Rainer Schnell

MISSING-DATA-PROBLEME IN DER EMPIRISCHEN SOZIALFORSCHUNG

Dissertation
zur Erlangung des akademischen Grades
eines Doktors der Sozialwissenschaft
der Ruhr-Universität Bochum

1986

INHALTSVERZEICHNIS

I. Einleitung 1

II. Eingrenzung des Themas 3

Anmerkungen 4

1. **Missing Data Muster** 5
1.1 Definitionen 5
1.1.1 "Missing at Random": MAR 5
1.1.2 "Missing at Random within Classes": MARC 7
1.1.3 "Censoring" und "Truncation" 8
1.1.4 Systematische MD-Muster 9
1.2 MAR als Element der klassischen Fehlermethodologie 10
1.3 Testmethoden zur Prüfung der MAR-Annahme 11
1.3.1 Verteilungsmuster fehlender Werte 13
1.3.2 Tests auf MAR 16
1.3.3 Empirische Ergebnisse zur Leistungsfähigkeit 20
1.4 Zusammenfassung 21

Anmerkungen zu Kapitel 1 22

2. **Ausfallursachen I: Datenerhebung** 24
2.1 Typologie fehlender Werte in der Datenerhebung 24
2.2 Ansätze einer allgemeinen Theorie des Interviews 25
2.2.1 Interviewbedingte MD 27
2.2.2 Befragtenbedingte MD 29
2.2.2.1 Elemente einer Erklärung der Befragtenreaktion 30
2.2.2.2 Rekonstruktion einer "WN"-Typologie 32
2.3 Empirische Ergebnisse 37
2.4 Konsequenzen in der Forschungspraxis 41

Anmerkung zu Kapitel 2 44

3. **Ausfallsursachen II: Datenerfassung und -editierung** 47
3.1 Codierfehler 47
3.2 Datenerfassung 48
3.3 Dateneditierung 49
3.3.1 Exkurs I: Multivariate "Outlier" und MD-Probleme 52
3.4 MD durch Datenverarbeitung: "Processing errors" 54
3.5 Zusammenfassung 55

Anmerkungen zu Kapitel 3 57

4. **Korrekturverfahren bei systematischen Ausfällen** 59
4.1 "Restriction of Range" - Korrekturen 59
4.1.1 Terminologie 60
4.1.2 Folgen der Selektion 60
4.1.3 Korrekturformeln 62
4.1.4 Probleme der Anwendung der Korrekturformeln 63
4.1.4.1 Arbeiten zur Verteilungsfunktion von R 63
4.1.4.2 Arbeiten zum Bias von R 65

4.1.5 Das Problem der unbekannten Selektionsverfahren 68
4.1.6 Lösungsversuche 69
4.1.7 Zusammenfassende Bewertung 70

4.2 Korrekturverfahren für Regressionen..... 71
4.2.1 Auswirkung der Selektion auf Regressionsparameter... 72
4.2.2 Korrekturverfahren..... 76
4.2.3 Probleme der Korrekturverfahren..... 77
4.2.4 Zusammenfassung..... 78

Anmerkungen zu Kap.4..... 80

5. **Missing data Techniken**..... 83
5.1 Parameter-Schätzverfahren..... 84
5.1.1 Berechnugsmethoden von Korrelationsmatrizen bei MD.. 84
5.1.2 Techniken der singulären Korrelationsmatrizen..... 86
5.1.3 Maximum-Likelihood-Schätzungen..... 87
5.1.3.1 ML-Schätzer f}r den bi- und trivariaten Fall..... 88
5.1.3.2 ML-Schätzer f}r Panel-Daten..... 88
5.1.3.3 EM-Algorithmus..... 89
5.2 MD-Ersetzungstechniken..... 92
5.2.1 Nicht-informative Ersetzungen..... 95
5.2.1.1 Ersetzung durch "Expertenratings"..... 96
5.2.1.2 Ersetzung durch den Mittelwert..... 96
5.2.2 Informative Ersetzungstechniken..... 97
5.2.2.1 Das Konzept der Imputationsklassen..... 97
5.2.2.1.1 Die Wahl der Imputationsklassen..... 98
5.2.2.1.2 Exkurs II: MD in der Clusteranalyse..... 99
5.2.2.1.3 Bestimmung der Imputationsklassen mit AID..... 102
5.2.2.1.4 Kritik des Imputationsklassen-Modells..... 104
5.2.2.2 Ersetzung durch den Gruppenmittelwert..... 106
5.2.2.3 Ratio-Schätzung..... 107
5.2.2.4 Zellen-Ratio-Schätzung: "Array"-Prozedur..... 107
5.2.2.5 Cold-deck-Techniken..... 108
5.2.2.6 Hot-deck-Techniken..... 109
5.2.2.7 Spezielle Hot-Deck-Techniken..... 112
5.2.2.8 Exkurs III: "Zero-Spike"-Methoden..... 113
5.3 Multivariate Ersetzungstechniken..... 115
5.3.1 Hauptkomponenten-Verfahren: Methode nach DEAR..... 115
5.3.2 Regressionsschätzverfahren: Methode nach BUCK..... 116
5.3.3 Weitere Regressionsschätzverfahren..... 117
5.3.4 Spezielle Modifikation..... 118
5.4 Exkurs IV: Missing Data in Skalen und Indizes..... 120
5.5 Exkurs V : Missing Data auf normalem Meßniveau..... 123
5.6 Zusammenfassende Bewertung..... 126

Anmerkungen zu Kapitel 5..... 128

6. **Effizienz der MD-Techniken bei der Kovarianzschätzung**.. 138
6.1 Analytische Ergebnisse..... 139
6.2 Simulationsstudien..... 141
6.2.1 Arbeiten unter der MAR-Annahme..... 142
6.2.2 Systematische Ausfälle..... 150
6.3 Zusammenfassung..... 153

Anmerkungen zu Kapitel 6..... 154

7. **Design der Simulationsstufe**..... 158
7.1 Ziel der Studie..... 158
7.2 Prinzip der Simulation..... 159
7.3 Beschreibung der simulierten Datensätze..... 160
7.4 Korrelationsstruktur..... 162

7.5 Ausfallmodelle..... 164
7.5.1 Modell 1: Missing at Random (MAR)..... 165
7.5.2 Modell 2: Hazard-Modell (MAT)..... 165
7.5.3 Modell 3: Zwei-Sample-Modell (MARC)..... 166
7.6 Missing-Data-Verfahren..... 167
7.7 Abhängige Variablen..... 169
7.7.1 Maße der Güte der Schätzung der Korrelationen..... 169
7.7.2 Maße der Güte der Schätzung fehlender Werte..... 171
7.8 Zusammenfassung des Designs..... 173

Anmerkungen zu Kapitel 7..... 174

8. **Auswertung der Simulation**..... 177
8.1 Randauszählung und Transformationen..... 177
8.2 Inerkorrelation der abhängingen Variablen..... 180
8.3 Ähnlichkeit der Reaktionskurve der Verfahren..... 183
8.4 Einflüsse der experimentellen Faktoren..... 196
8.4.1 Ergebnisse der Varianzanalysen..... 196
8.4.2 Ergebnisse der Regressionen..... 200
8.5 Relative Effiziens der Verfahren..... 210
8.6 Genauigkeit der Schätzungen..... 211
8.7 Zusammenfassung..... 213

Anmerkungen zu Kapitel 8..... 218

9. **Multiple Imputation (MI)**..... 220
9.1 MI unter einem Ausfallmechanismus..... 221
9.2 MI unter verschiedenen Ausfallmechanismen..... 222
9.3 Sensitivitätsuntersuchungen mit MI..... 225
9.3.1 Praktische Probleme bei der Durchführung..... 226
9.3.2 Abgrenzung und Konsequenzen in der Forschungspraxis.. 227

Anmerkungen zu Kapitel 9..... 229

Anhang..... 230

A. **Technische Details der Simulation**..... 230
A.1 Generierung der Datensätze..... 230
A.2 Iterationszahl..... 231
A.3 Einzelheiten des Programms..... 231
A.4 Durchführung..... 233
A.5 Auswertung..... 233

B. **Programmlisting**..... 234

C. **Vortests zweier Hot-Deck Verfahren**..... 247

D. **Tabellen zur relativen Effiziens der Verfahren**..... 249

E. **Realisierung der MD-Techniken mit SPSS-X**..... 255
E.1 MD-Ersetzungstechniken..... 256
E.2 Analyse von MD-Mustern..... 259
E.3 Hazard-Modelle..... 260

F. **Literaturverzeichnis**..... 263

VERZEICHNIS DER ABBILDUNGEN

8-1 Plot der Verteilung von WRDEV..... 179
8-2 Plot der Verteilung nach der LN-Transformation (LRSDEV) 179
8-3 Plot: Mittlere Abweichung der geschätzten Daten..... 185
8-4 Plot: Korrelationsabweichungsindex (ZMKS):MAR-Modell... 187
8-5 Plot: Korrelationsabweichungsindex (ZMKS):Hazard-Modell 188
8-6 Plot: Korrelationsabweichungsindex (ZMKS):MARC-Modell.. 189
8-7 Dendrogramm: ZMKS-MAR-Modell..... 191
8-8 Dendrogramm: ZMKS-Hazard-Modell..... 191
8-9 Dendrogramm: ZMKS-MARC-Modell..... 192
8-10 Dendrogramm: ZMKS - Total..... 192
8-11 Dendrogramm: HIT-MAR-Modell..... 194
8-12 Dendrogramm: HIT-Hazard-Modell..... 194
8-13 Dendrogramm: HIT-MARC-Modell..... 195
8-14 Dendrogramm: HIT-Total..... 195

E-1 Druckausgabe eines MD-Pattern mit SPSS-X..... 260

VERZEICHNIS DER TABELLEN

7-1 Vorgegebene Populationskorrelationsstruktur..... 163
8-1 Randauszählung der abhängigen Variablen..... 178
8-2 Korrelationsmatrix (EMD EMAD ERSD HIT)..... 180
8-3 Ergebnis der Faktoranalyse..... 180
8-4 Korrelationsmatrix (ZMKS LRSDEV RDEV)..... 181
8-5 Ergebnis der Faktoranalsye..... 181
8-6 Vollständige Korrelationsmatrix (Verfahren 5 - 10)..... 182
8-7 Ergebnis der Faktoranalyse..... 182
8-8 Patternmatrix der Faktoranalyse..... 198
8-9 Zusammenfassung der Varianzanalysen..... 201
8-10 Regressionen ZMKS: MAR..... 201
8-11 Regressionen ZMKS: Hazard-Modell..... 202
8-12 Regressionen ZMKS: MARC..... 202
8-13 Regressionen LRSDEV: MAR..... 203
8-14 Regressionen LRSDEV: Hazard-Modell..... 203
8-15 Regressionen LRSDEV: MARC..... 203
8-16 Regressionen RDEV: MAR..... 204
8-17 Regressionen RDEV: Hazard-Modell..... 204
8-18 Regressionen RDEV: MARC..... 205
8-19 Regressionen HIT : MAR..... 207
8-20 Regressionen HIT : Hazard-Modell..... 207
8-21 Regressionen HIT : MARC..... 208
8-22 Regressionen ERSD: MAR..... 208
8-23 Regressionen ERSD: Hazard-Modell..... 208
8-24 Regressionen ERSD: MARC..... 209
8-25 Korrelationsmatrizen mit mittleren Abweichungen > 10... 217

I. Einleitung

Kaum ein Datensatz im Rahmen sozialwissenschaftlicher Forschung enthält alle Daten, die zu erheben beabsichtigt war. Das Problem unvollständiger Datensätze findet im Gegensatz zu seiner Verbreitung in der Forschungspraxis allerdings nur einen geringen Niederschlag in der Lehrbuchliteratur und in noch schwächerem Ausmaß in veröffentlichten Forschungsberichten. Zwar gehört mittlerweile die zumindst formale Beachtung des damit verwandten Problems der Totalausfälle in Samplesurveys zum Standard bei der Beurteilung empririscher Untersuchungen, für "partiell" fehlende Daten ("missing data": MD) gilt dies jedoch nicht. In der Forschungspraxis wird das Vorliegen eines ernsthaften MD-Problems häzfig erst dann bemerkt, wenn ein Statistikprogrammpaket auf Grund seiner Voreinstellungen[1] an einem Datensatz scheitert: Clusteranalysen bei Datensätzen mit mehreren hundert Befragten und einem dutzend Variablen werden mit der Melden "no cases remained" abgebrochen[2] multiple Regressionen werden mit "nahezu singulären" Korrelationsmatrizen gerechnet[3] oder die Programme brechen mit rätselhaften Fehlermeldungen den Lauf ab[4]. Zumeist wird dann zu ad-hoc-Lösungen gegriffen, die weder auf statistischen Modellen beruhen noch durch eine soziologische "Theorie" gestützt werden. Darüber hinaus wird das Problem durch eine uneinheitliche Terminologie und fragwürdige "Selbstverständlichkeiten" (wie beispielsweise der Wertung der "weiß-nicht"-Antwort als fehlender Wert) eher verdunkelt.

Diese Arbeit versucht einen Überblick über die vorhandene Literatur[5] zur Entstehung fehlender Werte, deren Implikationen für Datenanalysen und die Methoden zu ihrer Handhabung zu geben. Nach den notwendigen terminologischen Klärungen und der Definition grundlegender Begriffe werden im ersten Kapitel die statistischen Verfahren zur Überprüfung der Annahme, daß Daten "zufällig" fehlen ("missing at random": MAR), dargestellt und diskutiert. Die MAR-Annahme liegt fast allen statistischen "Korrekturverfahren" für fehlende Werte (einschließlich dem Ignorieren des Problems) zugrunde. Die beiden folgenden Kapitel zeigen hingegen mögliche systematische Ausfallmechanismen im Rahmen empirischer Sozialforschung. Zunächst werden

die "sozial Regelmäßigkeiten" bei der Entstehung fehlender Werte im Befragungsprozeß dargestellt und anhand ausgewählter empirischer Arbeiten die generelle Unhaltbarkeit der Annahme des "zufälligen" Fehlens aufgezeigt. Der zweite Ursachenkomplex fehlender Werte - Datenerfassung und Dateneditierung - wird im dritten Kapitel ebenfalls mit dieser Schwerpunktsetzung erörtert. Das vierte Kapitel stellt die in der Literatur zu findenden Ansatzpunkte statistischer Korrekturverfahren für systematisch fehlende Werte dar. Hierzu werden die Koeffizienten bei Beschränkungen des Wertebereichs durch Selektionseffekte ("restriction of range".Korekturen) diskutiert. Diese, in der Regel isoliert als Sezialproblem angesehenen, Korrekturen finden in den neueren Arbeiten zur Korrektur von Regressionen unter einem bestimmten Ausfallmechanismus ("sample selection bias") einen allgemeineren Bezugsrahmen. Der zweite Teil des Kapitels stellt einige Ansätze dieser Korrekturverfahren für Regressionen dar. Das fünfte Kapitel wendet sich Kompensations- und Ersetzungstechniken für fehlende Werte zu. Die Brauchbarkeit und Leistungsfähigkeit der Ersetzungsverfahren insbesondere in Hinsicht auf sozialwissenschaftliche Forschungen wird im sechsten Kapitel beurteilt. Da die vorliegenden Arbeiten hierzu nicht ausreichend erschienen, wurde eine Computersimulation "sozialwissenschaftlicher Datensätze" mit unterschiedlichen Ausfallmodellen durchgeführt, in der zehn Ersetzungsverfahren getestet wurden. Die Beschreibung des Designs der Simulation ist Gegenstand des Kapitel 7, die Auswertung wird in Kapitel 8 dargestellt. Das Kapitel 9 enthält eine Erörterung der als Sensitivitätstests gegenüber verschiedenen Annahmen über den Ausfallmechanismus nutzbaren mehrfachen Ersetzungen fehlender Werte ("multiple imputations": MI). die Anhänge stellen die technischen einzelheiten der Simulation und die Durchführung einiger MD-Techniken mit SPSS-X dar.

II. Eingrenzung des Themas

In der Literatur zum Problem unvollständiger Datenmatrizen finden sich eine Reihe verschiedener Bezeichnungen für das Problem: missing data, incomplete data, fragmentary data, partial data, omitted data, spoilt data, missing values, missing units, itemomits etc. Weiterhin deutet die Unterschiedlichkeit der Stichworte, unter denen Arbeiten zum Thema erscheinen (u.a. itemnonresponse, nonresponse, response-errors, undecided repomdents, don't -know-research sowie nichtsubstantielle Antworten6 die unterschieldichen Kontexte an, in denen MD-Probleme diskutiert werden.
Ebenso unterschiedlich sind die Sachverhalte, für die der Begriff "missing data" gebraucht wird. In der Literatur finden sich Beispiele für die Benutzung des Begriffs für "intentional" fehlende Werte, also für Daten, die nicht erhoben werden sollten. Derartige Beispiele sind Fragen, die innerhalb eines Surveys Teilen des Samples nicht vorgelegt wurden, fehlende Daten der "ausgerollten" Sampleteile in Rotationssamples für einen Erhebungszeitpunkt und nicht-realisierte Zellen eines varianzanalytischen Untersuchungsplans[7]. Ebenso finden sich Probleme mit unvollständigen Listen der zu untersuchenden Population, Panelsterblichkeitsprobleme und Probleme der Beschränkung des Wertebereichs (im Folgenden stets: "Restriction of Range-Probleme") unter der Bezeichnung "missing data"[8].

Gegenstand dieser Arbeit sind hingegen fehlende Werte, soweit sie nicht Intention des Untersuchungsplans sind. Da empirische Sozialforschung hier als Überprüfungsmethode sozialwissenschaftlicher Theorien aufgefaßt wird, beschränkt sich die Arbeit auf die potentiellen Einflüsse fehlender Werte auf die analyse von Variablenbeziehungen[9]. Der Schwerpunkt der Arbeit liegt bei den Auswirkungen fehlender Werte auf Korrelationsmatrizen als dem Ausgangspunkt der meisten multivariaten Verfahren.

Anmerkungen

1. "Bei zu hohem Anteil fehlender Werte können (...) allerdings Probleme auftreten, die zum Programmabbruch führen." (BOLLINGER u.a 1983:33).

2. Die BMDP-Clusteranalysen besitzen die Löschung aller Fälle mit fehlenden Werten als Voreinstellung.

3. SPSS rechnet nach einer einzeiligen Warnung, daß die Korrelationsmatrix "nearly singular" sei, auch noch mit Determinanten, die um 10^{-14} liegen.

4. Die CYBER-175 Version von BMDP - die nicht mit "double precision" arbeitet- bricht mit einem "underflow error" ab und erstellt vor dem Zusammenbruch noch einen vollständigen CPU-Dump.

5. Da nur sehr wenige deutschprachige Arbeiten zum Thema existieren, sich andererseits aber insbesondere in Kanada und in den USA in den letzten Jahren eine umfangreiche Spezialliteratur entwickelt hat, ist eine Vorbemerkung zur Terminologie erforderlich: Um Begriffe wie z.B. "Dopplungsverfahrenzuweisungsklasse" für "imputationclass" zu vermeiden, wird eine Reihe von Begriffen nicht in der deutschen Übersetzung, sondern im Original gebraucht.

6. Vgl. SHIH (1981:1.1), ferner KALSBEK (1980:131).

7. Vgl. RUBIN (1976:581) und RUBIN (1972:136).

8. z.B. MAYS (1976).

9. Die Literatur zu Missing-Data-Problemen ist nahezu unüberschaubar geworden, soweit es sich um Ausfälle unter dem MAR- Modell handelt. Insbesondere zu Regressionen und Varianzanalysen ("missing plots") mit MD existiert eine umfangreiche Literatur. Aspekte des MD-Problems im Zusammenhang mit Varianzanalysen werden in dieser Arbeit nicht behandelt, die Literatur zu Regressionen nur teilweise. Ausführliche Übersichten finden sich bei AFIFI/ELASHOF (1966), HAMILTON (1975) und ANDERSON/BASILEVSKY/ HUM (1983), speziell zu MANOVA vgl. ANDERSON (1976). Arbeiten zur Faktorenanalyse finden sich bei WOODBURY/SILER (1966), DE LIGNY u.a. (1981), FINKBEINER (1979) und BROWN (1981); zu Log-Linearen Modellen FUCHS (1982), LITTLE (1983); zur Diskriminanzanalyse vgl. JACKSON (1968), CHAN/DUNN (1972), BOHANNON (1976a,1976b) und vor allem NORDHEIM (1978); zur Zeitreihenanalyse vgl. DUNSMUIR (1981) und PALM/NIJMAN (1983). Arbeiten zu sehr speziellen Problemen stellen z.B. die Abhandlungen von BROWN (1981) zu soziometrischen Daten und SGLE (1967) zu t-Tests dar. Bibliographien bieten HAMILTON (1975) und DALENIUS (1977) sowie ANDERSON/ BASILEVSKY/HUM (1983), BOGESTRÖM/LARSSON/LYBERG (1983) und KALTON (1983).

1. Missing Data Muster

In diesem Kapitel werden zunächst die Definitionen einiger Ausfallmechanismen vorgestellt. Daran schließt sich die Beschreibung und Kritik der Verfahren an, die in der anwendungsorientierten Literatur zur Untersuchung der Ausfallmechanismen in einem vorliegenden Datensatz vorgeschlagen werden. Insbesondere werden die "Tests auf Zufälligkeit von Missing Data" ausführlich erläutert.

1.1 Definitionen

Von zentraler Bedeutung für die Analyse fehlender Werte ist die Frage nach der Art des Ausfalls. Sie zielt auf den die fehlenden Werte erzeugenden Prozeß (im folgenden: "missing data generating process": MDGP). Der einfachste Fall ist dann gegeben, wenn zwischen dem fehlende Werte generierenden Prozeß und den Werten der Untersuchungseinheiten keinerlei Zusammenhang besteht: Die Daten fehlen dann zufällig ("missing at random": MAR). Die Diskussion um diesen Ausfallmechanismus und dessen Definition nimmt den überwiegenden Teil der Literatur ein. Andere Ausfallmechanismen werden in der Literatur nur sehr selten erörtert; dementsprechend existiert bisher keine Systematik bei der Definition anderer Ausfallmechanismen.

1.1.1 "Missing at Random": MAR

ANDERSON/BASILEVSKY/HUM (1983:416) unterscheiden in der Literatur vier verschiedene Definitionen für MAR :

Typ 1: Die MD werden als MAR bezeichnet, wenn das "Muster" der MD "gestreut" scheint, also keine Konzentration fehlender Werte in einem Teil der dann räumlich gedachten Datenmatrix vorhanden ist. Muster "räumlicher" Konzentration entstehen überwiegend durch das Forschungsdesign.

Typ 2: Es sei W(i,j) eine n*p Indikatorenmatrix, wobei W(i,j)=1, falls für den Fall i auf der Variablen j gültige Werte vorhanden sind; und W(i,j)=0, falls für den Fall i auf der Variablen j keine gültigen Werte vorhanden sind. Die MD werden dann als MAR bezeichnet, wenn für jede Menge aus W die Wahrscheinlichkeit für das gemeinsame Fehlen gleich dem Produkt der einzelnen Wahrscheinlichkeiten ist:

1-1. E(W(i,1),W(i,2)...W(i,p)) = E(W(i,1)) E(W(i,2))...E(W(i,p))

MAR in diesem Sinn impliziert also die Unabhängigkeit des Fehlens einer Variablen von jeder anderen Variablen für ein gegebenes Individuum.

Typ 3: Falls die fehlenden Daten für eine gegebene Variable Y dieselbe Verteilung besitzen wie der gesamte Datenvektor, wird von MAR gesprochen: Bezeichnet R einen Responseindikator, so gilt MAR falls p(R Y)= p(R) für beliebiges Y[1/]. Diese Definition bezieht sich auf den Wertebereich von Y und schließt Zusammenhänge zwischen dem Ausfallmechanismus und dem Wertebereich aus. Da Daten, für die MAR nach Typ 3 gegeben ist, als Zufallsstichprobe aus einer Zufallsstichprobe aufgefaßt werden können, bleiben alle Eigenschaften der Parameterschätzer erhalten.

Typ 4 : Falls zwischen den fehlenden Werten zweier Vektoren kein systematischer Zusammenhang besteht, wird für diese zwei Vektoren von MAR gesprochen.

Diese vier Typen werden von ANDERSON/BASILEVSKY/HUM(1983:417) als "logical independent" in dem Sinn bezeichnet, daß kein Typ einen anderen impliziert. Läßt man den Typ 1 außer Betracht (es handelt sich überwiegend um intentional fehlende Werte) kann empirisch in der Regel nicht von der Unabhängigkeit ausgegangen werden.

Exakte mathematische Definitionen von "missing at random" und "observed at random" hat RUBIN(1976) vorgelegt. Nach der umgangssprachlichen Fassung

dieser Definitionen (LITTLE/RUBIN 1983:216) wird MAR angenommen, wenn der datengenerierende Prozeß nicht von den interessierenden Variablen abhängt. Genauer: Wenn die Wahrscheinlichkeit für fehlende Werte nicht von den "wahren" Werten abhängt, handelt es sich um MAR, und die beobachteten Werte sind "observed at random". Hängt der datengenerierende Prozeß nicht von den "wahren" Werten der MD ab, sondern von den beobachteten Werten, handelt es sich um MAR, aber die beobachteten Daten sind nicht "observed at random". Falls der Prozeß von den Werten der MD abhängt, handelt es sich um "nonrandomly missing data"[2/]. MARINI/OLSEN/RUBIN (1980:316) bezeichnen das gemeinsame Vorliegen von "missing at random" und "observed at random" als "missing completely at random".

Im Rahmen dieser Arbeit wird "missing at random" im Sinne der Definition des Typs 2 von ANDERSON/BASILEVSKY/HUM benutzt, falls nicht explizit auf andere Definitionen Bezug genommen wird. Die Definition des Typs 2 bei ANDERSON/BASILEVSKY/HUM stammt von GLASSER (1964:834) und hat sich in der Literatur weitgehend durchgesetzt[3/].

1.1.2 "Missing at random within classes": MARC

Die Definition nach Typ 3 kann so modifiziert werden, daß ein realistischeres Modell als MAR entsteht: Für eine gegebene Variable werden für unterschiedliche Subpopulationen unterschiedliche Ausfallwahrscheinlichkeiten angenommen, wobei innerhalb der Subpopulationen MAR angenommen wird. Dieses Modell wird als "missing at random within classes" (MARC) bezeichnet (SANTOS 1981:17). Daten werden in bezug auf eine bestimmte Variable Y als MARC be- zeichnet, wenn die Subpopulation mit vorhandenen Daten der Klasse h (h=1,..H) eine Zufallsstichprobe der Größe R(h) (wobei R(h)=R) der Totalpopulation der Größe N(h) dieser Klasse darstellt.

Falls die MARC-Annahme korrekt ist (und die Klassen ohne Fehler identifiziert wurden), gilt die oben für MAR erwähnte Konstanz der Schätzereigenschaften

innerhalb der Klassen. Ist die Klassifizierungsvariable die Variable Y selbst, ermöglicht das MARC Modell die Modellierung von Ausfallprozessen, bei denen die Ausfallwahrscheinlichkeit eine Funktion des Wertebereichs der Variablen ist ("Censoring"-Problem).

1.1.3 "Censoring" und "Truncation"

Ein Sample wird als "censored" bezeichnet, wenn der Wert mindestens einer interessierenden Variablen für einige Fälle nicht beobachtet wurde und die Ausfallwahrscheinlichkeit mit dem tatsächlichem Wertebereich der Variablen zusammenhängt. Der Unterschied zu fehlenden Werten liegt darin, daß Informationen über die relative Größe der "censored observations" in bezug auf die vorhandenen Werte vorliegen (MC COOL 1982:389-390). Betrachtet man die ihrer Größe gemäß geordneten Werte x(1), x(2)...x(n) einer Variablen, ergeben sich folgende drei häufigste Formen des Censoring (MC COOL 1982:396):

- links ("from below") : Wegfall der r kleinsten Werte :
  x(1)...x(r);
- rechts ("from above"): Wegfall der r größten Werte :
  x(n-r+1)...x(n);
- symmetrisch : Wegfall der jeweils r größten und
  r kleinsten Werte:
  x(1)...x(r),x(n-r+1)...x(n).

Der Unterschied zwischen "truncation" und "censoring" liegt darin, daß bei "truncation" Fälle mit Werten oberhalb oder unterhalb eines Schwellenwertes nicht zur Kenntnis gelangen: Im Gegensatz zum "censoring" können sie nicht gezählt werden (MOSES 1972:196).

1.1.4 Systematische MD-Muster

Da der weitaus größte Teil der Literatur auf MAR beschränkt ist, existieren nur vereinzelte Definitionen nicht zufälliger Ausfallmuster. HILL/DIXON (1981:57) unterscheiden in ihrer Simulationsstudie "truncated" und "related"-MD-Muster. Ein Ausfallmechanismus besitzt ein "truncated"-Muster dann, wenn ein Wert größer als der Populationsmittelwert auf der Variablen x mit der Wahrscheinlichkeit p fehlt. Als "related" wird ein Muster bezeichnet, bei dem ein Wert auf der Variablen x(i) mit der Wahrscheinlichkeit p fehlt, wenn die Variable x(j) den Populationsmittelwert überschreitet[4/].

NORDHEIM (1978b:36) bezeichnet als "probit missingness" den Fall, daß ein Wert einer MD-behafteten Variablen x mit der Wahrscheinlichkeit

1-2 $$p(x) = F((x-1)/t) \quad ; \quad -\infty < 1 < \infty \quad , \quad 0 \leq t$$

ausfällt, wobei F die kumulative Fläche der Standardnormalverteilung darstellt. Ist t = 0, so handelt es sich um "censoring from above" am Punkt x = l. Geht t -> oo, so hängt die Wahrscheinlichkeit des Ausfalls nicht mehr von x ab: In diesem Fall entspricht "probit missingness" dem MAR-Modell.

NORDHEIM (1978:58) definiert daneben noch "logistic missingness", bei der ein Datum mit der Wahrscheinlichkeit

1-3 $$p(x) = 1/(1+\exp(a+b\cdot x))$$

ausfällt; "logistic missingness" entspricht bis auf die rechnerisch einfachere logistische Verteilung der "probit missingness".

1.2 MAR als Element der klassischen Fehlermethodologie

Die intensive analytische Berücksichtigung von MD-Problemen gehört bislang nicht zu den Standardverfahren bei der Datenanalyse, wie sich an der hohen Zahl rein deskriptiver Arbeiten zum Zusammenhang demographischer Variablen und MD zeigen läßt (z.B. LEVERKUS-BRÜNING 1966, MAURANA 1976). Entsprechend finden sich in der Lehrbuchliteratur nur vereinzelt Hinweise auf potentielle MD-Probleme; selbst im Zusammenhang mit der Anwendung multivariater Verfahren sind dort keine Erörterung von MD-Problemen (z.B. COOLEY/LOHNES 1971, OVERALL/KLETT 1972, GAENSSLEN/ SCHUBÖ 1976, FAHRMEIR/HAMERLE 1984, DILLON/GOLDSTEIN 1984) oder nur sehr kurze Hinweise (z.B. BORTZ 1977) enthalten.

In der Forschungspraxis wird in der Regel nach einer ersten Überprüfung der Randauszählung erst dann einem potentiellen MD-Problem nachgegangen, wenn eine Standardprozedur unerwartete Resultate zeigt, z.B. negative Eigenwerte einer Korrelationsmatrix. Die Hauptursache dieser Herangehensweise an das MD-Problem dürfte in den hohen subjektiven Kosten einer erweiterten Analyse zu suchen sein. So betonen MARINI/OLSEN/RUBIN (1980:317) sehr deutlich den Kostenaspekt des "Mehraufwandes", den die Verwerfung der MAR-Annahme nach sich zieht: Diese Entscheidung erfordert die explizite Modellierung des Ausfallmechanismus.

Der erwartete Nutzen einer MD-Analyse erscheint hingegen im allgemeinen gering; die gängigen Auswertungs- und Analyseverfahren werden für robust genug gehalten, um nach den ohnehin üblichen Verstößen gegen andere Modellannahmen (vgl. KRIZ 1981) auch noch die Verletzung der MAR-Annahme zu überstehen: Bei nur geringen Ausfallraten pro Variable wird keine Veränderung substantieller Beziehungen durch systematische Ausfälle erwartet.

1.3 Testmethoden zur Prüfung der MAR-Annahme

> "Although the temptation to assume randomness about missing data is nearly overwhelming, probably the safest thing to do is to test it."
>
> TABACHNIK/FIDELL(1983:69)

Testverfahren auf "Zufälligkeit" des Auftretens fehlender Werte richten sich auf den Zusammenhang zwischen dem Auftreten von MD und den jeweils interessierenden Variablen der inhaltlichen Fragestellung ("Kerntheorie"). Mit den in einem unvollständigen Sample enthaltenen Daten lassen sich nur Teilaspekte der strengen MAR-Annahme ("missing completely at random") überprüfen. Diese prinzipiellen Grenzen der Tests sollen an den vier MAR-Typen erläutert werden.

Die MAR-Annahme in bezug auf die Abwesenheit stabiler MD-Muster (Typ 1 MAR) ist zwar mit den vorhandenen Daten prüfbar, aber MAR- Typ 1 erlaubt keine Aussagen über den Zusammenhang des Auftretens von MD mit Variablen der Kerntheorie.

Die MAR-Annahme des Typs 2 (Erwartungswert für das gemeinsame Auftreten von MD ist gleich dem Produkt der Einzelwahrscheinlichkeiten) ist mit den vorhandenen Daten prüfbar. Die Definition und die Durchführung der Tests zum MAR Typ 2 schließen (prinzipiell) jegliche stochastische Abhängigkeit der MD voneinander aus. Damit ist zwar im Sinne der Definition RUBIN's gewährleistet, daß die Daten "zufällig fehlen", da aber die Tests notwendigerweise auf die Indikatormatrix beschränkt bleiben müssen, ist es immer noch möglich, daß die gesamte Stichprobe verzerrt ist: Die Daten könnten "nicht zufällig beobachtet" worden sein (Verletzung der Annahme des "observed at random"). In diesem Fall wären die Daten "nicht vollständig zufällig fehlend". Dieses Problem betrifft allerdings nur Stichproben mit extremer Selektion und wird daher hier nicht weiter verfolgt.

Die MAR-Annahme des Typs 2 erlaubt aber, daß zwischen dem Fehlen einer Variablen und ihrem Wertebereich oder dem Wertebereich einer anderen Variablen ein Zusammenhang besteht: Die Tests für MAR-Typ 2 reagieren auf "censoring" - Probleme nur dann, wenn es sich um multivariate "censoring"-Probleme handelt. Das am häufigsten genannte Beispiel für einen systematischen Ausfallmechanismus zeigt die Wichtigkeit dieser Einschränkung des Tests: Das vermutete Steigen der Verweigerungen mit der Höhe des Einkommens kann durch Tests auf MAR-Typ 2 nicht ausgeschlossen werden.

Die MAR-Annahmen des Typs 3 (Fälle mit MD sind eine Zufallsstichprobe aller Fälle) und des Typs 4 (kein systematischer Zusammenhang zwischen den fehlenden Werten zweier Vektoren) sind mit den vorhandenen Daten nicht prüfbar: Da die Daten der Fälle mit MD unbekannt sind, kann über deren Verteilung nichts gesagt werden. Das von RUBIN (1980:4-5) erwähnte Beispiel einer Stichprobe, deren Nonrespondenten ausschließlich Werte oberhalb des Mittelwertes einer Variablen und deren Respondenten ausschließlich Werte unterhalb des Mittelwertes besitzen, zeigt, daß die vorhandenen Daten keine Entscheidung über die MAR-Annahme erlauben, solange nicht "heroische Annahmen"[5/]über die Nonrespondenten gemacht werden. Das Argument läßt sich problemlos auf Itemnonresponse übertragen: Ohne die Voraussetzung der Konstanz der Variablenbeziehungen "externer", zur Prüfung herangezogener Variablen bei Fällen mit und ohne MD kann kein Test erfolgen.

Es dürfte deutlich geworden sein, daß Tests auf MAR jeweils nur einen Aspekt der potentiellen Ausfallmechanismen erfassen können. Grobe Verstöße gegen die MAR-Annahme dürften durch Tests auf Typ2 und Typ3 zu entdecken sein. Die Untestbarkeit der strengen MAR- Annahme zeigt aber, daß negativ verlaufende Tests lediglich eine notwendige Bedingung für die Akzeptanz der MAR-Annahme sein können.

In der sozialwissenschaftlichen Forschungspraxis sind allerdings weniger "falsch negativ" verlaufende Tests das Problem, als die hohe Wahrscheinlichkeit, schon durch sehr einfache Tests grobe Verstöße gegen die MAR-Annahme zu entdecken.

Damit verschiebt sich die eigentlich interessierende Frage: Es ist weniger von Interesse, ob die Daten nicht-zufällig fehlen, sondern ob der Effekt der nicht-zufällig fehlenden Daten stark genug ist, um die Schlußfolgerungen über Variablenbeziehungen zu verändern. Ein signifikantes Ergebnis eines MAR-Tests bedeutet noch nicht in jedem Fall, daß die substantiellen Ergebnisse völlig in Frage gestellt werden müssen[6/].

Die Methoden zur Untersuchung des möglichen Einflusses der MD auf die Ergebnisse der Analysen zur "Kerntheorie" sind Gegenstand des Kapitels 9, hier werden zunächst lediglich die in der Literatur weit gestreut zu findenden und bisher noch nicht systematisch dargestellten Tests auf MAR beschrieben.

1.3.1 Verteilungsmuster fehlender Werte

> "If BMDPAM is available to you, you can learn more about your missing data than you might ever want to know."
>
> TABACHNIK/FIDELL(1983:69)

Der erste Schritt bei einer MD-Analyse besteht in der Regel in der Ermittlung des Anteils fehlender Werte im Datensatz. Sofern die "missing values" korrekt spezifiziert wurden, liegt dieses Ergebnis mit der Randauszählung bei allen Statistikstandardpaketen vor. Für die Überprüfung der MAR-Annahme nach der Definition des Typs 1 sind insbesondere "graphische" Methoden nützlich. Mit dem BMDP-Programm PAM (DIXON u.a. 1981) steht ein Programm zur Verfügung, daß den graphischen Ausdruck der MD-Indikatorenmatrix in verfeinerter Form ermöglicht. Neben den Indikatoren der fehlenden Werte werden Werte außerhalb des spezifizierten Wertebereichs der jeweiligen Variablen im Plot mit speziellen Symbolen ausgewiesen. PAM erlaubt maximal 100 Variablen in diesem "cases by variables plot" darzustellen. Bei hohen Fall- oder Variablenzahlen wird der Output allerdings sehr schnell uninterpretierbar, da das Programm selbst keine Ordnungsmöglichkeit der Patterns bietet. Der Datensatz kann natürlich entsprechend einer beliebigen Variablen geordnet werden. Eine

interessante Möglichkeit ist die Berechnung einer Pattern-Indikatorvariablen mit einer externen Prozedur und anschließender Sortierung nach dieser Variablen. Eine Möglichkeit hierzu besteht aus zwei Schritten:

1. Berechnung der dichotomen Indikatorvariablen "fehlend/vorhanden" für jede Variable;

2. Berechnung der Patternvariablen mittels geometrischer Kodierung (MOSER/KALTON 1971:421)[7/]

$$PV = \sum_{i=1}^{p} a_i \cdot 2^{i-1}$$

; p = Variablenzahl

; a(i) = Indikatorvariable für Var.

(0-1 codiert)

Zusammen mit der vorhandenen Gruppierungsmöglichkeit in PAM wird der Output nach der Sortierung wesentlich übersichtlicher, da dann alle Fälle mit demselben Pattern hintereinander liegen.

Eine weitere Möglichkeit der übersichtlichen Darstellung stellt die Anwendung von RUBIN's "Factorization Table" dar (RUBIN 1974:468)[8/]. Nach der Erstellung der Indikatormatrix werden alle Zeilen der Matrix, die dasselbe (0-1)-Muster aufweisen, zusammengefaßt. Derselbe Prozeß wird für alle Spalten wiederholt[9/]. Die entstehende "reduced incompleteness matrix" enthält nur noch verschiedene Pattern und ist damit wesentlich leichter interpretierbar. Leider enthält kein Standardstatistikpaket diese Möglichkeit als eigenständige Prozedur, es ist aber mit den Datentransformationsroutinen möglich, sehr ähnliche Tabellen zu erstellen[10/].

Zur leichteren Interpretation der Indikatorenmatrix können natürlich auch Clusteranalysen benutzt werden (FRANE 1978:28). Damit liegt auch die Benutzung der Konfigurations-Frequenz-Analyse (KRAUTH/LIENERT 1973)

nahe und bildet einen Übergang zu den Test- verfahren (vgl. LÖSEL/WüSTENDÖRFER 1974:345).

Einen einfachen Ansatz zur Überprüfung der MAR-Annahme nach Typ 2 stellt bei ausreichender Fallzahl (KIM/CURRY 1977:219 sprechen von mindestens ca. 20 fehlenden Werten pro Variable) ein Chi-Quadrat-Test dar. LÖSEL/WÜSTENDÖRFER (1974:345) erwähnen einen Chi-Quadrat-Test über die Gleichverteilung der MD-Zahl der Spalten der Datenmatrix. Weiterhin legen sie die Prüfung der durch die Summe der MD pro Fall gebildeten Variablen auf Poissonverteilung nahe.

Ein Chi-Quadrat-Test über eine reduzierte Indikatorenmatrix findet sich bei KIM/CURRY (1977:219). Ein Beispiel: Gegeben seien 3 Variablen. Berechnet werden nun die Anteilswerte für nur auf Variable 1, nur auf Variable 2 und nur auf Variable 3 fehlende Fälle (q1, q2, q3) sowie für jede Variable der Anteil vorhandener Werte (p1, p2, p3). Dann ergibt sich für den erwarteten Anteil für nur auf Variable 1 fehlender Fälle:

$$m1= q1 * p2 * p3 * n;$$

entsprechend für nur auf Variable 2 fehlende Fälle:

$$m2 = q2 * p1 * p3 * n$$

und für nur auf Variable 3 fehlende Fälle:

$$m3 = q3 * p2 * p1 * n.$$

Der erwartete Anteil von auf jeder Variablen fehlenden Fällen ist dann: (1-p1 * p2 * p3) * n-m1-m2-m3; der erwartete Anteil vollständiger Fälle: p1 * p2 * p3 * n. Anschließend wird ein normaler Chi-Quadrat-Test mit 4 Freiheitsgraden berechnet. Sind k Variablen zu testen, ergeben sich bei dieser Reduktion k+2 Klassen, demzufolge k+1 Freiheitsgrade (KIM/CURRY 1977:219).

Wichtig bleibt festzuhalten, daß die Untersuchung der Indikatorenmatrix allein bestenfalls eine Aussagen über die MAR-Annahme nach der Definition der Typen 1 und 2 erlaubt (FRANE 1978:28).

1.3.2 Tests auf MAR

Tests auf MAR im Sinne der Definitionen 3 und 4 basieren auf der Verknüpfung der Datenmatrix mit der Indikatormatrix. Besteht zwischen den Matrizen Unabhängigkeit, wird MAR angenommen.

Eine der einfachsten Methoden des Tests auf MAR besteht in der Dichotomisierung einer Variablen mit hohem MD-Anteil in eine 0-1 Indikatorvariable "fehlend-nichtfehlend" und der Durchführung von t-Tests für intervallskalierte Variablen bzw. Chi-Quadrat-Tests für nominalskalierte Variablen (FRANE 1978:27,MÖNTMANN/BOLLINGER/ HERRMAN 1983:90-94)[11/]. Grobe Verstöße gegen die MAR-Annahme in Hinsicht auf univariate Differenzen der mit fehlenden Werten behafteten Fälle von den vollständigen Fällen können durch diese Tests aufgezeigt werden, sofern die dichotomisierte Variable mit dem MD-erzeugenden Prozeß verbunden ist ("This is frequently sufficient to dissolve any illusions that data are missing at random" FRANE 1978:27).

Eine naheliegende Erweiterung der Chi-Quadrat-Tests ist die Anwendung log-linearer Modelle auf die dichotomisierte Variable und die zu testenden Variablen des Datensatzes (MÖNTMANN/BOLLINGER/ HERRMAN 1983:94-95) - falls BMDP4F (oder ein anderes Programm zur log-linearen Analyse) nicht zur Verfügung steht, stellt eine Kreuztabellierung den ersten Schritt dar.

Die allen diesen Tests zugrunde liegende Idee der Benutzung der dichotomisierten Indikatorvariablen - deren Wahl theoretisch über Hypothesen des Ausfallmechanismus begründet werden müßte - läßt sich natürlich auf alle Testverfahren, die auf Unterschiede zweier Populationen zielen, übertragen.

Augenfällig ist die Anwendung der Diskriminanz- und Varianzanalyse (gegebenenfalls mit den Gruppierungsoptionen der Programme; FRANE 1978:28) und die Anwendung des BOX-M-Tests (Hypothese gleicher Kovarianzstruktur zweier Populationen)[12/]. Neben den Überprüfungen, ob die Kovarianzstrukturen

"global" Differenzen zwischen den Gruppen der Indikatorvariablen aufweisen, sind spezielle Tests bei bestimmten Fragestellungen möglich. WERTS/ROCK/ GRANDY (1979:201-204) demonstrieren z.B. die Überprüfung einer postulierten Faktorenstruktur in zwei Gruppen mit unterschiedlichen MD-Mustern (vollständig vs. eine Variable mit MD)[13/]. Die Analyse wird hier mit COFAMM (einem Programm zur konfirmatorischen Faktorenanalyse von SÖRBOM/JÖRESKOG) vorgenommen, vergleichbare Analysen sind mit LISREL VI ebenfalls möglich (Vgl. JÖRESKOG/ SÖRBOM 1984, Kapitel V und SÖRBOM 1979). Stehen LISREL oder COFAMM nicht zur Verfügung, lassen sich Faktorenstrukturen zumindest grob vergleichen, wenn die klassischen Faktorenanalysen getrennt nach den Gruppen der MD-Indikatorvariablen durchgeführt und die Lösungen mit einem Maß der Faktorenkongruenz verglichen werden (vgl. u.a. HARMAN 1976:341-347).

Alle bisher in diesem Abschnitt genannten Verfahren haben einen Schwachpunkt gemeinsam: Ist die Indikatorvariable nicht mit dem "eigentlichen" MD-Prozeß verbunden - m.a.W. existieren mehrere MD generierende Prozesse, von denen z.B. nur einer systematisch ist - so gehen diese Tests fehl: Sie zeigen MAR, weil der generierende Prozeß verfehlt wurde. Beispielsweise könnte neben einem systematischen MD-Prozeß ein unsystematischer MD-Prozeß existieren, der mehr Ausfälle produziert: Falls auf Grund eines syntaktischen Datenfehlers (falsches Format etc.) eine Variable eine extrem hohe Ausfallzahl besitzt, wird diese (nicht zum systematischen Prozeß gehörende) Variable für die Tests benutzt.

Eine andere Möglichkeit des globalen Tests der Abhängigkeit der Indikatormatrix und der Datenmatrix besteht in der Durchführung einer kanonischen Korrelation[14/]. Werden "signifikante" kanonische Korrelationen zwischen der Indikatormatrix und der Variablenmatrix festgestellt, so wird die MAR-Annahme verworfen. Dabei entsteht ein "technisches" Problem, das bei der Betrachtung der Grundgleichung[15/] der kanonischen Korrelation deutlich wird:

1-5 $$R = R(yy)^{-1} R(yx) R(xx)^{-1} R(xy)$$

Neben der Korrelationsmatrix der Indikatoren werden die Korrelationsmatrix der Variablen und die Variablen-Indikatoren - Korrelationsmatrix benötigt. Problemlos ist lediglich die erste Matrix, die anderen haben MD-Probleme. FRANE (1978:28) empfiehlt die Durchführung der kanonischen Korrelation nach der Ersetzung fehlender Werte durch Schätzwerte, wobei er die begrenzte Nützlichkeit dieser Form der Analyse betont[16/].

HILL/DIXON (1981:60) erwähnen in ihrer Simulationstudie eine interessante Variante der Korrelationstechniken zur Analyse von MD-Mustern. Sie erstellen sortierte Korrelationsmatrizen der Variablen und Indikatoren unter Verwendung eines MD-Korrelationsprogramms und der BMDP "allvalue"-Option (pairwise) mit dem BMDP-Programm P4M (Faktorenanalyse), das lediglich zur Sortierung und Konstrastierung der Korrelationsmatrizen im Output verwendet wird. Unter der MAR-Annahme müssen die Indikatoreninterkorrelationen und die Indikatorvariableninterkorrelationen nicht signifikante Werte aufweisen.

Zuletzt sollen noch zwei Möglichkeiten erwähnt werden, die stärker auf die Betrachtung fehlender Werte als Explanandum zielen. MÖNTMANN/BOLLINGER/HERMAN (1983:96-97) erörtern die Benutzung einer schrittweisen logistischen Regressionsprozedur (BMDPLR) mit dichotomer Indikatorvariable als abhängiger Variablen und substantiellen Variablen als Prädiktoren, wobei hier unterschiedliche Meßniveaus möglich sind. Eine gute Vorhersage der Indikatorvariablen ermöglicht (für die dem Indikator entsprechende Variable) die Ablehnung der MAR-Annahme und eine (empiristische) "Erklärung" der Ursache des Fehlens. Diese Methode hat zweifellos Vorteile gegenüber einfachen Methoden[17/]. Allerdings muß betont werden, daß BMDPLR für die Prädiktorvariablen nur die Menge der auf diesen Variablen vollständigen Fälle bei der Berechnung benutzt, was zu erheblichen Verzerrungen führen kann, falls z.B. MARC vorliegt[18/].

COHEN (1968:438) stellte eine Methode vor, die explizit die Information, die im Sachverhalt "fehlend" steckt, mit in die Analyse "substantieller" Zusammenhänge aufnimmt. Das von COHEN/ COHEN (1975:265-290) weiter ausgearbeitete

Verfahren weist weit über den hier interessierenden Test auf MAR hinaus. An dieser Stelle soll allerdings nur die Testanwendung diskutiert werden. In der einfachsten Form besteht die Methode nach COHEN in der Ersetzung fehlender Werte durch die Mittelwerte der vorhandenen Fälle der jeweiligen Variablen und der Bildung jeweils einer zusätzlichen (0-1)-kodierten Indikatorvariablen pro Variable: Die den Variablen entsprechende Indikatormatrix wird bei der Analyse mit der Datenmatrix zu einer Matrix kombiniert. Die Indikatorvariablen werden in die Regressionsmodelle mit aufgenommen.

COHEN/COHEN (1975:286) erwähnen die Möglichkeit, daß die Indikatoren nahezu perfekt korrelieren: In einem solchen Fall lassen sich keine Regressionen berechnen. Für diesen Fall schlagen sie die Verwendung einer einzigen Indikatorvariablen ("Anzahl fehlender Werte") vor, in einer Fußnote erwähnen sie darüber hinaus die Möglichkeit, die Indikatorenmatrix einer Faktorenanalyse zu unterziehen und die "factor scores" in die substantiellen Untersuchungen mit einzubeziehen[19/]. Faktorenanalysen mit dichotomen Variablen weisen besondere Probleme auf (DINGMAN 1958:13, MACRAE 1970:151-159), die aber in diesem Zusammenhang vernachlässigbar erscheinen (BORGATTA 1965). Die unbedingte Ablehnung faktorenanalytischer Techniken für "echt alternative Variablen, wie etwa das Geschlecht" (GAENSSLEN/SCHUBÖ 1976:269), erscheint zu orthodox (vgl. KIM/NIE/VERBA 1977).

In diesem Zusammenhang ist zuletzt noch ein Vorschlag von RUMMEL (1970:266) interessant. RUMMEL schlägt die Berechnung eines "missing data measures" - Anzahl fehlender Werte pro Fall - als zusätzlicher Variablen vor. Bei einer Faktorenanalyse ist dann die Ladung der zusätzlichen Variablen auf dem ersten Faktor bedeutsam. Ist die Ladung hoch, so kann es sich um einen "substantiellen" Effekt handeln. Weiterhin empfiehlt RUMMEL eine Aufteilung der Fälle in Gruppen hoher bzw. niedriger Anzahlen fehlender Werte und die anschließende Durchführung von, nach diesen Gruppen getrennten Faktorenanalysen. Bei Faktorstabilität sind die Ergebnisse unabhängig von der

Zahl fehlender Werte. Als dritte Möglichkeit schlägt RUMMEL eine orthogonale Rotation der Faktoren vor, so daß der erste Faktor kolinear zum "missing data measure" liegt. Zwar können die anderen Faktoren dann als unabhängig von der Anzahl fehlender Werte betrachtet werden (RUMMEL 1970:267), aber das garantiert in keiner Weise die Unabhängigkeit der gefundenen Faktoren vom generierenden Prozeß - die Anzahl fehlender Werte ist nur bedingt als Indikatorvariable für den generierenden Prozeß brauchbar[20/].

1.3.3 Empirische Ergebnisse zur Leistungsfähigkeit

HILL/DIXON (1981:59-60) berichten von einer Simulationsstudie, die die Empfindlichkeit unterschiedlicher MAR-Tests gegenüber verschiedenen Ausfallmechanismen und Verteilungen überprüfen sollte. Die Mittelwertdifferenzentests mit den Indikatorvariablen als Gruppierungsvariable sowie die "geordneten" Korrelationsmatrizen der Indikatoren und der Variablen zeigten für alle nicht MAR-Bedingungen deutlich die Verletzung der MAR-Annahme. Vergleichbare Ergebnisse wurden durch 2-Weg-Varianzanalysen erzielt, bei denen insbesondere die Interaktionseffekte der Indikatoren geprüft wurden. Als weiteres zur Aufdeckung systematischer Ausfallmechanismen erfolgreiches Verfahren führen sie Kreuztabellen an.

Clusteranalysen der Indikatormatrix zeigten keine brauchbaren Resultate. Die Benutzung schrittweiser logistischer Regressionen und Diskriminanzanalysen wird von ihnen wegen der Notwendigkeit der Schätzung der MD als umständlich abgelehnt.

Mit dieser Arbeit von HILL/DIXON liegen die ersten Ergebnisse zur Leistungsfähigkeit verschiedener Verfahren vor, die als Tests für systematische Ausfallmechanismen vorgeschlagen werden. Trotz der prinzipiell unvollständigen Untersuchungsmöglichkeiten der MD- Tests sind Untersuchungen dieser Art

notwendig, wenn die "naive" Nullhypothese des zufälligen Ausfalls den Ausgangspunkt der
Analyse bilden soll. Die Ergebnisse der Untersuchungen der Tests auf Zufälligkeit von MD können sich aber immer nur auf bestimmte Ausfallmechanismen beziehen, eine Zufallsstichprobe aus dem Universum aller denkbaren Ausfallmechanismen ist nicht möglich. Auch aus diesem Grund sind die Eigenschaften der Testverfahren prinzipiell nicht vollständig angebbar. Wie bereits angedeutet wurde, sind Ausfallmechanismen konstruierbar, die zumindest jeweils von einigen der Testverfahren nicht entdeckt werden können.

1.4 Zusammenfassung

Dieses Kapitel sollte die begrenzten Möglichkeit der Tests auf MAR verdeutlicht haben. Zwar stellt die Anwendung der Testverfahren einen Fortschritt gegenüber dem völligen Ignorieren potentieller MD-Probleme dar, aber Ergebnise der Tests sind allenfalls in dem Sinne interpretierbar, daß ein zufälliges Fehlen mit Sicherheit ausgeschlossen werden kann: Ein im Rahmen empirischer Sozialforschung nahezu triviales Ergebnis. Negativ verlaufende MAR-Tests dagegen stellen allenfalls notwendige Bedingungen der Akzeptierung der MAR-Annahme dar. Die Entwicklung verbesserter Testmöglichkeiten (vgl. z.B. NORDHEIM 1978:195) dürfte lediglich dazu führen, Anwender bei negativ verlaufenden Tests von der Robustheit der Schlußfolgerungen hinsichtlich der Tests der Kerntheorie zu überzeugen und damit wiederum potentielle Selektions-effekte zu verharmlosen, insbesondere wenn die Begrenzungen der MAR-Tests undiskutiert bleiben. Tests auf MAR stellen somit keine "Lösung" von MD-Problemen dar. Die Arbeiten von COHEN zeigen aber eine Möglichkeit auf, MD-Probleme zumindest als Explanandum empirischer Forschung nutzbar zu machen. Die Notwendigkeit der Entwicklung sozialwissenschaftlicher Ausfalltheorien soll nun in den beiden folgenden Kapiteln gezeigt werden. Der zweite Teil des Kapitels 4 kommt noch einmal kurz auf die Testproblematik zurück und belegt die Notwendigkeit der Ausfalltheorien aus einer anderen Richtung.

Anmerkungen zu Kapitel 1

1. RUBIN(1980:48) bezeichnet einen Responsemechanismus unter dieser Bedingung als "unconfounded".

2. Für die Einzelheiten muß auf RUBIN (1976:582-585) verwiesen werden.

3. Vgl. auch FRANE (1978:27). In den Arbeiten von Mathematikern wird in der Regel auf RUBIN (1976) Bezug genommen.

4. VAN GUILDER/AZEN (1981:53) benutzten in ihrer Simulationsstudie dieselben Definitionen, indem sie den Anteil (f) fehlender Werte mit der Ausfallwahrscheinlichkeit p als p = 3*f bestimmten.

5. Carl Morris, zitiert nach LITTLE (1983:401)

6. Vgl. in anderem Zusammenhang DWYER (1983:4):" The main function of significance tests has been to make trivial relationships appear important."

7. MOSER/KALTON hatten die Anwendung dieser Formel für die Erfassung von "multiple response questions" vorgeschlagen, vgl. hierzu
auch PAYNE (1977:47).

8. Das Ziel der Faktorisierung bei RUBIN ist nicht die Erstellung der "reduced incompleteness matrix", sondern die Faktorisierung der Likelihoodfunktion in ein Produkt der "complete data" und "incomplete data" Likelihoods mit verschiedenen Parametern. Die zu bildende "Factorizing Table" erlaubt die Identifizierung derjenigen Parameter, die mit den Standardtechniken für vollständige Daten geschätzt werden können, und derjenigen Parameter, für die spezielle MD-Techniken zur Schätzung verwandt werden müssen; vgl. RUBIN (1974:473).

9. Ein ausführliches Beispiel für die Erstellung der Matrizen findet sich bei HAMILTON (1975b:B1-B9). Ein FORTRAN-Programm zur Erstellung der Matrizen ist Gegenstand einer Arbeit von HINKINS (1980a; das Listing findet sich bei HINKINS 1980b:22-37).

10. Ein Beispiel mit SPSS-X findet sich im Anhang E.

11. Einzelheiten der technischen Durchführung finden sich bei MÖNTMANN/BOLLINGER/HERMAN (1983). Die Durchführung ist mit BMDP8D seit Version 81 besonders einfach. FRANE (1978:27) erwähnt die Benutzung von Hotellings-T-Quadrat.

12. Zum BOX-M-Test allgemein vgl. MARDIA/KENT/BIBBY(1979:140-141), zur Durchführung und der extremen Sensitivität des Tests vgl. TABACHNIK/FIDELL(1983:232-233); mit anderer Schwerpunktsetzung: EVANS/COOLEY/PISERCHIA (1979). Zum Vergleich von Korrelationsmatrizen vgl. auch JENNRICH(1970), STEIGER (1980), PEKALA/KUMAR (1985).

13. Das Verfahren liefert u.a. eine Schätzung der gemeinsamen Kovarianzmatrix, die von allen verfügbaren Werten Gebrauch macht. Die Autoren diskutieren keine Einschränkung des Anwendungsbereichs dieser Art der Kovarianzmatrixschätzung bei fehlenden Werten. Probleme entstehen spätestens dann, wenn kein "monotones Muster" herstellbar ist, d.h. wenn die MD stark in der Matrix "streuen". Das Problem besteht dann in der Wahl einer urspünglichen Schätzung der Kovarianzmatrix der unvollständigen Fälle. Da der Ansatz die gemeinsame Kovarianzmatrix entsprechend der Fallzahl "gewichtend" schätzt, wird den Fällen auf den unvollständigen Variablen die selbe Kovarianzstruktur unterstellt.

14. FRANE(1978:28), ein Beispiel bei MÖNTMANN/BOLLINGER/HERMAN (1983:98-101).

15. TABACHNIK/FIDELL (1983:151). Es existieren mehrere Formen dieser "Grundgleichung", die hier vorgestellte ist in diesem Zusammenhang anschaulicher als andere (vgl.z.B. COOLEY/LOHNES 1971:168-179 oder MARDIA/KENT/BIBBY 1979:282-288).

16. Bei MÖNTMANN/BOLLINGER/HERMAN (1983:100) wird dies nicht problematisiert: "Die Missing-Data-Korrelationsmatrix wird dort aus der BMDP-Datei eingelesen". Auf die Annahmen bei der Berechnung dieser Matrix gehen die Verfasser nicht ein.

17. Ein Vorteil liegt in der Nähe zu den Tests auf "Sample Selection Bias", vgl. hierzu Kapitel 4.

18. MÖNTMANN/BOLLINGER/HERMAN (1983) diskutieren keinerlei Einschränkungen dieser Methode.

19. Anwendungsbeispiele für die Methode von COHEN finden sich bei STUMPF (1978) und HUTCHESON/PRATHER (1981).

20. LÖSEL/WÜSTENDORFER (1974:346) zitieren RUMMEL(1970:260), der auf das Problem negativer Eigenwerte und überhöhter Ladungen bei Verwendung von "Pairwise"-Korrelationsmatrizen eingeht. Die Darstellung eines Problems bei RUMMEL gerät bei LÖSEL/WÜSTENDÖRFER zu einer "Möglichkeit zur Analyse von MD-Strukturen" - nach der Darstellung des Problems bricht die Erörterung ab. Von einem MD-Analyseverfahren kann kaum gesprochen werden, und als Test auf MAR erscheint das "Verfahren" genau so aussagekräftig wie eine Fehlermeldung eines zusammenbrechenden Programms.

2. Ausfallursachen I: Datenerhebung

"I have always believed that blanks do not occur randomly"

PREGIBON (1978:492)

Das Interview[1/] stellt im Rahmen der empirischen Sozialforschung[2/] z.Z. die vorherrschende Methode der Datenerhebung dar[3/] (MEULEMANN/REUBAND (1984:7). Die folgenden Ausführungen beschränken sich daher ausschließlich auf diese Form der Datenerhebung. Die Darstellung der "Fehlerquellen" bei der Datenerhebung bezieht sich hier weiterhin lediglich auf die Entstehung "nicht-substantieller Antworten" im Interviewprodukt, d.h. dem mehr oder weniger vollständig ausgefüllten Fragebogen. Unter "nicht-substantiellen" Antworten sollen operational "Daten" verstanden werden, die zwar zu erheben Absicht der Untersuchenden war, aber im Interviewprodukt nicht in einer Form vorliegen, die Aussagen über das Meßziel erlauben.

Im folgenden soll zunächst eine Typologie fehlender Werte dargestellt werden; danach wird eine allgemeine soziologische Theorie der Datenerhebung in ihrer Anwendung auf fehlende Werte erörtert und schließlich eine Reihe empirischer Ergebnisse vorgestellt, die für einen Großteil der "fehlenden Werte" in sozialwissenschaftlichen Erhebungen die Unhaltbarkeit der MAR-Annahme belegen oder zumindest demonstrieren, daß das Auftreten keiner Form der MD (in sozialwissenschaftlichen Datensätzen) als unabhängig von den substantiellen Beziehungen vorausgesetzt werden kann.

2.1 Typologie fehlender Werte in der Datenerhebung

Die im Interviewprodukt beobachtbaren "fehlenden Werte" sollen zunächst grob nach ihren möglichen Ursachen unterschieden werden. Eine einfache Typologie der beobachtbaren Phänomene könnte z.B. folgende Typen fehlender Werte umfassen[4/]:

- ausgelassene Fragen, Filterfehler der Interviewer ("Skip-Fehler"),
- Fehlcodierungen der Interviewer, die einen fehlenden Wert trotz vorliegender "gültiger" Reaktionen protokollieren ("Codier-Fehler"),
- Verweigerung einzelner Fragen durch die Befragten (KA),
- "Weiß nicht" - Antworten der Befragten (WN)[5/].

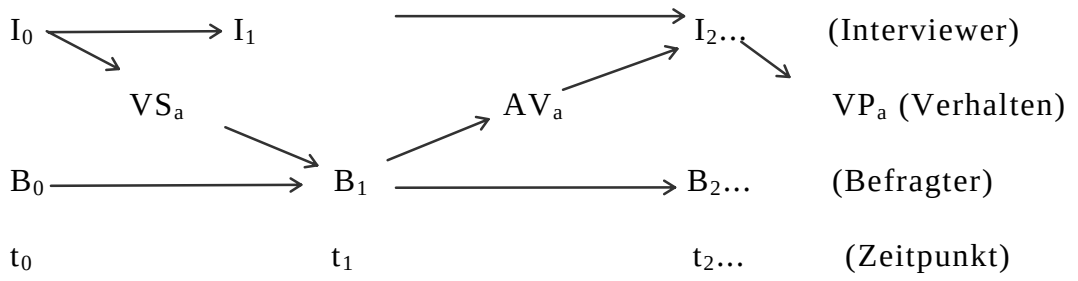
Zunächst sei festgehalten, daß für alle der genannten "Formen" "fehlender Werte" leicht systematische Ursachen postuliert werden können, insbesondere auch Ursachen, die mit der jeweils interessierenden theoretischen Fragestellung systematisch zusammenhängen. Wären diese postulierten Prozesse tatsächlich gegeben, so hätte ein Ausschluß der Fälle mit "fehlenden Werten" zwangsläufig eine Verzerrung der Ergebnisse zur Folge[6/]. Zwar existieren für jeden der genannten Mechanismen Untersuchungen, die die Existenz dieser Mechanismen zeigen, diese "Effekte" sind jedoch weder immer zu beobachten, noch geben diese Untersuchungen Hinweise darauf, unter welchen Bedingungen das Auftreten von MD mit substantiellen Beziehungen konfundiert ist.

2.2 Ansätze einer allgemeinen Theorie des Interviews

Im folgenden Abschnitt soll ein theoretischer Ansatz zur Erklärung der Entstehung "fehlender Werte" während der Datenerhebung in Surveys vorgestellt werden, der zum einen einen Anwendungsfall einer allgemeinen soziologischen Handlungstheorie darstellt und zum anderen die Möglichkeit einer Integration der Fülle von Einzelbefunden zumindest prinzipiell erlaubt. In diesem Zusammenhang soll das von ESSER (1975,1983,1984) entwickelte Grundmodell des Interviewverhaltens auf die hier interessierende Problemstellung eingegrenzt und unter dem Gesichtpunkt der Erklärung der MD spezifiziert werden.

ESSER's Grundmodell des Interviewverhaltens geht von der "idealen" Interviewsituation aus, in der nur zwei Akteure (Interviewer und Befragter) agieren[7/]. Bei starker Vereinfachung des Prozesses kann dieser wie folgt dargestellt werden:

Schema 1: Modell des Interviewprozesses (ESSER 1984:30)



Unter den Bedingungen der "internen Zustände" des Interviewers (I_0) und des Befragten (B_0) zum Zeitpunkt t_0 beginnt das Interview mit der Abgabe des Stimulus zur Frage a (VS_a). Der Befragte reagiert mit AV_a, was nach der Interpretation durch den Interviewer protokolliert wird (VP_a). Die Handlungen beider Akteure während des Interviews beeinflussen die internen Zustände beider, eine gewisse Konstanz der inneren Zustände wird aber unterstellt.

ESSER (1984:31) unterscheidet damit drei Explananda: 1. Die Reaktionen des Befragten, 2.die Stimulierungen, Beeinflussungen und Reaktionen des Interviewers und 3. das Vercodungshandeln. Insoweit als Resultate dieser Handlungen im Interviewprodukt "fehlende Werte" erscheinen, sollen einige Bedingungen im Rahmen einer kognitiven Handlungstheorie erklärt werden.

Das Grundprinzip einer handlungstheoretischen Erklärung[8/] besteht in der Annahme, daß Personen in jeder Situation diejenige Reaktionsmöglichkeit wählen, die ihnen subjektiv am besten geeignet erscheint, die Situation, die ihnen die stärkste Bedürfnisbefriedigung gewährt, herzustellen[9/]. Damit stellt sich für

jedes oben genannte Explanandum die Frage nach den Motivationen, Erwartungen, Kognitionen und "Strategien" der Zielerreichung. Da eine vollständige handlungstheoretische Erklärung jeweils auf eine spezielle Fragestellung bezogen sein müßte, soll im folgenden sowohl für interviewerbedingte MD (Stimulusabgabe und Vercodungshandeln) als auch für befragtenbedingte MD nur eine Aufzählung von handlungsrelevanten Bedingungen und Möglichkeiten zur Entstehung von MD erfolgen. Daß die jeweiligen speziellen Modelle bei einem gegebenen "inhaltlichen" Problem auf Rückkoppelungsprozesse (Interviewerverhalten wirkt auf Befragtenverhalten und umgekehrt) einzugehen hätte, müßte durch das Modell von ESSER deutlich geworden sein. Insbesondere die Möglichkeit, daß die "internen Zustände" während des Interviews Wandlungen unterworfen sein können, ist nochmals hervorzuheben.

2.2.1 Interviewerbedingte MD

Zunächst soll stark vereinfachend die Frage nach den Motivationen des Interviewers betrachtet werden. Neben dem nahezu unwahrscheinlichen Fall der Internalisierung einer "Professionsethik des Interviewers" (ESSER 1984:43), bei der, wie immer bei internalisierten Normen, schon die Ausführung einer der Norm ("keine Fälschungen, keine Fehlvercodungen, keine sachfremden Beeinflussungen") entsprechenden Handlung diese als belohnend empfunden wird, sind u.a. Motivationen zu nennen, die auf die durch Sozialkontakte allgemein möglichen Befriedigungen zielen. Im Rahmen der Praxis empirischer Sozialforschung scheinen, insbesondere wenn man die Rekrutierungsmöglichkeiten für Interviewer beachtet, zwei andere Aspekte von größerer Bedeutung zu sein: Die bedeutenste Motivation der Interviewer dürfte über Interesse am Forschungsgegenstand hinaus das Interesse am Gelderwerb sein.

Daneben sind noch Motivationen zu nennen, die mit der Selbstpräsentierung der Interviewer zusammenhängen, sowie Mechanismen, die aus der Tatsache resultieren, daß Interviewer (wie jedes informationsverarbeitende System) eine begrenzte Verarbeitungskapazität besitzen und daher mit

"komplexitätsreduzierenden" Strategien (selektive Wahrnehmung, Konsistenzbedürfnis usw.) arbeiten. Ist weder eine "Professionsethik" internalisiert noch das "Interesse" am Forschungsgegenstand oder die Motivationen der Selbstpräsentation kongruent mit den Forderungen der Professionsethik, dann sind als "Kosten" eines intentionalen "Fehlverhaltens" des Interviewers fast ausschließlich solche Maßnahmen zu nennen, die der Interviewer als Reaktion seiner Auftraggeber auf ein entdecktes Fehlverhalten erwartet. Die subjektive Kostenerwartung wird damit eine direkte Funktion der wahrgenommenen Effektivität der Kontrollen.

Die notwendigen kognitiven Fähigkeiten einer Tätigkeit als Interviewer prinzipiell vorausgesetzt, ergeben sich zwei Möglichkeiten der Entstehung "fehlender Werte" durch den Interviewer: Entweder stellt der Interviewer eine Frage nicht, oder er codiert eine Antwort fälschlich (bewußt oder unbewußt) als fehlenden Wert.

