Satz von GAUSS-MARKOFF.

Minimierung der Residuen: "Unter allen Prinzipien, die zu diesem Zweck vorgeschlagen werden können, halte ich keines für allgemeiner, genauer und einfacher in der Anwendung als dasjenige, ... bei dem die Summe der Quadrate der Fehler minimiert wird." (LEGENDRE 1805; Übersetzung F.P. nach SMITH 1959:577).

Die Minimierung der Summe der Residuenquadrate $\|Y - X.\hat{b}(Y)\|^2$ führt nach Differentiation zu den "Normalgleichungen"

$X^TX.\hat{b}(Y) = X^TY$, die lösbar sind (22:41 und 15.iv:33):

$X^T\underline{X(X^TX)^+}X^TY = X^TX^{T+}X^TY = X^TY$, mit den Lösungen (23:42, 15.iv:33):

$\hat{b}(Y) \in \{ X^+Y + (I_p - X^+X)z : z \in R^p \}$; beim klassischen Fall $\mathrm{rk}\, X = p$ gibt das den einzigen Schätzer (16.iv:34)

$\hat{b}(Y) = X^+Y = (X^TX)^{-1}X^TY$.

Maximum-Likelihood-Methode: GAUSS erwähnt die Methode der kleinsten Quadrate zum ersten Mal im Jahre 1809 in der Theoria motus corpus coelestium, Art. 179 (1887:103f; Werke VII:245f). Er nimmt an, daß die Komponenten Y_ν unabhängig und normalverteilt sind mit Mittelwerten μ_ν und gleicher Varianz σ^2; unter solch "ideal statistical conditions" (ANSCOMBE 1961:2) maximiert er die Dichte. Bei dieser Maximum-Likelihood-Methode gelangt man wiederum zu den Normalgleichungen.

Satz von GAUSS-MARKOFF: Hat Y unkorrelierte Komponenten mit gleicher Varianz σ^2, dann führt nach dem Satz von GAUSS-MARKOFF (32.iii:62) die Minimierung der Varianz unter allen erw.-treuen linearen Schätzfunktionen ebenfalls zum besten Schätzer X^+Y. Diese zweite Begründung wählte GAUSS im Jahre 1821 in der Arbeit Theoria combinationis observationum erroribus minimis obnoxiae, pars prior, Art. 21 (1887:25ff; Werke IV:24f).

Jede dieser Rechtfertigungen hat ihre Nachteile. Die erste Art empfiehlt den Kleinste-Quadrate-Schätzer wegen seiner Plausibilität und nicht, weil er in einem statistischen Sinne optimal sei. Die zweite Art fordert zum Beweis der Optimalität ideale Bedingungen, die anzunehmen man nur selten berechtigt sein wird. Und auch die dritte Art setzt zur Begründung des Schätzers mehr voraus, als zur Existenz nötig ist. Denn berechnen kann man den Kleinste-Quadrate-Schätzer X^+Y ohne Streuungsannahme allein aus dem Mittelwert-Design X.

Mit dem zweistufigen Risiko:
Minimierung der Norm unter allen Min.Bias-Schätzern
haben wir ein entscheidungstheoretisches Konzept für die Methode der kleinsten Quadrate entwickelt, das zum Nachweis der Optimalität nicht mehr voraussetzt als zur Berechnung des Schätzers sowieso benötigt wird.-

In GAUSS-MARKOFF-Modellen vom allgemeinsten Typ (Y,Xb,n,p) ist die Min.Norm-Min.Bias-Schätzung, d.h. die Methode der kleinsten Quadrate, optimal. Allerdings kann man Besseres tun, wenn man Voraussetzungen über die Streuung hinzunimmt:

Kap. 5: Minimum Varianz - Minimum Bias - lineare Schätzfunktionen

In diesem Kapitel gehen wir ganz ähnlich vor wie im vorigen, setzen aber ein GM-Modell (Y, Xb, n, p, S^2) mit bekannter Streuung S^2 voraus. Sei Lb wieder eine lineare Funktion des Regressionsparameters b : $L \in M(l,p)$, und $\hat{L}Y$ eine lineare Schätzfunktion für Lb : $\hat{L} \in M(l,n)$.

5.1 Minimum Varianz - Minimum Bias - lineare Schätzfunktionen

Mit der Streuung von Y : $DY = S^2$ ist auch die des Schätzers $\hat{L}Y$ gegeben: $D\hat{L}Y = \hat{L}.DY.\hat{L}^T = \hat{L}S^2\hat{L}^T$. Somit können unter allen Min.Bias- Schätzern diejenigen mit kleinster Varianz gesucht werden, dies ist üblicherweise komponentenweise gemeint:

Def. 9: (Min.Var.-Min.Bias-Schätzer für Lb)

$\hat{L}Y$ heiße "Minimum Varianz - Minimum Bias - lineare Schätzfunktion für Lb ", "Min.Var.-Min.Bias-Schätzer für Lb ", falls

$\hat{L}Y$ Min.Bias-Schätzer für Lb ist, und

$\hat{L}Y$ komponentenweise minimale Varianz unter allen Min.Bias-Schätzern für Lb hat: $\forall \lambda \in \{1,\dots,l\}$

$\mathrm{Var}\,(\hat{L}Y)_\lambda = \inf \{ \mathrm{Var}\,(\check{L}Y)_\lambda : \check{L}Y$ Min.Bias-Schätzer für $Lb \}$ _/

Komponentenweise Varianzminimierung kann gleichwertig notiert werden mit: $\mathrm{Diag}\ D\hat{L}Y \leqq \mathrm{Diag}\ D\check{L}Y$, wobei $\leqq$ komponentenweise gemeint ist oder in dem Sinne, daß $\mathrm{Diag}\,(D\check{L}Y - D\hat{L}Y)$ p.s.d. ist.

Der folgende Satz 30 führt komponentenweise Minimierung der Varianz zurück auf eine einfache Aufgabe der linearen Algebra: Komponentenweise Varianzminimierung ist dasselbe wie Minimierung einer passenden EUKLIDschen (Halb-)Norm. Diese Aussage kann als Rechtfertigung dafür dienen, daß man in GAUSS-MARKOFF-Modellen nur EUKLIDische Normen betrachtet.

<u>Satz 30: Varianzminimierung als Normminimierung</u>

(AITKEN 1935:46) $\hat{L}Y$ Min.Var.-Min.Bias-Schätzer für Lb

$\Leftrightarrow$ $\operatorname{tr} D\hat{L}Y = \inf \{ \operatorname{tr} D\check{L}Y :$
$: \check{L}Y$ Min.Bias-Schätzer für $Lb \}$

$\Leftrightarrow$ $\| \hat{L}S \| = \inf \{ \| \check{L}S \| :$
$: \check{L}Y$ Min.Bias-Schätzer für $Lb \}$

<u>Bew.:</u> Die zweite Äquivalenz ist trivial, weil
$\operatorname{tr} D\hat{L}Y = \operatorname{tr} \hat{L}S(\hat{L}S)^T = \|\hat{L}S\|^2$.

Bei der ersten Äquivalenz sieht man sofort, daß mit komponentenweise kleinster Varianz auch $\operatorname{tr} D\hat{L}Y = \Sigma \operatorname{Var} (\hat{L}Y)_\lambda$ minimal ist. Daß die Umkehrung auch gilt, liegt an der affinen Struktur der Min.Bias-Schätzer (27:54). Nehmen wir (indirekt) an, $\check{L}Y$ sei ein Min.Bias-Schätzer für Lb und für eine Komponente λ sei $\operatorname{Var} (\check{L}Y)_\lambda < \operatorname{Var} (\hat{L}Y)_\lambda$. Nach Satz 27:54 haben $\hat{L}, \check{L}$ Darstelllungen
$\hat{L} = LX^+ + \hat{Z}M$, $\check{L} = LX^+ + \check{Z}M$. Setze $E_{\lambda\lambda} = E^{(1,1)}_{\lambda\lambda}$.

Betrachten wir nun $\hat{L}$, korrigiert um die λ-te Zeile von $\check{L}$:

$$\bar{L} := \hat{L} + E_{\lambda\lambda}(\check{L}-\hat{L}) = LX^+ + \hat{Z}M + E_{\lambda\lambda}\check{Z}M - E_{\lambda\lambda}\hat{Z}M =$$
$$= LX^+ + (\hat{Z}+E_{\lambda\lambda}(\check{Z}-\hat{Z}))M .$$

$\bar{L}Y$ ist nach 27:54 ein Min.Bias-Schätzer für Lb. Wegen $\operatorname{tr} E_{\lambda\lambda}AE_{\lambda\lambda} = \operatorname{tr} E_{\lambda\lambda}A = \operatorname{tr} AE_{\lambda\lambda}$, $A \in M(1)$, gilt für $D\bar{L}Y$:

$$\operatorname{tr} D\bar{L}Y = \operatorname{tr} \bar{L}S^2\bar{L}^T =$$
$$= \operatorname{tr} (\hat{L} + E_{\lambda\lambda}\check{L} - E_{\lambda\lambda}\hat{L})S^2(\hat{L}^T + \check{L}^TE_{\lambda\lambda} - \hat{L}^TE_{\lambda\lambda}) =$$
$$= \operatorname{tr} \hat{L}S^2\hat{L}^T + \operatorname{tr} \hat{L}S^2\check{L}^TE_{\lambda\lambda} - \operatorname{tr} \hat{L}S^2\hat{L}^TE_{\lambda\lambda} + \operatorname{tr} E_{\lambda\lambda}\check{L}S^2\hat{L}^T +$$
$$+ \operatorname{tr} E_{\lambda\lambda}\check{L}S^2\check{L}^T - \operatorname{tr} E_{\lambda\lambda}\check{L}S^2\hat{L}^T -$$
$$- \operatorname{tr} E_{\lambda\lambda}\hat{L}S^2\hat{L}^T - \operatorname{tr} E_{\lambda\lambda}\hat{L}S^2\check{L}^T + \operatorname{tr} E_{\lambda\lambda}\hat{L}S^2\hat{L}^T =$$
$$= \operatorname{tr} D\hat{L}Y + 0 + 0 + 0 + \operatorname{Var} (\check{L}Y)_\lambda - \operatorname{Var} (\hat{L}Y)_\lambda <$$
$$< \operatorname{tr} D\hat{L}Y + 0 \text{ , Widerspruch.}$$

Auf grund dieses Satzes führt die Suche nach Min.Var.-Min.Bias-Schätzern zu einer Minimierungsaufgabe solcher Art, wie sie in Kap. 3 (23:42) besprochen wurde:

Satz 31: Hauptsatz über Min.Var.-Min.Bias-Schätzer für Lb

In einem GM-Modell (Y, Xb, n, p, S^2) mit bekannter Streuung sei Lb eine lineare Funktion des Regressionsparameters b: $L \in M(1,p)$; setze $M = I_n - XX^+$. Dann gilt:

i (Alle Min.Var.-Min.Bias-Schätzer für Lb)

Die Menge aller Min.Var.-Min.Bias-Schätzer für Lb ist

$\{ LX^+(I_n - S(MS)^+)Y + Z(M - MS(MS)^+)Y : Z \in M(1,n) \}$.

ii (Fast sichere Einzigkeit)

Unter allen Min.Var.-Min.Bias-Schätzern für Lb ist

$LX^+(I_n - S(MS)^+)Y = LX^+(I_n - S^2(MS^2M)^+)Y$ fast sicher einzig.

iii (Einzigkeit)

$LX^+(I_n - S(MS)^+)Y$ ist der einzige Min.Var.-Min.Bias-Schätzer für Lb genau dann, wenn $\operatorname{im} M \subset \operatorname{im} S^2$.

Bew.: i. $\hat{L}Y$ Min.Var.-Min.Bias-Schätzer für Lb

$\Leftrightarrow$ $\hat{L}Y$ Min.Bias-Schätzer für Lb $\wedge$ $\|\hat{L}S\| = \inf$

30:59 $\Leftrightarrow$ $\exists \hat{Z} \in M(1,n)$: $\hat{L} = LX^+ + \hat{Z}M$ $\wedge$ $\|LX^+S + \hat{Z}MS\| = \inf\limits_{Z \in M(1,n)} \|LX^+S + ZMS\|$

27:54 $\Leftrightarrow$ $\exists \hat{Z} \in M(1,n)$: $\hat{L} = LX^+ + \hat{Z}M$ $\wedge$ $\|SM\hat{Z}^T - (-SX^{T+}L^T)\| = \inf$

$\Leftrightarrow$ $\hat{L} = LX^+ + \hat{Z}M$ $\wedge$ $\hat{Z}^T \in \operatorname{Lsg}(\|SMx - (-SX^{T+}L^T)\| = \inf)$

$\Leftrightarrow$ $\hat{L} = LX^+ + \hat{Z}M$ $\wedge$ $\exists Z \in M(n,1)$: $\hat{Z}^T = (SM)^+(-SX^{T+}L^T) + (I_n - (SM)^+(SM))Z$

24.ii:43 $\Leftrightarrow$ $\hat{L} = LX^+ + \hat{Z}M$ $\wedge$ $\exists Z \in M(1,n)$: $\hat{Z} = -LX^+S(MS)^+ + Z(I_n - MS(MS)^+)$

$\Leftrightarrow$ $\exists Z \in M(1,n)$: $\hat{L} = LX^+(I_n - S(MS)^+) + Z(M - MS(MS)^+)$.

17.v:34

ii. I. Die Gleichheit folgt aus (17.v:34; 15.iv:33):

$S\underline{(MS)^+} = S.S\underline{M(MS.SM)^+} = S^2(MS^2M)^+$.

II. Zur f.s.-Einzigkeit zeigen wir, daß bei jedem Wahrscheinlichkeitsmaß $P \in \mathfrak{P}$ die Zufallsgröße $Z(M - MS(MS)^+)Y$ P-f.s. verschwindet:

$E_p(\underline{M - MS(MS)^+})Y = (M - MS(MS)^+).\underline{M.E_pY} = 0$, und

$D_p(M - MS(MS)^+)Y = \underline{(I_n - MS(MS)^+).M.S}S.M.(I_n - MS(MS)^+) = 0$.

iii I. $M - MS(MS)^+ = \operatorname{proj}(./\operatorname{im} M \cap \ker S^2)$; denn:

$M - MS(MS)^+ = M(I_n - MS(MS)^+) = (I_n - MS(MS)^+)M$ (17.v:34),

daher mit 2.ii:16, 18.i:35 und $y = My$ $\forall y \in \operatorname{im} M$ und 1.ii:15

$M - MS(MS)^+ = \operatorname{proj}(./\operatorname{im} M \cap \ker SM) = \operatorname{proj}(./\operatorname{im} M \cap \ker S) =$

$= \operatorname{proj}(./\operatorname{im} M \cap \ker S^2)$.

II. $LX^+(I_n - S(MS)^+)$ ist einzig genau, falls $Z(M - MS(MS)^+) = 0$ für alle Z, d.h. $M - MS(MS)^+ = \operatorname{proj}(./\operatorname{im} M \cap \ker S^2) = 0$, d.h. $\operatorname{im} M \cap \ker S^2 = \{0\}$, d.h. $\operatorname{im} M \subset \ker^\perp S^2 = \operatorname{im} S^2$ (1.i:15). _/

Die Zerlegung $LX^+Y - LX^+S(MS)^+Y + Z(M-MS(MS)^+)Y$ kann man so deuten: Ein Min.Var.-Min.Bias-Schätzer ist die Summe aus

- dem Min.Norm-Min.Bias-Schätzer (Kleinste-Quadrate-Schätzer) LX^+Y ,
- einem "Korrektur-Schätzer" $-LX^+S\underline{(MS)^+}Y = -LX^+S(MS)^+.MY$, der den streuungsrelevanten Anteil MY nach Vorschrift des gegebenen Streuung-Designs S^2 wichtet, und
- einem nicht notwendig eindeutigem "Null-Schätzer" $Z(M-MS(MS)^+)Y$, der nicht nur im Mittel, sondern fast sicher Null ist.

Die drei Summanden stehen senkrecht aufeinander:

$$\mathrm{tr}\ LX^+(LX^+S(MS)^+)^T = \mathrm{tr}\ LX^+\underline{(SM)^+}SX^{T+}L^T = \mathrm{tr}\ L\underline{X^+M}(SM)^+X^{T+}L^T = 0,$$

$$\mathrm{tr}\ LX^+(Z(M-MS(MS)^+)^T = \mathrm{tr}\ LX^+(M-MS(MS)^+)Z^T = 0 ,$$

$$\mathrm{tr}\ LX^+S(MS)^+(Z(M-MS(MS)^+)^T = \mathrm{tr}\ LX^+S\underline{(MS)^+(M-MS(MS)^+)}Z^T = 0 .$$

Insbesondere hat $LX^+(I_n-S(MS)^+)Y$ minimale Norm unter allen Min.Var.-Min.Bias-Schätzern für Lb .

An $S(MS)^+ = S^2(MS^2M)^+$ sieht man, daß man die Wurzel S der Streuungsmatrix S^2 zur Rechnung nicht braucht (15.iv:33).

Wird der zum Mittelwert komplementäre Raum $\mathrm{im}\ M = \mathrm{im}^\perp X$ ganz von der Wichtung S^2 erfaßt: $\mathrm{im}\ M \subset \mathrm{im}\ S^2$, d.h. überdecken Mittelwert-Design X und Streuung-Design S^2 den ganzen Stichprobenraum: $\mathrm{im}\ X + \mathrm{im}\ S^2 = R^n$, so ist keine Vielfachheit möglich, und es liegt Einzigkeit vor.

Ist $\hat{L}Y$ ein Min.Var.-Min.Bias-Schätzer für Lb , so ist $\hat{L}XX^+Y$ der Min.Norm-Min.Bias-Schätzer (Kleinste-Quadrate-Schätzer) für Lb ; denn

$$LX^+(I_n-S(MS)^+)XX^+Y + Z(M-MS(MS)^+)XX^+Y = LX^+XX^+Y -0+0 = LX^+Y .$$

Min.Var.-Min.Bias-Schätzer sind invariant gegenüber positiven Vielfachen σ^2 der Streuungsmatrix S^2 ; dies folgt aus $\sigma S(\sigma MS)^+ = \sigma\sigma^{-1}S(MS)^+ = S(MS)^+$ (17.iii:34) .

Von daher gewinnt die Schätzung des Streuungsparameters σ^2 besondere Bedeutung (Satz 44:83).

Geben wir nun einige Beispiele für Min.Var.-Min.Bias-Schätzer:

Satz 32: Beispiele für Min.Var.-Min.Bias-Schätzer

i (Schätzung von b : $L = I_p$)

$X^+(I_n - S(MS)^+)Y$ ist Min.Var.-Min.Bias-Schätzer für b , er ist genau dann erw.-treu, wenn X Spaltenrang p hat.

ii (Schätzung von Xb : $L = X$)

$XX^+(I_n - S(MS)^+)Y$ ist erw.-treuer Min.Var.-Min.Bias-Schätzer für Xb , er ist invariant unter allen Designs X mit demselben Bild im X .

iii (Satz von GAUSS-MARKOFF)

Y habe unkorrelierte Komponenten mit gleichen Varianzen σ^2, X habe Spaltenrang p . Dann ist der Kleinste-Quadrate-Schätzer $X^+Y = (X^TX)^{-1}X^TY$ der einzige und erw.-treue Min.Var.-Min.Bias-Schätzer (BLUE) für b .

Bew.: i und ii mit Hauptsatz 31:60 und Satz 26:53.
iii. Für das GM-Modell $(Y, Xb, n, p, \sigma^2 I_n)$, $\mathrm{rk}\, X = p$ gilt (16.iv:34)
$X^+(1_n - \sigma I_n(\sigma M)^+)Y = X^+Y - \sigma\sigma^+\underline{X^+M} = X^+Y - 0 = (X^TX)^{-1}X^TY$. _/

Der Satz von GAUSS-MARKOFF repräsentiert eine allgemeinere Fragestellung: Wann kann man besonders einfache Min.Var.-Min.Bias-Schätzer wählen? Im folgenden Satz geben wir die vollständige Antwort für zwei Spezialfälle:
"AITKEN-Schätzung": Wann ist der Min.Norm-Min.Bias-Schätzer (Kleinste-Quadrate-Schätzer) im AITKEN-Modell $(S^+Y, S^+Xb, n, p, S^+S)$ ein Min.Var.-Min.Bias-Schätzer im ursprünglichen GM-Modell (Y, Xb, n, p, S^2) ?
"Satz von GAUSS-MARKOFF-RAO": Wann hat der Min.Norm-Min.Bias-Schätzer des ursprünglichen Modells gleichzeitig minimale Varianz?

Satz 33: Spezielle Darstellungen für Min.Var.-Min.Bias-Schätzer

i (AITKEN-Schätzung)

$(S^+X)^+S^+Y = (X^TS^{2+}X)^+X^TS^{2+}Y$ ist genau dann Min.Var.-Min.Bias-Schätzer für b (im GM-Modell (Y, Xb, n, p, S^2)), wenn $\mathrm{im}\, X \subset \mathrm{im}\, S^2$.

ii (Satz von GAUSS-MARKOFF-RAO)

(RAO 1967:364) X^+Y ist genau dann Min.Var.-Min.Bias-Schätzer für b , wenn $\mathrm{im}\, S^2M \subset \mathrm{im}\, M$, d.h. $\mathrm{im}\, S^2X \subset \mathrm{im}\, X$.

<u>Bew.:</u> i. Die Gleichheit gilt nach 15.iv:33 und 17.iv:34 :
$(S^+X)^+S^+Y = (\ (S^+X)^TS^+X\ )^+\ (S^+X)^TS^+Y = (X^TS^{2+}X)^+X^TS^{2+}Y$.

$\Rightarrow$: Sei $(S^+X)^+S^+Y$ Min.Var.-Min.Bias-Schätzer, also für ein Z :
$(S^+X)^+S^+ = X^+(I_n-S(MS)^+) + Z(M-MS(MS)^+)$.