Eine nichterfolgte Fragestellung kann zunächst als "zufälliges" Überschlagen erfolgen; systematische Mechanismen sind aber auch hier konstruierbar: Erfolgt z.B. das Interview unter vom Befragten gesetzten zeitlichen Limitierungen, steigt die Wahrscheinlichkeit solcher Fehler.

Wesentlich bedeutender in Hinsicht auf systematische Effekte sind jedoch falsche Filterführungen des Interviewers. Neben kognitiven Schwierigkeiten erscheinen zwei Ursachen besonders wichtig: Erstens kann sich eine Erwartungshaltung beim Interviewer aufbauen, die ihn bestimmte Fragenkomplexe als von vornherein unzutreffend klassifizieren läßt, und zweitens kann der Interviewer durch seine Praxis lernen, daß bestimmte Fragen eine Reihe von Nachfragen bedingen: Übersprungene Fragen innerhalb von Filtern verkürzen das Interview. Da solche "Fälschungen" nur sehr schwer nachweisbar sind, die subjektive Kostenerwartung also auf Grund der niedrigen Sanktionswahrscheinlichkeit gering ist und die Entdeckungswahrscheinlichkeit (u.U. für den Interviewer erkennbar) mit Merkmalen des Befragten zusammenhängen kann (Sprachbarriere, Erreichbarkeit) sind

systematische Zusammenhänge der Ausfälle mit Untersuchungsfragestellungen möglich.

Im Zusamenhang mit Skip-Fehlern soll ein letzter Punkt genannt werden: Der Interviewer kann sich durch das Stellen einer Frage in seiner Selbstpräsentation gefährdet sehen; dieser Effekt dürfte besonders starke Selektionsfolgen besitzen, da die subjektiv empfundene Notwendigkeit der Selbstpräsentation mit wahrgenommenen Merkmalen der Befragten zusammenhängen kann.

Läßt man intentionale und triviale Fehler bei der Codierung außer acht (unlesbare, mehrdeutige Codierung, Schreibfehler) bleiben als MD-Ursachen beim Codierungshandeln der Interviewer erwartungskonforme Fehlvercodungen. Die Wahrscheinlichkeit einer MD- Codierung dürfte sowohl von Interviewer- als auch von Befragtenmerkmalen abhängig sein. Neben dem Konsistenzbedürfnis des Interviewers spielt die Ambiguität der Antwort des Befragten eine Rolle (ESSER 1983:50; BUNGARD/LÜCK 1974:50). Finden zudem Interviewerrückfragen nur unzureichend statt (aus denselben Gründen wie das Nichtstellen der Frage überhaupt), sind systematische Ausfallmechanismen bei gegebener Fragestellung und Stichprobe konstruierbar.

2.2.2 Befragtenbedingte MD

"...we too often settle for comic-book statistics on
infinitely complex matters of opinion."

BOGART(1967:345)

In der oben angeführten "Typologie" wurde zwischen expliziter Verweigerung und der "Weiß-Nicht"-Antwort unterschieden. In der Forschungspraxis lassen sich auf verschiedenen Ebenen zwei Tendenzen in der Behandlung dieser "fehlenden Werte" isolieren: die Zusammenfassung aller "nicht-substantiellen Antworten" zu "Itemnonresponse" bzw. "Missing Data" in der Datenanalyse und die Erstellung von Typologien der "Weiß nicht"- Antwort nach vermuteten Ursachen in der

stärker methodenorientierten Literatur. In diesem Abschnitt soll zunächst gezeigt werden, daß die Typologien der "Weiß-Nicht"-Antwort sehr einfach durch eine handlungstheoretische Interpretation "rekonstruierbar" sind und die explizite Verweigerung einen Spezialfall darstellt.

2.2.2.1 Elemente einer Erklärung der Befragtenreaktion

Auch hier soll zwischen Motivationen, Kognitionen und Widerständen/Barrieren[10/] als Elementen einer handlungstheoretischen Erklärung unterschieden werden[11/]. Der Motivationsaspekt im Zusammenhang mit der Befragtenreaktion müßte sich auf die Klärung der Ursachen beziehen, warum im Rahmen eines Surveys Fragen überhaupt beantwortet werden. Dieses Problem hat allerdings bisher in der Literatur kaum eine befriedigende Erklärung gefunden.

DILLMAN (1978:12-18) versucht eine Erklärung mit den Austauschtheorien von Homans, Blau, Thibaut und Kelley. Zentrale Annahme hierbei ist die Vermutung, daß ein gegebenes Verhalten umso wahrscheinlicher gezeigt wird, je niedriger dessen Kosten liegen und je höher der zu erwartendes Nutzen liegt, der vom Interaktionspartner vermittelt werden kann. Neben dem "Vertrauen", das in den Interaktionspartner in Hinsicht auf die Erfüllung seiner Obligationen gesetzt werden muß, stellt sich die Frage nach dem "Nutzen" der Beantwortung einer Frage. Damit wird aber das Problem lediglich verschoben: Da explizite "Belohnungen" nur sehr selten im Rahmen einer Interviewsituation möglich sind, bleiben DILLMAN in Anlehnung an Blau nur Spekulationen über positiv erlebte Interaktionen und die Stützung des Wertsystems des Befragten durch die Befragung als strategisch im Rahmen einer Untersuchung nutzbare "Motivationen"[12/].

BATESON (1984:112-113) unterstellt die Beantwortung einer Frage als "normal result" des Zusammentreffens zweier Fremder in einer Situation, in der einer der Akteuere um eine Information bittet, sofern keine widersprechenden sonstigen Motivationen ("sensitive questions") vorliegen. Eine Interviewsituation stellt aber

gerade keine "normale" Situation dar, BATESON's Erklärung bleibt daher unbefriedigend. ESSER (1983:60) leitet hingegen die Beantwortung von Fragen aus der Bereitschaft zur Meinungsäußerung als Folge eines generalisierten Verhaltensstils ab[13/]. Die eigentliche Frage bleibt daher ungeklärt.

Damit verbleibt hier nur die Möglichkeit, mögliche Einlußfaktoren aufzulisten. Als Faktoren, die positiv auf die Motivation Fragen zu beantworten wirken, lassen sich z.B. Bedürfnisse der Selbstpräsentierung, Reziprozitätsnormen und Konsistenzbedürfnisse hinsichtlich der Selbstwahrnehmung nennen. Andererseits sind Motivationen denkbar, die daraus resultieren, daß eine einmal begonnene Interaktion durch eine "ungewöhnliche" Reaktion eine Störung erfahren könnte und dieses durch die Beantwortung vermieden werden soll. Solche "Deutungen" beantworten aber kaum die eigentliche Frage nach der Ursache.

Hier soll die Beantwortung einer Frage bei Abwesenheit störender Faktoren als (wie auch immer zu erklärende) Basismotivation ("Beantwortungsnorm") vorausgesetzt werden.

Der kognitive Aspekt bezieht sich auf das Vorhandensein kognitiver Fähigkeiten des Befragten: Voraussetzung der Befragung ist die prinzipielle Fähigkeit des Befragten, die Frage überhaupt zu verstehen, und zum anderen die Existenz der gewünschten Information.

Der Bereich der Widerstände und Barrieren umfaßt u.a. alternative Ziele, die der zu Befragende in der Befragungssituation verfolgt, sowie insbesondere befürchtete Konsequenzen einer Fragenbeantwortung bzw. Antwortverweigerung.

Die wesentlichsten Elemente einer handlungstheoretischen Erklärung sind damit genannt. Die angemessene Darstellung einer konkreten handlungstheoretischen Erklärung bestünde nun in einer Formalisierung des Ansatzes (z.B. ESSER 1984:36-39). Das Problem einer solchen Formalisierung liegt jedoch in der

notwendigen Explizierung der Handlungsziele, über die bei einer Formalisierung die Summenindices laufen. M.a.W.: Soll für ein gegebenes Problem eine explizite Erklärung erfolgen, so ist die explizite Angabe der möglichen Handlungsziele (einschließlich deren Bewertungen, Erwartungen über Realisationswahrscheinlichkeiten usw.) erforderlich. Da eine vollständige Aufzählung möglicher Handlungsziele unmöglich ist, verbleibt neben der "leeren" formalen Darstellung der Versuch, Hinweise für eine handlungstheoretische Erklärung der in der Literatur zu findenden Typologien der "nichtsubstantiellen" Befragtenreaktion zu geben. Dies ist die Aufgabe des nächsten Abschnitts.

Wichtig bleibt festzuhalten, daß jegliche Form fehlender Daten, die im Erhebungsprozeß entstehen, einer handlungstheoretischen Erklärung bedürfen, die sich jeweils auf die gegebene Interviewsituation (einschließlich des jeweiligen Fragengegenstandes) beziehen muß: Obwohl damit die prinzipielle "Form" der Erklärung der MD gegeben ist, muß eine explizite Erklärung u.U. die jeweilige "Kerntheorie" mit einbeziehen. Eine Erklärung der MD muß also für jedes spezielle MD-Problem als zusätzliche "Instrumententheorie" in der Datenanalyse neu erfolgen.

2.2.2.2 Rekonstruktion einer "WN"-Typologie

In aller Regel ist aus der "nicht-substantiellen" Antwort kein Hinweis auf deren Ursache möglich. Vorhandene Typologien fassen daher vermutete, teilweise sehr heterogene Ursachenkomplexe zusammen.

In der Literatur finden sich Differenzierungen der WN nach vermuteten Gründen der Abgabe einer "nichtsubstantiellen" Antwort. So unterscheiden z.B. COOMBS/COOMBS (1976:497) nach "lack of information", "don't want to know" (in Anlehnung an BOGART 1967), verdeckter Verweigerung und Unentschiedenheit. LEVERKUS-BRÜNING (1966:22-29) unterscheidet neben den "Verweigerern" und den "Unentschiedenen" zusätzlich "die Nichtinformierten" und "die Meinungslosen", wobei das Trennkriterium zwischen diesen beiden

Typen in der Existenz einer richtigen ("normierten") Antwort besteht. Ausgehend von der Typologie von LEVERKUS-BRÜNING soll für die einzelnen Typen jeweils kurz skizziert werden, wie die bei der Entstehung vermuteten Prozesse handlungstheoretisch erklärt werden können.

Die Verweigerung einzelner Fragen kann als offene Verweigerung oder als verdeckte Verweigerung ("Weiß-Nicht") auftreten. Beide beruhen bei gegebenen kognitiven Voraussetzungen auf der Dominanz einer anderen Handlungstendenz als der durch die "Basismotivation" geforderten. Die Ursachen können z.B. vermittelt über den Frageinhalt in der wahrgenommenen Gefährdung anderer Handlungsziele liegen. Dazu gehören beispielsweise die Gefährdung sozialer Interaktionen (Anwesenheitseffekte Dritter) und der Selbstpräsentation ("sensitive Fragen"). Gelingt es z.B. dem Befragten nicht, die Interviewsituation als von allen anderen Aspekten seiner sozialen Interaktion losgelöstes und in Hinsicht auf diese folgenloses Ereignis zu interpretieren, und erachtet er die Situation zumindest potentiell in ihren wahrgenommenen Konsequenzen für sein Alltagsleben relevant, ist die Verweigerung einzelner Fragen erwartbar. Diese Mechanismen werden in der Literatur zumeist nur indirekt über das Problem der Anonymitätszusicherung erörtert.

Bei einer Verweigerung spielen also vor allem die Existenz anderer Handlungsziele, deren Wichtigkeit und die erwartete Wahrscheinlichkeit, mit der eine Beachtung der Beantwortungsnorm die Realisation anderer wichtiger Handlungsziele verhindern könnte, eine Rolle. Die Entscheidung zwischen "offener" bzw. "verdeckter" Verweigerung hängt wiederum von erwarteten Konsequenzen in der Folge der jeweiligen Handlung ab. Bei (für den Befragten) überwiegend undurchsichtigen Konsequenzen der Interviewsituation ist eine verdeckte Verweigerung - bei gleichzeitiger Abwesenheit anderer Motivationen wie z.B. internalisierter "Aufrichtigkeitsnormen" - wahrscheinlicher. Die hier nur angedeutete große Zahl der Parameter bei der Bestimmung der Stärke der Handlungstendenz macht deutlich, daß zum einen die Konstruktion systematischer Ausfälle über Verweigerung unproblematisch ist und zum anderen die Bestimmung der Handlungstendenzen für ein gegebenes Problem jeweils neu zu

leisten ist. Demzufolge können Versuche, systematische Variationen des Antwortverhaltens als von groben "externen" Variablen (Anonymitätszusicherung, Anwesenheitseffekte) abhängig zu demonstrieren, fehlgehen, weil in der Aggregierung die Mechanismen zu einer Auslöschung einzelner Effekte führen können.

Als weiterer Typ der "WN" bei LEVERKUS-BRÜNING soll nun "Uninformiertheit" als Ursache betrachtet werden. Das Trennkriterium (Existenz einer richtigen Antwort) ist nur im Zusammenhang mit "Faktenfragen" sinnvoll anwendbar. In analytischen Surveys besitzen "Faktenfragen" zwei Funktionen: Einmal dienen sie als Quelle für prinzipiell auch auf anderem Wege zu erlangende Informationen (Freizeitgewohnheiten u.ä.), zum anderen als Messung einer "latenten" Variablen. Beide Fälle müssen unterschieden werden, da unterschiedliche Implikationen für ihre Handhabung erwachsen. Dem Befragten stellen sich allerdings beide "Formen" der Faktenfrage gleich. Die Funktion der Frage sollte (erkennt der Befragte die Legitimität[14/] der Frage an) für sein Handeln keine Bedeutung besitzen. Im Fall einer "nicht-substantiellen Antwort" auf eine Faktenfrage fehlen (unter der Voraussetzung der Existenz der Basismotivation und der geringeren Handlungstendenz in bezug auf alternative Ziele) die kognitiven Voraussetzungen zur Beantwortung der Frage.

Der Fall, daß auf eine Frage nach einem extern validierbaren Sachverhalt keine "substantielle" Antwort gegeben wird - "Weiß nicht" als "Befragungsfehler" oder "Mißerfolg" (ZEISEL 1970:48) - hat in der Literatur vor allem unter Aspekten der Gedächtnispsychologie Beachtung gefunden (SUDMAN/BRADBURN 1974:67-92). Da "Vergessenskurven" interindividuell und intraindividuell (situational) verschieden sind, erscheint die Konstruktion systematischer Ausfälle über diese Mechanismen möglich. Eine vollständige Erklärung müßte jedoch das "Vergessen" selbst zum Explanandum machen. Solche Effekte können zumindest

prinzipiell durch die Konstruktion der Erhebungsinstrumente gemildert werden ("aided recall"; notfalls die Erhebung über andere Quellen). Dieser Aspekt soll daher hier nicht weiter verfolgt werden.

Sonstige Formen der "Faktenfrage" zielen nicht auf das Ausmaß des Wissens, sondern sollen eine "latente" Dimension messen: z.B. bei Sprachtests, in denen dann ein "Weiß-nicht" eine nahezu eindeutige Interpretation besitzt. In diesem Fall stellen WN keinesfalls "fehlende Werte" dar. Die Zuweisung eines theoretisch begründeten Wertes in der Datenanalyse erscheint unter solchen Umständen legtitim[15/].

Die Entstehung der WN durch "Unentschiedenheit" ist nur bei einer fehlenden Mittelkategorie bei den Antwortvorgaben möglich (das Problem kann nur bei ordinalen Fragen auftreten). Entsteht das WN tatsächlich durch "Unentschiedenheit", dann müßte die Einführung einer neutralen Mittelkategorie solche Effekte ausschließen können[16/]. Trotzdem sollen mögliche Ursachen der "Unentschiedenheit" erwähnt werden. So kann die Wahl der WN als Substitut für eine fehlende "neutrale" Kategorie als Strategie des Befragten betrachtet werden, um befürchtete Konsequenzen einer "inhaltlichen" Reaktion zu vermeiden. Zum anderen können (wiederum) alternative Ziele in der Befragunssituation wichtig werden, z.B. die Präsentation als "komplexe Persönlichkeit", der unterschiedliche Positionen zu der in Frage stehenden Problematik bekannt sind und eine Abwägung unmöglich ist ("Norm der 'intellektuellen Komplexität'", vgl. ESSER 1983:61).

Schließt man alle bisher genannten Sonderfälle und den Fall, daß eine bestimmte Einstellung das Explanandum darstellt, aus, dann bekommt das "weiß nicht" eine Indikatorfunktion für konzeptuelle Probleme innerhalb der Untersuchungsfragestellung. Das im Rahmen der "Meinungsforschung" mit "Non-Attitude" bezeichnete Problem[17/] besitzt in analytischen Surveys seine Entsprechung bei der Messung subjektiver Wahrnehmungen und Überzeugungen. Dieses Problem stellt den Kern dessen dar, was in der Regel als "Meinungslosigkeit" bezeichnet wird. Es besteht in der Verletzung einer der Voraussetzungen

der Datenerhebung, nämlich der Annahme des Vorhandenseins entsprechender Kognitionen bei den Befragten.

Auf diesem Hintergrund sind dann Korrelate der "Meinungslosigkeit" mit mangelnder Vertrautheit mit dem Gegenstand oder mangelndem Interesse am Gegenstand nahezu trivial. Da empirische Sozialforschung sich nicht in "Meinungsforschung" erschöpft, sondern primär der Überprüfung sozialwissenschaftlicher Theorien dient und somit Punktschätzungen nur von untergeordneter Bedeutung sind, muß dieses Problem als eine konzeptionelle Schwierigkeit verstanden werden. Im Regelfall sollen "Meinungsfragen" im Rahmen empirischer Sozialforschung zur Messung einer latenten Variablen dienen, falls nicht das Vorhandensein einer "Meinung" selbst das Explanandum darstellt (in diesem Fall ist das Nicht- Vorhandensein ein gültiger Meßwert). Ist die Messung einer latenten Variablen das Ziel der Erhebung einer Frage, dann ist weder ein Ausschluß der Fälle mit WN noch ihre (wie auch immer durchgeführte) Zuweisung eine legitime Verfahrensweise: In diesem Zusammenhang zeigt das "WN" an, daß die vom Konstrukteur des Instruments unterstellte Dimension bei den Befragten keine Entsprechung besitzt: Das Meßziel existiert nicht. Es handelt sich also nicht um einen "Respondent Error", sondern um einen Fehler bei der Konzeption der Messung. Die einzige Lösungsmöglichkeit bei einem gegebenen Datensatz kann dann nur in der Benutzung anderer theoretischer Konzepte liegen.

Zusammenfassend soll noch einmal betont werden, daß eine Reihe von Problemen mit "fehlenden Werten" in konzeptuellen Problemen der Operationalisierung der Untersuchungsfragestellung begründet liegen und die Kennzeichnung als "fehlende Daten" damit unzutreffend ist. Ein anderer Teil mag durch die spezielle Konstruktion des Meßinstrumentes bedingt werden, so daß die kognitiven Voraussetzungen einer Befragung für einen Teil der Population nicht gegeben sind. Hier sind insbesondere Konstruktionsfehler durch unverständliche oder mehrdeutige

Fragen erwähnenswert[18/]. Daß solche Effekte der Frageform sich u.U. nur in bestimmten Teilen der Population zeigen können, ist nahezu selbstverständlich[19/].

Die Erörterung der einzelnen "Typen" des "Itemnonresponse" dürfte die Vielzahl, der bei einer Erklärung der Ausfälle zu beachtenden Parameter gezeigt haben. Systematische Ausfallmodelle zu erstellen, bereitet daher kaum Probleme. Dadurch aber, daß sehr viele sich evtl. gegenseitig kompensierende Mechanismen bei der Entstehung von "Itemnonresponse" beteiligt sind, wird die große Zahl widersprüchlicher Ergebnisse einfacher Korrelationen nichtsubstantieller Antworten mit in der Regel einfachen demographischen Merkmalen oder einfachen Merkmalen der Interviewsituation verständlich. Eine Erklärung des "Itemnonresponse" muß für jedes spezielle Problem jeweils neu erfolgen, sollen Hypostasierungen von "Typen" der Befragten, die (aus Gründen, die sich der Erfassung bei dieser Herangehensweise anscheinend entziehen) überwiegend "nicht-substantiell" antworten, vermieden werden[20/].

2.3 Empirische Ergebnisse

Abschließend sollen einige empirische Ergebnisse als Beleg für das Ausmaß des MD-Problems in der sozialwissenschaftlichen Forschungspraxis herangezogen werden. Zum anderen soll der Eklektizismus[21/] bei der Auswahl "erklärender Variablen" im Zusammenhang mit "Itemnonresponse" verdeutlicht werden.

Die Literatur zu "Itemnonresponse" ist bei weitem nicht so umfangreich wie die Literatur zu Totalausfällen, ein Umstand, auf den schon FERBER (1966) in seiner klassischen Arbeit zu Itemnonresponse in einer schriftlichen Befragung hinwies. In der Tradition dieser Arbeit existieren nunmehr zwar eine Anzahl neuerer Arbeiten, die aber weitgehend theorielos vorgehen und widersprüchliche Resultate erbringen (z.B. MAURANA 1976, HUGHES 1979, DURAND/GUFFEY/PLANCHON 1983, OMURA 1983). In der Regel werden alle Formen "nicht-substantieller" Antworten zusammengefasst und Korrelationen mit lediglich "demographischen" Variablen berichtet. Selten hingegen wird

beispielsweise versucht, den Einfluß des Interviewers auf die Itemnonresponseqoute zu ermitteln[22/].

Aufgrund der oben erwähnten Probleme ist der Nachweis solcher Effekte nicht unproblematisch. So konnten SINGER/KOHNKE-AGUIRRE (1979:251-253) nur eine schwache (und nur in einem Fall signifikante) nicht-monotone Beziehung zwischen der Itemnonresponserate und der vom Interviewer erwarteten "Schwierigkeit" der Items zeigen. In einer Nachfolgeuntersuchung konnten SINGER/FRANKEL/ GLASSMAN (1983:76-77) in einem Telefonsurvey keine signifikante Beziehung zwischen Interviewererwartung und Itemnonresponse belegen, obwohl die Itemnonresponseraten bei einzelnen Items sehr hohe Werte erreichten, beispielsweise bei der Frage nach dem Einkommen mit 7% bis 11.9% und der Frage nach dem Marihuanakonsum der Freunde mit 8.7% bis 10.7%.

Nach der oben dargestellten Theorie sind die Motivationen des Befragten in der Befragungssituation für die Beantwortung der Fragen von zentraler Bedeutung. Die Frage nach der Motivation der Befragten wird in der amerikanischen Literatur erst seit dem Beginn der Diskussion um die "Respondent Burden" (BRADBURN 1979) ausführlicher erörtert; das Interesse an diesem Problem scheint erst durch die steigenden Ausfallraten in amerikanischen Surveys erwacht zu sein (vgl. STEEH 1981). BRADBURN (1979:50-52) führt Länge des Interviews, wahrgenommene Wichtigkeit des Surveys (sowie einen evtl. Interaktionseffekt der beiden Variablen), Aufwand (Dokumentenvergleich etc.), Streß durch unangenehme Fragen und die Häufigkeit der Befragung als mögliche determinierende Variablen an. Die empirische Überprüfung der Hypothesen ist mit Schwierigkeiten behaftet, so konnten SHARP/FRANKEL(1983) keinen Einfluß der "Respondent Burden" auf Itemnonresponse, der in ihrer Studie im Mittel bei 3% (0.5% Verweigerungen) lag, feststellen. Die subjektiv wahrgenommene Belastung konnte durch die erwähnten Variablen zudem nur unzureichend erklärt werden[23/].

Zum Problem der Anonymitätszusicherung als Beispiel für eine äußerst indirekte Operationalisierung befürchteter Konsequenzen existieren seit dem Beginn der Diskussion um das Problem des "informed consent" deutlich mehr Arbeiten. Erwähnt seien in diesem Zusammenhang nur die Arbeiten von SINGER. SINGER (1978:153) konnte bei nur wenigen Items einen "signifikanten" Zusammenhang zwischen dem Grad zugesicherter Vertraulichkeit der Angaben und dem Ausmaß des Itemnonresponse nachweisen.

In den Zusammenhang mit Faktoren, die die Motivation, Fragen zu beantworten beeinflussen können, gehört das Problem der "unangenehmen Fragen". Nach o.a. Ausführungen ist Nonresponse bei unangenehmen Fragen ein Resultat aus der Tatsache, daß andere Ziele in der Befragungssituation wichtiger werden als die Befolgung der "Beantwortungsnorm", z.B. Selbstpräsentation oder Konsequenzbefürchtungen. Demzufolge sind starke situationale Schwankungen zu erwarten. Dies soll an Hand einiger Arbeiten zum Vergleich verschiedener Erhebungsmethoden bei einer der klassischen "unangenehmen Fragen", der Frage nach dem Einkommen[24/] illustiert werden. JORDAN/MARCUS/REEDER (1980:212-213) geben 12% MD auf Einkommensfragen in einem Haushaltsinterview gegenüber 21%, von denen 14.5% als "don't know" codiert waren, in einem Telefoninterview an. ROGERS (1976:56) gibt 6% Verweigerungen und 6% "Don't know" auf Einkommensfragen bei einem persönlichen Interview gegenüber 11% Verweigerungen und 9% "Don't know" in einem Telefoninterview an. GROVES/KAHN (1979:87) geben bei einem Vergleich zwischen einem Survey mit persönlichem Interview und drei Telefoninterviews als Mittelwert bei 12 Einstellungsfragen für die Telefoninterviews zwischen 1.5% und 2.7% "Not Ascertained"-Codes gegenüber 1.4% mit persönlichem Interview an. Eine Zusammenfassung der Verweigerungen, "Don't know" und "Other"- Codes in Hinsicht auf eine Reihe von Fragen zu Einkommen und Steuern zeigte für die Telefoninterviews zwischen 5.3% und 36.7% nicht-substantieller Antworten gegenüber 2.5%-13.3% in den persönlichen Interviews (GROVES/KAHN 1979:99). Die tendentiell höheren Ausfallquoten für das Telefoninterview zeigen sich bei fast allen Variablen der Untersuchung (GROVES/KAHN 1979:106), die Verfasser betonen allerdings die schwierige

Verallgemeinerbarkeit dieses Befundes. Als vergleichbares Ergebnis gibt SIEMIATYCKI (1979:241) in einem Vergleich zwischen schriftlicher, telefonischer und persönlicher Befragung 8.8%, 18.9% und 14.1% Ausfälle in Hinsicht auf das Haushaltseinkommen an. BELL(1984:212) konnte in einem Telefoninterview die Nonresponserate bei einer Einkommensfrage von 26% auf 4% senken, indem eine kategoriale Vorgabe gegeben wurde. SUDMAN u.a. (1977:180) geben 1% "No answer" für alle Items ihres Surveys mit Ausnahme von Beischlaf, Masturbation und Einkommen an, bei der die Rate "under 5%" liegt.

Abschließend sollen einige Arbeiten zu in der Literatur vermuteten unspezifischen Determinanten, insbesondere Bildung, erwähnt werden. CONVERSE(1976) konnte neben Bildung lediglich "Personal Involvement" als signifikanten Prädiktor für "No Opinion" demonstrieren, weder sprachliche Schwierigkeit der Items noch Frageform zeigten deutliche Effekte. FRANCIS/BUSCH(1975:210) geben für drei Surveys zwischen 1960 und 1968 an, daß jeweils mehr als 10% der Befragten auf mehr als 15% der Fragen nichtsubstantielle Antworten gaben. Sie zeigen stabile, wenn auch schwache Zusammenhänge zwischen der Anzahl nicht-substantieller Antworten und den "Variablen" Alter, Bildung, Einkommen, Geschlecht, "Race", "Involvement" und "Political Efficacy" (FRANCIS/BUSCH 1975:215; 217)[25/]. CRAIG/McCANN (1978:286) geben für ihre schriftlichen Umfragen zwischen 2.9% und 8% Itemnonresponse an. Sie konnten Zusammenhänge mit "Alter" zeigen, aber nur schwache oder keine Zusammenhänge mit Bildung, Einkommen und Länge des Fragebogens.

Seit den Arbeiten von SCHUMAN/PRESSER gibt es mehrere Arbeiten, die "WN" vor der eigentlichen Frage auszufiltern versuchen. Schon BOGART(1967:337) wies auf die Notwendigkeit der Vorfilterung hin, in größerem Umfang werden solche Versuche aber erst seit der Arbeit von SCHUMAN/PRESSER (1978) vorgenommen. SCHUMAN/PRESSER (1978:271) geben eine mittlere Steigerung um 20-25% der WN-Rate durch explizite Vorfilter an. BISHOP/OLDENDICK/TUCHFARBER (1983: 535) berichten aus ihren Surveys größere Effekte der Vorfilter (4.5%-46.9%) und deutliche Unterschiede zwischen

den Items und der Form der Filter. Zu ihren Ergebnissen gehört, daß der Effekt des Vorfilters von der "Vertrautheit" mit dem Gegenstand abhängt: Je unvertrauter der Gegenstand, desto stärker der Effekt des Filters. Nachdem BISHOP/OLDENDICK/TUCHFARBER(1983:544) "Bildung" als "erklärende" Variable ablehnen ("this crude demographic variable"), bieten sie eine interessante Hypothese an: Liegt bei einem Befragten zu einem Item keine ausgeprägte "Meinung" vor, dann reagiert dieser auf vertraute Schlüsselworte innerhalb des Items und bewertet diese. Ein in diesem Zusammenhang interessantes Ergebnis findet sich bei BISHOP u.a. (1980:203). Bei einer Frage nach einem nicht-existierenden Gesetz ergab sich mit Filterung der Personen ohne Meinung ein Ansteigen geäußerter Meinungen in der höchsten Bildungsgruppe, in einer Bedingung ohne vorherige Ausfilterung lag die höchste Bildungsgruppe niedriger als die anderen Gruppen[26/].

Zusammenfassend soll festgestellt werden, daß die zitierten Arbeiten deutlich zeigen, daß von einer theoretisch befriedigenden Erklärung von "Itemnonresponse" bisher nicht die Rede sein kann. Es lassen sich sowohl Belege für extreme situationale Schwankungen der Itemnonresponsequoten finden, die sich durch die ad-hoc-Erklärungen kaum erfassen lassen, als auch Tendenzen, bestimmte schwache Zusammenhänge immer wieder zu finden, ohne diese erklären zu können: Dies ist vermutlich für jeweils spezielle Problematiken nur durch eine handlungstheoretische Erklärung im Zusammenhang mit der jeweiligen Kerntheorie zu leisten[27/].

2.4 Konsequenzen in der Forschungspraxis

In der sozialwissenschaftlichen Forschungspraxis ist die Zusammenfassung aller "nicht substantiellen Anworten", seien es explizite Verweigerungen oder "Weiß-nicht"-Antworten, zur "Missing data"-Kategorie üblich[28/]. Obgleich diese Praxis in der Literatur häufiger kritisiert wurde (z.B. FERBER 1956:583, ZEISEL 1970:57, LEVERKUS-BRÜNING 1966:22-30, ESSER 1975:278), hat diese Kritik bis heute (selbst in der Forschungspraxis von Methodologen) kaum sichtbare

Konsequenzen gezeigt. Die Qualifizierung aller nicht-substantiellen Anworten als "missing data" hat neben dem Aspekt eines nicht beabsichtigten Datenverlustes (nicht alle nicht- substantiellen Antworten geben keine Information) durch den Ausschluß der "missing data" aus der Datenanalyse einen Datenselektionseffekt. Diesen Selektionsaspekt des Problems betont ESSER (1975:278) deutlich:

"In der Auswertung wird die 'Restkategorie' meist - insbesondere bei Korrelationsrechnungen - nicht berück sichtigt. 'Meinungslosigkeit' wirkt so wie ein Ausfall mit ähnlichen Konsequenzen für die faktisch verwertete Datenbasis: Meinungslosigkeit verzerrt die Ergebnisse nach Maßgabe ihrer Nicht-Zufälligkeit; von einem unsystematischen Auftreten von Meinungslosigkeit kann jedoch - nach allen bekannten Ergebnissen - nicht die Rede sein."

Die Konsequenzen seien noch einmal stärker hervorgehoben: Durch den Selektionseffekt von "Meinungslosigkeit" wird bei der Datenanalyse die Annahme des zufälligen Ausfalls trotz gegenteiliger Datenlage aufrechterhalten, falls nicht explizit eine Neudefinition der Population, für die Verallgemeinerungen gelten sollen, erfolgt. Dies wäre nur dann vertretbar, wenn das Ausmaß der "WN"-Antworten so gering wäre, daß unter keinen Umständen ein "Effekt" auf die substantiellen theoretischen Beziehungen zu erwarten ist. Daß eben dies in vielen Fällen in der sozialwissenschaftlichen Forschung nicht der Fall ist, zeigt z.B. die Demonstration des Effektes der Annahmen über "Don't know"-Antworten bei RAPOPORT (1979:811). Neben dem üblichen Ausschluß der "WN" berechnete er die Stabilität einer Reihe von politischen Einstellungen in einem Panel nach Zuweisung von Zahlen aus einem Gleichverteilungszufallszahlengenerator und nach Zuweisung gemäß der Verteilung der vorhandenen Werte zu den Fällen mit "WN"[29/]. Die Tau-b Koeffizienten zeigten durchgängig einen erheblichen Rückgang, weiterhin stützten die Daten damit eine Alternativhypothese.

Mag die Abschätzung der Effekte bei einfachen bivariaten Analysen noch möglich sein, so gilt dies für komplexere Analyseverfahren kaum noch: Ob ein gutangepaßtes Modell bei Annahme anderer als zufälliger Ausfälle bei 5% Ausfällen pro Variable noch als gutangepaßt bezeichnet werden kann, dürfte sich einer ad-hoc Einschätzung entziehen.

Damit stellt sich das Problem der Abschätzung der möglichen Effekte unterschiedlicher Annahmen über den Ausfallprozeß. Auf Möglichkeiten dieser Tests und evtl. "Korrekturen" wird nach der Erörterung eines anderen Ursachenkomplexes "fehlender Werte" zurückgekommen.

Anmerkungen zu Kapitel 2

1. Genauer: Das mündliche, persönliche ("face to face"), standardisierte Einzel-Interview, vgl. FREY (1983:55)

2. In der gesamten vorliegenden Arbeit wird "empirische Sozialforschung" immer synonym mit "quantitativ nomothetisch orientierte Sozialforschung" benutzt: Die immer unterstellte Erhebungsform sind Surveys mit analytischer Orientierung, in denen Interesse an "Punktschätzungen" nur am Rande besteht und die hauptsächliche Intention in der Überprüfung sozialwissenschaftlicher Theorien liegt.

3. Mit der weiteren Verbreitung des Telefons und "rückkoppelungsfähiger neuer Medien" wird sich, wie teilweise z.Z. in den USA in bezug auf das Telefoninterview, möglicherweise eine Verschiebung ergeben.

4. Vgl. auch PREGIBON (1978:492), der zusätzlich MD auf Grund der Dateneditierung erwähnt. Diese "blanks" sind Gegenstand des Kapitels 3.

5. Das Problem der "Weiß nicht"-Antworten ist nur sehr begrenzt als MD-Problem anzusprechen; eine genauere Einschätzung wird an anderer Stelle gegeben.

6. Es sind zwar Bedingungen konstruierbar, unter denen verzerrte Randverteilungen unverzerrte Zusammenhänge bewahren (vgl. z.B. ESSER 1983:22-25), siehe dazu aber Kapitel 4.

7. Hierbei handelt es sich lediglich um eine Vereinfachung zur Darstellung des Prozesses; die Aufnahme zusätzlicher Akteure zur Erklärung von "Anwesenheitseffekten Dritter" ist unproblematisch. Vgl. hierzu - weitgehend ohne theoretischen Hintergrund und der klassischen Fehlermethodologie des Interviews verpflichtet - BRADBURN/SUDMAN(1981:134-146) und REUBAND (1984).

8. Zur kognitiven Handlungstheorie vgl. vor allem ATKINSON(1975), ferner LANGENHEDER(1975), ESSER(1980)

9. Zur Kritik der Einwände vgl. OPP(1979:82-83)

10. Zumeist wird zwischen Widerständen (als interne "Barrieren") und "externen" Barrieren unterschieden, was hier aber nicht geschehen soll.

11. Vgl. dazu BATESON (1984:104-115), der in bezug auf den Informanten "understanding", "ability" und "willingness" unterscheidet, ohne sich explizit auf eine Handlungstheorie zu beziehen.

12. Wobei einige Erklärungsmuster bei DILLMAN (1978:14) nahezu tautologisch anmuten: "The fact that some people enjoy answering questionaires regardless of content may explain why a certain percentage of response can be obtained on almost any topic."

13. Vgl. in diesem Zusammenhang auch GROVES/KAHN (1979:228): "If the norms of openess and courtesy between strangers are not enough to carry the interview, the question of benefits arises." Die Autoren betrachten die Akzeptanz allgemeiner Ziele als wesentliche Voraussetzung der Durchführung von Surveys. Die im Text geäußerte Kritik trifft auch hier.

14. Der Aspekt der "Legitimität" einer Frage hat in der Literatur bisher kaum Beachtung gefunden. Gemeint ist der Versuch des Befragten, einen für ihn nachvollziehbaren Zusammenhang einzelner Fragen zum von ihm vermuteten Befragungsziel herzustellen. Gelingt dies nicht, so kann die "Legitimität" der Fragestellung bezweifelt werden, m.a.W. das Bedürfnis nach einer stabilen Kognition kann wichtiger werden als die "Beantwortungsnorm".

15. Vgl. ferner den Exkurs in Kapitel 5.

16. Vgl. dazu allerdings SCHUMAN/PRESSER (1981:170). Sie können den Effekt zwar in 13 von 16 Vergleichsuntersuchungen nachweisen, die Veränderung des Anteils liegt aber im Mittel bei nur 2%. Sie betonen, daß es neben der Interpretation, daß "Weiß-nicht"-Antwortende eine Mittelposition gewählt hätten - wäre diese verfügbar gewesen - die umgekehrte Interpretation, also höhere "WN"- Anteile durch die Zusatzkategorie, möglich ist.

17. Die wichtigsten Arbeiten hierzu stammen von CONVERSE (1970). Vgl. auch FAULKENBERRY/MASON (1978:533) und SCHUMAN/PRESSER(1981: 147-160).

18. Eine Möglichkeit, diese Fehlerquelle durch Ausschaltung mehrdeutiger Items mit einer Skalogrammanalyse anzugehen, zeigen COOMBS/COOMBS (1976).

19. Einen Überblick über neuere Arbeiten zu den "Effekten" der Frageform geben KALTON/SCHUMAN (1980), SCHUMAN/PRESSER (1981), SUDMAN/BRADBURN (1983) und ausführlich MOLENAAR(1982).

20. Solche Hypostasierungen sind in der Literatur häufig zu finden, z.B. bei FAULKENBERRY/MASON (1978) oder in etwas anderem Zusammenhang bei INNES (1977), LAGA/LAUFFER (1980:234) und HATTIE (1983), die die "Neigung", Meinungsfragen nicht zu beantworten, ("Meinungslosigkeit" bei LAGER/LAUFER) als Response-Set betrachten.

21. Vgl. z.B. den nahezu unstrukturierten Hypothesenkatalog über eine Vielzahl von Fragen im Zusammenhang mit dem Foschungsinterview bei VAN DER ZOUWEN/DIJKSTRA (1982:224-234).

22. In Hinsicht auf nicht gestellte Fragen vgl. BRENNER (1981: 140-141), BRENNER (1982:150-152), HAGENAAR/HEINEN (1982:103) sowie DeLAMATER (1982:34), die die Vermeidung von Fragen erwähnen, welche zu vielen Nachfragen führen.

23. Vgl. dagegen die auf Grund der studentischen Stichprobe und des benutzten Instrumentes (einem 1-2 stündigen Persönlichkeitsfragebogen) nur sehr begrenzt verallgemeinerbaren Arbeit von HATTIE (1983: 1044), der lediglich eine starke Korrelation (r=.52) zwischen Position des Items und Zahl der "Omits" zeigen konnte.

24. Zu "unangenehmen Fragen" vgl. auch KOOLWIJK(1969) , BRADBURN u.a. (1978) sowie den Sammelband von BRADBURN/SUDMAN (1981).

25. Zum Zusammmenhang von Bildung, subjektiver Bedeutsamkeit eines Items und Itemnonresponse vgl. auch SMITH (1982).

26. Vgl. ferner HAWKINS/CONEY (1981).

27. So ist z.B. das Ergebnis von BISHOP u.a.(1980:206), daß Personen mit hohem "interpersonal trust" weniger häufig vorgeben zu einem nicht-existierenden Gegenstand eine Meinung zu besitzen, handlungstheoretisch erklärbar, ohne auf komplizierte Selbst-konzept-Erklärungen zurückgreifen zu müssen.

28. Ein willkürlich herausgegriffenes Beispiel für die Gleichsetzung von "Don't know" mit "Missing" findet sich bei SCHMIDT (1982:240-241,250-252). Der Anteil "fehlender Werte" schwankt bei den dort angegebenen Variablen zwischen .2% und 5.8%. SCHMIDT vermutet zwar, daß die "missing cases" nicht zufällig verteilt sind und damit möglicherweise einen Effekt auf die Parameterschätzungen im LISREL-Modell besitzen (S.241), aber eine Trennung der "weiß nicht"-Antworten von anderen "Missing Data" wird nicht vorgenommen, die Durchführbarkeit empirischer Tests der Einflußstärke dieses Effekts wird verneint.

29. Diese Art der Ersetzung ruft allerdings methodische Bedenken hervor. Insbesondere die Zuweisung der Zufallszahlen zu den "WN" stellt nicht das schlechtest denkbare Ausfallmodell dar, vgl. hierzu Kapitel 9.

3. Ausfallursachen II : Datenerfassung und Dateneditierung

"If there are any ways in which data can be lost, they will be."

COHEN/COHEN (1975:265)

Üblicherweise wird der Prozeß, bei dem nach Abschluß einer Datenerhebung aus dem Datenmaterial ein auswertbarer Datensatz erstellt wird, als Datenerfassung und Datenbereinigung bezeichnet. Da zu diesen Arbeitsschritten in der Regel auch die Erstellung eines Grundbestandes an Indizes gehört, seien alle diese Schritte hier zusammenfassend als "Datenmanagement" bezeichnet[1/]. Fehlende Werte bedingen zum einen Schwierigkeiten beim Datenmanagement, zum anderen stellt das Datenmanagement selbst einen der Mechanismen dar, die fehlende Werte produzieren.

Datenerfassung im Rahmen sozialwissenschaftlicher Forschung erfolgt z.Z. in der Regel noch mit der traditionellen, an der Verwendung von Lochkarten orientierten Methode, d.h. der Prozess besteht aus der Kodierung vorliegender Fragebögen in rein numerische Codes, die zuerst auf entsprechende Erfassungsbögen und dann auf ein maschinenlesbares Medium übertragen werden[2/]. Unter der Annahme dieser Vorgehensweise sollen zunächst einige mögliche Ursachen für die Entstehung fehlender Werte bei der Codierung kurz erwähnt werden.

3.1 Codierfehler

Die Prozesse, bei denen MD durch die Verkodung entstehen können, werden hier auf einen potentiell systematischen Mechanismus beschränkt[3/].

Die wohl wichtigste Ursache für die Entstehung systematisch fehlender Werte bei der Codierung ist im Auftreten unvorhergesehener Merkmalsausprägungen zu sehen. Jeder Konstruktion eines Erhebungsintrumentes liegen Annahmen über

die Realität zugrunde, die falsch sein können. Dementsprechend können bei der Konstruktion des Instrumentes unvollständige Klassifikationen (BATESON 1984: 82-90) und falsch geplante Filterführungen entstehen.

Unvollständige Klassifikationen und unvorhergesehene oder "unwahrscheinliche" Merkmalskombinationen lassen zusammen mit den notwendigerweise immer unvollständigen Anweisungen an die Kodiererhier einen Interpretationsspielraum für die Kodierer.BATESON (1984:126) erwähnt drei Möglichkeiten der Reaktion der Kodierer: Zuweisen eines MD-Codes, "coding by fiat" und "classification by fiat". "Classification by fiat" bezieht sich auf die dokumentierte Zuweisung bestimmter Werte nach feststehenden Regeln, "coding by fiat" auf die nichtdokumentierte Zuweisung nach nichtfestgelegten Regeln. Da die beiden letztgenannten Reaktionen einen formal gültigen Wert produzieren, wird auf sie hier nicht eingegangen. Die einfachste Lösung für die Kodierer besteht in solchen Fällen in der willkürlichen Vergabe eines MD-Codes ("Trifft nicht zu"- Code oder ein "Keine Angabe"-Code, falls überhaupt eine Differenzierung vorgenommen wurde). Die Gefahr systematischer "Ausfälle" dieser Art kann ohne intensive Kontrolle der Kodierer nicht ausgeschlossen werden[4/].

3.2 Datenerfassung

Erfolgt die Dateneingabe auf die oben erwähnte traditionelle Weise, kann - bei strikter Trennung der Kodierung von der Datenerfassung ("Lochung") - eine systematisch mit inhaltlichen Fragestellungen verbundene Erzeugung fehlender Werte mit nahezu völliger Sicherheit ausgeschlossen werden. MD können dann zusätzlich nur durch "Fehllochungen" entstehen, die bei einer unabhängig vorgenommenen vollständigen Verifikation, die allerdings sehr selten erfolgt, relativ sicher entdeckt werden können: SCHWAB/ SITTER (1974:13)[5/] geben 2-4% Fehllochungen als Durchschnittswerte an. MINTON (1969:1265) gibt bei vollständiger Verifikation eine verbleibende Fehlerquote von 4.1% der Fehllochungen in einigen Studien an. Die Gefahr systematischer Erzeugung von MD bei strikter Trennung von der Kodierung ist folglich verschwindend gering.

Datenerfassung wird daher hier nicht als zusätzliche Quelle fehlender Werte behandelt[6/].

3.3 Dateneditierung

> "Although the introduction of errors into the datais almosta certain occurrence, most researchers tend to view discoveryof errors with surprise,deal them with frustration, and blame the for unexpected delays in carrying out their analyses."
>
> KARWEIT/MEYERS (1983:392)

Intensive Datenprüfungen und Datenbereinigungen sind in der regulären empirischen Sozialforschung eher als Ausnahmen zu sehen, die zumeist nur in Großprojekten, Datenarchiven und Studien mit speziellen Designproblemen durchgeführt werden.

In vielen Projekten scheint der Regelfall in der Herstellung der allernotwendigsten Bedingungen für eine Datenanalyse zu bestehen, d.h. die "records" werden auf Vollständigkeit und Abfolge kontrolliert und gegebenenfalls bemerkte unzulässige alphanumerische Zeichen beseitigt (meist durch einen MD-Code). Nach einer Randauszählung werden "unmögliche" Werte in der Regel zusätzlich als "missing" codiert, damit gilt die Datenbereinigung als abgeschlossen. Werden noch zusätzliche Konsistenztests vorgenommen, überwiegt bei weitem ebenfalls die Ersetzung durch einen MD-Code, fällt ein Fall mehrfach auf, so wird der gesamte Fall gelöscht[7/].
Dateneditierung wird allgemein als eine der mühsamsten und unangenehmsten Arbeiten im Rahmen empirischer Forschung angesehen (ROISTACHER 1976:301). Die Beschränkung auf elementare Datenprüfung ist somit aus einer Reihe von Gründen verständlich: So sind z.B. die notwendigen umfangreichen Tests und Korrekturen mit den in der Regel zur Verfügung stehenden Statistikprogrammpaketen nur sehr umständlich durchzuführen und beanspruchen

daher unverhältnismäßig große Zeit- und Personalressourcen. Weiterhin basieren unvollständige Datenprüfungen auf der Annahme, daß die Datenfehler zufälligen Charakter besitzen und die Löschung weniger fehlerhaften Daten kaum einen Einfluß auf die Ergebnisse haben kann. Sowohl das Ausmaß fehlerhafter Daten als auch der notwendige Aufwand zu ihrer Beseitigung werden jedoch zumeist unterschätzt.

In der Literatur finden sich relativ wenige Hinweise auf die Fehlerraten in sozialwissenschaftlichen Datensätzen. ELDER (1979: 6) berichtet von 8-65% der Fälle verschiedener Datensätze, die von einem Editprogramm als fehlerhaft erkannt wurden und INFRATEST (im Methodenbericht zum "Sozio-ökonomischen Panel") von 29.5% durch ein Prüfprogramm entdeckten fehlerhaften Haushalts-fragebögen bzw. 35.6% fehlerhaften Personenfragebögen (SONDERFOR-SCHUNGSBEREICH 3 1985:81). Die Ursachen der Fehler lagen überwiegend in Interviewerfehlern (Filterfehler: 70.7%, bzw. 56.9%) und in Fehlern durch "nicht plausible Angaben" (14.5% bzw. 26.4%). Da neben den bereits besprochenen "Erhebungsfehlern" auch Fehler, die durch das Datenmanagement hervorgerufen werden, möglicherweise systematisch mit inhaltlichen Fragestellungen zusammen-hängen können, besteht die Gefahr systematischer Verzerrungen durch die Löschung fehlerhafter Werte. Dies soll an den wichtigsten Arten von "Datenfehlern" im folgenden kurz gezeigt werden.

ROISTACHER (1976:305) unterscheidet zwischen strukturellen, syntaktischen und semantischen Fehlern. Strukturelle Fehler bestehen aus der Nichtübereinstimmung der dem System übergebenen Datenbeschreibung mit der Datenstruktur. Syntaktische Fehler entsprechen Daten, deren Form nicht mit der korrekten Datenbeschreibung übereinstimmt, beispielsweise alphanumerische Zeichen in rein numerischen Datensätzen. Unter semantischen Fehlern versteht ROISTACHER zum einen Wertebereichsfehler, zum anderen Dateninkonsistenzen ("unconditional semantic" vs. "conditional semantic").

Ein Aspekt des Zusammenhangs zwischen MD und strukturellen Fehlern wird im übernächsten Abschnitt behandelt, der größte Teil struktureller Fehler ist im Zusammenhang mit fehlenden Daten bedeutungslos. Syntaktische Fehler entstehen fast ausnahmslos durch Fehler bei der Datenerfassung. Bei einer regulären Datenprüfung fallen syntaktische Fehler meist auf, ebenso reagieren die meisten Programmpakete mittlerweile auf syntaktische Fehler durch entsprechende Fehlermeldungen und nehmen automatisch die Zuweisung zu einem "System missing"- Code vor[8/] oder brechen die Verarbeitung ab. Systematische Ausfälle durch syntaktische Datenfehler dürften extreme Ausnahmen darstellen. Semantische Fehler in Form von Verstößen gegen Wertebereichsannahmen können in der Regel von den Analysesystemen nicht automatisch bemerkt werden. Zur Erklärung der Entstehung dieser Fehler muß zwischen Variablen mit festgelegtem Wertebereich (geschlossene kategoriale Fragen) und Variablen ohne bei der Konstruktion des Instrumentes festgelegten Wertebereich unterschieden werden. Bei Variablen mit festgelegtem Wertebereich können "auffällige" Daten nur durch Fehler bei der Datenerfassung, durch nichtdokumentierte Codes, die von den Kodierern in Fällen mit unerwarteten Merkmalen hinzugefügt wurden, und durch einfache Fehler bei der Kodierung entstehen. Die beiden letztgenannten Ursachen eröffnen die Möglichkeit eines systematischen Ausfalls, sofern "auffällige" Werte auf einen MD- Code gesetzt werden. Das Problem "auffälliger" Werte bei nicht festgelegtem Wertebereich entspricht dem klassischen univariaten "Ausreißer"-Problem und soll daher nicht weiter erörtert werden (vgl. ANSCOMBE 1972). Die Gefahr systematischer Verzerrung durch Löschung ist offensichtlich.

Ein weiteres Problem im Zusammenhang mit der Löschung auffälliger Daten sind Fälle, bei denen Inkonsistenzen bemerkt werden. In der Praxis werden zwar in der Regel kaum umfangreiche Konsistenzprüfungen vorgenommen, trotzdem soll auf die Möglichkeit der Verzerrung durch eine nach der Entdeckung der Inkonsistenzen vorgenommene Löschung aufmerksam gemacht werden.

Neben den sonstigen Fehlerquellen, die bisher erörtert wurden, können systematische Ursachen für die beobachteten Inkonsistenzen nicht ausgeschlossen werden. Neben systematischen Fehlern der Interviewer und Kodierer sind sowohl systematische nichtintendierte Fehler der Befragten (Erinnerungsfehler u.a.) als auch intendierte Fehler denkbar: Erscheint den Befragten die "Beantwortungsnorm" während des Interviews zwingend, so kann durch inkonsistente Angaben eine "verdeckte" Verweigerung erfolgen. Die Löschung einzelner Werte oder des gesamten Falls kann unter diesen Bedingungen systematische Verzerrungen der Ergebnisse hervorbringen[9/].

3.3.1 Exkurs I: Multivariate "Outlier" und MD-Probleme

> "Whenoutliers are present,one has either statistical problems if they are not eliminated, or moral problems if they are..."
>
> TABACHNICK/FIDELL (1983:74)

In den seltensten Fällen erfolgt vor Beginn einer Datenanalyse eine Suche nach Fällen mit extremen Merkmalsausprägungen, die über den univariaten "Wertebereichstest" hinausgeht. Wird während der Datenanalyse eine solche Anomalie offensichtlich, erfolgt in der Regel die Löschung des entsprechenden Falls oder gelegentlich einzelner Variablen. Neben den sonstigen in der Lehrbuchliteratur dargestellten Möglichkeiten, "Outlier" zu handhaben (erneute Datenprüfung, Datentransformation, Recodierung auf Maximalwerte), existiert zumindest eine alternative Technik, die angewandt werden kann, falls die Annahme, daß es sich um "zufällige" Meß- bzw. Datenfehler handelt, begründet erscheint. Zusätzlich illustriert das Verfahren zum einen den Einsatz von MD-Ersetzungstechniken bei einem speziellen MD-Problem, zum anderen die enge Verbindung von "Edit" und "Imputation" (Daten-"Bearbeitung" und "Datenersetzung").

Die meisten Verfahren zur Entdeckung multivariater "Outlier", wie z.B. die Verfahren von Wilks, Siotani und Hawkins[10/] benötigen eine Schätzung der Kovarianzmatrix. Im Falle fehlender Werte ergibt sich das Problem der Schätzung der Kovarianzmatrix aus einem Datensatz mit fehlenden und evtl. verzerrten Werten. In diesem Zusammenhang stellen LITTLE/SMITH (1983:519) eine interessante Anwendung eines iterativen MD-Ersetzungsverfahrens dar, bei dem die Schätzung fehlender Werte über multiple Regressionen erfolgt (eine Form des EM-Algorithmus, vgl. Kapitel 5.1.2). Im ersten Schritt des Verfahrens wird der EM-Algorithmus durchgeführt, als Ergebnis liegt eine Schätzung der Kovarianzmatrix und die mit multiplen Regressionen "vervollständigte" Datenmatrix vor. Dann erfolgt die Bestimmung der "outlying cases" über die Berechnung der Mahalanobis-Distanz. LITTLE/SMITH weisen alle Fälle mit einem p-Wert kleiner als .01 als "Outlier" zurück[11/]. Die verbleibenden Fälle werden für eine neue Schätzung der Kovarianzmatrix benutzt. Für jeden Fall, der als "Outlier" identifiziert wurde, wird dann ein Set von Variablen ermittelt, das die große Mahalanobis-Distanz des Falles bedingt. Die Ermittlung der "einflußreichen" Variablen erfolgt durch schrittweisen Ausschluß derjenigen Variablen, die am stärksten in die Distanz eingehen. Alle Variablen des Sets werden für den aktuellen Fall als MD betrachtet. Nachdem alle Fälle bearbeitet wurden, werden die MD erneut über den EM-Algorithmus ersetzt und die letzte Schätzung der Kovarianzmatrix berechnet.

Das Verfahren setzt neben den oben erwähnten Annahmen über die Entstehung der "outlier" multivariate Normalverteilung voraus, was seiner Anwendung für sozialwissenschaftliche Datensätze Grenzen setzt. Gegenüber den sonstigen Verfahren weist es aber eine Reihe von Vorzügen auf (z.B. MD-Schätzung, keine Plot-Technik).

3.4 MD durch Datenverarbeitung: "Processing errors"

> "Elektronische Datenverarbeitung ist ein sehr komplizierter Vorgang..."
>
> ALEMANN (1977:121)

Die Hauptursachen für die (oft nicht einmal bemerkte) Entstehung fehlender Werte während einer Datenanalyse sind in der Neukonstruktion von Indizes zu suchen, aber auch in der Nichtbeachtung der Voreinstellungen der Statistikprogrammpakete. Im Zusammenhang mit bestimmten Voreinstellungen einiger Programmpakete können Probleme z.B. durch die Datentransformationsprogramme entstehen. Beispielsweise war in SPSS bis zur Version 9 jeder nicht explizit deklarierte Wert bei der Initialisierung einer neuen Variablen gültig, was gelegentlich dazu führte, daß nach einer Datentransformation keine MD mehr existierten (WILKE 1983: 234). Mit SPSS-X führt die Initialisierung einer neuen Variablen zunächst zur Zuweisung des "System missing"-Codes und entspricht damit den Voreinstellungen der meisten anderen Pakete: PSTAT, SAS, GENSTAT und BMDP (vgl. SLUIS 1981:62). Diese Verfahrensweise führt dazu, daß nach nicht völlig durchdachten Datentransformationen nur Bruchteile des Samples über gültige Werte verfügen. Gerade durch den "automatischen" Ausschluß von Wertekombinationen, an die der Anwender nicht dachte, entstehen damit Möglichkeiten zu systematisch fehlenden Werten.

Ein weiteres Problem bilden die sich akkumulierenden MD-Anteile, bei den in der Regel im Laufe der Datenanalyse immer weiter verschachtelten Indizes. Da weder SPSS noch BMDP (im Gegensatz zu SAS) als Standardpakete genauso wenig wie die Anwender eine automatische Dokumentierung der Datentransformationen vornehmen, läßt sich gelegentlich durch den "automatischen" Ausschluß "fehlender Werte" kaum noch rekonstruieren, für welchen Teil einer ehemals großen Stichprobe eine Analyse gerechnet wurde.

Von den vielen Fehlermöglichkeiten struktureller Daten ehler erscheint hier nur eine von Interesse zu sein: Die Angabe eines MD-Codes in der Datenbeschreibung, der einem gültigen Wert entspricht. Fehler dieser Art entstehen in der Regel durch die Benutzung impliziter Zuweisungen ("TO"-Konventionen), insbesondere falls der Code "0" sowohl für fehlende Werte als auch für die Ausprägung eines Merkmals benutzt wird. "Ausfälle" dieser Art sind systematisch und nachträglich nicht nachzuweisen.

Die gängigen Programmpakete besitzen eine Reihe weiterer Eigenheiten bei der Behandlung fehlender Werte, deren einzelne Auflistung hier zu weit führen würde. Die Gefahr erheblicher Verzerrungen durch den Ausschluß von Fällen mit nicht erwarteten Wertekombinationen ist insbesondere unter den Rahmenbedingungen, unter denen Datenanalysen in der empirischen Sozialforschung stattfinden, ständig gegeben. Dies gilt vornehmlich dann, wenn auf Grund der Zeitersparnis auf die Dokumentation jeder Datenmodifikation und die vollständige Protokollierung eines Analyselaufs verzichtet wird.

Ohne dies durch Daten belegen zu können, sei die Vermutung geäußert, daß "processing errors" weit häufiger Artefakte produzieren als jeder andere Prozeß bei der Datenanalyse und weitaus schwieriger nachzuweisen sind als andere Artefaktquellen[12/].

3.5 Zusammenfassung

Dieses Kapitel mußte eine Reihe von wichtigen Aspekten im Zusammenhang von MD und dem Datenmanagement vernachlässigen. Vor allem anderen ist die enge Verknüpfung der MD-Techniken und der Dateneditierung, die insbesondere von den Statistikern von Statistics Canada in ihren Arbeiten im Anschluß an FELLEGI/HOLT immer wieder betont worden ist, zu nennen (G.SANDE 1979, I.G. SANDE 1982). Die Entwicklung statistischer Modelle, die diese Verbindung berücksichtigen (BIEMER 1980, PLATEK/GRAY 1983) steckt noch in den Anfängen, wenn es auch in Hinsicht auf einige Anwendungen integrierte Ansätze

gibt (FREUND/HARTLEY 1967, LITTLE/SMITH 1983). Eine brauchbare Theorie der "Datenkonstruktion", die auch sozialwissenschaftliche Theorieelemente mit einbezieht, fehlt bisher (BATESON 1984).

Für die folgenden Kapitel bleibt festzuhalten, daß jede Phase der Datenerhebung und der Datenanalyse fehlende Werte produzieren kann. Die entstehenden MD können fast immer auch systematische Ursachen besitzen, die mit inhaltlichen Fragestellungen zusammenhängen könnten. Die übliche Praxis des Datenmanagements ignoriert nahezu immer die in diesem Kapitel erwähnten Quellen zusätzlicher fehlender Werte. Dieser Praxis liegen zwei ungeprüfte Annahmen zugrunde:

- das Ausmaß der Fehler sei zu gering, um einen Einfluß auf die Ergebnisse zu besitzen,
- und der Einfluß der Datenfehler erfolge unsystematisch.

Ohne explizite Prüfung beider Annahmen (und deren Dokumentation) für ein gegebenes Problem können systematische Verzerrungen durch "fehlende" Werte allerdings nicht ausgeschlossen werden.

Anmerkungen zu Kapitel 3

1. WILKE(1983:211) zählt zum Datenmanagement auch die Änderung von Datenstrukturen. Der Entstehung von MD durch die Änderung von Datenstrukturen soll im folgenden nicht nachgegangen werden, da es sich entweder um Daten handelt, die ursprünglich nicht erhoben werden sollten, oder um MD, die durch Programmier- oder Konzeptfehler entstehen.

2. Die Entwicklung neuer Erhebungstechniken im Zusammenhang mit anderen Dateneingabemöglichkeiten scheint geeignet, Fehlerqouten in diesem Zusammenhang zu senken. Neben der Benutzung "intelligenter" Terminals, optischer und magnetischer Lesegeräte usw. ist vor allem CATI ("Computer Assisted Telephone Interviewing") zu nennen (vgl. KÜCHLER 1982, FREY 1983 und die gesamte Ausgabe von SMR 12,2,1983). Durch neue Techniken entstehen auch immer neue Fehlerquellen - ein Aspekt, der hier nicht weiter verfolgt werden soll. Zu den "neueren" Techniken allgemein: KARWEIT/MEYERS (1983).

3. Bewußte Falschvercodung (vgl. ESSER 1983:47-50), "skip-" und "digit preference"- Fehler u.ä. (vgl.NAUS 1975:132-136) bleiben hier unbeachtet. Vgl. zu fast allen hier nicht beachteten Fehlerquellen die ausführliche, wenn auch veraltete Abhandlung von ZARKOVICH(1966). Zu den wenigen Arbeiten zur Reliabilität und Validität der Kodierung, vor allem offener Fragen, siehe MORRISSEY (1974) und KALTON/ STOWELL (1979).

4. Zur Kontrolle von Codern vgl. BATESON (1984:125-127).

5. zitiert nach KARWEIT/MEYERS (1983:388).

6. Zu Fehlern bei der Datenerfassung vgl. die Literatur im nächsten Abschnitt und die Arbeiten von DEMING(1942) und MINTON(1969).

7. Die Trennung von Datenerhebungspersonal und Datenanalytikern, die sich z.B. bei der großen Zahl von Studien zeigt, bei denen bis zum "bereinigten Magnetband" alle Arbeiten z.B. von außeruniversitären Institutionen durchgeführt werden, wirft eine Reihe von Problemen im Zusammenhang mit fehlenden Werten, insbesondere solchen, die auf konzeptionellen Fehlern beruhen, auf. Da dies überwiegend zu einer "Soziologie der empirischen Sozialforschung" gehört(die es trotz des Titels der Arbeit von KREUTZ 1972 bisher nicht gibt) sollen solche Probleme hier ausgeklammert werden.
Der sich andeutenden Tendenz zur Auslagerung des Datenmanagements an spezielle Institute entspricht die Literaturlage: Zu den praktischen Problemen des Datenmanagements existiert bisher kaum deutschsprachige Lehrbuchliteratur (WILKE 1983:213). Die meisten Verfasser beschränken sich auf sehr kurze Hinweise (ALLERBECK 1972:51-53;72-78, 1977:171-175, ALEMAN 1977:108-116, HOLM 1975). In der darüber hinaus verfügbaren Literatur werden diese Aspekte stärker betont (SONQUIST/DUNKELBERG 1977, PAYNE 1978, WARWICK/ LININGER 1975). Keine der genannten Arbeiten entspricht allerdings den mittlerweile verfügbaren Techniken. Arbeiten, die auf die veränderten Möglichkeiten der Rechenanlagen eingehen, existieren bisher (mit Ausnahme von KARWEIT/MEYERS 1983) nicht. Im Gegensatz zu den zahlreichen Arbeiten amtlicher Statistiker (z.B. HILL 1978) zeigen sich in der sozialwissenschaftlichen Methodenliteratur auch kaum Ansätze zu einer Rezeption der "technischen" Literatur zur Dateneditierung wie z.B. NORDBOTTEN (1963), VARLEY (1969), NAUS/JOHNSON/MONTALVO (1972), FREUND/HARTLEY (1967) und vor allem FELLEGI/HOLT (1976) sowie LIEPINS (1980;1983). Eine ausgearbeitete Theorie der "Datenkonstruktion" existiert bisher nur in Ansätzen (BATESON 1984).

8. Probleme ergeben sich z.Z. noch durch uneinheitliche Reaktionen auf "blanks": SPSS-9 in der CDC-Version erkennt ein Blank als "-0" (!), andere Pakete als selbstständigen Wert oder automatisch als fehlenden Wert, vgl. hierzu z.B. SLUIS (1981:61-62) und WILKE (1983:232-241).

9. Zur Durchführung von Konsistenztests mit Standardpaketen vgl. ROISTACHER (1976:311-318), vgl. auch SONQUIST/DUNKELBERG (1977: 215-230, LANSING/MORGAN 1971:236-237. Das Problem inkonsistenter Angaben ist in der sozialwissenschaftlichen Literatur zu Interviewproblemen bisher weitgehend vernachlässigt worden.

10. Eine ausgezeichnete Darstellung der "Outlier"-Problematik (und der genannten Verfahren) findet sich zusammen mit einer ausführlichen Bibliographie bei BECKMANN/COOK (1983). Eine formalere Darstellung gibt BARNETT (1983).

11. Sie verwenden eine für die MD korrigierte Form der Prüfgröße, die nicht theoretisch begründet ist.

12. Vgl. KRIZ (1981). KRIZ betont zwar, daß es neben der falschen Anwendung statistischer Konzepte andere "Artefakte bei der Datenauswertung" gibt (KRIZ 1981:75), und kritisiert zu Recht diese Einengung, aber eine weiterführende Diskussion findet sich nicht. ROISTACHER (1976:306) erwähnt hingegen beispielsweise explizit die gröbste Fehlerquelle: Eine methodisch einwandfreie Verarbeitung des falschen Datensatzes. Wesentlich häufiger als die Verwechslung der Datensätze dürften Fehler durch übersehene Selektionsanweisungen innerhalb eines umfangreichen "Jobs" sein.

4. Korrekturverfahren bei systematischen Ausfällen

In diesem Kapitel werden zwei Ansätze zur Korrektur von auf Grund von Ausfällen verzerrten Schätzern der Variablenzusammenhänge dargestellt. Im Gegensatz zu den meisten anderen Verfahren, die in dieser Arbeit erörtert werden, beziehen sich die Methoden auf die Korrektur der Parameter und nicht auf die "Rekonstruktion" der Rohdaten (vgl. hierzu Kap. 5).

Das erste Verfahren, Korrekturformeln für Korrelationskoeffizienten bei Beschränkungen des Wertebereichs ("range-restriction"- Probleme), wird meist als ein Problem intentional fehlender Werte ("missing-by-design") aufgefaßt. Allerdings sind erstens (unter sehr restriktiven Annahmen) ex-post-Anwendungen denkbar, und zweitens lassen sich die Effekte selektiven Ausfalls unter dem "worst reasonable model", nämlich der Annahme, daß alle Ausfälle oberhalb oder unterhalb eines Schwellenwertes liegen, abschätzen.

Das zweite Verfahren, die Korrektur von OLS-Regressionen durch die Aufnahme einer rekonstruierten, den Ausfall determinierenden Variablen, ist aus mehreren Gründen in diesem Zusammenhang relevant. Neben seiner konzeptionellen Nähe zu speziellen Formen der range-restriction-Korrektur wird hiermit ein exaktes, nicht auf der MAR-Annahme beruhendes, Modell des Ausfalls aufgestellt. Darüber hinaus ergeben sich bedeutsame Implikationen für sozialwissenschaftliche Untersuchungen im allgemeinen (BERK 1983).

4.1 "Restriction of Range" - Korrekturen

Das "Restriction of Range"-Problem entsteht in folgender Situation: In einem größeren Sample werden eine Reihe von Tests erhoben. Auf Grund der Ergebnisse eines dieser Tests werden alle Probanden selektiert, die einen bestimmten Schwellenwert über- bzw. unterschreiten. Für weitere Analysen steht nunmehr nur noch das kleinere (restricted) Sample zur Verfügung[1/]. Probleme ergeben sich,

wenn weiterhin Interesse an den Parametern des "unrestricted" Samples besteht, aber nur Daten für das "restricted" Sample vorhanden sind.

4.1.1 Terminologie

Im folgenden wird diejenige Variable, die zur Entscheidung für die Selektion benutzt wurde, als explizite Selektionsvariable X bezeichnet. Eine mit X korrelierte Variable Y erfährt durch die Selektion ebenfalls eine Veränderung (incidential selection)[2/]. Eine mit X und Y korrelierte Kriteriumsvariable wird mit Z bezeichnet. Mit kleinen Buchstaben bezeichnete Größen beziehen sich auf das restricted Sample, so z.B. ist sy die Standardabweichung der Variablen y im restricted Sample. Die Korrelation im unrestricted Sample wird durch rho (bzw. rXY), Pearson's Korrelationskoeffizient mit r und ein korrigierter Korrelationskoeffizient mit R benannt. Der Quotient sX/sx wird als Selektionsratio, die Stärke der Selektion als lambda bezeichnet, wobei 1- lambda den Anteil des noch verbleibenden Samples darstellt.

4.1.2 Folgen der Selektion

Zu den offensichtlichen Folgen der Selektion gehört die Veränderung des Mittelwerts und der Standardabweichung von x. Der Mittelwert eines beliebigen Segmentes einer standardnormalverteilten Variablen läßt sich über

4-1 $$\bar{x} = \frac{f(z_1) - f(z_2)}{q}$$

(RYDBERG 1963:52)

berechnen, wobei q dem Anteil des Segmentes an der gesamten Verteilung und f(z) der Ordinate der Wahrscheinlichkeitsdichtefunktion der Standardnormalverteilung der unteren bzw. oberen Intervallgrenze entspricht.

Im Fall eines restricted Sample mit einem Schwellenwert über 0 vereinfacht sich 4-1 zu

4-2 $$\bar{x} = \frac{f(z)}{1 - F(z)}$$

(RYDBERG 1963:53)

wobei F(z) die Fläche der Standardnormalverteilung im Intervall (-∞,z) bezeichnet.
Die Standardabweichung in einem beliebigen Intervall kann mit

4-3 $$s = \sqrt{1 + \frac{f(z_1)z_1 - f(z_2)z_2}{q} - \left(\frac{f(z_1) - f(z_2)}{q}\right)^2}$$

berechnet werden (RYDBERG 1963:53)[3/].

Die Veränderung in der "incidential selection"-Variablen y hängt von der Höhe der Korrelation im unrestricted Sample ab: Je höher die Korrelation, desto stärker verändert sich y und s. Für standardnormalverteilte Variablen gilt[4/]:

4-4. $E(y) = r_{XY} * E(x)$

4.5 $E(s_y)^2 = r_{XY}^2 E(s_x^2) + 1 - r_{XY}^2$

4-6 $$s_y = \sqrt{r_{XY}^2 \cdot s_x + 1 - r_{XY}^2}$$

(nach JOHNSON/KOTZ 1972:112, Notation verändert; für nicht standardnormalverteilte Variablen: GULLIKSEN 1950:138).

Korrelationen, die mit den Daten des restricted Samples berechnet werden, stellen keine unverzerrten Schätzer von rho dar. Die Korrelation zwischen x und y wird sich immer gegen Null verschieben, dagegen ist im bivariaten Fall ein

Vorzeichenwechsel nicht möglich. Im Dreivariablenfall mit zwei Prädiktoren (X,Y), einer Kriteriumsvariablen (Z) und Selektion mit X sind sowohl Vorzeichenwechsel als auch die Inflationierung von ryz möglich, falls rXY und rXZ hoch sind bzw. die Differenz rXY und rXZ groß ist und rYZ einen bestimmten Wert überschreitet[5/].

4.1.3 Korrekturformeln

PEARSON legte 1903 eine Korrekturformel für den bivariaten Fall vor. Unter der Voraussetzung einer linearen Beziehung zwischen X und Y und der Homoskedastizität der Fehler sowie expliziter Selektion mit X[6/] (GULLIKSEN 1950:137) gilt:

$$4\text{-}7 \qquad R_{xy} = \frac{s_X \cdot r_{xy}}{\sqrt{s_X^2 \cdot r_{xy}^2 + s_x^2 - s_x^2 \cdot r_{xy}^2}}$$

Unter den genannten Voraussetzungen ist die Formel exakt. GROSS/ PERRY (1983:115) machen darauf aufmerksam, daß die Literatur zum "restriction of range"- Problem in einen mathematischen und in einen statistischen Teil gegliedert werden kann. Der mathematische Teil (unter den die meisten der hier zitierten Arbeiten fallen) vernachlässigt alle Stichprobenprobleme[7/]. Auf diese zusätzlichen Probleme wird hier ebenfalls nicht weiter eingegangen.

Der hier geschilderte Fall wird in der Literatur als "case 1" bezeichnet. Es existieren eine Reihe weiterer Formeln sowohl für diesen Fall als auch für den Fall 2 (Selektion mit X, Varianzen für Y bekannt) und für den Fall 3 (Selektion mit Z, Varianzen für Z bekannt) (GUILFORD 1956:320-321). Für den interessanten Fall unbekannter Varianzen, aber bekanntem selection ratio findet sich eine Formel bei SANDS u.a. (1978:748, in Anlehnung an KELLEY 1947); das Problem unbekannter Selektionsvariablen, aber bekannten Y und Z Varianzen

behandeln BRYANT/GOKHALE (1972:307), den Fall der Selektion mittels gewichteter Prädiktorensummen ELSHOUT/ROE (1972:138).

BROUWER/VIJN (1980a,1980b) entwickelten 4 neue Bayes-Schätzer[8/] zur Berechnung von Korrelationskoeffizienten in Stichproben mit MD, die zwarnicht explizit für Rangerestriktionprobleme konzipiert, aber im Rahmen einer Simulationsstudie für solche Probleme eingesetzt wurden[9/]. Ein weiterer Versuch, das Problem mit Bayes- Methoden anzugehen, findet sich bei GROSS/PERRY (1983). Formeln für multivariate Probleme finden sich bei GULLIKSEN(1950: 158- 172) und RYDBERG (1963:87-128)[10/].

4.1.4 Probleme der Anwendung der Korrekturformeln

Die Anwendung der Korrekturverfahren stößt auf eine Reihe von Problemen: Bias[11/] und Verteilungsfunktionen der Schätzer sind unbekannt, die Kenntnis von Grundgesamtheitsparametern wird vorausgesetzt, eine univariate Selektion mit einer bekannten Selektionsvariablen sowie "complete truncation" (alle Fälle, die auf einer Variablen einen bestimmten Schwellenwert überschreiten, werden selektiert) wird angenommen. Einige dieser Probleme sollen im folgenden kurz behandelt werden.

4.1.4.1 Arbeiten zur Verteilungsfunktion von R

Die statistischen Eigenschaften von R bezüglich Bias, Effizienz und Konsistenz sind (analytisch) unbekannt. Weder die Verteilungsfunktion noch Formeln für den Standardfehler sind verfügbar (FORSYTH 1971:116).

R hängt von drei Parametern ab: Stichprobengröße, rho und lambda. Die bei der Herleitung der Verteilungsfunktion entstehenden mathematischen Probleme haben bisher eine analytische Lösung verhindert (GULLICKSON/HOPKINS 1976:10). FORSYTH (1971) unternahm daher Simulationsstudien, um eine empirische

Verteilungsfunktion von R zu bestimmen. Er variierte die Samplegröße (25,50,100) des selektierten Samples, lambda (.9, .75, .5,. 25) und rho (.8, .5). Für jede der 24 Verteilungsfunktionen wurden 1000 Replikationen berechnet. Für R wurden über Fisher-z-Transformationen drei Konfidenzintervalle (.9, .95, .99) berechnet und der Anteil derjenigen Intervalle ermittelt, die den Populationswert enthielten. Der Vergleich mit dem nominalen Vertrauensniveau erlaubt die Abschätzung der Güte der Approximation der unbekannten Verteilungsfunktion von R durch die Verteilungsfunktion von r. Das empirisch ermittelte Vertrauensniveau lag bei allen 24 Verteilungsfunktionen unterhalb des nominalen Niveaus. FORSYTH (1971:121) zeigte drei Trends auf:

- für gegebenes rho und lambda hängt der Fehler nicht von n ab;
- mit zunehmender Selektion steigt der Fehler;
- mit zunehmendem rho sinkt der Fehler.

GULLICKSON/HOPKINS (1976) führten Simulationsstudien für 5 verschiedene Testgrößen der Hypothese rho=0 durch. Die übliche Fisher-z-Transformation gehörte zu denjenigen Formeln, die die größten Fehler in bezug auf das nominale Signifikanzniveau aufwiesen (GULLICKSON/HOPKINS 1976:12). Ihre Simulationen zur Abschätzung von Konfidenzintervallen zeigten deutlich, daß die Zunahme der Selektion eine Zunahme der Varianz von R bedingt. Lediglich falls n sehr groß wird und |R| gegen eins geht, sind die Effekte der Selektion vernachlässigbar[12/].

ELSHOUT u.a. (1979:574f) argumentierten, daß unter den ohnehin notwendigen Annahmen der Linearität der Regression und der Homoskedastizität eine Nullkorrelation im selektierten Sample eine Nullkorrelation in der Population

impliziert, und schlagen daher die einfache t-Statistik

$$4\text{-}8 \qquad t = \frac{r_{xy} \cdot \sqrt{n-2}}{\sqrt{1-r_{xy}^2}}$$

mit n-2 Freiheitsgraden vor.

4.1.4.2 Arbeiten zum Bias von R

Zentral für eine Bewertung der korrigierten Korrelationskoeffizienten ist der Bias, die Differenz zwischen dem korrigierten Korrelationskoeefizienten und dem "wahren" "Populationswert".

Im Rahmen von Simulationsstudien übliche Maße sind das arithmetische Mittel und der Median des Bias über eine Sequenz der Simulation mit konstanten Parametern. GROSS/KAGEN (1983:390) benutzen daneben den mittleren quadratischen Fehler (MQF),

$$4\text{-}9 \qquad MQF = \frac{\sum_{i=1}^{nk} (R - rho)^2}{nk} \qquad nk = \text{Anzahl der Replikationen}$$

der außer dem Bias auch die Stichprobenvarianz des Schätzers berücksichtigt. BOUWER/VIJN(1980:9) fanden zwischen dem MQF und dem mittleren absoluten Fehler (MAF) in Hinsicht auf die Resulate nur sehr geringe Differenzen, so daß sie nur mit dem MAF arbeiten. GREENER/OSBURN (1979,1980) berechneten die Differenzen der Fisher-z-Werte der Korrelationen im restricted und unrestricted Sample.

GREENER/OSBURN (1979) untersuchten das Verhalten von R anhand zweier realer Datensätze, u.a. denen des Projekts TALENT. Interessanterweise erfüllte keine der ausgewählten TALENT-Verteilungen die Annahmen der Linearität und Homoskedastizität. Daher variierten sie nicht nur lambda, sondern auch die Verteilungsform mittels einiger Datentransformationen um Linerarität bzw. Homoskedastizität zu erzielen. Um beide Voraussetzungen zu erfüllen wurden für gegebene x-Werte mit einem Zufallszahlengenerator neue y-Werte erzeugt[13/]. Bei einem unrestricted r < .2 war der korrigierte Wert nicht genauer als der unkorrigierte, allerdings zeigte sich unabhängig von lambda für r eine konsistente Unterschätzung und für R eine Tendenz zur Überschätzung des wahren Wertes. Für .6 < r < .8 war R der bessere Schätzer, und dies umso mehr, je größer lambda. R neigte auch hier zur Überschätzung der unrestricted Korrelation. Der Medianfehler von R wurde um so größer, je höher lambda war, blieb aber niedriger als der Medianfehler von r. Die Verletzung der Linearitätsannahme wirkte sich stärker auf den Medianfehler aus als die Verletzung der Homoskedastizitätsannahme. GREENER/OSBURN (1980) vertieften die Analyse des Verhaltens von R unter Verletzung der Linearitäts- und der Homoskedastizitätsbedingungen im Rahmen einer umfangreichen Simulationsstudie. Bei konstantem n=4000 (!) und über jeweils 5 Replikationen variierten sie in .1 Schritten lambda und rho zwischen 0.1 und 0.9. Die Daten wurden für 7 verschiedene Verteilungen (bivariat-normal, zwei ogivenförmige und zwei spezielle) generiert. Die beiden speziellen Verteilungen ("fan" und "football") wurden in je zwei verschieden Versionen erzeugt[14/]. Die Linearitätsannahme wird am stärksten durch die ogivenförmigen Verteilungen verletzt, "fan"- und "football"-förmige Verteilungen verletzen in unterschiedlicher Weise die Homoskedastizitätsannahme: in den "fan"- Verteilungen steigt die bedingte Varianz in y mit wachsendem x an, die "football"- Verteilungen besitzen die größten bedingten Varianzen um den Mittelwert. Für den Fall aller erfüllten Voraussetzungen war R der bei weitem bessere Schätzer als r, wobei weder rho noch lambda dieses Resultat beeinflußten. In den ogivenförmigen Verteilungen hing der ermittelte Fehler von R nicht von rho, dagegen stark von lambda ab. Bei zunehmender Selektion erbrachte R eine konsistente Unterschätzung von rho, die allerdings immer geringer war als der Fehler von r. In den beiden

"fan"-Verteilungen nahm die Unterschätzung mit steigendem rho und lambda zu, wiederum war der Fehler geringer als für r. Zu einer Überschätzung von rho durch R kam es in den "football"- Verteilungen als Funktion von rho und lambda. Bis zu lambda = .5 zeigt sich eine tendenzielle Unterschätzung mit wesentlich geringeren Fehlern als die Fehler, die durch r entstehen.

Die Simulationsstudie von BROUWER/VIJN (1980a;1980b) erbrachte mit wachsendem rho bei einem lambda von 0.1 bzw. 0.5 in den Fällen 1 und 2 für R kleinere MAF als für r (1980b:27). Für relativ kleines rho war r der bessere Schätzer (1980a:198-199).

GROSS/KAGEN (1983) berechneten in ihrer Simulation Bias und MQF für r und R bei n=50,100,200 ; rho=.3,.5,.7 ; lambda =.1 bis .9 und bivariater Normalverteilung über je 2000 Replikationen. In allen 81 Verteilungen stellen beide Koeffizienten negativ verzerrte Schätzer dar, wobei r immer stärker verzerrte als R. Das Ergebnis der MQF-Berechnung ergab dagegen in 31 der 81 Fälle für r einen kleineren MQF. Die Differenzen liegen in einigen Fällen in der Größenordnung von 100%. Die Resultate für r sind umso besser, je kleiner rho und je extremer die Selektion ist. Der Vorteil gegenüber R geht aber rasch mit wachsendem rho verloren und besteht dann nur noch bei extremer Selektion.

GROSS (1982) versuchte zu zeigen, daß die restriktiven Annahmen, die zur Anwendung der Korrektur notwendig sind, gelockert werden können, falls die von ihm neu entwickelte Größe Q gleich eins ist, wobei Q folgendermaßen definiert ist (GROSS 1982:798):

4-10 $$Q = \sqrt{s_E^2 / s_e^2 /(B/b)} \text{ ; wobei :}$$

$$B = s_{XZ} / s_x^2 \text{ ; } b = s_{xz} / s_x^2 \text{ ;}$$

$$s_E^2 = s_Z^2 \cdot \left(1 - r_{XZ}^2\right) \text{ ; } s_e^2 = \left(1 - r_{xz}^2\right)$$

Falls die Voraussetzungen der Homoskedastizität und Linearität gegeben sind, ist B/b = 1, var(E)/var(e) = 1 und damit Q=1 . Ist Q größer als 1, wird R eine Überschätzung von rXZ darstellen, ist Q kleiner als 1, wird R zu einer Unterschätzung von rXZ führen. Falls vermutet werden kann, daß ein gegebenes Problem der Bedingung Q=1 genügt, ist die Korrektur mit der Pearsonformel korrekt. Wichtiger als die kaum realistische Anwendung von Q zur Überprüfung der Annahmen ist die damit mögliche Einsicht, daß trotz Verletzung beider Annahmen die Korrektur möglich ist, falls Q = 1 durch einen Ausgleich beider Verletzungen angenommen werden kann. Bei dem Vorliegen einer kurvilinearen Beziehung wird in der Regel B größer als b sein. Falls lambda groß ist (nur ein kleines restricted Sample verbleibt), wird var (E) größer als var (e) sein. Diese beiden Faktoren können sich ausgleichen (GROSS 1982: 799), so daß die Korrektur den "wahren" Wert erbringt.

4.1.5 Das Problem der unbekannten Selektionsvariablen

Wie oben bereits erwähnt wurde, stößt die Anwendung der Korrekturformeln auf eine Reihe von Problemen. Das Hauptproblem besteht in der Praxis darin, daß die "wirkliche" Selektionsvariable unbekannt ist (LORD/NOVICK 1968:144). LINN (1968:70) weist auf die Unbekanntheit bzw. Unmeßbarkeit der Selektionsvariablen ins- besondere bei Selbstselektionsprozessen hin. An einem Beispiel zeigt LINN, daß die Benutzung einer "incidential selection"- Variablen als explizite Selektionsvariable zu einer Unterschätzung von rho führt, wobei der Bias allerdings geringer ist als der Bias von r. Der Bias von R stellt eine Funktion des Selektionsratios der "wahren" Selektionsvariablen dar (LINN 1968:71).

GROSS/FLEISCHMAN (1983) untersuchten mit sechs "realen" Datensätzen[15/] und jeweils fünf simulierten Selektionsprozessen den Bias von R unter Verletzung aller Annahmen: Linearität, Homoskedastizität und ausschließlicher Selektion mit einer bekannten Variablen. Sie benutzten zwei Prädiktorvariablen mit einer Inter-korrelation von .6 und eine Kriteriumsvariable. Die Prädiktor-Kriteriumskorrelation lag zwischen .36 und .63. Nicht alle möglichen

Kombinationen der Verletzung der Annahmen waren gegeben. Die unterschiedlichen Selektionsprozesse wurden durch unterschiedlich gewichtete Summen der beiden Prädiktoren und variierendes lambda simuliert. Als Biasmaß wurde der prozentuale Fehler verwendet. Je weniger die zur Korrektur benutzte Variable die Selektionsentscheidung beeinflußte, desto geringer wurde im allgemeinen der Betrag von R. In fast allen Datensätzen war der prozentuale Fehler von R geringer als der Fehler von r. Der Fehler von R war nur bei geringem lambda vertretbar, generell wuchs der Fehler mit der Anzahl fehlender Werte. Weiterhin traten Fälle auf, in denen r den geringeren Bias besaß. Zusammenfassend stellten die Autoren fest, daß R nicht robust gegenüber der Verletzung der Annahmen ist. Falls keine vorherige Information über die Verteilung vorhanden sei, könnten durch den Gebrauch der Korrekturformel größere Fehler entstehen.

4.1.6 Lösungsversuche

LORD/NOVICK (1968:144) zeigen zwar, daß die Selektion nicht explizit über X erfolgen muß, um die Korrektur durchzuführen. In diesem Fall ist aber die Annahme der Linearität nur schwer zu rechtfertigen. Ein Lösungsversuch stamt von MEREDITH (1958)[16/]. Die Voraussetzung für MEREDITH's faktorenanalytischen Ansatz zur Rekonstruktion einer expliziten Selektionsvariablen ist die perfekte multiple Korrelation der Prädiktoren mit der Selektionsvariablen; m.a.W.: Die Selektionsvariable muß eine Linearkombination der Prädiktoren sein (RYDBERG 1963:115). Diese sehr restriktive Annahme ist in der Praxis kaum akzeptabel.
Die Formel von BRYANT/GOKHALE (1972:307)

$$\text{4-11} \qquad R_{XY} = r_{xy} \cdot \sqrt{\frac{s_x^2}{s_X^2} \cdot \frac{s_y^2}{s_Y^2}} + \sqrt{\left(1 - \frac{s_x^2}{s_X^2}\right) \cdot \left(1 - \frac{s_y^2}{s_Y^2}\right)}$$

liefert eine Schätzung von rho unter der Normalverteilungsannahme und der Bekanntheit der Populations- und Samplevarianzen zweier "incidential selection"-Variablen[17/].

Ein sehr interessanter Ansatz findet sich bei ROE (1979). Um das Problem der unbekannten Selektionsvariablen zu vermeiden, wird hier eine Selektionsvariable rekonstruiert. ROE berechnet eine multiple Regression der Prädiktoren mit der dichotomen Variablen "selektiert-nicht selektiert". Die durch die Regression "vorhergesagten" Werte bilden die rekonstruierte Selektionsvariable, die für die Korrektur benutzt wird[18/]. ROE arbeitet mit der Formel von ROE/ELSHOUT(1972):

$$4\text{-}12 \qquad R_{xy} = \frac{\left(r_{xy} - r_{xy} \cdot r_{xz}\right)/H_y + r_{XY} \cdot r_{xz} \cdot H_x}{\sqrt{1 + r_{xz}^2\left(H_x^2 - 1\right)}}$$

(ROE 1979:556), wobei H dem Quotienten der Standardabweichung des unrestricted Samples und des restricted Samples entspricht.

4.1.7 Zusammenfassende Bewertung

Es konnte gezeigt werden, daß die Korrektur der Korrelationskoeffizienten im "range-restriction"-Problem nicht robust gegenüber den Verletzungen der Annahmen ist. Falls vorherige Informationen über die Form der Verteilungen und des Zusammenhangs gegeben sind und eine Variable als Selektorvariable identifiziert werden kann, ist die Anwendung der Korrektur zumindest als Ergänzung des Standardverfahrens (der Berechnung von r) sinnvoll. Im hier interessierenden Zusammenhang der "Korrektur" von Korrelationen mit missing data im allgemeinen sind diese Korrekturen begrenzt als ein Modell des Ausfalls brauchbar. Dieser Versuch wird aber nur durch den Vergleich mit anderen Modellen sinnvoll. Der Mangel dieses speziellen Modells liegt im Fall der Übertragung auf generelle MD-Situationen in der Annahme der "complete

truncation" und den besprochenen Schwierigkeiten mit der in der Regel unbekannten (Selbst-) Selektionsvariablen. Die Versuche zur Lösung des Problems entweder über ein (vermutetes) Selektionsratio (SANDS u.a. 1978) bzw. über das Selektionsratio der "incidential selection"-Variablen (BRYANT/GOKHALE 1972) oder über den Versuch der Rekonstruktion der Selektionsvariablen (ROE 1979) zeigen prinzipiell den Lösungsweg über die Modellbildung an. Das Durchspielen verschiedener Annahmen an einem Datensatz kann so immerhin die Empfindlichkeit der mit den traditionellen Methoden gewonnenen Ergebnisse gegenüber der Verletzung der MAR-Annahme demonstrieren.

Wie sich im folgenden Kapitel zeigen wird, läßt sich das "selection of range"-Problem als Spezialfall systematischer Ausfälle allgemein auffassen. Insbesondere wird sich die Möglichkeit der Spezifizierung von Ausfallmodellen mit nicht beobachteten Schwellenwerten zeigen. Die bisher erwähnten Versuche der Rekonstruktion der Selektorvariablen bilden eine Vorstufe zu den im folgenden Kapitel zu erörternden neueren Korrekturverfahren von HECKMAN.

4.2 Korrekturverfahren für Regressionen

In diesem Kapitel sollen einige der in den letzten Jahren überwiegend von Ökonomen entwickelten Methoden zur Korrektur von OLS-Regressionen bei Problemen der Sampleselektivität dargestellt werden. Obgleich sowohl das Problem systematisch verzerrter Samples als auch die Korrekturmethoden erhebliche Bedeutung für sozialwissenschaftliche Datenerhebungen besitzen, haben die entsprechenden Arbeiten bei Soziologen bisher kaum Beachtung gefunden[19/]. Neben der eigentlichen Anwendung bieten die Korrekturverfahren eine exakte Modellierung der Ausfallprozesse und die Möglichkeit, die im Kapitel 4.1 dargestellten Korrekturen für "restriction of range" -Probleme als Spezialfälle eines umfassenderen Modells aufzufassen.

4.2.1 Auswirkung der Selektion auf Regressionsparameter

Die Darstellung in diesem Abschnitt beschränkt sich auf Populationsparameter[20/].

In einer Population gelte das normale Regressionsmodell

4-13 $y = x' \cdot \beta + u$

wobei multivariate Normalverteilung, serielle Nullkorrelationen von u sowie E(u) = 0 gegeben sei. Eine Subpopulation werde dadurch selektiert, daß alle Fälle außerhalb eines Wertebereichs Y unbeachtet bleiben (explizite Selektion mit y).

Unter diesen Voraussetzungen gilt (GOLDBERGER 1981:359-360):

4-14 $\beta^* = \lambda \cdot \beta$

4-15 $\lambda = \theta / \left(1 - \rho^2 \cdot (1 - \theta)\right)$

4-16 $\alpha^* = \left(1 - \lambda \cdot \rho^2\right) \cdot \mu_y^*$

4-17 $\rho_*^2 = \lambda \cdot \rho^2$

wobei μ_y^* den Mittelwert von y in der selektierten Population, ρ^2 den Determinationskoeffizienten und θ den Quotienten der Varianz von y im selektierten Sample und der Varianz von y in der gesamten Population darstellt. Unter den genannten Voraussetzungen führt die explizite Selektion mit der abhängigen Variablen zu einer proportionalen Veränderung der Regressionskoeffizienten (GOLDBERGER 1981:361)[21/].

Eine spezielle Form der Selektion ergibt sich bei einem fixen Schwellenwert c auf Y, jenseits dessen alle Fälle verloren gehen: Truncation. Dann gilt unter jeglicher Verteilungsform der Regres
soren:

4-18 $E^* \cdot (y \mid x) = x' \cdot \beta - \omega \cdot r(z)$

4-19 $z = (c - x' \cdot \beta)/\omega$

4-20 $r(.) = f(.)/F(.)$, (GOLDBERGER 1981:361)

wobei w die Standardabweichung der Fehler in (4-13), $f(.)$ die Dichtefunktion und $F(.)$ die Verteilungsfunktion der Standardnormalverteilung darstellt. Die neue Variable z stellt lediglich die durch die Standardabweichung der Fehler gewichtete Differenz zwischen dem auf Grund der "conditional expectation function" erwarteten Wert und dem Truncationpunkt c dar. Das Ratio f(.)/F(.) ist eine Hazardfunktion, wie sie z.B. bei der Analyse von "survival data" üblich ist[22/].

Zwei anschauliche Interpretationen der Hazardfunktion bieten sich an (BERK/RAY 1982:362). Zum einen stellt die Hazardfunktion das Risiko des Ausfalls einer Beobachtung dar: das Risiko der Überschreitung des Schwellenwertes. Zum anderen kann das Ratio als der erwartete Wert der Fehler nach der Selektion verstanden werden. Falls das Modell vor der Selektion korrekt spezifiziert wurde, war der Erwartungswert der Fehler gleich Null. Nach der Selektion ist der Erwartungswert ungleich Null. Somit liegt ein Spezifikationsfehler vor: Der erwartete Wert der Fehler bildet einen zusätzlichen, nichtorthogonalen Regressor, der von dem spezifizierten Modell ignoriert wird[23/]. Selbst wenn man bereit sein sollte, die Ergebnisse einer Analyse strikt auf das vorliegende Datenmaterial zu begrenzen, sind die Ergebnisse inkorrekt, da das lineare Modell unangebracht ist.
GOLDBERGER (1981:362) zeigt, daß ohne die Voraussetzung der multivariaten Normalverteilung keine Aussage über die Veränderung der

Regressionskoeffizienten möglich ist. Akzeptiert man hingegen diese Voraussetzung, so folgt wiederum die proportionale Veränderung der Regressionskoeffizienten und in diesem Fall die proportionale Verminderung der Koeffizienten (GOLDBERGER 1981:363).

Wesentlich interessanter als der Fall expliziter Selektion mit der abhängigen Variablen ist der Fall der "incidential selection". Neben der "substantiellen" Gleichung muß nun eine Gleichung für den Selektionsprozeß selbst erstellt werden: Zwei verschiedene Kausalmodelle müssen spezifiziert werden (falls der Selektionsprozeß mit der substantiellen Gleichung modelliert werden kann, handelt es sich um explizite Selektion, vgl. BERK/ RAY 1982:367; HECKMAN 1979:155). Gegeben sei

4-21 $y_1 = x' \cdot \beta_1 + \mu_1; \quad y_2 = x' \cdot \beta_1 + \mu_2$

wobei multivariate Normalverteilung vorausgesetzt wird und die Residuen ebenfalls normalverteilt (mit dem Erwartungswert Null und der gemeinsamen Kovarianzmatrix Omega) sind. Wichtig bleibt festzuhalten, daß es demzufolge eine von Null verschiedene Kovarianz der Fehlerterme gibt.

Das Gleichungssystem stellt zunächst lediglich ein traditionelles "seemingly unrelated equation"-Modell dar[24/]. Es muß noch einmal betont werden, daß beide Gleichungen als in der unselektierten Population korrekt spezifiziert angesehen werden und daß zwischen den Fehlern der beiden Gleichungen eine Kovarianz ungleich Null besteht. Die Selektionsregel sei ($y_2 > c$), wobei c ein fester Truncationpunkt sei. Dann ist der Erwartungswert von y nicht nur eine Funktion der Regressoren, sondern auch eine Funktion einer den Effekt der Selektion modellierenden Hazardrate:

4-22 $E^*(y_1|x) = x' \cdot \beta_1 - (\omega_{12}/\omega_2) \cdot r(z_2),$

wobei

4-23 $z_2 = (c - x' \cdot \beta_2)/\omega_2$

(BERK/RAY 1982:369). Das Ratio der Kovarianz der Fehler zu der Standardabweichung der Fehler in der Selektionsgleichung ist natürlich ein Regressionskoeffizient, der als Gewichtung der Hazardrate dient. Falls die Kovarianz der Fehler gleich Null ist, verschwindet der Einfluß der Hazardrate und damit auch der Selektionseffekt.

GOLDBERGER zeigt für die Regressionsgleichungen[25/]

4-24 $$\beta_1^* = \beta_1 - \phi\left(\omega_{12}/\sigma_2^{\,2}\right)\cdot\beta_2$$

wobei

4-25 $$\phi = (1-\theta_2)/\left(1-\rho_2^{\,2}\cdot(1-\theta_2)\right)$$

und

4-26 $$\theta_2 = 1+\mathrm{r}(\mathrm{c}/\sigma_2)$$

Selbst unter der Annahme multivariater Normalverteilung läßt sich also keine proportionale Veränderung der Regressionsparameter unter "incidential selection" zeigen.

Zusammenfassend läßt sich folgendes festhalten: Explizite Selektion, Truncation und "incidential selection" führen zu einer Veränderung der Regressionsparameter, wobei sich unter der Annahme multivariater Normalverteilung lediglich für Truncation eine proportionale Verminderung der Regressionskoeffizienten zeigen läßt. Die Veränderungen führen in jedem Fall zur ein Fehlspezifi-
zierung des Regressionsmodells, selbst wenn eine ausdrückliche Beschränkung auf die vorliegenden Daten erfolgt.

4.2.2 Korrekturverfahren

GREENE (1981) behandelt die Korrektur einer OLS-Regression für den Fall, daß eine explizite Selektion mit der abhängigen Variablen erfolgte und alle ausgeschlossenen Fälle mit dem als bekannt vorausgesetzten Schwellenwert kodiert wurden ("limited dependent variable model)": das sogenannte Tobit-Modell[26/] nach TOBIN (1958). GREENE kommt zu einem erstaunlich einfachen Resultat. Unter der Annahme multivariater Normalverteilung[27/] und der Beobachtung der Fälle, die oberhalb des Schwellenwertes liegen, wird zur Korrektur lediglich der Anteil der "nonlimit observations" benötigt. Die Korrektur erfolgt durch einfache Divison der OLS- Regressionskoeffizienten durch den Anteilswert der "nonlimit observations"[28/].

Die z.Z. populärste Korrekturmethode bei OLS-Regressionen geht auf die Arbeiten von HECKMAN (1976, 1979) zurück[29/]. HECKMAN (1979:154) geht von einem Gleichungssystem aus, wie es oben (Gleichung 4-21) beschrieben wurde. Fehlen die Daten für den Fall i auf der Variablen y_1 dann, wenn y_{2i} einen Schwellenwert - z.B. Null - unterschreitet, kann, falls y_{2i} unbekannt[30/] ist, eine einfache Korrektur durchgeführt werden. Zunächst muß die Wahrscheinlichkeit für das Überschreiten des Schwellenwertes für das gesamte Sample bestimmt werden. Zwar ist y_{2i} nicht bekannt, aber das Ergebnis des Selektionsprozesses, die entsprechende dichotome Indikatorvariable (fehlend/nicht fehlend). Einen Schätzwert für y_{2i} erhält man durch eine Probit-Analyse[31/] der Indikatorvariablen und der Berechnung der Schätzwerte mit den erhaltenen Parametern der Probitanalyse. Die Schätzwerte für die Selektionsvariable werden als zusätzliche Variable ("hazard rate instrument") in die Regressionsgleichung für die "substantielle" Gleichung mit aufgenommen.

Die Probitanalyse wird durch die Annahme der bivariaten Normalverteilung der Fehlerterme bedingt; wird eine andere Verteilung unterstellt, ergeben sich zwei interessante Alternativen: Bei Annahme einer rechteckigen Verteilung bzw. der Annahme einer linearen Funktion zwischen den Fehlertermen der "substantiellen" Gleichung und den Fehlertermen der Selektionsgleichung ist ein "linear

probability model", also die Benutzung einer OLS-Regression anstelle der Probit-Gleichung, möglich und bei Annahme einer bivariaten logistischen Verteilung der Fehlerterme die Benutzung einer rechnerisch einfacheren logistischen Regression[32/] (BERK/ RAY 1982:387-391). Wird beim Probit-Modell "vorhanden" als "1" und "fehlend" als "0" codiert, erfolgt beim linearen Modell die Codierung umgekehrt, wobei lediglich die Konstante 1.0 von jeder vorhergesagten "Wahrscheinlichkeit" subtrahiert werden muß. Im Logit-Modell wird wie im linearen Modell "fehlend" als "1" codiert[33/]. Die durch das Logit-Modell vorhergesagten Wahrscheinlichkeiten stellen direkt das Hazard-Instrument für die substantielle Gleichung dar[34/].

HECKMAN (1979:158) schlägt vor, den Regressionskoeffizienten des Hazard-Instruments gegen Null zu testen, um die Nullhypothese ("keine Sampleselektivitätseffekte") zu prüfen. Der Test ist unproblematisch, da im Gegensatz zu den Schätzungen der Standardfehler der anderen Regressionskoeffizenten der Standardfehler des Regressionskoeffizienten der Instrumentalvariablen korrekt ist[35/]. Die in Kapitel 2 besprochenen und kritisierten Tests auf MAR erfahren hier eine interessante Erweiterung, insbesondere wenn man den Ansatz von COHEN/ COHEN (1975) mit der Benutzung einer Indikatorvariablen bei Multikollinearitätsproblemen der Einzelindikatoren als Methode mit der expliziten Berücksichtigung der "likelihood of missingness" interpretiert.

4.2.3 Probleme der Korrekturverfahren

Die Korrektur von OLS-Regressionen mit einem Hazard-Instrument beinhaltet einige Probleme. Zunächst einmal muß - neben den restriktiven Verteilungsannahmen - von der korrekten Spezifizierung des Regressionsmodells ausgegangen werden. Die Korrektur basiert auf der Annahme, daß die Kovarianz der Fehlerterme ausschließlich auf den Selektionseffekten beruht. Geht die Kovarianz hingegen auf einen Spezifikationsfehler zurück, dann ist das

Korrekturverfahren unangebracht. Der Anwendung der Korrektur muß somit die Überprüfung der Regressionsgleichung vorausgehen.

Ein weiteres Problem besteht in der häufig zu findenden hohen Multikollinearität des Instruments und den unabhängigen Variablen der substantiellen Gleichung (BERK/RAY 1982:386). Neben der korrekten Spezifizierung der substantiellen Gleichung ist selbstverständlich eine möglichst gute Anpassung der Selektionsgleichung eine wichtige Voraussetzung der Korrektur: Erklärt die Selektionsgleichung nur sehr wenig, dann erfolgt durch die Korrektur lediglich die Einführung einer neuen Konstanten.

Ebenso erweist sich die Hetereoskedastizität der Fehlerterme als problematisch. Die Schätzer der Regressionskoeffizienten sind zwar unverzerrt, aber ineffizient (HECKMAN 1979:157). HECKMAN (1976:483) erwähnt die Möglichkeit der Benutzung eines GLS-Schätzers, der aber gegenüber der ML-Schätzung immer noch ineffizient ist.

Gegenüber anderen Problemen des Modells sind diese Schwierigkeiten jedoch von untergeordneter Bedeutung. Das Hauptproblem dürfte bei der praktischen Durchführung darin liegen, daß nicht nur Daten auf der abhängigen Variablen systematisch fehlen, sondern ebenso Daten auf den unabhängigen Variablen - und dies eventuell aufgrund eines anderen, aber ebenso systematischen Prozesses. Bisher existieren keine Modelle, die unterschiedliche systematische Ausfallmechanismen für abhängige und unabhängige Variablen berücksichtigen.

4.2.4 Zusammenfassung

Die Entwicklung der Korrekturverfahren für OLS-Regressionen hat in der psychologischen und ökonometrischen Literatur im Gegensatz zur soziologischen Literatur sehr rasch ihren Niederschlag gefunden. Anwendungen sind z.B. in der Evaluationsforschung (BARNOW/CAIN/GOLDBERGER 1980) auf Grund der nicht zufälligen Zuweisungsmöglichkeit zu den "experimentellen Bedingungen"[36/]

in der Testpsychologie beim Restriction-of-Range-Problem (OLSON/ BECKER 1983), in der Therapieforschung (WELCH/FRANK/COSTELLO 1983) und vor allem in der ökonomischen Literatur (z.B. LEUTHOLD 1981) zu finden. Ohne Zweifel stellt die explizite Betonung der Entwicklung von Ausfallmodellen und Theorien über systematische Ausfälle im Zusammenhang mit der gesteigerten Sensitivität vieler Sozialwissenschaftler gegenüber den potentiell möglichen extremen Verzerrungen der substantiellen Beziehungen durch systematische Ausfallmechanismen einen Fortschritt gegenüber der unreflektierten Annahme des "missing at random" dar. Die Korrekturverfahren besitzen zwar erhebliche Probleme bei der praktischen Durchführung, aber die Notwendigkeit, eine jeweils auf eine spezielle substantielle Fragestellung bezogene explizite Ausfalltheorie zu entwickeln, ist durch die statistischen Korrekturmodelle vielleicht deutlicher betont worden, als es durch die eklektischen soziologischen Arbeiten zum Zusammenhang von Nonresponse und Globalvariablen möglich war[37/].

Anmerkungen zu Kapitel 4

1. Das ist die übliche Situation bei der Validierung eines Eignungstests.

2. Die Begriffe "incidential selection" und "explicit selection" gehen auf GULLIKSEN (1950:130) zurück.

3. In anderer Notation finden sich die Formeln auch bei JOHNSON/ KOTZ (1970:80;83), dort findet sich auch eine Tabelle, die die Veränderung des erwarteten Wertes und der Standardabweichung zeigt.

4. Unter der Voraussetzung, daß zudem bivariate Normalverteilung vorliegt und damit die bedingte Verteilung von y für gegebenes x normal mit dem Erwartungswert rXY * x und der Standardabweichung $(1-rXY^2)^{1/2}$ ist (JOHNSON/KOTZ 1972:112) .

5. ELSHOUT/ROE (1973:60). Die Autoren zeigen sich durch die Möglichkeit der Inflation der Korrelationskoeffizienten im restricted Sample überrascht. Allgemein vermittelt die Literatur den Eindruck, daß die Koeffizienten im restricted Sample kleiner werden als im unrestricted Sample.

6. PEARSON führte weiterhin die Normalverteilungsannahme an. Sie ist aber zur Herleitung nicht erforderlich, wie LAWLEY (1943/ 1944:30) zeigte (vgl. auch: GREENER/OSBURN 1980:338). PEARSON vermerkte dies (1912:437) in einer selten zitierten Arbeit (vgl. RYDBERG 1963:25).

7. In der gesamten Literatur werden Samplingprobleme in bezug auf die Stichprobenkoeffizienten entweder überhaupt nicht erwähnt oder als vernachlässigbar klein angesehen (vgl. RYDBERG 1963:100- 103). Die im statistischen Bereich entwickelten Schätzer (COHEN 1955,1957) erfordern lediglich die Ersetzung der Parameter in der Korrekturformel durch Samplekoeffizienten (vgl. GROSS/PERRY 1983: 115).

8. Auf die Formeln der Schätzer wird hier wegen ihres Umfangs nicht eingegangen. Sie finden sich bei BROUWER/VIJN (1980b:3-9). Die Verfasser stellen für die umfangreichen Berechnungen das FORTRAN-Programm RMISDAT zur Verfügung.

9. Die Bayes-Schätzer erbrachten in der Simulation unter der MAR- Annahme im Vergleich zu r (pairwise deletion) selbst mit informativen priori-Verteilungen keine entscheidenden Verbesserungen (maximal .03) (BROUWER/VIJN 1980a:197;1980b:15-26).

10. Hinweise auf Computerprogramme finden sich bei MAYS(1978:420) und ALLIGER/ALEXANDER(1984:677-678). Eine anschauliche "geometrische" Darstellung gibt SKINNER (1984).

11. Eine Übersicht über empirische Arbeiten bis 1963 findet sich bei RYDBERG (1963:204;Tabelle 8.41) .

12. Für praktische Zwecke stellen GULLICKSON/HOPKINS (1976) Nomogramme zur Verfügung.

13. Hier stellt sich dann allerdings die Frage, warum nicht gleich mit simulierten Datensätzen gearbeitet wurde.

14. Zur Veranschaulichung: ogivenförmige Verteilungen ergeben sich beispielsweise durch die kumulierte Häufigkeitsverteilung einer normalverteilten Variablen; "fan"-Verteilungen besitzen für wachsendes x größer werdende bedingte Varianzen (also y-Varianzen für gegebenes x); "football"-Verteilungen besitzen in der Nähe des x Mittelwertes die größten bedingten y-Varianzen, so daß sich für die "fan"-Verteilungen das Bild eines Fächers und für die "football"-Verteilungen das Bild eines -amerikanischen- Fußballs ergibt .

15. GROSS/FLEISCHMAN (1983:235) schreiben unverständlicherweise den Ergebnissen ihrer "realen" Datensätzen eine größere Generalisierbarkeit zu als den Ergebnissen simulierter Datensätze. Bei sorgfältiger Konstruktion der Simulation ist das Gegenteil zu erwarten.

16. Eine kurze Darstellung des Verfahrens findet sich bei RYDBERG (1963:37-38,115-118).

17. Die Annahmen sind bei weitem nicht so restriktiv wie ROE (1979:551;Fußnote 2) sie darstellt : "...some definite assumptions regarding the selector variable have to be made."

18. Die Darstellung bei GROSS/PERRY (1983:117) ist exakter als die Originalarbeit.

19. Vgl. BERK/RAY (1982:352). Diese Arbeit und BERK (1983) stellen die ersten lehrbuchartigen Übersichten zu diesem Thema von Soziologen dar.

20. Dieser Abschnitt lehnt sich eng an GOLDBERGER (1981) und BERK/RAY (1982) an.

21. Eine ausführliche Kommentierung findet sich bei BERK/RAY (1982:359-360).

22. Zu Hazardfunktionen allgemein vgl. LAWLESS (1982:11-27).

23. Anschauliche Graphiken findet man hierzu bei BERK/RAY (1982: 363).

24. Vgl. z.B. PINDYCK/RUBINFELD (1981:323-324).

25. Zur Ableitung vgl. GOLDBERGER (1981:364)

26. Sowohl für das Tobit-Modell als auch für das "truncated regression"- Problem existiert die Möglichkeit der ML-Schätzung, aber diese Methoden setzten iterativ arbeitende spezielle Software voraus und sollen daher hier nicht weiter erörtert werden. Zum Problem der Regressionsanalyse in "censored samples" vgl. SCHMEE/HAHN (1979), die eine " iterative least squares"-Technik (ILS) anwenden, die, wie AITKIN (1981) zeigt, bis auf die Schätzung der Fehlervarianz mit der ML-Schätzung über den EM-Algorithmus übereinstimmt. Zur ML-Schätzung siehe auch SWAN (1969) und WOLYNETZ (1979), zum EM-Algorithmus vgl. Kapitel 5. AITKIN (1981:163) gibt für dieses Problem an, daß der EM-Algorithmus mit einem 9-Zeilen-Macro in GLIM (BAKER/NELDER 1978) implementierbar sei. Eine einfach anzuwendende Alternative zu ML-Schätzern für den Fall der "truncated regression" findet sich bei OLSEN (1980), vgl. dagegen AMEMIYA (1973, 1974). Eine einfachere Darstellung geben JUDGE/GRIFFITHS/HILL/LEE (1980:609-616).

27. GREENE (1981:510) berichtet von den Ergebnissen einer Simulationsstudie, die die Robustheit der Korrektur gegenüber der Verteilungsannahme zeigt (Gleichverteilung/binäre Variablen). Lediglich extrem schiefe Verteilungen bedingen starke Verzerrungen.

28. Eine übersichtliche Darstellung anderer Schätzer findet sich bei BERK/RAY (1982:374), ebenso ein Beispiel. Vgl. auch GREENE (1983).

29. Vgl. auch NELSON (1977), siehe hierzu LITTLE (1983:390).

30. Ist die Variable bekannt, kann sie direkt zur Schätzung benutzt werden, vgl. HECKMAN (1979:157, Fußnote 3).

31. Einführungen in die Probit- und Logitanalyse finden sich bei PINDYCK/RUBINFELD (1981:273-312) und HANUSHEK/JACKSON (1977:179- 216). Vertiefungen stellen die Arbeiten von MCFADDEN (DOMENCICH/ MCFADDEN 1975:47-125; MCFADDEN 1974) dar. Einen Literaturbericht über neuere Entwicklungen gibt MANSKI (1981). Das in diesem Zusammenhang häufig zitierte Werk von FINNEY (1971) ist fast nur noch von historischem Interesse.

32. Standardsoftware für die Probitanalyse ist erst mit Release 2.0 von SPSS-X allgemein verfügbar. PROBIT berechnet Probit- und Logit-Modelle (vgl. SPSS-X Release 2.0 Update Documentation for IBM VM/CMS, SPSS INC. (1984:92-119).

33. Zur Durchführung vgl. BERK (1983:393-396).

34. BERK/RAY (1982:388) erwähnen eine von ihnen durchgeführte Simulationsstudie, die sehr hohe Korrelationen zwischen den verschiedenen Instrumenten nahelegt.

35. MELINO (1982) zeigt, daß der Test wünschenswerte asymptotische Eigenschaften besitzt, vgl. auch POIRIER/RUUD (1983), die u.a. die Empfindlichkeit des Tests gegenüber Verteilungsannahmen betonen.

36. Vgl. hierzu auch HAUSMAN/WISE (1977).

37. Interessanterweise existiert eine weitgehend unbeachtet gebliebene frühe Arbeit von EKLUND (1959), in der explizite (pfad- analytische) Ausfallmodelle für eine Reihe systematischer Ausfallmechanismen in der angewandten Statistik entwickelt wurden.

5. Missing Data Techniken

> "The arsenal available for attacking data which con tain in complete observations includes: 1/adapting the analysis to accomodate missing data, 2/ exclu ding all cases with missing values, 3/replacement of missing values with estimated values."
>
> VACEK/ASHIKAGA (1980:326)

Für die statistische Analyse von unvollständigen Datensätzen existieren eine Reihe spezieller Schätztechniken. Fast alle diese Verfahren setzen die MAR-Annahme voraus. In der Literatur werden mehrfach Taxonomien für diese Techniken vorgeschlagen (HAMILTON 1975:13; STEWART 1982:395; LITTLE/RUBIN 1983:49); prinzipiell kann zwischen Methoden zur Schätzung fehlender Werte und Methoden zur Parameterschätzung in Datensätzen mit fehlenden Werten unterschieden werden. Diese Unterscheidung basiert auf dem Kriterium, ob nach der Anwendung der Methode ein vervollständigter Datensatz, bei dem die fehlenden Werte durch Schätzungen ersetzt wurden (Imputation), existiert oder nicht.

Im folgenden sollen zunächst einige Parameterschätzmethoden erörtert werden; darauf folgt eine Darstellung der wichtigsten Substitutionsverfahren für fehlende Werte. Den Abschluß des Kapitels bildet eine Einschätzung zur Brauchbarkeit der MD-Techniken im Rahmen empirischer Sozialforschung. Zuvor soll abgrenzend noch eine der am häufigsten angewandten Methoden zur Handhabung des MD-Problem auf der Ebene der Erhebungseinheiten erwähnt werden: das Gewichtungsverfahren.

Gewichtungsverfahren sind nur für Unit-Nonresponse-Probleme (oder Item-Nonreponseprobleme mit einer interessierenden Variablen) praktikabel, da die Gewichte jeweils nur für eine Variable angemessen sind. Sollen multivariate Analysen erfolgen, entsteht das Problem, daß für eine Einheit mehrere Gewichte existieren (SANTOS 1981:18). Gewichtungsverfahren sind somit nur dann von Interesse, wenn es um die Schätzung univariater Statistiken geht; im Rahmen

analytischer Surveys sind Gewichtungsverfahren für Ausfälle auf Itemebene untauglich und werden im folgenden nicht mehr angesprochen[1/].

5.1 Parameter-Schätzverfahren

Zunächst sollen zwei MD-Techniken dargestellt werden, die primär keine Ersetzung fehlender Werte intendieren. Neben den bekannten "missing data"-Korrelationen sollen vor allem die Maximum-Likelihoodschätztechniken für Korrelationskoeffizienten bei MD etwas ausführlicher dargestellt werden.

5.1.1 Berechnungsmethoden von Korrelationsmatrizen bei MD

Die Methoden zur Berechnung von Korrelationsmatrizen aus Datensätzen mit fehlenden Werten sind sowohl selbst als MD-Technik von Bedeutung, als auch insoweit sie für viele der im folgenden beschriebenen Verfahren den Ausgangspunkt bilden. Diese vor allem in der sozialwissenschaftlichen Literatur häufig in Anlehnung an das BMDP-Programm 8D irreführend als "Missing data Korrelationen" (z.B. MÖNTMANN/BOLLINGER/HERRMANN 1983:91) bezeichneten Methoden stellen in vielen Programmpaketen die einzigen MD-Techniken dar, die verfügbar sind.

Die einfachste Methode besteht im Ausschluß aller Fälle mit fehlenden Werten aus den Berechnungen. Diese Matrizen werden von jedem Statistikprogrammpaket berechnet, aber jeweils anders bezeichnet. In BMDP werden die entsprechenden Optionen "COMPLETE" genannt, in SAS "NOMISS" und in SPSS "LISTWISE". Diese Methode wird im folgenden als "LISTWISE" bezeichnet.

MATTHAI(1951:148) erwähnt u.a. die Möglichkeit, aus den paarweise vorhandenen Werten die Kovarianzen zu berechnen. Die Mittelwerte bei der Berechnung der Kovarianz werden dabei nur aus den denjenigen Werten berechnete, für die der betreffende Fall auf der jeweils anderen Variablen

ebenfalls über gültige Werte verfügt. Werden dann die Varianzen aus allen verfügbaren Werten der Variablen berechnet, sind Korrelationskoeffizienten außerhalb des Wertebereichs (-1,+1) möglich. In der Literatur findet sich häufig die Bezeichnung als "GLASSER's method" (z.B. HAITOVSKY 1967:68). Diese Methode wird in BMDP als "COVPAIR" bezeichnet.

Eine weitere Möglichkeit stellt die Berechnung des Mittelwertes pro Variable aus allen verfügbaren Daten dieser Variablen und entsprechend der nichtbedingten Varianzen bei paarweiser Berechnung der Kovarianzen dar. Innerhalb der Standardpakete ist diese Methode nur in BMDP vorhanden und wird als "ALLVALUE" bezeichnet. Korrelationskoeffizienten, die mit "ALLVALUE" berechnet werden, können außerhalb des Wertebereichs (-1,+1) liegen.

Die andere in Statistikprogrammpaketen neben "COMPLETE" in der Regel vorhandene Methode berechnet für jedes Variablenpaar die bedingten Mittelwerte und Varianzen bei paarweiser Berechnung der Kovarianzen. Die Koeffizienten liegen immer im Wertebereich (-1,1) In BMDP werden die entsprechenden Optionen als "CORPAIR" bezeichnet, in SPSS als "PAIRWISE". Im folgenden wird diese Methode als "PAIRWISE" bezeichnet[2/].

In der Literatur werden die verschiedenen Methoden[3/] der paarweisen Berechnung nicht immer deutlich unterschieden[4/].

Neben diesen vier wichtigsten Methoden finden sich in der Literatur eine Reihe weiterer Vorschläge zur Berechnung von Korrelations- bzw. Kovarianzmatrizen bei fehlenden Werten[5/]. Vor allem in der frühen MD-Literatur wird die Berechnung der Korrelationsmatrix aus den Daten, nachdem die fehlenden Werte durch die Mittelwerte der jeweiligen Variablen ersetzt wurden, ausführlich erörtert. Diese Methode wird zumeist als "WILK's method"[6/] bezeichnet und ist in einigen Statistikpaketen, z.B. in BMDPAM als "MEAN"- Option und in SPSS-X REGRESSION, verfügbar. Im folgenden wird diese Methode als "MEAN" bezeichnet.

5.1.2 Techniken bei singulären Korrelationsmatrizen

Die durch "ALLVALUE", "CORPAIR" und "PAIRWISE" entstehenden Korrelationsmatrizen können bei hohem Anteil fehlender Werte und niedrigen Fallzahlen und/oder systematischen Ausfällen eventuell nicht mehr positiv semi-definit sein[7/], so daß z.B. Standardpakete zur multiplen Regression mit einer Fehlermeldung abbrechen. So ergab z.B. eine von SCHWERTMAN/ALLEN (1979:194) durchgeführte Simulationsstudie[8/] bei n=80 und 10% fehlender Werte 78% singuläre Varianz-Kovarianzmatrizen.

In dem hier interessierenden Zusammenhang gibt es zwei Möglichkeiten das Problem anzugehen. Zum einen kann eine Ridge-Prozedur angewandt werden, zum anderen ist ein "smoothing" der Korrelationsmatrix möglich. Bei einer Ridge-Regression wird eine kleine positive Zahl k zu jedem Element der Diagonale der Korrelationsmatrix addiert (FREUND/MINTON 1979:140). Dies reduziert eventuell vorhandene Varianzinflationsfaktoren, so daß bei geeigneter Wahl von k die Matrix keine negativen Eigenwerte mehr aufweist. Die Schätzungen der Koeffizienten[9/] sind nicht mehr unverzerrt; weiterhin wird k meist über Iterationen ermittelt[10/].

Neben dieser Methode ist ein "Glätten" der Matrix ("Smoothing") möglich. Das Prinzip beruht auf der Rekonstruktion einer Korrelationsmatrix P aus der Diagonalmatrix der positiven Eigenwerte E und der zugehörigen Matrix der Eigenvektoren V der ursprünglichen Korrelationsmatrix R. Bezeichnet man mit C das Produkt V'EV (wobei C die Form einer Kovarianzmatrix besitzt) und mit F die Diagonalmatrix der Quadratwurzeln der Diagonalelemente von C, dann wird die geglättete Matrix P aus der Matrix R mit

5-1 $$P = F^{-1} V`EVF^{-1}$$

ermittelt (FRANE 1978:31). Die Matrix P ist positiv semi-definit mit Diagonalelementen gleich Eins. Das Verfahren kann durch Ersetzen des Kriteriums, daß nur Eigenwerte größer Null verwendet werden sollen, durch ein

anderes Kriterium (so z.B. die Forderung nach Eigenwerten, die größer einer bestimmten Konstanten sein sollen) modifiziert werden. Eine Gewichtung der Variablen ist ebenso möglich (FRANE 1978:31) wie die Kombination der Ridge-technik mit dem Glätten der Korrelationsmatrix. Diese Möglichkeit besteht beispielsweise in BMDPAM.

Die bisher mangelnde Untersuchung beider Techniken unter verschiedenen Ausfallmodellen erlaubt keine Abschätzung der Konsequenzen der Benutzung der Verfahren[11/].

5.1.3 Maximum-Likelihood-Schätzungen

Richtet sich das Interesse weniger auf die Ersetzung fehlender Werte durch Schätzwerte[12/] als auf die Schätzung des Mittelwertvektors und der Kovarianzmatrix aus den Daten eines Samples mit fehlenden Werten, so bieten sich ML-Schätzungen an[13/]. Alle im folgenden beschriebenen Ansätze setzen MAR und multivariate Normalverteilung voraus.

Die Likelihood-Funktion für ein Sample mit fehlenden Werten ist wesentlich komplizierter als die Likelihood-Funktion eines vollständigen Samples (LITTLE 1983:350). Aus diesem Grund existieren nur für Spezialprobleme explizite Lösungen. Die Arbeiten zum bi- und trivariaten Problem werden im nächsten Abschnitt erwähnt. Für den Fall eines speziellen Musters der Ausfälle kann die Likelihood-Funktion faktorisiert werden. Da die Faktoren der Funktion dann unabhängig voneinander maximiert werden können, sind in diesem Fall explizite Lösungen möglich. Die Anwendung dieser Methode auf Panel-Daten ist Gegenstand des zweiten Abschnitts.
Existiert in den Daten kein spezielles Muster der fehlenden Werte (nur bei einem "monotonen" Muster ist eine Faktorisierung möglich), kann die Likelihood-Funktion nur durch ein iteratives Verfahren maximiert werden. Ein Ansatz, aus dem für das hier interessierende Problem eine durchführbare Lösung hergeleitet werden kann, ist der EM-Algorithmus. Die Anwendung einer speziel-

len Form des EM-Algorithmus zur Schätzung fehlender Werte und der Kovarianzmatrix unter Vorliegen multivariater Normalverteilung ist Gegenstand des dritten Abschnitts.

5.1.3.1 ML-Schätzer für den bi- und trivariaten Fall

Für eine Reihe spezieller Probleme zur Parameterschätzung beim Vorliegen unvollständiger Datensätze existieren Maximum-Likelihood-Schätzer. WILKS (1932) entwickelte für den Fall bivariater Normalverteilung ML-Schätzer für Mittelwert, Standardabweichung und den Korrelationskoeffizienten. MATTHAI (1951) erweiterte den Ansatz von WILKS auf den trivariaten Fall. Die Entwicklung expliziter ML-Schätzer bei MD wird in der Regel möglich, wenn die MD ein "monotones" ("nested") Muster aufweisen: Ein MD-Muster wird als monoton bezeichnet, wenn die Datenmatrix so angeordnet werden kann, daß für jeden Fall das Vorhandensein der Variablen j das Vorhandensein der Variablen i ($i < j$) impliziert (LITTLE 1983: 357). LORD (1955) und EDGETT (1956) entwickelten für Spezialfälle des trivariaten Falls ML-Schätzer, die Arbeit von ANDERSON (1957) umfaßt beide Fälle[14/]. Keiner der speziellen Schätzer hat in der Forschungspraxis Verbreitung gefunden[15/].

5.1.3.2 ML-Schätzer für Panel-Daten

Eines der Hauptprobleme der Durchführung von Paneluntersuchungen ist die Panelsterblichkeit. Allgemein wächst die Ausfallquote mit der Zahl der Befragungswellen (ARMINGER 1976:139-141). In der Regel weisen die Ausfälle ein monotones Muster auf: Nach dem ersten Ausfall einer Einheit ist es unwahrscheinlich, daß diese Einheit in späteren Wellen wieder verfügbar wird.

Für den Fall monotoner Panelausfälle[16/] multivariater Normalverteilung und der MAR-Annahme[17/] entwickelten MARINI/OLSEN/ RUBIN (1980) eine nichtiterative ML-Schätzmethode für die Mittelwerte und die Kovarianzmatrix.

Die Methode[18/] besteht in der schrittweisen Berechnung der Kovarianzmatrizen, wobei von dem Variablenblock mit der höchsten Responsequote ausgegangen wird. Die Variablen des Variablenblocks mit der zweithöchsten Responsequote werden dann einer multiplen Regression mit den Variablen des ersten Blocks unterworfen. Die Variablen des dritten Blocks werden einer multiplen Regression mit den Variablen des zweiten Blocks unterworfen usw.. Die ML-Schätzer der Kovarianzmatrix und des Mittelwertvektors werden jeweils durch "sweeping"[19/] der geschätzten Matrizen berechnet.

Die zugrunde liegende Idee dieser ML-Schätzer basiert darauf, daß die durch das monotone Muster ermöglichten Fraktionierungen der Likelihood-Funktion unabhängig voneinander maximiert werden können. Jeder der oben beschriebenen Schritte der Regression eines Variablenblocks auf die Variablen des vorhergehenden Blocks mit höherer Responsequote maximiert den jeweiligen Teil der Likelihood-Funktion. MARINI/OLSEN/RUBIN betonen, daß zur Berechnung der ML-Schätzer keine durch Regressionen berechneten Ersetzungen notwendig sind und alle Berechnungen mit einfachen Matrizenoperationen durchgeführt werden können[20/].

5.1.3.3 EM-Algorithmus

"An algorithm must be seen to be believed..."

KNUTH (1973:4)

Der "expectation-maximization" (EM)-Algorithmus stellt eine allgemeine Methode zur Berechnung von Maximum-Likelihood-Schätzungen dar, die sich prinzipiell auf jedes MD-Problem anwenden läßt. Die ersten Anwendungen des EM-Algorithmus gehen in die fünfziger Jahre zurück; die erste ausführliche Darstellung gaben ORCHARD/ WOODBURY (1972), deren Bezeichnung "MIP": "Missing information principle"[21/] sich gelegentlich in der Literatur findet. DEMPSTER/LAIRD/RUBIN (1977) führten den Begriff "EM" ein und beschrieben die Konvergenzeigenschaften[22/]. Die Bezeichnung des generellen Ansatzes bei DEMPSTER/LAIRD/RUBIN als "Algorithmus" hat generell Anlaß

zur Kritik gegeben (HASSELBLAD/CREASON/STEAD 1983:184): Von der Arbeit von DEMPSTER/LAIRD/RUBIN zu einem Algorithmus ist es mehr als ein Schritt[23/].

Der EM-Algorithmus arbeitet iterativ; jede Iteration besteht aus einem E-Schritt, in dem die "averaged log-likelihood" berechnet wird, und einem M-Schritt, in dem diejenigen Parameter gesucht werden, die die "averaged log-likelihood" maximieren (vgl. LITTLE 1983:369-371). Im hier interessierenden Zusammenhang ist nur eine der Anwendungsmöglichkeiten des EM-Algorithmus wichtig. Unter der Annahme multivariater Normalverteilung und einer einfachen Zufallsstichprobe stellen der Vektor der Mittelwerte und die Kovarianzmatrix die suffizienten Statistiken dar. In diesem Fall besteht der E-Schritt des EM-Algorithmus darin, die Erwartungswerte der suffizienten Statistiken unter den aktuellen Schätzungen des Mittelwertvektors und der Kovarianzmatrix zu finden. Der M-Schritt besteht aus der Berechnung der neuen Schätzungen. Ein fehlender Wert wird durch eine lineare Regression mit allen vorhandenen Variablenwerten dieses Falles geschätzt.
Der vollständige EM-Algorithmus sieht dann folgendermaßen aus:

1. Berechnung einer ursprünglichen Schätzung des Mittelwertvektors und der Kovarianzmatrix aus den vollständigen Fällen.
2. Für jeden Fall mit fehlenden Werten: Partitionieren der Matrizen in einen vollständigen und einen unvollständigen Teil.
3. Schätzung der fehlenden Werte durch multiple Regression mit allen vorhandenen Variablen unter Benutzung der geschätzten Mittelwerte und der Kovarianzmatrix.
4. Berechnung eines neuen geschätzten Mittelwertvektors und einer neuen geschätzten Kovarianzmatrix.
5. Korrektur der Kovarianzmatrix: Für jeden Fall mit fehlenden Werten wird die residuale Kovarianz der Variablen zu dem entsprechenden Element der Kovarianzmatrix addiert.
6. Berechnen eines Konvergenzkriteriums[24/]. Falls keine Konvergenz: Wiederholung der Schritte 2-6; sonst Abbruch.

Alle Schritte des Algorithmus mit Ausnahme von Punkt 5 stellen lediglich Anwendungen üblicher "complete data methods" dar. Die Korrektur der Kovarianzmatrix bedarf einiger Erläuterungen. Zur Berechnung eines Elementes der neuen geschätzten Kovarianzmatrix werden auch Elemente der Datenmatrix benutzt, die über multiple Regressionen geschätzt wurden. Die Diagonalelemente der Kovarianzmatrix (Varianzen) würden ohne Korrektur dadurch unterschätzt, die Kovarianzen würden unter- oder überschätzt. Die Korrektur erfolgt dadurch, daß bei der Berechnung der Elemente der Kovarianzmatrix zu jedem Summanden (d.h.: für jeden Fall) die residuale Kovarianz[25/] addiert wird. Die residuale Kovarianz zweier Variablen ist die Kovarianz der Residuen beider Variablen bezüglich der zur Vorhersage benutzten Variablen. Damit hängt die residuale Kovarianz von dem Muster der fehlenden Werte ab: Die residuale Kovarianz variiert zwischen den Fällen mit unterschiedlichen MD-Mustern, weil zu den Vorhersagen jeweils eine andere Menge von Prädiktoren benutzt wird, nämlich die Variablen, die in dem jeweiligen MD-Muster vorhanden sind. Die residuale Kovarianz ist nur dann ungleich Null, wenn beide Variablen für den entsprechenden Fall fehlen, da ansonsten mindestens eine der beiden Variablen zur Prädiktorenmenge der anderen Variablen gehört. In letzterem Fall wird die Kovarianz zwischen Residuen berechnet, aus denen der durch die jeweils andere Variable linear erklärbare Anteil eliminiert ist: Diese Kovarianz ist Null. Bei der Berechnung der Varianz (den Diagonalelementen) fehlen bei unvollständigen Fällen immer "beide" Variablen oder keine der "beiden" Variablen, daher ist für jeden Fall mit fehlenden Werten auf einer bestimmten Variablen eine Korrektur der Varianz dieser bestimmten Variablen erforderlich.

Wie BEALE/LITTLE (1975:134-137) zeigen, ist der EM-Algorithmus identisch mit der multiplen Regressionsersetzung nach BUCK (vgl. Kap. 5.6.), sofern diese iteriert angewandt wird. Ein wenig beachtetes Nebenergebnis dieser Herleitung bei BEALE/LITTLE ist, daß die Annahme multivariater Normalverteilung nicht notwendig ist.

Der EM-Algorithmus wird, zumindest unter der MAR-Annahme, in der neueren Literatur allgemein als das unter allen Umständen zu bevorzugende Verfahren angesehen; häufig stellt er das einzige überhaupt erwähnte MD-Verfahren dar (z.B. SEBER 1984). Trotz der hervorragenden Ergebnisse des EM-Algorithmus in den meisten der bisher vorliegenden Simulationsstudien (vgl. Kapitel 6) können z.Z. keine Angaben über die Robustheit gegenüber der Verletzung der MAR-Annahme gemacht werden.

5.2 MD-Ersetzungstechniken

> "Die Vorgehensweise ist einfach. Man wählt eine unter vier Möglichkeiten der Schätzung fehlender Werte aus. (...) Mit den vollständigen Daten läßt sich anschließend jedes Programm ohne Verlust von Fällen rechnen."
>
> BOLLINGER/HERRMANN/MÖNTMANN(1983:33)

Neben den bisher erörterten Parameterschätzverfahren existieren Verfahren, die fehlende Werte durch Schätzwerte ersetzen ("Imputation") und damit "vervollständigte" Datensätze, die keine fehlenden Werte enthalten, produzieren. Vor der Darstellung der wichtigsten Ersetzungstechniken soll zunächst auf die prinzipiellen Argumente für und gegen die Ersetzungstechniken eingegangen werden.

In der Literatur dominieren zwei Argumente für den Einsatz von Ersetzungsverfahren. Zum einen wird die leichtere Handhabbarkeit vollständiger Datensätze hervorgehoben. So müssen den "Endnutzern" von Datensätzen (Auftraggebern, Sekundäranalytikern, Lesern von Forschungsberichten) nicht ständig die bei verschiedenen Auswertungen variierenden Fallzahlen, auf denen die Auswertungen basieren, erklärt werden (z.B. I.G. SANDE 1979:240). Durch Ersetzungstechniken können z.B. inkonsistente Randverteilungen bei wechselnden Kreuztabellierungen in einer Publikation vermieden werden (KALTON 1983:65).

Darüber hinaus verfügen z.B. neu entwickelte Auswertungstechniken in ihren ersten Programmrealisationen als einziger MD-Behandlung zumeist nur über den völligen Ausschluß von Fällen mit fehlenden Werten, wenn überhaupt auf evtl. unvollständige Daten Rücksicht genommen wird[26/]. Vor der Anwendung solcher Programme[27/] muß also der Datensatz entweder durch Ersetzungsverfahren oder durch Löschung "vervollständigt" werden.

Im Zusammenhang mit der Erstellung eines bereinigten Datensatzes bezieht sich eine weitere Begründung des Einsatzes von Ersetzungstechniken auf die im Kapitel 3 angesprochene Verbindung zwischen Datenmanagement als Vorbereitung der Datenanalyse und den Ersetzungstechniken: Um überhaupt zu einem analysierbaren Datensatz zu gelangen, müssen die Daten zumindest elementaren Konsistenzbedingungen genügen. Ist dies nicht der Fall, bleibt als Alternative zur Ersetzungdie Löschung aller an der Entstehung von Inkonsistenzen beteiligten Daten einer Erhebungseinheit. Dies kann die Löschung eines nicht mehr tolerierbar großen Teils der Daten bedingen (I.G.SANDE 1982:145-146)[28/].

Zum anderen wird für die Verwendung von Ersetzungsverfahren angeführt, daß durch die Verwendung geschätzter Daten eine Reduktion des Nonresponse-Bias erreicht werden kann (KALTON 1983:65; FORD 1983:187). Zentral für dieses Argument ist die Akzeptanz des MARC-Modells, m.a.W. die Annahme, daß innerhalb bestimmter Klassen eine grundlegende Ähnlichkeit zwischen Respondenten und Nonrespondenten[29/] besteht, so daß Ersetzungen für Nonrespondenten auf der Grundlage der Daten der Respondenten erfolgen können (vgl. KALTON 1983:16). Die mit dieser Annahme verbundenen Probleme sind Gegenstand der weiteren Ausführungen zur generellen Problematik informativer Ersetzungsverfahren (Kapitel 5.2.2).
Trotz dieser einsichtigen Argumente sind Ersetzungstechniken aus verschiedenen Gründen nicht unumstritten. Die Spannweite der Argumente gegen die Ersetzungstechniken reicht von Bedenken in Hinsicht auf Mächtigkeit statistischer Tests, da Ersetzungsverfahren die effektive Stichprobengröße nicht vergrößern, sondern verkleinern (SANTOS 1981:18-19), bis zum Vorwurf, daß

Imputation die Abkapselung der Surveys von der "sozialen Wirklichkeit" bedeutet (BATESON 1984:138), da der ausschließliche Träger von Informationen der Befragte sei und Ersetzungen auf Vermutungen des "Erhebenden" basieren. Dieses Argument wird zumeist durch die Feststellung gestützt, daß in der Datenanalyse jedes Datum als unabhängig von allen anderen Daten erhobene Information betrachtet wird; eben diese grundlegende Annahme werde aber durch die MD-Ersetzung verletzt (BANISTER 1980:1; BATESON 1984:138).

Ein weiteres Argument bezieht sich auf die Möglichkeit, daß Ersetzungen nicht zur erhofften Reduktion, sondern zur Vergrößerung des Bias führen können (KALTON 1983:65). Die Güte der Ersetzungsverfahren hängt von (zumeist implizit gelassenen) Annahmen bei der Ersetzung ab (SANTOS 1981:19). Werden die Ersetzungen in Hinsicht auf eine Biasreduktion univariater Statistiken optimiert, können größere Verzerrungen in Hinsicht auf die Kovarianzstruktur resultieren (I.G. SANDE 1982:147). Die möglichen Verzerrungen der Kovarianzstruktur sind vor allem dann gegeben, wenn Variablen, die zur Durchführung der Ersetzungen benutzt wurden, später in der Datenanalyse mit den vervollständigten Variablen korreliert werden[30/].

Derartige Artefakte können nur durch die eindeutige Kennzeichnung aller Ersetzungen durch spezielle Indikatorvariablen (Flagvariablen)[31/] vermieden werden. Darüber hinaus müssen auch für rein deskriptive Auswertungen die vorgenommenen Ersetzungen ausreichend dokumentiert werden, um dem Argument des Vortäuschens einer hohen Datenqualität[32/] zu begegnen.

Die Berücksichtigung solcher Indikatorvariablen im Verlauf der Datenanalyse bzw. die Dokumentation der technischen Vorgehensweise und der theoretischen Annahmen, auf denen die Ersetzungen basieren, gehören auf Grund der dargestellten Kritik notwendigerweise zur Durchführung einer MD-Ersetzung. Die meisten der genannten Kritikpunkte werden durch diese Art der Dokumentation gegenstandlos. Die prinzipielle Möglichkeit des Mißbrauchs[33/] der Ersetzungsver

fahren besteht im selben Umfang wie bei jeder anderen Technik.

Im Gegensatz zu den oben angesprochenen Parameterschätzverfahren, die fast ausschließlich auf der MAR-Annahme beruhen, stellen die meisten Ersetzungstechniken einen Versuch zur Reduktion der Verzerrungen durch systematische Ausfälle dar. Damit verbleiben vor allem zwei Probleme: zum einen das Problem der den Ersetzungen zugrunde liegenden Ausfallmodelle und zum anderen die Abschätzung des Einflusses der Ersetzungen und der ihnen zu Grunde liegenden Annahmen auf die letztlich beabsichtigten Parameterschätzungen. Auf diese Probleme wird später zurückgekommen (Vgl. Kap. 5.2. und Kap. 9).