I. Rechtsmultiplikation mit SS^+X gibt:

$$(S^+X)^+\underline{S^+SS^+}X = (S^+X)^+S^+X =$$
$$= X^+(I_n-S(MS)^+)SS^+X + Z(M-MS(MS)^+).M.SS^+X =$$
$$= X^+S(I_n-(MS)^+MS)S^+X + 0 .$$

Rechtsmultiplikation mit X gibt:

$$(S^+X)^+S^+X = X^+(I_n-S(MS)^+)X + Z(M-MS(MS)^+)X = X^+X -0+0 .$$

Zusammen heißt das:

$$X^+X = (S^+X)^+S^+X = X^+S(I_n-(MS)^+MS)S^+X .$$

II. Die letzte Gleichung wird von links mit S^+X multipliziert:

$$S^+XX^+X = S^+X =$$
$$= S^+\underline{X.X^+}S(I_n-(MS)^+MS)S^+X =$$
$$= S^+(I_n-M)S(I_n-(MS)^+MS)S^+X =$$
$$= (\ S^+S - S^+S(MS)^+MS\ \underline{- S^+MS + S^+MS(MS)^+MS}\ )S^+X =$$
$$= S^+SS^+X - S^+S\underline{(MS)^+MS}S^+X -0 =$$
$$= S^+X - \underline{S^+S}SM(SM)^+S^+X =$$
$$= S^+X - SS^+SM(SM)^+S^+X =$$
$$= S^+X - \underline{SM(SM)^+}S^+X =$$
$$= S^+X - (MS)^+MSS^+X \text{ , also}$$

$$(MS)^+MSS^+X = 0 .$$

III. Mit diesem Ergebnis zurück nach I zeigt, daß

$$X^+X = X^+SS^+X - X^+S.\underline{(MS)^+MSS^+X} = X^+SS^+X -0 \text{ , also}$$
$$XX^+ = X.X^+X.X^+ = XX^+SS^+XX^+ .$$

IV. Es folgt nun $SS^+XX^+ - XX^+ = 0$; denn:

$$\mathrm{tr}\ (SS^+XX^+-XX^+)(SS^+XX^+-XX^+)^T = \mathrm{tr}\ (SS^+XX^+-XX^+)(XX^+SS^+-XX^+) =$$
$$= \mathrm{tr}\ SS^+XX^+SS^+ - SS^+XX^+ - XX^+SS^+ + XX^+ =$$
$$= \mathrm{tr}\ SS^+XX^+(SS^+-I_n) - XX^+SS^+ + XX^+SS^+XX^+ =$$
$$\overset{III}{=} \mathrm{tr}\ SS^+XX^+(SS^+-I_n) + XX^+SS^+(XX^+-I_n) =$$
$$\underset{kom.}{=} \mathrm{tr}\ \underline{(SS^+-I_n)SS^+}XX^+ + \underline{(XX^+-I_n)XX^+}SS^+ = 0 .$$

$SS^+XX^+ = XX^+$ bedeutet aber nach dem Teilraumkriterium 22:41 gerade $\mathrm{im}\ X \subset \mathrm{im}\ S = \mathrm{im}\ S^2$ (1.ii:15) .

$\Leftarrow$: Sei $\mathrm{im}\ X \subset \mathrm{im}\ S^2 = \mathrm{im}\ S$, d.h. $SS^+X = X$. Dann ist

V. $(S^+X)^+ = X^+S(I_n-(MS)^+MS)$; denn:

(PK1) $S^+\underline{X.X^+}S(I_n-(MS)^+MS) = S^+(I_n-M)S(I_n-(MS)^+MS) =$

$$= S^+S - S^+S(MS)^+MS\ \underline{- S^+MS + S^+MS(MS)^+MS} =$$
$$= S^+S - SS^+SM(SM)^+ = S^+S - SM(SM)^+ \in \mathrm{Sym}(n) ;$$

(PK2) $X^+S(I_n-(MS)^+MS).S^+X = X^+\underline{SS^+X} - X^+S(MS)^+M\underline{SS^+X} = X^+X -0 \in \mathrm{Sym}(p)$

(PK3) $S^+X.X^+S(I_n-(MS)^+MS).S^+X = S^+X.X^+X = S^+X$;

(PK4) $X^+S(I_n-(MS)^+MS).S^+X.X^+S(I_n-(MS)^+MS) = X^+X.X^+S(I_n-(MS)^+MS)$.

VI. $X^+SS^+ = X^+$; denn:

$\operatorname{im} X^{+T} \underset{18.ii:35}{=} \operatorname{im} X \subset \operatorname{im} S$, d.h. $SS^+X^{+T} = X^{+T}$, d.h. $X^+SS^+ = X^+$.

VII. $(MS)^+MSS^+ = (MS)^+$; denn:

$MS = S - \underline{X}X^+S = SS^+.S - SS^+.XX^+.S = SS^+MS$, und

$(MS)^+ = (SS^+MS)^+ \underset{17.v:34}{=} (SS^+MS)^+SS^+ = (MS)^+SS^+ = (MS)^+MSS^+$.

VIII. Zusammen gibt das:

$$(S^+X)^+S^+ \underset{V}{=} X^+S(I_n-(MS)^+MS)S^+ = X^+SS^+ - X^+S(MS)^+MSS^+ =$$

$$\underset{VI,VII}{=} X^+-X^+S(MS)^+ = X^+(I_n-S(MS)^+) .$$

Also ist $(S^+X)^+S^+Y$ Min.Var.-Min.Bias-Schätzer für b .

ii. I. Zur zweiten Äquivalenz:

$\operatorname{im} S^2X \subset \operatorname{im} X = \ker M \Leftrightarrow MS^2X = 0 \Leftrightarrow X^TS^2M = 0 \Leftrightarrow$

$\Leftrightarrow \operatorname{im} S^2M \subset \ker X^T \underset{1.i:15}{=} \operatorname{im}^\perp X = \operatorname{im} M$.

II. Zur ersten Äquivalenz:

X^+Y Min.Var.-Min.Bias-Schätzer für b

$\Leftrightarrow \exists Z\in M(p,n)$: $X^+ = X^+(I_n-S(MS)^+) + Z(M-MS(MS)^+) =$

$= X^+ - X^+S(MS)^+ + Z(M-MS(MS)^+)$

$\Leftrightarrow (M-MS(MS)^+)x = (X^+S(MS)^+)^T$ lösbar

$\Leftrightarrow (M-MS(MS)^+)(M-MS(MS)^+)(X^+S(MS)^+)^T \underset{22.41}{=}$

$(M-MS(MS)^+)(X^+S(MS)^+)^T = (X^+S(MS)^+)^T$ (16.ii:34)

$\Leftrightarrow X^+S\underline{(MS)^+M} - X^+S\underline{(MS)^+MS(MS)^+} = X^+S(MS)^+$

$\Leftrightarrow X^+S(MS)^+ = 0$

$\Leftrightarrow \operatorname{im} S(MS)^+ \subset \ker X^+ \underset{18.ii:35}{=} \ker X^T$

$\Leftrightarrow X^TS(MS)^+ = 0$

$\Leftrightarrow \operatorname{im} (MS)^+ \underset{18.ii:35}{=} \operatorname{im} (MS)^T = \operatorname{im} SM \subset \ker X^TS$

$\Leftrightarrow X^TS^2M = 0$

$\Leftrightarrow \operatorname{im} S^2M \subset \ker X^T = \operatorname{im}^\perp X = \operatorname{im} M$. ✓

Diesen Darstellungssatz kann man anschaulich deuten:

i. Zu einem S^+-transformierten Modell kann man genau dann übergehen, wenn das Mittelwert-Design X vollständig mit(pseudo)invertiert wird: $\operatorname{im} X \subset \operatorname{im} S^2$.

ii. Die vom Streuung-Design abhängende Wichtung $X^+S(MS)^+.MY$ verschwindet genau dann, wenn S^2 auf $\operatorname{im} M$ nichts zur Mittelwert-Schätzung beiträgt: $S^2(\operatorname{im} M) \subset \operatorname{im} M$.

Satz 34: Beispiele zu Satz 33

i (AITKEN 1935)

X habe Spaltenrang p, und S^2 sei p.d. . Dann ist $(X^TS^{-2}X)^{-1}X^TS^{-2}Y$ der einzige und erw.-treue Min.Var.-Min.Bias-Schätzer (BLUE) für b .

ii (LEWIS & ODELL 1966)

S^2 sei p.d. . Dann ist $(X^TS^{-2}X)^+X^TS^{-2}Y$ der einzige Min.Var.-Min.Bias-Schätzer für b .

iii (33.i als Beispiel für 33.ii)

$(S^+X)^+S^+Y$ ist Min.Var.-Min.Bias-Schätzer für b im AITKEN-Modell (S^+Y,S^+Xb,n,p,S^+S) .

Bew.: iii. Der Min.Norm-Min.Bias-Schätzer $(S^+X)^+S^+Y$ des AITKEN-Modells hat nach 33.ii genau dann minimale Varianz (im AITKEN-Modell), wenn im $S^+S.S^+X \subset$ im S^+X ; dies gilt immer. _/

Wegen des Satzes von GAUSS-MARKOFF (32.iii:62) könnte man 33.ii:62 auch so formulieren: Wann hat der Min.Var.-Min.Bias-Schätzer für b im GM-Modell $(Y,Xb,n,p,\sigma^2 I_n)$ gleichzeitig minimale Varianz in einem zweiten GM-Modell (Y,Xb,n,p,S^2) ?

Satz 35: Robustheit gegen Fehler im Streuung-Design

Gegeben seien zwei GM-Modelle (Y,Xb,n,p,S^2) , (Y,Xb,n,p,T^2) mit gleichem Mittelwert-Design X und verschiedenem Streuung-Design S^2,T^2 .

i (Robustheit von $X^+(I_n-S(MS)^+)Y$)

$X^+(I_n-S(MS)^+)Y$ ist Min.Var.-Min.Bias-Schätzer für b im GM-Modell (Y,Xb,n,p,T^2) genau dann, wenn im $T^2M \subset \ker X^+(I_n-S(MS)^+)$.

ii (Robustheit aller Min.Var.-Min.Bias-Schätzer)

(RAO 1971c:387) Jeder Min.Var.-Min.Bias-Schätzer für b im GM-Modell (Y,Xb,n,p,S^2) ist auch Min.Var.-Min.Bias-Schätzer im GM-Modell (Y,Xb,n,p,T^2) genau dann, wenn im $T^2M \subset$ im S^2M .

Bew.: I. Wir geben zunächst ein Kriterium, wann $X^+(I_n-S(MS)^+)Y + Z(M-MS(MS)^+)Y$ minimale Varianz im GM-Modell (Y,Xb,n,p,T^2) hat. Dies ist genau dann der Fall, wenn die Gleichung $X^+(I_n-S(MS)^+)+Z(M-MS(MS)^+) = X^+(I_n-T(MT)^+)+x(M-MT(MT)^+)$, bzw. $(M-MT(MT)^+)x = (X^+T(MT)^+ - X^+S(MS)^+ + Z(M-MS(MS)^+))^T$ lösbar ist.

D.h. (22:41) $(M-MT(MT)^+)(\ldots)^T = (\ldots)^T$, oder transp oniert

$X^+T(MT)^+ - X^+S(MS)^+ + Z(M-MS(MS)^+) =$

$= X^+T\underline{(MT)^+M} - X^+S\underline{(MS)^+M} + Z\underline{(M-MS(MS)^+)M} -$

$- X^+T\underline{(MT)^+MT(MT)^+} + X^+S\underline{(MS)^+MT(MT)^+} - Z\underline{(M-MS(MS)^+)MT}(MT)^+$,

d.h. $0 = - X^+T(MT)^+ + X^+S(MS)^+.T(MT)^+ - Z(M-MS(MS)^+).T(MT)^+$,

d.h. $0 = - \left(X^+(I_n-S(MS)^+)+Z(M-MS(MS)^+) \right).T(MT)^+$,

mit 18.ii:35 ($\operatorname{im}(MT)^+ = \operatorname{im}(MT)^T = \operatorname{im} TM$) gibt das als Kriterium

$0 = \left(X^+(I_n-S(MS)^+) + Z(M-MS(MS)^+) \right).T^2M$.

II. i folgt sofort für $Z=0$: $X^+(I_n-S(MS)^+).T^2M$ verschwindet genau bei $\operatorname{im} T^2M \subset \ker X^+(I_n-S(MS)^+)$.

III. Nach I, II ist ii gleichwertig dazu, daß

$\operatorname{im} T^2M \subset \ker X^+(I_n-S(MS)^+)$ <u>und</u> $\forall Z\in M(p,n)\ \ Z(M-MS(MS)^+).T^2M = 0$.

Letzteres ist äquivalent zu $(M-MS(MS)^+).T^2M = 0$, d.h.

$\operatorname{im} T^2M \subset \ker M-MS(MS)^+ \underset{1.i:15}{=} \operatorname{im}^\perp M-MS(MS)^+ \underset{31.iii(I):60}{=} (\operatorname{im} M\cap\ker S^2)^\perp =$

$= \operatorname{im}^\perp M + \ker^\perp S^2 = \operatorname{im} X + \operatorname{im} S^2$.

Behauptung ii gilt also genau dann, wenn

$\operatorname{im} T^2M \subset (\operatorname{im} X + \operatorname{im} S^2) \cap \ker X^+(I_n-S(MS)^+)$.

IV. $\operatorname{im} X \cap \ker X^+(I_n-S(MS)^+M) \underset{MX=0}{=} \operatorname{im} X \cap \ker X^+ \underset{18.ii:35}{=}$

$= \operatorname{im} X \cap \ker X^T = \operatorname{im} X \cap \operatorname{im}^\perp X = \{0\}$.

V. $\operatorname{im} S^2 \cap \ker X^+(I_n-S(MS)^+) = \operatorname{im} S \cap \ker X^+(I_n-S(MS)^+) =$

$= \{ y\in R^n : \exists x\in R^n \ \ y=Sx \ \wedge \ X^+(I_n-S(MS)^+)Sx = 0 \} =$

$= \{ Sx\in R^n : x \in \operatorname{Lsg}(X^+S(I_n-(MS)^+MS)x=0) \} =$

$= \{ Sx\in R^n : x \in \operatorname{Lsg}(\underline{XX^+}S(I_n-(MS)^+MS)x=0) \} =$

$= \{ Sx\in R^n : x \in \operatorname{Lsg}((I_n-M)S(I_n-(MS)^+MS)x=0) \} =$

$= \{ Sx\in R^n : x \in \operatorname{Lsg}(S(I_n-(MS)^+MS)x -0x \ =0) \} =$

$= \left\{ Sx\in R^n : x \in \{ (I_n-(S(I_n-(MS)^+MS))^+S(I_n-(MS)^+MS))z : z\in R^n \} \right\} =$

$= \{ Sz - S.\underline{(S(I_n-(MS)^+MS))^+}S(I_n-(MS)^+MS)z : z\in R^n \} =$

$= \{ Sz - S(I_n-(MS)^+MS).(S(I_n-(MS)^+MS))^+.S(I_n-(MS)^+MS).z : z\in R^n \} =$

$= \{ Sz - S(I_n-(MS)^+MS)z : z \in R^n \} =$

$= \{ Sz - Sz + S(MS)^+MSz : z \in R^n \} =$

$= \operatorname{im} S(MS)^+MS = S(\operatorname{im}(MS)^+MS) \underset{18.i:35}{=} S(\operatorname{im} SM) = \operatorname{im} S^2M$.

Nach III, IV, V gilt ii genau dann, wenn $\operatorname{im} T^2M \subset \{0\}+\operatorname{im} S^2M$. _/

Wie lautet ein Kriterium für Robustheit gegen Fehler im Mittelwert-Design bei gleichem Streuung-Design und für Robustheit gegen Fehler in Mittelwert- <u>und</u> Streuung-Design? Die Antworten darauf stehen noch aus; für die speziellen Streuungsmatrizen σ^2I_n vergleiche RAO & MITRA (1971:162-165).

5.2 Von LEGENDREs "Méthode des moindres quarrés" zu RAOs "Unified Theory of Linear Estimation"

LEGENDRE (1805) beschrieb die "Méthode des moindres quarrés" als erster, gab ihr den Namen und empfahl sie wegen ihrer Einfachheit (vgl. Abschnitt 4.5).

GAUSS kannte nach eigenem Bekunden (GAUSS an LAPLACE; Werke X,1:373) dieses Verfahren zwar schon 1795 und hatte die von ihm in der Theoria motus dargestellte Maximum-Likelihood-Begründung schon 1798 gefunden; aber seine Veröffentlichung im Jahre 1809 kam nach der LEGENDREs. Wem auch immer die Methode der kleinsten Quadrate zugerechnet werden mag (vgl. PLACKETT 1972), die beiden in Abschnitt 4.5 dargestellten klassischen wahrscheinlichkeitstheoretischen Rechtfertigungen hat zweifellos GAUSS beigetragen.

MARKOFF widmete der Methode der kleinsten Quadrate das 7. Kapitel seiner "Wahrscheinlichkeitstheorie" (1912), und im "Satz von GAUSS-MARKOFF" wird sein Name mitgenannt. Aber eher halte ich PLACKETTs Urteil (1949:460) für angemessen: "... MARKOFF, who refers to GAUSS' work, may perhaps have clarified assumptions implicit there but proved nothing new."

Beim Satz von GAUSS-MARKOFF setzt man meistens voraus ein GM-Modell $(Y,Xb,n,p,\sigma^2 I_n)$, rk $X = p$. Dann ist der Kleinste-Quadrate-Schätzer $X^+Y = (X^TX)^{-1}X^TY$ ein BLUE, d.h. er hat minimale Varianz unter allen erw.-treuen linearen Schätzfunktionen.

Tatsächlich bewies GAUSS in seiner Theoria combinationis observationum, pars prior, Art. 19 mehr (1887:23; Werke IV:21f). Er nimmt (mit unseren Bezeichnungen) eine p.d. Diagonalmatrix als Streuungsmatrix S^2 und berechnet den besten Schätzer, indem er auf das mit S^{-1} – transformierte Modell die Methode der kleinsten Quadrate anwendet.

AITKEN (1935) benutzte die Matrizenschreibweise und zeigte damit für p.d. Streuungsmatrizen S^2 allgemein, daß die Methode der kleinsten Quadrate optimale Schätzer liefert, wenn man sie nur auf das mit S^{-1}-transformierte Modell anwendet.

RAO (1945:9) führt – unter Verweis auf Raj Chandra BOSE – den Begriff der erw.-treuen Schätzbarkeit ein. CHIPMAN (1964:1094ff) überträgt die Abstandsminimierung $\|\hat{L}X-L\|=\inf$ auf die statistische Fragestellung und erweitert damit das Konzept der erw.-treuen Schätzbarkeit zur Min.Bias-Schätzung.

1965 beantwortet RAO (1967:364) die Frage, wann der Kleinste-Quadrate-Schätzer minimale Varianz hat (33.ii:62). Der beharrliche Einsatz von verallgemeinerten Matrixinversen – seit (1955) – führt ihn schließlich zu einer Formulierung der Schätztheorie in GM-Modellen, in der auf Rangannahmen verzichtet werden kann: eine "Unified Theory of Linear Estimation" (1971c).

WATSON (1972) behandelt die Mittelwert-Schätzung, ohne explizit verallgemeinerte Matrixinverse zu benutzen[20].–

RAO & MITRA (1971:148) erwähnen, daß $\mathrm{im}\, X \subset \mathrm{im}\, S^2$ hinreichend für AITKEN-Schätzung ist (33.i:62); hier haben wir auch die Notwendigkeit dieser Bedingung gezeigt.

Die Darstellung $X^+(I_n-S(MS)^+)Y$ für "den" BLUE von erw.-treu schätzbarem b gibt ALBERT (1972:90). RAO & MITRA (1971:148) nennen eine andere Formel; in ihrer Formel kommen generalisierte Inverse A^- vor, eine spezielle Wahl für A^- ist die Pseudoinverse A^+. Mit dieser Wahl wird der Zusammenhang der Schätzer von RAO & MITRA und ALBERT ganz einfach: sie sind gleich[21]. Der Beweis folgt aus der Darstellung für $X^T(S^2+XX^T)^+$ gemäß 21:39 :

<u>Satz 36: Der BLUE von RAO & MITRA</u>

$$X^+(I_n-S(MS)^+) \;=\; (\, X^T(S^2+XX^T)^+X\,)^+.X^T(S^2+XX^T)^+$$

<u>Bew.:</u> Setzen wir $B := X^+(I_n-S(MS)^+)$, dann gilt $X^+XB = B$, und $BX = X^+X$. Daher mit Satz 21:39 und 18.i:35 :

$$(\, X^T(S^2+XX^T)^+X\,)^+.X^T(S^2+XX^T)^+ = (\,(I_p+BS^2B^T)^{-1}\underline{B.X}\,)^+.(I_p+BS^2B^T)^{-1}\underline{B} =$$
$$= (\,(I_p+BS^2B^T)^{-1}X^+X\,)^+.(I_p+BS^2B^T)^{-1}X^+X.B =$$
$$= \mathrm{proj}(B/\mathrm{im}\, X^+X.(I_p+BS^2B^T)^{-1}) = \mathrm{proj}(B/X^+X(\,\mathrm{im}\,(I_p+BS^2B^T)^{-1}\,))=$$
$$= \mathrm{proj}(B/\mathrm{im}\, X^+X) = X^+XB = B\;.$$

_/

---

20 Eine Bibliographie zur Methode der kleinsten Quadrate steht bei MILLER (1973:719-726).

21 "At this time [März 1973], the question concerning the relationship between the matrices in [ALBERT 1972] and [RAO & MITRA 1971] remains open." (ALBERT 1973:183).