5.2.1 Nicht-informative Ersetzungen

Unter nicht-informativen Ersetzungstechniken sollen hier (in Anlehnung an SANTOS 1981:30) Methoden verstanden werden, die nicht auf direkte Weise versuchen, die im Datensatz vorhandene Redundanz zur Ersetzung zu nutzen: Die nicht-informativen Techniken sind prinzipiell anwendbar, ohne Daten des unvollständigen Datensatzes über zur in Frage stehenden (unvollständigen) Variablen zusätzlich vorhandene Variablen zu verwenden. Neben der Ersetzung durch eine aus einem Zufallszahlengenerator gezogene Zahl (vgl. Exkurs IV) und der Ersetzung durch Daten eines zufällig gezogenen vollständigen Falles ("Random Imputation", vgl. SANTOS 1981:22) sind vor allem die Ersetzung durch "subjektiv" "wahrscheinliche" Werte ("Expertenratings") und die Mittelwerterset zung zu den nicht-informativen Ersetzungstechniken zu zählen.

5.2.1.1 Ersetzung durch "Expertenratings"

Das "Expertenratingverfahren" besteht aus der Ersetzung eines fehlenden Wertes durch einen Wert, der von einem Experten als "wahrscheinlichster" Wert bezeichnet wird. RUMMEL (1970:262-263) unterscheidet zwischen "partial" bzw. "complete estimates" und "ratings". Partielle Schätzungen laufen auf den Gebrauch von "durchschnittlichen Werten"[34/] hinaus, die ohne Informationen über den zu schätzenden Fall stattfinden. "Rating" bezieht sich auf die durch Informationen gestützte Schätzung eines Wertes. Auf dieses nichtstandardisierte Verfahren wird im folgenden nicht mehr eingegangen.

Vom "Rating" muß die sogenannte "deductive imputation", die sich gelegentlich in der MD-Literatur findet (z.B. KALTON 1983:68), unterschieden werden. Ist die Deduktion eines fehlenden Wertes möglich, kann also beispielsweise die Summe des Einkommens aus vorhandenen Daten einzelner Einkommensquellen berechnet werden, dann handelt es sich in der hier benutzten Terminologie weder um "Rating" noch um fehlende Werte.

5.2.1.2 Ersetzung durch den Mittelwert

Die Ersetzung fehlender Werte durch den Mittelwert der vorhandenen Werte der jeweiligen Variablen geht auf WILKS (1932) zurück. WILKS befaßte sich mit Parameterschätzungsproblemen im bivariaten Fall bei MD, für die er ML-Schätzer für drei Spezialfälle entwickelte und mit einfacheren Schätzern (WILKS 1932:178-179) verglich. Die in der Literatur zu findende Gleichsetzung von "Wilks' method"[35/] mit der Mittelwertersetzung ist ungerechtfertigt. Neben der Ersetzung durch den Spaltenmittelwert der Datenmatrix findet sich bei DEAR (1959)[36/] der Vorschlag der Ersetzung durch den Gesamtmittelwert der Datenmatrix.

5.2.2 Informative Ersetzungstechniken

Als informative Ersetzungstechniken sollen hier Verfahren bezeichnet werden, die direkt auf einer Redundanzannahme beruhen und dementsprechend den unvollständigen Fällen Werte zuschreiben, wie sie durch die vorhandenen Daten nahegelegt werden, und sich dabei auf vorhandene Daten des jeweiligen unvollständigen Falles zur Schätzung stützen. Obwohl auch die multivariaten Ersetzungstechniken zu den informativen Ersetzungsmethoden gehören, werden sie getrennt dargestellt[37/] die analytische Trennung der Verfahren voneinander ist zum einen ohnehin weitgehend willkürlich, da fast jedes Verfahren als "degenerierte" Form einer multiplen Regression aufgefaßt werden kann (vgl. SANTOS 1981:20-25) und zum anderen historische Gründe diese Trennung nahelegen: Die hier unter informativen Techniken vorzustellenden Methoden stammen weitgehend aus der Praxis der amtlichen Statistik, die multivariaten Ersetzungstechniken überwiegend aus der angewandten Forschungsstatistik.

5.2.2.1 Das Konzept der Imputationklassen

Die Definition des MARC-Modells als Ausfallmechanismus, bei dem die Respondenten innerhalb festgelegter Klassen eine einfache Zufallstichprobe aus der Population der jeweiligen Klasse darstellen (SANTOS 1981:17), legt die Benutzung von Ersetzungstechniken nahe, die jeweils für diese Klassen getrennt angewendet werden. Diese Klassen werden in der Missing Data-Literatur in der Regel als "imputationclasses"[38/] bezeichnet.

Damit stellt sich die Frage nach der Konstruktion von Imputationklassen und nach Kriterien zur Beurteilung der Güte der Imputationklasseneinteilung[39/].

Das MARC-Modell und die auf das MARC-Modell gegründeten Imputationklassentechniken basieren auf zwei Annahmen:

- Es existieren reale Subgruppen, die sich sowohl in Hinsicht auf die interessierenden Variablen als auch in ihren Responsewahr scheinlichkeiten unterscheiden (vgl. CHAPMAN 1983:448);
- die Responsewahrscheinlichkeiten sind in jeder Subgruppe homo gen und von Null verschieden (vgl. OH/SCHEUREN 1983:180).

Die Annahme der homogenen Responsewahrscheinlichkeiten[40/] enthält die Annahme, daß der Mechanismus, der innerhalb einer Zelle über Response/Nonreponse entscheidet, zumindest mit den interessierenden abhängigen Variablen nicht zusammenhängt. Die Annahme, daß zumindest ein Teil der Subgruppe Daten liefert (Responsewahrscheinlichkeit größer Null), schließt Truncation-Probleme aus. Das Imputationklassenkonzept ist auf solche Probleme nicht anwendbar (im Gegensatz zu Censoring-Problemen, vgl. SANTOS 1981:17 und GREENLESS/REECE/ZIESCHANG 1982).

Betont werden muß, daß das Imputationklassenmodell, auch wenn es nicht direkt als Konsequenz des MARC-Modells konzeptualisiert wird (z.B. KALTON/SANTOS 1983:16) und lediglich die Erstellung in Hinsicht auf die zu vervollständigenden Variablen vermutlich homogener Gruppen, die aus Respondenten und Nonrespondenten bestehen, beabsichtigt, die MARC-Annahme enthält: Nur unter der MARC-Annahme kann von einem Imputationprozeß Biasreduktion erwartet werden.

5.2.2.1.1 Die Wahl der Imputationklassen

Bei RIZVI (1983:306) finden sich eine Reihe praxisbezogener Hin- weise zur Wahl von Imputationklassen. Zunächst einmal müssen die- jenigen Variablen, die zur Konstruktion der Zellen benutzt werden sollen sowohl für die Respondenten als auch für die Nonrespondenten bekannt sein (in der Praxis schränkt dies die Zahl der verwendbaren Variablen zumindest für nachträgliche Festlegungen der

Zellen sehr stark ein). RIZVI[41/] hält eine gröbere Tabellierung mehrerer, möglichst wenig korrelierter Variablen für effektiver als eine detailierte Einteilung einer Variablen. Weiterhin erwähnt er die Möglichkeit, für verschiedene Gruppen verschiedene Variablen zur Klassifizierung zu benutzen. Da RIZVI die Anforderungen an eine Einteilung dahingehend zusammenfaßt, daß die Einteilung in Hinsicht auf die "outcome variables" intern so homogen und zwischen den Klassen so heterogen wie möglich sein sollte und zusätzlich die Responseraten differieren sollten, betrachtet er Clusteranalysen und insbesondere AID als angebrachte Techniken zur Ermittlung der Zellen.

5.2.2.1.2 Exkurs II: MD in der Clusteranalyse

Naheliegend ist der Versuch der Bestimmung der Imputationklasse mittels clusteranalytischer Verfahren. Im "Idealfall" könnte die Klassifizierung damit "automatisch" erfolgen. Dabei ergeben sich jedoch theoretische Probleme sowohl in Hinsicht auf die formale Modellierung der ablaufenden Prozesse als auch in Hinsicht auf dann unüberprüfbar werdende, dem Algorithmus zu unterstellende "soziologische Hypothesen". Die erfolgende Klassifizierung kann z.B. nicht auf Homogenität von handlungsrelevanten Variablen in bestimmten Gruppen, sondern lediglich auf rein zufälligen, sich nur in diesem speziellen Datensatz findenden "Klumpungen" von Meßfehlern beruhen.

Darüber hinaus ergeben sich Probleme mit eventuell fehlenden Werten auf den zur Klassifikation herangezogenen Variablen. Dieses Problem ist natürlich bei der zunehmenden Verwendung der Clusteranalyse auch von eigenständigem Interesse. Die Frage nach fehlenden Werten hat in der Clusteranalyseliteratur[42/] nur wenig Beachtung gefunden. Demzufolge arbeiten die meisten Programmpakete auch mit der einfachsten Lösung des Problems als Voreinstellung: Jeder Fall, der auch nur auf einer einzigen Variablen fehlt, wird aus den Berechnungen völlig ausgeschlossen (BMDP, SPSS-X).
WISHART (1978:282) bemerkt die Unangemessenheit der "listwise deletion" Praxis, da ja gerade eine Klassifikation beabsichtigt ist. Er betrachtet die

"pairwise"-Methode für Clusteralgorithmen, die als Distanz- bzw. Ähnlichkeitstransformationen definiert sind (single linkage, average linkage, etc.) als akzeptabel[43/]. Für die auf Minimierung einer Funktion abzielenden Algorithmen sieht er theoretische Schwierigkeiten. Bei Verfahren zur Minimierung der quadrierten euklidischen Distanz kann die Summe der quadrierten Abweichung eines Clusters in drei Teile zerlegt werden (WISHART 1978:284; Gleichung 6): eine Komponente für die Quadratsumme der vollständigen Fälle, eine MD- Komponente und eine Komponente für die (gewichtete) Differenz der Mittelwerte der vorhandenen und der fehlenden Werte. "Pairwise" benutzt nur den ersten Summanden, "clustermeans" setzt den zweiten und dritten Summanden gleich Null: Damit unterschätzen beide Verfahren die Gesamtsumme. Unter der Annahme, daß die Mittelwerte der vollständigen Fälle und der Fälle mit MD gleich sind, schlägt er für eine Reihe von Verfahren minimale Veränderungen vor. Da hier insbesondere systematische Ausfälle interessieren, werden die Einzelheiten nicht erörtert (WISHART 1978:285-286).

Für den Fall systematisch fehlender Werte hebt WISHART die Wichtigkeit zuverlässiger Schätzungen der MD hervor. Er benutzt den in der Regel überwiegend im Rahmen der MAR-Annahme diskutierten EM-Algorithmus (vgl. Kapitel 5.1.) innerhalb der Cluster.

Der damit von WISHART vorgeschlagene Algorithmus besteht aus vier Schritten:

STEP 1 : Gewinnung einer ursprünglichen Clustereinteilung mit dem k-means-Algorithmus. Für jedes Cluster werden die MD durch die Clustermittelwerte ersetzt.

STEP 2 : Schätzung der Kovarianzmatrix für jedes Cluster unter Benutzung der im STEP 1 vervollständigten Datenmatrix.

STEP 3 : Für jedes Cluster: Berechnung partieller Regressions koeffizienten mit den vorläufigen Mittelwerten und Ko varianzen. Dann Schätzung der MD mit der Regressions gleichung innerhalb jedes Clusters.

STEP 4 : Berechnung einer neuen Kovarianzmatrix und des Mittel wertvektors für jedes Cluster. Wiederholung der Schritte 3 und 4 bis ein Konvergenzkriterium erreicht ist.

Die Darstellung der Arbeit von WISHART dürfte die technischen Schwierigkeiten, die sich aus der Verwendung der Clusteranalyse in diesem Zusammenhang dann ergeben, wenn die zur Klassifikation benutzten Variablen selbst unvollständig sind, ersichtlich gemacht haben.

Die oben angesprochenen theoretischen Probleme der Benutzung clusteranalytischer Techniken zur "Gewinnung" von Imputationklassen werden bei der nun folgenden Erörterung der hierzu zumeist benutzten Technik, der AID-Methode, deutlich.

5.2.2.1.3 Bestimmung der Imputationklassen mit AID

Einen Spezialfall divisiver Clusteralgorithmen stellt das unter dem Namen AID (Automatic Interaction Detector) bekannt gewordene Programm von SONQUIST/MORGAN dar. AID-Techniken basieren auf der Erstellung einer Reihe sukzessiver Splits einer abhängigen Variablen in disjunkte Subgruppen, deren Zwischengruppenvarianz jeweils durch Splits maximiert wird[44/]. Für die folgenden Darstellungen muß festgehalten werden, daß AID aus einem Set hochkorrelierter Variablen stets den stärksten Prädiktor wählt, da ausschließlich nach Reduktion der Fehlervarianz selektiert wird.

KALTON/SANTOS (1983) und CHAPMAN (1983) benutzen auf zwei verschiedene Arten[45/] Modifikationen der AID-Technik zur Bestimmung der Imputationklassen, die die mit der Verwendung von AID in diesem Zusammhang verbundenen Probleme deutlich hervortreten lassen.

KALTON/SANTOS (1983:84-87, 93-96) berichten von der Benutzung des AID-Nachfolgeprogramms SEARCH zur Bestimmung der "imputation control variables", wobei sie diejenige Variable, für die Ersetzungen vorgenommen werden sollen, als abhängige Variable verwenden. KALTON und SANTOS halten SEARCH für ein "valuable tool for guiding the choice of the imputation classes" (S.110), sofern die Fallzahl groß ist und die Splits nur so lange fortgesetzt werden, wie die Fallzahlen in den entstehenden Subgruppen ausreichend groß sind.

Neben der Gefahr, daß das Verfahren auf ein zufälliges Pattern in den Daten zu stark reagiert, sehen sie ein weiteres Problem in der Benutzung derselben, mit SEARCH gewonnenen Kontrollvariablen für verschiedene abhängige Variablen. Dieses Problem ließe sich entweder über eine Kompromißlösung aus unabhängigen SEARCH-Analysen oder mit Hilfe der multivariaten Version MAID-M (Multivariate Automatic Interaction Detection - Monitored) lösen.

Die bei KALTON/SANTOS angesprochenen Probleme bei der Benutzung einer AID-Technik berühren jedoch nicht den Kern des Problems bei der Bestimmung der Imputationklassen. Das Hauptproblem besteht zwar technisch in der Suche nach homogenen Gruppen, gesucht wird aber nach in Hinsicht auf den Ausfallmechanismus homogenen Gruppen.

AID ist nur dann in der Lage "sinnvolle" Imputationklassen zu entdecken, wenn

- die MARC-Annahme korrekt ist und
- zwischen Split-Variablen und Ausfallmechanismus ein enger Zu sammenhang besteht.

In der Literatur werden diese Probleme zumeist anders beurteilt. OH/SCHEUREN (1983:157-158) beispielsweise halten zwar einerseits die Bestimmung der Subpopulationen in Hinsicht auf konstante Responsewahrscheinlichkeiten für das fundamentale Problem. Andererseits betrachten sie aber selbst bei falscher Spezifizierung des Responsemechanismus die Wahl von Gruppen, die in Hinsicht auf die interessierenden Variablen für die Nonrespondenten kleine Binnengruppenvarianzen und große Zwischengruppenvarianzen zeigen, als unkritisch.

Die Arbeit von CHAPMAN (1983) stellt in dieser Hinsicht eine Verbesserung dar. Abhängige Variable für die AID-Analysen bildet die (0-1)-Response/Nonresponse-Variable. Neben einer Reihe von interessanten Nebenergebnissen (geringe Varianzerklärung durch standarddemographische Variablen, Instabilität der AID-Splits (S.442: Tab.2) erwähnt CHAPMAN eine einfache Splittechnik, die die Standardvariablen zugrunde legt und die erzielten Kovarianzen zwischen den Zellengewichten (in der Regel reziproke Responserate der Zelle) und den Zellenmittelwerten ("relcov") zur Wahl der Klassen heranzieht. Ein Vergleich von AID-Splits der Response/ Nonresponse-Variablen

mit den erzielten Relcovs kann damit Gruppen hervorbringen, die homogene Responsewahrscheinlichkeiten und möglichst starke Mittelwertdifferenzen in bezug auf die zu vervollständigende Variable besitzen.

Die Frage nach der Gültigkeit der beiden oben erwähnten Annahmen bleibt aber hiervon unberührt. Beide Annahmen können mit den vorhandenen Daten nicht geprüft werden, eine Begründung kann nur über eine theoretische Erklärung[46/] des Ausfallprozesses erfolgen. AID kann zwar bei Entwicklung eines Ausfallmodells nützlich sein, eine automatische Erstellung der Imputationklassen erscheint aber nur dann vertretbar, wenn das Ausmaß fehlender Werte so gering ist, daß selbst bei Annahme des "ungünstigsten" Ausfallmechanismus der zu erwartende Einfluß auf die Ergebnisse so gering ist, daß die Wahl der Imputationmethode und der zugewiesenen Werte praktisch beliebig erfolgen könnte.

5.2.2.1.4 Kritik des Imputationklassen-Modells

> "Eine weitere Überlegung zeigt indessen, daß diese einfachen statistischen Erwägungen, die wir ausprobiert haben, nicht viel taugen."
>
> (S.LEM: Das Katastrophenprinzip, (1983:35))

In der Literatur scheint sich nur eine radikale Kritik auf das Konzept homogener bzw. konstanter[47/] Responsewahrscheinlichkeiten zu finden: DALENIUS (1983:414) neigt zur Ablehnung aller Ansätze, die auf Responsewahrscheinlichkeiten basieren. DALENIUS führt neben der Unmöglichkeit der Berechnung dieser Wahrscheinlichkeiten im Rahmen eines Surveys vor allem soziologische Argumente gegen das Konzept an, insbesondere die mögliche Existenz einer Reihe von Responsewahrscheinlichkeiten für jedes Individuum [48/]. So wird z.B. die Interviewsituation als determinierender Faktor erwähnt[49/].

Die zentrale Annahme des Imputationklassenkonzepts stellt die Vermutung dar, daß Gruppen existieren, die in sich völlig homogen in bezug auf die interessierenden Variablen sind und deren Mitglieder sich nur rein zufällig für oder gegen eine Antwort entscheiden[50/]. Da nach der Erhebung für jedes Sampleelement das Ergebnis des Entscheidungsprozesses bekannt ist, kann zwar für in bezug auf vorhandene Variablen homogene Gruppen die "Antwortwahrscheinlichkeit" für diese Gruppen berechnet werden; die Betrachtung dieser "Aggregatstatistik" als Individualmerkmal hingegegen ist eine Annahme, die durch die Daten nicht gerechtfertigt werden kann[51/]. Die Konsequenzen der Annahme zeigen sich deutlich, wenn es um die empirische Bestimmung der Klassen geht (s.w.o.): Schon aus Gründen der Praktikabilität werden die Klassen häufig durch die Kreuztabellierung demographischer Variablen gewonnen[52/]. An dieser Stelle ergeben sich eine Reihe von möglichen Einwänden:

-Ein empirisches Argument für den Schluß von der Homogenität in bezug auf demographische Variablen auf Homogenität der Antwortwahrscheinlichkeiten kann nicht vorgebracht werden.
-Läuft die Begründung für die Imputationklassenbildung über das Argument der vermuteten Homogenität der Klasse in bezug auf die interesserienden Variablen, so kann hierfür kein empirisches Argument angeführt werden, das aus dem aktuellen Datensatz stammt. Wird das Argument durch andere Datensätze gestützt, bleibt angesichts dieser - bestenfalls unvollständigen-Erklärung[53/] immer noch die Frage nach der Gültigkeit der Behauptung der Konstanz der Randbedingung.
- Auf eine Erklärung des Ausfallmechanismus wird verzichtet.

Beide Homogenitätsannahmen bedienen sich impliziter Hypothesen über die "Wirkung" demographischer Variablen auf andere Merkmale der Befragten. Hierbei werden wie in der traditionellen "Variablensoziologie"[54/] Globalvariablen benutzt. Die Ergebnisse sind vergleichbar: niedrige Korrelationen, Instabilität der Ergebnisse, "conditional relationships" (ROSENBERG 1968:106). Die zwar relativ beständigen, aber schwachen Zusammenhänge der Globalvariablen mit in der Regel in analytischen Surveys interessierenden Variablen sind dadurch

verständlich, daß lediglich "Aggregationseffekte" handlungsrelevanter Variablen erhoben werden[55/]. So lassen sich vermutlich die z.B. bei CHAPMAN (1983:442) erwähnten niedrigen erklärten Varianzen dadurch erklären, daß die den Ausfallprozeß determinierenden Variablen nur indirekt erfaßt worden sind[56/].

Falls ein prinzipiell nicht kumulationsfähiges, empiristisches Vorgehen vermieden werden soll, müssen also die in den Imputationmodellen impliziten Hypothesen expliziert werden, um den Anwendungsbereich dieser Modelle abschätzen zu können: Ändern sich die den Aggregationseffekt bestimmenden Prozesse, ist ein rein empirisch gewonnenes Modell des Prozesses unbrauchbar.

Operational bedeutungslos wären die zu entwickelnden Ausfallmodelle nur dann, wenn entweder das Ausmaß der Ausfälle so gering ist, daß sich keinerlei Effekte der MD auf beliebige Statistiken oder ein starker "Aggregationseffekt" eines komplexen Mechanismus auf eine allgemein verfügbare Variable zeigen ließe.

5.2.2.2 Ersetzung durch den Gruppenmittelwert

Die beiden einfachsten Anwendungen des Imputationklassenkonzeptes zur Ersetzung fehlender Werte stellen die naheliegenden Modifikationen der entsprechenden nicht-informativen Techniken dar. Die Anwendung der Mittelwertersetzung innerhalb der jeweiligen Imputationklassen, also die Ersetzung durch den Mittelwert der Zelle einer Klassifikation der Untersuchungseinheiten, wird in der Literatur zumeist als "Cellmeansubstitution", gelegentlich als "Groupmeansolution" bezeichnet. Ähnlich wie die Verwendung der nicht-informativen Ersetzung durch einen zufällig gezogenen vorhandenen Wert (Random Imputation) anstelle der Mittelwertersetzung um die Varianz der zu vervollständigenden Variablen zu erhalten, kann die zufällige Ziehung innerhalb der Imputationklasse erfolgen: "Random imputation within classes" (KALTON/ KASPRZYK 1982:23).

5.2.2.3 Ratio-Schätzung

Besteht zwischen zwei Variablen x und y eine sehr hohe Korrelation und ist x in der Regel "more observed" (RUBIN 1974:468) als y, bietet sich als einfaches Schätzverfahren die Ratioschätzung an. Hierbei wird für den vollständigen Teil des Samples das Ratio der beiden Variablen berechnet (R = @y/@x) und für den unvollständigen Teil des Samples die MD geschätzt, indem die vorhandenen Werte der einen Variablen mit dem Ratio multipliziert[57/]. werden (y' = Rx). Fehlt für einen Fall der entsprechende Wert auf x, so kann entweder das Ratio von y und einer anderen mit y hochkorrelierenden Variablen z benutzt werden, oder man verwendet eine mit x hochkorrelierende Variable als Näherung für den fehlenden Wert auf x. Das Verfahren kann modifiziert[58/] werden, um Editrestriktionen Rechnung zu tragen. KALTON (1983:80) legt die Verwendung der Ratioschätzung für fehlende Paneldaten nahe.

5.2.2.4 Zellen-Ratio-Schätzung: "Array"-Prozedur

Eine Verfeinerung der Ratio-Schätzung erhält man, wenn das Ratio nicht über alle Fälle, sondern innerhalb der Imputationklassen berechnet wird. In der von FORD/KLEWENO/TORTORA (1980:252) beschriebenen Version werden zwei Variablen (a_1,a_2) zur Klassifizierung verwendet. Für jede Zelle der durch a_1 und a_2 gebildeten Tabelle wird das Ratio der zu vervollständigenden Variablen y zu einer Variablen a_3 berechnet und dieses zellenspezifische Ratio zur Schätzung von y verwandt.

Bei der von FORD/KLEWENO/TORTORA beschriebenen Version der Array-Prozedur handelt es sich um ein "One pass"-Verfahren, d.h. der Datensatz wird nur einmal gelesen. Dies setzt die Initialisierung der Zelleninhalte mit Schätzungen

der Ratios (z.B. Schätzungen auf Grund vorheriger Surveys) sowie eine sequentielle Veränderungsmethode der Zelleninhalte ("Updating") voraus. Das Updating erfolgt bei FORD/KLEWENO/TORTORA über eine Gewichtungsformel, die dem alten Zelleninhalt das doppelte Gewicht wie dem neu berechneten Ratio des neuen Falls zuordnet, um die Gefahr der Verzerrung durch Ausreißer zu vermindern. Das Verfahren beginnt mit einer ursprünglichen Schätzung der Zellenratios und arbeitet sequentiell durch den Datensatz. Ist die in Frage stehende Variable für einen neuen Fall vorhanden, so wird für die zugehörige Zelle $R(a_1,a_2)$ das Ratio $r = y/a_3$ berechnet und die Zelle neu mit $2 * R(a_1,a_2) + r) /3$ besetzt. Ist y für den neuen Fall nicht vorhanden, erfolgt die Schätzung $y' = a_3 * R(a_1,a_2)$. In dieser Version ist das Verfahren natürlich nicht invariant gegenüber der Abfolge der Fälle im Datensatz.

5.2.2.5 Cold-deck-Techniken

Die insbesondere in der amtlichen Statistik bedeutendsten Ersetzungstechniken für fehlende Werte werden unter dem Begriff der "Deck"-Techniken zusammengefaßt. Es wird zwischen Cold- und Hot-Deck-Techniken unterschieden; die Hot-Deck-Techniken bilden vor allem unter dem Einfluß der theoretischen Arbeiten von Statistics Canada in der neueren Literatur den Schwerpunkt der z.Z. verwendeten Ersetzungstechniken in der amtlichen Statistik. Auf Grund der historischen Bedeutung sollen zunächst aber die Cold-Deck-Techniken erwähnt werden.

Werden MD einer Erhebungseinheit pro Variable durch die Zuweisung eines Wertes aus einer vor Durchführung der Ersetzung festgelegten Verteilung ersetzt, so wird dieses Verfahren als "cold deck" bezeichnet (DEINIGER/KERLER 1971:298; CHAPMAN 1976:245).

Die Verteilung wird in der Regel aus einer früheren Untersuchung der selben Population übernommen. Die zur Ersetzung benutzte Variable wird mit Klassifikationsvariablen in Imputationklassen ("Deck" bei ZINDLER 1967:210)

mit einem Wert der festgesetzten Verteilung als Zelleninhalt, z.B. dem Median der entsprechenden Variablen in einer früheren Untersuchung, eingeteilt. Das Verfahren wird für jede zu vervollständigende Variable getrennt durchgeführt, für jede unvollständige Variable kann eine andere Klassifikation bestehen. Die Ersetzung erfolgt für jede unvollständige Erhebungseinheit mit dem Wert der Zelle, die den Klassifikationsvariablen des Falls entspricht. Die Zuordnung erfolgt deterministisch: Jeder Fall innerhalb der jeweiligen Imputationklasse erhält denselben, vor der Ersetzung festgelegten Wert. Die Cold-deck-Technik stellt einen Spezialfall der stochastischen "Monte-Carlo"-Methode (ZINDLER 1967:223) dar, bei der für jede Imputationklasse eine Verteilungsfunktion gespeichert wird. Die MD-Ersetzung erfolgt über eine zufällige Ziehung gemäß der Verteilungsfunktion der jeweiligen Imputationklasse[59/].

5.2.2.6 Hot-deck-Techniken

Es gibt keine feststehenden Definitionen der Hot-deck-Verfahren (FORD 1983:185-186); den Definitionsversuchen gemein ist die Zuschreibung des stochastischen Charakters des Verfahrens, die Verwendung von Daten des aktuellen Datensatzes zur Zuschreibung und die Betonung des Dopplungscharakters der Ersetzung. Die Probleme der Definition werden deutlich, wenn zwischen zwei Grundtypen unterschieden wird: den sequentiellen und den simultanen Verfahren.

Die sequentiellen Verfahren ("traditional hot-deck": KALTON/ KASPRZYK 1982:23) beginnen mit der Festlegung der Imputationklassen und einer ersten Besetzung der Zellen. Für diese Startwertefestlegung existieren mehrere Vorschläge; einer der einfachsten besteht in der Zuweisung des Zellenmittelwertes aus früheren Untersuchungen. Der Datensatz wird dann sequentiell abgearbeitet. Für jeden Fall des Datensatzes wird die zugehörige Imputationklasse mit den Klassifikationsvariablen[60/] bestimmt. Ist ein gültiger Wert für den Fall auf der zu vervollständigenden Variablen vorhanden, so wird der aktuelle Wert dieses Falles ("Donor")[61/] zum Inhalt der Imputationklasse.

Fehlt für den aktuellen Fall der Wert der Variablen, so wird dem Fall ("Rezipient" oder "Kandidat") der Wert der Imputationklasse zugewiesen (der Name "Hot-Deck" geht auf diese andauernde Veränderung des Zelleninhalts durch im aktuellen Datensatz vorhandenen Werte zurück).

Simultane Verfahren beruhen auf der Zuweisung von Werten eines Falles, der die geringste Distanz in bezug auf die "matching variables"[62/] zum zu vervollständigenden Fall besitzt.

Die Schwierigkeit einer Definition der Hot-Deck-Techniken resultiert aus den fließenden Übergängen der Verfahren: Fast jede Technik besitzt als "degenerierten" Spezialfall zumindest eine Methode, die üblicherweise nicht als Hot-Deck-Technik bezeichnet wird; selbst die Unterscheidung zwischen sequentiellen und simultanen Verfahren ist nicht völlig trennscharf.

Als Beispiel sei das von COX (1980) vorgestellte "weighted sequential" Hot-Deck erwähnt, das zwar sequentiell vorgeht, aber über ein sequentielles gewichtetes Samplingverfahren verfügt, so daß jeder potentielle Donor die Chance hat, tatsächlicher Donor zu werden, und die zugewiesen Werte dieselben gewichteten Statistiken besitzen, die auch dem vollständigen Sampleteil zukommen. Das Verfahren setzt mehrfachen[63/] Zugriff zu den Daten voraus und ist damit kein One-pass-Verfahren wie das traditionelle Hot-Deck.

Ausgehend von den einfachen Typen der traditionellen Hot-Deck- Technik lassen sich Abwandlungen und Probleme der Verfahren darstellen.

Eines der Hauptprobleme der Deck-Techniken liegt darin begründet, daß sequentielle Techniken gegenüber der Abfolge der Fälle im Datensatz nicht invariant sind. Die Wünschbarkeit dieser Eigenschaft ist umstritten[64/]. Einerseits gehen damit nur schwer kontrollierbare Effekte in den Imputationprozeß ein,

andererseits kann der Abfolgeeffekt die Wirkung einer nicht explizit definierten zusätzlichen Klassifikationsvariablen haben und damit einen besseren Match bedingen.

Eine andere Schwierigkeit liegt in den eventuell spezifizierten Konsistenztests begründet. Falls mehrere Felder pro Fall durch Daten unterschiedlicher Donoreinheiten vervollständigt werden, ist es möglich, daß die vervollständigten Fälle den Editbedingungen nicht mehr genügen (SANDE 1979:244). Die einfachste Lösung des Problems besteht in der Zuweisung aller Werte des Donors für alle MD-Felder eines Falles. Es ist möglich, daß selbst hierbei Fälle auftreten, die den Editbedingungen nicht genügen. Die einzige verbleibende Möglichkeit stellt dann das Löschen und gemeinsame Ersetzen mehrere Felder dar (SANDE 1979:245).

Bei differenzierten Einteilungen in Imputationklassen kann weiterhin das Problem auftreten, daß kein "exact match" zwischen einer Reihe von zu vervollständigenden Fällen und den ursprünglichen Imputationklassen möglich ist oder daß die Imputationklasse "leer" ist, d.h. daß kein Fall mit gültigen Werten auftritt. CODER (1978) und WELNIAK/CODER (1980:422) beschreiben für dieses Problem ein "flexible matching imputation"-System, daß den Match hierarchisch herbeiführt, indem bei unmöglichem Match auf der ersten Stufe dann sukzessiv Imputationklassen zusammengefaßt werden, bis ein Match möglich ist.

Das traditionelle sequentielle Hot-Deck wird überwiegend für kategoriale Daten benutzt. SANDE (1979:246) führt die relativ seltene Benutzung sequentieller Hot-Deck-Techniken für die Ersetzung numerischer Daten[65/] überwiegend auf Schwierigkeiten mit den Editbedingungen zurück. Werden numerische Variablen als Matchvariablen benutzt, ergeben sich zwei Möglichkeiten: Entweder müssen die Variablen klassifiziert werden (KALTON/KASPRZYK 1982:22), oder es müssen Distanzfunktionen spezifiziert werden. Im Falle der Verwendung von Distanzfunktionen handelt es sich zumeist[66/] um Versionen der simultanen Hot-Deck-Technik, damit um Verfahren mit einer "nearest neighbor rule", d.h. die MD pro Fall werden durch die Werte desjenigen Falles ersetzt, der die geringste

Distanz aufweist[67/] (ergeben sich "ties", erfolgt eine Zufallsauswahl aus den Fällen mit gleicher Distanz; FORD 1976:326). Für die Spezifizierung der Distanzfunktion gibt es eine Reihe von Vorschlägen, so z.B. maximale absolute Differenz über alle Felder, euklidische Distanz etc.. Auch eine Gewichtung der Variablen ist möglich. Interessant ist die Möglichkeit, numerische und kategoriale Distanzfunktionen zu kombinieren (SANDE 1979:248-249)[68/]. Die Wahl der Distanzfunktion scheint nach den Erfahrungen von Statistics Canada keinen sehr großen Einluß auf die Leistungsfähigkeit des Imputationsystems zu besitzen (SANDE 1979:248). Relativ oft findet sich in der Literatur der Vorschlag, die Häufigkeit, mit der ein Fall als Donor benutzt wird, mit in die Distanzfunktion aufzunehmen, um so die Möglichkeit extrem häufiger Benutzung eines Falls als Donor zu verhindern. Dies erscheint aus Gründen der "Varianzerhaltung" und der Gefahr der Zuweisung von Extremwerten wünschenswert. Andererseits kann die gleichmäßige Donorbenutzung einen Bias nach sich ziehen[69/].

5.2.2.7 Spezielle Hot-Deck-Techniken

Die Verwendung von Distanzfunktionen zur Bestimmung des Donors für einen gegebenen Rezipienten erlaubt die Entwicklung zahlreicher spezieller Hot-Deck-Versionen (SANDE 1983:345). Zwei dieser Vorschläge haben größere Aufmerksamkeit gefunden.

VACEK/ASHIKAGA (1980:326) beschreiben eine Verfeinerung der Hot-Deck-Technik mit der "Nearest Neighbor Rule". Die Methode besteht in der Ersetzung fehlender Werte eines Falls i durch die vorhandenen Werte desjenigen Falls j, der die geringste Distanz (auf den vorhandenen Variablen des Falls i) zum Fall i aufweist. Sie schlagen die Verwendung der Mahalanobis-Distanz[70/] folgender Form vor:

5-2 $$d_{ij} = (X_i - X_j)' \cdot s^{-1}(X_i \cdot X_j) ,$$

wobei X_i den 1*k-Vektor der vohandenen Werte des Falls i, X_j den entsprechenden 1*k-Vektor des vollständigen Falls j und S^{-1} die Inverse der k*k Kovarianzmatrix, die aus den vollständigen Fällen berechnet wird, darstellen.

Ein Hot-Deck-Verfahren mit einer ungewöhnlichen Distanzfunktion wird von FORD/KLEWENO/TORTORA (1980:253-254) beschrieben: die "PRINCOMP"-Methode. Das Verfahren beginnt mit einer Partitionierung des Datensatzes in eine Gruppe mit vollständigen Daten (S1) und Gruppen mit unvollständigen Daten (S2,S3...SP)[71/] entsprechend dem Muster der fehlenden Werte. Für jede Gruppe, die jeweils ein MD-Muster aufweist, wird die erste Hauptkomponente berechnet. Anschließend wird für jedes Element aus S1 und für jedes Element aus der aktuellen Gruppe der Hauptkomponentenwert berechnet. Dann wird für jedes Element aus der unvollständigen Gruppe dasjenige Element in S1 gesucht, das die geringste absolute Differenz des Hauptkomponentenwertes aufweist. Die fehlenden Werte des Falles aus der unvollständigen Gruppe werden durch die Werte des Falles aus S1 ersetzt, der dem unvollständigen Fall am nächsten ist. Das Verfahren stellt damit eine Version eines Hot-Deck-Verfahrens mit "nearest neighbor rule" dar.

5.2.2.8 Exkurs III: "Zero-Spike"-Methoden

PREGIBON (1976) macht auf ein Problem bei der Schätzung fehlender Werte bei "typischen" Surveydaten aufmerksam. Viele quantitative Surveyvariablen besitzen in ihrer Häufigkeitsverteilung ein lokales Maximum an der Stelle Null, m.a.W. die Variable hat für viele Fälle die Ausprägung "Null", weil die Variable für diese Fälle nicht zutrifft. Fehlt nun ein Wert für die Variable, stellt sich die Frage, ob "trifft nicht zu" zutrifft oder ob ein Wert "existiert" und geschätzt werden sollte. Die Imputationaufgabe besteht dann aus zwei Schritten: Schätzung, ob ein Wert zugewiesen werden sollte, und gegebenenfalls Schätzung des fehlenden Wertes.

Eine einfache Version[72/] einer Zero-Spike-Prozedur für jeweils eine MD-behaftete Variable x beginnt mit der Erstellung einer trichotomen Indikatorvariablen z. Für jeden Fall i ist $z_i = 0$, falls $x_i = 0$; $z_i = 1$, falls $x_i > 0$; $z_i = 2$, falls x_i fehlt. Weiterhin wird für die vollständigen Fälle eine dichotome Indikatorvariable y, wobei $y_i=0$, falls $x_i = 0$; $y_i = 1$, falls $x_i > 0$ erstellt. Die Bildung der Indikatorvariablen erfolgt für jede in die Prozedur einbezogene Variable (k-mal). Damit ergeben sich maximal 2_k Muster der vollständigen Fälle ("domains") - Indikatorvektoren der Länge k - und 3_k-2_k Muster der unvollständigen Fälle ("states"); wiederum Indikatorvektoren der Länge k.

Unter der Annahme, daß jeder Fall des Samples zu einem Domain gehört[73/] kann die bedingte Wahrscheinlichkeit, daß ein Fall mit einem bestimmten State aus einem bestimmten Domain stammt, mit dem Bayestheorem berechnet werden[74/]. Jeder Fall i mit fehlenden Werten auf der Variablen x wird nun mit den bedingten Wahrscheinlichkeiten der Domains für den gegebenen State einem Domain zugewiesen[75/]. Weist der Domainvektor[76/] an der Stelle der Variablen x eine Null auf, so wird der fehlende Wert x_i durch Null ersetzt, ansonsten wird x_i mit einem der üblichen Schätzverfahren geschätzt[77/].

SCHIEBER (1978:214) behandelt ein ähnliches Problem: die Schätzung von fehlenden Angaben zu "Social Security benefits". Hierbei verwendet SCHIEBER drei Methoden zur Bestimmung, ob einem Fall ein von Null verschiedener Wert zugewiesen werden soll (Empfänger von "benefits" oder Nichtempfänger).

Die erste Methode besteht aus einer Logit-Analyse der vollständigen Fälle mit der dichotomen "Empfänger"-Variablen als abhängiger Variablen. Die am vollständigen Sample gewonnenen Koeffizienten werden dann für die Gleichung der unvollständigen Fälle verwendet. Fälle mit einer vorhergesagten "Wahrscheinlichkeit" größer als .5 werden als Empfänger betrachtet.

Die zweite Methode besteht aus der Ziehung aus einem (0,1)-Zufallszahlengenerator für jeden unvollständigen Fall. Liegt der gezogene Wert über dem Anteilswert der "Nichtempfänger" im vollständigen Sample innerhalb der Imputationklasse, zu der der unvollständige Fall gehört, wird der Fall als "Empfänger" betrachtet.

SCHIEBER (1978:214) verwendete als dritte Methode lineare Modelle anstatt des Logit-Modells. Beide Modellformen erbrachten in dem Datensatz von SCHIEBER die Zuschreibung des "Empfängerstatus" für nahezu alle unvollständigen Fälle. Trotz dieses schlechten Ergebnisses erscheint die Logit-Methode theoretisch wünschenswerter als die Ziehung aus einem (0,1)-Generator, da das Logit-Modell auf eine explizite Modellierung abzielt[78/].

5.3 Multivariate Ersetzungstechniken

Im folgenden Abschnitt sollen die wichtigsten Ersetzungsverfahren, die unmittelbar auf der Verwendung traditioneller multivariater Methoden basieren, dargestellt werden, ohne daß explizit auf die Übergänge sowohl zwischen einzelnen multivariaten Techniken als auch zu den bereits besprochenen ML-Schätzern oder den informativen Ersetzungsverfahren als "degenerierten" Spezialfällen Bezug genommen wird.

5.3.1 Hauptkomponenten-Verfahren: Methode nach DEAR

DEAR (1959) wird eine Ersetzungsmethode zugeschrieben, die auf der Berechnung der Hauptkomponenten basiert[79/]. Ausgehend von der "mean imputed"-Korrelationsmatrix[80/] wird die erste Hauptkomponente der Matrix berechnet. Der normierte Eigenvektor g wird dann zur Schätzung verwandt, wobei mit der standardisierten Datenmatrix Z gearbeitet wird:

5-3 $$\hat{z}_{ij} = g_j \cdot \sum_{k=1}^{p} z_{ik} \cdot g_k$$

Hierbei bezeichnet z den standardisierten Wert des Falls i auf der Variablen j und g_j das zugehörige Element des Eigenvektors[81/]. Fehlt bei Berechnung ein z_{ik}, wird für die Berechnung der Mittelwert angenommen: $z_{ik} = 0$.

Eine Reihe von Modifikationen liegen nahe[82/]. Zuerst kann jede der einfachen Schätzungen der Korrelationsmatrix als Anfangsmatrix verwandt werden: LISTWISE, PAIRWISE, MEAN[83/]. Weiterhin können mehr als nur eine Hauptkomponente benutzt werden, wobei das Verfahren im Grenzfall fast in die multiple Regressionsschätzung übergeht[84/]. Darüber hinaus kann das Verfahren iterativ durchgeführt werden, also die Berechnung so lange fortgesetzt werden, bis zwei aufeinanderfolgende Korrelationsmatrizen keine wesentlichen Differenzen zeigen[85/] (GLEASON/STAELIN 1975:238- 239).

5.3.2 Regressionsschätzverfahren: Methode nach BUCK

Zu den am häufigsten zitierten Methoden der MD-Schätzung gehört die multiple Regressionsschätzung nach BUCK(1960). BUCK berechnet die p*p-Korrelationsmatrix für die vollständigen Fälle und berechnet dann für jeden Fall i und für jede fehlende Variable j ($j=1,2...m_i$) die Regressionsschätzungen mit allen verfügbaren Variablen ($p-m_i$) des Falls i. Falls jede Kombination fehlender Variablen möglich ist und k Variablen fehlen, müssen maximal #### Regressionsgleichungen berechnet werden (BUCK 1960:303). Da die vorhergesagten Werte alle auf der Regressionsgraden liegen, ist eine Verzerrung der dann aus den geschätzten Daten berechneten Korrelationsmatrix zu erwarten (der geschätzte Mittelwertvektor ist bei MAR unverzerrt). BUCK entwickelte daher eine Korrektur der geschätzten Kovarianzmatrix. Bezeichnet man die aus den vollständigen und den geschätzten Fällen berechnete Kovarianzmatrix mit A

und fehlt die Variable x_j für einen Anteil l_j des Samples, dann wird die Varianz a_{jj}' korrigiert, indem $a_{jj}' = a_{jj}' + l_j / c_{jj}$ berechnet wird, wobei c_{jj} das Diagonalelement in A^{-1} darstellt (BUCK 1960:304)[86/]. Diese von BUCK beschriebene Korrektur [87/] wird in der Regel in der Literatur nicht angeführt, sondern jede multiple Regressionsschätzung als "BUCK's method" bezeichnet[88/].

5.3.3 Weitere Regressionsschätzverfahren

Eine Reihe von Modifikationen und Erweiterungen des Ansatzes von BUCK wurden in der Literatur vorgeschlagen[89/]. Unabhängig von BUCK entwickelte WALSH (1961) ein multiples Regressionsverfahren, das die gesamte vorhandene Sampleinformation benutzt. Das Verfahren ist höchst aufwendig[90/] und scheint niemals getestet worden zu sein (GLEASON/STAELIN 1975:232). Wie oben erwähnt, geht BUCK von der Korrelationsmatrix der vollständigen Fälle aus. Schon vor BUCK wurde eine multiple Regressionsvorhersage von FEDERSPIEL/ MONROE/GREENBERG (1959) vorgeschlagen[91/]. Ausgehend von einer durch Mittelwertersetzung vervollständigten Datenmatrix werden alle fehlenden Werte durch multiple Regressionschätzungen ersetzt. Aus dem vervollständigten Datensatz wird eine neue Korrelationsmatrix berechnet und die Schätzung der MD solange wiederholt, bis ein Konvergenzkriterium erreicht wird.

Neben der Möglichkeit der Iteration, die sich z.B. in den Programmen von BERGER (1979) und O'GRADY(1982) finden, können verschiedene Ausgangsmatrizen Verwendung finden. Außer der Matrix der vollständigen Fälle finden sich die MEAN-Matrix und die PAIRWISE-Matrix (beide bei BERGER 1979 und O'GRADY 1982). Da die PAIRWISE-Matrix häufig nicht positiv-semidefinit ist, empfiehlt FRANE (1978:31) ein "Glätten" der Ausgangsmatrix (vgl. Kapitel 5.1.2). GLEASON/STAELIN (1975:245) benutzen die Ergebniskorrelationsmatrix[92/] ihrer Version des Hauptkomponentenverfahrens als Ausgangsmatrix.

Andere Formen der Regression sind ebenfalls denkbar. FRANE (1976:413-414; 1978:32) schlägt die Benutzung einer schrittweisen Regression als Schätzverfahren vor[93/] um einen "Overfit" durch die multiple Regression zu vermeiden. Ebenfalls von FRANE (1978: 32) stammt der Vorschlag, die einfache Regressionsvorhersage nicht nur durch die Korrelation der Prädiktorvariablen mit der zu vervollständigenden Variablen steuern zu lassen, sondern auch die paarweise vorhandenen Fallzahlen bei der Auswahl zu berücksichtigen, um die Auswahl gegenüber Verzerrungen durch extrem kleine Fallzahlen robuster zu machen. Von CHAN/DUNN (1972:474) stammt eine Modifikation der multiplen Regressionsschätzung[94/]. Sie schlagen vor, zunächst alle Fälle mit genau einem fehlenden Wert über die BUCK-Prozedur zu schätzen und dann die ursprünglich vollständigen Fälle mit den nun vervollständigten Fälle gemeinsam als Ausgangsdatenmatrix für eine Schätzung aller Fälle mit genau zwei fehlenden Werten zu verwenden. Die vervollständigte Datenmatrix dient dann zur Vorhersage der Fälle mit genau drei fehlenden Werten usw., bis alle fehlenden Werte ersetzt sind. HAMILTON (1975:25-26) schlägt eine weitere "multistage" Regressionschätzung vor: Ausgehend von einer a-prioiri-Reihenfolge der Wichtigkeit ("importance") der Variablen (y1...yp) werden zunächst die fehlenden Werte der Variablen y1 mit BUCK's Methode geschätzt. Aus der ursprünglichen vollständigen Datenmatrix und den nun vervollständigten Fällen wird eine neue Kovarianzmatrix für die Schätzung der Variablen y2 mit der BUCK-Methode berechnet usw., bis alle fehlenden Werte ersetzt sind.

5.3.4 Spezielle Modifikationen

Für das oben angesprochene Problem der erwarteten Varianzunterschätzung der Variablen durch die multiple Regressionsvorhersage der fehlenden Werte gibt es einen weiteren Lösungsvorschlag, der nicht auf die Korrektur der Kovarianzmatrix abzielt, sondern die MD-Schätzungen verändert. Die multiplen Regressiongleichungen werden wie üblich berechnet und auch die Schätzung fehlender Werte erfolgt zunächst auf die übliche Weise. Zu jedem geschätzten

Wert wird dann ein "random residual" addiert, so daß die geschätzten Werte um die Regressionsgerade streuen. Anders formuliert: Während bei "deterministic imputation" der Fehlerterm bei der Regressionsvorhersage gleich Null gesetzt wird, so wird bei "stochastic imputation" der Fehlerterm aus einem Zufallsprozeß gewonnen. Für die Gewinnung des Fehlerterms existieren mehrere Vorschläge. KIM/CURRY (1977:233) erwähnen die Addition der mit einer nicht näher spezifizierten Zufallszahl multiplizierten Residualvarianz, SANTOS(1981:21) beschreibt die Ziehung aus einem N(0,S)-Generator, wobei S der Residualvarianz der Regression bei den vollständigen Fällen entspricht. SANTOS (1981:21) schlägt weiterhin vor, für jeden unvollständigen Fall einen vollständigen Fall zufällig auszuwählen und das Residuum dieses Falls zur Schätzung des unvollständigen Falls zu addieren. Bei KALTON/ KASPRZYK (1982:24) findet sich der Vorschlag, das Residuum einem vollständigen Fall mit ähnlichen Werten auf den unabhängigen Variablen zu entnehmen, um eventuellen Nichtlinearitäten Rechnung zu tragen. Bei einem exakten "match" wird damit dem unvollständigen Fall der Wert des vollständigen Falles zugewiesen. SCHIEBER (1978:214) verwendet eine Hot-Deck-Technik mit Residuen als Zelleninhalt[95/], LITTLE/SAMUHEL (1983:419) benutzen eine systematische Auswahl einer nach Klassen gegliederten und der Größe nach sortierten Residualvariablen der vollständigen Fälle.

Neben der Möglichkeit der Korrektur der Varianzunterschätzung finden sich Vorschläge zur Einbeziehung nichtlinearer Terme (TIMM) sowie die Benutzung der MD-Häufigkeit als weiterer Prädiktorvariablen. HAITOVSKY(1968:68) zitiert eine mündliche Bemerkung von A.E.BEATON, der den Vorschlag machte, im Falle der Verletzung der MAR-Annahme zusätzliche Indikatorvariablen zu berechnen und in die Regressionsgleichungen Interaktionsterme aufzunehmen, um unterschiedliche Steigungen der Regressionsgraden für verschiedene Subgruppen zu berücksichtigen[96/]. Die Erwähnung anderer Regressionstechniken als entweder der einfachen Regression bzw. der multiplen Regression nach BUCK ist allerdings selten, die entsprechenden Verfahren[97/] scheinen zudem bisher nicht systematisch getestet worden zu sein.

5.4 Exkurs IV: Missing Data in Skalen und Indizes

In vielen Surveys stellen Skalen und Indizes einen Großteil der Variablen. Neben den bereits erwähnten Ersetzungsmethoden finden sich in der Literatur speziell für Skalen und Indizes weitere Methoden.

Leistungstests

Leistungstests ("speed-and-power-tests") sollen die Fähigkeiten eines Probanden auf einer latenten Variablen messen. In der Regel gilt es, eine möglichst große Zahl von Aufgaben in begrenzter Zeit richtig zu lösen. Unter diesen Bedingungen ist die Ersetzung eines "fehlenden" Wertes trivial: Wurde das entsprechende Item nicht gelöst, wird (falls der Testwert die Summe der richtig gelösten Items ist) das Item als falsch gelöst gewertet. Falls die Probanden die Logik des Tests verstanden haben, liegt es in ihrem eigenen Interesse, die von ihnen nicht bearbeiteten Teile des Tests durch pures Raten auszufüllen[98/]. Die damit notwendige Korrektur für geratene Antworten benachteiligt dann allerdings diejenigen Probanden, die nicht geraten haben, sondern keine Lösung anboten. Da die Item-response-Theorie auf diese Art der Skalierung nur dann angewandt werden kann, wenn kein Zeitlimit gegeben ist, rät LORD (1980:226) von der Verwendung der Zeitlimits in diesem Zusammenhang ab. Falls der Test nach LORD für die geratenen Antworten mit

5-4 $$y = x - \frac{w}{A-1}$$

(y = korrigierter Score
x = Anzahl richtiger Items
w = Anzahl falscher Items
A = Anzahl der Antwortmöglichkeiten pro Item)

(LORD 1980:226)

korrigiert wurde, können nicht beantwortete Items nicht einfach ignoriert werden, da sonst ein Proband einen hohen Wert einfach dadurch erreichen kann, daß er lediglich diejenigen Items beantwortet, deren richtige Antwort er mit Sicherheit kennt (LORD 1980:228). Die Zuweisung zufälliger Werte zu den nicht beantworteten Items ist somit auf Grund der angebenen Formel "unfair". LORD (1974) bietet eine Lösung dieses Problems über eine Änderung der Likelihoodfunktion. Weiterhin legte LORD 1983 eine Arbeit mit einem interessanten Modell des "item omitting" vor. Im Rahmen dieses Modells entwickelt er folgenden Maximum-Likelihood-Schätzer für die "Fähigkeit" des Probanden:

5-5 $$\hat{R} = \frac{1}{n} \cdot \left(V + \frac{W}{1-C} \right)$$

(n = Anzahl der Items

V = Anzahl der "omits"

W = Anzahl der falschen Items

C = 1/A).

Für die Einzelheiten muß auf LORD (1983:481) verwiesen werden. Tests dieser Art werden in Surveys allerdings nur selten benutzt. Für das in der Praxis häufiger vorkommende zuerst genannte Problem ist das Zuweisen des Wertes "nicht gelöst" einfach und ausreichend.

Likert-Skalen

Obwohl die Likert-Skalierung zu den am weitesten verbreiteten Skalierungsmethoden gehört (CRANO/BREWER 1975:232), finden sich in der Literatur kaum Hinweise auf die Handhabung fehlender Werte. Das in der Praxis übliche Verfahren scheint die Löschung der Fälle mit fehlenden Werten (LISTWISE) zu sein[99/].

Daneben finden sich Hinweise auf die Benutzung der von SONQUIST/ DUNKELBERG (1977) erwähnten Methoden (z.B. OMURA 1983:321). Sie beschreiben die Ersetzung durch den Mittelwert für die vorhandenen Items und das Ersetzen durch Ziehung aus einem Pseudo- Zufallszahlengenerator. Bei dieser

Methode wird die Wahrscheinlichkeit für das gegebene Antwortmuster berechnet und als Parameter an den Generator übergeben[100/] (SONQUIST/DUNKELBERG 1977: 270-271). Weiterhin erwähnen sie das Ersetzen durch eine Zufallszahl aus einem Generator derselben Verteilungsform wie der Verteilung der vorhandenen Werte. MEYERS (1973:332) schlägt ebenso die Verwendung von Zufallszahlen aus einem Normalverteilungsgenerator mit den Parametern des vollständigen Datensatzes vor. Unter der MAR-Annahme stellen diese Methoden lediglich die Einführung zusätzlicher unsystematischer Meßfehler dar, die zur Unterschätzung der Korrelationskoeffizienten führen.

Spezielle Indizes

Aufgrund ihrer relativ häufigen Verwendung in Surveys (insbesondere zur Indexkonstruktion) sollen hier Fragebatterien des Typs "vorhanden/nicht vorhanden" erwähnt werden, obwohl sie nicht zu den Skalen im engeren Sinne gehören. Obgleich nur bei schriftlichen Befragungen bei dieser Art Fragen MD vorkommen sollten (mit Ausnahme von Verweigerungen), ergeben sich in der Praxis häufig doch fehlende Werte. ERDOS/MORGAN (1970:205)[101/] schlagen für schriftliche Befragungen die Zuweisung des Wertes "nicht vorhanden" für Fragen dieses Typs vor, sofern andere Fragen auf derselben Fragebogenseite beantwortet wurden. Diese schwer zu rechtfertigende Verfahrensweise dürfte implizit vielen Kodierplänen auch in Surveys zugrunde liegen.

Imputation vs. Löschung

Zu der allgemeinen Problematik, ob und wann ein Fall aus den Berechnungen ausgeschlossen werden soll, wenn er auf einigen Skalen-Items fehlende Werte besitzt, existiert nur sehr wenig Literatur. Diese Entscheidung hängt natürlich von der Art der Skala, dem Untersuchungsziel und der Art des Ausfalls ab. Zum Problem des Ausfalls unter der MAR-Annahme hat PROCTOR(1974;1978) zwei Arbeiten vorgelegt. Nach einer Reihe von Annahmen (homogene

Interitemkorrelationsstruktur, Varianzhomogenität, Normalverteilung, Ersetzung fehlender Werte durch den Gesamtmittelwert der Stichprobe) entwickelt PROCTOR (1978:209) eine "optimale" Teststatistik, die er dann mit der aus der Mittelwertersetzung resultierenden Teststatistik vergleicht. Unter den genannten Voraussetzungen kommt er zu dem Ergebnis, daß ein Fall immer dann gelöscht werden sollte, falls mehr als 50% der Daten fehlen (PROCTOR 1978:210). Interessant an PROCTOR`s Arbeit ist vor allem der Versuch des Vergleichs der "power" der Handlungsalternativen. Ein wichtiges Ergebnis ist beispielsweise die Abhängigkeit der optimalen Löschungsentscheidung von der Reliabilität der Skala. Unter den bereits erwähnten Bedingungen steigt der Anteil der "tolerierbaren", d.h. der ohne "power"-Verlust in Kauf zu nehmenden und zu ersetzenden fehlenden Werte als nichtlineare Funktion der Reliabilität der Skala an (PROCTOR 1974:408).

Für sozialwissenschaftliche Datensätze sind solche Empfehlungen auf Grund der idealisierenden Annahmen weitgehend unbrauchbar. Andere vorgeschlagene Schwellenwerte, wie z.B. mehr als 1/3 fehlender Werte pro Fall (LORD; s.o.), sind rein willkürlich.

Neben der Alternative "Ersetzung vs. Löschung" gibt es noch die Möglichkeit des "Flagging". Hierbei wird der fehlende Wert durch den Mittelwert (oder eine Schätzung) ersetzt und eine zusätzliche Variable gebildet, die zur Kennzeichnung der ersetzten Werte benutzt wird ("flagging of imputed values"). Die Flagvariablen werden dann bei allen Analysen mit berücksichtigt (vgl. hierzu auch die im Kapitel 1 dargestellte Methode von COHEN/COHEN).

5.5 Exkurs V: Missing data auf nominalem Meßniveau

Die zu Beginn des Kapitels eingeführte Unterscheidung zwischen informativen und nicht-informativen Ersetzungen legt es nahe, diese Unterscheidung auch auf Ersetzungsmethoden für nominale Daten zu übertragen. Die Anwendung eines Redundanzkonzepts auf die Schätzung fehlender Daten auf nominalem Meßniveau führt über die PRE-Interpretation nahezu unmittelbar auf die Benutzung des

Bayes-Theorems. In der Literatur finden sich hierzu eine Reihe von Arbeiten, ohne daß in der Regel dieser Zusammenhang hergestellt wird.

Die bekannteste Arbeit in diesem Zusammenhang stammt von PRESS/ YANG (1974)[102/]. Sie entwickeln einen Bayes-Punktschätzer für die Anteilswerte einer Variablen bei MD unter Berücksichtigung zusätzlich vorhandener Informationen über das Antwortverhalten auf anderen Variablen. PRESS/YANG zielen insbesondere auf "weiß nicht"-Antworten als verdeckte Verweigerung, die sie von den "weiß nicht"-Antworten auf Faktenfragen unterscheiden (PRESS/YANG 1974:58). Ausgehend von einer unveröffentlichten Arbeit von GEISSER gelangen sie zu folgendem Bayes-Punktschätzer:

$$5\text{-}3 \qquad \hat{q}_i = \frac{(n_i + a_i) + \sum_{j=1}^{m} P \cdot (PI_i | z_j)}{m + \sum_{j=1}^{M} (n_j + a_j)}$$

(PRESS/YANG 1974:60; Gleichung 2.8)

wobei gilt: M = Anzahl der Kategorien auf der Variablen mit fehlenden Werten, n_i = Anzahl der klassifizierten Fälle in Kategorie i der interessierenden Variablen, m = Anzahl der Fälle mit fehlenden Werten, z_j = Antwortvektor des Befragten j auf den zusätzlichen Variablen, a_j = Hyperparameter (die im Falle nicht- informativer Prioriverteilungen gleich "1" gesetzt werden) und P(PI z_j) = "marginal predictive probability" für die Klassifizierung des Befragten j mit dem Antwortmuster z_j.

PRESS/YANG betonen, daß für die Punktschätzung eine tatsächliche Klassifizierung nicht notwendig ist, und verweisen für dieses Problem auf GEISSER (1966). Der Schätzer von PRESS/YANG kann über seine Konstruktion im Falle von "vague priors" lediglich minimale Differenzen zur "naiven" Schätzung über

das Bayes-Theorem erbringen. Die Bewertung dieses Schätzers hängt also von den zu spezifizierenden Hyperparametern ab.

PRESS/YANG diskutieren die Frage nach der Auswahl der zusätzlichen Variablen im Rahmen einer kurzen Darstellung der Berechnung der "discriminatory power" möglicher zusätzlicher Variablen und erörtern insbesondere das Problem der "Nullwahrscheinlichkeit" bei zu hoher Variablenzahl bzw. Kategorienzahl (PRESS/YANG 1974:65). Obwohl explizit in Hinsicht auf sozialwissenschaftliche Anwendungen argumentiert wird, findet sich bis auf den relativierenden Satz, daß nicht an eine Anwendung auf "wirkliche" "weiß nicht"-Antworten gedacht ist, keine Diskussion der Anwendungsbedingungen[103/], insbesondere die dem Ansatz implizite MAR-Annahme wird nicht angesprochen.

Eine vergleichsweise einfache Imputationtechnik wurde von POHL/ BRUNO (1978) vorgestellt. POHL/BRUNO berechnen für eine Reihe von "ähnlichen" dichotomen Variablen auf Grund der vollständigen Fälle die Wahrscheinlichkeiten für das Auftreten jedes möglichen Antwortmusters. Für jede unvollständige Variable werden diese Wahrscheinlichkeiten mit dem Bayes-Theorem in a-posteriori-Wahrscheinlichkeiten umgerechnet. Den Fällen mit fehlenden Werten wird dann der Wert mit der höchsten bedingten Wahrscheinlichkeit zugewiesen.

Die Folgen der Ersetzung bleiben bei POHL/BRUNO undiskutiert. Die von den Autoren vorgeschlagene Ersetzungsregel bedingt für die Punktschätzung eine Überschätzung des Anteilswertes des Musters mit der höchsten bedingten Wahrscheinlichkeit, dies ließe sich, wie bei SONQUIST/DUNKELBERG, durch eine Ziehung aus einem Generator mit der berechneten a-posteriori-Wahrscheinlichkeit als Parameter vermeiden.

DANIEL u.a. (1982) demonstrieren eine Anwendung des Bayestheorems auf Unit-Nonresponse. Sie erwähnen die Möglichkeit der Anwendung auf

Itemnonresponse, erörtern das Problem der Auswahl der Prädiktoren und sehr kurz die MAR-Annahme. Für das von ihnen bearbeitete Problem benutzen sie dieselbe Imputationregel wie POHL/ BRUNO, für Itemnonresponse ist damit ihre Methode[104/] mit dem Vorgehen von POHL/BRUNO identisch.

Die hier besprochenen Arbeiten stellen lediglich Beispiele für die Verwendung des Bayestheorems bei der Schätzung fehlender Werte auf nominalen Meßniveau dar, andere Techniken müssen hier unerwähnt bleiben[105/].

Zusammenfassend soll festgehalten werden, daß die bisher existierenden Ansätze auch auf nominalem Meßniveau weitgehend auf der MAR- bzw. MARC-Annahme basieren und die Kritik dieser Ausfallmodelle somit auch diese Techniken trifft. Die Annahme, daß die Heranziehung zusätzlicher Information eine Verbesserung der Schätzung ermöglicht, impliziert zumindest die Ablehnung der "traditionellen Lösung"[106/] des Problems (Ignorieren der missing values), d.h. die Ablehnung der Gleichverteilungsannahme. Die Verbesserung hängt natürlich davon ab, daß, wie weiter oben ausgeführt, die implizit vorausgesetzte Homogenität innerhalb der Imputationklasse gegeben ist.

5.6 Zusammenfassende Bewertung

Die Darstellung der wichtigsten Kompensationstechniken und Ersetzungsmethoden und ihrer impliziten Annahmen sowie der generellen Problematik der Ersetzungsmethoden sollte sowohl die Spannweite der Lösungsvorschläge wie der Probleme der Techniken illustrieren. Obwohl die meisten Autoren die Begrenzung der Techniken in Hinsicht auf die MAR-Annahme betonen und weithin die Gültigkeit der MAR-Annahme für sozialwissenschaftliche Surveys bezweifeln, wurden kaum Ansätze entwickelt, die auf systematische Ausfallmechanismen Bezug nehmen. Die Implementierung "automatischer" Ersetzungstechniken wie in BDPAM oder die unreflektierte Verwendung einer "Maximum Likelihood"-Schätzung kann zu einem zu sorglosen Umgang mit systematischen Ausfallmechanismen verleiten.

Zwei Hauptprobleme ergeben sich aus der bisherigen Praxis: zum einen die Gefahr der ungerechtfertigten Verallgemeinerung der Ergebnisse der Leistungsfähigkeitsuntersuchungen der MD-Verfahren unter der MAR-Annahme auf andere Ausfallmechanismen, zum anderen die Abdrängung des MD-Problems zu einem rein technischen Problem.

Der Untersuchung der Verallgemeinerbarkeit der Ergebnisse in Hin- sicht auf die Leistungsfähigkeit der Verfahren sind die Kapitel 6-8 gewidmet; die Untersuchung der Möglichkeiten zur Abschätzung potentieller Verzerrungen ist Gegenstand des Kapitels 9.

Generell bleibt festzuhalten, daß nahezu alle in diesem Kapitel vorgestellten Techniken explizit oder implizit auf der MAR- bzw. MARC-Annahme basieren, ohne daß dies in der Forschungspraxis ausreichend berücksichtigt wird.[107/]. Keine "automatisch" arbeitende MD-Kompensationstechnik kann die Gefahr systematischer Verzerrungen ausschalten. Auf Grund der dargelegten theoretischen Einwände gegen das MARC-Modell und der früher dargelegten - überwiegend empririsch begründeten - Zweifel an der MAR-Annahme wird hier die Legitimation der Ersetzungstechniken lediglich in der technischen Unterstützung bei der Beurteilung verschiedener theoretischer Annahmen über den Ausfallprozeß und möglicher Auswirkungen auf die Datenanalyse gesehen. Die Ersetzungstechniken werden somit als derzeit einzig gangbarer Weg zu theoretisch begründeten Sensitivitätsuntersuchungen (vgl. dazu Kapitel 9) angesehen.

Anmerkungen zu Kapitel 5

1. Einen Überblick über Gewichtungsverfahren geben OH/SCHEUREN (1983).

2. FEDERSPIEL (1959) (zitiert nach HAMILTON 1975:21) und GLASSER (1964:836) berechneten aus den paarweise vorhandenen Werten die Kovarianzen. GLASSER (1964) schlug einen Schätzer für Regressionen mit fehlenden Werten auf den unabhängigen Variablen vor, die explizite Angabe eines Korrelationskoeffizienten findet sich nicht bei GLASSER. Die naheliegende Anwendung der Kovarianzformel auf die Diagonalelemente der Matrix führt zu COVPAIR. Die Methode wird in der Literatur meist GLASSER (1964) zugerechnet.

3. Bei ENGELMAN (1982) findet sich ein sehr effizientes FORTRAN- Programm zur Berechnung aller vier Matrizen.

4. z.B. STEWART (1982:397) bezeichnet ALLVALUE als PAIRWISE, wobei diese von ihm GLASSER zugeschriebene Methode alle Daten verwendet. GLASSER verwandte aber bedingte Mittelwerte bei der Berechnung der Kovarianzen.

5. z.B. BOAS (1967). DEMPSTER (1971:343) erwähnt die Möglichkeit, nur diejenigen Fälle aus der Analyse auszuschließen, die auf bestimmten Variablen MD besitzen, und spezielle Techniken auf die verbleibenden MD anzuwenden.

6. Die von WILKS entwickelten ML-Schätzer bleiben zumeist unerwähnt.

7. Zu den Auswirkungen der paarweisen Berechnung der Kovarianzmatrix auf die Eigenwertstruktur vgl. HEIBERGER (1977:46). FRANE (1978:30) weist auf Rundungsfehler bei nahezu singulären Matrizen hin.

8. Die Studie von SCHWERTMAN/ALLEN (1979) basiert auf einem "realen" Datensatz, dessen Kovarianzstruktur als Startwert für Zufallszahlengeneratoren benutzt wurde. Die Ergebnisse gelten streng genommen nur für Stichproben mit vergleichbarer Kovarianzstruktur.

9. Vgl. z.B. FREUND/MINTON (1979:138-148), kritisch zur Ridge- Regression: ROZEBOOM (1979).

10. Vgl. KENNDEY/GENTLE (1980:540-543).

11. Eine formale Beschreibung des Verfahrens findet sich, wie auch eine kleine Simulationsstudie zum Verhalten des Algorithmus bei SCHWERTMAN/ALLEN (1979); vgl. auch das FORTRAN-Programm bei HUSEBY/SCHWERTMAN/ALLEN (1980:304-308).

12. Der EM-Algorithmus liefert auch MD-Ersetzungen.

13. Allgemein zur ML-Schätzung unter MD vgl. ORCHARD/WOODBURY (1972:699-702), DEMPSTER/LAIRD/RUBIN (1977:passim), HINKINS(1979: 16-39) und LITTLE (1983:348-351).

14. Ausführlichere Darstellung finden sich bei AFIFI/ELASHOFF (1966:601-603), SHIH (1981:2.1-2.16) und ANDERSON/BASILEVSKY/HUM (1983:436-439). Die Bedeutung der Arbeit von ANDERSON (1957) liegt in der Partitionierung der Likelihood-Funktion, Verallgemeinerungen auf andere Fälle gibt RUBIN (1974), vgl. auch LITTLE (1983:356). Darstellungen der ML-Schätzungen für diese Spezialprobleme geben u.a. HAITOVSKY (1965:2-5), TIMM (1969:2-7), LYNCH (1972:10-12) und MORRISON (1976:120-123).

15. TIMM (1969:6) führt als Kritikpunkte vor allem an, daß keine MD-Schätzung erfolgt und alle Verfahren in besonderem Maße die Normalverteilungsannahme voraussetzen.

16. Falls das Muster der MD nur annähernd monoton ist, legen MARINI/OLSON/RUBIN (1980:322-323) die Löschung derjenigen Beobachtungen (nicht Fälle) nahe, die außerhalb des monotonen Musters liegen. Insofern keine systematischen Differenzen zwischen den Fällen mit und den Fällen ohne monotonem Muster bestehen, erfolgt durch die Löschung (weniger) Fälle außerhalb des Musters nur eine leichte Vergrößerung der Mutungsintervalle. Als weitere Möglichkeit, die Daten in das geforderte Muster zu bringen, erwähnen sie Ersetzungstechniken.

17. MARINI/OLSON/RUBIN (1980:316-317) unterscheiden zwischen "missing at random" und "missing completely at random". Letzteres wird gleichbedeutend mit RUBIN (1976) als "missing at random" und "observed at random" gebraucht (vgl. Kap. 1). Sie bezweifeln die Gültigkeit der "missing completely at random"-Annahme für Panels, argumentieren aber, daß durch die Kontrolle einer Reihe von Variablen die Konstanz der Variablenbeziehungen eine vernünftige Annahme darstellt. Unter "missing at random" verstehen sie, daß sich Respondenten und Nonrespondenten mit den selben Werten der beobachteten Variablen auf den nicht beobachteten Variablen nicht unterscheiden.

18. Eine exaktere Darstellung würde den Rahmen dieser Arbeit sprengen. Eine ausführlichere Erörterung gibt LITTLE (1983:358- 364), die allerdings nicht ausreicht, um die Berechnungen durchzuführen. Hierzu muß auf MARINI/OLSON/RUBIN (1980:330-333) zurückgegangen werden.

19. Der "Sweep"-Operator stellt lediglich eine Reihe modifizierter elementarer Matrixoperationen dar, vgl. ausführlich GOODNIGHT (1979) sowie KENNEDY/GENTLE (1980:288-294).

20. Eine kritische Darstellung von MD-Techniken unter MAR im Zusammenhang mit Panels geben GRILLICHES/HALL/HAUSMAN(1978). Ihre Arbeit enthält einen Vergleich mehrerer Schätztechniken bei systematischen Ausfällen in einem Panel.

21. Eine leicht verständliche Darstellung des MIP gibt LAIRD (1985).

22. Vgl. LITTLE/RUBIN (1983:53), zum Konvergenzbeweis vgl. dagegen BOYLES (1983). Auf die vielfältigen Anwendungsmöglichkeiten und die Beziehung des EM-Algorithmus zu anderen Verfahren kann hier nicht eingegangen werden, vgl. z.B. LITTLE (1983:368-369).

23. Die erstaunlich niedrige Zahl der in der Literatur nachweisbaren Programme für das hier interessierende Problem - ESTMAT, MISSMLE, SCICON, BMDPAM (und nach HASSELBLAD/CREASON/STEAD 1983: 199: GENSTAT) - zeigt, daß die Umsetzung in einen programmierbaren Algorithmus keineswegs eine triviale Aufgabe darstellte. Lediglich BMDPAM und GENSTAT sind vergleichsweise allgemein verfügbar. Wenn man von unveröffentlichten Research Reports (wie z.B. HINKINS/HAMILTON 1976) absieht, ist das Flußdiagramm erst seit HASSELBLAD/CREASON/STEAD (1983:200) allgemein zugänglich: Dreizehn Jahre nach dem nur schwer zugänglichen ersten Flußdiagramm bei (WOODBURY/HASSELBLAD 1970). Das Diagramm von HINKINS/HAMILTON (1976:13) ist für eine Programmierung wesentlich geeigneter, die Darstellung der Berechnungen ist die bisher detaillierteste.

24. z.B. HINKINS/HAMILTON (1976:7) verwenden die absolute größte relative Veränderung eines beliebigen Elementes sowohl des Mittelwertvektors als auch der Kovarianzmatrix zwischen zwei vollständigen Iterationen. Abbruchkriterien sind mehr als 100 Iterationen oder ein Konvergenzkriterium kleiner als .001.

25. Zur partiellen Kovarianzmatrix allgemein vgl. SCHLITTGEN/ STREITBERG (1984:446-447).

26. KALTON (1983:65) führt als weiteres Argument die höheren Rechenzeiten komplexer Parameterschätzverfahren bei unvollständigen Datensätzen an.

27. Die meisten in Lehrbüchern veröffentlichten Programme, die evtl. für spezielle Auswertungen herangezogen werden könnten, berücksichtigen fehlende Werte nicht, z.B. COOLLEY/LOHNES (1971) oder OVERALL/KLETT (1972).

28. Im Zusammenhang mit Problemen der Dateneditierung ergibt sich ein weiteres Argument: Falls Datenerhebung und Datenanalyse personell getrennt erfolgt, kann für eine sinnvolle Ersetzung u.U. spezielles Wissen über die Umstände der Datenerhebung erforderlich sein, so daß die Ersetzung von den "data collectors" zu erfolgen hat. FORD bemerkt(1983:189): "Most data analysts do not want the reponsibility of adjusting for missing data. Their interests pertain to mainstream statistics which emphasize estimation and analysis given the ability to obtain an entire data set."

29. Obwohl hier in Hinsicht auf Itemnonresponse argumentiert wird, wird als vereinfachende Sprachregelung die Begrifflichkeit "Respondenten" und "Nonrespondenten" benutzt.

30. Vgl. hierzu insbesondere die Arbeit von ROCKWELL (1975) sowie die Diskussion in KREWSKI/PLATEK/RAO (1981:481-497).

31. Ein Katalog derjenigen Sachverhalte, die im Imputationprozeß der Dokumentation bedürfen ("auditing") findet sich bei I.G. SANDE (1983:346-347).

32. Zumeist richtet sich die Kritik gegen die Ersetzungsverfahren vor allem gegen Imputationverfahren bei Zensuserhebungen. z.B. macht BANISTER (1980:18) darauf aufmerksam, daß der U.S. Census 1970 von 4.9 Millionen (!) "Personen" "Daten" enthält, die durch Ersetzungsmethoden gewonnen wurden. In bestimmten Subgruppen erreicht der Anteil geschätzter Daten für einige Fragen bis zu 20% aller vorhandenen Daten (BANISTER 1980:16).

33. Einen solchen Mißbrauch von MD-Ersetzungsverfahren stellt z.B. die Arbeit von BRONNER (1983) dar. BRONNER wendet eine Art Hot-Deck mit einer Distanzfunktion auf vollständig fehlende Teile von Fragebögen an. BRONNER glaubt, die Kosten der Marktforschung senken zu können, indem er Variablen von Personen anderer Untersuchungen "ähnlichen" Personen einer aktuellen Untersuchung zuspielt.

34. STEINHAUSEN/LANGER (1977:177) legen für Einstellungsfragen in der empirischen Sozialforschung die Antwort "teils/teils" nahe.

35. z.B. bei BERGER (1979).

36. Zitiert nach HAMILTON (1975:20) und AFIFI/ELASHOFF (1966: 598).

37. Zu den Differenzen, Übergängen und Anwendungsschwerpunkten von MD-Kompensationsverfahren unter idealen Annahmen existiert nunmehr eine umfangreiche Literatur. Stellvertretend kann auf OH/ SCHEUREN (1983), KALTON/SANTOS (1983) und PLATEK/GRAY (1983) verwiesen werden.

38. Im hier interessierenden Zusammenhang sind sowohl die Unterschiede in bezug auf Unit- und Itemnonreponse als auch zwischen "Hotdeck-matching variables", "weighting class"-Variablen und "imputation control variables" unwichtig.

39. Analytische Resultate zur Beurteilung des Gewichtungs- bzw. Mittelwertersetzungsverfahrens unter der Annahme, daß die Zellen den Ausfallmechanismus exakt modellieren, liegen für univariate und teilweise für multivariate Statistiken vor (vgl. OH/SCHEUREN 1983 und SANTOS 1981). Vgl. in diesem Zusammenhang auch THOMSEN (1973, 1975).

40. Zum Konzept der Responsewahrscheinlichkeiten vgl. PLATEK/ SINGH/TREMBLAY (1977:7-16), für einen analytischen Vergleich verschiedener Imputationverfahren bei durch Drittvariablen bedingten Responsewahrscheinlichkeiten vgl. ANDERSON (1979).

41. Vgl. ebenso HERTEL (1976:465): Die Klassifizierungsvariablen sollten untereinander unkorreliert sein und eine hohe Korrelation zu der zu vervollständigenden Variablen aufweisen. HERTEL argumentiert (in Anlehnung an KISH 1953:192-193), daß die Erhöhung der Zahl der Klassifikationsvariablen wenig Verbesserung bringt und eine hohe Zahl leerer Zellen nach sich zieht, und hält dementsprechend 2-3 Klassifikationsvariablen für ausreichend.

42. STEINHAUSEN/LANGER (1977:177) erwähnen die Ersetzung durch den "inhaltlich" "wahrscheinlichsten" Wert, durch die Erwartungswerte (Mittelwert der Variablen bzw. Cluster), durch Regressionsschätzung und die Änderung der Distanzfunktion durch paarweise vorhandene Werte. Ein FORTRAN-Programm mit der Ersetzung fehlender Werte durch Clustermittelwerte findet sich bei HARTIGAN (1975:72).

43. WISHART erwähnt nicht die theoretischen Probleme bei der Interpretation der reduzierten Distanzfunktion: "Ähnlichkeit" im Rahmen der gegeben Daten kann bei systematischem Ausfall zu erheblichen Verzerrungen führen.

44. Zur AID-Technik allgemein vgl. ausführlich SONQUIST/BAKER/ MORGAN (1971), kurze Übersichten bei DILLON/GOLDSTEIN (1984: 183-186) und BIERVERT/DIERKES/WALZEL (1969), eine kritische Darstellung gibt FIELDING (1979). Entsprechende FORTRAN-Programme finden sich bei HARTIGAN (1975).

45. Neben den unterschiedlichen abhängigen Variablen bildet die konkrete Zielsetzung einen weiteren Unterschied. KALTON/SANTOS suchten Imputationklassen, CHAPMAN "weighting classes"; im hier interessierenden Zusammenhang ist dieser Unterschied unbedeutend.

46. Die im Sinne der Relcovs effektivsten Variablen in dem von CHAPMAN analysierten Survey waren Alter und Geschlecht (Tabellen 7-10, S.451-454). CHAPMAN empfiehlt darüber hinaus, die ethnische Zugehörigkeit und die geographische Region miteinzubeziehen. Alle genannten Variablen sind soziologische Globalvariablen, die der Explikation bedürfen.

47. Die Annahme konstanter Responsewahrscheinlichkeiten stellt eine Verschärfung dar: Angenommen wird eine einzige Superpopulation, aus der zufällig die endlichen Populationen der Respondenten und Nonrepondenten gezogen werden. Bei wiederholten Ziehungen antworten Respondenten immer, Nonrepondenten nie. Die Annahme homogener Responsewahrscheinlichkeiten läßt wechselndes Verhalten ("at random") zu, vgl. KALTON/SANTOS (1983:15).

48. Die Nähe dieser Bemerkung von DALENIUS zur handlungstheoretischen Interpretation des Befragtenverhaltens ist offensichtlich, vgl. hierzu Kapitel 2.

49. Dies stellt nur ein Beispiel der zumeist von Statistikern stark vernachlässigten "nonsampling errors" dar. Vgl. in diesem Zusammenhang die Diskussionsbeiträge im Anschluß an ein Thesenpapier von DALENIUS (1981) in KREWSKI/PLATEK/RAO (1981:77-84).

50. Die letzte Konsequenz dieser Annahme wird bei Hot-Deck- Verfahren mit "nearest neighbor rule" gezogen: Die fehlenden Daten eines Nonrespondenten werden durch die Daten eines Respondenten ersetzt, der in bezug auf die verfügbaren Daten die größte Ähnlichkeit besitzt. Im Extremfall besitzen alle "matching variables" den selben Wert; die Responsereaktion kann nicht erklärt werden.

51. Vgl. DALENIUS (1983:412):"I am at a complete loss to understand how valid estimates can be computed on the basis of the data collected in the course of a real-life survey...". Unter idealen Bedingungen ließe sich die individuelle Antwortwahrscheinlichkeit in einem Panel bestimmen; das für Panels typische Ausfallmuster ist aber ein "monotones" Muster.

52. So findet sich bei ENGEL(o.J.:22) eine von der Problematik der Resonsewahrscheinlichkeiten völlig losgelöste Bestimmung der Imputationklassen für ein HOT-DECK-Verfahren: Benutzt wird Alter/ Geschlecht bzw. "Humankapitalbestand" (Schuljahre). Zwar erfolgt eine Darstellung des Zusammenhangs zwischen Verweigerungen und Interviewvariablen und "Hintergrundmerkmalen", aber ein Zusammenhang zwischen diesen Darstellungen und der Bestimmung der Imputationklassen wird nicht erkennbar. Der Versuch einer Modellierung des Ausfallprozesses wie z.B. bei GREENLESS/REECE/ZIESCHANG (1982) unterbleibt.

53. Zu Erklärungen mit impliziten Gesetzen vgl. OPP (1976:152- 157).

54. Zur Kritik vgl. H.ESSER (1981), E.ESSER (1982:178f).

55. Vgl. LANGENHEDER(1975:61): "... alle diese Variablen stellen typische Kombinationen bestimmter Einflußfaktoren dar und weisen deshalb oft positive oder negative Korrelationen mit bestimmten anderen Variablen auf."

56. Die von ESSER (1979:16) in Hinsicht auf die "kollektivistischen" Annahmen der empirischen Sozialforschung formulierte Kritik, daß "... die bloße Zugehörigkeit von Akteuren zu kategorialen Kollektiven (...) zu den theoretisch bedeutsamen Variablen zählt und über solche Zugehörigkeitsmerkmale irgendwelche abhängigen Variablen 'kausal' erklärt werden könnten", trifft ebenso die implizit gelassenen soziologischen Grundlagen des Imputation-klassenkonzept.

57. Ratioschätzung kann als Spezialfall einer Regressionsschätzung mit nur einer Prädiktorvariablen und keiner Konstanten aufgefaßt werden (KALTON 1983:80).

58. Eine Beschreibung des Verfahrens und möglicher Modifikationen findet sich bei FORD/KLEWENO/TORTORA (1980:252).

59. In der amtlichen Statistik ist u.a. ein weiteres Verfahren üblich: Doppeln (ZINDLER 1967:211-217). Beim Doppeln werden alle Einheiten mit Datenfehlern von der Aufbereitung ausgeschlossen und jeder ausgeschlossene Fall durch einen zufällig ausgewählten vollständigen und korrekten Fall ersetzt, d.h. der entsprechende ausgewählte Fall wird gedoppelt. Eine Aufspaltung in Imputationklassen ist möglich, wenn auch damit nicht immer ein Präzisionsgewinn verbunden ist (ZINDLER 1967:217).

60."control or matching variables" bei KALTON (1983:67)

61. Die Literatur zu diesen Imputationtechniken bedient sich weitgehend einer eigenen Terminologie, die wichtigsten Begriffe werden im folgenden eingeführt.

62. In Anlehnung an FORD (1983:186-187) wird für "qualitative" Variablen, die die Imputationklassen bilden, der Ausdruck "Klassifikationsvariablen" benutzt, während die Variablen, die bei der Berechnung der Distanzfunktionen Verwendung finden, als "Match-variablen" bezeichnet werden. Die Benennung "simultane Verfahren" findet sich nicht in der Literatur. Sequentielle Verfahren sind "one pass"-Methoden, d.h. der Datensatz wird nur einmal gelesen und geschrieben, simultane Verfahren setzen dagegen in irgendeiner Form immer einen "random access" zu den Daten voraus (SANDE 1979:245 bezeichnet simultane Verfahren als "random choice"- Methoden).

63. COX (1980:724) gibt 4 Datenzugriffe an, geht allerdings von einem Sortiergang aus, der bei anderer Implementierung nicht erfolgen muß.

64. Vgl. u.a. FRANKEL/BANKS (1979:103), KALTON/KASPRZYK (1982: 23), SANDE (1982:149), FORD (1983:196).

65. Eine verbreitete Anwendung finden numerische Hot-Decks lediglich bei Einkommensdaten, so z.B. SPIERS/KNOTT (1969), ONO/MILLER (1969) und SCHIEBER (1978).

66. In diesem Fall (metrische Daten, Distanzfunktion, sequentielle Bearbeitung) entspricht die Zuordnung eines Falles dem Leader- Algorithmus der Clusteranalyse.

67. Bei FORD (1976:326) findet sich der Vorschlag, den Mittelwert der beiden "nächsten" Fälle zu verwenden: "Two-Closest"-Methode.

68. SANDE (1979:247) betont die Notwendigkeit der Transformation der Variablen bei schiefen Verteilungen.

69. "The main reason for limiting donor usage may be the pacification of nervous clients" (SANDE 1979:250), vgl. u.a. auch FORD (1983:197).

70. Zur Mahalanobis-Distanz allgemein und insbesondere zur Kritik vgl. STEINHAUSEN/LANGER (1977:59-61).

71. Die Darstellung bei FORD/KLEWENO/TORTORA (1980) beschreibt den Spezialfall mit zwei unvollständigen Variablen.

72. Für die Details der Methode muß auf PREGIBON (1976) verwiesen werden. Dort findet sich sowohl eine sehr elegante Formalisierung des Problems und der Zero-Spike-Prozedur als auch eine interessante Erörterung des Problems der Festlegung der Wahrscheinlichkeiten für den vermuteten "non-random blank injection mechanism" (PREGIBON 1976:71).

73. M.a.W.: Jede in der Population existierende Kombination, die im unvollständigen Sampleteil vorkommt, muß im vollständigen Sampleteil existieren, vgl. PREGIBON (1976:73).

74. Vgl. hierzu auch Exkurs V.

75. Für den Fall, daß "subjektive" Informationen für die Spezifizierung der gemeinsamen Verteilungsfunktion vorliegen, gibt PREGIBON (1976:75-78) Hinweise für eine Konstruktion der Verteilung, für die die Spezifizierung von einem Parametervektor der Länge k ausreicht.

76. Wichtig festzuhalten bleibt der stochastische Charakter der Zuweisung; die Zuweisung zu dem Domain mit der jeweils größten bedingten Wahrscheinlichkeit würde einen Bias bedingen (PREGIBON 1976:74).

77. PREGIBON (1976:74) schlägt explizit eine Regressionsschätzung vor. FORD/KLEWENO/TORTORA (1980:253) zitieren in diesem Zusammenhang PREGIBON falsch, indem sie ihm bei der Schätzung Beliebigkeit der Schätzmethode unterstellen. Sie benutzen in ihrer Version der zero-spike-Prozedur eine Ratioschätzung.

78. Die Darstellung bei CHAPNEY/BELL (1982:432) ist irreführend : "The predicted values are then used to define the distance function". Die Verfahren von SCHIEBER (1978:214) benutzen keine Distanzfunktionen.

79. Die Originalarbeit von DEAR (1959) war leider nicht zugänglich, so daß die Details der ursprünglichen Methode aus der Sekundärliteratur entnommen werden mußten.

80. Es ist unklar, welche Korrelationsmatrix DEAR benutzt. Diese Beschreibung (MEAN-Korrelationsmatrix als Ausgangsmatrix) basiert auf HAMILTON (1975:27). In der Anwendung der Methode benutzt HAMILTON (1975b:11) die LISTWISE-Matrix, ebenso TIMM in seiner Simulation (1969:24). GLEASON/STAELIN (1975: 236-237;241-242;244) wenden sich u.a. auf Grund der Ergebnisse von TIMM gegen die generelle Verwendung der LISTWISE-Matrix, es bleibt aber undeutlich, welche Matrix sie als Ausgangspunkt in ihrer Simulation verwenden.

81. Der vollständige Formalismus findet sich bei TIMM (1969:22- 25), TIMM (1970:419-421), GLEASON/STAELIN (1975:232-234) und bei ANDERSON/BASILEVSKY/HUM (1983:442-444).

82. Zu den Modifikationen vgl. HAMILTON (1975:27-28), ANDERSON/ BASILEVSKY/HUM (1983:444).

83. Eine andere Möglichkeit der Bestimmung der Hauptkomponenten bei fehlenden Werten in der Datenmatrix findet sich bei WIBERG (1976). Dieses Verfahren scheint noch nicht für die DEAR-Methode benutzt worden sein.

84. Der Beweis findet sich bei GLEASON/STAELIN (1975:235-236). Falls die ersten Eigenwerte der Anfangsmatrix relativ groß gegenüber den restlichen Eigenwerten sind, ist die zu erwartende Differenz sehr gering.

85. FRANE (1976:412) schreibt der Methode nach DEAR in der von GLEASON/STAELIN vorgeschlagenen Form einen höheren Rechenzeitaufwand als allen von ihm erörterten Regressionsmethoden, einschließlich der Stepwise-Regression, zu. Diese Äußerung ist angesichts der Tatsache, daß die Methode keiner wiederholten Partitionierung der Variablen bedarf und weiterhin keine vollständige Hauptkomponentenanalyse erforderlich ist, unverständlich. Wird die erste Hauptkomponente über die Methode nach Hotelling (vgl. PAWLIK 1971:67-80) berechnet, so stabilisieren sich die Schätzungen in der Regel nach nur wenigen Iterationen. Obwohl diese Methode als numerisch instabiler gilt als die "singular-value"- Dekomposition (vgl. KENNDEDY/GENTLE 1980:566-567), läßt sich ihre Verwendung gerade in diesem Zusammenhang rechtfertigen. Der Rechenzeitbedarf liegt bei dieser Implementierung weit unter dem der Methode von BUCK.

86. Eine andere Darstellung der Korrektur findet sich bei PREGIBON (1976:86), der sich auf ORCHARD/WOODBURY (1972:705) bezieht.

87. Weiterführende Überlegungen zur Korrektur im Anschluß an ORCHARD/WOODBURY (1972) finden sich bei BEALE/LITTLE (1975:134- 137) und CONNIFFE (1983), eine einfache Darstellung des Problems bieten KIM/CURRY (1977:231-232).

88. Ohne Korrektur wird das Verfahren z.B. bei GLEASON/STAELIN (1975:236), HAMILTON (1975:23), TIMM(1970:421-422) dargestellt.

89. Eine Modifikation findet sich bei AFIFI/ELASHOF (1969:350- 351).

90. WALSH (1961:205-208) benötigt für eine unvollständige Darstellung des Verfahrens mehr als 3 Seiten, daher wird auf eine kommentierte Wiedergabe verzichtet.

91. Die Originalarbeit war leider nicht zugänglich, zitiert wird nach JACKSON (1968:837-839) und ANDERSON/BASILEVSKY/HUM (1983: 458).

92. GLEASON/STAELIN (1975:244) begründen ihre Entscheidung mit einem Rechenzeitargument. KOOPMAN (1976) zeigte eine einfache Transformation der zu berechnenden Regressionsgleichungen, die eine erhebliche Rechenzeitersparnis erbringt, so daß dieses Argument an Geltung verliert.

93. FRANE hat vier Möglichkeiten der Regressionsschätzung in BMDP im Programm PAM implementiert. Neben der einfachen Regressionsschätzung mit der höchstkorrelierenden vorhandenen Variablen (SIMPLE) existieren die Optionen TWO-STEP, STEP und REGR. TWO-STEP wählt zusätzlich zur höchstkorrelierenden Variablen die Variable mit der höchsten Partialkorrelation aus, für die ein Wert vorhanden ist und ein "F-to-enter"-Kriterium nicht unterschritten wird. STEP arbeitet ausgehend von der höchstkorrelierenden Variablen die Variablen in der Reihenfolge ihrer "Partialkorrelationen" ab, bis das F-to-enter-Kriterium unterschritten wird. Die Vorhersage erfolgt dann mit allen in der Gleichung befindlichen Variablen. REGR stellt eine multiple Regression mit allen verfügbaren Variablen dar.

94. CHAN/DUNN (1972:474) bezeichnen ihr Verfahren als Modifikation des Verfahrens von BUCK. Es ist unklar, ob sie die Korrektur benutzen.

95. Das Verfahren wurde von SCHEUREN vorgeschlagen, vgl. SCHIEBER (1978:218).

96. Diese Stelle bei HAITOVSKY stellt eine der wenigen frühen Erwähnungen von Korrekturverfahren für nicht zufällig fehlende Werte dar. Um so erstaunlicher ist die fast völlige Ignorierung des Vorschlags in der sozialwissenschaftlichen Literatur.

97. Das trifft auf die Vorschläge von FRANE ebenso zu wie auf die nach den Arbeiten von COHEN und RUMMEL (vgl. Kap. 1.3.2) naheliegende, wenn auch von den Autoren nicht explizit erwähnte Verwendung der MD-Indikatorvariablen zur Schätzung. Dies scheint erst in der Folge der in Kap. 4 dargestellten Arbeiten zu Hazardinstrumenten erwartbar zu sein.

98. LORD (1974:249) zitiert empirische Arbeiten, die eine höhere Rate fehlender Werte berichten, als zu erwarten war. Bei mehr als einem Drittel fehlender Werte rät LORD (1974:248) zum Löschen des Falls.

99. Das entspricht beispielsweise der Voreinstellung der SPSS- Prozedur RELIABILITY.

100. Die Nähe zu den Cold-deck-Verfahren ist ebenso offensichtlich wie der Übergang zu den direkt auf dem Bayes-Theorem basierenden Ersetzungstechniken, vgl. hierzu Exkurs V.

101. Zitiert nach HUGHES (1978:594).

102. Die Arbeit verwendet eine bemerkenswert schwierige Notation; POHL/BRUNO zitieren die Arbeit von PRESS/YANG lediglich um das Wort "undecided" für nichtsubstantielle Antworten in der Literatur zu belegen, eine Anwendung des Verfahrens in der sozialwissenschaftlichen Literatur ist nicht nachzuweisen.

103. Zur Kritik vgl. vor allem NORDHEIM (1978:44; 52-54).

104. Die Anwendung des Bayestheorems für Probleme dieser Art ist naheliegend, wie die Darstellung der Arbeiten von SONQUIST/ DUNKELBERG (1977) und POHL/BRUNO (1978) zeigte. DANIEL u.a. (1982:59) scheinen die erwähnten Arbeiten entgangen zu sein: Bei ihnen handelt es sich um eine "new method of adjusting for nonresponse"; ihre Literaturhinweise beziehen sich auf Arbeiten zum Bayes-Theorem in der computergestützten Diagnose. Vgl. in diesem Zusammenhang die Diskussion von Bayes-Verfahren zur Klassifikation sowie der Nutzung der Diskriminanzanalyse bei OVERALL/KLETT (1972:329-412).

105. Mit dem hier behandelten Problem sind Versuche zur Parameterschätzung bei dichotomen Daten natürlich eng verwandt. Die Behandlung der verschiedenen Schätzer würde den Rahmen dieser Arbeit sprengen. Ein Übersichtsartikel zu diesem Problemkreis fehlt bisher, daher sei auf die Arbeiten von ELASHOFF/ELASHOFF (1974), ALBERT (1982), KAUFMAN/KING (1973) und für eine Bayes' sche Behandlung der Schätzung von "population totals" auf SMOUSE (1982) verwiesen. Als Beispiel für die wachsende Zahl von Arbeiten zu missing data in log-linearen Modellen siehe FUCHS (1982). Im Zusammenhang mit Aggregattabellen müssen noch die in Literatur weitgehend unbeachtet gebliebenen Arbeiten von YOLKINA/ZAGORUIKO (1978/1982) erwähnt werden. Die von ihnen entwickelten Algorithmen ZET-75 und VANGA füllen "leere" Zellen in Kreuztabellen auf, arbeiten also im Gegensatz zu den hier überwiegend behandelten Methoden auf Aggregatebene. Abgesehen von der exakten algorithmischen Formulierung ist die von ihnen hergestellte Verbindung zum "pattern-recognition"- Problem theoretisch interessant. Für die von ihnen behandelten Probleme geben sie eine hohe Effizienz der Algorithmen bis zu 30% fehlender Werte an. In der Literatur sind keine Anwendungen ihrer Methode nachzuweisen.

106. Eine mittlerweile klassische Untersuchung zu den Folgen der traditionellen Lösung stellt BIRNBAUM/SIRKEN (1950) dar. Ein Beispiel für die Anwendung der Gleichverteilungsannahme findet sich bei MUSTAFI (1981).

107. Dies gilt ebenso für die am weitesten verbreiteten Technik: Ausschluss aller Fälle mit fehlenden Werten (die "Do-nothing"- Technik; SANTOS 1981:22).

6. Effizienz der MD-Techniken bei der Kovarianzschätzung

Die in der Literatur und vor allem in der Forschungspraxis vorherrschende Auffassung, das MD-Problems als technisches Problem zu behandeln führt zur Frage nach der Effizienz der MD-Verfahren. Diese Frage setzt entweder die MAR-Annahme voraus oder die Annahme, daß ein eventueller systematischer Ausfallmechanismus nicht so stark mit den interessierenden Konstrukten zusammenhängt, daß eine Änderung substantieller Beziehungen möglich erscheint. Daneben könnte die Annahme erfolgen, daß ein MD-Verfahren in der Lage sei, einen systematischen Ausfallmechanismus zu kompensieren. Nur unter einer dieser Prämissen stellt sich die Frage nach einem "optimalen" MD-Verfahren, die in der Literatur einen großen Raum einnimmt[1/].

Aufgabe dieses Kapitels ist es, die vorliegenden Arbeiten zur Effizienz verschiedener MD-Techniken bei der Schätzung von Korrelationsmatrizen vorzustellen. Dabei wird sich zeigen, daß für einen Großteil der Verfahren des Kapitels 5 keine analytischen Ergebnisse in Hinsicht auf ihre Auswirkungen auf Kovarianzstrukturen vorliegen. Der größte Teil der gesamten MD-Literatur befaßt sich mit Parameterschätzungsmethoden in Regressionsmodellen unter der MAR-Annahme[2/]. Nur ein sehr kleiner Teil der Arbeiten ist hier von Interesse, da zumeist auf sehr spezielle MD-Muster bzw. Regressionsmodelle Bezug genommen wird[3/].

Der bedeutendste Teil der Literatur besteht aus Simulationsstudien, die das Verhalten einiger weniger Verfahren unter verschiedenen Bedingungen zeigen sollen. Mit wenigen Ausnahmen erfolgen die Studien explizit unter der MAR-Annahme; die Kritik der vorliegenden Arbeiten wird zeigen, daß selbst unter der MAR-Annahme kaum durch vorliegende Studien gestützte Aussagen über die Wahl eines "optimalen" MD-Verfahrens möglich sind. Das Problem zeigt sich bei Einbeziehung systematischer Ausfallmechanismen noch

wesentlich deutlicher: Abgesehen von sehr wenigen Ausnahmen existiert hierzu keine Literatur; die Auswirkungen systematischer Ausfälle auf die meisten MD-Techniken sind unbekannt.

6.1 Analytische Ergebnisse

Selbst unter der MAR-Annahme sind analytische Berechnungen zur Leistungsfähigkeit der verschiedenen MD-Techniken mit extremen Schwierigkeiten behaftet (SANTOS 1981:27, ANDERSON/BASILEVSKY/ HUM 1983:459), demzufolge existieren nur für sehr einfache Fälle exakte analytische Ergebnisse für Parameterschätzverfahren[4/]. Die damit verbundenen Probleme nehmen insbesondere bei den MD-Ersetzungsverfahren zu, wenn der Effekt auf multivariate Statistiken Gegenstand des Interesses wird. Mit Ausnahme der Arbeit von SANTOS (1981) existieren für die meisten der in Kapitel 5 dargestellten Ersetzungsverfahren[5/]keine analytischen Ergebnisse über ihren Effekt auf multivariate Statistiken, sondern lediglich für einfache univariate Statistiken (fast ausschließlich Mittelwert bzw. "totals")[6/]. Das Verhalten sowohl der Schätzverfahren als auch der Ersetzungsverfahren unter anderen Ausfallmodellen als MAR ist analytisch weitgehend unbekannt.

Die Darstellung der analytischen Ergebnisse wird daher hier auf die kurze Erwähnung zweier neuerer Arbeiten beschränkt, die in den Literaturübersichten bisher nicht enthalten sind: zum einen auf die Arbeit von SANTOS und zum anderen auf die Arbeit von BROWN (1983), der für Faktorenanalysen einige interessante Ergebnisse für Standardschätzer erhielt.

BROWN (1983) legt eine Studie zur asymptotischen Effizienz für eine Reihe von Kovarianzschätzern unter MAR vor in Hinsicht auf die Schätzung von Faktorenladungen. BROWN untersuchte LISTWISE, PAIRWISE, den EM-Algorithmus und eine einfache Regressionsmethode (für Variablenpaare getrennt

durchgeführte Regressionen, vgl. RATKOWSKY 1974). BROWN zeigt, daß mit Ausnahme von LISTWISE alle Verfahren konsistente Schätzer darstellen, falls mit wachsendem n gegen unendlich die Zahl der paarweise vorhandenen Werte ebenfalls gegen unendlich geht . LISTWISE stellt einen konsistenten Schätzer dar, falls zusätzlich die Zahl vollständiger Beobachtungen gegen unendlich geht. BROWN definiert eine Reihe von Indizes, die die "Dichte" der vorhandenen Beobachtungen (in bezug auf die Indikatorenmatrix) erfassen sollen. Er untersucht die Effizienz der Schätzer in bezug auf die Kommunalitäten unter Berücksichtigung verschiedener Muster. Bei balancierten Mustern (ein Muster heißt balanciert, wenn nur wenige Muster vorhanden und "gleichmäßig verteilt" sind; exakter bei BROWN 1983:270-271; 276) ergaben BROWN's Analysen die deutliche Überlegenheit von PAIRWISE gegenüber LISTWISE, solange die standardisierten Ladungen kleiner als .9 liegen. In nichtbalancierten Mustern sind die Ergebnisse weder für LISTWISE noch für PAIRWISE besonders gut, PAIR-WISE wird aber umso besser, je niedriger die Kommunalitäten liegen (BROWN 1982:281).

SANTOS (1981,1981b) untersuchte den Einfluß einer Reihe von MD-Ersetzungsverfahren unter MAR und MARC für den bi- und trivariaten Fall mit fehlenden Werten auf einer Variablen in Hinsicht auf die Kovarianzschätzung. Gegenstand der Arbeit sind u.a. die Schätzung auf der Grundlage der vollständigen Fälle, Mittelwertersetzung, Zellenmittelwertersetzung, Einfach-Regression, einfache Regression mit zufälligem Residuum, multiple Regression und multiple Regression mit zufälligem Residuum. Die Darstellung der Ergebnisse würde den Rahmen der Arbeit sprengen, bis auf einige wenige Ergebnisse muß daher auf SANTOS (1981) verwiesen werden.

Unter der MAR-Annahme stellen die vorhandenen Werte nur eine einfache Zufallsstichprobe aller Werte dar; demzufolge ist die Kovarianzschätzung mit den vollständigen Fällen unverzerrt. Da die Mittelwertersetzung und die zufällige Zuweisung die Kovarianz der Fälle mit fehlenden Werten mit Null annehmen, entsteht ein Bias, der proportional zum Anteil fehlender Werte ist. Unter MARC ist der Bias für die Schätzung auf der Grundlage der vollständigen Fälle eine

Funktion der Differenz der Zellenresponseraten (falls die Zellenmittelwerte auf beiden Variablen unterschiedlich sind und die Kovarianzen beider Variablen in allen Zellen nicht gleich Null sind).

Auch die Arbeit von SANTOS zeigt, daß die Gewinnung analytischer Ergebnisse fast nur unter extrem vereinfachenden Annahmen möglich ist. Außer den in der Literatur häufig zu findenden Warnungen vor der Benutzung der Mittelwertersetzungsverfahren, falls Kovarianzschätzungen beabsichtigt sind, gibt es für die meisten der in Kapitel 5 dargestellten Verfahren keine auf analytische Ergebnisse gestützten Empfehlungen. Für systematische Ausfallmechanismen und komplexere multivariate Probleme existieren für keine MD- Technik analytische Ergebnisse.

6.2 Simulationsstudien

Das Verhalten der MD-Verfahren hängt von einer großen Zahl von Bedingungen ab, so z.B. dem Ausfallmechanismus, dem MD-Anteil, dem MD-Muster, den Variablenzusammenhängen, der Zahl der Fälle, der Zahl der Variablen usw.. Um die Effekte der großen Zahl von Parametern ohne analytische Versuche dennoch zu studieren, wurden für die meisten MD-Verfahren Simulationsstudien durchgeführt. Simulationsstudien sind bei diesen Problemen die einzige Möglichkeit, zu verallgemeinerbaren Aussagen zu gelangen. Der Vergleich verschiedener Verfahren an einem Datensatz läßt die Spezifika des jeweiligen Datensatzes unkontrolliert, Verallgemeinerungen sind nicht möglich[7/]. In noch stärkerem Ausmaß trifft dies auf Studien zu, die von vornherein mit unvollständigen Datensätzen arbeiten[8/]. In diesen Fällen ist kein Vergleich mit den "wahren" Werten möglich. Nur falls alle das Verhalten der MD-Verfahren beeinflussenden Bedingungen kontrolliert werden können, kann ein Vergleich verschiedener Verfahren zu prinzipiell verallgemeinerbaren Aussagen gelangen (HAITOVSKY 1965:4, ANDERSON/BASILEVSKY/ HUM 1983:446).

Es gibt bisher keine Arbeit, die einen Gesamtüberblick über die vorgenommenen Simulationen zu geben versucht, wenn auch mit den Arbeiten von HAMILTON (1975:41-47); ANDERSON/BASILEVSKY/HUM (1983:459-472) Ansätze hierzu existieren. Ein Überblick wird zudem einmal durch die fast immer unzureichende Dokumentation der Studien[9/] und zum anderen dadurch erschwert, daß Replikationen mit nur wenigen anderen variierten Parametern nicht existieren. Die Simulationsstudien sind kaum vergleichbar (ANDERSON/ BASILEVSKY/HUM 1983:467), es gibt weder einen Set von Standardverfahren, der immer mit einbezogen wird, noch einen Standard bei der Parameterwahl oder der Wahl der Effizienzmaße[10/]. Orthogonale Designs bilden die Ausnahmen, häufig werden Wiederholungen mit unterschiedlichen Löschungsmustern an einem Datensatz als unabhängige Iterationen innerhalb einer Zelle des Designs angesehen. M.a.W.: Ein Teil der Simulationen genügt elementaren Anforderungen an Monte-Carlo-Studien nicht[11/].

Im folgenden werden zunächst einige Arbeiten unter der MAR-Annahme etwas ausführlicher dargestellt und kritisiert, die in ihrer Rezeption besonders wichtig wurden[12/].

6.2.1 Arbeiten unter der MAR-Annahme

Zu den einflußreichsten Arbeiten gehören HAITOVSKY (1969), TIMM (1970), GLEASON/STAELIN (1975), BEALE/LITTLE (1975), KIM/CURRY (1977) und LITTLE (1979). Die Arbeit von HAITOVSKY wird in der Regel herangezogen, um die Überlegenheit von LISTWISE gegenüber PAIRWISE zu stützen (z.B. LITTLE 1979:76), die Arbeit von KIM/ CURRY (1979) um das Gegenteil zu belegen (z.B. HOLM 1979:152), BEALE/LITTLE (1975) und LITTLE (1979) werden zitiert, um die Überlegenheit des EM-Algorithmus zu demonstrieren.

HAITOVSKY (1966,1969) unterwarf in seiner Simulation 5 Verfahren einer Untersuchung an verschiedenen Datensätzen mit je 10 Wiederholungen. Vier Verfahren zeigten keine signifikanten Differenzen:
HAITOVSKY (1969) berichtet nur über den Vergleich von LISTWISE und COVPAIR[13/]. Die Datenstrukturen variierten n zwischen 100 und 1000, die Zahl unabhängiger Variablen (2,4,5) und drei Löschungsmuster (darunter ein systematisches) sowie den Anteil fehlender Werte. Abhängige Variablen waren die mittleren Fehlerquadrate der Regressionskoeffizienten des spezifizierten Modells. Das Design war nicht orthogonal.

HAITOVSKYS Ergebnis zeigte eine deutliche Überlegenheit der LIST-WISE-Methode gegenüber COVPAIR, wobei nur in den Fällen, in denen der Anteil vollständiger Fälle unter 10% sank, COVPAIR bessere Ergebnisse erzielte.

KIM/CURRY (1977:223) führen die hohen multiplen R^2's und r's, sowie Unklarheiten über die MD-Anteile als Kritik an den Simulationen HAITOVSKY's an. Der erste Einwand stimmt zwar für die meisten Modelle, Tabelle 5 bei HAITOVSKY (1969:75) zeigt bei einem R^2 = .158 aber dieselbe Tendenz wie alle anderen Modelle. Neben den vielen Unklarheiten in technischen Details bleiben als Kritikpunkte die hohen Interkorrelationen der unabhängigen Variablen, die Nichtorthogonalität des Designs, die unsystematische Variation der Interkorrelationen und die Beschränkung des systematischen Ausfalls auf ein spezielles Regressionsmodell (mit R^2 = 0.992 und 0.778 (!)) bei einem n=100 bestehen. HAITOVSKY's Studie stützt damit die weitreichenden Konsequenzen, die in der MD- Literatur daraus gezogen wurden, nur begrenzt[14/].

TIMM (1969,1970) untersuchte das Verhalten der Mittelwertersetzung, von LISTWISE und der Methoden BUCK's und DEAR's unter MAR bei n = 50, 100, 200; p= 2, 5, 10, dem Anteil fehlender Werte (1%, 10%, 20%) und 9 verschiedenenen Kovarianzmatrizen mit mittleren Interkorrelationen zwischen .189 und .834. TIMM benutzte rein zufällige Löschung der MD. Abhängige Variablen waren die Spur der Differenzenmatrix der Korrelations- und Kovarianzmatrizen. Für jede Bedingung wurden zwei Replikationen durchgeführt. Alle Ver-

fahren erbrachten stets positiv definite Schätzungen. LISTWISE und die Mittelwertersetzungsmethode neigten zur Unterschätzung der mittleren Interkorrelation, DEAR zur Überschätzung und BUCK zu einer leichten Überschätzung (mit Ausnahme von p = 2).

TIMM erstellte eine umfangreiche Liste, welche Methode unter welchen Bedingungen als "optimales" Verfahren zu wählen sei. Insgesamt brachten BUCK und DEAR unter nahezu allen Bedingungen die besten Ergebnisse, insbesondere bei mittleren und hohen Interkorrelationen. DEAR erbrachte bei extrem hohen Ausfallraten und hohen Korrelationen bessere Ergebnisse als BUCK.

Als Kritikpunkte an der Studie ist die niedrige Zahl der Replikationen, der extrem hohe MD-Anteil von 20% und die Struktur der Kovarianzmatrizen[15/] zu nennen. Lediglich drei der neun Kovarianzmatrizen verfügen über mehr als einen Eigenwert größer 1.0. Aus diesem Grund (hohe Multikollinearität bzw. klar durch einen Faktor dominierte Eigenwertstruktur) ist insbesondere in den 10-Variablenbedingungen das gute Ergebnis für BUCK und DEAR wenig überraschend (vgl. TIMM 1969:60, Tab. XVII).

GLEASON/STAELIN (1975) untersuchten die Mittelwertersetzungsmethode, DEAR, eine teilweise iterative Version der Hauptkomponentenmethode, COVPAIR[16/] und eine modifizierte BUCK-Version. Die BUCK-Prozedur wurde nicht mit der LISTWISE-Matrix begonnen, eine Kovarianzkorrektur wird von ihnen nicht erwähnt. Die Hauptkomponentenmethode benutzte mehr als eine Komponente, wenn die "root mean square"-Korrelation größer als .16 wurde. Überschritt diese 0.30 wurde sie zweimal iteriert. Die Eingabematrix für die Regressionsmethode bestand aus der endgültigen Schätzung der Hauptkomponentenmethode. Sie variierten n (50,100,250), die Zahl der Variablen (10,15,30), den MD-Anteil (4%,12%,20%) und die "root mean square"-Korrelation zwischen .111 und .552. Das Design war nicht orthogonal. Abhängige Variablen waren eine modifizierte Version des Spurkriteriums der Korrelationsmatrizen und u.a. ein standardisiertes "root-mean-square"-Residuum der geschätzten Daten[17/].

GLEASON/STAELIN bestätigen im wesentlichen die Ergebnisse, die TIMM erzielte. An den Einzelergebnissen fällt insbesondere die Ähnlichkeit der Reaktionskurven der Hauptkomponentenmethode und der Regression auf, die beide umso besser wurden, je stärker die Zusammenhänge in der Matrix wurden[18/]. Bei dem Vergleich der Mittelwertersetzungsmethode und COVPAIR bei der Schätzung der Startmatrix der Verfahren kommen sie zu dem Ergebnis, daß ALLVALUE bei stärkeren Zusammenhängen in der Matrix zu besseren Ergebnissen gelangt.

Ein direkter Vergleich der Ergebnisse mit denen anderer Studien ist nicht möglich: GLEASON/STAELIN legen ihre Ergebnisse in Form von multiplikativen Regressionsgleichungen vor. Für den Korrelationsindex liegen die R^2's zwischen .71 für die Regressionsmethode und .92 für DEAR, bei dem Datenindex zwischen .37 für DEAR und .66 für die Hauptkomponentenmethode. Neben der ungewöhnlichen Darstellung der Resultate bleiben weitere Kritikpunkte: Das Design ist nicht orthogonal, die 20%-MD-Bedingung unrealistisch, LISTWISE wurde nicht berücksichtigt, zwei Verfahren wurden entscheidend modifiziert und die Benutzung der Ausgabematrix des Hauptkomponentenverfahrens als Eingabematrix des Regressionsverfahrens legt ein Artefakt nahe. Schlußfolgerungen lassen sich aus der veröffentlichten Studie nur sehr begrenzt ableiten.

BEALE/LITTLE (1975) verglichen LISTWISE, BUCK, "iterated BUCK" (EM-Algorithmus) und drei weitere Modifikationen des "iterated Buck" (u.a. mit Gewichtung der unvollständigen Fälle) bei n=50, 100, 200 und MD-Anteilen zwischen 5% und 40% (rein zufällige, unabhängige Löschung) in 7 Regressionsmodellen mit 3-5 Variablen mit R^2 zwischen .0888 und .9904. Sie wiederholten jede Bedingung zehnmal. Abhängige Variable war das Abweichungsquadrat der geschätzten und der tatsächlichen Werte der abhängigen Variablen im jeweiligen Regressionsmodell.

BEALE/LITTLE erzielten durchgängig bessere Ergebnisse mit BUCK und EM als mit LISTWISE, EM brachte fast immer eine Verbesserung von BUCK durch die Iteration. Die Modifikationen des "iterated BUCK" zeigten kaum Unterschiede.
Außer gegen die Begrenzung auf prinzipiell 3 Methoden und die teilweise extremen Ausfallquoten richtet sich die Hauptkritik gegen die Kovarianzstrukturen, die in der Simulation benutzt wurden. Selbst in dem Modell mit dem niedrigsten R^2 von .0888 mit vier Variablen liegen die Interkorrelationen der Prädiktoren um 0.9. Alle Modelle sind durch hohe Multikollinearität gekennzeichnet (vgl. ANDERSON/BASILEVSKY/HUM 1983:468). Verallgemeinerungen auf andere Kovarianzstrukturen erscheinen problematisch.

LITTLE (1979) verglich LISTWISE mit der ML-Schätzung nach ORCHARD/WOODBURY (die nahezu mit dem EM-Algorithmus über "iterated BUCK" identisch ist) bei n=50 bzw. 100 in einem 4-Variablen Regressionsmodell mit R^2 zwischen .55 und .94. LITTLE benutzte 6 spezielle Löschungsmuster mit zwischen 60% und 15% vollständiger Fälle. Für jede Bedingung wurden 50 Datensätze erzeugt. LITTLE gelangt zu der Schlußfolgerung, daß die Benutzung des ML-Schätzers bei hohen Fallzahlen und nicht allzu hohen MD-Anteilen einen Gewinn gegenüber der LISTWISE darstellt.
Die Studie von LITTLE erlaubt aufgrund der Beschränkung auf zwei Verfahren, wegen der untypischen Löschungsmuster und Kovarianzstrukturen keine weitergehenden Schlußfolgerungen[19/].

KIM/CURRY (1977) verglichen PAIRWISE und LISTWISE in einem Regressionsmodell mit der BLAU/DUNCAN (1967:169) Korrelationsmatrix. Bei konstantem n = 1000 löschten sie 10% der Fälle rein zufällig und wiederholten den Prozeß zehnmal. Abhängige Variable war ein Abweichungsindex der einzelnen Regressions- und Korrelationskoeffizienten, bzw. der Kovarianzmatrizen. Bei einer Wiederholung des Experiments mit variierenden MD-Anteilen (1%, 2%, 5%, 10%) und einer Wiederholung mit variierender Variablenzahl (2-5) gelangten sie stets zu einem besseren Ergebnis für PAIRWISE.

Es handelt sich bei der Arbeit von KIM/CURRY nur in sehr begrenztem Sinn um eine Simulationsstudie, das Modell ist auf sehr große Fallzahlen und eine bestimmte Kovarianzstruktur beschränkt. Die Autoren sehen diese Begrenzung ihrer Studie und leiten lediglich die Schlußfolgerung ab, daß in bestimmten Situationen PAIRWISE gute Resultate zeigen kann. Eine generelle Empfehlung läßt sich aus der Studie nicht herleiten[20/].

Der Vollständigkeit halber sollen noch die Arbeiten von CHAN/DUNN erwähnt werden, die Kovarianzschätzer im Rahmen der Diskriminanzanalyse untersuchten. CHAN/DUNN (1972) verglichen LISTWISE, COVPAIR, MEAN, BUCK, eine modifizierte BUCK-Version und DEAR mit homogenen Korrelationsmatrizen unter MAR mit 100 Iterationen (CHAN/DUNN 1972:474 erwähnen 250 zusätzliche "randomly chosen correlation matrices", deren Ergebnisse aber so ähnlich waren, daß sie von ihnen nicht wiedergegeben werden). Berichtet wird von den Ergebnissen mit 2,3,4 und 8 Variablen, einem MD-Anteil von .2 mit dem Anteil korrekt klassifizierter Fälle als abhängiger Variablen. Im bivariaten Fall zeigte sich BUCK bei kleinen Determinanten überlegen, DEAR bei mittleren und MEAN bei großen Determinanten. Bei höheren Variablenzahlen änderte sich lediglich für große Determinanten das beste Verfahren, hierbei erbrachte LIST-WISE das beste Ergebnis. Die Begrenzungen des Designs in Hinsicht auf Schlußfolgerungen für Kovarianzmatrizenschätzungen sind offensichtlich. Dasselbe gilt für die Nachfolgeuntersuchung CHAN/ GILMAN/DUNN (1976), die weitere Modifikationen der Grundtechniken testete und zu ähnlichen Ergebnissen gelangte, wenn auch die modifizierte BUCK-Version bei diesem Experiment überwiegend bessere Ergebnisse erbrachte[21/].

Zu den Hot-Deck-Verfahren existieren fast ausschließlich Arbeiten zu den Auswirkungen auf univariate Statistiken[22/]. Die beiden umfangreicheren Arbeiten zu ihren Auswirkungen auf die Kovarianzstrukturen sollen im folgenden dargestellt werden[23/].

VACEK/ASHIKAGA (1980) testeten in einer Simulation mit variierter Fallzahl (n=20, 50, 100), variierter Variablenzahl (2, 3, 5, 10) und homogenen

Korrelationsmatrizen (r = 0.0, 0.04, 0.16, 0.36, 0.64, 0.81, 0.99) BUCK's Methode und ihre Version der "nearest neighbor rule". Sie simulierten 10%, 20% und 30% fehlender Werte, wobei sie zwei Löschungsmuster benutzten. Das erste Muster war ein monotones Muster, das zweite erlaubte einen fehlenden Wert pro Fall (immer auf derselben Variablen); nur die Ergebnisse des zweiten Musters werden berichtet[24/]. Sie geben 1000 (!) Iterationen an, wobei nicht klar ist, ob dies pro Zelle gilt. Sie verglichen u.a. die mittleren Fehlerquadrate der Kovarianzmatrizen.

Die "Nearest neighbor rule" erzielte mit steigender Variablenzahl und steigendem MD-Anteil auch in kleinen Stichproben zumeist bessere Ergebnisse als das Regressionsverfahren. Beide Verfahren zeigten eine deutliche Verbeserung, wenn der Zusammenhang in der Korrelationsmatrix anstieg. Bei n = 100 und starken Zusammenhängen waren die Ergebnisse für das Regressionsverfahren immer leicht besser als für die "nearest neighbor rule"[25/]. Das Regressionsverfahren erbrachte in Hinsicht auf die Varianzen im 10 Variablenfall bei abnehmender Korrelation und zunehmendem MD- Anteil sehr schlechte Schätzungen.

Da der Effekt der Stärke der Zusammenhänge auf die "nearest neighbor rule" nicht so stark war wie auf das Regressionsverfahren, halten die Autoren die Anwendung insbesondere dann für angebracht, wenn multivariate Analysen beabsichtigt sind und die Korrelationsstruktur unbekannt ist.

Neben den sehr kleinen Fallzahlen muß vor allem die Begrenzung des MD-Musters kritisiert werden. Auch die nicht berichtete zweite Form des MD-Musters entspricht keinem "typischen" MD- Muster und muß daher ebenfalls als Bewertungsgrundlage abgelehnt werden. Die Simulation kann daher nicht als verallgemeinerbar angesehen werden.

KAISER (1983) untersuchte den Einfluß verschiedener Hot-Deck Verfahren in kleinen Stichproben auf die Kovarianzstruktur. Er untersuchte drei Hot-Deck Versionen: Sequentielles Hot-Deck (jeder Donor wurde nur einmal benutzt), Nearest-Neighbor-Hot-Deck (bei Äquidistanz Mittelwert der Donors) und die

Methode BAILAR's (1978). Das Design war ein (3*4*3)-faktorielles Design, wobei der Anteil unvollständiger Records (.1 - .3) und die Anzahl fehlender Werte pro Record (1-4) variiert wurden. Erzeugt wurden 500 (!) Datensätze pro Zelle, jeder Datensatz umfaßte 30 Fälle und 10 Variablen, 2 davon waren die Schichtungsvariablen. Die Interkorrelation der zehn Variablen war fixiert, die Interkorrelationen lagen zwischen .19 und .47., die Mediankorrelation der MD-behafteten Variablen mit den Schichtungsvariablen lag bei .27, die Interkorrelation der Schichtungsvariablen bei .25, KAISER benutzte als abhängige Variablen den Likelihood-Ratio-Test, das modifizierte Spur-Kriterium von TIMM und das standardisierte Residuum von GLEASON/STAELIN[26/]. In Hinsicht auf die hier überwiegend interessierenden Ergebnisse der Kovarianzschätzungen erbrachte die Distanztechnik unter allen Bedingungen mit dem Spur-Kriterium die besseren Ergebnisse. Dasselbe Ergebnis zeigte das auf die Güte der geschätzten Daten bezogenen Kriterium. Der Likelihood-Ratio-Test zeigte vergleichsweise die Überlegenheit des sequentiellen Hot-Decks gegenüber den anderen Techniken. Allerdings steigt der Anteil der vervollständigten Stichproben, die eine signifikante Abweichung ($p < .05$) von der tatsächlichen Kovarianzstruktur aufweisen, mit dem Anteil unvollständiger Records und der MD-Zahl innerhalb der Records deutlich an. Selbst in den Bedingungen mit 10% unvollständiger Records und nur einem fehlenden Wert pro Record liegt der Anteil der signifikanten Abweichungen bei 17%, um im Extremfall bei 30% unvollständiger Records mit 4 fehlenden Werten auf bis zu 40% anzusteigen. KAISER rät daher von der Anwendung von Hot-Deck Techniken ab, falls multivariate Analysen beabsichtigt sind.

Die Kritik an der Studie muß die extrem niedrige Fallzahl und die fixierte Korrelationsstruktur sowie die Beschränkung der MD- Muster und Anteile erwähnen. Ebenso bedauerlich ist der völlige Ausschluß aller anderen MD-Verfahren[27/]. Die extremen Abweichungen der Kovarianzmatrizen unter MAR sollten allerdings ausreichen, um die Hot-Deck-Verfahren in Hinsicht auf Kovarianzmatrizenschätzungen bei kleinen Stichproben als unbrauchbar ansehen zu lassen.

6.2.2 Systematische Ausfälle

VAN GUILDER/AZEN (1981) unternahmen mit ihrer Simulationsstudie den Versuch, das Verhalten des EM-Algorithmus (EM), der LISTWISE- und der ALLVALUE-Methode unter anderem unter verschiedenen Ausfallmechanismen zu untersuchen. Da auch die Güte der Ersetzungen untersucht werden sollte, wurden die MD durch multiple Regressionen mit den LISTWISE- bzw. ALLVALUE-Kovarianzmatrizen geschätzt. Als abhängige Variablen der Simulation benutzten sie zwei Maße, von denen sich eines auf die Güte der geschätzten Regressionskoeffizienten (C1) und das andere (C2) auf die Güte der MD-Schätzungen[28/] bezieht[29/].

Die Stichprobengröße betrug 100, die Variablenzahl 4; zugrunde lag ein Regressionsmodell mit allen Regressionsparametern gleich eins. Die drei unabhängigen Variablen folgten in den Bedingungen stets der gleichen Verteilungen ("i.i.d."): normal, gemischt normal, lognormal. Variiert wurde sowohl das multiple R^2 (.5, .9) als auch die homogene Interkorrelation (.25, .75). Der Anteil fehlender Werte wurde auf 5% und 25% fixiert; drei Ausfallmechanismen wurden benutzt: MAR, "truncated" und "related".

"Truncated" bestand aus der Löschung des Wertes auf den Variablen 2 oder 3 mit der Wahrscheinlichkeit p, falls der Wert der jeweiligen Variablen über dem Median lag, "related" bestand aus der Löschung des Wertes mit der Wahrscheinlichkeit p auf der Variablen 2 oder 3, falls Variable 1 über dem Median lag[30/].

Unter den 5%-MD-Bedingungen lag für C1 und für C2 ALLVALUE im allgemeinen schlechter als LISTWISE und EM, unabhängig von der Verteilung und dem Ausfallmechanismus (MDGP), solange die Korrelation hoch lag[31/]. Unter der 25%-MD-Bedingung war EM im allgemeinen sowohl für C1 als auch für C2 unter "truncated" das beste Verfahren, LISTWISE und ALLVALUE zeigten eine starke Abhängigkeit von der Stärke der Korrelation: Je höher diese lag, umso besser war LISTWISE. EM zeigte für hohes R^2 schlechteres Verhalten als bei

niedrigem R^2, unabhängig vom MDGP. Die besten Ergebnisse für EM ergaben sich unter MAR, die schlechtesten unter "truncated"[32/]. Das schlechte Ergebnis für ALLVALUE führen die Autoren auf die gelegentlich nahezu singulären Kovarianzmatrizen in den Bedingungen mit hohen Korrelationen zurück. Sie berichten von einer kleineren Simulationstudie, in der eine Ridge-Prozedur für ALLVALUE benutzt wurde, die für C1 (nicht für C2) eine Verbesserung erbrachte, die ALLVALUE teilweise bessere Resultate zeigen ließ als LISTWISE.

Neben der ungewöhnlich unvollständigen Beschreibung sowohl der Simulation als auch der Ergebnisse sind eine Reihe von Einwänden gegen die Arbeit von VAN GUILDER/AZEN notwendig. Zunächst sind die benutzten Simulationsparameter extrem: 25% fehlender Werte wird im allgemeinen kaum toleriert werden, ebenso ungewöhnlich hoch sind multiple R^2 von .9, zumindest in der sozialwissenschaftlichen Forschung.

Die Berechnung von C1 für ALLVALUE und LISTWISE unabhängig von den MD-Schätzungen ist zumindest unrealistisch: Im Prinzip werden von ihnen 5 Methoden getestet (EM, LISTWISE, ALLVALUE sowie multiple Regressionen ausgehend von LISTWISE und ALLVALUE: MRC; MRP), berichtet wird für C1 von EM, LISTWISE und ALLVALUE, für C2 von EM, MRC und MRP (wobei MRC einen nach dem ersten Schritt abgebrochenen, unkorrigierten EM darstellt). Weiterhin stellt die Einschränkung, keine Fälle mit MD auf allen Variablen zuzulassen, eine schwer zu rechtfertigende Begrenzung dar. Neben den sonst möglichen Einwänden[33/] sind vor allem die Schlußfolgerungen der Autoren kritisierbar: Aus den schlechten Ergebnissen der Verfahren unter verschiedenen MDGP's und den deutlichen Interaktionseffekten (MDGP*Methode) leiten sie Empfehlungen für die Benutzung der Verfahren in bestimmen Fällen her, wobei sie nach Parameterschätzungs- und MD-Ersetzungsabsicht unterscheiden. An keiner Stelle (außer zur Identifikation des "besten" Verfahrens) gehen sie auf die notwendige Modellierung des Ausfallprozesses ein. Der MDGP wird gleichwertig wie alle anderen Simulationsparameter behandelt: Nach der Identifikation des MDGP erfolgt die Wahl des Verfahrens.

SANTOS (1981a,1981b) berichtet von einer Simulation mit zwei realen Datensätzen, bei der LISTWISE, Mittelwertersetzung, Zellenmittelwertersetzung, "random cell" und ein multiples Regressionsverfahren mit und ohne Zufallsresiduum getestet wurden. Als Ausfallmechanismus wurde ein MARC-Modell simuliert. Bei einem n = 1036 wurden 10.1% der Daten der abhängigen Variablen im Regressionsmodell ("hourly rate of pay") gelöscht. Der MD-Anteil in den vier Zellen schwankte zwischen 3.6% und 17.5%. Der zweite Datensatz mit n = 400 wurde ebenfalls mit vier Zellen simuliert, deren MD-Anteil der abhängigen Variablen ("quarterly earnings") zwischen 45.2% und 78.9% lag, so daß der gesamte MD-Anteil 55% betrug. Von jedem Datensatz wurden 10 Replikationen benutzt. Verglichen wurden die Ergebnisse der Regressionkoeffizienten.

LISTWISE lieferte durchgängig bessere Ergebnisse als die Mittelwertersetzung. Durch die sorgfältige Konstruktion der Zellen zeigten auch die beiden Zellentechniken stets bessere Ergebnisse als die Mittelwertersetzung. Die Regressionsverfahren zeigten für diejenigen Variablen, die zur Imputation benutzt wurden, stets bessere Ergebnisse als für die Variablen, die nicht zur Imputation herangezogen wurden. In bezug auf die R^2-Statistik ergab sich kein einheitliches Bild. SANTOS (1981:120-121) warnt ausdrücklich vor Verallgemeinerungen des Ergebnisses hinsichtlich der Güte von LISTWISE. Die Begrenzungen des Designs liegen vor allem in der Modellierung des MARC-Modells. Zwar wurden die Klassen so gebildet, daß ihre MD-Anteile denen der Hauptuntersuchung, aus der der Experimentaldatensatz stammt, entsprechen, aber durch das Design wurde MARC in reiner Form erzwungen: Im Datensatz der Hauptuntersuchung bestehende Differenzen zwischen den vollständigen und unvollständigen Fällen bleiben unberücksichtigt. Diese von LITTLE (1981c:153) formulierte Kritik trifft dann, wenn die Untersuchung nicht als Simulation, sondern als Einzelfallstudie aufgefaßt wird. Bei einer Betrachtung als Simulation bestehen die üblichen Kritikpunkte: unrealistische Löschung nur auf einer Variablen, keine Variation der Kovarianzstruktur und der MD-Anteile. Wie auch immer man die Arbeit auffaßt, Verallgemeinerungen der Befunde sind nicht möglich.

6.3 Zusammenfassung

Trotz der kaum noch zu überblickenden Literatur zu MD-Problemen, insbesondere zur Kovarianzschätzung und multiplen Regression, kann festgehalten werden, daß die Frage nach einem "optimalen" MD-Verfahren auf der Grundlage der bisher vorliegenden Literatur nicht beantwortet werden kann.

Weder existiert, mit Ausnahme der in Kapitel 4 angesprochenen Probleme, Literatur zur Schätzung bei systematischen Ausfällen, noch können aus den vorliegenden Simulationsstudien eindeutige Empfehlungen zur Wahl eines MD-Ersetzungsverfahrens selbst unter MAR gegeben werden. Der Großteil der Verfahren ist nicht systematisch getestet worden, analytische Bewertungen sind mit extremen Schwierigkeiten behaftet und die meisten Simulationen, wie gezeigt wurde, nicht verallgemeinerbar. Die Auswirkungen der Ersetzungsverfahren auf die Kovarianzstruktur sind selbst unter MAR weitgehend unbekannt.

Das wichtigste Ergebnis dieses Kapitels ist die Feststellung, daß die in der Literatur zu findenden Empfehlungen

- alle auf der MAR-Annahme beruhen,
- weitgehend ohne analytische Ergebnisse gegeben wurden,
- überwiegend auf Simulationen basieren, die nur sehr schwer verallgemeinerbar sind[34/].

Alle Versuche, die unter MAR gewonnenen Ergebnisse zur Leistungsfähigkeit verschiedener MD-Techniken auf Probleme anzuwenden, bei denen systematische Ausfälle wahrscheinlich sind, müssen zusätzlich zu den oben beschriebenen Problemen die Annahme machen, daß die relative Leistungsfähigkeit der Verfahren unter anderen Ausfallmechanismen erhalten bleibt, also kein "Interaktionseffekt" zwischen Ausfallmechanismus und Leistungsfähigkeit der MD-Verfahren besteht. Um die Frage nach diesem Interaktionseffekt zu beantworten wurde die Simulationsstudie, die in den beiden folgenden Kapiteln beschrieben wird, unternommen.

Anmerkungen zu Kapitel 6

1. Beispiele hierfür sind die Arbeiten von TIMM (1970), HAMILTON (1975), GLEASON/STAELIN (1975), STEWART (1982) und zuletzt ANDERSON/BASILEVSKY/HUM (1983).

2. Die Literatur ist nahezu unüberschaubar umfangreich. Zu den frühen Ansätzen vgl. ausführlich AFIFI/ELASHOFF (1966). HAMILTON (1975) enthält nahezu alle Arbeiten bis 1974, von den neueren kurzen Übersichten ist ANDERSON/BASILEVSKY/HUM (1983) die vollständigste, wenn auch im hier interessierenden Zusammenhang wichtige Arbeiten (z.B. GLEASON/STAELIN 1975 und VAN GUILDER/AZEN 1981) fehlen und mehrere Fehler vorhanden sind.

3. Auf spezielle Ansätze zu MD-Problemen in multiplen Regressionsmodellen unter MAR kann hier nicht eingegangen werden, vgl. DAGENAIS (1972), PRESS/SCOTT (1974,1976), SINGH/SEDRANSK (1983: 229-235), ferner MEHTA/SWAMY (1974) zu Bayes-Ansätzen, sowie DAGENAIS (1971,1972), GOURIEROUX/MONFORT (1981), vor allem KMENTA (1981) und BALESTRA/KMENTA (1984). Als ein Beispiel sei CONNIFFE (1983) erwähnt, der den Gewichtungsansatz bei BEALE/ LITTLE (1975) (erfolglos) auszubauen versucht. CONNIFFE hält die Suche nach anderen als den iterativen ML-Schätzern für sinnvoll, da diese mit hohem Rechenaufwand verbunden sind und bestimmte Verteilungsannahmen voraussetzen (CONNIFFE 1983:269). CONNIFFE's Ansatz setzt allerdings ebenfalls Verteilungsannahmen voraus. CONNIFE (1983b) bringt analytische Ergebnisse für einige spezielle Schätzer im Regressionsmodell mit zwei unabhängigen Variablen, von denen eine MD-behaftet ist.

4. Zu analytischen Effizienzberechnungen unter MAR in einfachen Fällen für die Parameterschätzverfahren vgl. FIERING (1962), GLASSER (1964), AFIFI/ELASHOFF (1967,1969), KELEJIAN (1969) sowie DONNER (1982).

5. Sieht man von den gelegentlichen Bemerkungen zur Mittelwertersetzung und deren Einfluß auf Korrelationskoeffizienten ab. Vgl. z.B. HERTEL (1976:461-468), der dem Zellenmittelwertverfahren im Gegensatz zum Gesamtmittelwertverfahren eine Inflationierung der Korrelationskoeffizienten unterstellt.

6. Analytische Arbeiten zu Hot-Deck Verfahren finden sich bei ZINDLER (1965), SZAMEITAT/ZINDLER (1965) und BAILAR/BAILAR (1978, 1983).

7. So z.B. BUCK (1960), hierzu HAITOVSKY (1965:33): "But a single example certainly does not prove a thing."

8. Zu den Arbeiten mit diesem Ansatz gehören JACKSON (1968) und GUERTIN (1968). LÖSEL/WÜSTENDÖRFER (1974:342-354) stellen die "stärker pragmatisch orientierten Vergleichsuntersuchungen" kritiklos den Simulationsstudien gegenüber.

9. Zur Kritik der Darstellung von Simulationen allgemein vgl. HAUCK/ANDERSON (1984:214-215). Eine Liste derjenigen Parameter, die der Dokumentation in technischer Hinsicht bedürfen, findet sich bei HOAGLIN/ANDERSON (1975).

10. TIMM (1969:97-98) erwähnt, daß die Entscheidung, welches Verfahren "optimal" für eine gegebene Problemstellung ist, nicht invariant gegenüber dem Bezug auf Kovarianzmatrizen, Korrelationsmatrizen oder einem Eigenwertkriterium ist. In der Literatur wird dieses Problem nur selten thematisiert, die Schlußfolgerungen in Hinsicht auf "optimale" Verfahren werden in der Regel davon unabhängig gezogen.

11. Einige Arbeiten, die besonders unvollständig dokumentiert wurden, seien als Beispiel erwähnt: VAN GUILDER/AZEN (1981) definieren ihre abhängigen Variablen nicht, GLEASON/STAELIN (1975) präsentieren ihre Ergebnisse in Form angepaßter Regressionsgleichungen (mit teilweisen sehr schlechtem "fit"); den Arbeiten von TIMM (1969,1970) und VACEK/ASHIKAGA (1980) sind nur mit Mühe die Iterationszahlen zu entnehmen. Andere Studien stellen nur begrenzt Simulationen dar, z.B. FORD (1976), FORD/KLEWENO/TORTORA (1980) sowie KALTON (1983).

12. Eine Reihe von Arbeiten insbesondere zum bivariaten Fall werden hier vernachlässigt, z.B. die häufig zitierte Simulationsstudie von AFIFI/ELASHOFF (1969b), die 5 Schätzer im bivariaten Fall in kleinen Stichproben unter MAR testeten. Eine konsistente Regressionsmethode zeigte nahezu immer bessere Ergebnisse als die Berechnung auf der Grundlage der vollständigen Fälle. Ebenso bleibt die Simulationsstudie von vier Kovarianzschätzern im bivariaten Fall von SRIVATAVA/ZAATAR (1973) ausgeklammert. COVPAIR zeigte die besten Ergebnisse, wenn $r < .3$ gegeben war, ansonsten war ab $r>.5$ und einer hohen Zahl vollständiger Beobachtungen der ML-Schätzer überlegen. Insgesamt empfehlen die Autoren auf Grund Ihrer Ergebnisse PAIRWISE. Ebenso wird auf Grund der Beschränkung auf den Fall einer Regression mit zwei unabhängigen Variablen, von denen eine MD-behaftet ist, auf die Darstellung der Arbeit von DONNER/ROSNER (1982) verzichtet. Sie verglichen bei n=100 LISTWISE, COVPAIR und den ML-Schätzer von EDGETT mit MD-Anteilen zwischen 0.1 und 0.5 und Korrelationen zwischen 0.2 und 0.8. Der ML-Schätzer zeigt sich in fast allen Bedingungen überlegen. Die Arbeit von DAGENAIS (1972) zu einem sehr ähnlichen Problem entfällt unter anderem wegen der Beschränkung zweier Schätzer auf den vorliegenden Spezialfall. DAGENAIS (1972) testete u.a. COVPAIR und LISTWISE nur im trivariaten Fall, wobei jede unvollständige Einheit lediglich auf einer der beiden unabhängigen Variablen fehlende Werte besaß.