Min.Var.-Min.Bias-Schätzer sind die besten linearen Schätzfunktionen bzgl. des beschriebenen zweistufigen Risikos. "The word 'best' here is used in the purely formal sense of 'minimum variance' and does not imply approbation. On the contrary, if the minimum bias is large, it makes little sense to minimize the variance. For precisely this reason, an alternative criterion of minimum mean square error is taken up in [Kap. 6]." (CHIPMAN 1964:1096):

Kap. 6: Minimum mean square error - lineare Schätzfunktionen

In diesem Kapitel rechnen wir weiter in einem GM-Modell (Y,Xb,n,p,S^2) und schätzen lineare Funktionen Lb, $L\in M(1,p)$. Bias und Varianz eines Schätzers $\hat{L}Y$ für Lb, $\hat{L}\in M(1,n)$, werden gleichzeitig berücksichtigt, indem ihre Summe minimiert wird.

6.1 Minimum mean square error - lineare Schätzfunktionen

Bisher wurde auf einer ersten Stufe der Bias $\|\hat{L}X-L\|$ und dann erst mit $\|\hat{L}S\|$ die Varianz minimiert. Es liegt die Alternative nahe, beide Größen gleichzeitig zu minimieren: $\|\hat{L}S\|^2 + \|\hat{L}X-L\|^2 = \inf$; oder allgemeiner: ihre gewichtete Summe zu minimieren: $\|\hat{L}S\|^2 + \beta^2\|\hat{L}X-L\|^2 = \inf$, $\beta^2 > o$.

Es gibt eine bessere Begründung. Für das GAUSSsche Risiko (mean square error) gilt $E_b\|\hat{L}Y-Lb\|^2 = \|\hat{L}S\|^2 + \|(\hat{L}X-L)b\|^2$; denn $E_b\|\hat{L}Y-Lb\|^2 = E_b \operatorname{tr} \hat{L}Y(\hat{L}Y)^T - 2\hat{L}Y(Lb)^T + Lb(Lb)^T =$
$= \operatorname{tr} E_b\hat{L}Y(\hat{L}Y)^T - E_b\hat{L}Y(E_b\hat{L}Y)^T + E_b\hat{L}Y(E_b\hat{L}Y)^T - 2(E_b\hat{L}Y)(Lb)^T + Lb(Lb)^T =$
$= \operatorname{tr} D\hat{L}Y + \operatorname{tr} (E_b\hat{L}Y - Lb)(E_b\hat{L}Y - Lb)^T = \|\hat{L}S\|^2 + \|(\hat{L}X-L)b\|^2$.

Die Lösungen der Minimierungsaufgabe $\|\hat{L}S\|^2 + \|(\hat{L}X-L)b\|^2 = \inf$ hängen vom unbekannten Parameter b ab. Um Gleichmäßigkeit in b zu erreichen, wird wieder - wie in Abschnitt 4.3, 4.4 - die Standardabschätzung 13.iv:28 versucht:
$\|\hat{L}S\|^2 + \|(\hat{L}X-L)b\|^2 \leqq \|\hat{L}S\|^2 + \|\hat{L}X-L\|^2.\|b\|^2$.
(Für eindimensionale Parameter b gilt Gleichheit.)

Wechseln wir - dieses Vorgehen ist in der Literatur nicht üblich - zu der Minimierungsaufgabe $\|\hat{L}S\|^2 + \|\hat{L}X-L\|^2 \cdot \|b\|^2 = \inf$, so erreichen wir zwar nicht Unabhängigkeit vom zu schätzenden Regressionsparameter b , aber immerhin gelten die Lösungen gleichmäßig auf p-dimensionalen Sphären $S_p(\beta)$, $S_p(\beta) := \{ b\in R^p : \|b\| = \beta \}$. Damit ist gerechtfertigt die

Def. 10: i ($S_p(\beta)$-lokale GAUSS-Schätzer für Lb) $\forall \beta\in R_+$
$\hat{L}Y$ heiße "Minimum mean square error - lineare Schätzfunktion lokal auf $S_p(\beta)$ für Lb ", " $S_p(\beta)$-lokaler GAUSS-Schätzer für Lb ", falls

$$\|\beta^{-1}\hat{L}S\|^2+\|\hat{L}X-L\|^2 = \inf \{ \|\beta^{-1}\check{L}S\|^2+\|\check{L}X-L\|^2 : \check{L}\in M(1,n) \} .$$

ii (GAUSS-Schätzer für Lb)
$\hat{L}Y$ heiße "GAUSS-Schätzer für Lb ", falls
$\hat{L}Y$ $S_p(\beta)$-lokaler GAUSS-Schätzer für Lb ist für alle $\beta>0$.

GAUSS-Schätzer (ii) werden diejenigen lokalen GAUSS-Schätzer genannt, die nicht vom Lokalitätsparameter β abhängen.

Zur Berechnung lokaler GAUSS-Schätzer haben wir in Satz 25:44 ein Kriterium zur Minimierung quasi-innerer Produkt gezeigt:

Satz 37: Hauptsatz über $S_p(\beta)$-lokale GAUSS-Schätzer für Lb
In einem GM-Modell (Y,Xb,n,p,S^2) mit bekannter Streuung sei Lb eine lineare Funktion des Regressionsparameters b : $L\in M(1,p)$; setze $M = I_n-XX^+$, $B = X^+(1_n-S(MS)^+)$.
Dann gilt für alle positiven β :
i (Alle $S_p(\beta)$-lokalen GAUSS-Schätzer für Lb)
Die Menge aller $S_p(\beta)$-lokalen GAUSS-Schätzer für Lb ist

$$\{ L(I_p+\beta^{-2}BS^2B^T)^{-1}BY + Z(M-MS(MS)^+)Y : Z\in M(1,n) \} .$$

ii (Fast sichere Einzigkeit)
Unter allen $S_p(\beta)$-lokalen GAUSS-Schätzern für Lb ist
$L(I_p+\beta^{-2}BS^2B^T)^{-1}BY$ fast sicher einzig.
iii (Einzigkeit)
$L(I_p+\beta^{-2}BS^2B^T)^{-1}BY$ ist der einzige $S_p(\beta)$-lokale GAUSS-Schätzer für Lb genau dann, wenn im $M \subset$ im S^2 .

Bew.: Die Behauptungen ii, iii folgen wie im Hauptsatz 31:60, da die vielfachheit-erzeugenden Nullschätzer $Z(M-MS(MS)^+)Y$ dieselben sind. Es bleibt also nur i zu beweisen.

Setzen wir zur Abkürzung $S_\beta := \beta^{-1}S$. Es ist zu minimieren
$\|\hat{L}S_\beta\|^2 + \|\hat{L}X-L\|^2 = \operatorname{tr} \hat{L}S_\beta{}^2\hat{L}^T + \hat{L}X(\hat{L}X)^T - \hat{L}XL^T - \hat{L}XL^T + LL^T$.

I. $V(A,B) := \operatorname{tr} AS_\beta{}^2B^T + AX(BX)^T - AXL^T - BXL^T + LL^T$
ist quasi-inneres-Produkt; denn:

(Q1) $V(A,B)-V(0,B) = \operatorname{tr} AS_\beta{}^2B^T + AX(BX)^T - AXL^T$ ist linear in A .

(Q2) $V(A,B) = V(B,A)$ klar.

(Q3) $V(A,A) = \|AS_\beta\|^2 + \|AX-L\|^2 \geqq 0$.

II. Nach 25:44 ist $V(A,A)$ minimal genau dann, falls $\forall C \in M(1,n)$
$0 = V(A,C)-V(A,0) = V(C,A)-V(0,A) =$
$= \operatorname{tr} CS_\beta{}^2A^T + CX(AX)^T - CXL^T = \operatorname{tr} C.(S_\beta{}^2A^T+X(AX)^T-XL^T)$;
weil $\operatorname{tr} AB^T$ ein inneres Produkt ist, gilt äquivalent:
$0 = S_\beta{}^2A^T + XX^TA^T - XL^T$, $(S_\beta{}^2+XX^T)A^T = XL^T$. Diese Gleichung ist lösbar in A^T ; um das zu zeigen, beweisen wir erst

III. $I_n-(S_\beta{}^2+XX^T)^+(S_\beta{}^2+XX^T) = M-MS(MS)^+$; denn:
Es ist $\ker S_\beta{}^2+XX^T = \ker S^2 \cap \ker X^T$; mit $(S_\beta{}^2+XX^T)x = 0$ verschwindet nämlich auch
$x^T(S_\beta{}^2+XX^T)x = x^TS_\beta{}^2x + x^TXX^Tx = \|S_\beta x\|^2 + \|X^Tx\|^2$,
also ist $X^Tx = 0$ und mit $S_\beta x = \beta^{-1}Sx = 0$ auch $S^2x = 0$.
Mit 18.i:35 und 31.iii(I):60 folgt nun:
$I_n-(S_\beta{}^2+XX^T)^+(S_\beta{}^2+XX^T) = \operatorname{proj}(./\ker S_\beta{}^2+XX^T) =$
$= \operatorname{proj}(./\ker X^T \cap \ker S^2) = \operatorname{proj}(./\operatorname{im}^\perp X \cap \ker S^2) =$
$= \operatorname{proj}(./\operatorname{im} M \cap \ker S^2) = M-MS(MS)^+$.

IV. $(S_\beta{}^2+XX^T)x = XL^T$ ist lösbar; denn:
$(S_\beta{}^2+XX^T)^+(S_\beta{}^2+XX^T).XL^T = (I_n - (I_n-(S_\beta{}^2+XX^T)^+(S_\beta{}^2+XX^T))).XL^T =$
$= XL^T - \underline{(M-MS(MS)^+)XL^T} = XL^T - 0$ und 22:41 .

V. A minimiert also $V(A,A)$ genau dann, wenn
$A^T \in \operatorname{Lsg}(\ (S_\beta{}^2+XX^T)x=XL^T\ ) =$
$\overset{II}{=} \{\ (S_\beta{}^2+XX^T)^+XL^T + (I_n - (S_\beta{}^2+XX^T)^+(S_\beta{}^2+XX^T))Z : Z \in M(n,1)\ \}$,
mit III und der Formel 21:39 ist dies dasselbe wie
$A \in \{\ LX^T(S_\beta{}^2+XX^T)^+ + Z(M-MS(MS)^+) : Z \in M(1,n)\ \} =$
$= \{\ L(I_p+BS_\beta{}^2B^T)^{-1}B + Z(M-MS(MS)^+) : Z \in M(1,n)\ \} =$
$= \{\ L(I_p+\beta^{-2}BS^2B^T)^{-1}B + Z(M-MS(MS)^+) : Z \in M(1,n)\ \}$. _/

Mean square error - Schätzer werden auch von RAO (1971c:389) diskutiert. Allerdings macht er keine Verwendung der Standardabschätzung 13.iv:28 und minimiert $\|\hat{L}S\|^2+\|\hat{L}Xb-Lb\|^2$. Mit $Lbb^TX^T(S^2+Xbb^TX^T)^+Y$ erhält er daher einen Schätzer, der nur im Punkt b optimal ist. Man kann sein Ergebnis sofort aus V ablesen, wenn man das Lexikon $\begin{array}{c|c} \beta & 1 \\ X & Xb \\ L & Lb \end{array}$ setzt.

Aus praktisch-heuristischen Überlegungen heraus untersuchen HOERL & KENNARD (1970:57; vgl. auch MARQUARDT 1970) die Klasse der "Ridge estimator": $\{ (I_p+k(X^TX)^{-1})^{-1}X^+Y : k \geq o \}$. Dabei gehen sie aus von einem klassischen GM-Modell $(Y,Xb,n,p,\sigma^2 I_n)$, rk $X = p$. HOERL & KENNARD bemerken (1970:55), daß "(the procedure) can be used to obtain a point estimate with a smaller mean square error."

LINDLEY & SMITH (1972:11) erreichen dieselben Schätzer über einen BAYESschen Ansatz.

Im klassischen Fall $S^2 = \sigma^2 I_n$, rk $X = p$ wird der $S_p(\beta)$-lokale GAUSS-Schätzer für b gerade
$(1_p+\beta^{-2}\sigma^2 X^+X^{+T})^{-1}X^+Y = (1_p+(\sigma/\beta)^2(X^TX)^{-1})^{-1}X^+Y$,
also ein Ridge estimator. Daraus läßt sich eine gewisse Rechtfertigung für das hier gewählte Vorgehen ableiten.

6.2 Beziehungen zwischen den Schätzkonzepten

Die Abkehr vom üblichen Weg erlaubt es, die Beziehungen mit den beiden anderen Schätzkonzepten formelmäßig herzustellen. Das ist ein bedeutender Vorteil der hier entwickelten sphären-optimalen GAUSS-Schätzer gegenüber den klassischen punkt-optimalen Minimum mean square error-Schätzern.

Es ist schon erwähnt worden (Seite 61), wie man von einem Min.Var.-Min.Bias-Schätzer zum Min.Norm-Min.Bias-Schätzer gelangt. Deshalb reicht nun zu zeigen, wie Min.Var.-Min.Bias-Schätzer und lokale GAUSS-Schätzer auseinander hervorgehen:
i. Aus einem beliebigen Min.Var.-Min.Bias-Schätzer kann man zu jedem $\beta > o$ einen $S_p(\beta)$-lokalen GAUSS-Schätzer berechnen.
ii. Aus nur einem (!) lokalen GAUSS-Schätzer erhält man einen Min.Var.-Min.Bias-Schätzer:

Satz 38: Min.Var.-Min.Bias- und lokale GAUSS-Schätzer

i (lokale GAUSS-Schätzer aus Min.Var.-Min.Bias-Schätzern)
Sei $\hat{b}Y$ ein Min.Var.-Min.Bias-Schätzer für b.
Dann gilt für alle $\beta>0$:
$(I_p+\beta^{-2}\hat{b}S^2\hat{b}^T)^{-1}\hat{b}Y$ ist $S_p(\beta)$-lokaler GAUSS-Schätzer für b.

ii (Min.Var.-Min.Bias-Schätzer aus lokalen GAUSS-Schätzern)
Sei ein $\beta>0$ gegeben und $\hat{b}Y$ ein $S_p(\beta)$-lokaler GAUSS-Schätzer für b.
Dann ist $(\hat{b}X)^+\hat{b}Y$ Min.Var.-Min.Bias-Schätzer für b.

Bew.: Setze $B = X^+(I_n-S(MS)^+)$.

i. Nach Hauptsatz 31:60 hat $\hat{b}$ eine Darstellung
$\hat{b} = B + Z(M-MS(MS)^+)$. Wegen $(M-MS(MS)^+)S = 0$ gilt damit

$$(I_p+\beta^{-2}\hat{b}S^2\hat{b}^T)^{-1}\hat{b} = (I_p+\beta^{-2}(BS+0)(BS+0)^T)^{-1}(B+Z(M-MS(MS)^+)) =$$
$$= (I_p+\beta^{-2}BS^2B^T)^{-1}B + \tilde{Z}(M-MS(MS)^+),$$

nach Hauptsatz 37:70 ist dies ein $S_p(\beta)$-lokaler GAUSS-Schätzer.

ii. Aus Teil V:71 des Beweises zu Hauptsatz 37:70 folgt, daß $\hat{b}$ die Darstellung hat:
$\hat{b} = X^T(S_\beta{}^2+XX^T)^+ + Z(M-MS(MS)^+)$. Mit der Formel von RAO & MITRA (36:68) gibt das:

$$(\hat{b}X)^+\hat{b} = (X^T(S_\beta{}^2+XX^T)^+X + 0)^+X^T(S_\beta{}^2+XX^T)^+ + \tilde{Z}(M-MS(MS)^+ =$$
$$= X^+(I_n-S_\beta(MS_\beta)^+) + \tilde{Z}(M-MS(MS)^+) =$$
$$= B + \tilde{Z}(M-MS(MS)^+)$$

_/

Als Anwendung dieses Satzes können wir spezielle Darstellungen für lokale GAUSS-Schätzer finden:

Satz 39: Spezielle Darstellungen für lokale GAUSS-Schätzer

i (Über AITKEN-Schätzung) $\forall \beta\in R_+$
$(I_p+\beta^{-2}(X^TS^{2+}X)^+)^{-1}(X^TS^{2+}X)^+X^TS^{2+}Y$ ist $S_p(\beta)$-lokaler GAUSS-Schätzer für b genau dann, wenn $\operatorname{im} X \subset \operatorname{im} S^2$.

ii (Über den Satz von GAUSS-MARKOFF-RAO) $\forall \beta\in R_+$
$(I_p+\beta^{-2}X^+S^2X^{T+})^{-1}X^+Y$ ist $S_p(\beta)$-lokaler GAUSS-Schätzer für b genau dann, wenn $\operatorname{im} S^2M \subset \operatorname{im} M$.

iii (= Min.Var.-Min.Bias-Schätzer)
$X^+(I_n-S(MS)^+)Y$ ist GAUSS-Schätzer für b genau dann, wenn $\operatorname{im} SX \subset \operatorname{im} SM$.

Bew.: Setze $B = X^+(I_n - S(MS)^+)$.

i. $\Rightarrow$: Nach 38.ii:73 erhält man den Min.Var.-Min.Bias-Schätzer

$$((I_p+\beta^{-2}(X^TS^2+X)^+)^{-1}(S^+X)^+S^+.X)^+.(I_p+\beta^{-2}(X^TS^2+X)^+)^{-1}\underline{(S^+X)^+S^+} =$$

$$= ((I_p+\beta^{-2}(X^TS^2+X)^+)^{-1}(S^+X)^+S^+X)^+(I_p+\beta^{-2}(X^TS^2+X)^+)^{-1}(S^+X)^+S^+X(S^+X)^+S^+ =$$

$$\underset{18.i:35}{=} \operatorname{proj}(\ (S^+X)^+S^+ \ /\ \operatorname{im} (S^+X)^+S^+X.(I_p+\beta^{-2}(X^TS^2+X)^+)^{-1}\ ) =$$

$$= \operatorname{proj}(\ (S^+X)^+S^+ \ /\ \operatorname{im} (S^+X)^+S^+X\ ) = (S^+X)^+S^+X(S^+X)^+S^+ =$$

$= (S^+X)^+S^+$, mit 33.i:62 folgt $\operatorname{im} X \subset \operatorname{im} S^2$.

$\Leftarrow$: Nach 33.i:62 und 38.i:73 erhält man aus dem Min.Var.-Min.Bias-Schätzer $(S^+X)^+S^+Y$ den lokalen GAUSS-Schätzer, der angegeben ist, wegen (17.vi:34 und 17.iv:34) :

$$(\underline{S^+}X)^+S^+.\underline{S^2.S^+}(S^+X)^{+T} = (\underline{S^+S}.S^+X)^+.\underline{S^+S.S^+S}.(S^+X)^{T+} =$$

$$= (S^+S.S^+X)^+(S^+X)^{T+} = (S^+X)^+(S^+X)^{T+} = (X^TS^2+X)^+ .$$

ii. $\Rightarrow$: Mit 38.ii:73 errechnet man den Min.Var.-Min.Bias-Schätzer

$$((I_p+\beta^{-2}X^+S^2X^{+T})^{-1}X^+.X)^{-1}(I_p+\beta^{-2}X^+S^2X^{+T})X^+.XX^+ =$$

$$= \operatorname{proj}(\ X^+ \ /\ \operatorname{im} X^+X.(I_p+\beta^{-2}X^+S^2X^{+T})^{-1}\ ) = \operatorname{proj}(X^+/\operatorname{im} X^+X) =$$

$X^+XX^+ = X^+$, mit 33.ii:62 : $\operatorname{im} S^2M \subset \operatorname{im} M$.

$\Leftarrow$: Mit 33.ii:62 und 38.i:73 .

iii. I. BY ist GAUSS-Schätzer genau bei $BS = 0$; denn:
Bei $BS = 0$ ist BY nach Hauptsatz 37.i:70 GAUSS-Schätzer.
Sei nun BY GAUSS-Schätzer, etwa mit der Darstellung

$$B = (I_p+\beta^{-2}BS^2B^T)^{-1}B + Z(M-MS(MS)^+) .$$

Rechtsmultiplikation mit S gibt:

$$BS = (I_p+\beta^{-2}BS^2B^T)^{-1}BS + \underline{Z(M-MS(MS)^+)S} = (I_p+\beta^{-2}BS^2B^T)^{-1}BS .$$

Multipliziert man mit $I_p+\beta^{-2}BS^2B^T$ von links:

$$(I_p+\beta^{-2}BS^2B^T)BS = BS + \beta^{-2}BS^2B^TBS = BS \text{ , also}$$

$$BS^2B^TBS = 0 \ ,\quad SB^T.BS.SB^TBS = 0 \ ,\quad SB^TBS = 0 \ ,\quad BS = 0 .$$

II. $0 = BS = X^+(I_n-S(MS)^+)S = X^+S(I_n-(MS)^+MS)$

$\Leftrightarrow \quad (I_n-(MS)^+MS)SX^{+T} = 0 \quad \Leftrightarrow \quad (I_n-(MS)^+MS)SX = 0 \quad \Leftrightarrow$

$\Leftrightarrow \ SX = (MS)^+MS.SX \underset{18.ii:35}{=} SM(SM)^+.SX \ \Leftrightarrow \ \operatorname{im} SX \subset \operatorname{im} SM$. _/

Satz 40: Beispiele für $S_p(\beta)$-lokale GAUSS-Schätzer

i (Gleiche Varianzen, unkorreliert) $\forall \beta \in R_+$
Im GM-Modell $(Y,Xb,n,p,\sigma^2 I_n)$ ist $(I_p+(\sigma/\beta)^{-2}(X^TX)^+)^{-1}X^+Y$ der einzige $S_p(\beta)$-lokale GAUSS-Schätzer.

ii (Gleiche Mittelwerte, gleiche Varianzen, unkorreliert) $\forall \beta \in R_+$
Im GM-Modell $(Y,1_n\mu,n,1,\sigma^2 I_n)$ ist

$\dfrac{n\beta^2}{n\beta^2+\sigma^2}.\bar{Y}$ der einzige $\{\pm \beta\}$-optimale GAUSS-Schätzer für μ .