13. Bei HAITOVSKY werden die Schätzer anders benannt.

14. Zwar ist ein mit einem realen Datensatz gewonnenes Ergebnis in diesem Zusammnhang gegenüber den Ergebnissen einer Simulation nur von untergeordneter Bedeutung, aber ein bei HAITOVSKY (1966:40) erwähntes und in der Literatur nicht zitiertes Beispiel ist dennoch interessant. Im realen Datensatz zeigten alle von HAITOVSKY getesteten Verfahren bessere Ergebnisse als LISTWISE: Das Ergebnis steht im Widerspruch zu HAITOVSKY's eigener Simulation. Ebenso bemerkenswert ist, daß die Ansätze zur Simulation systematischer Ausfälle bei HAITOVSKY in der Literatur nicht weiter verfolgt wurden.

15. Drei der neun Matrizen besaßen sehr kleine Determinanten (.001,.002, .04E-8), in den 3 Matrizen der 10-Variablenbedingung existierten 2-mal nur je ein Eigenwert größer 1.0.

16. GLEASON/STAELIN bezeichnen COVPAIR als GLASSER, vgl. Kap. 5.

17. Demzufolge ist die Fußnote 8 bei ANDERSON/BASILEVSKY/HUM (1983:468) falsch: Iterative Hauptkomponentenmethoden wurden von TIMM (1970:436) vorgeschlagen und von GLEASON/STAELIN (1975:238- 239) sowohl vorgeschlagen als auch getestet.

18. GLEASON/STAELIN geben für hohe Korrelationen eine mittlere Abweichung der geschätzten Daten von 3% der Standardabweichungen.

19. Zur Anwendung des EM-Algorithmus vgl. auch PREGIBON (1976). Er verglich u.a. EM und BUCK, "zero spike" sowie LISTWISE an einem realen Datensatz, der nicht-normalverteilte Variablen enthielt und mit teilweise systematischen MD-Mustern versehen wurde. Eine Verbesserung von BUCK durch Iteration war nicht zu beobachten (PREGIBON 1976:95). Die Arbeit ist allerdings explizit als "case study" und nicht als Simulation benannt und kann daher nicht verallgemeinert werden.

20. Die Autoren legen vielmehr die Benutzung des EM-Algorithmus nahe. Weiterhin betonen sie, daß das Verhalten von PAIRWISE und des EM-Algorithmus für große Stichproben und mittlere Korrelationen nicht bekannt sei. Es sei erwähnt, daß sie HAITOVSKY den Test von PAIRWISE unterstellen, die Arbeit von HAITOVSKY bezieht sich aber auf COVPAIR.

21. Hier soll noch die Arbeit von HESTER (1976) erwähnt werden, die eine reine Monte-Carlo-Studie mit Variation der Verteilungsform darstellt. HESTER untersuchte COVPAIR, Mittelwertersetzung und eine Regressionsersetzung mit zusätzlicher Variablen in normalen und lognormalen Verteilungen und konstanten, sehr hohen MD- Anteilen in einem Fünf-Variablen-Regressionsmodell. Erwähnt werden soll nur, daß COVPAIR nicht sehr robust in lognormalen Verteilungen erschien (HESTER 1976:17) und der Anteil nicht positiv definiter Matrizen mit wachsendem n zurückgeht (HESTER 1976:15). Ebenso soll eine weitere Simulationsstudie von HEIBERGER (1977,1978) mit p=4, n=20 und und einem MD-Anteil von 0.1 unter MAR zur Eigenwertstruktur von mit COVPAIR ("pairwise" bei HEIBERGER) und LISTWISE geschätzten Kovarianzmatrizen erwähnt werden.

22 Z.B. FORD (1976) untersuchte Ratio-Schätzer, Regressionsschätzer, "Cellmean" und Hot-Deck Verfahren mit einer nicht näher beschriebenen Simulation in ihrem Einfluß auf univariate Statistiken, ebenso COX/FOLSOM (1978). Sie verglichen an einem realen Datensatz den Einfluß von Gewichtungsverfahren und Hot-Deck. Vgl. ferner die Arbeit von SCHIEBER (1978), der Hot-Deck-Verfahren und zwei "two-stage" Prozeduren für Einkommensdaten (allerdings unter simulierten MAR) verglich. Weitere Beispiele für empirische Arbeiten zum Einfluß von Hot-Deck-Verfahren auf univariate Statistiken stellen RIZVI (1983) und CHAPMAN (1983) dar. FORD (1983) bietet einen Überblick über eine Reihe von Arbeiten. Die Verfasser stimmen darin überein, daß der Einfluß der Hot-Deck- Verfahren auf Kovarianzstrukturen weitgehend unbekannt ist (RIZVI 1983:331, FORD 1983:207). In Hinsicht auf univariate Statistiken stellt FORD (1983:206-207) für die von ihm besprochenen Vergleichsstudien fest, daß sie kein einheitliches Ergebnis dahingehend zeigen, daß ein Verfahren den anderen überlegen sei.

23. Vgl. FORD/KLEWENO/TORTORA (1980). Sie verglichen in einer bedingt als Simulation anzusehenden Arbeit Ratioschätzer, ESTMAT, Zero-Spike und PRINCOMP an einem realen Datensatz. Sie zogen zwei nicht-MAR-Löschungsmuster mit ein und verglichen u.a. die Kovarianzschätzungen. ESTMAT zeigte in ihrer Studie die schlechtesten Ergebnisse (vgl. zur Anwendung von ESTMAT: HUDDLESTON/ HOCKING 1978). Die Arbeit ist allerdings, wie die Autoren selbst betonen (vgl. S.256), u.a. auf Grund der extrem schiefen Verteilungen nicht verallgemeinerbar. Zur Kritik vgl. CHILTON (1980).

24. Die Autoren geben nicht explizit ein MAR-Löschmuster an.

25. Dieses Ergebnis wird nicht explizit erwähnt, vgl. VACEK/ ASHIKAGA (1980:329-330; Tab. 4-6).

26. Damit stellt KAISER erstmals eine gewisse Kontinuität in dieser Forschungsrichtung her. Die Simulation von KAISER ist ohnehin im Design und der Dokumentation beispielhaft.

27. Die sinnvolle Anwendbarkeit des Likelihood-Ratio-Tests kann anders beurteilt werden.

28. VAN GUILDER/AZEN dokumentieren lediglich C1 und C2 (ohne Details, keine Formeln). LITTLE (1981:62) bezieht sich in seiner Diskussion auf drei Maße, wobei offensichtlich zwei "Parametermaße" und ein "Ersetzungsmaß" benutzt wurden.

29. Das Maße C1 bezieht sich auf die LISTWISE- und ALLVALUE- Matrizen; die Berechnung der Kovarianzmatrizen nach den Regressionen hätte andere Ergebnisse gezeigt. Es bleibt festzuhalten, daß die Kovarianzmatrizen der vervollständigten Datensätze nicht mit den beiden Ausgangsmatrizen übereinstimmen.

30. Die Zahl unabhängiger Replikationen läßt sich dem Text nicht entnehmen, die Rede ist von 50 Datensätze einer "'onetime generation'".

31. Die Autoren berichten nur vereinzelt von Signifikanztests.

32. Einzelergebnisse dieser Art werden nur für EM berichtet.

33. Z.B. weist LITTLE (1981:61) darauf hin, daß die Ausfallmechanismen in der Studie von VAN GUILDER/AZEN dieselben Regressionskoeffizienten in den vollständigen wie in den unvollständigen Fällen bedingen. LITTLE glaubt, daß unter Bedingungen, bei denen diese Annahme nicht gegeben ist, der EM-Algorithmus eine deutliche Verbesserung gegenüber LISTWISE darstellt.

34. Dieser Punkt soll noch einmal an der Literaturübersicht von ANDERSON/BASILEVSKY/HUM (1983) demonstriert werden. Ihre Ablehnung der PAIRWISE Methode begründen sie mit den Simulationsstudien von DAGENAIS (1971,1972) und HAITOVSKY (1967). HAITOVSKY wurde bereits ausführlich kritisiert, insbesondere auf Grund der von ihm benutzten Kovarianzmatrizen. DAGENAIS (1971) läßt exakt einen fehlenden Wert pro Fall zu, DAGENAIS (1972) beschäftigt sich mit dem trivariaten Fall, bei dem ebenfalls maximal ein fehlender Wert pro Fall zugelassen ist.

7. Design der Simulationstudie

In Kapitel 6 wurde gezeigt, daß für einen Teil der MD-Ersetzungsverfahren bisher keine empirischen Ergebnisse zu ihrer Leistungsfähigkeit vorliegen und mit wenigen Ausnahmen lediglich Arbeiten unter der MAR-Annahme existieren. Um einerseits die Beurteilung der MD-Verfahren untereinander zu erleichtern und andererseits Kriterien zur Beurteilung des prinzipiellen Herangehens an MD-Probleme mit Ersetzungstechniken zu gewinnen, wurde eine umfangreichere Simulationsstudie als notwendig angesehen. Das vorliegende Kapitel beschreibt die wesentlichen Designentscheidungen bei der Konzeption der durchgeführten Simulation.

7.1 Ziel der Studie

Die Studie sollte die Unterschiede zwischen einer Reihe von MD-Ersetzungstechniken in Hinsicht auf die Verzerrung von Korrelationsmatrizen und der geschätzten Daten bei verschiedenen Ausfallmechanismen ermitteln. Die Anlage der Simulation erfolgte mit einem praktischen Interesse: Unter welchen Bedingungen erbringt welches Verfahren die geringsten Verzerrungen? Kann von den Simulationen und analytischen Ergebnissen unter der MAR-Annahme auf das Verhalten der Verfahren unter in den Sozialwissenschaften realistischeren Ausfallmodellen verallgemeinert werden? Wie reagieren die Verfahren auf nicht durch eine inhaltliche Theorie geleitete Anwendungen[1/]?

Gegenstand der Simulation

Die Fragestellung wurde auf zwei Aspekte eingeengt: Erstens auf eine mögliche Aussage über die Güte der Schätzung der fehlenden Werte und zweitens auf die Frage nach dem Erhalt der "Korrelationsstruktur". Die Wahl von Korrelationsmatrizen als Erhebungseinheiten erfolgte aufgrund der Anschaulichkeit und der Bedeutung der Korrelationsmatrizen als

Ausgangsmaterial der meisten multivariaten Verfahren. Darüber hinaus hätten alle speziellen statistischen Techniken spezielle abhängige Variablen erfordert, z.B. Maße der Ähnlichkeit der extrahierten Faktorenstruktur, falls das Interesse sich auf die Auswirkungen fehlender Werte in der Faktorenanalyse gerichtet hätte. Für die Verwendung von Korrelationsmatrizen spricht weiterhin deren Verwendung in der multiplen Regression (und damit in der Pfadanalyse), der Faktorenanalyse sowie in LISREL-Modellen. Daher erfolgte die Simulation lediglich mit dem Ausgangsmaterial der Verfahren unter der Annahme, daß ein aus verzerrten Korrelationen berechnetes Modell nicht korrekt sein kann, solange nicht die Ursache der Verzerrung mit in das Modell aufgenommen wird.

7.2 Prinzip der Simulation

Die Verwendung realer Datensätze zur Ermittlung der Eigenschaften statistischer Verfahren schied aus einer Reihe von Gründen aus, da die das Verhalten der Verfahren bestimmenden Parameter entweder unbekannt sind oder in nicht ausreichendem Maß variiert werden können. Damit verblieb als einziges - zumindest prinzipiell verallgemeinerbares - experimentelles Verfahren die Simulation.

In dieser Simulation wurde entsprechend den noch darzustellenden Vorüberlegungen mit Hilfe eines Simulationsprogramms zunächst ein simulierter Datensatz erzeugt, der jeweils einem realen Datensatz, wie er in der Forschungspraxis vorkommen könnte, entsprechen sollte, dessen "Eigenschaften" aber vollständig bekannt waren, so z.B. die Populationskorrelationen. Für diesen Datensatz wurde dann die Stichprobenkorrelation berechnet und anschließend - entsprechend den zu simulierenden Ausfallmodellen - ein Teil der Daten als "fehlend" ausgewiesen. Auf diesen unvollständigen Datensatz wurden dann eine Reihe unterschiedlicher MD-Ersetzungsverfahren angewandt und jeweils für jede Ersetzungstechnik die Abweichungen der geschätzten Daten von den bekannten "wahren" Daten und die Abweichungen der aus den geschätzten Daten berech-

neten Korrelationsmatrizen von den bekannten "wahren" Stichprobenkorrelationsmatrizen berechnet.

Ingesamt wurden vier Faktoren, von denen ein Einfluß auf die Leistungsfähigkeit der MD-Ersetzungstechniken erwartet wurde, systematisch variiert:
- Umfang der Stichprobe,
- mittlere Interkorrelation in der Population,
- Anteil fehlender Werte pro Variable in der Stichprobe
- und der Ausfallmechanismus.

Jeder dieser Faktoren besaß drei Faktorstufen, so daß bei einem vollständigen faktoriellen Design (3 * 3 * 3 * 3 =) 81 verschiedene Bedingungen entstanden. Für jede der Bedingungen wurden insgesamt 10 voneinander unabhängige Datensätze erzeugt, um die Stichprobenvariabilität zu simulieren. Damit waren also 810 Datensätze, die jeweils einem realen Datensatz in der Forschungspraxis entsprechen, Grundlage der Simulation.

7.3 Beschreibung der simulierten Datensätze

Um die oben formulierten Fragen zu beantworten, sollte die Simulation mit Datensätzen arbeiten, die den in der Forschungspraxis vorkommenden Datensätzen in möglichst vielen Beziehungen (Fallzahl, Variablenzahl, Interkorrelation, Meßniveau und Verteilung) entsprachen.

Die Datensätze umfassten 100, 300 und 500 Fälle; mit der Ausnahme "nationaler Surveys" dürften diese Zahlen den Bereich der in der analytisch orientierten empirischen Sozialforschung vorkommenden Fallzahlen abdecken[2/].

Die Zahl der Variablen wurde rein willkürlich auf 10 festgelegt. Zwar wäre eine höhere Variablenzahl wünschenswert, um den z.B. bei Standardanwendungen der Faktorenanalyse üblichen umfangreicheren Korrelationsmatrizen näher zu

kommen, aber erstens wurde von einer weiteren Erhöhung der Variablenzahl kein starker Effekt auf die Verbesserung der Imputationschätzungen erwartet und zweitens hätte die Erhöhung der Variablenzahl dramatische Auswirkungen auf die Rechenzeit gehabt.

Fast alle der im Kapitel 5 behandelten Ersetzungsverfahren setzen multivariat-normal verteilte metrische Daten voraus. Die vorliegenden Simulationsstudien arbeiten mit wenigen Ausnahmen ausschließlich mit solchen Datensätzen. Da die hier durchgeführte Simulation sich jedoch stärker auf die - nicht durch eine inhaltliche Theorie geleitete - Anwendung der Verfahren auf im Rahmen sozialwissenschaftlicher Forschung vorkommende typische Datensätze beziehen sollte, erschien es notwendig, die simulierten Datensätze auch in Hinsicht auf die Verteilungsform und die "Stetigkeit" der Messung an die in der Praxis vorherrschenden Datensätze anzunähern. Da die Variation der Verteilungsform allerdings die Einführung eines vollständig neuen Faktors in das Design bedingt hätte, wurde auf die Variation dieses Faktors verzichtet[3/]. Die Variablen der Datensätze sind Realisationen von in der Population multivariatnormal verteilten Variablen, die mit einem Zufallszahlengenerator mit vorgegebener Kovarianzstruktur erzeugt wurden[4/].

Durch die Festlegung der Korrelationsmatrizen als Erhebungseinheit war das Meßniveau als metrisch vorgegeben. Es verblieb aber das Problem, daß im Regelfall zumindest die Eingabedaten keine kontinuierlichen Variablen darstellen[5/]. Weiterhin liefern eine Reihe der Ersetzungsverfahren kontinuierliche Werte als Substitute.

In diesem Zusammenhang ist nur ein Aspekt des Problems wesentlich: die Verzerrung eines Pearson-Korrelationskoeffizienten zwischen zwei "eigentlich" metrischen Variablen durch eine diskrete Werte erbringende Messung [6/]. Die zu diesem Problem vorliegenden Simulationsstudien kommen übereinstimmend zu dem Ergebnis, daß unter der Voraussetzung ursprünglich normalverteilter Variablen und einer großen Zahl von Kategorien die entstehenden Verzerrungen des Korrelationskoeffizienten gering sind (JOHNSON/ CREECH 1983: 406;

BOLLEN/BARB 1981:237; O'BRIEN 1979:855). Ebenso übereinstimmend empfehlen die Autoren den Gebrauch von mindestens fünf Kategorien (BOLLEN/BARB 1981:236; JOHNSON/CREECH 1983:406). Für die Simulation wurden daher Variablen mit sieben Kategorien verwendet[7/].

Da die Pseudo-Zufallszahlengeneratoren der verwendeten FORTRAN- Bibliothek kontinuierliche Werte liefern, wurden die Variablen transformiert. Die Gruppierung erfolgte nach dem von BOLLEN/ BARB (1981:233-234) vorgeschlagenen Verfahren: Innerhalb des +/- 3-sigma Bereiches lagen mehr als 99.7% der Fälle. Dieser 6- Einheiten Bereich wurde durch die Anzahl der Kategorien dividiert, um die Zahl der Standardeinheiten pro Kategorie zu bestimmen. Der mittlere Bereich wurde symmetrisch um Null gelegt.

7.4 Korrelationsstruktur

Wie bereits erwähnt, sollten Datensätze mit je zehn Variablen simuliert werden. Diese zehn Variablen, auf denen fehlende Werte möglich sein sollten, entsprechen den Variablen einer an dem jeweiligen Datensatz zu überprüfenden Kerntheorie und werden im folgenden als "interne Variablen" bezeichnet.

Zusätzlich zu diesen zehn internen Variablen wurden drei weitere Variablen erzeugt, auf denen keine fehlenden Werte generiert wurden. Diese drei zusätzlichen Variablen werden im folgenden als "externe Variablen" bezeichnet. Sie sollten übliche demographische Variablen, wie "Geschlecht" und "Alter", simulieren. Der Sinn der Simulation dieser drei zusätzlichen Variablen liegt zum einen in der Notwendigkeit begründet, auch auf allen internen Variablen der Kerntheorie fehlende Fälle zu schätzen[8/] und zum anderen benötigen zwei der untersuchten MD-Ersetzungsverfahren Klassifikationsvariablen, die sonst nur über zusätzliche Berechnungen hätten ermittelt werden können. Für die externen Variablen wurde angenommen, daß auf ihnen keine fehlenden Werte existieren. Diese idealisierende Annahme wurde in Kauf genommen, weil zumindest für

adressenbasierte Surveys für die im Adressenmaterial steckende Information bzw. für Interviewerschätzungen keine prinzipielle Erhebbarkeitsbarriere existiert.

Schließlich erforderte die Simulation eines der drei Ausfallmechanimen die Erzeugung einer weiteren Variablen, die aber lediglich bei der Generierung der fehlenden Werte benutzt wurde und nicht Bestandteil der Datensätze war. Auf diese Variable wird bei der Darstellung der Ausfallmechanismen zurückgekommen.

Die Analysen der Abweichungen der geschätzten Korrelationen beziehen sich auf die (10 * 10)-Korrelationsmatrizen der internen Variablen mit ihren jeweils 45 Korrelationskoeffizienten. Sämtliche Matrizen besaßen für die internen Variablen eine vollständig homogene Interkorrelationsstruktur. Diese Struktur wurde wegen ihrer Einfachheit und der Nähe zu Interitemkorrelationsmatrizen gewählt[9/]. Die mittlere Interkorrelation (CM) nahm die Werte 0.2, 0.4 und 0.6 an und dürfte damit den in der Regel in der Praxis vorkommenden Größenordnungen entsprechen.

Tabelle 7-1: Vorgegebene Populationskorrelationsstruktur

	externe Variablen			interne Variablen				Selektions-variable
	V1	V2	V3	V4	V5	...	V13	(V14)
V1		.2	.2	.15	.15		.15	.15
V2			.2	.15	.15		.15	.15
V3				.15	.15		.15	.15

Tabelle 7-1 zeigt die vorgegebene Struktur, die allen Matrizen zugrunde lag. Die externen Variablen weisen eine niedrige homogene Interkorrelation von 0.2 auf und korrelieren mit den internen Variablen jeweils mit 0.15.

Die Setzung dieser Korrelationen hat einen theoretisch interessanten Hintergrund. Unter der Voraussetzung, daß die Globalvariablen nur deshalb in empirischen Untersuchungen mit kognitiven Variablen Korrelationen aufweisen, weil in der Regel typische Kombinationen von Merkmalsausprägungen vermittelt über jeweils typische handlungsrelevante Mechanismen, z.B. spezielle stabile Motivationen vermittelnde Sozialisationspraktiken, jeweils ähnliche Strukturen hervorbringen (LANGENHEDER 1975:61)[10/], wurden diese Korrelationen konstant niedrig angesetzt.

7.5 Ausfallmodelle

Drei verschiedene Ausfallmechanismen, die jedoch nicht als die "häufigsten" gelten können, wurden simuliert. Neben MAR sollten lediglich zwei andere Mechanismen, die im Rahmen bestimmter soziologischer Fragestellungen plausibel scheinen, zum Vergleich herangezogen werden.

Die Generierung der fehlenden Daten erfolgte für alle Mechanismen unabhängig voneinander; "Mischungen" der Ausfallmechanismen, wie sie in den meisten Datensätzen in der Praxis vorkommen dürften, wurden nicht vorgenommen. In allen drei Ausfallmodellen wurde der Prozentsatz fehlender Werte pro Variable dreistufig variiert: 1%, 5% und 10%[11/]. Die Generierung erfolgte für jede Variable vollständig unabhängig von allen anderen Variablen, so daß auch Fälle mit auf allen internen Variablen fehlenden Werten möglich waren. Die Prozentsätze wurden so gewählt, daß ein direkter Vergleich mit einer Reihe veröffentlichter Simulationsstudien möglich wurde.

7.5.1 Modell 1: Missing at Random (MAR)

Dieses Modell bietet den Bezugspunkt für den Vergleich der Verfahren bei systematischen Ausfallmechanismen. Nach der Definition von MAR (vgl. Kapitel 1.1.1) kann nur der Prozentsatz fehlender Werte innerhalb des Modells variiert werden.

7.5.2 Modell 2 : Hazard-Modell (MAT)

Als zweites Modell wurde ein Hazard-Modell gewählt ("missing above threshold"; MAT). In diesem Modell fehlen Daten dann, wenn eine nur indirekt gemessene Variable einen Schwellenwert überschreitet. Der Ausfallmechanismus sollte nicht deterministisch sein, um den in sozialwissenschaftlichen Datensätzen unrealistischen Fall des "complete truncation" zu vermeiden[12/]. Das realisierte Modell zieht eine Zufallsstichprobe aus denjenigen Fällen des generierten Datensates, die einen in Abhängigkeit vom Anteil fehlender Werte bestimmten Schwellenwert auf der latenten Variablen überschreiten[13/].

In diesem Modell sind offensichtlich zwei Faktoren bedeutsam: Der Anteil fehlender Werte und die Korrelationen der internen Variablen mit der latenten Selektionsvariablen. Der Anteil fehlender Werte wurde wie im MAR-Modell mit 1%, 5% und 10% vorgegeben. Da mit homogenen Korrelationsmatrizen gearbeitet wurde, sollte das Homogenitätskriterium auch auf die Korrelationen mit der Selektionsvariablen angewandt werden, so daß diese Korrelationen in der Population als gleich groß vorgegeben wurden. Die Höhe dieser Korrelation wurde jeweils mit einer leicht stärkeren Korrelation als der mittleren Interkorrelation (r = CM + 0.1) angenommen. Diese Festlegung erfolgte, um die Komplexität des Designs nicht weiter zu steigern.

Das durch diese Festlegung entstehende Modell einer homogenen Interkorrelationsstruktur mit einer latenten Variablen schien darüber hinaus einer Reihe von typischen Datensätzen zu entsprechen: Die Interpretation der entstehenden Korrelationsmatrix als Inter-Item-Korrelationsmatrix in Hinsicht auf die latente Selektionsvariable liegt nahe. Das Modell könnte zum Beispiel Ausfallmechanismen bei Leistungstests, bei denen entweder oberhalb eines Leistungslevels durch wahrgenommene Trivialität des Tests oder unterhalb eines Leistungslevels durch Frustration, Ausfälle auftreten, entsprechen[14/].

7.5.3 Modell 3 : Zwei-Sample-Modell (MARC)

Dieses Modell (MARC: Missing at Random within Classes) simuliert Stichproben, deren Elemente aus zwei verschiedenen Populationen stammen. Die Populationen unterscheiden sich sowohl in Hinsicht auf die Art und Höhe des Zusammenhangs der internen Variablen, als auch in ihren Ausfallwahrscheinlichkeiten. Innerhalb der Populationen wird MAR angenommen. Bei diesem Modell sind offensichtlich die Faktoren:

- Differenz der Variablenbeziehungen,
- Differenz der Ausfallwahrscheinlichkeiten,
- Anteilswerte der Sampleteile

von Bedeutung. Die systematische Variation aller drei Faktoren hätte den Rahmen der Arbeit gesprengt[15/]. Daher wurde der Versuch unternommen, ein "worst reasonable model" (DEMPSTER 1978:30) zu konstruieren, also ein Modell, das vermutlich sehr starke Verzerrungen in den zu schätzenden Parametern bedingt. Um die Faktorenzahl zu begrenzen, wurden die Faktoren "Differenz der Anteilswerte" und "Differenz der Variablenbeziehungen" auf zwar extreme, aber noch plausible Werte festgelegt.

Der Extremfall in einem einfachen MARC-Modell wird dann erreicht, wenn zwei gleichgroße Sampleteile vorhanden sind, wobei in einem Sampleteil starke Zusammenhänge und niedrige Ausfallwahrscheinlichkeiten und in dem anderen Sampleteil niedrige Zusammenhänge und hohe Ausfallwahrscheinlichkeiten bestehen. Im realisierten Modell wurden die Sampleanteile mit 0.8 und 0.2 vorgegeben. Der Wert 0.2 schien (vollkommen subjektiv) den größten Anteil eines Samples darzustellen, der im Rahmen einer "routinemäßigen" Datenanalyse eine "faire" Chance hat, in Hinsicht auf sein deviantes Reaktionsmuster "unentdeckt" zu bleiben. Die Chance der Nicht- Entdeckung wurde dadurch simuliert, daß keine Variable der simulierten Datensätze mit der unbekannten Klassifikationsvariable perfekt korreliert, anders gesagt: Die Klassifikationsvariable ist nicht Bestandteil der Datensätze. Das Verhältnis der Ausfallwahrscheinlichkeiten wurde ebenso willkürlich mit 1:10 festgelegt, um ein möglichst stark verzerrendes Modell erhalten. Die MD-Anteile wurden so bestimmt, daß sich im zusammengefassten Sample 1%, 5% und 10% fehlende Daten pro Variable ergaben. Die Differenz der mittleren Interkorrelation wurde indirekt über die variierende Interkorrelation[16/] im Sampleteil 1 bei konstanter 0.1 - Korrelation[17/] im Sampleteil 2 simuliert.

7.6 Missing-Data-Verfahren

Aus den in Kapitel 5 beschriebenen Verfahren wurden zunächst zwölf Verfahren ausgewählt. Die Auswahl beruhte auf zwei Kriterien: Das Verfahren sollte entweder weit verbreitet oder von besonderem theoretischen Interesse sein. Nach einer Reihe von Voruntersuchungen wurden zwei Verfahren ausgeschieden, deren Ergebnisse so weit unter denen der anderen Verfahren lagen, daß eine vollständige Prüfung unangebracht schien (vgl. Anhang C).

Die folgenden Verfahren wurden damit in der Hauptuntersuchung verwendet:

1. - LISTWISE (complete cases)
2. - PAIRWISE (paarweise vorhandene Werte)
3. - MEAN (Mittelwertersetzung)
4. - CELLMEAN (Zellenmittelwertersetzung)
5. - HOT-DECK (Doppelungsverfahren)
6. - REGRESSION (Einfachregression)
7. - HAZARD (Zweifachregression mit MD-Indikator)
8. - DEAR (Hauptkomponentenverfahren)
9. - MULT (Multiple Regression, BUCK-ähnlich)
10. - ILS (iterierte multiple Regression, EM-ähnlich)

Einige Verfahren bedürfen kurzer Erläuterungen. CELLMEAN verwendet für jede der internen Variablen die Zellenmittelwerte der vorhandenen Fälle einer Variablen aus einer mit V1 und V2 gebildeten Crossbreaktabelle. HOT-DECK benutzt für jede Variable V1 und V2 als "Imputationclassvariable"[18/]. REGRESSION benutzt die jeweils höchstkorrelierende Variable zur Vorhersage. HAZARD verwendet neben der von REGRESSION benutzten Variablen als zusätzliche Variable die Anzahl fehlender Werte als Prädiktorvariable. HAZARD ist das einzige Verfahren, daß bisher nicht in Simulationen getestet wurde. DEAR ersetzt fehlende Werte durch Werte der ersten Hauptkomponente, die aus der PAIRWISE-Matrix berechnet wurde. MULT stellt eine multiple Regression mit allen verfügbaren Werten jedes Falles dar, wobei die LISTWISE-Matrix den Ausgangspunkt bildete. ILS benutzt die Resultate von MULT als Startwerte und wiederholt vollständige multiple Regressionen, bis ein Konvergenzkriterium erreicht wird[19/]. Die meisten Verfahren sind in Standardprogrammpaketen entweder vorhanden oder realisierbar (vgl. Anhang E).

7.7 Abhängige Variablen

Zwei Aspekte der Leistungsfähigkeit der MD-Techniken sollten untersucht werden: Zum einen die Güte der Schätzung der Korrelationsmatrizen und zum anderen die Güte der Schätzung der fehlenden Daten. Für beide Aspekte wurden mehrere Variablen definiert.

7.7.1 Maße der Güte der Schätzung der Korrelationen

> "One should be sceptical about any measure claiming to summarize the r(r+1)/2 paired differences between Σ and Σ' but due to the lack of a metric when comparing matrices some sacrifices must be made."
>
> PREGIBON (1976:88)

Zentral für die Bewertung der Leistungsfähigkeit der Verfahren war die Frage nach der "Ähnlichkeit" der jeweils geschätzten mit der "wahren" Korrelationsmatrix. Damit verschob sich das Problem auf die Definition der "Ähnlichkeit" zweier Korrelationsmatrizen[20/]. Da nicht die Verteilungsfunktion eines Schätzers untersucht werden sollte, wurde "Ähnlichkeit" in Hinsicht auf aus den Korrelationsmatrizen zu ziehende Konsequenzen aufgefaßt: Nicht die Güte der Reproduktion der einzelnen Koeffizienten, sondern die Güte der Reproduktion der Struktur der Matrix war von Interesse. Damit ergaben sich drei Möglichkeiten: Statistische Tests der Kovarianzmatrizen, Vergleich der Eigenwerte bzw. Determinanten und "Ähnlichkeitsmaße".

Die naheliegende Verwendung eines statistischen Tests zur Überprüfung der Übereinstimmung der geschätzten mit der "wahren" Korrelationsmatrix schied auf Grund der Abhängigkeit der Stichproben aus. Ein summarischer Test für auf diese Art "verbundene" Stichproben scheint bisher nicht bekannt zu sein[21/].

Im Anschluß an TIMM (1969:27; 1970:426) wurde auf die Verwendung der Eigenwerte bzw. der Determinanten verzichtet. TIMM (1970:426) erwähnt als Maße die Determinante der Differenzenmatrix der "wahren" Korrelationsmatrix (P) und der geschätzten Matrix (R) und das Ratio der Determinanten von R und P. Da die Determinanten gleich dem Produkt der Eigenwerte sind, lehnt TIMM diese beiden Maße wegen ihrer Verzerrung durch kleine Eigenwerte ab[22/]. Damit verbleibt die Möglichkeit der Konstruktion von "Ähnlichkeitsmaßen". Die Definition der Ähnlichkeit zweier Korrelations matrizen müßte auf die Resultate der jeweils von diesen Matrizen ausgehenden multivariaten Verfahren bezogen werden[23/].

Gegen dieses Vorgehen sprach sowohl die notwendige Vergleichbarkeit mit früheren Untersuchungen als auch die Schwierigkeit der Konstruktion von Ähnlichkeitsmaßen für die Ergebnisse der jeweiligen Verfahren. Da dieses Vorgehen eine Problemverschiebung dargestellt hätte, wurden im Rahmen der nur explorativ gedachten Simulation die Vergleiche auf der Ebene der Korrelationsmatrizen durchgeführt[24/].

Um die am Rande interessierende Reproduktion der einzelnen Koeffizienten zu messen, wurde die mittlere Abweichung der Koeffizienten als

$$\text{7-1} \qquad \text{RDEV} = \frac{\sum_{i=1}^{p-1} \sum_{j=i+1}^{p} RT(i,j) - R(i,j)}{p(1-p)}$$

berechnet, um ein Vorzeichen für die Richtung der Verzerrung zu erhalten, wobei p der Variablenzahl, RT der "wahren" und R der geschätzten Matrix entspricht. TIMM (1970:426) verwendet die euklidische Norm der Differenzenmatrix : Tr ((RT-R) (RT-R)'); eine additive Funktion der Eigenwerte, die der Summe der quadrierten Abweichung entspricht; GLEASON/STAELIN (1975:242) verwenden eine leichte Modifikation, die "root mean square deviation" der Nichtdiagonalelemente. In der Simulation wurde dieses Maß als

$$7\text{-}2 \qquad RSDEV = \frac{\sum_{i=1}^{p-1}\sum_{j=i+1}^{p}\left(RT(i,j)-R(i,j)\right)^2}{p(1-p)} RSDEV$$

berechnet. Die durch die Wurzeltransformation von RSDEV gebildete Variable WRDEV entspricht dem "root mean square residual". Der "overall deviation index" von KIM/CURRY (1977:224) ist hiermit identisch.

Als drittes Ähnlichkeitsmaß der "Struktur" der Matrix wurde (wegen seiner unmittelbaren Anschaulichkeit) der Korrelationskoeffizient (MKS) der beiden als Datenvektoren aufgefaßten Korrelationsmatrizen berechnet[25/]. Da der Korrelationskoeffizient invariant gegenüber linearen Transformationen ist, bezieht sich dieses "Ähnlichkeitsmaß" nicht auf die Reproduktion einzelner Koeffizienten, sondern auf den Erhalt der Struktur[26/] genauer: auf den Erhalt relativer Größenordnungen[27/].

7.7.2 Maße der Güte der Schätzung fehlender Werte

Neben der Genauigkeit der Schätzung der Korrelationsmatrix sollte die Güte der Schätzung der fehlenden Werte untersucht werden. Mit Ausnahme des Hot-Decks liefern die Verfahren keine ganzzahligen Werte als Schätzung, sondern "reals". Da die Eingabewerte ganzzahlig waren, schien die Kategorisierung der geschätzten Werte sinnvoll. Da sich eines der Hauptargumente für die Verwendung von MD-Ersetzungsverfahren auf die leichte Handhabung vollständiger Datensätze bezieht, kann die Entscheidung zur Kategorisierung auch hierüber gestützt werden. Von den vorhandenen Simulationsstudien nehmen lediglich GLEASON/STAELIN (1975:244) explizit eine Kategorisierung vor, sie halten die Entscheidung für willkürlich. Jeder geschätzte Wert wurde daher zur nächsten ganzen Zahl gerundet, bevor er für die Berechnung der Ähnlichkeitsmaße der Daten benutzt wurde (im Gegensatz dazu wurden die für

die Berechnung der Korrelationsmatrix benutzten geschätzten Werte nicht gerundet).

GLEASON/STAELIN (1975:243) benutzen das "root-mean-square standardized residual", bei dem die Abweichungen mit der Varianz der Variablen standardisiert werden. Verzichtet man auf die Standardisierung mit der Varianz, gelangt man zur "root mean square deviation", die zum Beispiel von KALTON/SANTOS (1983:88) verwendet wurde. Das Maß wurde in der vorgelegten Studie als

$$\text{7-3} \qquad \text{ERSD} = \sqrt{\frac{\sum\sum\left(\text{DT}(i,j)-\text{D}(i,j)\right)^2}{\text{M}}}$$

berechnet, wobei DT die vollständige "wahre" Datenmatrix, D die geschätzte Datenmatrix und M die Anzahl fehlender Werte in der Matrix bezeichnet. Daneben ist die mittlere Abweichung

$$\text{7-4} \qquad \text{EMD} = \frac{\sum \text{DT}(i,j)-\text{D}(i,j)}{\text{M}}$$

von Interesse. Auf Grund der vermuteten Verteilung der Fehler wurde auch die mittlere absolute Abweichung

$$\text{7-5} \qquad \text{EMAD} = \frac{\sum |\,\text{DT}(i,j)-\text{D}(i,j)\,|}{\text{M}}$$

berechnet. Auch dieses Maß findet sich bei KALTON/SANTOS (1983: 88). Da die Werte gruppiert wurden, kann auch die Frage nach dem Anteil "richtig" geschätzter Werte beantwortet werden. Hierzu wurde die Variable HIT berechnet,

wobei HIT der durch die Anzahl fehlender Werte dividierten Anzahl der "exakten" Schätzungen entspricht.

7.8 Zusammenfassung des Designs

Nach den dargestellten Überlegungen ergab sich folgendes fünffaktorielles Design:

1 - MTYPE: Ausfallmechanismus (MAR, MAT, MARC)
2 - N : Fallzahl (100, 300, 500)
3 - CM : Interkorrelation (.2, .4, .6)
4 - MPER : Anteil fehlender Werte (1%, 5%, 10%)
5 - IPROG: MD-Verfahren (LIST,PAIR,MEAN,CELL,DECK,REGR,HAZD, DEAR,MULT,ILS)

Das Simulationsprogramm erzeugte für jede Kombination der Faktoren 1-4 ("Zellen") einen Datensatz, der dann den 10 bzw. 6 "Behandlungen", den MD-Verfahren, unterworfen wurde. Für jede Zelle des Designs wurden insgesamt 10 voneinander vollständig unabhängige Datensätze erzeugt. Das Design stellt damit ein fünffaktorielles (3 * 3 * 3 * 3 * 10) balanciertes Design mit wiederholter Messung nur des letzten Faktors dar. Mit den 81 Zellen und 10 Wiederholungen wurden insgesamt 810 vollständig voneinander unabhängige Datensätze erzeugt. Die die Korrelationsmatrizen betreffenden abhängigen Variablen wurden für alle 10 Verfahren berechnet; die die geschätzten Daten betreffenden abhängigen Variablen dagegen nur für die Verfahren 5-10, da weder PAIRWISE noch LISTWISE Ersetzungen vornehmen und der Einsatz von MEAN und CELLMEAN als MD-Ersetzungsverfahren zwar möglich, aber nicht sinnvoll schien. Für die die geschätzten Daten betreffenden Variablen reduzierte sich das Design auf ein (3 * 3 * 3 * 3 * 6) - Design.

Anmerkungen zu Kapitel 7

1. Die Entwicklung einer empirischen Verteilungsfunktion der resultierenden Schätzer war nicht beabsichtigt. Solche Verteilungsfunktionen sind mit vertretbarem Aufwand nur unter der MAR-Annahme simulierbar, da ansonsten die Zahl der zu berücksichtigenden Parameter extreme Rechenzeiten erfordert, solange nicht zu speziellen "variance reduction techniques" gegriffen wird. Die Verwendung der "swindle"-Techniken hätte den Rahmen der Arbeit gesprengt, vgl. KLEIJNEN (1974).

2. Die SPSS Usage-Reports zeigen für die Jahre 1982-1984 in den Rechenzentren Köln und Bochum, daß ca. 90% der Datensätze weniger als 200 Fälle umfassten.

3. Mit Ausnahme der Arbeiten zu fehlenden Einkommensdaten existiert bisher zur Schätzung metrischer, nicht normalverteilter Variablen keine größere Simulationsstudie.

4. Einzelheiten finden sich im Anhang A.

5. Das hierdurch entstehende Problem führt unmittelbar auf die "weak measurement vs. strong statistics" - Debatte Ende der sechziger Jahre, vgl. die Arbeiten in dem Sammelband von LIEBERMANN (1971).

6. Eine Klassifikation der entstehenden Fehler stammt von O'BRIEN, zitiert bei JOHNSON/CREECH (1983:399).

7. Dies betrifft die Variablen, auf denen MD auftreten konnten. Eine der drei "demographischen" Variablen wurde dichotomisiert (V1), die beiden anderen Variablen wurden einer Gruppierung mit fünf Kategorien unterworfen.

8. I.d.R. werden Simulationsstudien so programmiert, daß solche Fälle nicht auftreten können (so z.B. DAGENAIS 1973:323: "...each incomplete observation involved only one missing value."). Um realitätsnah zu bleiben, wurden auch solche Fälle in die Analysen aufgenommen.

9. Vergleichende Untersuchungen mit anderen Korrelationsstrukturen fehlen bisher, die Implikationen dieser Entscheidung sind unklar.

10. Hier zeigt sich eine weitere Problematik der Bestimmung der Imputationklassen: Je höher der Grad gesellschaftlicher Differenzierung wird, desto differenzierter müßten die Imputationklassen werden.

11. Der Anteil wurde exakt bestimmt, siehe Anhang A-1.

12. Falls der Ausfall deterministisch erfolgen würde, entstünde ein multivariates Restriction-of-Range-Problem. Wie im Kapitel 4.1 gezeigt wurde, stellen die beiden Annahmen der complete truncation und die Bekanntheit der Selektionsvariablen die Hauptprobleme der Restriction-of-Range-Korrekturen dar. Aus diesem Grund wurden keine solchen Korrekturen vorgenommen.

13.Einzelheiten finden sich im Anhang A.

14. Es wird nicht behauptet, daß ein derart simples Modell die erwähnten Prozesse hinreichend genau abbilden kann. Hier deutet sich die Komplexität der notwendigen inhaltlichen Ausfall-theorien an.

15. Die Fragestellung bezog sich nicht auf die Entwicklung von Ausfallmodellen und deren Auswirkungen, sondern auf den Vergleich zu der Reaktionskurven der Verfahren unter MAR mit anderen Modellen.

16. Hier vermischen sich leider zwei Faktoren: Die Differenz der beiden Sampleteile und der "Informationsgehalt" der jeweiligen Korrelationsmatrizen, der direkt auf die Fähigkeit zur Schätzung fehlender Werte wirkt.

17. Die Null-Korrelation wurde vermieden, um die Wahrscheinlichkeit der Benutzung der externen Variablen zur Vorhersage zu vermindern.

18. Die Benutzung der durch V1 und V2 gebildeten Imputationklassen für CELLMEAN und HOT-DECK sollte die "mechanische" Anwendung der Verfahren simulieren. Daß die Verfahren bei sorgfältiger, theoretisch fundierter Wahl der Imputationklassen wesentlich bessere Ergebnisse zeigen könnten, ist möglich.

19. Einzelheiten finden sich in den Anhängen A und B.

20. Ein vergleichbares Problem stellt sich bei der Beurteilung von LISREL-Modellanpassungen. LISREL VI bietet hier drei verschiedene Indices an:
- eine Chi-Quadrat basierte Testgröße für die ML-Schätzung,
- einen "goodness-of-fit-index" (GFI),
- das "root-mean-square-residual" (RMR).

Die Chi-Quadrat-Statistik wird aus dem Minimum der Likelihood- Funktion berechnet und läßt sich aus dem Likelihood-Ratio-Test herleiten. Der GFI ist eine standardisierte Form der Abweichungsquadrate, dessen Verteilungsfunktion unbekannt ist. RMR entspricht exakt der Definition des hier verwendeten Index WRDEV (s.w.u.), vgl. DWYER (1983:242, 250-252) sowie JÖRESKOG/SÖRBOM (1984:I-39 - I-41).

21. EVANS/COOLEY/PISERCHIA (1979:473). Die Verwendung des BOX-M- Tests (MARDIA/KENT/BIBBY 1979:140-141) scheitert an der Verbundenheit der Stichproben. Falls man die Nachteile dieses Tests (wenig robust, extreme Sensitivität; vgl. SPECHT/WARREN 1976:51 sowie TABACHNIK/FIDELL 1983:233) berücksichtigt, gibt es eine interessante Anwendungsmöglichkeit für denselben im Zusammenhang mit MD-Imputationverfahren, auf die EVANS/COOLEY/ PISERCHIA hinweisen. Man kann das Sample in zwei Gruppen aufteilen: Die erste Gruppe besteht aus den auf allen Variablen vollständigen Fällen, die zweite Gruppe aus den Fällen, bei denen Ersetzungen vorgenommen wurden. Die getrennten Varianz-Kovarianzmatrizen werden mit dem BOX-M-Test auf Homogenität getestet. Eine ermittelte Inhomogenität deutet dann auf einen Imputationbias hin (EVANS/COOLEY/ PISERCHIA 1976:471). Die Verfasser legen in diesem Fall eine Verwerfung des Ersetzungsverfahrens nahe. Allerdings ist in einem Fall eine andere Schlußfolgerung möglich: Handelt es sich um einen systematischen Ausfall, kann die Inhomogenität der Matrizen auf eine erfolgreiche Ersetzung hinweisen, falls das jeweilige Imputationverfahren in der Lage ist, Nicht-MAR-Ausfälle zu schätzen.

22. Die naheliegende "Strukturgleichheitsprüfung" mit einem auf Eigenwerten basierenden Maß wurde u.a. wegen ihrer unanschaulichen Interpretation abgelehnt. Ein Beispiel für die Verwendung der Eigenwertvektoren für ein Distanzmaß von Kovarianzmatrizen findet sich bei PREGIBON (1976:88).

23. Neben dem Vergleich der Korrelations- bzw. Kovarianzmatrizen basieren die am häufigsten verwendeten Maße auf dem Vergleich von Regressionsparametern.

24. Da in der Literatur die Empfehlungen für die Verwendung von MD-Kompensationsverfahren in der Regel unabhängig davon erfolgt, ob die Ergebnisse durch den Vergleich von Korrelationsmatrizen oder Regressionskoeffizienten gewonnen wurden, scheint Konsens zu bestehen, daß die Wahl der "Erhebungseinheit" nicht kritisch ist.

25. Da es sich um den Bravais-Pearson-Koeffizienten handelt, wird die Formel, die zur Berechnung benutzt wurde, nicht wiedergegeben. Im Programm (Anhang B) findet sich die Berechnung in der Subroutine RESULT vor dem Statement 10.

26. Die in der Literatur zu findenden Matrixkorrelationskoeffizienten sind für andere Verwendungszwecke gedacht, vgl. CRAMER/ NICEWANDER (1979). Zu den notwendigen Bedingungen nach RENYI, die ein Korrelationsmaß zweier Matrizen erfüllen muß, vgl. RAMSAY/ BERGE/STYAN (1984:406).

27. Für die Berechnung der Abweichung zwischen der "wahren" und der geschätzten Korrelationsmatrizen wurden die nicht-transformierten Korrelationskoeffizienten benutzt. Die sonst übliche Fisher-z-Transformation (KENDALL/STUART 1977:419) ist in diesem Fall unangebracht. Diese Transformation erfolgt, um der sich mit wachsendem n nur langsam der Normalverteilung annähernden Verteilung von r (KENDALL/STUART 1977:416) eine approximativ normalverteilte Prüfgröße zu erhalten. Zwar können z-transformierte Korrelationskoeffizienten als intervallskaliert aufgefasst werden (BORTZ 1977:260-261), aber keinesfalls ist (wie die Darstellung bei BORTZ nahelegt) die Mittelung von Korrelationskoeffizienten immer sinnlos (vgl. JÄGER 1974:422-423 zur Mittelung von Korrelationskoeffizenten: Das arithmetische Mittel von Korrelationskoeffizienten ist um so günstiger, je höher n wird und n in den Stichproben annähernd gleich ist). Dies ist insbesondere dann der Fall, wenn die über die Transformation erfolgende höhere Gewichtung hoher Korrelationen (Konfidenzintervalle um hohe Korrelationen sind bezogen auf niedrige Korrelationen kleiner) unerwünscht ist. Ein Beispiel dafür bietet die Meta-Analyse von Korrelationen in verschiedenen Studien (HUNTER/SCHMIDT/JACKSON 1982:41-42). In diesen Fällen führt die Verwendung der Transformation zu einem positiven Bias. Die Verwendung der Transformation würde zur Konfundierung der Nichtlinearität der Transformation mit einem experimentellen Faktor des Designs führen.

8. Auswertung der Simulation

In diesem Kapitel werden die Ergebnisse der in Kapitel 7 beschriebenen Computersimulation dargestellt. Die zentrale Frage richtet sich dabei auf die Verallgemeinerbarkeit der Aussagen über die Leistungsfähigkeit der MD-Ersetzungstechniken unter der MAR-Annahme bei Annahme anderer Ausfallmechanismen.

8.1 Randauszählung und Transformationen

Die Randauszählung zeigt bei fast allen Variablen eine starke Abweichung von der Normalverteilung. Erwartbar war dies insbesondere bei der Variablen MKS. Da MKS durch die Berechnung eines Korrelationskoeffizienten gebildet wurde, schien die Transformation mit der üblichen Fisher-z-Transformation von Korrelationskoeffizienten angebracht: Die transformierte Variable weist eine Schiefe von .327 (S.E. 0.027) gegenüber einer Schiefe von -2.083 der untransformierten Variablen auf.

In Hinsicht auf die häufige Verwendung der untransformierten Wurzel der Spur der Differenzenmatrix der Korrelationskoeffizienten erstaunlich (wenn auch erwartbar) ist die extreme Schiefe der Verteilung (vgl. Abb. 8-1). Bei einer solchen Schiefe stellt der Mittelwert kein brauchbares Maß der Zentraltendenz dar; folglich wurde die Variable logarithmisch transformiert (LRSDEV).

Nach diesen beiden Transformationen weichen alle Variablen immer noch mit $p < .05$ von der Normalverteilung ab, die Plots zeigen aber eine deutliche Näherung, so daß auf weitere Transformationen verzichtet wurde. Obgleich die Variablen nicht normalverteilt sind, sind ihre Abweichungen von der Normalverteilung so gering, daß kaum Effekte auf die Ergebnisse der Analyse erwartet wurden. Für die Analyse wurden somit die in Kapitel 7 definierten Variablen EMD, EMAD, ERSD, HIT, RDEV, ZMKS und LRSDEV benutzt.

Tabelle 8-1: Randauszählung der abhängigen Variablen

VARIABLE	MEAN	STANDARD DEVIATION	RANGE	SKEWNESS
RDEV	0.00549442	0.02862442	0.35008000	1.20199838
RSDEV	0.00148727	0.00280790	0.03855551	4.86460649
WRDEV	0.02997912	0.02426097	0.19382490	1.92237102
LRSDEV	-3.79915716	0.78099822	4.34501874	-0.07860049
MKS	0.91107870	0.08820195	0.73800000	-2.08307650
ZMKS	1.78785099	0.57370700	3.53301996	0.32695034
EMD	0.18275312	0.37906931	3.96667000	1.71410435
EMAD	0.95344989	0.25016384	2.76667000	1.47939937
ERSD	1.24893577	0.27889238	2.72846000	1.29858899
xIT	0.32235819	0.09087816	0.70000000	-0.16162738

Abbildung 8-1: Plot der Verteilung von WRDEV

```
COUNT  MIDPOINT
  179       .00  ****        .
 2253       .01  ******************:***************************
 1876       .02  *************************:**************
 1345       .03  ***************************:*
  865       .04  ******************      .
  575       .05  ************      .
  344       .06  *******     .
  207       .07  ****  .
  141       .08  **:
   91       .09  :*
   65       .10  *
   62       .11  *
   34       .12  *
   23       .13
   17       .14
   10       .15
    4       .16
    1       .17
    4       .18
    3       .19
    1       .20
              I....+....I....+....I....+....I....+....I....+....I....+....I
              0         6        12        18        24        30
```

Abbildung 8-2: Plot der Verteilung nach der LN-Transformation (LRSDEV)

```
COUNT  MIDPOINT
    0     -6.30
    1     -6.05  .
   16     -5.80  *.
   81     -5.55  ***.
  195     -5.30  ******:*
  472     -5.05  ***********:*******
  554     -4.80  ******************:****
  504     -4.55  *********************     .
  692     -4.30  ****************************      .
  945     -4.05  ****************************************.
  901     -3.80  *************************************    .
 1028     -3.55  ****************************************:**
  884     -3.30  ************************************:*
  763     -3.05  ***************************:****
  492     -2.80  ******************:*
  294     -2.55  ***********:
  172     -2.30  ******:
   85     -2.05  ***.
   18     -1.80  *.
    3     -1.55  .
    0     -1.30
               I....+....I....+....I....+....I....+....I....+....I....+....I
               0         3         6         9        12        15
```

Anmerkung: Die Symbole ´.´ und ´:´ geben eine überlagerte Normalverteilung wieder

8.2 Interkorrelation der abhängigen Variablen

Die Korrelationsmatrix der abhängigen Variablen EMD, EMAD, ERSD und HIT sowie das Ergebnis der Faktorenanalyse (Tab. 8-2 und 8-3) zeigt eine hohe Interkorrelation der Variablen; dementsprechend wird mit der Hauptkomponentenmethode nur ein Faktor extrahiert. Die vergleichsweise niedrigen Korrelationen der Variablen EMD und die entsprechend niedrige Kommunalität erfordert die getrennte Analyse für den Variablenblock EMAD, ERSD und HIT einerseits und EMD andererseits.

Tabelle 8-2: Korrelationsmatrix (EMD EMAD ERSD HIT)

	EMD	EMAD	ERSD	HIT
EMD	1.00000			
EMAD	.43627	1.00000		
ERSD	.39094	.97734	1.00000	
HIT	-.34448	-.82526	-.71245	1.00000

Tabelle 8-3: Ergebnis der Faktorenanalyse

VARIABLE	COMMUNALITY	FACTOR	EIGENVALUE	PCT OF VAR
EMD	.32420	1	2.92124	73.0
EMAD	.96053			
ERSD	.87849			
HIT	.75802			

Die Korrelationsmatrix der die Korrelationsmatrizen betreffenden abhängigen Variablen RDEV, LRSDEV und ZMKS zeigt eine vergleichbare Struktur: Nur ein Faktor kann extrahiert werden, die Kommunalität der mittleren Abweichung ist wiederum relativ niedrig, eine getrennte Analyse erscheint erforderlich.

Tabelle 8-4: Korrelationsmatrix (ZMKS LRSDEV RDEV)

	ZMKS	LRSDEV	RDEV
ZMKS	1.00000		
LRSDEV	-.84095	1.00000	
RDEV	-.13233	.29406	1.00000

Tabelle 8-5: Ergebnis der Faktorenanalyse

VARIABLE	COMMUNALITY	FACTOR	EIGENVALUE	PCT OF VAR
ZMKS	.84086	1	1.93857	64.6
LRSDEV	.91385			
RDEV	.18386			

Für die Verfahren 5-10 (HOTDECK, REGRESSION, HAZARD, DEAR, MULT, ILS) wurden alle abhängigen Variablen erhoben. Die durchgeführte Extraktion mit der Hauptkomponentenmethode erbachte zwei Faktoren, die nach obliquer Rotation eine Faktorenkorrelation von - 0.22166 zeigten. Der zweite Faktor wird durch die Spur der Differenzenmatrix der Korrelationsmatrizen und durch den Korrelationskoeffizienten der Korrelationsmatrizen gebildet; der erste Faktor besteht aus allen anderen Variablen, einschließlich der mittleren Abweichung der Korrelationskoeffizienten.

Einige der Analysen wurden daher für die beiden Aspekte (Abweichung der Korrelationsmatrizen einerseits und Abweichung der geschätzten Daten andererseits) getrennt durchgeführt, wobei entsprechend den oben angeführten Ergebnissen innerhalb der beiden Variablenblöcke gegebenenfalls weiter differenziert wurde.

Tabelle 8-6: Vollständige Korrelationsmatrix (Verfahren 5-10)

	EMD	EMAD	ERSD	HIT	ZMKS	LRSDEV
EMD	1.00000					
EMAD	.43627	1.00000				
ERSD	.39094	.97734	1.00000			
HIT	-.34448	-.82526	-.71245	1.00000		
ZMKS	-.00177	-.27682	-.32684	.16537	1.00000	
LRSDEV	.11847	.37013	.40840	-.25482	-.86504	1.00000
RDEV	.47431	.49877	.50633	-.35282	-.16313	.22347

Tabelle 8-7: Ergebnis der Faktorenanalyse

VARIABLE	COMMUNALITY	FACTOR	EIGENVALUE	PCT OF VAR
EMD	.45758	1	3.62797	51.8
EMAD	.91007	2	1.57546	22.5
ERSD	.85597			
HIT	.67952			
ZMKS	.92774			
LRSDEV	.90976			
RDEV	.46279			

Tabelle 8-8: Patternmatrix der Faktorenanalyse

	FACTOR 1	FACTOR 2
EMD	.69164	.20544
EMAD	.91077	-.14644
ERSD	.85509	-.21136
HIT	-.81374	.04296
ZMKS	.03390	.97014
LRSDEV	.09159	-.92932
RDEV	.68052	.00108

8.3 Ähnlichkeit der Reaktionskurven der Verfahren

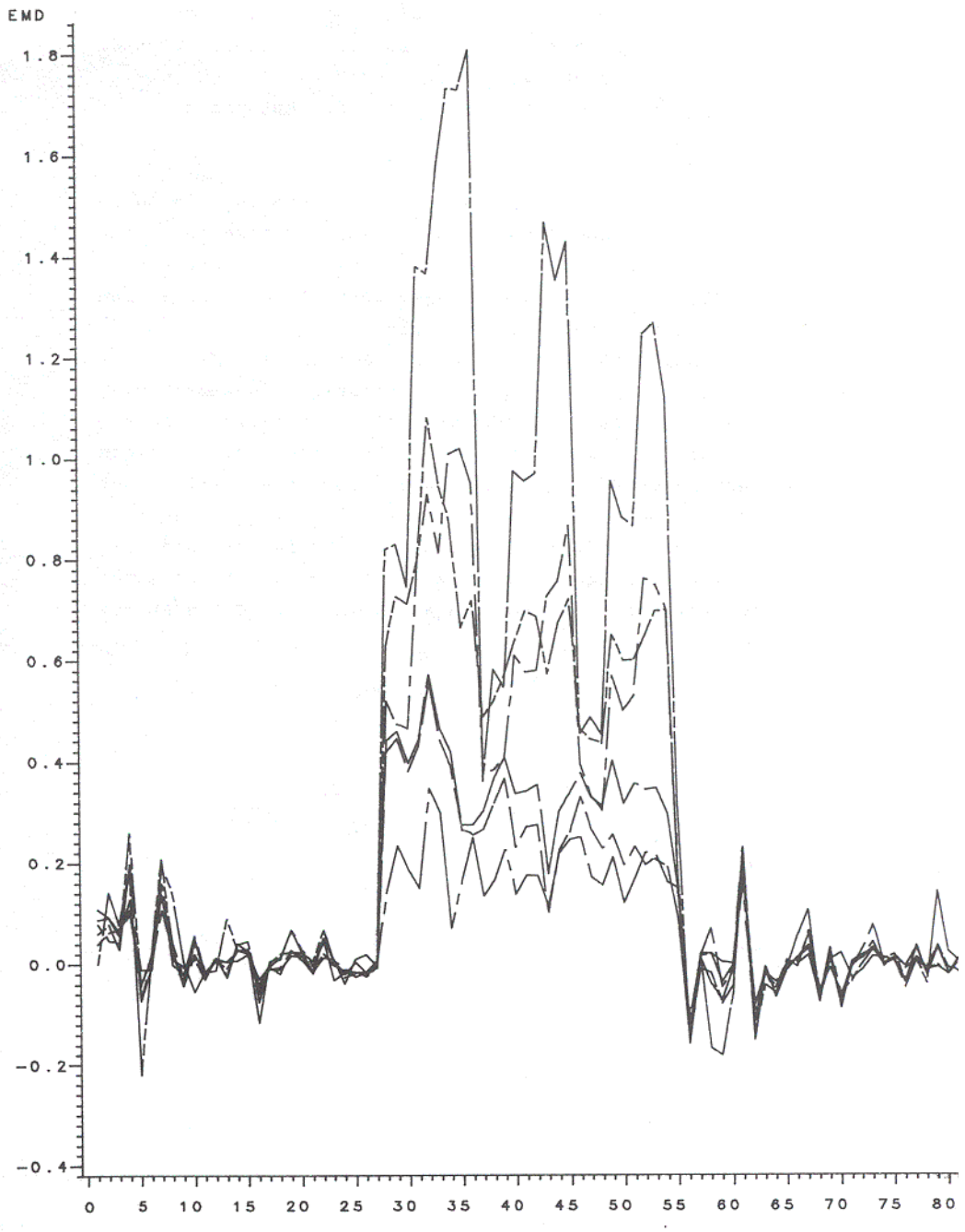
Die Hauptaufgabe der Simulation bestand in der Beantwortung der Frage, ob Aussagen, die auf Grund der Leistungsfähigkeit eines MD-Verfahrens unter der MAR-Annahme gemacht wurden, auch Geltung unter anderen, im Rahmen sozialwissenschaftlicher Forschung vermutlich realistischeren Ausfallmodellen Geltung beanspruchen können.

Ein Aspekt dieser Aufgabe bestand bei Betrachtung der zehn bzw. sechs Meßpunkte (die MD-Verfahren) pro "wahrer" Korrelationsmatrix darin, festzustellen, ob in Hinsicht auf die Güte der Schätzung eine konstante Differenz zwischen den Verfahren existiert. Darüber hinaus ließ die Konstruktion der Ersetzungsverfahren die Existenz "natürlicher" Gruppen vermuten, die in sich homogen auf die experimentellen Bedingungen reagieren. Diese Fragen wurden mit unterschiedlichen Methoden am Datensatz untersucht.

Die Mittelwerte der jeweils 10 Messungen einer der abhängigen Variablen konnten bei dem gegebenen Design für jedes MD-Verfahren unter den 81 verschiedenen experimentellen Bedingungen als ein "Verlaufsprofil" aufgefaßt werden, so daß je nach abhängiger Variable entweder 6 oder 10 Verlaufsprofile entstehen. Jede im Plot erkennbare "Kreuzung" der Verlaufsprofile entspricht damit einem Interaktionseffekt zwischen einem MD-Verfahren und den experimentellen Bedingungen: Wäre die Annahme korrekt, daß die relative Leistungsfähigkeit der MD-Verfahrens unabhängig von den experimentellen Bedingungen ist, dann müßten die Reaktionskurven der Verfahren parallel verlaufen. Hätten die experimentellen Bedingungen keinen Einfluß auf die Leistungsfähigkeit, müßten die Reaktionskurven zudem parallel zur X-Achse verlaufen. Wäre die Wahl eines MD-Verfahrens beliebig, also alle Verfahren gleich effizient, dann müßten alle Profile durch <u>eine</u> Reaktionskurve darstellbar sein.

Um einen ersten Überblick zu erhalten, wurden die Mittelwerte der Güte der Schätzungen der MD-Verfahren gegen die experimentellen Bedingungen geplottet[1/]. Vier dieser Plots sollen als Beispiele der beiden wichtigsten Ergebnisse dargestellt werden[2/]. Die Abbildung 8-3 zeigt den Plot der mittleren Abweichung der von den MD-Verfahren geschätzten fehlenden Werte gegen alle 81 experimentellen Bedingungen, wobei hohe Werte auf der Y-Achse schlechte Schätzungen der Daten (große Abweichungen) bedeuten. Der Plot zeigt vor allem[3/], daß der Einfluß des Ausfallmechanismus den aller anderen Faktoren übertrifft: Mit dem Wechsel des Ausfallmechanismus zwischen Bedingung 27 und Bedingung 28 von MAR zum Hazardmodell zeigt sich ein extremes Anwachsen der Schätzfehler; weiterhin zeigt sich die unterschiedliche Empfindlichkeit der Verfahren gegenüber dem Ausfallmechanismus durch die vergleichsweise großen Mittelwertdifferenzen zwischen dem schlechtesten und dem besten Verfahren.

Abbildung 8-3: Plot: Mittlere Abweichung der geschätzten Daten



Die Abbildungen 8-4 bis 8-6 stellen die Mittelwerte der MD- Verfahren in Hinsicht auf die Variable ZMKS, die den Fischer-Z- Werten der Korrelationskoeffizienten zwischen den durch die MD- Verfahren geschätzten und den "wahren" Korrelationskoeffizienten entspricht, dar. Die Plots wurden für jeden Ausfallmechanismus getrennt erstellt; hohe Werten auf der Y-Achse entsprechen nun sehr guten Schätzungen.

Sehr deutlich wird der geringe Einfluß der Stichprobengröße durch die nur kleinen Steigerungsänderungen aufeinanderfolgender Meßpunkte, ebenso der stärkere Einfluß der mittleren Interkorrelation und vor allem des Anteils fehlender Werte, der sich in den großen Steigungsänderung (alle 9 Einheiten von COND) äußert. Die extremen Verzerrungen aller Verfahren bei Kombination aller ungünstigen Faktoren (kleine Stichproben, niedrige Zusammenhänge, hohe Ausfälle) wird ebenso deutlich wie das relativ homogene Reaktionsverhalten der Verfahren: Die Reaktionskurven zeigen einen recht ähnlichen Verlauf, wenn auch von Parallelität der Profile nicht gesprochen werden kann. Die Plots zeigen die Existenz von Interaktionseffekten zwischen den Verfahren und den experimentellen Bedingungen, aber ebenso die relative Schwäche der Interaktionseffekte, wenn man den Ausfallmechanismus ausnimmt [4/].

Abbildung 8-4: Plot: Korrelationsabweichungsindex (ZMKS): MAR-Modell

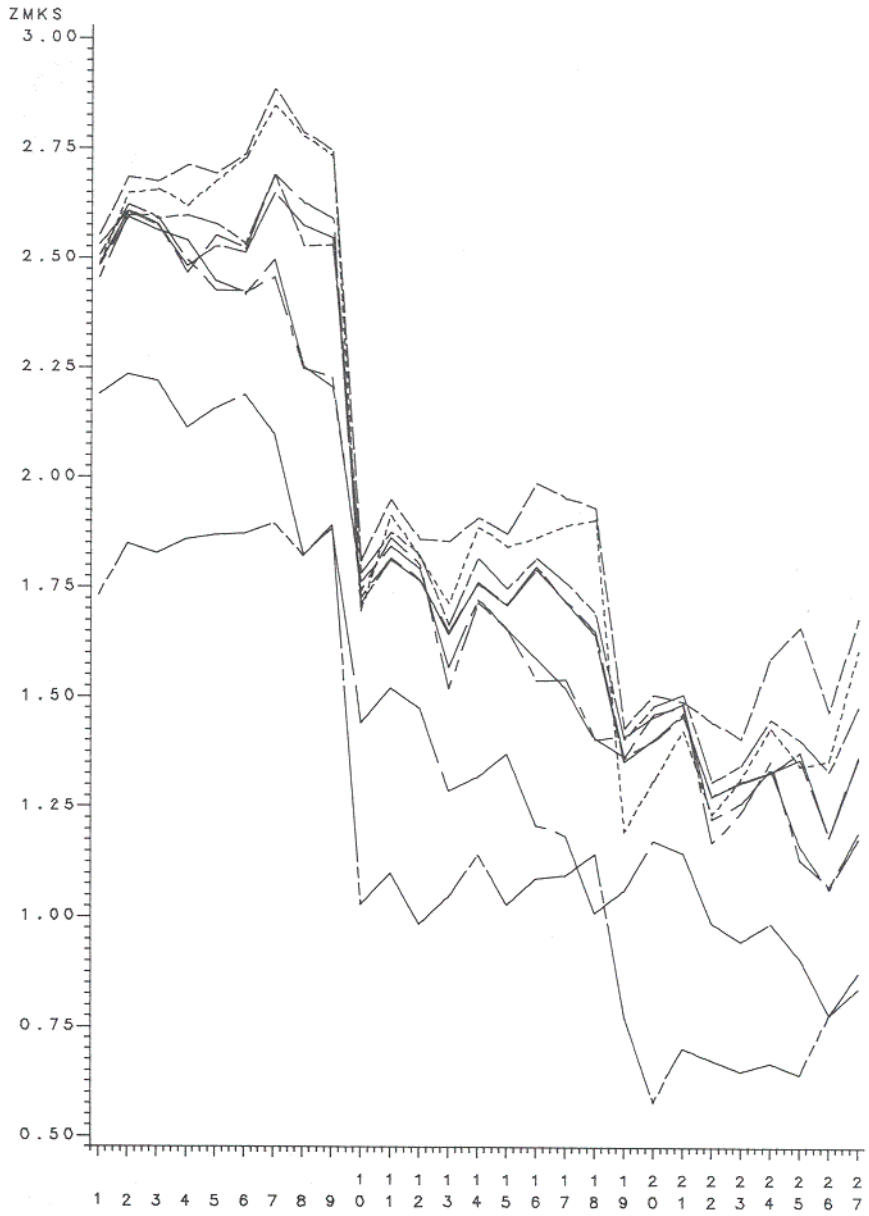


Abbildung 8-5: Plot: Korrelationsabweichungsindex (ZMKS): Hazard-Modell

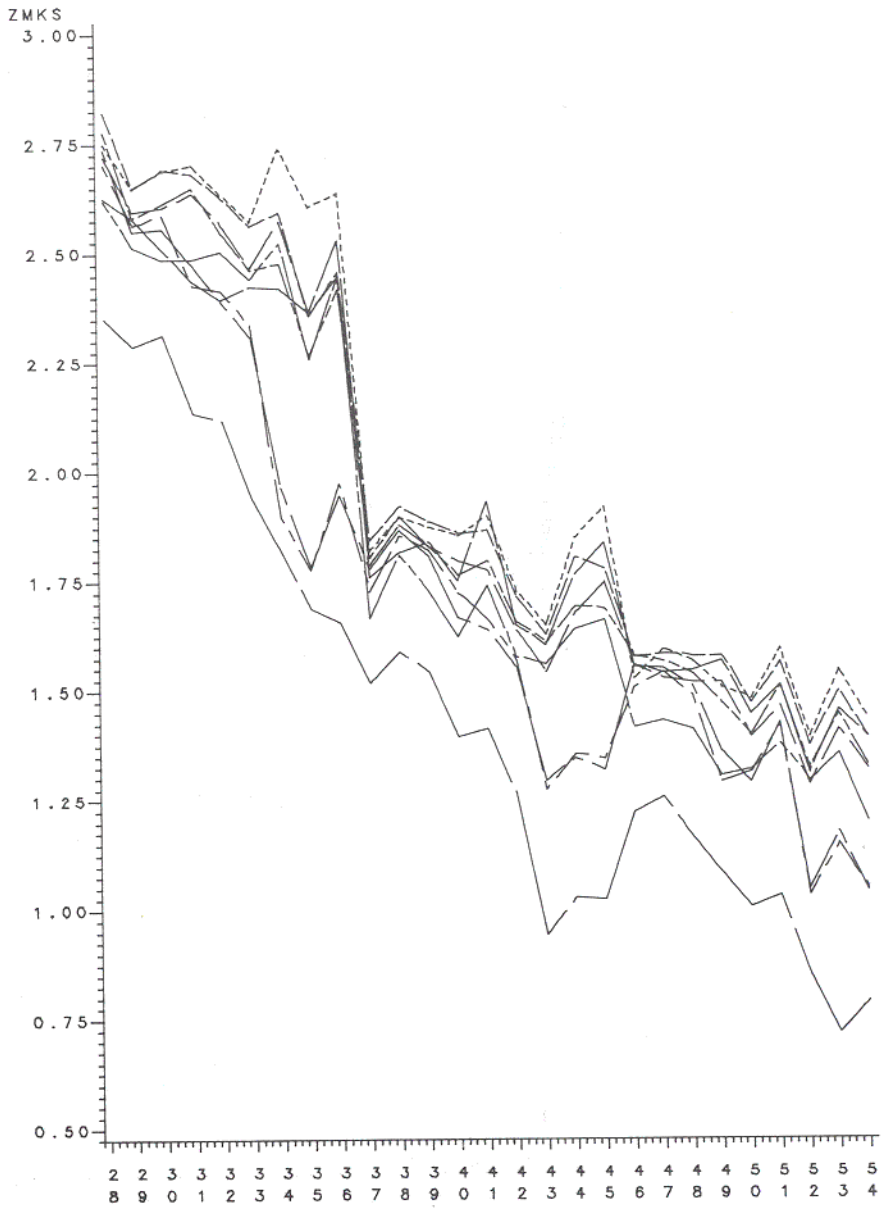
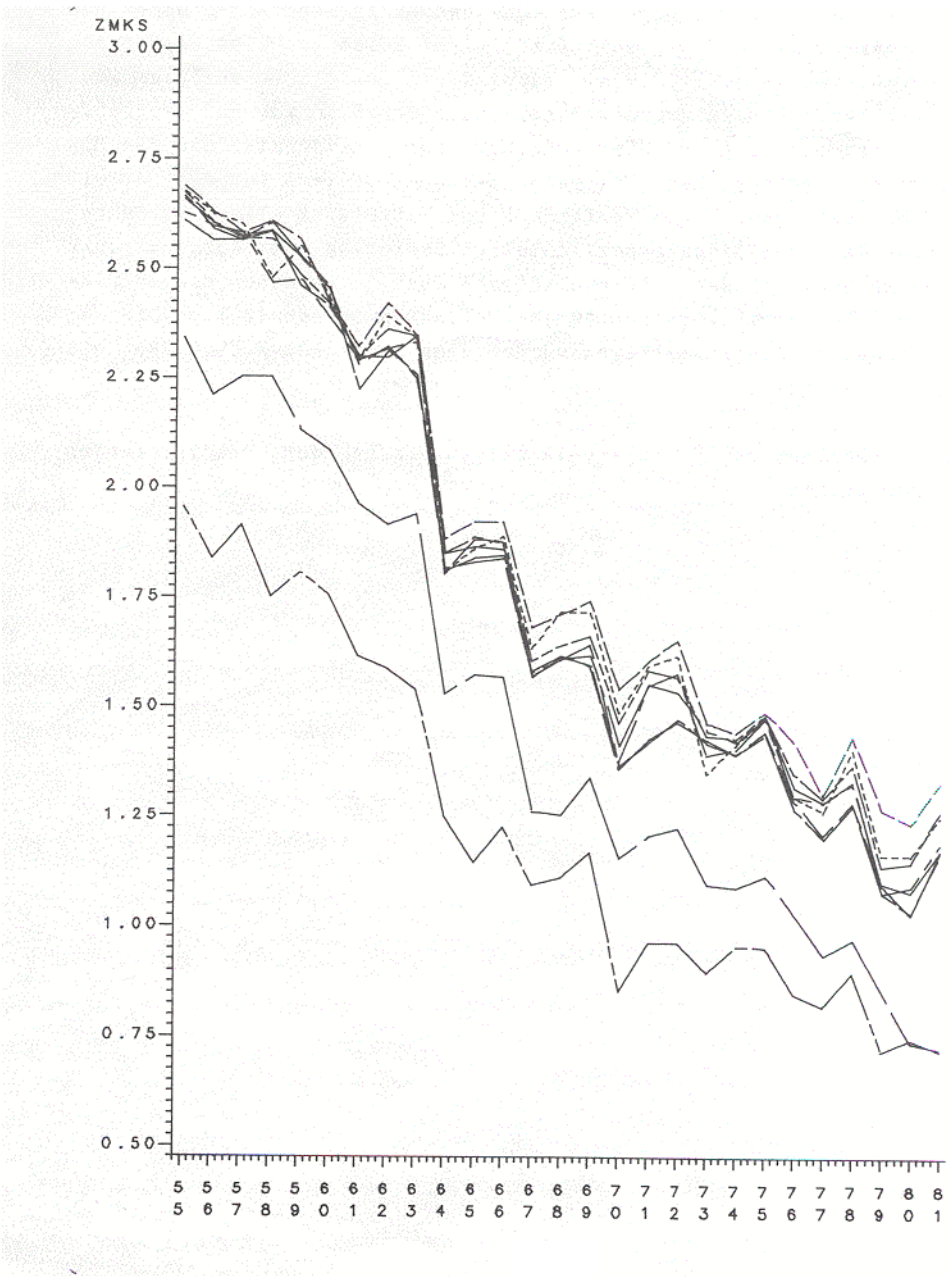


Abbildung 8-6: Plot: Korrelationsabweichungsindex (ZMKS): MARC-Modell



Trotz der schon durch die Plots deutlich gewordenen Interaktionseffekte bleibt die Frage nach der Ähnlichkeit der Reaktionskurven der Verfahren in Hinsicht auf voneinander isolierbare Gruppen von Verfahren. Diese Frage sollte mit einer Reihe von Clusteranalysen beantwortet werden[5/], u.a. einer Serie von Clusteranalysen nach Variablen. Hierzu wurden die, jeder der 810 "wahren" Korrelationsmatrizen zugeordneten zehn bzw. sechs Meßpunkte (die den MD-Verfahren entsprechen) fallweise z-standardisiert (um die Abweichung des jeweiligen Verfahrens vom Mittelwert aller Verfahren unter der jeweiligen experimentellen Bedingung zum Ausgangspunkt zu machen) und dann die euklidische Distanz zwischen den Variablen und damit zwischen den MD-Verfahren berechnet [6/]. Die so gewonnenen Distanzmatrizen wurden dann (mit SPSS-X CLUSTER) mit der Average-Linkage Methode [7/] geclustert.