Bew.: i ist eine Anwendung von 39.ii:73 . ii folgt aus i und

$$(1+\sigma^2\beta^{-2}n^{-1})^{-1} = \frac{1}{1+\sigma^2/n\beta^2} = \frac{n\beta^2}{n\beta^2+\sigma^2} \quad \text{(vgl.29.iii:56).} \qquad \_/$$

6.3 Die wichtigsten Ergebnisse zur Mittelwert-Schätzung

Gemeinsam ist der Min.Norm-Min.Bias-, Min.Var.-Min.Bias- und lokalen GAUSS-Schätzung die wichtige

Regel zur Schätzung linearer Funktionen (28:55; 31.i:60; 37.i:70):

> Den besten Schätzer $\hat{L}Y$ für eine lineare Funktion Lb erhält man, indem man die Funktion L anwendet auf den besten Schätzer $\hat{b}Y$ von b : $\hat{L}Y = L.\hat{b}Y$.
> (Dabei kann $\hat{L}Y$ erw.-treu sein, ohne daß $\hat{b}Y$ das ist.)

Bei der Schätzung von b kann man von einem optimalen Schätzer zum andern wechseln (38:73; Seite 61):

Formeltabelle: Zusammenhänge zwischen den Mittelwertschätzern

Aus gegebenem $\hat{b}Y$ als ... erhält man für ... die Darstellung:	Min.Norm-Min.Bias-Schätzer	Min.Var.-Min.Bias-Schätzer	$S_p(\beta)$-lok. GAUSS-Schätzer
Min.Norm-Min.Bias-Schätzer	$\hat{b}Y$	$\hat{b}XX^+Y$	$(\hat{b}X)^+\hat{b}XX^+Y$
Min.Var.-Min.Bias-Schätzer	–	$\hat{b}Y$	$(\hat{b}X)^+\hat{b}Y$
$S_p(\beta)$-lokale GAUSS-Schätzer	–	$(I_p+\beta^{-2}\hat{b}S^2\hat{b}^T)^{-1}\hat{b}Y$	$\hat{b}Y$

Bei der Schätzung von b gibt es für Min.Var.-Min.Bias-Schätzer in Spezialfällen besondere Darstellungen (33:62; 35:65):

Übersichtstabelle: Eigenschaften zur Min.Var.-Min.Bias-Schätzung

Schätzer	Eigenschaft	Kriterium
$X^+(I_n-S(MS)^+)Y$	einzig	im M $\subset$ im S^2
$(S^+X)^+S^+Y$	Min.Var.-Min.Bias-Schätzer im GM-Modell (Y,Xb,n,p,S^2)	im X $\subset$ im S^2
X^+Y		im S^2M $\subset$ im M
$X^+(I_n-S(MS)^+)Y+$ $+Z(M-MS(MS)^+)Y$	Min.Var.-Min.Bias-Schätzer im GM-Modell (Y,Xb,n,p,T^2)	im T^2M $\subset$ im S^2M

* * *

Teil III

Schätzung der Streuungsmatrix in GAUSS-MARKOFF-Modellen

GAUSS-MARKOFF-Modelle werden erweitert zu linearen Modellen linearer Regression für Mittelwert und Streuung. Gesucht sind optimale Schätzfunktionen für lineare Funktionen des Streuung-Regressionsparameters.

Zu der transformierten Zufallsgröße, die aus der ursprünglichen entsteht durch eine Mittelwert-Invarianz erzeugende Transformation und durch Übergang zum KRONECKERquadrat, gehört wieder ein GAUSS-MARKOFF-Modell. In diesem "abgeleiteten" GAUSS-MARKOFF-Modell ist der Regressionsparameter für den Mittelwert derselbe wie der für die Streuung im ursprünglichen Modell (Kap. 7). Mit dieser "Streuung-Mittelwert-Korrespondenz" werden die Ergebnisse der Mittelwert-Schätzung übertragen und zur Streuung-Schätzung verwendet (Kap. 7 - 9).

Alle Schätzer werden explizit angegeben, in Spezialfällen vereinfacht und untereinander in Beziehung gesetzt. Beispiele aus der Literatur verdeutlichen die einzelnen Schätzverfahren und lassen erkennen, wie die Streuung-Mittelwert-Korrespondenz die Schätztheorie für GAUSS-MARKOFF-Modelle weiter vereinheitlicht.

Kap. 7: Minimum Norm - Minimum Bias - invariante quadratische Schätzfunktionen

Zunächst führen wir GAUSS-MARKOFF-Modelle zur Schätzung der Streuungsmatrix ein (7.1). Dann leiten wir die Streuung-Mittelwert-Korrespondenz her (7.2) und beginnen, die Ergebnisse der Mittelwert-Schätzung zu übertragen (7.3 , 7.4).

7.1 GAUSS-MARKOFF-Modelle zur Streuungsregression

Def. 11: (GM-Modell $(Y,Xb,n,p,\Sigma\, t_\varkappa D_\varkappa)$)
Ein 5-Tupel $(Y,Xb,n,p,\Sigma\, t_\varkappa D_\varkappa)$ heiße "GAUSS-MARKOFF-Modell mit unbekannter Streuung aus span $D_\varkappa$" - kurz: GM-Modell $(Y,Xb,n,p,\Sigma\, t_\varkappa D_\varkappa)$ -, falls
(Y,Xb,n,p) ein GM-Modell ist,
(GMM5) $(D_\varkappa)_{\varkappa=1,\dots,k} \in \mathrm{Sym}(n)^k$ und
(GMM6) die Streuungsmatrix von Y in span $D_\varkappa$ liegt:
$$DY \in \operatorname{span}_{\varkappa=1}^{k} D_\varkappa \quad . \qquad \_/$$

Def. 12: (GM-Modell $(Y,Xb,n,p,\Sigma\, t_\varkappa D_\varkappa, F^2)$)
Ein 6-Tupel $(Y,Xb,n,p,\Sigma\, t_\varkappa D_\varkappa, F^2)$ heiße "GAUSS-MARKOFF-Modell mit unbekannter Streuung aus span $D_\varkappa$ und bekannten vierten Momenten F^2" - kurz: GM-Modell $(Y,Xb,n,p,\Sigma\, t_\varkappa D_\varkappa, F^2)$ -, falls
$(Y,Xb,n,p,\Sigma\, t_\varkappa D_\varkappa, F^2)$ ein GM-Modell mit unbekannter Streuung aus span $D_\varkappa$ ist und
(GMM7) Y die gemischten zentralen vierten Momente F^2 hat:
$$D(Y-EY)\otimes(Y-EY) = F^2 \quad . \qquad \_/$$

(GMM6) bedeutet genauer: span $D_{\mathfrak{B}}Y$ = span $D_\varkappa$ (vgl. Seite 49). - Wegen (GMM6) existiert ein Koeffizientenvektor $t \in R^k$ mit $DY = \Sigma\, t_\varkappa D_\varkappa$, t heißt "Streuung-Regressionsparameter".

Der Fundamentaldefekt linearer Modelle läßt sich bei der Annahme eines linearen Modells für die Streuung nicht vermeiden: In span $D_\varkappa$ liegen mit p.s.d. auch negativ semidefinite Matrizen[22].

---

22 Es gibt auch unsinnigeDesignannahmen. In span $\begin{pmatrix}+1 & 0\\ 0 & -1\end{pmatrix}$ etwa gibt es keine nichttrivialen p.s.d. Matrizen.

7.2 Die Streuung-Mittelwert-Korrespondenz

Es sei ein GM-Modell $(Y,Xb,n,p,\Sigma\, t_\varkappa D_\varkappa)$ mit unbekannter Streuung aus span $D_\varkappa$ gegeben. Wir diskutieren das abgeleitete Modell und die Streuung-Mittelwert-Korrespondenz und leiten daraus die Klasse der "natürlichen" Schätzfunktionen für eine lineare Funktion Lt des Regressionsparameters t her, $L\in M(1,k)$.

Das abgeleitete Modell
- wird aus dem ursprünglichen Modell $(Y,Xb,n,p,\Sigma\, t_\varkappa D_\varkappa)$ hergeleitet,
- ist selbst ein GM-Modell und
- hat t zum Regressionsparameter für den Mittelwert.

Das so etwas möglich ist, kann nur auf den ersten Blick überraschen. Denn tatsächlich ist die Streuungsmatrix DY als Erwartungswert definiert, nämlich als Erwartungswert der Zufallsmatrix $(Y-EY)(Y-EY)^T$.

Der Anteil $XX^+Y = \mathrm{proj}(Y/\mathrm{im}\, X)$ von Y , der in das Bild von X fällt, kann ganz durch die Mittelwert-Regression erklärt werden. Es erscheint daher sinnvoll, die Schätzung der Streuung nur auf den Anteil $MY = \mathrm{proj}(Y/\mathrm{im}^\perp X)$ zu stützen (b-invarianz-erzeugende Transformation): Aus $(Y-EY)(Y-EY)^T$ erhält man die neue Zufallsmatrix $MY(MY)^T$.

Die Schätztheorie für den Mittelwert arbeitet mit vektorwertigen Zufallsgrößen. Also geht man noch über zu

$$\mathrm{vec}\, MY(MY)^T \underset{13.i:28}{=} MY\otimes MY .$$

Zu der transformierten Zufallsgröße $MY\otimes MY$ gehört wieder ein GM-Modell:

$$\begin{aligned}
E\, MY\otimes MY &\underset{M.EY=0}{=} E\, M(Y-EY)\otimes M(Y-EY) = \\
&\underset{12.iii:26}{=} E\, M\otimes M.(Y-EY)\otimes(Y-EY) = \\
&\underset{13.i:28}{=} E\, M\otimes M.\mathrm{vec}\,(Y-EY)(Y-EY)^T = \\
&= M\otimes M.\mathrm{vec}\, E\,(Y-EY)(Y-EY)^T = \\
&= M\otimes M.\mathrm{vec}\, DY = \\
&= M\otimes M.\mathrm{vec}\,\Sigma\, t_\varkappa D_\varkappa = \\
&= \underset{(n^2,n^2)}{M\otimes M}.\Sigma\, \underset{(n^2,1)}{(\mathrm{vec}\, D_\varkappa)}.\underset{(1,1)}{t_\varkappa} \;;
\end{aligned}$$

wir ordnen das Streuung-Design anders an:

$D := [\operatorname{vec} D_\varkappa]_{\varkappa=1,\ldots k} \in M(n^2,k)$

und vereinfachen damit die Schreibung:

$E\ MY \otimes MY = M \otimes M.Dt$.

Wegen der Modellannahme (GMM6): $\operatorname{span} D_{\mathfrak{B}} Y = \operatorname{span} D_\varkappa$ gilt hier: $\operatorname{span} E_{\mathfrak{B}} MY \otimes MY = \operatorname{im} M \otimes M.D$.

Das heißt: Das 4-Tupel $(MY \otimes MY,\ M \otimes M.Dt,\ n^2,k)$ ist ein GM-Modell zur Mittelwert-Regression und hat als Regressionsparameter t , den Regressionsparameter für die Streuung im ursprünglichen Modell:

Def. 13: (Abgeleitetes Modell)

Das GM-Modell $(MY \otimes MY,\ M \otimes M.Dt, n^2,k)$ heiße "Ableitung des GM-Modells $(Y,Xb,n,p,\Sigma\ t_\varkappa D_\varkappa)$", "abgeleitetes GM-Modell" .

Schätzung der Streuung ist eine Frage des Standpunktes; t kann angesehen werden als Regressionsparameter für
- die Streuung im ursprünglichen Modell, bzw. für
- den Mittelwert im abgeleiteten Modell.

Dies nennen wir die "Streuung-Mittelwert-Korrespondenz".

Ist Lt eine lineare Funktion des Regressionsparameters t : $L \in M(1,k)$, dann bilden die linearen Schätzfunktionen $\hat{L}.MY \otimes MY$, $\hat{L} \in M(1,n^2)$, für Lt im abgeleiteten Modell die Klasse der "natürlichen" Schätzfunktionen für Lt im ursprünglichen Modell, wenn man sie in der Variablen Y liest:

$\{\ Q(Y) = \hat{L}.MY \otimes MY : \hat{L} \in M(1,n^2)\ \}$. -

Wir sind hier anders vorgegangen als in der Literatur üblich. Es bleibt zu zeigen, daß dieselbe Klasse von Schätzfunktionen zugrundeliegt, die auch sonst gewählt wird.

Es sei Lt eine Linearform des Regressionsparameters t : $L \in M(1,k)$. Unter den reellwertigen Funktionen wählt man gewöhnlich diejenigen $Q(Y)$ als Schätzfunktionen für Lt aus,
- die von Y nur über MY abhängen: $Q(Y) = Q(MY)$ – solche Funktionen heißen invariant –, und
- die quadratisch sind, d.h. die von einer Bilinearform $\bar{Q}(x,y)$ abstammen: $Q(Y) = \bar{Q}(Y,Y)$.

Einerseits sind die von uns als natürlich bezeichneten Schätzfunktionen $Q(Y) := \hat{L}.MY \otimes MY$ invariant: $Q(MY) = \hat{L}.M^2Y \otimes M^2Y = \hat{L}.MY \otimes MY = Q(Y)$, und quadratisch: $\bar{Q}(x,y) := \hat{L}.Mx \otimes My$ ist bilinear mit $Q(Y) = \bar{Q}(Y,Y)$.

Andererseits ist jede invariante quadratische Form $Q(Y)$ eine in unserem Sinne natürliche Schätzfunktion[23].
$Q(Y)$ stammt per Definition von einer Bilinearform $\bar{Q}(x,y)$ ab; da nun das KRONECKERprodukt ein Tensorprodukt ist (11:25), faktorisiert $\bar{Q}$: $\bar{Q}(x,y) = \bar{Q}^*(x \otimes y)$, wobei $\bar{Q}^*$ eine Linearform auf dem R^{n^2} ist: $\hat{L} := \bar{Q}^* \in M(1,n^2)$. Somit gilt $\bar{Q}(x,y) = \hat{L}.x \otimes y$ und per Definition $Q(Y) = \hat{L}.Y \otimes Y$. Zudem ist $Q(Y)$ als invariant vorausgesetzt: $Q(Y) = Q(MY) =$ $= \hat{L}.MY \otimes MY$. Diese Darstellung zeigt, daß $Q(Y)$ eine in unserem Sinne natürliche Schätzfunktion ist. –

In dieser Arbeit schätzen wir nicht nur Linearformen, sondern allgemeiner lineare Funktionen $Lt : L \in M(1,k)$. In offenbar konsistenter Sprechweise nennen wir eine in unserem Sinne natürliche Schätzfunktion $Q(Y) = \hat{L}.MY \otimes MY$ eine "<u>invariante quadratische Schätzfunktion für Lt</u>", oder kürzer: einen "(Streuungs-)<u>Schätzer</u>".

<u>7.3 Erwartungstreue und Minimum Bias - inv. quadr. Schätzfunktionen</u>

Gegeben sei ein GM-Modell $(Y,Xb,n,p,\Sigma\, t_\varkappa D_\varkappa)$ mit unbekannter Streuung aus span $D_\varkappa$, eine lineare Funktion Lt des Regressionsparameters t : $L \in M(1,k)$, und eine invariante quadratische Schätzfunktion $Q(Y) = \hat{L}.MY \otimes MY$ für Lt : $\hat{L} \in M(1,n^2)$. Wir übertragen die Ergebnisse der Abschnitte 4.2 und 4.3 .

Der Regressionsparameter b für den Mittelwert spielt jetzt keine Rolle mehr, da man sich auf invariante Funktionen beschränkt: $Q(Y) = Q(MY) = Q(MY-M.EY) = Q(\ M(Y-EY)\ ) = Q(Y-EY)$. Ohne Beschränkung der Allgemeinheit kann im folgenden Y als zentriert angenommen werden: $EY = 0$.

---

23 Nur an dieser Stelle wird die Idee des Tensorenkalküls - die Korrespondenz zwischen bilinearen und linearen Funktionen (3:19) - voll eingesetzt.
Die wiederholt benutzten Ergebnisse aus den Abschnitten 1.3 (Das KRONECKERprodukt) und 1.4 (Die Funktion vec) erscheinen als sinnvoll und naheliegend, auch ohne daß sie in den Tensorenkalkül eingebettet werden.

$Q(Y) = \hat{L}.MY \otimes MY$ ist eine lineare Schätzfunktion für Lt im abgeleiteten Modell. Es ist daher wohldefiniert, wann

- $Q(Y)$ erw.-treu ist,
- Lt erw.-treu schätzbar ist,
- $Q(Y)$ ein Min.Bias-Schätzer für Lt ist und
- $Q(Y)$ ein Min.Norm-Min.Bias-Schätzer für Lt ist.

<u>Satz 41: Erw.-treue Schätzbarkeit</u>

i (Kriterium für Erw.-treue)
$Q(Y) = \hat{L}.MY \otimes MY$ ist genau dann erw.-treu,
wenn $\hat{L}.M \otimes M.D = L$.

ii (Erw.-treue Schätzbarkeit)
Lt ist genau dann erw.-treu schätzbar,
wenn $L.(M \otimes M.D)^{+}D = L$.

iii (Spezialfall $L = I_k$)
t ist genau dann erw.-treu schätzbar, wenn $\mathrm{rk}\, M \otimes M.D = k$.

<u>Bew.:</u> Anwendung des Satzes 26:53 auf das abgeleitete Modell unter Beachtung von $(M \otimes M.D)^{+}.M \otimes M = (M \otimes M.D)^{+}$ nach 12.v:26; 17.v:34 . _/

Sei t erw.-treu schätzbar; dann gilt ($\mathrm{rk}\, X = s$)

$$k = \mathrm{rk}\, M \otimes M.D \leq \mathrm{rk}\, M \otimes M \underset{12.ii:26}{=} \mathrm{rk}\, M \,.\, \mathrm{rk}\, M = (n-s)^2 ,$$

$\sqrt{k} \leq n-s$, $n \geq s + \sqrt{k}$: Notwendig zur erw.-treuen Schätzbarkeit von t ist also, daß mindestens $k^{\frac{1}{2}} + \mathrm{rk}\, X$ Beobachtungen vorliegen.

<u>Satz 42: Alle Min.Bias-Schätzer für Lt</u>

Die Menge aller Min.Bias-Schätzer für Lt ist
$\{ L.(M \otimes M.D)^{+}.Y \otimes Y + ZN.Y \otimes Y : Z \in M(1,n^2) \}$,
$N := M \otimes M - M \otimes M.D(M \otimes M.D)^{+}$.

<u>Bew.:</u> Nach Satz 27:54 ist die Klasse aller Min.Bias-Schätzer für Lt im abgeleiteten Modell

$$\{ L.(M \otimes M.D)^{+}.MY \otimes MY + Z(I_{n^2} - M \otimes M.D(M \otimes M.D)^{+}).MY \otimes MY : Z \in M(1,n^2) \} ,$$

wegen $MY \otimes MY = M \otimes M.Y \otimes Y$ (12.iii:26) und $(M \otimes M.D)^{+}.M \otimes M = (M \otimes M.D)^{+}$ (12.v:26; 17.v:34) ist dies dieselbe Menge wie in der Behauptung. _/

7.4 Minimum Norm - Minimum Bias - inv. quadr. Schätzfunktionen

Unter denselben Voraussetzungen wie in 7.3 werden die Ergebnisses des Abschnitts 4.4 übertragen.

Es ist wohldefiniert, wann $Q(Y) = \hat{L}.MY\otimes MY$ ein Min.Norm-Min.Bias-Schätzer für Lt ist; dabei betrachtet man $Q(Y)$ vom abgeleiteten Modell her. Doch auch aus der Sicht des ursprünglichen Modells ist es sinnvoll, die Norm unter allen Min.Bias-Schätzern zu minimieren:

Bezeichne F^2 die symmetrische (c^2,c^2)-Matrix der gemischten zentralen vierten Momente:
$F^2 = D\,(Y-EY)\otimes(Y-EY) \underset{EY=0}{=} D\,Y\otimes Y$.
Das Ziel, die Varianz von $Q(Y)$ zu minimieren, führt von
$\operatorname{tr} D\,Q(Y) = \operatorname{tr} D\,\hat{L}.MY\otimes MY = \operatorname{tr} \hat{L}.M\otimes M.\,D\,Y\otimes Y\,.\,M\otimes M.\hat{L}^T =$
$= \operatorname{tr} \hat{L}.M\otimes M.F^2.M\otimes M.\hat{L}^T = \|\hat{L}.M\otimes M.F\|^2$
über die Standardabschätzung 13.iv:28
$\|\hat{L}.M\otimes M.F\| \leqq \|\hat{L}.M\otimes M\|.\|F\|$
zur Minimierung der Norm
$\|\hat{L}.M\otimes M\|$.
Für die in 42:81 angegebenen Min.Bias-Schätzer gilt aber
$\hat{L}.M\otimes M = \hat{L}$, also ist zu minimieren $\|\hat{L}\|$.

Anwendung des Satz 28:55 ergibt sofort den

Satz 43: Hauptsatz über Min.Norm-Min.Bias-Schätzer für Lt

In einem GM-Modell $(Y,Xb,n,p,\Sigma\, t_\varkappa D_\varkappa)$ sei Lt eine lineare Funktion des Regressionsparameters t : $L\in M(1,k)$;
setze $D = [\operatorname{vec} D_\varkappa]_{\varkappa=1,\dots,k}$.
Dann ist $L.(M\otimes M.D)^+.Y\otimes Y$ der einzige Min.Norm-Min.Bias-Schätzer für Lt . _/

In Satz 44:83 berechnen wir den Min.Norm-Min.Bias-Schätzer, wenn die Streuungsmatrix DY ein Vielfaches einer vorgegebenen Streuungsmatrix S^2 ist: $DY \in \operatorname{span} S^2$; dieser Fall interessiert besonders deshalb, weil Min.Var.-Min.Bias-Schätzer für den Mittelwert invariant sind gegenüber solchen Vielfachen (Seite 61). Satz 45:84 gibt RAOs 1970-MINQUE-Theorie als Beispiel unserer Min.Norm-Min.Bias-Schätzung.

Satz 44: Beispiele für Min.Norm-Min.Bias-Schätzer

i ($DY \in \operatorname{span} S^2$)

Im GM-Modell $(Y,Xb,n,p,\sigma^2 S^2)$ ist $\|MS^2M\|^{2+}.Y^TMS^2MY$ der Min.Norm-Min.Bias-Schätzer für σ^2, er ist p.s.d.; er ist genau dann erw.-treu, wenn $S^2M \neq 0$.

ii (Gleiche Varianzen, unkorreliert)

Im GM-Modell $(Y,Xb,n,p,\sigma^2 I_n)$, $n > \operatorname{rk} X = s$ ist $(n-s)^{-1}Y^TMY$ der Min.Norm-Min.Bias-Schätzer für σ^2, er ist p.s.d. und erw.-treu.

iii (Gleiche Mittelwerte, gleiche Varianzen, unkorreliert)

Im GM-Modell $(Y,1_n\mu,n,1,\sigma^2 I_n)$, $n > 1$ ist die Stichprobenstreuung $(n-1)^{-1}.\Sigma\,(Y_\nu-\bar{Y})^2$ der Min.Norm-Min.Bias-Schätzer für σ^2, er ist p.s.d. und erw.-treu.

Bew.: Es ist $D = \operatorname{vec} S^2$.

i. I. Für den Min.Norm-Min.Bias-Schätzer für σ^2 (43:82) gilt:

$$
\begin{aligned}
(M\otimes M.D)^+.Y\otimes Y &= (M\otimes M.\operatorname{vec} S^2)^+.Y\otimes Y = \\
&= (\operatorname{vec} MS^2M)^+.Y\otimes Y = \\
&\overset{13.ii:28}{=} \|MS^2M\|^{2+}.\operatorname{vec}^T MS^2M.Y\otimes Y = \\
&\overset{16.i:34}{=} \|MS^2M\|^{2+}.\operatorname{vec}^T MS^2M.\operatorname{vec} YY^T = \\
&\overset{13.i:28}{=} \|MS^2M\|^{2+}.\operatorname{tr} MS^2MYY^T = \\
&\overset{13.iii:28}{\underset{kom.}{=}} \|MS^2M\|^{2+}.Y^TMS^2MY .
\end{aligned}
$$

Diese Schlußweise ist typisch für alle ähnlichen Fälle.

II. σ^2 ist erw.-treu schätzbar genau bei $\operatorname{rk} M\otimes M.D = 1$ (41.iii:81). Dabei:

$\operatorname{rk} M\otimes M.D = \operatorname{rk} M\otimes M.\operatorname{vec} S^2 = \operatorname{rk} \operatorname{vec} MS^2M = 1 \Leftrightarrow \operatorname{vec} MS^2M \neq 0 \Leftrightarrow$
$\Leftrightarrow MS^2M \neq 0 \Leftrightarrow SM \neq 0 \Leftrightarrow S^2M \neq 0$.

ii. Aus i und $M \neq 0 \Leftrightarrow 0 \neq \operatorname{rk} M = n-s \Leftrightarrow n > s$.

iii. Es ist $X = 1_n$, $s = 1$, $MY = Y-X.X^+Y = Y - \bar{Y}.1_n$.

Daher: $Y^TMY = (MY)^TMY = (Y-\bar{Y}.1_n)^T(Y-\bar{Y}.1_n) = \Sigma\,(Y_\nu-\bar{Y})^2$.

(Vgl. 29.iii:56; 40.ii:74). _/

RAO (1970, 1971a, 1972) entwickelte als erster eine Minimum Norm - Theorie zur Schätzung von Streuungskomponenten: MInimum Norm Quadratic Unbiased Estimation = MINQUE. Die Ergebnisse aus seiner ersten Arbeit (1970) werden von der gerade diskutierten Min.Norm-Min.Bias-Schätztheorie verallgemeinert:

Satz 45: RAOs 1970-MINQUE-Formel

Gegeben sei ein GM-Modell $(Y,Xb,n,p,\Sigma\, t_\nu E_{\nu\nu}^{(n,n)})$:
Y habe also unkorrelierte Komponenten mit ungleichen, unbekannten Varianzen: setze $M = I_n - XX^+$, $M_2 := (M_{\nu\mu}^{\ 2})_{(n,n)} \in \mathrm{Sym}(n)$; Lt sei eine Linearform : $L \in M(1,n)$. Dann gilt:

i (Erw.-treue Schätzbarkeit)
(RAO 1970:166) t ist genau dann erw.-treu schätzbar, wenn M_2 regulär ist.
(RAO 1970:168) Lt ist genau dann erw.-treu schätzbar, wenn $M_2x = L^T$ lösbar ist.

ii (Schätzer) (RAO 1970:166f)
Sei Lt erw.-treu schätzbar und λ eine Lösung von $M_2x = L^T$. Dann ist $Y^TM.\mathrm{Diag}\,\lambda.MY$ erw.-treuer Min.Norm-Min.Bias-Schätzer (MINQUE) für Lt .

ii* (Verallgemeinerung)
Sei λ eine Lösung von $\|M_2x - L^T\| = \inf$. Dann ist $Y^TM.\mathrm{Diag}\,\lambda.MY$ der Min.Norm-Min.Bias-Schätzer für Lt .

Bew.: Es ist $k = n$, $D = [\mathrm{vec}\, E_{\nu\nu}]_{\nu=1,\dots,n}$.

I. $D^T.M\otimes M.D = M_2$; denn:

$$D^T.M\otimes M.D = [\mathrm{vec}^T E_{\nu\nu}]_{\nu=1,\dots,n}.M\otimes M.[\mathrm{vec}\, E_{\mu\mu}]_{\mu=1,\dots,n} =$$
$$= (M_{\nu\mu}.M_{\nu\mu})_{(n,n)} = M_2 .$$

II. $(M\otimes M.D)^+.D = M_2^{\ +}M_2$; denn:

$$M_2^{\ +}M_2 = (D^T.M\otimes M.D)^+D^T.M\otimes M.D =$$
$$= \underline{(D^T.M\otimes M.M\otimes M.D)^+D^T.M\otimes M}.D \underset{15.iv:33}{=} (M\otimes M.D)^+.D .$$

III. Lt ist erw.-treu schätzbar genau bei $L.(M\otimes M.D)^+D = L$ nach 41.ii:81 . Es gilt:

$$L.(M\otimes M.D)^+D = L \underset{II}{\Leftrightarrow} L.M_2^{\ +}M_2 = L \Leftrightarrow M_2M_2^{\ +}L^T = L^T \underset{22:41}{\Leftrightarrow} M_2x = L^T \text{ lösbar} .$$

IV. t ist erw.-treu schätzbar genau, wenn jede Linearform erw.-treu schätzbar ist, d.h. nach dem gerade Gezeigten: M_2 ist regulär.- Oder direkt mit 41.iii:81 : t ist erw.-treu schätzbar genau bei $n = \mathrm{rk}\, M\otimes M.D \underset{\mathrm{GRAM}}{=} \mathrm{rk}\, D^T.M\otimes M.D \underset{I}{=} \mathrm{rk}\, M_2$, d.h. M_2 regulär.

Damit ist i gezeigt. Statt ii beweisen wir die allgemeinere Aussage ii*; es wird die typische Schlußweise von 44.i(I):83 verwendet:

V. $\forall a \in R^n$:

$$
\begin{aligned}
Y^T M.\mathrm{Diag}\ a.MY &= \mathrm{tr}\ Y^T M.\mathrm{Diag}\ a.MY = \\
&\underset{\text{kom.}}{=} \mathrm{tr}\ M.\mathrm{Diag}\ a.M.YY^T = \\
&\underset{13.iii:28}{=} \mathrm{vec}^T M.\mathrm{Diag}\ a.M\ .\ \mathrm{vec}\ YY^T = \\
&\underset{13.i:28}{=} \mathrm{vec}^T M.\mathrm{Diag}\ a.M\ .\ Y \otimes Y = \\
&\underset{13.ii:28}{=} (M \otimes M.\mathrm{vec}\ \mathrm{Diag}\ a)^T.Y \otimes Y = \\
&= (M \otimes M.\mathrm{vec}\ \Sigma\ a_\nu E_{\nu\nu})^T.Y \otimes Y = \\
&= (M \otimes M.\Sigma\ (\mathrm{vec}\ E_{\nu\nu}).a_\nu)^T.Y \otimes Y = \\
&= (M \otimes M.Da)^T.Y \otimes Y = \\
&= a^T.D^T.M \otimes M.Y \otimes Y\ .
\end{aligned}
$$

VI. $\forall\ \lambda \in \mathrm{Lsg}(\|M_2 x - L^T\| = \inf) \underset{24:43}{=} \{\ M_2{}^+ L^T + (I_n - M_2{}^+ M_2) z\ :\ z \in R^n\ \}$:

$$
\begin{aligned}
Y^T M.\mathrm{Diag}\ \lambda.MY &= \lambda^T.D^T.M \otimes M.Y \otimes Y = \\
&= LM_2{}^+ D^T.M \otimes M.Y \otimes Y + z^T \underline{(I_n - M_2{}^+ M_2) D^T.M \otimes M}.Y \otimes Y = \\
&\underset{I}{=} L.(D^T.M \otimes M.M \otimes M.D)^+ D^T.M \otimes M.Y \otimes Y + \\
&+ z^T(\ M \otimes M.D(I_n - (M \otimes M.D)^+.M \otimes M.D)\ )^T.Y \otimes Y = \\
&\overset{II}{\underset{15.iv:33}{=}} L.(M \otimes M.D)^+.Y \otimes Y + 0\ . \qquad \lrcorner
\end{aligned}
$$

Es wird sich herausstellen, daß die allgemeinere Formulierung der MINQUE-Theorie (RAO 1971a, 1972) <u>formal</u> als Min.Var.-Theorie behandelt werden kann: Zwar werden keine Annahmen über die vierten Momente gemacht und deshalb <u>ist</u> die MINQUE-Theorie <u>keine</u> Min.Var.-Theorie; trotzdem ermöglicht es die Aufgabenstellung, wie in der Min.Var.-Theorie zu rechnen. Wir ordnen die allgemeinere MINQUE-Theorie daher der Min.Var.-Min.Bias-Schätztheorie zu:

Kap. 8: Minimum Varianz - Minimum Bias - invariante quadratische Schätzfunktionen

Min.Var.-Min.Bias-Schätztheorie für die Streuung ist Min.Var.-Min.Bias-Schätztheorie für den Mittelwert im abgeleiteten Modell: die Ergebnisse des Kap. 5 werden übertragen (8.3). Vorher wird RAOs MINQUE-Theorie erörtert; dieses Verfahren wird nicht als Minimum Varianz - Schätzung begründet, kann aber formal so behandelt werden.

8.1 Minimum RAOnorm - Minimum Bias - inv. quadr. Schätzfunktionen

Zunächst wird das "MINQUE-Modell" entwickelt, auf das die Überlegungen aufbauen; daraus leiten wir eine neue Rechtfertigung der MINQUE-Theorie her, die besser in unser Konzept paßt, RAOs Begründung wird in Abschnitt 8.2 dargestellt. Dann deuten wir die MINQUE-Schätzung als Min.Var.-Min.Bias-Schätzung in einem "erweiterten Modell" mit passend gewählten vierten Momenten; das sich aus der Streuung-Mittelwert-Korrespondenz ergebende abgeleitete Modell wird zu einem GM-Modell mit bekannter Streuung: damit ist die MINQUE-Schätzung reduziert auf die Min.Var.-Min.Bias-Schätzung eines Mittelwerts. Schließlich geben wir die MINQUE-Schätzer explizit an.

Unsere Lösungen - wir nennen sie Minimum RAOnorm-Minimum Bias-Schätzer - gelten allgemeiner als RAOs MINQUE-Schätzer:
- Auf Rangannahmen wird verzichtet.
- Statt erw.-treuer werden Min.Bias-Schätzer gesucht.
- Es werden nicht nur Linearformen, sondern gleich lineare Funktionen geschätzt.

RAOs Modell (1971a:266; 1972:112) heiße "MINQUE-Modell"; es unterscheidet sich von den bisherigen dadurch, daß auch für die Residuen Y-EY ein Design angenommen wird, dieses bezeichnen wir als "Fehlerdesign".

Im MINQUE-Modell setzt sich der Beobachtungsvektor Y zusammen aus einem systematischen Anteil Xb und einem Störanteil $\Sigma U_\varkappa \xi_\varkappa$: $Y = Xb + \Sigma U_\varkappa \xi_\varkappa$. Dabei sind die $\xi_\varkappa$, $\varkappa = 1,\dots,k$, unkorrelierte Zufallsvektoren mit Werten in $R^{c_\varkappa}$, sie stellen die zufälligen Fehler dar. Es gelte die Verteilungsannahme

$E\,\xi_\varkappa = 0$, $D\,\xi_\varkappa = \sigma_\varkappa^2 I_{c_\varkappa}$, $\sigma_\varkappa^2$ unbekannt.

Die Fehlervariablen $\xi_\varkappa$ wirken über ein Fehlerdesign $U_\varkappa$:

$U_\varkappa \in M(n,c_\varkappa)$, $Y = Xb + \Sigma U_\varkappa \xi_\varkappa$.

Zur Abkürzung der Schreibung wird gesetzt:

$c := \Sigma c_\varkappa$; $U := [U_\varkappa]_{\varkappa=1,\dots,k} \in M(n,c)$; $\xi := \begin{bmatrix} \xi_1 \\ \vdots \\ \xi_k \end{bmatrix} \in R^c$;

$V_\varkappa^2 := U_\varkappa U_\varkappa^T \in \mathrm{Sym}(n)$, p.s.d. .

Damit ergibt sich:

$Y = Xb + U\xi$,

$EY = Xb + 0 = Xb$,

$DY = D\,U\xi = D\,\Sigma U_\varkappa \xi_\varkappa = \Sigma U_\varkappa . D\,\xi_\varkappa . U_\varkappa^T = \Sigma\,\sigma_\varkappa^2 . U_\varkappa U_\varkappa^T = \Sigma\,\sigma_\varkappa^2 V_\varkappa^2$;

ein MINQUE-Modell ist also ein GM-Modell

$(Y, Xb, n, p, \Sigma\, t_\varkappa V_\varkappa^2)$, $Y = Xb + U\xi$.

Wegen $E\,\xi = 0$ hat die Annahme eines Fehlerdesigns keine Auswirkungen auf den Mittelwert von Y : die Bedingungen für Invarianz und Erwartungstreue bleiben dieselben.

Aber jetzt liegt mehr Information vor im Sinne der Plausibilitätsüberlegungen zur Minimierung der Norm (Seite 82). Sei nämlich die (c^2,c^2)-Matrix G^2 der gemischten zentralen vierten Momente des Fehlervektors ξ gegeben:

$G^2 = D\,\xi\otimes\xi$.

Das Ziel, die Varianz von $Q(Y)$ zu minimieren, führt von

$$\begin{aligned} \mathrm{tr}\,D\,Q(Y) &= \mathrm{tr}\,D\,\hat{L}.M\otimes M.Y\otimes Y = \mathrm{tr}\,D\,\hat{L}.M\otimes M.U\otimes U.\xi\otimes\xi = \\ &= \mathrm{tr}\,\hat{L}.M\otimes M.U\otimes U.G^2.U^T\otimes U^T.M\otimes M.\hat{L}^T = \\ &= \|\,\hat{L}.M\otimes M.U\otimes U.G\,\|^2 \end{aligned}$$

über die Standardabschätzung 13.iv:28

$\|\,\hat{L}.M\otimes M.U\otimes U.G\,\| \leqq \|\,\hat{L}.M\otimes M.U\otimes U\,\| . \|G\|$

zur Minimierung von – setze $V^2 := UU^T = \Sigma U_\varkappa U_\varkappa^T = \Sigma V_\varkappa^2$ – :

$$\begin{aligned} \|\,\hat{L}.M\otimes M.U\otimes U\,\|^2 &= \mathrm{tr}\,\hat{L}.M\otimes M.U\otimes U.U^T\otimes U^T.M\otimes M.\hat{L}^T = \\ &= \mathrm{tr}\,\hat{L}.M\otimes M.V^2\otimes V^2.M\otimes M.\hat{L}^T = \\ &= \|\,\hat{L}.M\otimes M.V\otimes V\,\|^2 . \end{aligned}$$

<u>Def. 14:</u> (Min.RAOnorm-Min.Bias-Schätzer für Lt)

In einem GM-Modell $(Y,Xb,n,p,\Sigma\, t_\varkappa V_\varkappa{}^2)$ mit p.s.d. Streuung-Design: $V_\varkappa{}^2$ p.s.d., $\varkappa=1,\dots,k$ heiße $Q(Y) = \hat{L}.MY\otimes MY$ "<u>Minimum RAOnorm - Minimum Bias - invariante quadratische Schätzfunktion für Lt</u>", "<u>Min.RAOnorm-Min.Bias-Schätzer für Lt</u>", falls $Q(Y)$ Min.Bias-Schätzer für Lt ist und

$\| \hat{L}.M\otimes M.V\otimes V \| = \inf \{ \| \check{L}.M\otimes M.V\otimes V \| :$

$: \check{L}.MY\otimes MY$ Min.Bias-Schätzer für $Lt \}$;

$L\in M(1,k)$, $\hat{L}\in M(1,n^2)$, $V^2 = \Sigma\, V_\varkappa{}^2$. _/

Betrachten wir das "<u>erweiterte Modell</u>": das GM-Modell $(Y,Xb,n,p,\Sigma\, t_\varkappa V_\varkappa{}^2, V^2\otimes V^2)$ mit unbekannter Streuung aus span $V_\varkappa{}^2$ und bekannten vierten Momenten $V^2\otimes V^2$, so führt die Suche nach Min.Var.-Min.Bias-Schätzern zu derselben Minimierungsaufgabe

$\mathrm{tr}\; D\, \hat{L}.MY\otimes MY = \mathrm{tr}\; \hat{L}.M\otimes M.V^2\otimes V^2.M\otimes M.\hat{L}^T = \| \hat{L}.M\otimes M.V\otimes V \|^2 = \inf$.

Min.RAOnorm-Min.Bias-Schätzer im GM-Modell $(Y,Xb,n,p,\Sigma\, t_\varkappa V_\varkappa{}^2)$ fallen also zusammen mit den Min.Var.-Min.Bias-Schätzern im erweiterten Modell, deshalb werden sie hier im Kapitel über Minimum Varianz - Schätzung diskutiert.

Zu dem erweiterten Modell gehört das abgeleitete Modell $(MY\otimes MY,\ M\otimes M.Dt,\ n^2,k,\ MV^2M\otimes MV^2M)$, $D = [\,\mathrm{vec}\ V_\varkappa{}^2]_{\varkappa=1,\dots,k}$; denn $MY\otimes MY$ hat die Streuung

$D\ MY\otimes MY = M\otimes M.V^2\otimes V^2.M\otimes M = MV^2M\otimes MV^2M$.

Übertragung von Hauptsatz 31:60 gibt den

<u>Satz 46: Hauptsatz über Min.RAOnorm-Min.Bias-Schätzer für Lt</u>

In einem GM-Modell $(Y,Xb,n,p,\Sigma\, t_\varkappa V_\varkappa{}^2)$ sei Lt eine lineare Funktion des Regressionsparameters t : $L\in M(1,k)$; setze $D = [\,\mathrm{vec}\ V_\varkappa{}^2]_{\varkappa=1,\dots,k}$, $V^2 = \Sigma\, V_\varkappa{}^2$, $M = I_n - XX^+$, $N = M\otimes M - M\otimes M.D(M\otimes M.D)^+$. Dann gilt:

i (Alle Min.RAOnorm-Min.Bias-Schätzer für Lt)

Die Menge aller Min.RAOnorm-Min.Bias-Schätzer für Lt ist

$\{ L.(M\otimes M.D)^+(I_{n^2} - V\otimes V.(N.V\otimes V)^+).Y\otimes Y +$

$+ Z(N - N.V\otimes V.(N.V\otimes V)^+).Y\otimes Y : Z\in M(1,n^2) \}$.

ii (Einzigkeit)

$L.(M\otimes M.D)^+(I_{n^2} - V\otimes V.(N.V\otimes V)^+).Y\otimes Y$ ist der einzige Min.RAOnorm-Min.Bias-Schätzer für Lt genau dann, wenn $\mathrm{im}\ N \subset \mathrm{im}\ MV^2\otimes MV^2$.

<u>Bew.:</u> Zur Anwendung von Hauptsatz 31:60 auf das abgeleitete Modell $(MY\otimes MY,\ M\otimes M.Dt,n^2,k,MV^2M\otimes MV^2M)$ erhalten wir das Lexikon:

$$\begin{array}{l|l}
X & M\otimes M.D \\
M & I_{n^2} - M\otimes M.D(M\otimes M.D)^+ \\
S^2 & MV^2M\otimes MV^2M \\
MS^2M & N.V^2\otimes V^2.N \\
S(MS)^+ & M\otimes M.V\otimes V.(N.V\otimes V)^+ \ .
\end{array}$$

Zur vorletzten Zeile (13.ii:28; 17.v:34):

$$(I_{n^2}-M\otimes M.D.(M\otimes M.D)^+).M\otimes M.V^2\otimes V^2.M\otimes M.(I_{n^2}-M\otimes M.D(M\otimes M.D)^+) =$$
$$= N.V^2\otimes V^2.N \quad .$$

Zur letzten Zeile (15.iv:33; 13.ii:28):

$$S(MS)^+ = S^2(MS^2M)^+ \mid \ M\otimes M.V^2\otimes V^2.M\otimes M.(N.V^2\otimes V^2.N)^+ =$$
$$= M\otimes M.V\otimes V.\underline{V\otimes V.N.(N.V\otimes V.V\otimes V.N)^+} = M\otimes M.V\otimes V.(N.V\otimes V)^+ \ .$$

Damit:

$$L.(M\otimes M.D)^+(I_{n^2}-M\otimes M.V\otimes V.(N.V\otimes V)^+).M\otimes M.Y\otimes Y +$$
$$+ Z.(\ (I_{n^2}-M\otimes M.D(M\otimes M.D)^+) - (I_{n^2}-M\otimes M.D(M\otimes M.D)^+).M\otimes M.V\otimes V.$$
$$.(N.V\otimes V)^+\ ).M\otimes M.Y\otimes Y \ =$$
$$= L.(M\otimes M.D)^+.M\otimes M.(I_{n^2}-V\otimes V.(N.V\otimes V)^+).Y\otimes Y +$$
$$+ Z.(N - N.V\otimes V.(N.V\otimes V)^+).Y\otimes Y \quad .$$

ii. $\mathrm{im}\ MV^2M\otimes MV^2M = \mathrm{im}\ MV\otimes MV.(MV\otimes MV)^T = \mathrm{im}\ MV\otimes MV =$
$= M\otimes M(\mathrm{im}\ V\otimes V) = M\otimes M(\mathrm{im}\ V^2\otimes V^2) = \mathrm{im}\ MV^2\otimes MV^2$, nach 1:15 . _/

Mit diesem Hauptsatz haben wir rein formal auch schon die Frage nach Min.Var.-Min.Bias-Schätzern für Lt beantwortet; denn gerechnet wurde ja in einem GM-Modell mit bekannten vierten Momenten $F^2 = V^2\otimes V^2$.