Die Abbildungen 8-7 bis 8-14 zeigen einige der resultierenden Dendrogramme[8/].

Abbildung 8-7: Dendrogramm: ZMKS-MAR-Model

```
                          Cluster Distanz
Methode     *---------*---------*---------*---------*---------*
                      5        10        15        20        25
REGRESSION  -+-----+
             I     I
HAZARD      -+     +---+
                   I   I
PAIRWISE    -------+   +-----+
                       I     I
MEAN        ---+-------+     I
               I             I
CELLMEAN    ---+             +---------------------------------+
                             I                                 I
DEAR        ---+---+         I                                 I
               I   I         I                                 I
ILS         ---+   +---------+                                 I
                   I                                           I
MULT        -------+                                           I
                                                               I
LISTWISE    -----------------------------------+---------------+
                                               I
HOTDECK     -----------------------------------+
```

Abbildung 8-8: Dendrogramm: ZMKS-Hazard-Modell

```
                          Cluster Distanz
Methode     *---------*---------*---------*---------*---------*
                      5        10        15        20        25
MULT        -+
             I
ILS         -+---+---+
                 I   I
DEAR        -----+   +-------+
                     I       I
PAIRWISE    -----+---+       +-----+
                 I           I     I
REGRESSION  -----+           I     I
                             I     I
HAZARD      -----------------+     +-------------------------+
                                   I                         I
MEAN        -----+-------------+  I                         I
                 I             I  I                         I
CELLMEAN    -----+             +---+                         I
                               I                             I
LISTWISE    -------------------+                             I
                                                             I
HOTDECK     -------------------------------------------------+
```

Abbildung 8-9: Dendrogramm: ZMKS-MARC-Modell

```
                          Cluster Distanz
Methode     *---------*---------*---------*---------*---------*
                      5         10        15        20        25
REGRESSION  -+
             I
HAZARD      -+--------+
                      I
PAIRWISE    -+-+      +----------------------------------------+
             I I      I                                         I
MEAN        -+ +---+  I                                         I
               I   I  I                                         I
CELLMEAN    ---+   +--+                                         I
                   I                                            I
MULT        -+     I                                            I
             +---+ I          I                                 I
ILS         -+   +-+                                            I
                   I                                            I
DEAR        -------+                                            I
                                                                I
LISTWISE    -------------------------------+-------------------+
                                             I
HOTDECK     -------------------------------+
```

Abbildung 8-10: Dendrogramm: ZMKS - Total

```
                          Cluster Distanz
Methode     *---------*---------*---------*---------*---------*
                      5         10        15        20        25
MULT        -+
             I
ILS         -+-+------+
               I      I
DEAR        ---+      +---+
                      I   I
PAIRWISE    -----+----+   +---+
                 I         I   I
REGRESSION  -----+         I   +-------------------------------+
                           I   I                                 I
HAZARD      --------------+   I                                 I
                              I                                 I
MEAN        -+----------------+                                 I
             I                                                  I
CELLMEAN    -+                                                  I
                                                                I
LISTWISE    ----------------------------------------+---------I
                                                     I
HOTDECK     ----------------------------------------+
```

Die Dendrogramme (Abb. 8-7 bis 8-10) für ZMKS zeigen ebenso wie die Plots die Isolierung des Reaktionsmusters des HOTDECK-Verfahrens, das auf Grund seiner außergewöhnlich schlechten Ergebnisse immer ein eigenes Cluster bildet; ähnliches gilt für LISTWISE, wenn auch die Resultate wesentlich besser als für HOTDECK sind (s.w.u.). Auffällig ist die Zuordnung der Mittelwertverfahren zum LISTWISE-Cluster unter dem Hazard-Modell: Zwar bestehen noch starke Unterschiede zwischen den Verfahren, aber der Ausfallmechanismus schlägt auf die Leistungsfähigkeit der Verfahren durch, da sowohl die Mittelwertverfahren als auch LISTWISE hierbei systematisch die Mittelwerte unterschätzen. Daß die Mittelwertverfahren sehr ähnliche Profile besitzen, ist aufgrund der "mechanischen" Bestimmung der Imputationklassen und der niedrigen gesetzten Korrelationen der Imputationklassenvariablen mit den "internen" Variablen verständlich. Durch die Konstruktion des Hazard-Modells wird auch die Zuordnung der HAZARD-Methode nachvollziehbar: Da unter MAR die HAZARD-Methode durch den zusätzlichen Prädiktor "MD-Häufigkeit" keinen Informationsgewinn nutzen kann, sind die Resultate mit denen von REGRESSION nahezu identisch; wird der Prädiktor "informationsreicher", verbessert sich das Verfahren und erhält damit ein "selbständigeres" Profil. Da MULT und ILS sich lediglich durch die Iteration unterscheiden, ist die Nähe ihrer Reaktionsprofile wenig erstaunlich. Die Zuordnung von DEAR zum Cluster der multiplen Regressionsverfahren kann durch die der Simulation zugrunde liegende homogene Interkorrelationsstruktur der internen Variablen erklärt werden: Letztlich wird eine "latente" Variable simuliert, die von diesen Verfahren lediglich mit unterschiedlichen Ansätzen "approximiert" wird. Die Verallgemeinerung dieses Ergebnisses ist somit äußerst problematisch. Auf diesem Hintergrund wird dann auch die Zuordnung von PAIRWISE zu den die Information der Kovarianzstruktur stark nutzenden Verfahren im Hazardmodell verständlich; die "Totalausfälle" für PAIRWISE wirken relativ schwach in Richtung einer systematischen Verzerrung.

Abbildung 8-11: Dendrogramm: HIT-MAR-Modell

```
                          Cluster Distanz
Methode     *---------*---------*---------*---------*---------*
                     5        10        15        20        25
MULT        --+
              I
ILS         --+-------+-----+
                     I     I
DEAR        ----------+     +-----------------------------------I
                           I                                    I
REGRESSION  --+-------------+                                   I
              I                                                 I
HAZARD      --+                                                 I
                                                                I
HOTDECK     ----------------------------------------------------+
```

Abbildung 8-12: Dendrogramm: HIT-Hazard-Modell

```
                         Cluster Distanz
Methode     *---------*---------*---------*---------*---------*
                   5        10        15        20        25
MULT        --+
              I
ILS         --+--------------+
                             +-----+
HAZARD      -----------------+     I
                                   +--------------------------+
REGRESSION  -----------------+     I                          I
                             +-----+                          I
DEAR        -----------------+                                I
                                                               I
HOTDECK     ---------------------------------------------------+
```

Abbildung 8-13: Dendrogramm: HIT-MARC-Modell

```
                          Cluster Distanz
Methode     *---------*---------*---------*---------*---------*
                     5        10        15        20        25
MULT        --+
              +----------+
ILS         --+          +------+
                         I      I
DEAR        -------------+      +------------------------------+
                                 I                              I
REGRESSION  -------+------------+                              I
                   I                                            I
HAZARD      -------+                                            I
                                                                I
HOTDECK     ----------------------------------------------------+
```

Abbildung 8-14: Dendrogramm: HIT-Total

```
                          Cluster Distanz
Methode     *---------*---------*---------*---------*---------*
                     5        10        15        20        25
MULT        --+
              I
ILS         --+-------------+
                            I
DEAR        ----------------+-+-------------------------------+
                              I                               I
REGRESSION  -----------+------+                               I
                       I                                      I
HAZARD      -----------+                                      I
                                                              I
HOTDECK     ----------------------------------------------------+
```

Die Dendrogramme für die Anzahl exakter Schätzungen fehlender Werte (HIT) (vgl. Abb. 8-11 bis 8-14) zeigen in groben Zügen ein vergleichbares Bild: Die Isolation des HOTDECK-Verfahrens, die Nähe der multiplen Regressionsverfahren und die veränderte Zuordnung der HAZARD-Technik im Hazardmodell. Auffällig ist lediglich die Trennung von DEAR im Hazardmodell von den multiplen Regressionsverfahren: Wie später die Ergebnisse der Regressionen zeigen werden, ist dieses Resultat durch die Empfindlichkeit gegenüber der mittleren Interkorrelation, die hier direkt auf die Ausfall-wahrscheinlichkeiten wirkt, erklärlich.

Zusammenfassend kann festgehalten werden, daß die Verfahren, nahezu trivialerweise, durch ihre Konstruktion ähnliche Reaktionsmuster innerhalb bestimmter Gruppen besitzen[9/].

8.4 Einflüsse der experimentellen Faktoren

Um die Frage nach dem Einfluß der experimentellen Bedingungen auf das Verhalten der Verfahren zu beantworten, wurden zunächst Varianzanalysen und danach multiple Regressionen durchgeführt.

8.4.1 Ergebnisse der Varianzanalysen

Für jede der abhängigen Variablen wurde (mit SPSSX-MANOVA) eine eigene fünffaktorielle Varianzanalyse mit Meßwiederholung auf dem letzten Faktor berechnet[10/]. Eine Zusammenfassung der Ergebnisse findet sich in Tabelle 8-11.

Die Effekte der experimentellen Faktoren zwischen den Objekten sind für alle Variablen mit drei Ausnahmen hochsignifikant. Die nichtsignifikanten Effekte sind N bei RDEV und EMD, sowie MPER bei ERSD (s.u.).

Der Effekt der MD-Verfahren ist unter allen Bedingungen hochsignifikant. Jeder der Faktoren (Fallzahl, MDGP, mittlere Interkorrelation und Anzahl fehlender Werte) zeigt für jede der abhängigen Variablen hochsignifikante Interaktionseffekte mit den MD- Verfahren. Die einzige Ausnahme hiervon bildet die nichtsignifikante Interaktion zwischen Fallzahl und MD-Verfahren bei der mittleren Abweichung der geschätzten Daten. Die Höhe der Eta-Quadratkoeffizienten[11/] des Interaktionseffekts von MTYPE mit PROG (von .987 bei RSDEV bis .306 bei HIT) zeigt deutlich, daß die Wahl eines MD-Verfahrens gegenüber dem MDGP kritisch ist.

Um die Effekte zu verdeutlichen wurden anschließend an die Varianzanalysen multiple Regressionen gerechnet. Die Regressionen wurden für jede abhängige Variable, jeweils sowohl für die Programme als auch für die drei MDGP getrennt durchgeführt.[12/].

Tabelle 8-9: Zusammenfassung der Varianzanalysen

Variable: ZMKS

ZWISCHEN (MS:.00926 ; DF 2/729)

	F	SIGNIF.
MTYPE	40.342	< .001
N	3.655	< .050
MPER	1421.117	< .001
CM	94.191	< .001

BINNEN

	LAMBDA	MULT.-F	DF	SIGNIF.
PROG	.10747	665.337	9/ 721	< .001
PROG*MTYPE	.27745	71.979	18/1442	< .001
PROG*N	.76850	11.273	18/1442	< .001
PROG*MPER	.24809	80.728	18/1442	< .001
PROG*CM	.39193	47.853	18/1442	< .001

Variable: RDEV

ZWISCHEN (MS:.00046 ; DF 2/729)

	F	SIGNIF.
MTYPE	1763.151	< .001
N	1.646	n.s.
MPER	53.319	< .001
CM	64.007	< .001

BINNEN

	LAMBDA	MULT.-F	DF	SIGNIF.
PROG	.01020	7777.469	9/ 721	< .001
PROG*MTYPE	.01756	524.478	18/1442	< .001
PROG*N	.67944	17.078	18/1442	< .001
PROG*MPER	.02067	477.117	18/1442	< .001
PROG*CM	.04102	315.456	18/1442	< .001

Variable: RSDEV

ZWISCHEN (MS:.21860 ; DF 2/729)

	F	SIGNIF.
MTYPE	133.860	< .001
N	1026.398	< .001
MPER	6277.575	< .001
CM	83.906	< .001

BINNEN

	LAMBDA	MULT.-F	DF	SIGNIF.
PROG	.03607	2141.156	9/ 721	< .001
PROG*MTYPE	.01279	628.358	18/1442	< .001
PROG*N	.49944	33.247	18/1442	< .001
PROG*MPER	.55847	27.088	18/1442	< .001
PROG*CM	.24698	81.088	18/1442	< .001

Variable: HIT

ZWISCHEN (MS:.01301 ; DF 2/729)

	F	SIGNIF.
MTYPE	54.555	< .001
N	7.296	< .001
MPER	13.387	< .001
CM	44.608	< .001

BINNEN

	LAMBDA	MULT.-F	DF	SIGNIF.
PROG	.25906	414.719	5/ 725	< .001
PROG*MTYPE	.69353	29.114	10/1450	< .001
PROG*N	.96579	2.546	10/1450	< .010
PROG*MPER	.96437	2.655	10/1450	< .010
PROG*CM	.79536	17.588	10/1450	< .001

Variable: EMD

ZWISCHEN (MS:.17840 ; DF 2/729)

	F	SIGNIF.
MTYPE	829.913	< .001
N	2.850	< .100
MPER	18.864	< .001
CM	9.708	< .001

BINNEN

	LAMBDA	MULT.-F	DF	SIGNIF.
PROG	.19949	581.854	5/ 725	< .001
PROG*MTYPE	.11235	287.604	10/1450	< .001
PROG*N	.99074	.676	10/1450	n.s.
PROG*MPER	.80187	16.925	10/1450	< .001
PROG*CM	.52505	55.110	10/1450	< .001

Variable: EMAD

ZWISCHEN (MS:.05587 ; DF 2/729)

	F	SIGNIF.
MTYPE	101.889	< .001
N	11.964	< .001
MPER	15.382	< .001
CM	63.751	< .001

BINNEN

	LAMBDA	MULT.-F	DF	SIGNIF.
PROG	.10759	1202.687	5/ 725	< .001
PROG*MTYPE	.39816	84.794	10/1450	< .001
PROG*N	.87552	9.965	10/1450	< .001
PROG*MPER	.83830	13.369	10/1450	< .001
PROG*CM	.60302	41.725	10/1450	< .001

Variable: ERSD

ZWISCHEN (MS:.06505 ; DF 2/729)

	F	SIGNIF.
MTYPE	100.127	< .001
N	8.256	< .001
MPER	2.921	< .100
CM	81.234	< .001

BINNEN

	LAMBDA	MULT.-F	DF	SIGNIF.
PROG	.08811	1500.661	5/ 725	< .001
PROG*MTYPE	.36037	96.542	10/1450	< .001
PROG*N	.81722	15.398	10/1450	< .001
PROG*MPER	.78218	18.951	10/1450	< .001
PROG*CM	.56696	47.572	10/1450	< .001

8.4.2 Ergebnisse der Regressionen

Die Tabellen 8-10 bis 8-12 zeigen, daß der Anteil fehlender Werte den stärksten Einfluß auf die Reproduktion der Struktur der Korrelationsmatrix besitzt. Der Effekt ist stets signifikant und erklärt zwischen 81% und 55% der Varianz. Entgegen der Erwartung, daß die Reproduktion der Struktur umso besser gelingt, je höher die Fallzahl wird, zeigen die Regressionen, daß die Fallzahl lediglich für MULT und ILS unter MAR einen signifikanten Einfluß besitzt, der allerdings weniger als 1% der Varianz erklärt. Möglicherweise sind Effekte der Fallzahl lediglich bei kleineren Stichproben als den hier simulierten (mit einem Minimum von N = 100) feststellbar.

Die interessantesten Ergebnisse zeigt der Effekt der mittleren Interkorrelation. Unter MAR ist CM nur für diejenigen Verfahren signifikant, die die Information über die Kovarianzstruktur entweder überhaupt nicht oder am stärksten benutzen: Der Effekt ist für MEAN, CELLMEAN und HOTDECK negativ (mit ca. 4% erklärter Varianz) und für DEAR, MULT und ILS schwach positiv (weniger als 1% erklärter Varianz). Unter MARC wechseln die Vorzeichen des Effekts für DEAR, MULT und ILS: War die Schätzung unter MAR umso besser, je höher die Korrelation, so wird die Schätzung nun umso schlechter. Die Effekte sind mit ca. 4% erklärter Varianz allerdings schwach. Für DEAR trifft der Vorzeichenwechsel von CM unter dem Hazard-Modell ebenfalls zu, wobei der Effekt wiederum nur schwach ist (ca. 2%).

Auffällig sind weiterhin die starken Effekte von CM unter dem Hazard-Modell für MEAN, CELLMEAN und HOT-DECK, die zwischen 16% und 18% der Varianz erklären. Festgehalten werden sollte, daß für alle Verfahren unter den Nicht-MAR-Bedingungen der Effekt von CM stets negativ ist. Je stärker also die Zusammenhänge in den simulierten Stichproben werden, umso größer sind die resultierenden Verzerrungen in den geschätzten Korrelationsmatrizen in Hinsicht auf die relativen Größenordnungen der Koeffizienten bei systematischen Ausfällen.

Tabelle 8-10: Regressionen ZMKS: MAR

		--- Regressionsparameter---			--- Beta[13/] ---		
	R^2	CM	N	MPER	CM	N	MPER
LIST	.80	.1703	5.6806E-5	-12.5191*	.0539	.0182	-.8930
PAIR	.78	-.1020	4.3718E-5	-12.6858*	-.0316	.0136	-.8854
MEAN	.78	-.6982*	-5.1683E-5	-12.6525*	-.2106	-.0158	-.8606
CELL	.78	-.7101*	6.6323E-5	-12.6897*	-.2134	.0201	-.8601
DECK	.76	-.7485*	-3.4454E-5	-12.1811*	-.2290	-.0106	-.8401
REGR	.80	-.0761	3.0276E-5	-13.4929*	-.0224	.0090	-.8945
HAZD	.79	-.0909	3.6198E-5	-13.2334*	-.0272	.0109	-.8921
DEAR	.79	.3163*	6.9997E-5	-13.0991*	.0950	.0212	-.8870
MULT	.82	.3268*	3.0496E-4*	-14.5422*	.0898	.0846	-.9007
ILS	.78	.2917*	2.2961E-4*	-13.1041*	.0868	.0690	-.8789

Tabelle 8-11: Regressionen ZMKS: Hazard-Modell

		--- Regressionsparameter---			---- Beta ----		
	R^2	CM	N	MPER	CM	N	MPER
LIST	.74	-.3202*	5.1070E-5	-12.3244*	-.0989	-.0159	-.8584
PAIR	.73	-.5390*	-1.0974E-4	-11.6637*	-.1713	-.0352	-.8357
MEAN	.74	-1.3887*	-1.0719E-4	-10.5931*	-.4323	-.0337	-.7436
CELL	.75	-1.3729*	4.3843E-7	-10.7737*	-.4264	.0001	-.7545
DECK	.74	-1.3153*	1.6526E-4	-11.1336*	-.4003	-.0508	-.7639
REGR	.73	-.4593*	-1.0252E-4	-12.0867*	-.1423	-.0321	-.8446
HAZD	.70	-.3050*	4.9530E-5	-10.8768*	-.1034	.0170	-.8316
DEAR	.74	-.4657*	-9.0701E-4	-11.9077*	-.1468	-.0289	-.8460
MULT	.74	-.1660	1.1674E-5	-12.5194*	-.0507	.0036	-.8626
ILS	.74	-.0988	2.8567E-5	-12.2152*	-.0301	.0090	-.8593

* $p < .05$

Tabelle 8-12: Regressionen ZMKS: MARC

		-- Regressionsparameter--			---- Beta -----		
	R^2	CM	N	MPER	CM	N	MPER
LIST	.75	-.6730*	4.5634E-5	-9.9143*	-.2518	.0171	-.8309
PAIR	.82	-.7279*	4.9323E-5	-12.6198*	-.2240	.0153	-.8756
MEAN	.82	-.8147*	2.4602E-5	-12.8265*	-.2462	.0075	-.8739
CELL	.83	-.7870*	1.3221E-4	-12.7677*	-.2400	.0407	-.8780
DECK	.82	-.8500*	-5.8109E-5	-12.7943*	-.2568	-.0177	-.8715
REGR	.80	-.8836*	-3.4603E-7	-13.2124*	-.2554	-.0001	-.8612
HAZD	.80	-.8724*	4.2282E-5	-13.1237*	-.2530	.0123	-.8580
DEAR	.80	-.6373*	1.6027E-5	-12.3707*	-.1992	.0051	-.8715
MULT	.81	-.6644*	1.4506E-4	-12.8727*	-.2014	.0444	-.8797
ILS	.81	-.6800*	1.2763E-4	-12.5884*	-.2093	.0397	-.8736

Tabelle 8-13: Regressionen LRSDEV: MAR

		--- Regressionsparameter---			--- Beta ---		
	R^2	CM	N	MPER	CM	N	MPER
LIST	.84	-.9629*	-1.1882E-3*	15.4826*	-.2165	-.4276	.7848
PAIR	.86	-.9211*	-1.9085E-3*	13.5470*	-.2288	-.4789	.7585
MEAN	.84	.9488*	-9.9744E-4*	17.3717*	.2091	-.2221	.8631
CELL	.83	1.0144*	-1.2174E-3*	16.9411*	.2235	-.2710	.8416
DECK	.80	1.2893*	-9.0553E-4*	17.6938*	.2744	-.1947	.8489
REGR	.86	-.8788*	-1.6962E-3*	15.0841*	-.2085	-.4066	.8068
HAZD	.86	-.8513*	-1.7025E-3*	14.8283*	-.2048	-.4139	.8044
DEAR	.86	-.9880*	-1.1189E-3*	17.0338*	-.2209	-.2686	.8588
MULT	.87	-.7987*	-1.4610E-3*	18.3832*	-.1669	-.3085	.8663
ILS	.86	-.6971*	-1.3556E-3*	18.0608*	-.1488	-.2923	.8689

* $p < .05$

Tabelle 8-14: Regressionen LRSDEV: Hazard-Modell

		--- Regressionsparameter---			---- Beta ----		
	R^2	CM	N	MPER	CM	N	MPER
LIST	.80	1.8316*	-5.6512E-4*	15.9462*	.4026	-.1255	.7903
PAIR	.77	1.4794*	-7.2918E-4*	14.1360*	.3623	-.1804	.7805
MEAN	.84	2.4499*	-5.1742E-4*	15.2450*	.5272	-.1125	.7397
CELL	.84	2.4915*	-6.9414E-4*	14.8456*	.5398	-.1520	.7252
DECK	.82	2.1778*	-6.2563E-4*	15.6265*	.4704	-.1365	.7610
REGR	.75	.3027*	-8.6488E-4*	13.9458*	.0801	-.2314	.8325
HAZD	.81	-.8705*	-1.8425E-3*	11.4240*	-.2375	-.5079	.7028
DEAR	.75	.9853*	-1.1783E-3*	12.2501*	.2700	-.3622	.7568
MULT	.83	-1.0443*	-1.8922E-3*	13.1449*	-.2579	-.4722	.7320
ILS	.83	-.8528*	-1.7583E-3*	14.3427*	-.2059	-.4289	.7808

Tabelle 8-15: Regressionen LRSDEV: MARC

		-- Regressionsparameter--			---- Beta -----		
	R^2	CM	N	MPER	CM	N	MPER
LIST	.77	1.3685*	-1.0387E-3*	12.2574*	.3696	-.2834	.7464
PAIR	.87	.9747*	-1.3432E-3*	16.1436*	.2268	-.3158	.8469
MEAN	.84	.0029	-1.8889E-3*	12.5402*	.0008	-.5148	.7626
CELL	.85	.0363	-1.9963E-3*	12.6168*	.0096	-.5341	.7532
DECK	.84	.4583*	-1.4969E-3*	14.2248*	.1179	-.3892	.8253
REGR	.86	.5324*	-1.5580E-3*	15.6305*	.1271	-.3758	.8414
HAZD	.86	.5000*	-1.5852E-3*	15.4078*	.1202	-.3851	.8353
DEAR	.87	.5627*	-1.0985E-3*	17.2352*	.1290	-.2544	.8905
MULT	.88	.6390*	-1.2181E-3*	18.0716*	.1395	-.2687	.8895
ILS	.88	.6528*	-1.2239E-3*	18.0885*	.1424	-.2697	.8894

* $p < .05$

Tabelle 8-16: Regressionen RDEV: MAR

		--- Regressionsparameter---			--- Beta ---		
	R^2	CM	N	MPER	CM	N	MPER
LIST	.00	-3.6754E-3	-6.0151E-6	6.2132E-2	-.0224	-.0370	.08412
PAIR	.00	2.8913E-3	-3.5536E-7	4.7776E-3	.0831	-.0103	.03097
MEAN	.77	5.3887E-2*	1.0166E-6	3.7417E-1*	.4721	.0090	.73905
CELL	.74	5.7738E-2*	2.1921E-6	3.4574E-1*	.5135	-.0197	.69332
DECK	.78	9.4341E-2*	2.5419E-6	5.9511E-1*	.5088	.0139	.72364
REGR	.06	-1.8633E-3	1.5118E-6	3.5986E-2*	-.0584	.0478	.25410
HAZD	.07	-4.0301E-4	2.3431E-10	4.0836E-2*	-.0123	.0000	.28006
DEAR	.76	3.0893E-3	2.2324E-6	-2.2286E-1*	.0535	-.0390	-.86960
MULT	.52	-1.2395E-2*	-1.3411E-5*	-1.7288E-1*	-.2088	-.2282	-.6565
ILS	.72	-1.0724E-2*	-5.5509E-6*	-2.2272E-1*	-.1738	-.0909	-.8303

Tabelle 8-17: Regressionen RDEV: Hazard-Modell

		--- Regressionsparameter---			---- Beta ----		
	R^2	CM	N	MPER	CM	N	MPER
LIST	.70	.1113*	-6.0177E-6	.5910*	.5373	-.0294	.6433
PAIR	.67	.0691*	-4.5637E-6	.3508*	.5387	-.0359	.6163
MEAN	.78	.1289*	-4.7323E-6	.6142*	.6023	-.0223	.6468
CELL	.77	.1281*	-7.0764E-6	.5640*	.6220	-.0347	.6171
DECK	.77	.1462*	-4.1944E-6	.7198*	.5899	-.0171	.6547
REGR	.52	.0253*	1.2685E-6	.2310*	.3177	.0161	.6547
HAZD	.00	-.0039	-3.3146E-6	.0071	-.0900	-.0765	.0367
DEAR	.55	.0547*	-4.2296E-6	.1298*	.6570	-.0513	.3513
MULT	.05	-.0088*	-3.6089E-6	-.0223	-.1952	-.0808	-.1116
ILS	.41	-.0174*	-1.6987E-6	-.1359*	-.3181	-.0313	-.5591

* $p < .05$

Tabelle 8-18: Regressionen RDEV: MARC

		-- Regressionsparameter--			--- Beta ----		
	R^2	CM	N	MPER	CM	N	MPER
LIST	.64	-.1249*	-6.3048E-7	-.4228*	-.6369	-.0033	-.4863
PAIR	.76	-.0505*	2.8124E-6	-.2831*	-.5421	.0305	-.6856
MEAN	.11	-.0120*	1.8607E-6	.0032	-.3377	.0527	.0199
CELL	.09	-.0096*	-1.0323E-6	-.0300*	-.2587	-.0282	-.1837
DECK	.56	.0277*	4.7622E-6	.2361*	.3460	.0601	.6651
REGR	.64	-.0268*	3.1766E-6	-.1727*	-.4535	.0544	-.6596
HAZD	.61	-.0261*	8.3120E-7	-.1611*	-.4587	.0147	-.6377
DEAR	.86	-.0272*	1.5996E-6	-.3507*	-.3021	.0180	-.8786
MULT	.83	-.0357*	-2.4565E-6	-.3683*	-.3605	-.0250	-.8381
ILS	.84	-.0351*	-1.7524E-6	-.3708*	-.3540	-.0179	-.8438

* $p < .05$

Für die transformierte Spur der Differenzenmatrix (LRSDEV) ist der Effekt des Anteils fehlender Werte stets signifikant und erklärt zwischen 49% und 79% der Varianz. Der Effekt der Fallzahl ist im Gegensatz zu ZMKS für LRSDEV durchgängig signifikant, wobei das Vorzeichen den Erwartungen entspricht: Je höher die Fallzahl, desto geringer die Abweichungen. Die Effektstärke schwankt zwischen den Verfahren und den MDGP erheblich: Der Anteil erklärter Varianz liegt zwischen 1% und 28%. Mit den beiden Ausnahmen MEAN und CELLMEAN unter MARC ist der Effekt von CM immer signifikant und erklärt zwischen ca. 1% und 29% der Varianz. Auffällig sind unter MAR die negativen Vorzeichen aller Verfahren, die die Information der Kovarianzmatrix nutzen; die beiden Mittelwertverfahren und das HOTDECK besitzen dagegen für CM ein positives Vorzeichen. Unter dem Hazard-Modell besitzen alle Verfahren mit Ausnahme von HAZARD, MULT und ILS positive Vorzeichen von CM: Bei diesen Verfahren wird die Schätzung tendenziell besser, je höher der Zusammenhang in der Matrix wird; der Effekt ist jedoch mit 4% schwach, so daß eher von keinem ungünstigen Einfluß gesprochen werden kann. Unter MARC sind alle Vorzeichen von CM positiv, die Effekte jedoch sehr schwach, wobei nur PAIRWISE (5%) und LISTWISE (13%) deutliche Effekte zeigen. Die Interpretation keines ungünstigen

Einflusses von CM wird durch die vergleichbaren Resultate für ZMKS (s.o.) gestützt.

Für die mittlere Abweichung der Korrelationskoeffizienten unter MAR sind die durch die Regression erklärten Varianzen für LISTWISE und PAIRWISE gleich Null. Obwohl die Regressionen für REGRESSION und HAZARD signifikant sind, sind die Abweichungen mit 6%, bzw. 7% kaum erklärt. Betrachtet man die Mittelwerte der Verfahren unter MAR, so zeigen LISTWISE (GM:3.028E-3), PAIRWISE (GM:3.264E-4), REGRESSION (GM:2.279E-3) und HAZARD (GM:2.793E-3) im Gegensatz zu den anderen Verfahren keine systematische Über- oder Unterschätzung der Korrelationskoeffizienten. MEAN, CELLMEAN und HOTDECK zeigen dagegen systematische, wenn auch kleine (.0201, .0186, .0332) Unterschätzungen; DEAR, MULT und ILS hingegen Überschätzungen (-.012, -.01, -.012). Lediglich MPER zeigt durchgängig einen signifikanten Effekt auf die mittlere Abweichung. CM wirkt nur auf die "nichtinformativen" Verfahren in beachtlichem Umfang mit ca. 25% der Varianz. Unter dem HAZARD- Modell besitzen HAZARD und MULT geringe Abweichungen, die durch die Regressionen nicht oder nur zu 5% erklärt werden können. Für die anderen Verfahren werden zwischen 41% und 78% der Varianz erklärt. Die durchgängig signifikanten Effekte von MPER auf die anderen Verfahren erklären zwischen 31% und 42% der Varianz. Da von den Verfahren nur ILS eine leichte Überschätzungstendenz zeigt (-3.698E-3) sind alle Koeffizienten mit Ausnahme von ILS positiv. Die Fallzahl besitzt auf kein Verfahren einen Effekt in Hinsicht auf die mittlere Abweichung. Der Effekt von CM erklärt zwischen 3% und 38% der Varianz, wobei der Effekt auf die nichtinformativen Verfahren am stärksten wirkt. Unter MARC zeigen die Mittelwertverfahren keine erklärbaren Varianzen (< 1%). Alle anderen Verfahren zeigen deutliche Effekte von MPER (zwischen 23% und 77%) der erklärten Varianz. Das Vorzeichen von MPER wechselt gegenüber den anderen MDGP's

(mit Ausnahme des Vorzeichens des HOTDECK's): Alle Verfahren überschätzen im Mittel die Korrelationen. CM ist für alle Verfahren signifikant, wobei der Effekt sich auf LISTWISE am stärksten auswirkt.

Tabelle 8-19: Regressionen HIT : MAR

	--- Regressionsparameter---				---- Beta ----		
	R^2	CM	N	MPER	CM	N	MPER
DECK	.00	.0296	9.2426E-7	.1635	.0712	.0023	.0894
REGR	.05	.0989*	4.6264E-5	.0316	.2231	.1054	.0161
HAZD	.08	.1285*	1.3435E-5	.0384	.2924	.0309	.0197
DEAR	.28	.2842*	2.6455E-5	.0095	.5362	.0504	.0040
MULT	.29	.2581*	1.0685E-5*	-.2433*	.4941	.2067	-.1050
ILS	.29	.2520*	7.2697E-5*	-.0552	.5207	.1517	-.0257

Tabelle 8-20: Regressionen HIT: Hazard-Modell

	--- Regressionsparameter---				---- Beta ----		
	R^2	CM	N	MPER	CM	N	MPER
DECK	.18	-.1121*	1.3742E-5	.6344*	-.2672	.0331	.3411
REGR	.03	.0302	1.1516E-5	.3854*	.0660	-.0254	.1898
HAZD	.26	.2361*	7.6511E-5*	.5736*	.4363	.1429	.2391
DEAR	.11	-.0478	7.3732E-6	.7267*	-.0953	.0148	.3261
MULT	.25	.2450*	5.3953E-5*	.2521*	.4853	.1080	.1126
ILS	.29	.2724*	5.1750E-5	.3789*	.5132	.0985	.1610

* $p < .05$

Tabelle 8-21: Regressionen HIT: MARC

		--- Regressionsparameter---			---- Beta ----		
	R^2	CM	N	MPER	CM	N	MPER
DECK	.00	.0071	8.6100E-6	-.0116	.0177	-.0217	-.0066
REGR	.02	-.0730*	3.2085E-5	-.0310	-.1661	.0738	-.0159
HAZD	.00	-.0501	1.3963E-5	.0007	-.1151	.0324	.0003
DEAR	.04	-.1080*	-7.7710E-5	-.4833*	-.1471	-.1070	-.1485
MULT	.02	.0268	6.0932E-5*	.1424	.0662	.1551	.0793
ILS	.01	.0396	4.6529E-5	.0566	.0928	.1100	.0299

Tabelle 8-22: Regressionen ERSD: MAR

		--- Regressionsparameter---			---- Beta ----		
	R^2	CM	N	MPER	CM	N	MPER
DECK	.00	-.0700	8.7793E-5	-.1029	-.0663	.0840	-.0220
REGR	.31	-.4728*	-3.9292E-5	.0930	-.5649	-.0474	.0251
HAZD	.31	-.4912*	-2.6149E-5	-.0368	-.5668	-.0305	-.0096
DEAR	.64	-.7369*	5.2176E-6	.2246	-.7975	.0057	.0541
MULT	.59	-.7803*	-2.5226E-4*	1.1664*	-.6971	-.2277	.2350
ILS	.60	-.7574*	-1.4341E-4*	.4564*	-.7553	-.1445	.1026

Tabelle 8-23: Regressionen ERSD: Hazard-Modell

		--- Regressionsparameter---			---- Beta ----		
	R^2	CM	N	MPER	CM	N	MPER
DECK	.27	.7574*	8.5010E-5	-2.4242*	.4239	.0481	-.3059
REGR	.13	-.2838*	5.8911E-5	-.7978*	-.3081	.0646	-.1953
HAZD	.45	-.6556*	1.6391E-4*	-.8910*	-.6281	-.1587	-.1925
DEAR	.11	.0261	-3.0341E-5	-1.2860*	.0314	-.0368	-.3480
MULT	.54	-.7343*	-1.2477E-4*	.0411	-.7306	-.1254	.0092
ILS	.58	-.7692*	-1.2497E-4*	-.2033	-.7503	-.1232	-.0447

* $p < .05$

Tabelle 8-24: Regressionen ERSD: MARC

		--- Regressionsparameter---			---- Beta ----		
	R^2	CM	N	MPER	CM	N	MPER
DECK	.00	.0109	-2.7324E-5	.0161	.0104	-.0263	.0035
REGR	.05	.1518*	-1.1251E-4*	.3734	.1825	-.1367	.1021
HAZD	.03	.1309*	-1.4134E-4*	.0409	.1424	-.1554	.0100
DEAR	.03	-.1134*	-7.7293E-5	-.3378	-.1470	-.1012	-.0987
MULT	.09	-.0889	-2.2955E-4*	.2684	-.1081	-.2819	.0736
ILS	.06	-.0799	-1.9497E-4*	.2336	-.0976	-.2405	.0643

* p < .05

Die Tabelle 8-19 zeigt die Regressionen der einzelnen Verfahren für den Anteil korrekt geschätzter Werte (HIT). Die Regressionen sind unter MAR nur für DEAR, MULT und ILS in der Lage, beachtenswerte Varianzanteile von HIT zu erklären (ca.29%). Hierbei erklärt CM ca. 25% der Varianz, weder MPER noch N zeigen deutliche Einflüsse. Unter dem Hazard-Modell liegen die erklärten Varianzen zwischen 3% und 29%, der Effekt von CM erklärt für HAZARD, MULT und ILS zwischen 17% und 25% der Varianz, wobei kein Einfluß auf DEAR feststellbar ist. Überraschenderweise ist der Effekt von MPER positiv (wenn auch mit maximal 10% schwach): Tendenziell steigt der Anteil korrekt geschätzter Werte mit dem Anteil fehlender Werte. Unter dem MARC-Modell kann die Varianz von HIT mit den Regressionen nicht mehr erklärt werden (maximal 4%).

Betrachtet man die Ergebnisse der Regressionen für die mittlere quadratische Abweichung der geschätzten Werte (Tabelle 8-22), so steigen die erklärten Varianzen von Null (HOTDECK) auf 64% bei DEAR an. Die stärksten Effekte besitzt die Variable CM auf das Hauptkomponentenverfahren, bei dem sie 63% der Varianz erklärt. Sowohl die Fallzahl als auch der Anteil fehlender Werte zeigt lediglich für MULT und ILS schwache Effekte, wobei N negativ und MPER positiv wirkt. Unter dem Hazard-Modell schwanken die erklärten Varianzen zwischen 11% und 58%. Den stärksten Effekt besitzt CM (mit Ausnahme von DEAR) (9% bis 57%). Der Effekt von CM wirkt auf DEAR nicht, auf das

HOTDECK positiv und auf alle anderen Verfahren negativ in Hinsicht auf ERSD. Die Fallzahl besitzt keinen nennenswerten Einfluß, MPER lediglich für HOTDECK und DEAR (9% bzw. 12%). Das für HIT bereits erwähnte Ergebnis des tendenziell positiven Effekts des Anteils fehlender Werte auf die Güte der Schätzung zeigt sich auch für die mittlere quadratische Abweichung: Alle Vorzeichen von MPER (mit Ausnahme von MULT) sind negativ; der Effekt ist allerdings nur schwach. Unter dem MARC- Modell sinken die erklärten Varianzen auf Null bis 9%, so daß keine linear erklärbare Varianz vorliegt.

8.5 Relative Effizienz der Verfahren

Um die Frage nach dem "besten" Verfahren unter den verschiedenenen Bedingungen zu beantworten, wurden in Anlehnung an TIMM (1969:28;1970:427) relative Effizienzindices berechnet. Die relative Effizienz eines Verfahrens (in bezug auf eine bestimmte Variable) ist definiert als der Quotient zwischen dem Meßwert des jeweiligen Verfahrens und dem Maximum, bzw. Minimum der Meßwerte aller Verfahren unter der jeweiligen Bedingung. Für die Variablen ZMKS und HIT wurde zunächst für jede experimentelle Bedingung das Maximum, für alle anderen Variablen das Minimum gesucht und der Quotient berechnet. Für jede abhängige Variable wurde für jedes Verfahren die relative Effizienz mit SPSSX-REPORT zusammengefaßt. Aufgrund des Umfangs der Tabellen werden im Anhang D nur zwei vollständige Reports (für ZMKS und HIT) wiedergegeben.

In bezug auf die Effizienz bei der Anzahl exakt geschätzter Werte zeigt HOTDECK über alle Bedingungen stets die niedrigsten Werte. Auffällig ist allerdings das Ansteigen der Effizienz im MARC- Modell. Während REGRESSION und HAZARD sich unter MAR und MARC nur wenig voneinander unterscheiden, wird die Differenz beider Verfahren unter MAT deutlich: HAZARD gehört unter diesem Modell zu den drei besten Verfahren. DEAR stellt unter MAR eines der drei besten Verfahren dar, wobei die Effizienz immer mit steigendem Zusammenhang in der Korrelationsmatrix wächst. Unter MAT ändert sich das Vorzeichen dieser Beziehung: Je höher CM, desto desto niedriger die

Effizienz von DEAR. In diesem Modell gehört DEAR nicht mehr zu den besten Verfahren. Unter MARC verbessert sich DEAR wieder: Der Zusammenhang mit CM wird wieder positiv, unter den meisten Bedingungen liegt DEAR über allen anderen Verfahren. MULT und ILS ähneln sich unter allen Bedingungen sehr stark, eine leichte Verbesserung durch die Iteration ist in der Regel feststellbar. Beide Verfahren zeigen insgesamt die höchsten Effizienzen. Bemerkenswert ist vor allem die erstaunlich hohe Effizienz von ILS unter MAT bei hohem CM und 10% Ausfällen: Unter dieser Bedingung zeigt ILS stets die höchste Effizienz[14/].

Die Effizienz der Verfahren in Hinsicht auf die Reproduktion der relativen Größenordnungen der Korrelationskoeffizienten (ZMKS) zeigt die Ähnlichkeit von PAIRWISE, DEAR, MULT und ILS unter MAR (0.923, 0.984, 0.933 und 0.968) ebenso deutlich wie die außergewöhnlich schlechten Ergebnisse von LISTWISE (0.560): Lediglich HOT-DECK zeigt vergleichbar schlechte Ergebnisse (0.694). Interessant auf diesem Hintergrund ist die relative Verbesserung der Effizienz von LISTWISE unter MAT (0.882), wobei das HOTDECK die schlechtesten Ergebnisse (0.692) zeigt und DEAR deutlich absinkt (0.957). Sowohl REGRESSION als auch HAZARD zeigen unter MAT deutlich bessere Effizienzen: 0.923 bzw. 0.928. Unter MARC verschlechtert sich LISTWISE auf 0.629 und liegt damit im Mittel noch schlechter als HOTDECK mit 0.752. DEAR steigt im Mittel auf 0.978 und erzielt damit unter MARC fast durchgehend die besten Ergebnisse.

8.6 Genauigkeit der Schätzungen

Betrachtet man die absoluten Ergebnisse der Verfahren in Hinsicht auf die Schätzung fehlender Daten, so ist die der Anteil "exakt" geschätzter Daten überraschend hoch: Trotz der 7-stufigen Variablen liegt der Anteil "exakter" Schätzungen beim besten Verfahren niemals unter 28%, unter MAR liegt der Anteil im Durchschnitt für ILS bei 38%, unter MAT bei 36% und unter MARC bei 33%. Weitgehend unabhängig vom Anteil fehlender Werte erreicht HIT unter

hohem CM und N bis zu 45%. Auffällig ist das Absinken von HIT unter MARC ebenso wie die große Streuung der Verfahren (wie sie auch durch die Betrachtung der relativen Effizienzen deutlich wird). Aufgrund der Ersetzungsmethode verständlich erscheint die Empfindlichkeit von DEAR gegenüber dem Ausfallmechanismus: Unter MAT entspricht der Anteil exakter Schätzungen bei DEAR ungefähr dem Anteil von REGRESSION und gehört damit zu den schlechteren Verfahren (DEAR: MAR 39%, MAT 30%, MARC 34%).

Die mittlere Abweichung der geschätzten Daten unter MAR schwankt im Durchschnitt über alle Bedingungen und Verfahren zwischen 0.017 (DEAR) und 0.025 (REGRESSION), damit findet tendenziell eine leichte Unterschätzung statt. Unter MAT steigt die mittlere Abweichung der geschätzten Daten um mindestens eine Zehnerpotenz an: Die mittleren Unterschätzungen liegen zwischen 0.295 bei ILS und 1.05 bei HOTDECK. Unter MARC gehen die mittleren Abweichungen auf -0.01 (HAZARD) bis 0.002 (DEAR) zurück, wobei häufige Überschätzungen insbesondere bei kleineren Fallzahlen auffallen. Der Plot (Abb. 8-3) von EMD zeigte das drastische Ansteigen der mittleren Abweichung unter MAT sehr deutlich, ebenso die extremen Differenzen der Verfahren.

Auf der Ebene der Korrelationsmatrizen liegen unter MAR die mittleren Abweichungen zwischen -.012 (ILS) und 0.033 (HOTDECK), wobei duchgängig eine Unterschätzung stattfindet. Die Ausnahmen hiervon bilden DEAR, MULT und ILS bei denen durchgängig Überschätzungen stattfinden, wobei die Überschätzung mit dem Anteil fehlender Werte ansteigt. Unter MAT steigen die mittleren Abweichungen auf -0.04 (ILS) bis 0.049 (HOTDECK), wobei alle Verfahren mit Ausnahme von ILS Unterschätzungen der Korrelationskoeffizienten zeigen. ILS zieht tendenziell sehr kleine Überschätzungen nach sich, der Extremwert wird mit -0.016 bei N=300, CM=.6 und MPER=10% erreicht. Unter MARC zeigen alle Verfahren (mit Ausnahme von HOTDECK und MEAN)

deutliche Überschätzungen der Korrelationskoeffizienten: -0.038 (LISTWISE) bis -0.002 (CELLMEAN). HOTDECK und MEAN zeigen Unterschätzungen (0.0129 und 0.0002). Die größte mittlere Überschätzung erreicht LISTWISE mit -0.082 bei N=500, CM=.6 und MPER=10%.

Betrachtet man die Ergebnisse in bezug auf ZMKS, so fällt die die relative Konstanz der Ergebnisse von PAIRWISE, MULT und ILS auf: Die Koeffizienten liegen im Mittel zwischen 1.89 und 2.02, die Korrelationen zwischen geschätzten und "wahren" Korrelationsmatrizen liegen damit zwischen 0.955 und 0.97 und reproduzieren damit die "Struktur" der Matrizen weitgehend. Trotz dieser überwiegend guten Ergebnisse existieren für alle MD-Techniken Bedingungen, bei denen die Reproduktion nur ungenügend gelingt: Das Minimum erreicht ILS mit 1.20 (r=0.835) unter MARC bei N=300, CM=.6 und MPER=10%. Das schlechteste Ergebnis aller Verfahren unter allen Bedingungen erzielt LISTWISE mit .577 (r=0.52) unter MAR, N=300, CM=.2 und MPER=10%.

8.7 Zusammenfassung

Die wichtigsten Ergebnisse der Simulation können in drei Punkten zusammengefaßt werden:

1. Die MD-Verfahren bilden durch ihre Konstruktion Gruppen mit in sich relativ homogenen Reaktionsmustern. Die Ergebnisse der Clusteranalysen werden durch den Vergleich der Regressionsparameter der getrennt durchgeführten Regressionen ebenso bestätigt wie durch die Plots.

2. Die MD-Verfahren reagieren auf die vermuteten Einflußfaktoren "mittlere Interkorrelation" und vor allem "Anteil fehlender Werte" erwartungsgemäß; der Effekt der Fallzahl kann nicht so deutlich demonstriert werden.

3. Ein unter allen Bedingungen effizientes Schätzverfahren existiert nicht. Die Ergebnisse der Simulation stützen die Vermutung der Existenz der postulierten Interaktionseffekte zwischen Ausfallmechanismus und der Leistungsfähigkeit der

MD-Verfahren. Verallgemeinerungen von Ergebnissen über die Leistungsfähigkeit der MD-Verfahren ausgehend von den Resultaten unter MAR sind problematisch,

die Wahl eines "optimalen" MD-Verfahrens bei unbekanntem Ausfallmechanismus unmöglich. Die zentrale Frage der Simulation, die Frage nach der "Robustheit" der Verfahren gegenüber dem Ausfallmechanismus, muß negativ beantwortet werden: Die Clusterzugehörigkeiten ändern sich ebenso wie die Regressionsparametervorzeichen; die Interaktionseffekte in den Varianzanalysen sind dementsprechend signifikant; die Plots zeigen dieselbe Tendenz.

Neben diesen Hauptergebnissen lassen sich aus der Simulation weitere Resultate gewinnen, die für die sozialwissenschaftliche Forschungspraxis relevant sind.

So zeigt entgegen dem Großteil der Literatur (vgl. Kap. 6) die allgemein für MAR nahegelegte fallweise Löschung fehlender Werte (LISTWISE) gegenüber der PAIRWISE-Methode überwiegend enttäuschende Ergebnisse. In diesem Zusammenhang soll noch einmal auf die Ergebnisse von BROWN (1983) und FINKBEINER (1979) hingewiesen werden. FINKBEINER (1979:419) stellt fest: "If the researcher finds the CPO (PAIRWISE, R.S.) method usable in his application, it can be expected to produce relatively accurate results" und "If cost is a concern, then the MR (MEAN, R.S.) and CPO methods are to be recommended as being effective yet relatively inexpensive and easy to implement" (FINKBEINER 1979:420). FINKBEINER bezieht sich auf die Ergebnisse seiner Simulationsstudie in Hinsicht auf Faktorenanalysen; seine "CPO method" liegt bei den Ergebnissen besser als die Hauptkomponentenmethode und besser als MEAN und LISTWISE. BROWN argumentiert ebenfalls in Hinsicht auf Faktorenanalysen, allerdings mit analytischer Zielsetzung. BROWN (1983:286) stellt fest: "... PP (PAIRWISE, R.S.) performs reasonably well relative to ML provided the measure of sparseness is small" und (S.281) "Qualitatively, PP estimates become more efficient compared to DEL (LISTWISE, R.S.) as the signal-to-noise level in a variable decreases. Improvement of DEL over PP is realized only when one

variable is measured almost without error (communality above 0.80) and there is five percent or less missing data." Möglicherweise liegt in diesem letzten Ergebnis von BROWN die Erklärung sowohl für die vergleichsweise guten Resultate für PAIRWISE im Rahmen dieser Simulation als auch für die starke Verbreitung der PAIRWISE-Methode in der sozialwissenschaftlichen Forschungspraxis, die sich damit im Gegensatz zu den Ergebnissen der theoretischen Statistiker befindet (vgl. HEIBERGER 1977, LITTLE 1981).

Unter Beachtung der mehrfach erwähnten theoretischen Vorbehalte gegenüber einer Datenanalyse ohne explizite (und notwendige) Modellierung des Ausfallprozesses lassen sich aus der Simulation in ihrer Verallgemeinerbarkeit stark begrenzte Empfehlungen ableiten, falls eine Analyse ohne die im nächsten Kapitel noch darzustellenden "multiple Imputations" erfolgen muß: Die Ergebnisse der Simulation legen für homogene Korrelationsmatrizen die Verwendung von ILS, bzw. DEAR (je nach Ausfallmechanismus) nahe. Ist dies technisch nicht möglich, stellt PAIRWISE eine aktzeptable Alternative dar.

In bezug auf die "Leistungsfähigkeit" der Verfahren sollten die teilweise recht hohen mittleren Werte sollten nicht darüber hinweg täuschen, das bei bestimmten Bedingungen alle Verfahren so niedrige Werte nach sich ziehen, daß die "Struktur" der Matrix als zerstört angesehen werden kann: Die relative Größe der Koeffizienten wird nur ungenügend reproduziert; daher müssen alle Verfahren, die auf der relativen Größe aufbauen, notwendigerweise weitgehend irreführende Ergebnisse erbringen. Auch die sozialwissenschaftlich realistischen Modelle zeigen unter einigen Bedingungen so kleine Koeffizienten, daß z.B. eine Faktorenanalyse, gleich mit welcher Methode die Korrelationsmatrix berechnet wurde, mit Sicherheit sehr stark verzerrte Faktorenstrukturen zeigen würde. Diese potentiellen Gefahren durch systematische Ausfallmechanismen sollen anhand eines Nebenergebnisses der Simulation noch einmal demonstriert werden.

Die Standardbehandlung fehlender Werte im Rahmen empirischer Sozialforschung erschöpft sich in der Regel in der Entscheidung zwischen PAIRWISE und LISTWISE als Berechnungsmethoden [15/] der Korrelationsmatrizen (vgl. KIM/CURRY 1977:215).

Ein damit verbundenes Problem zeigt die Tabelle 8-25. Konventioneller Weise wird in der Pfadanalyse und in der konfirmatorischen Faktorenanalyse ein Modell verworfen, wenn die Abweichungen zwischen dem Modell und den beobachteten Koeffizienten größer als 0.1 sind (WEEDE/JAGODZINSKI 1977:321). Nun ist nur in den seltensten Fällen die Stichprobe Gegenstand des Interesses, dieses richtet sich dagegen auf Aussagen über Populationen - auch wenn inferenzstatistische Probleme zumeist vernachlässigt werden. So ist es durchaus möglich, ein Modell entsprechend dem oben genannten Kriterium anzupassen, aber die Anpassung trifft eben nur auf die jeweilige Stichprobe, nicht hingegen auf die Population zu, falls ein systematischer Ausfallmechanismus vorliegt und somit die beobachteten Korrelationen selbst mehr als .1 von den "wahren" Werten abweichen. Die Tabelle 8-20 zeigt für die beiden häufigsten Berechnungen von "MD-Korrelationen" den in der Simulation ermittelten Anteil derjenigen geschätzten ("beobachteten") Korrelationsmatrizen, bei denen die mittlere Abweichung vom wahren Wert größer als .1 ist. Der Anteil derart stark verzerrter Korrelationen erreicht unter MAT und 10% fehlender Werte bei LISTWISE 18.9%. Die Akzeptanz oder Verwerfung eines Modells wird dementsprechend häufig fehlerhaft sein. Die Tabelle bezieht sich nicht auf einzelne Koeffizienten (wie das Kriterium) sondern auf die <u>mittlere</u> Abweichung aller Koeffizienten.

Tabelle 8-25: Korrelationsmatrizen mit mittleren Abweichungen >.10

MDGP	Methode	% MD pro Variable (1-10%)	10%
MAR	LIST	0.7	2.2
	PAIR	0.0	0.0
MAT	LIST	6.7	18.9
	PAIR	0.4	1.1
MARC	LIST	3.7	7.8
	PAIR	0.0	0.0

(Angaben in Prozent)

Diese hohe Zahl potentieller Fehlentscheidungen und die durch die Simulation demonstrierte Existenz des Interaktionseffektes zwischen der Leistungsfähigkeit der MD-Verfahren und den Ausfallmechanismen stellen weitere Argumente für die Notwendigkeit der theoretischen Begründung vermuteter Ausfallmechanismen dar. Eine "automatische" Kompensation systematischer Ausfallmechanismen kann von den MD-Techniken nach den Ergebnissen der Simulation nicht erwartet werden. Damit stützen die empirischen Ergebnisse die theoretische Kritik an den Annahmen der Ersetzungsverfahren im Kapitel 5: Die auf den MAR- bzw. MARC-Annahmen basierten Techniken können bei Vorliegen anderer Ausfallmechanismen bzw. bei einer rein "mechanischen" Anwendung zu Fehlschlüssen führen. Die somit verbleibende Handlungsalternative bei der Analyse von Datensätzen, bei denen systematische Ausfallmechanismen nicht ausgeschlossen werden können, besteht in der explizit theoretisch geleiteten Ersetzung fehlender Werte. Da diese theoretischen Annahmen mit den vorhandenen Daten unüberprüfbar sind, verbleibt lediglich die Möglichkeit, unterschiedliche Hypothesen über den Ausfallprozeß in ihren Konsequenzen auf Entscheidungen über Akzeptanz oder Verwerfung substantieller theoretischer Hypothesen zu untersuchen. Den hierzu zur Verfügung stehenden Techniken ist das letzte Kapitel gewidmet.

Anmerkungen zu Kapitel 8

1. Nachdem die ersten Plots die relativen Einflußstärken der experimentellen Faktoren gezeigt hatten, wurde die Hilfsvariable COND konstruiert, deren Werte den 81 experimentellen Bedingungen entsprechen. Die Konstruktion erfolgte mit COND= (MTYPE-1)*27 + (MPER-1)*9 + (CM-1)*3 + N, so daß COND=1 der Bedingung "missing at random" (MTYPE=1), Anteil fehlender Werte pro Variable gleich einem Prozent (MPER=1), mittlere Interkorrelation gleich einem r von 0.2 (CM=1) in einer Stichprobe vom Umfang 100 (N=1) entspricht. In den Plots wechselt der Ausfallmechanismus also alle 27 Einheiten, der Anteil fehlender Werte alle 9 Einheiten, die mittlere Interkorrelation alle 3 Einheiten und die Stichprobengröße mit jeder Einheit auf der Hilfsvariablen COND. Die mit SAS- GRAPH erstellten Plots stellen die Mittelwerte der jeweiligen Verfahren über die 10 Wiederholungen pro Bedingung dar.

2. Die sehr ähnlichen Plots der anderen Variablen werden hier auf Grund ihres Umfangs nicht wiedergegeben.

3. Da aufgrund der großen Streuung bei der Skalierung der Plots mehrere Ziele der Darstellung unvereinbar waren, wurde auf eine exakte Kennzeichnung der Verfahren in den Plots verzichtet, um einen deutlichen Eindruck von der Stärke der Effekte vermittelt zu können.

4. Die Hypothese der Parallelität der Profile konnte durch einen formalen Test (mit SPSS-X MANOVA) für alle abhängigen Variablen mit $p < 0.001$ verworfen werden; zur Profilanalyse vgl. MORRISON (1976:205-216) und SPSS-X User's Guide (1983:535-539).

5. Weiterhin wurde eine Vergleichsanalyse mit SAS durchgeführt, bei der eine andere Datenstruktur zu Grunde gelegt wurde. In dieser hier nicht wiedergegebenen Analyse wurde aus der Simulationsergebnisdatei eine zunächst eine Datei mit 100 Fällen und 580 Variablen erstellt. Die 100 "Fälle" stellen die MD-Verfahren mit ihren Wiederholungen dar. Den Fällen wurden die abhängigen Variablen der Simulation unter den 81 experimentellen Bedingungen zugeordnet, damit stellen die Variablen in dieser Datenstruktur ein "Verlaufsprofil" des jeweiligen MD-Verfahrens dar. Um die Mittelwerte der Verfahren als Einheiten zu gewinnen wurde diese Datei aggregiert, dann wurde die Clusteranalyse mit der Ward- Methode durchgeführt. Die Ergebnisse dieser Art der Analyse, bei der jeweils alle abhängigen Variablen zur Clusterbildung benutzt wurden, stimmen mit den Ergebnissen der im Text beschriebenen Vorgehensweise weitestgehend überein.

6. Die Berechnungen wurden mit SPSS-X 2.1 PROXIMITIES durchgeführt. Die Ergebnisse sind gegenüber der Verwendung anderer Distanzmaße stabil.

7. Vgl. zur Profil-Cluster-Analyse ausführlich SCHLOSSER (1976: 51-151), zu hierarchischen Clusterverfahren STEINHAUSEN/LANGER (1977:73-100).

8. Im Vergleich ergaben sich nur minimale Differenzen für die verschiedenen abhängigen Variablen in den Clusterlösungen, daher werden nur die Ergebnisse für ZMKS berichtet. Das selbe gilt für die Ergebnisse bei den Abweichungen der geschätzten Daten, bei denen nur die Dendrogramme für die Variable HIT wiedergegeben werden.

9. Die umgekehrte Fragestellung, inwieweit die MD-Verfahren auf Grund ihrer Reaktionsprofile korrekt identifiziert werden können, wurde durch eine Reihe von Diskriminanzanalysen untersucht. Naheliegend ist hierbei die Benutzung der Reaktionsprofile als Prädiktoren. Da bei dann lediglich 100 Fällen und 27 Variablen ein Overfit naheliegt, wird diese Analyse hier nicht dargestellt; tendenziell konnten die Ergebnisse der Clusteranalysen bestätigt werden.

10. Dem Design entsprechend hätten univariate (Meßwiederholungs-) Varianzanalysen (GEISSER/GREENHOUSE 1959, WINER 1971:559) gerechnet werden können. Diese Varianzanalysen setzen aber die Homogenität der der Varianz-Kovarianzmatrizen der einzelnen Faktorstufen (vgl. BORTZ 1977:437-452) voraus. Die Annahme der "compound symmetry" (Diagonale der Matrix enthält $Sigma^2$, die nichtdiagonal Elemente sind gleich Rho * $Sigma^2$, vgl. auch WINER 1971:596) wurde für jede Variable einzelnen mit SPSSX-MANOVA über den BARTLETT-Sphericity-Test getestet. Zusätzlich wurde die Annahme der Homogenität der Varianzen über die Wiederholungen mit dem FMAX-Test überprüft. Ausnahmslos alle Tests zeigten hochsignifikante Verletzungen der Annahmen. Daher erfolgte die Auswertung mit dem multivariaten Ansatz (vgl. auch O'BRIEN/KAISER 1985:317).

11. Eta-Quadrat ergibt sich hier aus 1- Lambda. Die Summe aller Eta-Quadratkoeffizienten kann bei einer MANOVA eins überschreiten (vgl.TABACHNICK/FIDELL 1983:244).

12. Die Regressionen wurden getrennt, um die Aufnahme von 9 Dummy-Variablen für die Programme und 2 Dummy-Variablen für die MDGP und deren Interaktionsterme zu vermeiden, und somit eine überschaubare Darstellung zu gewährleisten.

13. Aufgrund des orthogonalen Designs sind die Beta-Quadrate direkt als erklärte Varianzen interpretierbar.

14. Auf dem Hintergrund der Arbeit von SCHMEE/HAHN (1979) ist dieses Ergebnis wenig überraschend: Eine ILS-Technik wird von ihnen explizit für "Censoring-Probleme" entwickelt.

15. vgl. HOLM (1979:150): "Für die abhängigen Variablen (...) lösen wir das Nonresponse-Problem so wie in Band 2 (...) beschrieben. Dies ist die Methode des 'paarweisen Ausscheidens'."

9. Multiple Imputation (MI)

D.B.RUBIN entwickelt seit 1977 in einer Reihe von Arbeiten (RUBIN 1977,1978,1979,1980,1983) das Konzept der "multiple imputation"[1/]. Ausgehend von der Tatsache, daß jede Analysestrategie beim Vorliegen eines MD-Problems ein Modell des Ausfallprozesses verwenden muß (in der Regel ein implizites MAR-Modell), besteht das Vorgehen nach einer multiplen Imputationstrategie aus drei Schritten.

Zunächst müssen explizite Modellierungen des Ausfallprozesses erfolgen. Entsprechend den Annahmen des Modells werden dann Ersetzungen der fehlenden Werte mit einem nichtdeterministischen MD-Ersetzungsverfahren vorgenommen (das Ergebnis ist ein vollständiger Datensatz, zu dessen Analyse Standardtechniken benutzt werden können). Dieser Schritt wird nun I-mal wiederholt; da das MD-Verfahren bei Wiederholungen nicht die selben Ersetzungen vornimmt, entstehen I vollständige (unterschiedliche) Datensätze. Im letzten Schritt werden die aus den I-Datensätzen berechneten Statistiken miteinander verglichen. Der Vergleich erlaubt (falls <u>ein</u> Modell des Ausfallprozesses benutzt wurde) die Simulation der Verteilung der Statistiken bei gegebenen Daten und gegebenem Ausfallprozeß. Werden hingegen unterschiedliche Annahmen über den Ausfallprozeß gemacht, d.h. also: mehrere explizite Modellierungen vorgenommen, erlaubt der Vergleich die Beurteilung der Sensitivität der Ergebnisse gegenüber unterschiedlichen Annahmen über den Ausfallprozeß.

RUBIN liefert die Grundlagen für eine explizite Bayes-Analyse beim Vorliegen eines MD-Problems. Die Entwicklung von Standard- Bayes-Techniken (insbesondere die Einbeziehung von Hyperparametermodellen bei systematischen Ausfällen) stellt für RUBIN den Schwerpunkt seiner Arbeit (und auch der Anwendung der MI) dar (vor allem RUBIN 1978:26). Die hier erfolgende Darstellung orientiert sich fast ausschließlich an praktischen Aspekten der MI, da

zum einen für systematische Ausfallmechanismen keine ausgearbeiteten Modelle[2/] existieren und selbst unter MAR-Annahmen die Berechnung der a-posteriori-Verteilung analytisch unlösbar werden kann. Zur Durchführung einer MI ist keine (im Sinne der Bayes- Statistik) explizite Modellierung erforderlich (DEMPSTER 1978:30, RUBIN 1980:13).

Obwohl der Schwerpunkt der bisherigen Arbeiten zu MI auf der Simulation der a-posteriori-Verteilung einer Statistik bei gegebenen Daten und angenommenem Ausfallmechanismus liegt und nur rudimentäre Ansätze zur Beurteilung der Sensitivität gewonnener Statistiken gebenüber unterschiedlichen Annahmen über den Ausfallmechanismus existieren, soll der MI-Ansatz etwas ausführlicher dargestellt werden, da MI eine relativ neue Methode darstellt, die erstmals eine explizite (im Sinne inhaltlicher Annahmen) Modellierung der Ausfallprozesse fordert.

9.1 MI unter einem Ausfallmechanismus

Die Kompensation der Varianzunterschätzung eines vervollständigten Datensatzes mit MI läßt sich am einfachen Fall der Mittelwertschätzung unter MAR darstellen. Werden die fehlenden Werte über einen Zufallsmechanismus (z.B. ein Hotdeck) ersetzt, dann ist zwar (bei wiederholten Ziehungen) der Erwartungswert für den Mittelwert der wahre Mittelwert, aber das Konfidenzintervall um den Mittelwert wird um die Wurzel des Quotienten der Anzahl vorhandener Werte und der Stichprobengröße zu klein sein, da die vorhandene Stichprobe kleiner ist als das n in der Berechnung des Konfidenzintervalls. Werden nun MI vorgenommen (d.h. werden voneinander unabhängige Wiederholungen der Ersetzungen durchgeführt) und für jeden der vervollständigten Datensätze die Statistiken berechnet, dann ist der Erwartungswert des Mittelwertes natürlich gleich dem wahren Mittelwert, aber das aus der mittleren Varianz der einzelnen Imputations und der Varianz der Mittelwerte über die Imputations berechnete Konfidenzintervall[3/] nähert sich dem wahren Wert, wenn es auch immer noch unterschätzt wird[4/].