Bevor wir RAOs Formeln herleiten, geben wir spezielle Darstellungen:

i. Wann ist der Min.Norm-Min.Bias-Schätzer für t ein Min.RAOnorm-Min.Bias-Schätzer?

ii. Wann ist der Min.Norm-Min.Bias-Schätzer für t im $(\sqrt{\ } MV^2M\otimes MV^2M\ )^+$-transformierten abgeleiteten Modell ein Min.RAOnorm-Min.Bias-Schätzer im ursprünglichen Modell?

iii. Wann ist der Min.Norm-Min.Bias-Schätzer für t im abgeleiteten V^+-transformierten Modell ein Min.RAOnorm-Min.Bias-Schätzer im ursprünglichen Modell?

<u>Satz 47: Spezielle Darstellungen für Min.RAOnorm-Min.Bias-Schätzer</u>

Unter denselben Voraussetzungen wie in Satz 46:88 gilt:

i (= Min.Norm-Min.Bias-Schätzer)

$(M\otimes M.D)^{+}.Y\otimes Y$ ist genau dann Min.RAOnorm-Min.Bias-Schätzer für t, wenn $\mathrm{im}\, MV^2M\otimes MV^2M.D \subset \mathrm{im}\, M\otimes M.D$.

ii (Über das transformierte abgeleitete Modell)

$(\,D^T(MV^2M\otimes MV^2M)^{+}D\,)^{+}.D^T(MV^2M\otimes MV^2M)^{+}.Y\otimes Y$ ist genau dann Min.RAOnorm-Min.Bias-Schätzer für t,

wenn $\mathrm{im}\, M\otimes M.D \subset \mathrm{im}\, MV^2\otimes MV^2$.

iii (Über das abgeleitete transformierte Modell)

Setze $M_V := I_n - V^{+}X(V^{+}X)^{+}$. Sei V^2 p.d. .

Dann ist $(M_VV^{-1}\otimes M_VV^{-1}.D)^{+}.V^{-1}\otimes V^{-1}.Y\otimes Y$ Min.RAOnorm-Min.Bias-Schätzer für t .

<u>Bew.:</u> i. Die Bedingung $\mathrm{im}\, S^2X \subset \mathrm{im}\, X$ aus 33.ii:62 entspricht $\mathrm{im}\, MV^2M\otimes MV^2M.M\otimes M.D = \mathrm{im}\, MV^2M\otimes MV^2M.D \subset \mathrm{im}\, M\otimes M.D$.

ii.Die Bedingung $\mathrm{im}\, X \subset \mathrm{im}\, S^2$ aus 33.i:62 entspricht (1.ii:15)

$\mathrm{im}\, M\otimes M.D \subset \mathrm{im}\, MV^2M\otimes MV^2M = \mathrm{im}\, MV\otimes MV.(MV\otimes MV)^T = \mathrm{im}\, MV\otimes MV =$

$= M\otimes M\ (\mathrm{im}\, V\otimes V) = M\otimes M\ (\mathrm{im}\, V^2\otimes V^2) = \mathrm{im}\, MV^2\otimes MV^2$.

iii. Das V^{-1}-transformierte ursprüngliche Modell ist

$(V^{-1}Y, V^{-1}Xb, n, p, \Sigma\, t_\varkappa V^{-1}V_\varkappa{}^2V^{-1})$,

dazu gehört das abgeleitete Modell

$(M_VV^{-1}Y\otimes M_VV^{-1}Y,\ M_VV^{-1}\otimes M_VV^{-1}.Dt,\ n^2, k)$.

Der Min.Norm-Min.Bias-Schätzer für t im abgeleiteten V^{+}-transformierten Modell ist

$(M_VV^{-1}\otimes M_VV^{-1}.D)^{+}.M_V\otimes M_V.V^{-1}\otimes V^{-1}.Y\otimes Y$,

nach 17.v:34 ist das der angegebene Schätzer.

Da V^2 als regulär vorausgesetzt ist, gilt die Bedingung $\mathrm{im}\, M\otimes M.D \subset \mathrm{im}\, MV^2\otimes MV^2 = \mathrm{im}\, M\otimes M$. Wir zeigen, daß sich die Darstellung aus ii zu der in iii angegebenen vereinfacht.

Zur Abkürzung setzen wir

$R := (MV^2M)^{+}$;

nach ii und 16.v:34 ist also

$(D^T.R\otimes R.D)^{+}D^T.R\otimes R.Y\otimes Y$

Min.RAOnorm-Min.Bias-Schätzer für t .

Es soll R berechnet werden.

I. $(MV)^{+} = M_VV^{-1}$; denn:

(PK1) $MVM_VV^{-1} = MV(I_n - V^{-1}X(V^{-1}X)^{+})V^{-1} = MVV^{-1}(I_n - X(V^{-1}X)^{+}V^{-1}) =$
$= M - 0 \in \mathrm{Sym}(n)$.

(PK2) $M_VV^{-1}MV = (I_n - V^{-1}X(V^{-1}X)^{+})V^{-1}(I_n - XX^{+})V =$
$= (I_n - V^{-1}X(V^{-1}X)^{+})V^{-1}V - 0 = M_V \in \mathrm{Sym}(n)$.

(PK3) $MV.M_V V^{-1}.MV = M.MV = MV$.

(PK4) $M_V V^{-1}.MV.M_V V^{-1} = M_V.M_V V^{-1} = M_V V^{-1}$.

II. $(MV^2M)^+ = (VM)^+(MV)^+$; denn beispielsweise:

(PK1) $MV^2M(VM)^+(MV)^+ = MV.VM(VM)^+(MV)^+ = MV.(MV)^+MV.(MV)^+ =$
$= MV(MV)^+ \in Sym(n)$.

III. $R = (MV^2M)^+ \underset{II}{=} (MV)^{+T}(MV)^+ \underset{I}{=} V^{-1}M_V.M_V V^{-1} = V^{-1}M_V V^{-1}$.

Damit folgt die angegebene Darstellung, da (15.iv:33)

$(D^T.R\otimes R.D)^+D^T.R\otimes R =$

$= (D^T.V^{-1}\otimes V^{-1}.M_V\otimes M_V.M_V\otimes M_V.V^{-1}\otimes V^{-1}.D)^+.D^T.V^{-1}\otimes V^{-1}.M_V\otimes M_V.V^{-1}\otimes V^{-1}$ =

$= (M_V\otimes M_V.V^{-1}\otimes V^{-1}.D)^+.V^{-1}\otimes V^{-1} = (M_V V^{-1}\otimes M_V V^{-1}.D)^+.V^{-1}\otimes V^{-1}$. _/

Satz 48: RAOs 1971-1972-MINQUE-Formel

In einem GM-Modell $(Y,Xb,n,p,\Sigma\ t_\varkappa V_\varkappa^{\ 2})$ sei Lt eine Linearform des Regressionsparameters t : $L\in M(1,k)$;
sei $V^2 = \Sigma\ V_\varkappa^{\ 2}$ p.d.; setze
$R := V^{-2}(I_n-X(X^TV^{-2}X)^+X^TV^{-2}) \in M(n)$,
$S := (tr\ RV_\varkappa^{\ 2}RV_\lambda^{\ 2})_{(k,k)} \in Sym(k)$. Dann gilt:

i (Erw.-treue Schätzbarkeit) (RAO 1972:114)
Lt ist genau dann erw.-treu schätzbar, wenn $Sx = L^T$ lösbar ist.

ii (Schätzer) (RAO 1971a:270)
Sei Lt erw.-treu schätzbar und λ eine Lösung von $Sx = L^T$. Dann ist $Y^TR.\Sigma\ \lambda_\varkappa V_\varkappa^{\ 2}.RY$ der einzige und erw.-treue Min.RAOnorm-Min.Bias-Schätzer (MINQUE) für Lt.

ii*(Verallgemeinerung)
Sei λ eine Lösung von $\|Sx-L^T\| = inf$. Dann ist $Y^TR.\Sigma\ \lambda_\varkappa V_\varkappa^{\ 2}.RY$ der einzige Min.RAOnorm-Min.Bias-schätzer für Lt .

Bew.: I. Setze $M_V = I_n-V^{-1}X(V^{-1}X)^+$, dann gilt:

$R = V^{-2}(I_n-X(V^{-1}X)^+V^{-1}) = V^{-1}(I_n-V^{-1}X(V^{-1}X)^+)V^{-1} =$
$= V^{-1}M_V V^{-1} = (MV^2M)^+ \in Sym(n)$

wie gerade in 47.I-III:90 bewiesen wurde.

II. $S = D^T.R\otimes R.D$; denn:

$S = (tr\ RV_\varkappa^{\ 2}RV_\lambda^{\ 2})_{(k,k)} = (tr\ RV_\varkappa^{\ 2}(V_\lambda^{\ 2}R)^T)_{(k,k)} =$

$\underset{13.iii:28}{=} (vec^T RV_\varkappa^{\ 2}.vec\ V_\lambda^{\ 2}R)_{(k,k)} =$

$\underset{13.ii:28}{=} (\ (R\otimes I_n.vec\ V_\varkappa^{\ 2})^T.I_n\otimes R.vec\ V_\lambda^{\ 2}\ )_{(k,k)} =$

$= (\ vec^T V_\varkappa^{\ 2}.R\otimes R.vec\ V_\lambda^{\ 2}\ )_{(k,k)} =$

$= D^T.R\otimes R.D$.

III. Mit V^2 ist auch $V^2 \otimes V^2$ regulär (12.ii:26); nach 46.ii:88 existiert genau ein Min.RAOnorm-Min.Bias-Schätzer: die Darstellungen in 46.i:88 und 47.ii:90 sind also gleich:

$(M\otimes M.D)^+(I_{n^2}-V\otimes V.(N.V\otimes V)^+) \;=\; (D^T.R\otimes R.D)^+D^T.R\otimes R$.

IV. $(N.V\otimes V)^+D = 0$; denn:

$(N.V\otimes V)^+D \;=\; (N.V\otimes V)^+.ND =$

$\overset{17.v:34}{=} (N.V\otimes V)^+.(M\otimes M - M\otimes M.D(M\otimes M.D)^+).M\otimes M.D = 0$.

V. $S^+S = (M\otimes M.D)^+D = D^T.M\otimes M.(D^T.M\otimes M)^+$; denn mit II-IV :

$S^+S = (D^T.R\otimes R.D)^+D^T.R\otimes R.D =$

$= (M\otimes M.D)^+(I_{n^2}-V\otimes V.(N.V\otimes V)^+).D =$

$= (M\otimes M.D)^+D - 0 = D^T.M\otimes M.(D^T.M\otimes M)^+$.

VI. ii. Mit 22:41; V; 41.ii:81 :

$Sx = L^T$ lösbar $\Leftrightarrow$ $SS^+L^T = L^T$ $\Leftrightarrow$ $LS^+S = L$ $\Leftrightarrow$

$\Leftrightarrow$ $L.(M\otimes M.D)^+D = L$ $\Leftrightarrow$ Lt erw.-treu schätzbar .

VII. $\forall a\in R^k$:

$Y^TR.\Sigma a_\varkappa V_\varkappa{}^2.RY \;=\; \mathrm{tr}\; R.\Sigma a_\varkappa V_\varkappa{}^2.R.YY^T \;=$

$= (\;\mathrm{vec}\; R.\Sigma a_\varkappa V_\varkappa{}^2.R\;)^T\; \mathrm{vec}\; YY^T =$

$= (\;\Sigma\; (\mathrm{vec}\; V_\varkappa{}^2)a_\varkappa\;)^T.R\otimes R.Y\otimes Y = a^T.D^T.R\otimes R.Y\otimes Y$,

mit derselben Schlußweise wie in Satz 45.V:85 .

VIII. Wir beweisen die allgemeinere Aussage ii* :

$\forall \lambda\in \mathrm{Lsg}(\|Sx-L^T\|=\inf) = \{\; (D^T.R\otimes R.D)^+L^T + (I_k-S^+S)z \;:\; z\in R^k \;\}$:

$Y^TR.\Sigma \lambda_\varkappa V_\varkappa{}^2.RY =$

$= L(D^T.R\otimes R.D)^+D^T.R\otimes R.Y\otimes Y +$

$+ z^T.(I_k-D^T.M\otimes M.(D^T.M\otimes M)^+).D^T.M\otimes M.R\otimes R.Y\otimes Y =$

$= L.(D^T.R\otimes R.D)^+D^T.R\otimes R.Y\otimes Y +0$,

dies ist nach 47.ii:90 der Min.RAOnorm-Min.Bias-Schätzer für Lt .

8.2 Exkurs: RAOs Aufgabenstellung zu seiner MINQUE-Theorie

RAO (1970; 1971a; 1972) schätzt mit quadratischen Formen $Q(Y) = Y^TAY$ Linearformen Lt des Regressionsparameters t , $L\in M(1,k)$, $A\in \mathrm{Sym}(n)$.

Für die Darstellung im abgeleiteten Modell erhält man:

$Y^TAY = \mathrm{tr}\; A.YY^T = \mathrm{vec}^TA.Y\otimes Y$, $\hat{L} = \mathrm{vec}^TA$.

Wie man leicht zeigt, ist unsere Invarianzbedingung: $A = MAM$ äquivalent zu der von RAO: $AX = 0$.

Für einen invarianten Schätzer Y^TAY , d.h. $A = MAM$, reduziert sich das Kriterium für Erw.-treue (41.i:81) zu:

$$L = \hat{L}.M\otimes M.D = vec^TA.M\otimes M.D = vec^TMAM.D = vec^TA.D =$$

$$= (vec^TA.vec\ D_\varkappa)_{\varkappa=1,\dots,k}^T = (tr\ AD_\varkappa)_{\varkappa=1,\dots,k}^T \quad ;$$

dies fordert RAO als Bedingung für Erw.-treue.

Für ein MINQUE-Modell $(Y,Xb,n,p,\Sigma\, t_\varkappa V_\varkappa^{\ 2})$, $Y = Xb+U\xi$ (Seite 87) begründet RAO (1971a:267f) die Normminimierung so:

Bei bekannten hypothetischen Variablen ξ wäre ein natürlicher Schätzer für eine Linearform $Lt = \Sigma\, L_\varkappa t_\varkappa$ gegeben durch

$$\Sigma\, L_\varkappa . \tfrac{1}{c_\varkappa} . \xi_\varkappa^{\ T} \xi_\varkappa \;=\; \xi^T \Delta\, \xi \quad , \quad \Delta := \begin{bmatrix} \frac{L_1}{c_1}.I_{c_1} & & \\ & \ddots & \\ & & \frac{L_k}{c_k}.I_{c_k} \end{bmatrix} \in \mathrm{Diag}(c)\ .$$

Tatsächlich wird aber geschätzt mit

$$Y^TAY = \xi^TU^TAU\xi + 2b^TX^TAU\xi + b^TX^TAXb \underset{\text{Invarianz}}{=} \xi^TU^TAU\xi + 0 + 0\ .$$

Die Abweichung $\xi^T(U^TAU - \Delta)\xi$ des tatsächlichen vom natürlichen Schätzer wird nun in gewissem Sinne klein gemacht, wenn $\|U^TAU - \Delta\|$ minimiert wird. Es ist aber

$$\begin{aligned}
\|U^TAU - \Delta\|^2 &= tr\ U^TAUU^TAU - 2tr\ U^TAU\Delta + tr\ \Delta\Delta = \\
&= tr\ AV^2AV^2 - 2\ tr\ A.\Sigma\, \tfrac{L_\varkappa}{c_\varkappa}.V_\varkappa^{\ 2} + tr\ \Delta\Delta = \\
&= tr\ AV^2AV^2 - 2\,\Sigma\, \tfrac{L_\varkappa}{c_\varkappa}.tr\ AV_\varkappa^{\ 2} + tr\ \Delta\Delta = \\
&\underset{\text{Erw.-treue}}{=} tr\ AV^2AV^2 - 2\,\Sigma\, c_\varkappa^{\ -1}.L_\varkappa^{\ 2} + tr\ \Delta\Delta = \\
&= tr\ AV^2AV^2 - 2\ tr\ \Delta\Delta + tr\ \Delta\Delta = \\
&= tr\ AV^2AV^2 - tr\ \Delta\Delta\ .
\end{aligned}$$

Die Minimierung von $\|U^TAU - \Delta\|$ ist also gleichwertig mit der Minimierung von $tr\ AV^2AV^2$. Für diesen Ausdruck gilt:

$$tr\ AV^2AV^2 = tr\ VAV.VAV = vec^TVAV\ vec\ VAV = vec^TA.V^2\otimes V^2.vec\ A =$$

$$= \|\ vec^TA.V\otimes V\ \|^2 = \|\ \hat{L}.V\otimes V\ \|^2\ .$$

Die Minimierung der Norm $\|\ vec^TA.V\otimes V\ \|$ unter Einschluß der Invarianzbedingung $A = MAM$ und der Bedingung für Erw.-treue stellt offenbar dieselbe Aufgabe wie die Suche nach erw.-treuen Min.RAOnorm- Min.Bias-Schätzern für Linearformen Lt .

8.3 Minimum Varianz - Minimum Bias - inv. quadr. Schätzfunktionen

Mit dem Lexikon $V^2 \otimes V^2 \mid F^2$ folgt aus 46:88 der

Satz 49: Hauptsatz über Min.Var.-Min.Bias-Schätzer für Lt

In einem GM-Modell $(Y, Xb, n, p, \Sigma\, t_\varkappa D_\varkappa, F^2)$ sei Lt eine lineare Funktion des Regressionsparameters t : $L \in M(1,k)$; setze $M = I_n - XX^+$, $D = [\mathrm{vec}\, D_\varkappa]_{\varkappa=1,\dots,k}$, $N = M\otimes M - M\otimes M.D(M\otimes M.D)^+$. Dann gilt:

i (Alle Min.Var.-Min.Bias-Schätzer für Lt)

Die Menge aller Min.Var.-Min.Bias-Schätzer für Lt ist $\{ L.(M\otimes M.D)^+(I_{n^2} - F(NF)^+).Y\otimes Y + Z(N - NF(NF)^+).Y\otimes Y : Z \in M(1,n^2) \}$.

ii (Fast sichere Einzigkeit)

Unter allen Min.Var.-Min.Bias-Schätzern für Lt ist $L.(M\otimes M.D)^+(I_{n^2} - F(NF)^+).Y\otimes Y$ fast sicher einzig.

iii (Einzigkeit)

$L.(M\otimes M.D)^+(I_{n^2} - F(NF)^+).Y\otimes Y$ ist der einzige Min.Var.-Min.Bias-Schätzer für Lt genau dann, wenn $\mathrm{im}\, N \subset \mathrm{im}\, M\otimes M.F^2$.

Als Beispiel für diesen Satz bringen wir ein Ergebnis von DRYGAS (1972). DRYGAS rechnet in demselben Modell wie HSU (1938): Ein "<u>HSU-Modell</u>" sei ein GM-Modell (Y,Xb,n,p) , in dem die Komponenten von Y sich bis zu den vierten Momenten wie unabhängige Variable verhalten, gleiche (unbekannte) Varianz σ^2 und gleiche bekannte Kurtosis β haben[24]: $E_{\sigma^2,\beta} Y_\nu^4 = \beta\sigma^4$.

Satz 50: Beispiel für Min.Var.-Min.Bias-Schätzer für Lt

(DRYGAS 1972:378) Gegeben sei ein HSU-Modell und eine inv. quadr. Schätzfunktion Y^TAY , $A \in \mathrm{Sym}(n)$, $A = MAM$. Dann gilt:

Y^TAY ist erw.-treuer Min.Var.-Min.Bias-Schätzer für σ^2 genau dann,

wenn $\mathrm{tr}\, A = 1$ und $2A + (\beta-3)M.\mathrm{Diag}\, A.M \in \mathrm{span}\, M$.

<u>Bew.:</u> Wir berechnen zunächst die zu einem HSU-Modell gehörende Matrix $F_{\sigma^2,\beta}^{\;2}$ der vierten Momente (I) und beweisen dann die Behauptung (II-IV).

---

24 Eine Verallgemeinerung dieses HSU-Modells untersucht RAO (1971b) .

I. $F_{\sigma^2,\beta}^{\;2} = \sigma^4(I_{n^2} + \Sigma\, E_{\mu\nu} \otimes E_{\nu\mu} + (\beta-3)\, \Sigma\, E_{\nu\nu} \otimes E_{\nu\nu})$,
wobei auch μ von 1 bis n läuft und $E_{\nu\mu}$ steht für
$E_{\nu\mu} = E_{\nu\mu}^{(n,n)} = e_\nu^n (e_\mu^n)^T$, d.i. die (ν,μ)-te (n,n)-Basismatrix;
setze $\delta_{ijkl} := \delta_{ij}\delta_{kl}\delta_{ik} = \begin{cases} 1 & \text{falls } i=j=k=l \\ 0 & \text{sonst} \end{cases}$; dann
gilt für die Matrix der vierten Momente in einem HSU-Modell,
wobei wegen der Beschränkung auf invariante Schätzer o.B.d.A.
Y als zentriert angenommen werden kann: $EY = 0$:

$$\mathrm{vec}^T F_{\sigma^2,\beta}^{\;2} = \mathrm{vec}^T\, D\, Y \otimes Y =$$
$$= \mathrm{vec}^T(E\, Y\otimes Y.(Y\otimes Y)^T - E\, Y\otimes Y.E^T Y\otimes Y) =$$
$$= (E\, Y_iY_jY_kY_l - E\, Y_iY_j.E\, Y_kY_l)_{l,k,j,i=1,\dots,n} =$$
$$= \sigma^4(\delta_{ij}\delta_{kl}+\delta_{ik}\delta_{jl}+\delta_{il}\delta_{jk} - 3\delta_{ijkl} + \beta\delta_{ijkl} - \delta_{ij}\delta_{kl})_{l,k,j,i=1,\dots,n} =$$
$$= \sigma^4(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk} + (\beta-3)\delta_{ijkl})_{l,k,j,i=1,\dots,n} =$$
$$= \sigma^4.\Sigma(\delta_{i\mu}\delta_{k\mu}\delta_{j\nu}\delta_{l\nu} + \delta_{i\mu}\delta_{l\mu}\delta_{j\nu}\delta_{k\nu} + (\beta-3)\delta_{\mu\nu}\delta_{i\mu}\delta_{j\mu}\delta_{k\nu}\delta_{l\nu})_{l,k,j,i} \;;$$

$$F_{\sigma^2,\beta}^{\;2} =$$
$$= \sigma^4.\Sigma\, [\delta_{i\mu}\delta_{k\mu}.(\delta_{j\nu}\delta_{l\nu})_{\substack{j=1,\dots,n\\ l=1,\dots,n}} + \delta_{i\mu}\delta_{k\nu}.(\delta_{j\nu}\delta_{l\mu})_{\substack{j=1,\dots,n\\ l=1,\dots,n}} +$$
$$+ (\beta-3)\delta_{\mu\nu}\delta_{i\mu}\delta_{k\nu}.(\delta_{j\mu}\delta_{l\nu})_{\substack{j=1,\dots,n\\ l=1,\dots,n}}]_{\substack{i=1,\dots,n\\ k=1,\dots,n}} =$$
$$= \sigma^4.\Sigma(E_{\mu\mu} \otimes E_{\nu\nu} + E_{\mu\nu} \otimes E_{\nu\mu} + (\beta-3).\delta_{\mu\nu}.E_{\mu\nu} \otimes E_{\mu\nu})$$
$$= \sigma^4(I_{n^2} + \Sigma\, E_{\mu\nu} \otimes E_{\nu\mu} + (\beta-3)\, \Sigma\, E_{\nu\nu} \otimes E_{\nu\nu}) .$$

II. Setze $F_\beta^{\;2} := F_{1,\beta}^{\;2}$. Dann
$M\otimes M.F_\beta^{\;2}.\mathrm{vec}\, A = \mathrm{vec}(2\, MAM + (\beta-3)\, M.\mathrm{Diag}\, A.M)$; denn (13:28) :

$$M\otimes M.F_\beta^{\;2}.\mathrm{vec}\, A \underset{I}{=} (M\otimes M + \Sigma\, ME_{\mu\nu} \otimes ME_{\nu\mu} + (\beta-3)\Sigma\, ME_{\nu\nu} \otimes ME_{\nu\nu})\mathrm{vec}\, A =$$
$$= \mathrm{vec}\, MAM + \Sigma\, \mathrm{vec}\, ME_{\mu\nu}A(ME_{\nu\mu})^T + (\beta-3)\, \Sigma\, \mathrm{vec}\, ME_{\nu\nu}A(ME_{\nu\nu})^T =$$
$$= \mathrm{vec}\, MAM + \mathrm{vec}\, M.\Sigma\, E_{\mu\nu}AE_{\mu\nu}.M + (\beta-3)\, \mathrm{vec}\, M.\Sigma\, E_{\nu\nu}AE_{\nu\nu}.M =$$
$$= \mathrm{vec}\, MAM + \mathrm{vec}\, MA^TM + (\beta-3)\, M.\mathrm{Diag}\, A.M =$$
$$= \mathrm{vec}(2\, MAM + (\beta-3)\, M.\mathrm{Diag}\, A.M) \;.$$

III. Y^TAY ist erw.-treu genau dann, wenn $\mathrm{tr}\, A = 1$; denn:
$Y^TAY = \mathrm{tr}\, A.MY(MY)^T = \mathrm{vec}^T A.MY\otimes MY$ ist nach 41.i:81 erw.-treu
genau bei $1 = \mathrm{vec}^T A.M\otimes M.D = \mathrm{vec}^T MAM.\mathrm{vec}\, I_n = \mathrm{tr}\, A$.

IV. Zur Behauptung: Mit $\hat{L} := \mathrm{vec}^T A$ und den Bezeichnungen des Hauptsatzes 49:94 ($D = \mathrm{vec}\, I_n$, etc.) folgt allgemeiner:

$$Y^TAY \text{ inv.} \wedge \|1-\mathrm{tr}\,A\| = \inf \wedge 2A + (\beta-3)M.\mathrm{Diag}\,A.M \in \mathrm{span}\,M$$

$$\underset{II}{\Leftrightarrow} A=MAM \wedge \|1-\mathrm{tr}\,A\| = \inf \wedge \exists\mu\in R\text{: } 2\,MAM + (\beta-3)M.\mathrm{Diag}\,A.M = \mu M$$

$$\Leftrightarrow A=MAM \wedge \|1-\mathrm{tr}\,MAM\|=\inf \wedge \exists\mu\in R\text{: } M\otimes M.F_\beta^{\,2}.\mathrm{vec}\,A = (\mathrm{vec}\,M).\mu$$

$$\Leftrightarrow A=MAM \wedge \|1-\mathrm{tr}\,MA\| = \inf \wedge (\mathrm{vec}\,M)x = M\otimes M.F_\beta^{\,2}.\hat{L}^T \text{ lösbar}$$

$$\overset{\text{kom.}}{\Leftrightarrow} A=MAM \wedge \|1-\mathrm{tr}\,MA\| = \inf \wedge \mathrm{vec}\,M.\mathrm{vec}^+M.M\otimes M.F_\beta^{\,2}.\hat{L}^T = M\otimes M.F_\beta^{\,2}.\hat{L}^T$$

$$\Leftrightarrow A=MAM \wedge \|\mathrm{vec}^T M.\hat{L}^T - 1\| = \inf \wedge$$

$$0 = (M\otimes M - \mathrm{vec}\,M.\mathrm{vec}^+M.M\otimes M).F_\beta^{\,2}.\hat{L}^T =$$

$$= (M\otimes M - M\otimes M.\mathrm{vec}\,I_n\,.\underline{(M\otimes M.\mathrm{vec}\,I_n)^+.M\otimes M})F_\beta^{\,2}.\hat{L}^T =$$

$$= (M\otimes M - M\otimes M.D(M\otimes M.D)^+).F_\beta^{\,2}.\hat{L}^T =$$

$$\overset{17.v:34}{=} NF_\beta^{\,2}\hat{L}^T$$

$$\Leftrightarrow \hat{L} = \hat{L}.M\otimes M \wedge \hat{L}^T \in \mathrm{Lsg}(\|\mathrm{vec}^T M.x - 1\|=\inf) =$$

$$= \{ \mathrm{vec}^{T+}M + (I_{n^2}-\mathrm{vec}^{T+}M.\mathrm{vec}^T M)z : z\in R^{n^2} \} \wedge$$

$$\wedge \hat{L}^T \in \mathrm{Lsg}(NF_\beta^{\,2}x=0) \overset{23:42}{=} \{ (I_{n^2}-(NF_\beta^{\,2})^+NF_\beta^{\,2})z : z\in R^{n^2} \}$$

$$\Leftrightarrow \hat{L} \in \{ \mathrm{vec}^+M + Z(M\otimes M-\mathrm{vec}\,M.\mathrm{vec}^+M) : Z\in M(1,n^2) \} =$$

$$= \{ (M\otimes M.D)^+ + ZN : Z\in M(1,n^2) \} \wedge$$

$$\wedge \hat{L}^T \in \{ (I_{n^2}-(NF_\beta^{\,2})^+NF_\beta^{\,2})z : z\in R^{n^2} \}$$

$$\Leftrightarrow \exists Z\in M(1,n^2)\text{: } \hat{L} = (M\otimes M.D)^+ + ZN \wedge (I_{n^2}-(NF_\beta^{\,2})^+NF_\beta^{\,2})x = \hat{L}^T \text{ lösbar}$$

$$\Leftrightarrow \exists Z\in M(1,n^2)\text{: } \hat{L} = (M\otimes M.D)^+ + ZN \wedge (I_{n^2}-(NF_\beta^{\,2})^+NF_\beta^{\,2})\hat{L}^T = \hat{L}^T$$

$$\overset{22:41}{\Leftrightarrow} \exists Z\in M(1,n^2)\text{: } \hat{L} = (M\otimes M.D)^+ + ZN \wedge (NF_\beta^{\,2})^+NF_\beta^{\,2}\hat{L}^T = 0$$

$$\Leftrightarrow \exists Z\in M(1,n^2)\text{: } \hat{L} = (M\otimes M.D)^+ + ZN \wedge \underline{NF_\beta^{\,2}.(NF_\beta^{\,2})^+NF_\beta^{\,2}}\hat{L}^T = NF_\beta^{\,2}\hat{L}^T = 0$$

$$\Leftrightarrow \exists Z\in M(1,n^2)\text{: } \hat{L} = (M\otimes M.D)^+ + ZN \wedge NF_\beta^{\,2}NZ^T = -\,NF_\beta^{\,2}(D^T.M\otimes M)^+$$

$$\Leftrightarrow \exists Z\in M(1,n^2)\text{: } \hat{L} = (M\otimes M.D)^+ + ZN \wedge$$

$$\wedge Z^T \in \mathrm{Lsg}(NF_\beta^{\,2}Nx=-NF_\beta^{\,2}(D^T.M\otimes M)^+) =$$

$$= \{ -(NF_\beta F_\beta N)^+NF_\beta F_\beta(D^T.M\otimes M)^+ + (I_{n^2}-(NF_\beta F_\beta N)^+NF_\beta F_\beta N)z : z\in R^{n^2} \}$$

$$= \{ -(F_\beta N)^+F_\beta(D^T.M\otimes M)^+ + (I_{n^2}-(F_\beta N)^+F_\beta N)z : z\in R^{n^2} \}$$

$$\overset{15.iv:33}{\Leftrightarrow} \exists Z\in M(1,n^2)\text{: } \hat{L} = (M\otimes M.D)^+ - (M\otimes M.D)^+F_\beta(NF_\beta)^+ + Z(N-(F_\beta N)^+F_\beta N) =$$

$$= (M\otimes M.D)^+(I_{n^2}-F_\beta(NF_\beta)^+) + Z(N-NF_\beta(NF_\beta)^+)$$ _/

Welche spezielle Darstellungen gibt es für Min.Var.-Min.Bias-Schätzer? Genauer:

i. Wann hat der Min.Norm-Min.Bias-Schätzer für t gleichzeitig minimale Varianz?

ii. Wann hat der Min.Norm-Min.Bias-Schätzer für t im transformierten abgeleiteten Modell minimale Varianz im ursprünglichen Modell?

iii. Wann hat der kürzeste Min.RAOnorm-Min.Bias-Schätzer für t minimale Varianz?

iv. Wann sind alle Min.RAOnorm-Min.Bias-Schätzer für t auch Min.Var.-Min.Bias-Schätzer?

Die beiden ersten Fragen sind schon in Satz 47:90 beantwortet worden; man setze das Lexikon $V^2 \otimes V^2 \mid F^2$.

Die beiden letzten Fragen kann man anders stellen:

iii. Wann hat der kürzeste Min.Var.-Min.Bias-Schätzer für t im erweiterten Modell (mit vierten Momenten $V^2 \otimes V^2$) minimale Varianz im ursprünglichen Modell (mit vierten Momenten F^2)?

iv. Wann sind alle Min.Var.-Min.Bias-Schätzer im erweiterten Modell Min.Var.-Min.Bias-Schätzer im ursprünglichen Modell?

Zum erweiterten Modell $(Y, Xb, n, p, \Sigma\, t_\varkappa V_\varkappa^2, V^2 \otimes V^2)$ gehört das abgeleitete Modell $(MY \otimes MY,\ M \otimes M.Dt, n^2, k,\ M \otimes M.V^2 \otimes V^2.M \otimes M)$; zum ursprünglichen Modell $(Y, Xb, n, p, \Sigma\, t_\varkappa V_\varkappa^2, F^2)$ gehört das abgeleitete Modell $(MY \otimes MY,\ M \otimes M.Dt, n^2, k,\ M \otimes M.F^2.M \otimes M)$. Die Antwort zu iii und iv steht daher im Satz 35:65 über Robustheit gegen Fehler im Streuung-Design bei gleichem Mittelwert-Design:

<u>Satz 51: Spezielle Darstellungen für Min.Var.-Min.Bias-Schätzer</u>

Gegeben sei ein GM-Modell wie in 49:94; für die Aussagen iii, iv sei $D_\varkappa = V_\varkappa^2$ p.s.d. . Dann gilt:

i (= Min.Norm-Min.Bias-Schätzer)
$(M \otimes M.D)^+ .Y \otimes Y$ ist genau dann Min.Var.-Min.Bias-Schätzer für t, wenn $\mathrm{im}\, M \otimes M.F^2.M \otimes M.D \subset \mathrm{im}\, M \otimes M.D$.

ii (Über das transformierte abgeleitete Modell)
$(\ D^T (M \otimes M.F^2.M \otimes M)^+ D\ )^+ .D^T (M \otimes M.F^2.M \otimes M)^+ .Y \otimes Y$ ist genau dann Min.Var.-Min.Bias-Schätzer für t, wenn $\mathrm{im}\, M \otimes M.D \subset \mathrm{im}\, M \otimes M.F^2$.

iii (= kleinstem Min.RAOnorm-Min.Bias-Schätzer)
$(M\otimes M.D)^+(I_{n^2}-V\otimes V(N.V\otimes V)^+).Y\otimes Y$ ist genau dann Min.Var.-Min.Bias-Schätzer für t ,
wenn $\operatorname{im} M\otimes M.F^2N \subset \ker (M\otimes M.D)^+(I_{n^2}-V\otimes V.(N.V\otimes V)^+)$.
iv (= allen Min.RAOnorm-Min.Bias-Schätzer)
Jeder Min.RAOnorm-Min.Bias-Schätzer für t ist genau dann auch Min.Var.-Min.Bias-Schätzer für t ,
wenn $\operatorname{im} M\otimes M.F^2N \subset \operatorname{im} MV^2\otimes MV^2.N$. _/

Ein Beispiel für diesen Satz ist eines der ältesten Ergebnisse, das zur Streuung-Schätzung bewiesen wurde, der

<u>Satz 52: Satz von HSU</u>

i(HSU 1938:100) Gegeben sei ein HSU-Modell (Seite 94)
mit $n > \operatorname{rk} X = s$; setze $M = I_n - XX^+$,
$M_2 = (M_{\nu\mu}^2)_{(n,n)} \in \operatorname{Sym}(n)$, $m := (M_{\nu\nu})^T_{\nu=1,\dots,n} \in R^n$.
Dann gilt:
$(n-s)^{-1}Y^TMY$ ist genau dann (erw.-treuer und p.s.d.) Min.Var.-Min.Bias-Schätzer für σ^2 ,
wenn $(\beta-3).M.\operatorname{Diag} M.M \in \operatorname{span} M$,
d.h. $\beta = 3$ oder $M_2m = \operatorname{tr} M_2.(\operatorname{tr} M)^{-1}.m$.

ii (Normalverteilte Komponenten)
In einem GM-Modell (Y,Xb,n,p) , $n > \operatorname{rk} X = s$ seien die Komponenten von Y unabhängig und normalverteilt mit gleichen Varianzen σ^2 .
Dann ist $(n-s)^{-1}Y^TMY$ Min.Var.-Min.Bias-Schätzer für σ^2 , er ist einzig, erw.-treu und p.s.d. .
iii (Gleiche Mittelwerte, gleiche Varianzen, gleiche Kurtosis)
Y habe unabhängige Komponenten mit gleichen Mittelwerten, gleichen Varianzen und gleicher Kurtosis.
Dann ist die Stichprobenstreuung $(n-1)^{-1}.\Sigma(Y_\nu-\bar{Y})^2$ erw.-treuer und p.s.d. Min.Var.-Min.Bias-Schätzer für die Varianz σ^2 .

<u>Bew.:</u> i. Nach 44.ii:83 ist $(n-s)^{-1}Y^TMY$ gerade der Min.Norm-Min.Bias-Schätzer für σ^2 . In unserer Sprechweise stellt sich also die Frage: Wann hat der Min.Norm-Min.Bias-Schätzer gleichzeitig minimale Varianz? Dafür gibt 51.i:97 das Kriterium $\operatorname{im} M\otimes M.F_\beta^2.M\otimes M.D \subset \operatorname{im} M\otimes M.D$. Dabei ist $D = \operatorname{vec} I_n$ und F_β^2 die in 50(1,II):95 angegebene Matrix der vierten Momente:

$\mathrm{im}\ M\otimes M.F_\beta^{\,2}.M\otimes M.\mathrm{vec}\ I_n \subset \mathrm{im}\ M\otimes M.\mathrm{vec}\ I_n \Leftrightarrow$

$\Leftrightarrow \mathrm{span}\ M\otimes M.F_\beta^{\,2}.\mathrm{vec}\ M \subset \mathrm{span}\ \mathrm{vec}\ M \Leftrightarrow$

$\Leftrightarrow \mathrm{vec}\ (\ 2\ M + (\beta-3).M.\mathrm{Diag}\ M.M\ ) \in \mathrm{span}\ \mathrm{vec}\ M \Leftrightarrow$

II:95

$\Leftrightarrow 2\ M + (\beta-3).M.\mathrm{Diag}\ M.M \in \mathrm{span}\ M \Leftrightarrow$

$\Leftrightarrow (\beta-3).M.\mathrm{Diag}\ M.M \in \mathrm{span}\ M$.

Zur zweiten Äquivalenz:

I. $\Rightarrow$: Sei $(\beta-3)M.\mathrm{Diag}\ M.M \in \mathrm{span}\ M$. Dann ist $\beta = 3$, oder es existiert eine reelle Zahl ρ mit:

$M.\mathrm{Diag}\ M.M = \rho M$, folglich – wegen $\sum\limits_\mu M_{\nu\mu}M_{\mu\nu} = (M^2)_{\nu\nu} = M_{\nu\nu}$) – :

$\rho\ \mathrm{tr}\ M = \Sigma\ M_{\mu\nu}M_{\nu\nu}M_{\nu\mu} = \Sigma\ M_{\nu\nu}^{\ 2} = \mathrm{tr}\ M_2$,

$\rho = \mathrm{tr}\ M_2.(\mathrm{tr}\ M)^{-1}$ und

$\rho m = (\Sigma\ M_{i\nu}M_{\nu\nu}M_{\nu i})_{i=1,\dots,n}^{\ \ T} = (\Sigma\ M_{i\nu}^{\ 2}M_{\nu\nu})_{i=1,\dots,n}^{\ \ T} = M_2 m$.

II. $\Leftarrow$: Sei $M_2 m = \rho m$, $\rho = \mathrm{tr}\ M_2.(\mathrm{tr}\ M)^{-1}$. Dann

$o = m^T M_2 m - \rho m^T m = \Sigma\ M_{\nu\nu}M_{\nu\mu}^{\ 2}M_{\mu\mu} - 2\ \rho\ \mathrm{tr}\ M_2 + \rho\,\mathrm{tr}\ M_2 =$

$= \Sigma\ M_{\nu\mu}M_{\mu\mu}M_{\mu\nu}M_{\nu\nu} - 2\ \rho\ \mathrm{tr}\ M.\mathrm{Diag}\ M + \rho\ \mathrm{tr}\ M_2\ .(\mathrm{tr}\ M)^{-1+1} =$

$= \mathrm{tr}\ M.\mathrm{Diag}\ M.\ M.\mathrm{Diag}\ M - 2\ \rho\ \mathrm{tr}\ M^2.\mathrm{Diag}\ M + \rho^2.\mathrm{tr}\ M =$

$= \mathrm{tr}\ (M.\mathrm{Diag}\ M.M)^2 - 2\mathrm{tr}\ \rho M.M.\mathrm{Diag}\ M.M + \mathrm{tr}\ (\rho M)^2 =$

kom.

$= \mathrm{tr}\ (M.\mathrm{Diag}\ M.M - \rho M)^2$, also

$M.\mathrm{Diag}\ M.M = \rho M \in \mathrm{span}\ M$.

ii. Die Normalverteilung hat die Kurtosis $\beta = 3$. Außer dem ist die Matrix $F_\beta^{\,2}$ der vierten Momente p.d. (vgl. I:95), nach 49.iii:94 ist der angegebene Schätzer der einzige.

iii. Bei gleichen Mittelwerten ist $X = 1_n$:

$M = I_n - \frac{1}{n}.1_{(n,n)}$, $\mathrm{Diag}\ M = \frac{n-1}{n}.I_n$, $M.\mathrm{Diag}\ M.M = \frac{n-1}{n}.M$.

(Vgl. 29.iii:56; 40.ii:74; 44.iii:83) _/

Kap. 9: Minimum mean square error - invariante quadratische Schätzfunktionen

9.1 Minimum mean square error - inv. quadr. Schätzfunktionen

Übertragung der Ergebnisse des Hauptsatzes 37:70 über lokale GAUSS-Schätzer mittels der Streuung-Mittelwert-Korrespondenz gibt den

Satz 53: Hauptsatz über $S_k(\tau)$-lokale GAUSS-Schätzer für Lt [25]

In einem GM-Modell $(Y, Xb, n, p, \Sigma\, t_\varkappa D_\varkappa, F^2)$ sei Lt eine lineare Funktion des Regressionsparameters t : $L \in M(1,k)$;

setze $M = I_n - XX^+$, $D = [\mathrm{vec}\, D_\varkappa]_{\varkappa=1,\dots,k}$,

$N = M\otimes M - M\otimes M.D(M\otimes M.D)^+$, $T = (M\otimes M.D)^+(I_{n^2} - F(NF)^+)$,

$S_k(\tau) = \{\ t \in R^k : \|t\| = \tau\ \}$.

Dann gilt für alle positiven τ :

i (Alle $S_k(\tau)$-lokalen GAUSS-Schätzer für Lt)

Die Menge aller $S_k(\tau)$-lokalen GAUSS-Schätzer für Lt ist

$\{\ L.(I_k + \tau^{-2} TF^2T^T)^{-1} T.Y\otimes Y + Z(N - NF(NF)^+).Y\otimes Y : Z \in M(1,n^2)\ \}$.

ii (Fast sichere Einzigkeit)

Unter allen $S_k(\tau)$-lokalen GAUSS-Schätzern für Lt ist

$L(I_k + \tau^{-2} TF^2T^T)^{-1} T.Y\otimes Y$ fast sicher einzig.

iii (Einzigkeit)

$L(I_k + \tau^{-2} TF^2T^T)^{-1} T.Y\otimes Y$ ist der einzige $S_k(\tau)$-lokale GAUSS-Schätzer für Lt genau dann, wenn $\mathrm{im}\, N \subset \mathrm{im}\, M\otimes M.F^2$.

_/

Wie in den vorangegangenen Kapiteln können alle Ergebnisse der Mittelwert-Schätzung übersetzt werden. Wir verzichten darauf, und bringen gleich ein Beispiel, wie Min.Var.-Min.Bias-Schätzer und lokale GAUSS-Schätzer für $t = \sigma^2$ auseinander hervorgehen. In diesem Beispiel ist $k = 1$: Unsere GAUSS-Schätzer fallen mit den üblichen zusammen. Die Abhängigkeit vom Lokalitätsparameter τ verschwindet hier: die Schätzfunktionen sind (globale) GAUSS-Schätzer im Sinne der Definition 10.ii:70 .

---

25 Quadratische Schätzfunktionen für Streuungsparameter unter dem üblichen GAUSS-Risiko betrachtet LAMOTTE (1973), die Lösungen stellt er zum Teil explizit dar.

Satz 54: Min.Var.-Min.Bias- und GAUSS-Schätzer für σ^2

(DRYGAS 1972:379f) Unter denselben Voraussetzungen wie in Satz 52.i:98 gilt für inv. quadr. Schätzfunktionen Y^TAY , $A \in Sym(n)$, $A = MAM$:

i (GAUSS-Schätzer aus Min.Var.-Min.Bias-Schätzern)
Sei Y^TAY ein Min.Var.-Min.Bias-Schätzer für σ^2 .
Dann ist $(1 + tr\ 2A^2+(\beta-3)A.Diag\ A\ )^{-1}.Y^TAY$ GAUSS-Schätzer für σ^2 .

ii (Min.Var.-Min.Bias-Schätzer aus GAUSS-Schätzern)
Sei Y^TAY GAUSS-Schätzer für σ^2 . Dann ist $(tr\ A)^{-1}.Y^TAY$ Min.Var.-Min.Bias-Schätzer für σ^2 .

iii (Vielfache von M) $\forall \lambda \in R_+$:
λY^TMY ist GAUSS-Schätzer für σ^2 genau dann, wenn $\beta = 3$ oder $M_2m = tr\ M_2.(tr\ M)^{-1}.m$, genau dann wenn $(\ n-s+2 + (\beta-3).tr\ M_2.(tr\ M)^{-1}\ )^{-1}.Y^TMY$ GAUSS-Schätzer für σ^2 .

<u>Bew.:</u> i. 38.i:73, angewendet auf $\hat{t} = vec^TA = vec^TA.M\otimes M$, gibt den ($\tau=\|t\|=\|\sigma^2\|=\sigma^2$) $\{\pm\sigma^2\}$-lokalen GAUSS-Schätzer für σ^2 :

$$(I_1+\sigma^{-4}\hat{t}F_{\sigma^2,\beta}{}^2\hat{t}^T)^{-1}\hat{t}.Y\otimes Y \underset{I:95}{=} (1+ vec^TA.M\otimes M.F_\beta{}^2.vec\ A)^{-1}.Y^TAY =$$

$$\underset{II:95}{=} (1+ vec^TA.vec\ 2A+(\beta-3)M.Diag\ A.M\ )^{-1}.Y^TAY =$$

$$\underset{13.iii:28}{=} (1 + tr\ 2A^2+(\beta-3)MAM.Diag\ A\ )^{-1}.Y^TAY \ .$$

ii. Mit 38.ii:72 erhält man den Min.Var.-Min.Bias-Schätzer $(\hat{t}.M\otimes M.D)^+\hat{t}.Y\otimes Y = (vec^TA.vec\ I_n)^+Y^TAY = (tr\ A)^+.Y^TAY$; wegen $M \neq 0$ ist σ^2 erw.-treu schätzbar (44.i:83) und also kann A nicht die Nullmatrix sein (50(III):95) .

iii. I. Aus dem GAUSS-Schätzer λY^TMY erhält man nach ii den Min.Var.-Min.Bias-Schätzer $(tr\ \lambda M)^{-1}Y^T(\lambda M)Y = (tr\ M)^{-1}Y^TMY =$
$\underset{2.i:16}{=} (rk\ M)^{-1}Y^TMY = (n-s)^{-1}Y^TMY$, und 52.i:98 ist anwendbar.

II. Mit 52.i:98 ergibt i den GAUSS-Schätzer:

$$(1 + tr\ 2(n-s)^{-2}M+(\beta-3)(n-s)^{-1}M.(n-s)^{-1}Diag\ M\ )^{-1}(n-s)^{-1}Y^TMY =$$

$$= (1+2(n-s)^{-2}(n-s)+(\beta-3)(n-s)^{-2}tr\ M_2)^{-1}(n-s)^{-1}Y^TMY =$$

$$= (n-s+2 + (\beta-3).tr\ M_2.(n-s)^{-1}\ )^{-1}.Y^TMY \ .$$

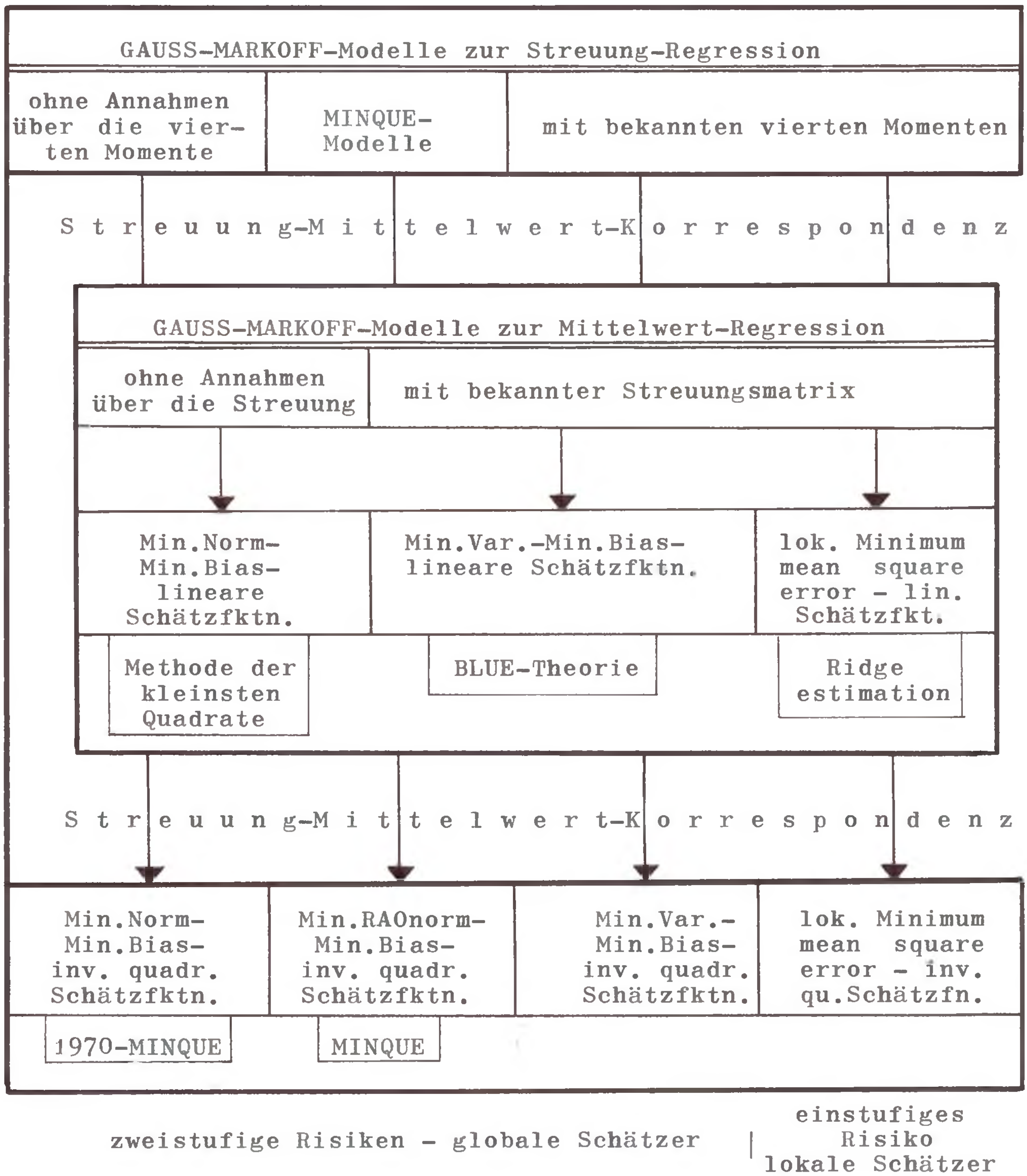
_/

9.2 Abschließender Überblick

Das Diagram zeigt die beiden Teile der Schätztheorie für GAUSS-MARKOFF-Modelle:

- Die Schätzung des Mittelwerts steht im Mittelpunkt.
- Die Schätzung der Streuung baut darauf auf.

Die einzelnen Schätzverfahren sind in das Gesamtbild eingeordnet.



Die Streuung-Mittelwert-Korrespondenz ergibt - als Anwendung der Mittelwert-Schätzung - eine wohlgegliederte Theorie zur Streuung-Schätzung. Sie ist aber auch für die Schätzung des Mittelwerts von Nutzen gewesen: RAO führt seine MINQUE-Theorie ein (1970) zur Schätzung von Streuungskomponenten: Zurückgespielt auf die Mittelwert-Schätzung gibt das die Minimum Norm-Minimum Bias-Schätzung. DRYGAS (1972:379f) beweist eine Beziehung zwischen GAUSS- und Min.Var.-Min.Bias-Schätzern für σ^2 (54:101); auch das hat ein Gegenstück schon auf der Ebene des Mittelwerts (38:73).

Eine erste - kursivgedruckte - Bemerkung in Richtung Streuung-Mittelwert-Korrespondenz macht RAO (1970:167): "Thus we have the interesting result that the MINQUE is a linear combination of the square of residuals." DRYGAS (1972:382) wird deutlicher: "[Eine invariante quadratische Schätzfunktion] is nothing else but a linear function of MYY^TM, and the best quadratic unbiased estimator of σ^2 is equal to the best linear estimator (linear in MYY^TM) of σ^2 [im abgeleiteten Modell]. The classical estimator $\widehat{\sigma^2} = (n-s)^{-1}Y^TMY$ is exactly the least squares-estimator of σ^2 [im abgeleiteten Modell]."

Die Zufallsgröße MYY^TM - als Vektor aufgefaßt - hat eine singuläre Streuungsmatrix und verlangt deshalb nach einer "ranglosen" Schätztheorie: Der bei DRYGAS durchscheinende Zusammenhang und RAOs Unified Theory of Linear Estimation geben Ziel und Mittel an die Hand, die Streuung-Schätzung auf die Schätzung des Mittelwerts eines geeignet abgeleiteten Modells zurückzuführen und so eine Streuung-Mittelwert-Korrespondenz zu entwickeln.

Offen bleiben drei Fragen, umschrieben mit den Schlagwörtern
Fundamentaldefekt: Wann führen die angegebenen Schätzer zu p.s.d. Schätzungen der Streuungsmatrix?
Feedback (RAO 1970:170): Welche Eigenschaften haben Mittelwert-Schätzer mit einer geschätzten Streuungsmatrix?
Feedforward: Was passiert mit den Streuung-Schätzern, wenn man auf Invarianz verzichtet und geschätzte Mittelwerte benutzt?

* * *

Literaturverzeichnis

AITKEN 1935 A.C. AITKEN
On Least Squares and Linear Combination of Observations
Proc. Roy. Soc. Edinburgh 55, 1934-5, 42-48
Zitiert auf S. 4, 59, 65, 67

ALBERT 1972 Arthur ALBERT
Regression and the MOORE-PENROSE Pseudoinverse
Academic Press, New York 1972
Zitiert auf S. 4, 5, 17, 32, 33, 68

ALBERT 1973 Arthur ALBERT
The GAUSS-MARKOV Theorem for Regression Models with Possibly Singular Covariances
SIAM J. Appl. Math. 24, 1973, 182-187
Zitiert auf S. 68

ANSCOMBE 1961 F.J. ANSCOMBE
Examination of Residuals
Proc. Fourth Berkeley Symp. Math. Statist. Prob. 1, 1961, 1-36
Zitiert auf S. 57

BEUTLER 1965 Frederick J. BEUTLER
The Operator Theory of the Pseudoinverse I: Bounded Operators
J. Math. Anal. Appl. 10, 1965, 451-470
Zitiert auf S. 33

CHIPMAN 1964 John S. CHIPMAN
On Least Squares with Insufficient Observations
J. Amer. Statist. Assoc. 59, 1964, 1078-1111
Zitiert auf S. 68, 69

CLINE 1964 Randall E. CLINE
Representations for the Generalized Inverse of a Partioned Matrix
SIAM J. Appl. Math. 12, 1964, 588-600
Zitiert auf S. 36

CLINE 1965 Randall E. CLINE
Representations for the Generalized Inverse of Sums of Matrices
SIAM J. Numer. Anal. 2, 1965, 99-114
Zitiert auf S. 38

DRYGAS 1972 Hilmar DRYGAS
The Estimation of Residual Variance in Regression Analysis
Math. Operationsforsch. Statist. 3, 1972, 373-388
Zitiert auf S. 5, 44, 94, 101, 103

GAUSS Werke — Carl Friedrich GAUSS
Werke
Hrsg. Königliche Gesellschaft der Wissenschaften zu Göttingen
Band IV, Göttingen 1873
Band VII, Göttingen 1906
Band X,1, Göttingen 1917
Zitiert auf S. 57, 67

GAUSS 1887 — Carl Friedrich GAUSS
Abhandlungen zur Methode der kleinsten Quadrate
In deutscher Sprache hrsg. von A. BÖRSCH, P. SIMON
Stankiewicz, Berlin 1887
Zitiert auf S. 57, 67

GREUB 1967 — W.H. GREUB
Multilinear Algebra
Springer, Berlin 1967
Zitiert auf S. 18, 20, 21, 23, 24, 26, 28

HOERL & KENNARD 1970 — Arthur E. HOERL, Robert W. KENNARD
Ridge Regression: Biased Estimation for Nonorthogonal Problems
Technometrics 12, 1970, 55-67
Zitiert auf S. 72

HSU 1938 — P.L. HSU
On the Best Quadratic Estimate of Variance
Statist. Res. Memo. 2, 1938, 91-104
Zitiert auf S. 5, 94, 98

LAMOTTE 1973 — L.R. LAMOTTE
Quadratic Estimation of Variance Components
Biometrics 29, 1973, 311-330
Zitiert auf S. 5, 100

LANG 1966 — Serge LANG
Linear Algebra
Addison-Wesley, Reading Mass. 1966
Zitiert auf S. 19, 22

LEGENDRE 1805 — Adrien-Marie LEGENDRE
Nouvelles Méthodes pour la Determination des Orbites des Comètes
Paris 1805
Englische Übersetzung des Anhangs S. 72-75
Sur la Méthode des moindres quarrés
in SMITH 1959, 576-579
Zitiert auf S. 56, 67

LEWIS & ODELL 1966 — T.O. LEWIS, P.L. ODELL
A Generalization of the GAUSS-MARKOV Theorem
J. Amer. Statist. Assoc. 61, 1966, 1063-1066
Zitiert auf S. 65

LINDLEY & SMITH 1972 — D.V. LINDLEY, A.F.M. SMITH
BAYES Estimates for the Linear Model
J. Roy. Statist. Soc. Ser. B 34, 1972, 1-41
(1-18 paper, 18-41 discussion)
Zitiert auf S. 72

MARKOFF 1912 — A.A. MARKOFF
Wahrscheinlichkeitsrechnung
Teubner, Leipzig 1912
Zitiert auf S. 67

MARQUARDT 1970 — Donald V. MARQUARDT
Generalized Inverses, Ridge Regression, Biased Linear Estimation, and Nonlinear Estimation
Technometrics 12, 1970, 591-612
Zitiert auf S. 72

MILLER 1973 — Kenneth S. MILLER
Complex Linear Least Squares
SIAM Rev. 15, 1973, 706-726
Mit Bibliographie
Zitiert auf S. 50, 68

MOORE 1920 — E.H. MOORE
On the Reciprocal of the General Algebraic Matrix (Abstract)
Bull. Amer. Math. Soc. 26, 1920, 394-395
Zitiert auf S. 30

PENROSE 1955 — R. PENROSE
A Generalized Inverse for Matrices
Proc. Cambridge Philos. Soc. 51, 1955, 406-413
Zitiert auf S. 30, 32, 33, 34, 35, 41, 42

PENROSE 1956 — R. PENROSE
On Best Approximate Solutions of Linear Matrix Equations
Proc. Cambridge Philos. Soc. 52, 1956, 17-19
Zitiert auf S. 13

PLACKETT 1949 — R.L. PLACKETT
A Historical Note on the Method of Least Squares
Biometrika 36, 1949, 458-460
Zitiert auf S. 67

PLACKETT 1972 — R.L. PLACKETT
Studies in the History of Probability and Statistics XXIX: The Discovery of the Method of Least Squares
Biometrika 59, 1972, 239-251
Zitiert auf S. 67

RADO 1956 — R. RADO
Note on Generalized Inverses of Matrices
Proc. Cambridge Philos. Soc. 52, 1956, 600-601
Zitiert auf S. 30

RAO 1945 — C. Radhakrishna RAO
Generalization of MARKOFF's Theorem and Tests of Linear Hypothesis
Sankhyā 7, 1945, 9-15
Zitiert auf S. 68

RAO 1955 C. Radhakrishna RAO
Analysis of Dispersion for Multiply Classified Data with Unequal Numbers in Cells
Sankhyā 15, 1955, 253-280
Zitiert auf S. 68

RAO 1967 C. Radhakrishna RAO
Least Squares Theory Using an Estimated Dispersion Matrix and its Applications to Measurement of Signals
Proc. Fifth Berkeley Symp. Math. Statist. Prob.1, 1967, 355-372
Zitiert auf S. 62, 68

RAO 1970 C. Radhakrishna RAO
Estimation of Heteroscedastic Variances in Linear Models
J. Amer. Statist. Assoc. 65, 1970, 161-172
Zitiert auf S. 5, 82, 83, 84, 92, 103

RAO 1971a C. Radhakrishna RAO
Estimation of Variance and Covariance Components - MINQUE-Theory
J. Multivariate Anal. 1, 1971, 257-275
Zitiert auf S. 5, 83, 85, 86, 91, 92, 93

RAO 1971b C. Radhakrishna RAO
Minimum Variance Quadratic Unbiased Estimation of Variance Components
J. Multivariate Anal. 1, 1971, 445-456
Zitiert auf S. 5, 94

RAO 1971c C. Radhakrishna RAO
Unified Theory of Linear Estimation
Sankhyā Ser. A 33, 1971, 371-394
Corrigenda: Sankhyā Ser. A 34, 1972, 194 und 288
Zitiert auf S. 4, 50, 51, 65, 68, 71, 103

RAO 1972 C. Radhakrishna RAO
Estimation of Variance and Covariance Components in Linear Models
J. Amer. Statist. Assoc. 67, 1972, 112-115
Zitiert auf S. 5, 83, 85, 86, 91, 92

RAO 1973 C. Radhakrishna RAO
Representations of Best Linear Unbiased Estimators in the GAUSS-MARKOV Model with a Singular Dispersion Matrix
J. Multivariate Anal. 3, 1973, 276-292
Zitiert auf S. 51, 52

RAO & MITRA 1971 C. Radhakrishna RAO, Sujit Kumar MITRA
Generalized Inverse of Matrices and its Applications
Wiley, New York 1971
Mit Bibliographie
Zitiert auf S. 4, 5, 32, 50, 68

SEARLE 1971 — S.R. SEARLE
Topics in Variance Components Estimation
Biometrics 27, 1971, 1-79
Zitiert auf S. 47, 51

SEELY 1970 — Justus SEELY
Linear Spaces and Unbiased Estimation
Ann. Math. Statist. 4l, 1970, 1725-1734
Zitiert auf S. 47

SMITH 1959 — David Eugen SMITH
A Source Book in Mathematics, Vol. 2
Dover, New York 1959
Zitiert auf S. 56, 105

URQUHARDT, WEEKS HENDERSEN 1973 — N.S. URQUHARDT, D.L. WEEKS, C.R. HENDERSEN
Estimation Associated with Linear Models:
A Revisitation
Comm. Statist. 1, 1973, 303-330
Zitiert auf S. 47

WATSON 1972 — G.S.WATSON
Prediction and the Efficiency of Least Squares
Biometrika 59, 1972, 91-98
Zitiert auf S. 68

WITTING 1966 — Hermann WITTING
Mathematische Statistik
Teubner, Stuttgart 1966
Zitiert auf S. 47