Um das Konfidenzintervall korrekt zu berechnen, schlägt RUBIN (1980:79-83) eine Korrektur vor. Diese "adjusted multiple imputations" werden dadurch erzeugt, daß für jede der unabhängigen Imputations i ein Bias b(i) aus einem N(0,S)-Generator gezogen wird, wobei N(0,S) der a-posteriori-Verteilung der Differenz des Populationsmittelwertes und des Stichprobenmittelwertes entspricht[5/]. Der Bias b(i) wird für jeden kompletten zu vervollständigenden Datensatz einmal bestimmt und zu jedem der zufällig gezogenen Werte der vorhandenen Fälle addiert, der einem unvollständigen Fall zugewiesen werden soll. Der Erwartungswert des Mittelwertes der Mittelwerte der einzelnen vervollständigten Datensätze entspricht wieder dem wahren Mittelwert, der Erwartungswert der Varianz entspricht der wahren Varianz und das Konfidenzintervall ist korrekt, falls die Zahl der unabhängigen Imputations hinreichend groß ist[6/].

9.2 MI unter verschiedenen Ausfallmechanismen

Der Vorteil, den MI bieten, zeigt sich im oben angeführten Beispiel, bei dem die tatsächlichen Parameter bekannt sind, nicht. Dieser wird erst bei Anwendungen deutlich, in denen die Parameter unbekannt sind. Dies gilt insbesondere für diejenigen Parameter, die den Ausfallmechanismus bestimmen.

Die oben dargestellte Vorgehensweise[7/] bei MI läßt sich für den Fall, daß kein Nonresponse-Bias erwartet wird und normalverteilte Variablen vorliegen, folgendermaßen durchführen (RUBIN 1979:529; 1980:84):

Gegeben sei ein Sample mit n Einheiten, von denen die Daten für n_1 Einheiten vorhanden sind. Für jede unabhängige Replikation (i=1,...,I) werden die folgenden Schritte durchgeführt:

1. Ziehung einer Zufallsvariablen x_i aus einem Chi-Quadrat -Generator mit n1-1 Freiheitsgraden,
2. Berechnung von S_i^2 als $s1^2/ x_i$,
3. Ziehung einer Zufallsvariablen z_i aus einem N(0,1)-Generator,
4. Berechnung von $b_i= z_i(S_i^2/n_1)^{1/2}$,
5. Zufälliges Ziehen von $(n-n_1)$-Werten aus den vorhandenen Daten,
6. Dividieren jedes Wertes durch $x_i^{1/2}$
7. Addieren von b_i zu jedem gezogenen Wert,
8. Ersetzen der MD durch die berechneten Werte.

Dieses Verfahren läßt sich leicht modifizieren, um unterschiedliche Annahmen über den Ausfallmechanismus zu modellieren und in ihren Konsequenzen abzuschätzen.

- Eine Annahme über einen von den Respondenten verschiedenen Mittelwert läßt sich über eine Konstante k_1 ("shift") modellieren, wobei im Schritt 7 statt b_i der Bias $b_i + k_1S_i$ addiert wird. Der Erwartungswert der Differenz der Mittelwerte beträgt k_1 Standardabweichungen.

- Eine Annahme über eine von den Respondenten verschiedene Streuung ("scale") läßt sich über die Division jedes Wertes durch $(x_i/k_2)^{1/2}$ in Schritt 6 modellieren. Der Erwartungswert der Standardabweichung liegt bei dem k_2-fachen der Standardabweichung der vorhandenen Werte.

- Eine Annahme über eine von den Respondenten verschiedene Verteilungsform ("shape"), die durch eine Funktion $k_3(.)$ dargestellt wird, läßt sich durch eine der Funktion proportionale Ziehung im Schritt 5 modellieren (als Beispiel für eine Funktion $k_3(.)$ gibt RUBIN 1979:530 $k_3(y_i) = y_i$ an, die Ziehung erfolgt also proportional zur Größe des Wertes; RUBIN erwähnt als Anwendung Einkommen).

Durch die Variation der drei Biaswerte lassen sich eine Vielzahl unterschiedlicher Mechanismen leicht simulieren und beschreiben: Eine effizientere Beschreibung eines Ausfallmodelles als durch drei Koeffizienten ist kaum vorstellbar. Beispielsweise läßt sich das MAR-Modell durch (k_1=0, k_2=1, k_3=const.) zusammenfassen.

Durch den Vergleich der Statistiken der unter verschiedenen Annahmen über den Ausfallmechanismus vervollständigten Datensätze lassen sich Aussagen über die Sensitivität der Ergebnisse gegenüber den Ausfallmechanismen gewinnen. MI liefert keine eindeutigen Resultate, sondern zeigt die Spannweite möglicher Ergebnisse auf[8/].

RUBIN betont die Notwendigkeit komplizierterer Modelle im multivariaten Fall. Als einfache Modifikation der oben beschriebenen Technik für große Stichproben schlägt RUBIN (HERZOG/RUBIN 1983: 224) (um die Ziehung aus einem Wishart-Generator zu vermeiden) die Vernachlässigung der Schritte 1 und 2 sowie die Ersetzung des Schrittes 3 durch die Ziehung aus einem multivariaten Normalverteilungsgenerator (mit den Mittelwerten und Kovarianzen des vorhandenen Samples) vor. Dementsprechend muß k_1 durch einen "shift vector", k_2 durch eine "scale transformation matrix" und k_3 durch eine Skalarfunktion eines Vektors ersetzt werden.

Abschließend soll noch auf eine von RUBIN (1980:86-87) in diesem Zusammenhang beschriebene Ersetzungsmethode für den Fall kategorialer Variablen hingewiesen werden, den "Bayesian Bootstrap" (zum Hintergrund vgl. RUBIN 1981:131): Die Variable Y kann K verschiedene Werte d annehmen. Für jede der unabhängigen Replikationen der Mehrfachersetzungen werden $K+n_1-1$ Werte u aus einem U(0,1)-Generator gezogen, geordnet und $K + n_1$ Differenzen gebildet ($u_0 = 0$; $u_{K+n1+1} = 1$). Es sei n_{1k} die Zahl der beobachteten Fälle mit der Ausprägung d_k. Die Summe der ersten $n_{1k}+1$ Differenzen sei q_k, usw. bis zur Summe q_K der letzten $n_{1k}+1$ Differenzen. Für jeden der n_0 fehlenden Wert wird der Wert d_k mit der Wahrscheinlichkeit q_k, $k = 1,\ldots,K$ zufällig gezogen.

9.3 Sensitivitätsuntersuchungen mit MI

Die Idee, mehrfache Ersetzungen für fehlende Daten vorzunehmen, um den Effekt verschiedener Ausfallmodelle zu testen, ist keineswegs neu. Explizit findet sich der Vorschlag z.B. schon bei FINNEY (1974:10). RUBIN hat dann jedoch die theoretischen Grundlagen einer vollständigen Bayes'-Analyse für die MI formuliert, deren Durchführung in der Praxis jedoch fast unmöglich ist, selbst wenn die Beschränkung auf "ignorable mechanisms" erfolgt: Selbst mit einer einzigen Variablen ist der Arbeitsaufwand außerordentlich hoch (HERZOG/RUBIN 1983:242). DEMPSTER (1978:30) betrachtet MI denn auch weniger als "... something to be put into the hands of the average user." Der Ansatz der expliziten Formulierung wird, solange keine Standardmodelle verfügbar sind, nur selten eine Anwendung bei einem konkreten Problem finden. Die praktische Nützlichkeit der Entwicklung von notwendigerweise einfachen Standardmodellen systematischer Ausfälle könnte aus der Sicht von Soziologen allerdings generell bezweifelt werden, da unterschiedliche Ausfallmechanismen und unterschiedliche Mischformen bei wechselnden inhaltlichen Theorien zumindest möglich erscheinen.

Die Durchführung der MI mit (in Hinsicht auf die Bayes-Grundlagen) impliziten Techniken, wie z.B. einem Hotdeck, ist hingegen vergleichsweise einfach. Wenn auch gegen die ursprünglichen Intentionen des MI-Ansatzes gerichtet, ist eine von den theoretischen Grundlagen der MI abgelöste Anwendung möglich. Bei dem derzeitigen Stand der Arbeiten zu systematischen Ausfällen scheint diese Anwendung der MI gerade für den "average user" die einzige Möglichkeit zu sein, ein MD-Problem in seinen möglichen Auswirkungen abzuschätzen.

9.3.1 Praktische Probleme bei der Durchführung

Die praktische Durchführung einer MI läßt sich in drei Schritten beschreiben[9/]. Zunächst müssen die unterschiedlichen Annahmen über den Ausfallprozeß explizit formuliert und in eine Form gebracht werden, die in das Imputationsystem eingehen kann (vgl. hierzu die drei Koeffizienten für das einfache Imputationsystem in Kap. 9.2). Dann müssen für alle explizierten Modelle die MI vorgenommen werden. Um die Variabilität innerhalb der Modelle beurteilen zu können, müssen auch innerhalb der Modelle Wiederholungen vorgenommen werden. Der letzte Schritt besteht aus dem Vergleich der mit den vervollständigten Datensätzen gewonnen Statistiken. Alle Schritte bieten z.Z. noch Schwierigkeiten.

Wie schon erwähnt, existieren bis auf extrem vereinfachte Beispiele keine brauchbaren Standardmodelle, an denen zumindest eine Orientierung möglich wäre. Die Modelle müssen völlig neu entwickelt werden. Da auch nur vereinzelte und in der vorhandenen Form in diesem Zusammenhang nicht brauchbare sozialwissenschaftliche ("inhaltliche") Modelle existieren, bleibt als relativ einfache Möglichkeit nur die Entwicklung simpler Modelle, die den "ungünstigsten Fall" eines systematischen Ausfalls für die jeweilige Fragestellung erfassen sollen. Diese (von DEMPSTER 1978:30 als "worst reasonable models (WRM's)" benannten) Modelle können nur die Extremwerte[10/] der interessierenden Statistiken zeigen, eine "realistische" Modellierung der Ausfallprozesse erfolgt nicht. Trotzdem stellen "WRM's" durch die durch sie möglichen Sensitivitätstest einen brauchbaren ersten Versuch dar.

Der zweite Schritt in einer MI besteht aus der Anwendung des Imputationsystems. Hier entstehen eine Reihe insbesonderer technischer Probleme. Zunächst einmal existieren bisher keine allgemein verfügbaren Imputationsysteme. Die notwendigen Berechnungen sind zwar mit erheblichem Aufwand fast immer mit einem der Standardpakete (SPSS-X, BMDP, SAS) durchführbar, aber spätestens bei der Notwendigkeit, eine Reihe vollständiger Datensätze zu verwalten, wird der Anwender in der Regel die "Kosten" der MI kaum noch zu tragen bereit sein.

Weiterhin erfordern schon sehr einfache "WRM's", wie z.B. das in der Simulation in Kapitel 7 verwendete MARC-Modell, die Erstellung spezieller Programme.

Der letzte Schritt einer MI besteht im Vergleich der gewonnenen Statistiken. Wie schon aus der Darstellung des einfachen Problems der Beurteilung der "Ähnlichkeit" zweier Korrelationsmatrizen (in Kap. 7) deutlich geworden ist, fehlen für die auf diese Art "verbundenden" Stichproben formale Beurteilungsgrößen weitgehend; wenn auch manche Probleme relativ einfach gelöst werden könnten, z.B. durch einen Vergleich der "goodness-of-fit"-Indices in LISREL-Modellen[11/].

Neben der Entwicklung von Sensitivitätsindices bleibt vor allem die Frage nach der Änderung der Entscheidungen über Akzeptanz oder Verwerfung einer konkreten Hypothese bei Annahme verschiedener Ausfallprozesse.

Zusammenfassend muß betont werden, daß die Durchführung einer MI z.Z. mit einer so großen Zahl praktischer Probleme behaftet ist, daß die Anwendung im Rahmen einer Standardanalyse weitgehend ausscheidet. Da MI bisher jedoch die einzige Möglichkeit zur Abschätzung der Konsequenzen eines MD-Problems darstellen, muß die Entwicklung sozialwissenschaftlicher Ausfalltheorien und eines effizienten[12/] Imputationsystems als dringend notwendig betrachtet werden.

9.3.2 Abgrenzung und Konsequenzen für die Forschungspraxis

In der Literatur finden sich gelegentlich Beispiele, die den Einfluß eines MD-Problems auf die Ergebnisse dadurch abzuschätzen versuchen, daß die Analyse mit verschiedenen MD-"Behandlungen", z.B. "pairwise"/"listwise"/"mean substitution", wiederholt wird. Stabilität der Ergebnisse soll die Vernachlässigbarkeit des MD- Problems garantieren (z.B. TABACHNIK/FIDELL 1983:72). Die mehrfache Durchführung statistischer Analysen mit unterschiedlichen MD-Techniken, beispielsweise der Vergleich der Ergebnisse einer Faktorenanalyse, deren Korrelationsmatrix einmal mit "pairwise" und dann mit "listwise" berechnet wurde, stellt allerdings keine MI dar. Zunächst sind simple

Ausfallmechanismen denkbar, deren Konsequenzen in beiden Fällen extreme Verzerrungen bedingen könnten (dieser "Test" ist zumindest nicht hinreichend zur Ablehnung einer Hypothese eines systematischen Ausfalls). Weiterhin ist weder eine Abschätzung der Variabilität innerhalb eines Ausfallmodells noch der Variabilität unter verschiedenen Modellannahmen möglich. Da keine (oder nur simple) Ersetzungen vorgenommen werden, bleibt dieser "Vergleich" ohnehin nur auf Analysetechniken beschränkt, die über einfache MD-Techniken verfügen. Das sozialwissenschaftlich wichtigere Argument beruht auf der nicht erfolgten expliziten (handlungstheoretischen) Modellierung alternativer Ausfallmechanismen. Der einfache Vergleich enthält keine sozialwissenschaftlich relevanten Hypothesen über den Ausfallmechanismus. Bringt der einfache Vergleich Differenzen, so muß zu ad-hoc-Annahmen und Techniken gegriffen werden, die weder einen Erkenntniszuwachs über systematische Ausfälle noch die Angabe der Ergebnisse unter einem "WRM" erlauben. Zu einer unter sozialwissenschaftlich explizierten Ausfallmodellen erfolgenden MI gibt es zur Zeit keine brauchbaren Alternativen.

MI bieten ebenso wenig wie alle anderen in dieser Arbeit dargestellten Methoden eine "Lösung" des MD-Problems. Es dürfte deutlich geworden sein, das alle Versuche einer Handhabung von MD auf eine Modellbildung nicht verzichten können. Gerade in der Entwicklung expliziter handlungstheoretisch begründeter Ausfallmodelle liegt eine Aufgabe für Soziologen bei der notwendigen Interpretation der Datenerhebung als sozialem Prozeß (vgl. ESSER 1983). Wie schon im Zusammenhang mit dem Scheitern der klassischen Fehlermethodologie des Interviews deutlich wurde, erfordert die Einbeziehung theoretischer Modelle in die Datenanalyse sowohl die Abkehr von der "Variablensoziologie" als auch von dem in der sozialwissenschaftlichen Forschungspraxis noch immer vorherrschenden naiven Empirismus: "Beobachtung ist stets <u>Beobachtung im Licht von Theorien..."</u> (POPPER 1971:31).

Anmerkungen zu Kapitel 9

1. Eng mit diesem Konzept verbunden ist das Konzept der "Repeated Replication Imputation Procedure" (RRIP) von KISH (1983) für Unitnonresponse.

2. "... there is essentially no literature on calculating posterior distributions under such special models" (RUBIN 1978:25)

3. Die Formeln finden sich bei RUBIN (1979:5211-525) und RUBIN (1980:73-75). Da dieser Aspekt der MI hier nicht im Mittelpunkt des Interesses steht, wird auf die Wiedergabe verzichtet.

4. Die Unterschätzung läßt sich durch die Benutzung von Stichprobenwerten anstelle der unbekannten Populationsparameter erklären, vgl. RUBIN (1979:525-527) und RUBIN (1980:78-79).

5. Wobei $S= s_1^2/n_1$; Varianz und Anzahl der vorhandenen Werte

6. Eine Tabelle der relativen Effizienz der "adjusted multiple imputations" in Abhängigkeit von der Anzahl unabhängiger Wiederholungen findet sich bei RUBIN(1980:82).

7. Diese Verfahrensweise unterscheidet sich nur durch die Ziehung der Chi-Quadratvariablen von der oben dargestellten Technik. Die Chi-Quadratvariable dient nur zur Korrektur in kleinen Stichproben, vgl. RUBIN (1979:530).

8. Vgl. WAINER (1978:31) :"... is more for telling us what we do not know, and perhaps cannot know, than anything else."

9. Die hier aufgeführten Schritte stimmen nicht mit den drei Schritten bei RUBIN (1980:122-133) überein: RUBIN unterscheidet "imputation task", "estimation task" und "modelling task". Die Modellierung bezieht sich auf die explizite Bayes-Analyse des für die Imputation benötigten Modells, die "estimation task" auf die Berechnung der posteriori-Verteilung des Modells und die Imputation auf die Ziehungen aus dieser Verteilung. Da hier "implizite" Modellierungen erörtert werden, ergeben sich Unterschiede.

10. Auf diesern Hintergrund lassen sich die Formeln zu Minimal- und Maximalwerten für Mittel- und Anteilswerte beim Vorliegen von Nonresponse (z.B. COCHRAN 1972:417-419, ZEH 1976:28-34, AIKEN 1981:1031-1036) als einfache Beispiele für "WRM's" interpretieren, bei denen jedoch keine Modellierung des Ausfallprozesses erfolgt. So lassen sich auch die in Kapitel 4 beschriebenen Methoden bei bewußter Verletzung der Annahmen als simple "WRM''-Technik ohne Modellierung betrachten. Beide "Extremwert'-Methoden liefern zudem keine oder nur simple Ersetzungen, so daß eine Anwendung auf komplexere Probleme nicht (direkt) möglich ist.

11. Ansätze zum Vergleich der Ergebnisse von Faktorenanalysen lassen sich aus den bei HARMAN (1976:341-347) beschriebenen Indices entwickeln.

12. Ein Problem eines Mi-Systems liegt in der Notwendigkeit begründet mit Pointern zu arbeiten (RUBIN 1978:21, 1980:120) um umfangreichere Datensätze noch verarbeiten zu können. Auch aus dieserm Grund empfiehlt sich die Implementierung als externe Prodzedur innerhalb eines Standardpaketes.

A. Technische Details der Simulation

Die Beurteilung der Ergebnisse einer Simulation erfordert eine Anzahl von Informationen über eine Reihe technischer Einzelheiten der Durchführung. HOAGLIN/ANDERSON (1975) schlugen einige Standardinformationen vor, die für jede veröffentlichte Simulation verfügbar sein sollten. Wie HAUCK/ANDERSON (1984:214-215) in ihrem Simulationsstudien-Survey bemängeln, werden diese Informationen von den wenigsten erschienen Arbeiten bereitgestellt. In diesem Anhang sollen daher wesentliche technischen Details dargestellt werden, um die Ergebnisse gegebenenfalls auch aus technischer Sicht kritisierbar zu machen.

A.1 Generierung der Datensätze

Die Datensätze, aus denen die Korrelationsmatrizen berechnet wurden, wurden durch das im Anhang B dokumentierte FORTRAN- Programm SIAM06 erzeugt.
Zur Generierung der Pseudo-Zufallszahlen wurden Subroutinen der NUMERICAL ALGORITHM GROUP (NAG) FORTRAN LIBRARY Mark 10 verwendet. Die allen Generatoren zugrunde liegende Subroutine G05CAF benutzt die "multiplicative congruential method" (KENNEDY/GENTLE 1980:136-149). Die Subroutine G05CBF wurde zum Setzen eines wiederholbaren Status von G05CAF benutzt[1/]. Die Erzeugung der Datensätze mit vorgegebener Kovarianzstruktur erfolgte mit den Subroutinen G05EAF und G05EZF. Die Erzeugung der multivariatnormalverteilten Variablen läuft über die Dekomposition der Kovarianzmatrix C, so daß $LL^T = C$. G05EAF erzeugt den Referenzvektor R (REFER im Programm), der den Mittelwertvektor A, die Anzahl der Dimensionen sowie L enthält. R wird an G05EZF übergeben. G05EZF erzeugt unabhängig standardnormalverteilte Variablen X und berechnet den Rückkehrvektor A + LX, der die geforderte Verteilung besitzt[2/]. Die Generatoren wurden durch ein Testprogramm mit 10 Variablen mit einer mittleren Interkorrelationen von 0 bis .9 in r=.1-Schritten bei n=100 und je 30 Replikationen überprüft.

---

1. In der ursprünglichen Testversion des Simulationsprogramms wurde die Subroutine G05CCF zum Setzen eines nichtwiederholbaren Status von G05CAF verwendet. Umfangreiche Tests erbrachten eine sehr hohe serielle Korrelation zwischen aufeinander folgenden "Datensätzen" und demzufolge auch der Korrelationsmatrizen. Dieser Effekt liegt in der Implementierung von G05CCF begründet: Die Subroutine liest den Wert der "real time clock" des Systems und leitet daraus den Startwert N der Subroutine G05CAF ab. Der Zugriff zur Systemzeit geschah vermutlich innerhalb des Simulationsprogramms zu schnell aufeinander, so daß nur minimale Änderungen der Systemzeit zu nur minimalen Veränderung von N und damit der Sequenz führten. Nach der Ersetzung von G05CCF durch G05CBF konnte keine serielle Korrelation mehr nachgewiesen werden.
2. Vgl. NAG (1983) NAGFLIB Dokumentation zu den genannten Routinen. Eine Darstellung der Methode findet sich bei KENNEDY/GENTLE (1980:228-229). Ein FORTRAN - Programm findet sich bei HURST/KNOP (1972), vgl. auch PAGE (1974).

Neben der Fallzahl (n=100, 300, 500) und der (homogenen) mittleren Interkorrelation (CM =.2, .4 und .6) wurde als dritter Faktor der Anteil fehlender Fälle (MPER) variiert. Drei Anteilswerte wurden erzeugt: 1%, 5% und 10% fehlende Werte pro Variable. Die Ausfallmechanismen wurden so programmiert, daß der Zielwert des Anteils fehlender Werte in jedem Datensatz exakt erreicht wurde. Hierzu wurde eine sequentielle Ziehung "ohne Zurücklegen" programmiert (vgl. KENNEDY/GENTLE 1980:238). Die Ziehung erfolgte für jede Variable unabhängig von allen anderen Variablen, so daß auch Fälle mit fehlenden Werten auf allen internen Variablen möglich waren.
Die drei realisierten Ausfallmechanismen (Missing at Random, Hazard-Modell, Missing at Random within Classes) wurden durch drei leicht unterschiedliche Generatoren simuliert. Der Ausfallmechanismus für MAR bestand aus einem einfachen sequentiellen Ziehungsvorgang, im MARC-Modell wurden für die beiden Gruppen mit je unterschiedlicher Kovarianzstruktur zwei verschiedene Generatoren des MAR-Typs verwendet. Für das Hazard-Modell wurde die zusätzliche Variable (mit dem Index 14) erzeugt, deren Korrelation mit den externen Variablen als 0.15 und mit den internen Variablen als (CM + .1) vorgegeben wurde. Die Implementierung umfaßte vier Schritte:

- Erzeugung der erweiterten Datenmatrix mit der Variablen 14, - Sortieren der Fälle nach V14,
- Bestimmung der doppelten Anzahl fehlender Werte (N2),
- Sequentielle Ziehung aus den oberen N2 Fällen für jede Variable.

A.2 Iterationszahl

Eine exakte Begründung der Iterationszahl pro Zelle kann nur über die angestebte Mächtigkeit der beabsichtigten statistischen Tests erfolgen (HAUCK/ANDERSON 1984:215). Da diese Studie keine Modellbildung beabsichtigte und keine expliziten apriori-Hypothesen geprüft werden sollten, die Arbeit somit lediglich explorativen Charakter besitzt, ist eine Legitimation über die Mächtigkeit nicht möglich. Somit verbleiben rein pragmatische Gründe: Rechenzeit und Plattenspeicherplatz. In Hinsicht auf die Stabilität der Ergebnisse ist natürlich eine möglichst hohe Iterationszahl wünschenswert. Unter Berücksichtigung der benötigten Rechenzeit wurde die Iterationszahl auf zehn festgesetzt. Damit wurden in der Simulationsstudie in den 81 Zellen des Designs mit je zehn Wiederholungen insgesamt 810 unabhängige "Datensätze" mit 100 bis 500 "Fällen" und je zehn Variablen erzeugt. Da jeder Datensatz mit Zehn verschiedenen MD-Techniken untersucht wurde, ergab sich eine Gesamtzahl von 8100 geschätzten Korrelationsmatrizen bei 810 verschiedenen wahren Korrelationsmatrizen.

A.3 Einzelheiten des Programms

Das zur Simulation benutzte Programm ist in Programmsegmente eingeteilt, die weitgehend voneinander unabhängig sind. Es handelt sich um die Subroutinen des ursprünglichen Simulationsprogramms. Auf die Programmierung von Unterprogrammen mußte in Hinsicht auf die Rechenzeit verzichtet werden. Die Parameterübergabe der Datenarrays beansprucht eine beachtliche Rechenzeit. Nachdem Tests extreme Rechenzeiten erbrachten, wurde das Programm vollständig umgeschrieben, um die Parameterübergabe zu minimieren (dies schien auch das Programm lesbarer zu halten, als die Version mit einem umfassenden "COMMON BLOCK"). Diese dritte Version des Programms lief ca. 45% schneller.

Die meisten Programmsegmente des Simulationsprogramms verwenden NAGFLIB-Subroutinen. Die Entscheidung, die in bezug auf die Rechenzeiten nicht optimalen NAGFLIB-Programme zu nutzen, wurde getroffen, um die Programmentwicklungs- und Testzeit zu verkürzen. Die Ergebnisse sämtlicher Programmsegmente des Simulationsprogramms wurden mit Ergebnissen anderer NAGFLIB - Routinen verglichen. Alle Routinen wurden mit Ergebnissen aus der Literatur überprüft. Um möglichst numerisch genaue Ergebnisse zu erhalten wurden alle REAL-Variablen als DOUBLE-PRECISION deklariert, bei den kritischen Variablen wurde E12.5 als Ein/Ausgabeformat benutzt.
LISTWISE und PAIRWISE wurde mit den NAGFLIB-Routinen realisiert; MEAN verwendete alle vorhandenen Werte einer Variablen für die Mittelwertersetzung. CELLMEAN baute für jede zu unvollständig Variable eine CROSSBREAK-Tabelle (mit 2*5 Zellen) aus den externen Variablen V1 und V2 auf. War die Zelle unbesetzt, wurde für die entspechende Zelle der 10-Zellen Tabelle die zweite Variable kollabiert. Wäre die Zelle immer noch leer, wäre für die Zelle der Mittelwert der Variablen verwendet worden, dieser Fall trat aber während der Simulation nicht auf. HOT-DECK benutzte ebenfalls die Variablen V1 und V2 als Imputationklassen, als Startwerte wurden die Variablenmittelwerte benutzt. Fehlende Daten wurden variablenweise ersetzt und nicht fallweise, d.h. ein unvollständiger Fall wurde eventuell durch gültige Werte mehrerer verschiedener Fälle komplettiert. Das Programmsegment REGRESSION schätzt fehlende Werte mit einer linearen Einfachregression mit der jeweils am stärksten korrelierenden Variablen, wobei die PAIRWISE-Korrelationsmatrix den Ausgangspunkt darstellt. REGRESSION weist gegenüber BMDPAM eine Besonderheit auf: Falls die unabhängige Variable für den i-ten Fall fehlt, erfolgt die Schätzung mit der nächst höchstkorrelierenden Variablen. BMDPAM greift auf die Mittelwertersetzung zurück, falls alle Variablen für den Fall i fehlen. Da im Simulationsprogramm die Variablen 1, 2 und 3 niemals fehlende Werte besitzen, erfolgt in REGRESSION zumindest immer eine Regressionsschätzung mit den externen Variablen. Dies setzt allerdings eine weitere Beschränkung in Kraft: Die Untergrenze für die von der Regression verwandte Information ist nicht die mittlere Interkorrelation (und damit im Extremfall Null), sondern die höchste Extern-Intern-Korrelation. Falls also entweder die höchste Interkorrelation der internen Variablen kleiner als die höchste Extern-Intern-Korrelation ist oder alle internen Variablen für den Fall i fehlen, erfolgt eine Schätzung mit den simulierten demographischen Variablen. HAZARD verwendet zusätzlich zu der von REGRESSION verwendeten Variablen die Anzahl fehlender Werte pro Fall zur Vorhersage. DEAR ersetzt fehlende Werte durch Werte der ersten Hauptkomponente, die nach der Methode Hotellings aus der PAIRWISE-Matrix berechnet wird (während der Simulation trat immer die Konvergenz ein). MULT verwendete die LISTWISE-Matrix als Ausgangspunkt der multiplen Regressionsschätzung, die mit den NAGFLIB-Routinen G02CEF

(Ordnen der Matrizen) und G02CGF(multiple Regression) realisiert wurde[3/]. ILS verwendete die geschätzte Datenmatrix von MULT und die zugehörige vollständige Korrelationsmatrix als Ausgangspunkt. ILS wurde mit den NAGFLIB-Routinen G02CFF (Ordnen der Matrizen) und G02CGF (multiple Regression) realisiert. ILS konvergierte immer innerhalb der Simulation mit weniger als 100 Iterationen. Konvergenzkriterium war eine maximale Veränderung zwischen zwei aufeinanderfolgenden Korrelationsmatrizenschätzungen kleiner als 0.0001.

A.4 Durchführung

Das Programm wurde auf den IBM 4341 des HRZ Essen unter VM/SP Service Level 0307 und SSI mit dem IBM-VS-FORTRAN Compiler Level 1.3.0 erstellt und gerechnet. Die Studie wurde im Batchbetrieb in zwei Durchläufen (MAR; MAT und MARC) gerechnet. Die Generatoren wurden für beide Jobs mit dem Startwert Null gestartet. Die Rechenzeit lag zusammen bei knapp über 15 CPU-Stunden, während der Simulation trat keine der im Programm vorgesehenen Fehlerbedingungen auf.

A.5 Auswertung

Die Ergebnisse der einzelnen Iterationen wurden vom Simulationsprogramm auf einer Platte gespeichert und anschließend auf einem Magnetband gesichert. Diese Datei wies für jede Iteration und jedes Verfahren ein Record auf, Analyseeinheiten waren damit geschätzte Korrelationsmatrizen. Die Datenstruktur entsprach damit einem "univariate setup" (vgl. SPSS-X Release 1.0 User's Guide:503), d.h. voneinander abhängige Untersuchungen derselben Variablen an derselben Untersuchungseinheit unter verschiedenen experimentellen Bedingungen werden durch mehrere Records dargestellt. Für eine vereinfachte Auswertung mit MANOVA wurde diese Datei zu einem "multivariate setup" umgeschrieben; diese Datei besitzt für jede Iteration ein Record, damit ist die Analyseeinheit jeweils eine "wahre" Korrelationsmatrix, die verschiedenen "Behandlungen" ausgesetzt wurde. Für die Erstellung der Crossbreaks wurde eine weitere Datei über AGGREGATE erstellt, die als Analyseeinheiten die Mittelwerte der Verfahren pro Zelle besitzt. Darüber hinaus wurde die Datei P100 erstellt, die die Verfahren als Analyseeinheiten mit je 81 Variablenblöcken, die den 81 experimentellen Bedingungen entsprechen, pro abhängiger Variablen enthält und für einige Clusteranalysen verwendet wurde.
Diese Dateien wurde mit SPSS-X Release 2.1 in der VM/CMS Version des HRZ Essen ausgewertet. Ein Teil der Analysen erfolgte mit SAS82 (ebenfalls unter VM/CMS), jede der Dateien besitzt ein daher ein SAS-Äquivalent.

---

3. In der Literatur wird allgemein eine Sortierung der Fälle nach MD-Mustern vorgeschlagen, um die Berechnung der Regressionen für jedes Muster nur einmal durchführen zu müssen. SIAM06 führt keine solche Sortierung durch, da ein Versuchsprogramm mit Sortierung bei den gegebenen Fallzahlen keine deutliche Verminderung der Rechenzeit zeigte.

B. Programmlisting

```
       PROGRAM SIAM06
      REAL*8 FNMISS, ORIGIN(500,15),A(500,15),B(500,15),RTRUE(15,15)
      INTEGER*4 MLIKE(500),N
      COMMON//FNMISS,ORIGIN,A,B,RTRUE,MLIKE,N
      REAL*8 REFER1(300),REFER2(300)
      COMMON/SET/REFER1,REFER2
      REAL*8 Z(15)
      COMMON/SCORE/Z
      INTEGER*4 IMTYPE,ILN,ICM,ILMPER,ITER
      COMMON/STATUS/ IMTYPE,ILN,ICM,ILMPER,ITER
      REAL*8 C(500,15), MEANCC (15),MEANTP(15)
      REAL*8 STDCC(15),STDTP(15), CORRCC (15,15)
      REAL*8 CORRPW(15,15),R(15,15), GR(15,15),XOXO(15),TRIAL(15)
      REAL*8 CELLS(2,5), CELLM(2,5)
      INTEGER*4 NCELL(2,5)
      REAL*8 SDH(15,15), XSSP(15,15),SSPCC(15,15),XMISS(15),COUNT(15,15)
      REAL*8 XBAR(15),STD(15),ROLD(15,15), COEFF(15,3)
      REAL*8 XSSP(15,15), XMEAN(15),XSTD(15),XR(15,15),CONST(3)
      INTEGER*4 MISS(15),IWORK(15),KORDER(15)
C     SORT ARRAYS
      REAL*8 SORT(500),SORTH(50)
      INTEGER*4 IND(500),INDW(50)
C     VARIABLE DECLARATIONS
      REAL*8 CM,MPER
      INTEGER*4 MITER,MTYPE
      REAL*8 ZERO,X,Y,CODEM,S,B1,B2
      INTEGER*4 INDEX,H,M,MT,IFAIL,CASECC,CASEPW,MAXIT
      INTEGER*4 I1,I2,ITEST,IFLAG,N1,NG,NS
      PARAMETER (CODEM=9.999,ZERO=0.0)
C     CONTROL ARRAYS
      REAL*8 LMPER(3),LCM(3)
      INTEGER*4 LN(3)
      DATA LMPER/0.01,0.05,0.1/
      DATA LCM/0.2,0.4,0.6/
      DATA LN/100,300,500/
C     INIT
      ITEST=1
      MAXIT=10
      CALL G05CBF(0)
C
C     MAIN LOOP
C
      DO 80000 MTYPE=1,3
      IMTYPE=MTYPE
      M=13
      MT=13
      IF(MTYPE.EQ.2)MT=14
      DO 80000 ILN  =1,3
      N=LN(ILN)
      DO 80000 ICM  =1,
      CM=LCM(ICM)
      DO 80000 ILMPER=1,3
      MPER=LMPER(ILMPER)
      WRITE(6,3)MTYPE,N,CM,MPER
3     FORMAT(1X,'MTYPE ',I1,'/ N ',I3,' / CM ',F2.1,' / MPER ',F3.2)
C
```

```
C     ITERATION LOOP
C     DO 50000 MITER=1,MAXIT
      ITER=MITER
C
C     GENERATE DATA MATRIX ORIGIN AND WORK ARRAY A
C
C     INIT MODELS ON FIRST ITERATION
      IF(MITER.EQ.1) THEN
      CALL SETGPR(MT,CM,MTYPE)
      IF(MTYPE.EQ.3) N1=NINT(DFLOAT(N)*0.8)
      END IF
      FNMISS=ZERO
      IFLAG=1
      DO 10 I = 1,N
      MLIKE(I)=0
      IF(MTYPE.EQ.3.AND.I.GT.N1)IFLAG=2
C     IFLAG=2 : MARC  MODEL,SECOND SAMPLE
      CALL GDATA(MT,IFLAG)
C     GENERATED DATA VECTOR : Z ; SIZE :MT
      IF(MTYPE.EQ.2)SORT(I)=Z(14)
      DO 10 J=1,M
      ORIGIN(I,J)=Z(J)
      A(I,J)=Z(J)
      B(I,J)=Z(J)
10    CONTINUE
C     DATA SETS READY;GENERATE MD
      IF(MTYPE.EQ.1)THEN
C     MISSING AT RANDOM
      NS=NINT(DFLOAT(N)*MPER)
      DO 11 J=4,M
      NG=0
      DO 11 I=1,N
      IF(G05CAF(X).LE.DFLOAT(NS-NG)/DFLOAT(N-I+1))THEN
      NG=NG+1
      A(I,J)=CODEM
      B(I,J)=CODEM
      MLIKE(I)=MLIKE(I)+1
      FNMISS=FNMISS+1.0
      END IF
11    CONTINUE
      END IF
C
      IF(MTYPE.EQ.2)THEN
C     HAZARD MODELL
      IFAIL=1
      CALL M01AJF(SORT,SORTH,IND,INDW,N,50,IFAIL)
         IF(IFAIL.NE.0)THEN
                        PRINT*,'M01AJF FAILED ',IFAIL
                        STOP 'HAZARD SORT'
                     END IF
       NS=NINT(DFLOAT(N)*MPER)
       H=2*NS
       DO 13 J=4,M
       NG=0
       DO 12 I1=N-H+1,N
       IF(G05CAF(X).LE.DFLOAT(NS-NG)/DFLOAT(N-I1+1))THEN
       I=IND(I1)
       NG=NG+1
       A(I,J)=CODEM
       B(I,J)=CODEM
```

```
        MLIKE(I)=MLIKE(I)+1
        FNMISS=FNMISS+1.0
        END IF
12      CONTINUE
13      CONTINUE
        END IF
C
C
      IF(MTYPE.EQ.3)THEN
C     MISSING AT RANDOM WITHIN CLASSES
      DO 19 J=4,M
      NG=0
      NS=NINT(DFLOAT(N1)*MPER/2.8)
      DO 18 I=1,N1
      IF(G05CAF(X).LE.DFLOAT(NS-NG)/DFLOAT(N1-I+1))THEN
      NG=NG+1
      A(I,J)=CODEM
      B(I,J)=CODEM
      MLIKE(I)=MLIKE(I)+1
      FNMISS=FNMISS+1.0
      END IF
18    CONTINUE
      NS=NINT(DFLOAT(N-N1)*MPER/.28)
      NG=0
      DO 19 I=N1+1,N
      IF(G05CAF(X).LE.DFLOAT(NS-NG)/DFLOAT(N-I+1))THEN
      NG=NG+1
      A(I,J)=CODEM
      B(I,J)=CODEM
      MLIKE(I)=MLIKE(I)+1
      FNMISS=FNMISS+1.0
      END IF
19    CONTINUE
      END IF
      IF(INT(FNMISS).NE.INT(DFLOAT(N)*MPER*10.0)) STOP 'MD FAILED'
      DO 20 I=1,N
      A(I,14)=MLIKE(I)
20    CONTINUE
C
C     COMPUTE TRUE CORRELATIONMATRIX RTRUE
C
      IFAIL=1
      CALL G02BAF(N,14,ORIGIN,500,MEANTP,STDTP,SSP,15,RTRUE,15,IFAIL)
      IF (IFAIL.NE.0) THEN
                         PRINT*,'G02BAF FAILED ',IFAIL
                         STOP 'TRUE CORRELATIONS'
                       END IF
C
C     COMPUTE 'LISTWISE' DELETION CORRELATION MATRIX CORRCC
C
      IFAIL=1
      DO 40 J=1,M
      MISS(J)=1
      XMISS(J)=CODEM
40    CONTINUE
      CALL G02BBF(N,M,A,500,MISS,XMISS,MEANCC,STDCC,SSPCC,15,
     1            CORRCC,15,CASECC,IFAIL)
      IF(IFAIL.NE.0) THEN
                         PRINT*,'G02BBF FAILED ',IFAIL
                         STOP 'LISTWISE'
```

```
                    END IF
      CALL  RESULT(CORRCC,1)
C
C     COMPUTE 'PAIRWISE' DELETION CORRELATION MATRIX CORRPW
C
      IFAIL=1
      DO 50 J = 1,14
      MISS(J)=1
      XMISS(J)=CODEM
50    CONTINUE
      CALL G02BCF(N,14,A,500,MISS,XMISS,MEANTP,STDTP,SSP,15,
     1            CORRPW,15,CASEPW,COUNT,15,IFAIL)
      IF(IFAIL.NE.0) THEN
                        PRINT*,'G02BCF FAILED ', IFAIL
                        STOP 'PAIRWISE'
                        END IF
       CALL RESULT(CORRPW,2)
C
C     COMPUTE 'PAIRWISE' MEANS AND STD.DEV FOR REGRESSIONS
C
      DO 81 J=1,14
      DO 81 K=1,14
      SDH(J,K)=ZERO
      GR(J,K)=ZERO
81    CONTINUE
      DO 82 I = 1,N
      DO 82 J = 1,14
      DO 82 K = 1,14
      IF(A(I,J).NE.CODEM.AND.A(I,K).NE.CODEM) THEN
      SDH(J,K)=SDH(J,K)+A(I,J)
      GR (J,K)=GR(J,K)+A(I,J)**2
      END IF
82    CONTINUE
      DO 83 J = 1,14
      DO 83 K = 1,14
      SDH(J,K)=SDH(J,K)/COUNT(J,K)
C     PAIRWISE MEANS READY
      X = GR(J,K) / COUNT(J,K) - SDH (J,K) **2
      IF (X.LE.ZERO) THEN
                       PRINT*,'VARIANCE ZERO PAIR ',J,K
                       STOP 'PAIRWISE MEANS'
                    END IF
      GR(J,K)= DSQRT(X)
C     PAIWISE STD.DEV READY
83    CONTINUE
C
C
C     REPLACE MISSING VALUES BY COLUMN MEANS
C
C
      DO 92 I = 1,N
      DO 91 J=4,M
      IF(A(I,J).EQ.CODEM) B(I,J)=MEANTP(J)
91    CONTINUE
92    CONTINUE
      IFAIL=1
      CALL G02BAF(N,M,B,500,XBAR,STD,SSP,15,R,15,IFAIL)
      IF (IFAIL.NE.0) THEN
                         PRINT*,'G02BAF FIALED ',IFAIL
                         STOP 'COMPLETE R - MEAN IMPUTATION'
```

```
                        END IF
      CALL RESULT (R,3)
C
C
C     REPLACE MISSING VALUES BY CELL MEANS
C
C
      DO 105 INDEX=4,M
      DO 100 J=1,2
      DO 100 K=1,5
      CELLS(J,K)=ZERO
      NCELL(J,K)=0
100   CONTINUE
C     COMPUTE CELL SUMS AND N
      DO 101 I = 1,N
      IF(A(I,INDEX).NE.CODEM) THEN
      CELLS(A(I,1),A(I,2))=CELLS(A(I,1),A(I,2))+A(I,INDEX)
      NCELL(A(I,1),A(I,2))=NCELL(A(I,1),A(I,2))+1
      END IF
101   CONTINUE
      DO 103 J=1,2
      DO 103 K=1,5
      IF(NCELL(J,K).GT.0) THEN
      CELLM(J,K)=CELLS(J,K)/NCELL(J,K)
      ELSE
C     EMPTY IMPUTATION CELL :2.TRY COLLAPS VAR2
      X=ZERO
      H=0
      DO 102 I1=1,5
      X=X+CELLS(J,I1)
      H=H+NCELL(J,I1)
102   CONTINUE
      IF(H.GT.0)THEN
      CELLM(J,K)=X/H
      ELSE
C     EMPTY CELL :3.TRY USE MEAN
      CELLM(J,K)=MEANTP(INDEX)
      END IF
      END IF
103   CONTINUE
C     CELL MEANS READY
      DO 104 I=1,N
      IF(A(I,INDEX).EQ.CODEM) B(I,INDEX)=CELLM(A(I,1),A(I,2))
104   CONTINUE
105   CONTINUE
      IFAIL=1
      CALL G02BAF(N,M,B,500,XBAR,STD,SSP,15,R,15,IFAIL)
      IF (IFAIL.NE.0) THEN
                          PRINT*,'G02BAF FAILED ',IFAIL
                          PRINT*,'COMPLETE R - M.I.C.'
                          STOP
                       END IF
      CALL  RESULT(R,4)
C
C     HOT-DECK
C
C
      DO 130 INDEX=4,M
      DO 121 J=1,2
      DO 121 K=1,5
```

```
      CELLS(J,K)=MEANTP(INDEX)
121   CONTINUE
      DO 122 I=1,N
      IF (A(I,INDEX).NE.CODEM)THEN
      CELLS(A(I,1),A(I,2))=A(I ,INDEX)
      ELSE
      B(I,INDEX)=CELLS(A(I,1),A(I,2))
      END IF
122   CONTINUE
130   CONTINUE
      IFAIL=1
      CALL G02BAF(N,M,B,500,XBAR,STD,SSP,15,R,15,IFAIL)
      IF(IFAIL.NE.0)THEN
                        PRINT*,'G02BAF FAILED ',IFAIL
                        STOP 'HOT-DECK'
                    END IF
      CALL  RESULT(R,5)
C
C
C     SIMPLE REGRESSION ESTIMATION
C
C
      DO 200 J = 4,M
C     EXTRACT CORRELATION VECTOR AND SORT
      DO 201 K = 1,M
      XMISS(K)=DABS(CORRPW(J,K))
201   CONTINUE
      IFAIL=1
      CALL M01AJF(XMISS,XOXO,MISS,IWORK,M,15,IFAIL)
      IF (IFAIL.NE.0) THEN
                          PRINT*,'M01AJF FAILED ',IFAIL
                          STOP 'SORT- SIMPLE REGRESSION'
                       END IF
      DO 202 I = 1,N
      IF(A(I,J).EQ.CODEM) THEN
      H=M
203   INDEX=MISS(H)
      IF(INDEX.EQ.J .OR. A(I,INDEX).EQ.CODEM) THEN
      H=H-1
      GO TO 203
      END IF
C     INDEX = VARIABLE INDEX OF HIGHEST CORRELATED VARIABLE
      Y= CORRPW(J,INDEX) * GR(J,INDEX) / GR(INDEX,J)
      B(I,J)= Y * (A(I,INDEX)-SDH(INDEX,J)) + SDH(J,INDEX)
C     2.ESTIMATE:  HAZARD MODEL
      S= CORRPW(INDEX,14)
      X = 1.0 - S**2
      IF(X.LE.ZERO)  STOP 'HAZARD  ZERO'
      B1=CORRPW(INDEX,J)-CORRPW(14,J)*S
      B2=CORRPW(14,J)-CORRPW(INDEX,J)*S
      B1= B1 / X * STDTP(J) / STDTP(INDEX)
      B2= B2 / X * STDTP(J) / STDTP(14)
      Y = MEANTP(J) -  B1 * MEANTP(INDEX) - B2 * MEANTP(14)
      C(I,J)= Y+ B1*A(I,INDEX) + B2*A(I,14)
      END IF
202   CONTINUE
200   CONTINUE
      IFAIL=1
      CALL G02BAF(N,M,B,500,XBAR,STD,SSP,15,R,15,IFAIL)
      IF (IFAIL.NE.0) THEN
```

```
                        PRINT*,'G02BAF FAILED ',IFAIL
                        STOP 'COMPLETE CORR - SIMPLE REGRESSSION'
                    END IF
      CALL  RESULT(R,6)
C
C     HAZARD ESTIMATE
C
      DO 250 I=1,N
      DO 250 J=4,M
      IF(A(I,J).EQ.CODEM)B(I,J)=C(I,J)
250   CONTINUE
      IFAIL=1
      CALL G02BAF(N,M,B,500,XBAR,STD,SSP,15,R,15,IFAIL)
      IF (IFAIL.NE.0) THEN
                        PRINT*,'G02BAF FAILED ',IFAIL
                        STOP 'COMPLETE CORR - HAZARD'
                    END IF
      CALL  RESULT(R,7)
C
C     DEAR
C
      DO 300 J=1,M
      TRIAL(J)=1.0
      XMISS(J)=100000.0
300   CONTINUE
C     FIND FIRST EIGENVECTOR; HOTELLINGS METHOD
      DO 310 K=1,100
C     ITERATION LOOP
      DO 301 J=1,M
      XOXO(J)=ZERO
301   CONTINUE
      DO 302 I1=1,M
      DO 302 J=1,M
      XOXO(I1)=XOXO(I1)+TRIAL(J)*CORRPW(I1,J)
302   CONTINUE
      S=ZERO
      DO 303 I1=1,M
      S=S+XOXO(I1)**2
303   CONTINUE
      S=1.0/DSQRT(S)
      DO 304 I1=1,M
      TRIAL(I1)=XOXO(I1)*S
304   CONTINUE
      X=ZERO
      DO 305 I1=1,M
      Y=DABS(XOXO(I1)-XMISS(I1))
      IF(Y.GT.X)X=Y
      XMISS(I1)=XOXO(I1)
305   CONTINUE
      IF(X.LT.0.001)GO TO 320
310   CONTINUE
      STOP 'DEAR:NO CONVERGENCE REACHED'
320   CONTINUE
C     CONVERGENCE REACHED:EIGENVECTOR=TRIAL
      DO 350 I=1,N
      IF (MLIKE(I).NE.0) THEN
      DO 325 J=1,M
      IF (A(I,J) .EQ. CODEM) THEN
      XOXO(J)=ZERO
      ELSE
```

```
      XOXO(J)=(A(I,J)-MEANTP(J))/STDTP(J)
      END IF
325   CONTINUE
      S=ZERO
      DO 330 J=1,M
      S=S+XOXO(J)*TRIAL(J)
330   CONTINUE
      DO 335 J=4,M
      IF (A(I,J).EQ.CODEM)B(I,J)=MEANTP(J)+S*TRIAL(J)*STDTP(J)
335   CONTINUE
      END IF
350   CONTINUE
      IFAIL=1
       CALL G02BAF(N,M,B,500,XBAR,STD,SSP,15,R,15,IFAIL)
      IF(IFAIL.NE.0)THEN
                         PRINT*,'G02BAF FAILED ',IFAIL
                          PRINT*,'DEAR'
                      END IF
      CALL  RESULT(R,8)
C
C
C     MULTIPLE REGRESSION
C
C
      DO 400 I =1,N
      IF (MLIKE(I).NE.0) THEN
C     PARTITION DATA VECTOR: MISSING / NONMISSING
      I1=0
      I2=0
      DO 402 J=1,M
      IF(A(I,J).EQ.CODEM) THEN
      I1=I1+1
      MISS(I1)=J
      ELSE
      I2=I2+1
      KORDER(I2)=J
      END IF
402   CONTINUE
      H=I2+1
      DO 410 L=1,I1
      INDEX=MISS(L)
      KORDER(H)=INDEX
C     FORM REDUCED VECTORS AND MATRICES
      IFAIL=1
      CALL G02CEF(M,MEANCC,STDCC,SSPCC,15,CORRCC,15,H,KORDER,XMEAN,
     1XSTD,XSSP,15,XR,15,IFAIL)
      IF (IFAIL.NE.0) THEN
                           PRINT*,'G02CEF FAILED ',IFAIL
                           PRINT*,'REDUCED MATRICES: MULT'
                           STOP
                        END IF
C     COMPUTE REGRESSION
      IFAIL=1
      CALL G02CGF(CASECC,H,I2,XMEAN,XSSP,15,XR,15,XOXO,
     1COEFF,15,CONST,GR,15,COUNT,15,ROLD,15,IFAIL)
      IF (IFAIL.NE.0) THEN
                           PRINT*,'G02CGF FAILED ',IFAIL
                           PRINT*,'MULTIPLE REGRESSION:MULT'
                           STOP
                         END IF
```

```
C     ESTIMATE
      Y=ZERO
      DO 420 J =1,I2
      Y=Y+ A(I,KORDER(J)) *COEFF(J,1)
420   CONTINUE
      B(I,INDEX)= Y + CONST(1)
410   CONTINUE
      END IF
400   CONTINUE
      IFAIL=1
      CALL G02BAF(N,M,B,500,XBAR,STD,SSP,15,ROLD,15,IFAIL)
      IF(IFAIL.NE.0)THEN
                        PRINT*,'G02BAF FAILED ',IFAIL
                         PRINT*,'FINAL ESTIMATE R - MULT'
                         STOP
                    END IF
      CALL  RESULT(ROLD,9)
C
C     ITERATED LEAST SQUARE
C
C     ITERATION LOOP FOR ILS
      DO 550 L=1,100
C     COMPUTE REGRESSION FOR EACH VARIABLE WITH MD
      DO 540 INDEX=4,M
C     RESTORE ORIGINAL ORDER
      DO 501 J=1,M
      XMEAN(J)=XBAR(J)
      XSTD(J)=STD(J)
      DO 501 K=1,M
      XR(J,K)=ROLD(J,K)
      XSSP(J,K)=SSP(J,K)
501   CONTINUE
C     ORDER INDEXVECTOR KORDER
      I1=0
      DO 502 J=1,M
      IF(J.NE.INDEX)THEN
      I1=I1+1
      KORDER(I1)=J
      END IF
502   CONTINUE
      KORDER(M)=INDEX
C     ORDER MATRICES ACCORDING TO KORDER
      IFAIL=1
      CALL G02CFF(M,KORDER,XMEAN,XSTD,XSSP,15,XR,15,IWORK,IFAIL)
      IF(IFAIL.NE.0)THEN
                        PRINT*,'G02CFF FAILED ',IFAIL
                        PRINT*,'ILS - ORDER'
                        STOP
                      END IF
C     COMPUTE REGRESSION
      I1=M-1
      IFAIL=1
      CALL G02CGF(N,M,I1,XMEAN,XSSP,15,XR,15,XOXO,COEFF,15,CONST,
     1            GR,15,CORRCC,15,COUNT,15,IFAIL)
      IF(IFAIL.NE.0)THEN
                        PRINT*,'G02CGF FAILED ',IFAIL
                        PRINT*,'ILS - REGRESSION'
                        STOP
                      END IF
C     ESTIMATE
```

```
      DO 510 I=1,N
      IF(A(I,INDEX).EQ.CODEM)THEN
      Y=ZERO
      DO 507 J=1,I1
      Y=Y+B(I,KORDER(J))*COEFF(J,1)
507   CONTINUE
      B(I,INDEX)=Y+CONST(1)
      END IF
510   CONTINUE
540   CONTINUE
C     CHECK CONVERGENCE
      IFAIL=1
      CALL G02BAF(N,M,B,500,XBAR,STD,SSP,15,XR,15,IFAIL)
      IF(IFAIL.NE.0)THEN
                        PRINT*,'G02BAF FAILED ',IFAIL
                        PRINT*,'ILS- IMPROVED ESTIMATE'
                        STOP
                    END IF
      X=ZERO
      DO 545 J=1,M-1
      DO 545 K=J+1,M
      Y= DABS( XR(J,K)-ROLD(J,K))
      IF(Y.GT.X)X=Y
      ROLD(J,K)=XR(J,K)
545   CONTINUE
      IF(X.LT.0.0001) GOTO 551
C     END OF ILS-ITERATION LOOP
550   CONTINUE
      PRINT*,'WARNING :ILS : NO CONVERGENCE AFTER 100 ITERATIONS'
551   CONTINUE
      CALL  RESULT(XR,10)
C
C     END OF ITERATION LOOP
C
50000 CONTINUE
C
C     END OF MAIN LOOP
C
80000 CONTINUE
      STOP '********* NORMAL END OF JOB ********** '
      END
C
C
C
C
C
C
C
      SUBROUTINE SETGPR(M,CM,MTYPE)
C     SET UP REFERENCE VECTOR REFER FOR  G05EZF
      REAL*8 REFER1(300),REFER2(300)
      COMMON/SET/REFER1,REFER2
      REAL*8 CMEAN(15),GPR(15,15)
      REAL*8 CM,CM2
      INTEGER*4 M,MTYPE
C     SET 'EXTERNAL' VARIABLE INTERCORRELATIONS
      GPR(1,2)=.2
      GPR(1,3)=.2
      GPR(2,3)=.2
C     SET 'EXTERNAL' AND 'INTERNAL' CORRELATIONS
```

```
      DO 10 J = 1,3
      DO 10 K = 4,M
10    GPR(J,K)=.15
C     ZERO MEANS AND BUILD UP GIVEN MATRIX
      DO 20 J = 1,M
      GPR(J,J)=1.0
20    CMEAN(J)=0.0
      DO 30 J = 4,12
      DO 30 K = J+1,13
      GPR(J,K)=CM
30    CONTINUE
      IF(MTYPE.EQ.2) THEN
C     THRESHOLD  MODEL
      CM2=CM+0.1
      DO 40  J= 4,13
      GPR(J,14)=CM2
40    CONTINUE
      END IF
      IFAIL=1
      CALL G05EAF(CMEAN,M,GPR,15,0.000001,REFER1,300,IFAIL)
      IF (IFAIL.NE.0) THEN
                         PRINT*, 'G05EAF FAIL' ,IFAIL
                         STOP 'REFERENCE GENERATOR 1'
                      END IF
C     REFERENCE VECTOR  1 FOR MAR AND MAT READY
      IF (MTYPE.EQ.3) THEN
C     MARC MODEL
      DO 50 J=4,12
      DO 50 K=J+1,13
      GPR(J,K)=0.1
50    CONTINUE
      CALL G05EAF(CMEAN,M,GPR,15,0.000001,REFER2,300,IFAIL)
      IF (IFAIL.NE.0) THEN
                         PRINT*, 'G05EAF FAIL' ,IFAIL
                         STOP 'REFERENCE GENERATOR 2'
                      END IF
      END IF
C     REFERENCE VECTOR  2 FOR MARC READY
      RETURN
      END
C
C
      SUBROUTINE GDATA(M,IFLAG)
C     GENERATE DATA-VECTOR Z SIZE M
      REAL*8 REFER1(300),REFER2(300)
      COMMON/SET/REFER1,REFER2
      REAL*8 Z(15)
      COMMON/SCORE/Z
      INTEGER*4  M,IFLAG,IFAIL
      REAL*8 Z7(7),Z5(5)
      DATA Z7/-2.143,-1.286,-.429,.429,1.286,2.143,10000.0/
      DATA Z5/-1.8,-.6,.6,1.8,10000.0/
      IF(IFLAG.EQ.1) THEN
C     USE REFERENCE VECTOR 1
      IFAIL=1
      CALL G05EZF(Z,M,REFER1,300,IFAIL)
      IF (IFAIL.NE.0) THEN
                         PRINT*,'G05EZF FAIL ',IFAIL
                          STOP 'MAIN GENERATOR 1'
                          END IF
```

```
      ELSE
C     USE REFERENCE VECTOR 2
      IFAIL=1
      CALL G05EZF(Z,M,REFER2,300,IFAIL)
      IF (IFAIL.NE.0) THEN
                         PRINT*,'G05EZF FAIL ',IFAIL
                          STOP 'MAIN GENERATOR 2'
                          END IF
      END IF
C     GROUP VECTOR Z
      IF (Z(1).LT.0.5) THEN
      Z(1)=1.0
      ELSE
      Z(1)=2.0
      END IF
      DO 20 J=2,3
      DO 10 K=1,5
      IF (Z(J).LE.Z5(K)) GO TO 15
10    CONTINUE
15    Z(J)=DFLOAT(K)
20    CONTINUE
      DO 40 J = 4,13
      DO 30 K = 1,7
      IF (Z(J).LE.Z7(K)) GO TO 35
30    CONTINUE
35    Z(J)=DFLOAT(K)
40    CONTINUE
      RETURN
      END
C
C
      SUBROUTINE RESULT(R,IPROG)
      REAL*8 FNMISS, ORIGIN(500,15),A(500,15),B(500,15),RTRUE(15,15)
      INTEGER*4 MLIKE(500),N
      COMMON//FNMISS,ORIGIN,A,B,RTRUE,MLIKE,N
      INTEGER*4 IMTYPE,ILN,ICM,ILMPER,ITER
      COMMON/STATUS/ IMTYPE,ILN,ICM,ILMPER,ITER
      REAL*8 R(15,15),ZERO,CODEM,FHIT,EMD,EMAD,ERSD,D
      REAL*8 DI,RSDEV,RDEV,MKS,SX,SY,S2X,S2Y,SXY,X,Y
      INTEGER*4 IPROG, HIT
      CODEM=9.99999
      ZERO=0.0
      EMD=ZERO
      EMAD=ZERO
      ERSD=ZERO
      HIT=0
      RDEV=ZERO
      RSDEV=ZERO
      SX=ZERO
      SY=ZERO
      S2Y=ZERO
      S2X=ZERO
      SXY=ZERO
      DO 10 J = 4,12
      DO 10 K = J+1,13
      DI= RTRUE(J,K)-R(J,K)
      RDEV=RDEV+DI
      RSDEV=RSDEV+DI**2
      X= R(J,K)
      Y= RTRUE(J,K)
```

```
      SX=SX+X
      SY=SY+Y
      S2X=S2X+X**2
      S2Y=S2Y+Y**2
      SXY=SXY+X*Y
 10   CONTINUE
      RDEV=RDEV/45.0
      RSDEV=RSDEV/45.0
      MKS=45.0 * SXY - SX*SY
      X=(45.0*S2X-SX**2)*(45.0*S2Y-SY**2)
      IF (X.LE.ZERO) STOP 'MKS - ZERO VARIANCE'
      MKS=MKS/DSQRT(X)
      IF(IPROG.LT.5)THEN
      EMAD=CODEM
      EMD =CODEM
      ERSD=CODEM
      FHIT=.9999
      ELSE
      DO 110 I=1,N
      IF(MLIKE(I).NE.0)THEN
      DO 100 J = 4,13
      IF(A(I,J).EQ.9.999)THEN
      D=ORIGIN(I,J)-DNINT(B(I,J))
      IF(DABS(D).LT.1.0)HIT=HIT+1
      EMD=EMD+D
      EMAD=EMAD+DABS(D)
      ERSD=ERSD+D**2
      END IF
100   CONTINUE
      END IF
110   CONTINUE
      IF(FNMISS.GT.ZERO) THEN
      EMD=EMD/FNMISS
      EMAD=EMAD/FNMISS
      ERSD=DSQRT(ERSD/FNMISS)
      FHIT=DFLOAT(HIT)/FNMISS
      END IF
      END IF
      WRITE(8,200)IMTYPE,ILN,ICM,ILMPER,ITER,IPROG,RDEV,RSDEV,MKS,
     1 EMD,EMAD,ERSD,FHIT
200    FORMAT(4I1,2I2,2E12.5,F7.4,3F12.5,F5.4)
      RETURN
      END
```

C. Vortests zweier Hot-Deck Verfahren

Zwei Verfahren wurden im Rahmen der Vortests ausgeschieden. Beide Verfahren sind Varianten der Hot-Deck-Technik.
Das erste Verfahren besteht aus einer Modifikation einer Clusteranalyse mit Leaderalgorithmus und euklidischer Distanzfunktion, die über alle 13 Variablen berechnet wird[4/]. Im ersten Durchgang werden die Clusterleader ermittelt. Die Methode setzt den ersten Fall im Datensatz immer als ersten Clusterleader ein. Jeder folgende Fall wird darauf hin geprüft, ob die Distanz zum Clusterleader unterhalb eines "thresholds" liegt. Ist die Distanz kleiner als der Threshold, so wird der Fall dem Cluster zugeordnet, andernfalls wird der Fall zu einem neuen Clusterleader. Die Ermittelung der Cluster erfolgt nur innerhalb des vollständigen Teils des Samples. Im zweiten Durchgang werden sämtliche fehlenden Werte eines Falles durch die Werte des ersten Clusterleaders ersetzt, zu dem die modifizierte Distanzfunktion unterhalb des Thresholds liegt[5/].
Das zweite Verfahren besteht aus einer Implementierung der "nearest neighbour rule". Für jeden Fall mit fehlenden Werten wird derjenige Fall als Donor gewählt, der die geringste euklidische Distanz (berechnet über die jeweils vorhandenen Werte) aufweist. Bei mehreren gleich entfernten potentiellen Donors wird jeweils der erste von Dateibeginn an gewählt[6/]. Diese Methode[7/] wurde in zwei Versionen getestet:

- sämtliche fehlenden Werte eines Falls werden durch die Werte des Donors ersetzt;
- potentielle Donors waren alle Fälle mit vollständigen Werten auf der jeweils betrachteten Variablen. In dieser Variante wurde das Verfahren für jede Variable einzeln durchgeführt.

---

4. Es wurde eine modifizierte Version der Subroutine QUICK von HARTIGAN (1975:82) benutzt. Bei HARTIGAN (1975:77-78) findet sich auch eine Kritik des Leaderalgorithmus.
5. Das Verfahren war in bezug auf den ersten Schritt iterativ: Der Threshold wurde vom Programm solange modifiziert, bis die vorgegebene Clusterzahl (10 und 20 Cluster wurden getestet) erreicht war. Ließ sich im zweiten Schritt ein Fall nicht zuordnen, wurde der Threshold vergrößert und damit die Clusterzahl verkleinert und neu mit dem ersten Schritt begonnen.
6.Das Problem des "minimum match" (Unmöglichkeit der Berechnung der Distanzfunktion bei vollständigem Fehlen aller Klassifikationsvariablen) wurde durch die Vorgabe der drei immer vollständigen Variablen umgangen.
7.Die Ermittelung des Falls mit minimaler Distanz geschah durch lineare Suche im Datensatz. Diese "brute force method" (FRIEDMAN u.a. 1977:22) zieht ungeheure Rechenzeiten nach sich. Im ungünstigsten Fall müssen ((n-1)/2)*(n/2) Vergleiche durchgeführt werden. Wird das Verfahren für jede Variable einzeln durchgeführt, erhöht sich der Aufwand entsprechend. Für praktikable Implementierungen müssen effizientere Algorithmen benutzt werden, z.B. Sortierungsalgorithmen oder K-D-Trees (vgl. hierzu FRIEDMAN u.a. 1977). Das von G.T.SANDE entwickelte Hot-Deck-Verfahren von Statistics Canada benutzt K-D-Trees.

Beide Verfahren wurden unter MAR mit Fallzahlen zwischen 100 und 500, MD-Anteilen zwischen 1% und 10% und mittleren Interkorrelationen zwischen 0.3 und 0.6 getestet. Die Ergebnisse waren in bezug auf die Korrelationsmatrizen auf allen Variablen um eine Zehnerpotenz schlechter als alle anderen Verfahren. Obwohl sich eine Tendenz zur Verbesserung mit wachsendem n zeigte, wurde auf die weitere Benutzung beider Verfahren verzichtet. Möglicherweise werden diese Verfahren erst bei extrem großen Fallzahlen, wie sie in Censuserhebungen vorkommen, konkurrenzfähig. Weiterhin hängt das Ergebnis von den verwendeten "Klassifikationsvariablen" ab. Die in den Programmen vorgenommene Benutzung aller Variablen könnte einen verzerrenden "Overfit" nach sich gezogen haben. In Hinsicht auf die immensen Rechenzeiten des zweiten Verfahrens und der schlechten Ergebnisse beider Verfahren bei relativ kleinen Fallzahlen wurden nach den Ergebnissen der Vortests keine weiteren Untersuchungen mit diesen beiden Methoden vorgenommen.

D. Tabellen zur relativen Effizienz der Verfahren

FISHER-R-ABWEICHUNGSINDEX : RELATIVE EFFIZIENZ DER VERFAHREN

MISSING DATA PROZESS	SAMPLE SIZE	MEAN INTER CORR	ANTEIL MD	LIST	PAIR	MEAN	CELL	DECK	REGR	HAZD	DEAR	MULT	ILS
M A R	100	.2	1%	.651	.943	.934	.923	.823	.951	.934	.960	.933	.941
			5%	.553	.962	.952	.941	.780	.929	.922	.977	.915	.941
			10%	.539	.950	.955	.923	.707	.922	.919	.965	.808	.917
			MEAN	.581	.952	.947	.929	.770	.934	.925	.967	.885	.933
		.4	1%	.667	.935	.913	.897	.759	.887	.891	.979	.944	.945
			5%	.559	.894	.842	.815	.692	.884	.880	.994	.917	.949
			10%	.467	.904	.846	.808	.686	.876	.876	.990	.846	.935
			MEAN	.564	.911	.867	.840	.713	.882	.883	.988	.902	.943
		.6	1%	.634	.902	.839	.826	.710	.900	.885	.967	.948	.948
			5%	.543	.904	.790	.763	.601	.892	.896	.989	.929	.950
			10%	.382	.842	.695	.677	.545	.815	.824	.996	.803	.940
			MEAN	.520	.883	.775	.755	.618	.869	.868	.984	.894	.946
		MEAN		.555	.915	.863	.841	.700	.895	.892	.980	.894	.941
	300	.2	1%	.681	.965	.955	.956	.820	.960	.959	.989	.975	.980
			5%	.558	.952	.936	.946	.770	.922	.921	.988	.969	.984
			10%	.377	.968	.952	.958	.773	.919	.919	.983	.852	.968
			MEAN	.538	.962	.948	.953	.788	.933	.933	.986	.932	.977
		.4	1%	.689	.948	.901	.893	.793	.937	.928	.990	.982	.986
			5%	.592	.940	.893	.889	.683	.911	.910	.990	.976	.980
			10%	.448	.930	.871	.857	.654	.904	.903	.963	.906	.975
			MEAN	.576	.939	.888	.880	.710	.917	.914	.981	.955	.981
		.6	1%	.650	.900	.803	.801	.651	.936	.917	.990	.987	.988
			5%	.560	.894	.774	.782	.599	.871	.872	.991	.959	.978
			10%	.531	.901	.720	.722	.519	.798	.797	.991	.922	.969
			MEAN	.580	.898	.766	.768	.590	.868	.862	.991	.956	.978
		MEAN		.565	.933	.867	.867	.696	.906	.903	.986	.948	.979
	500	.2	1%	.671	.953	.943	.954	.818	.948	.946	.983	.977	.979
			5%	.522	.975	.959	.962	.789	.944	.945	.993	.970	.989
			10%	.457	.966	.952	.950	.734	.936	.940	.955	.913	.989
			MEAN	.550	.965	.951	.955	.780	.943	.944	.977	.953	.986
		.4	1%	.680	.920	.879	.880	.796	.916	.913	.993	.989	.990
			5%	.540	.922	.874	.874	.725	.904	.904	.988	.972	.979
			10%	.418	.909	.837	.848	.614	.835	.833	.997	.895	.973
			MEAN	.546	.917	.864	.867	.712	.885	.884	.993	.952	.981
		.6	1%	.683	.913	.796	.804	.680	.936	.921	.990	.985	.986
			5%	.591	.869	.721	.723	.518	.849	.847	.993	.981	.984
			10%	.486	.864	.689	.696	.511	.795	.794	.979	.937	.986
			MEAN	.587	.882	.736	.741	.569	.860	.854	.987	.968	.986
		MEAN		.561	.921	.850	.855	.687	.896	.894	.986	.958	.984
	MEAN			.560	.923	.860	.854	.694	.899	.896	.984	.933	.968

FISHER-R-ABWEICHUNGSINDEX : RELATIVE EFFIZIENZ DER VERFAHREN

MISSING DATA PROZESS	SAMPLE SIZE	MEAN INTER CORR	ANTEIL MD	LIST	PAIR	MEAN	CELL	DECK	REGR	HAZD	DEAR	MULT	ILS
M A T	100	.2	1%	.902	.955	.941	.930	.815	.935	.910	.971	.946	.946
			5%	.890	.960	.946	.920	.808	.950	.935	.981	.968	.965
			10%	.879	.980	.966	.938	.758	.967	.969	.980	.947	.948
			MEAN	.890	.965	.951	.929	.794	.951	.938	.977	.954	.953
		.4	1%	.910	.966	.904	.886	.785	.960	.892	.976	.983	.981
			5%	.856	.926	.878	.905	.733	.942	.917	.976	.972	.980
			10%	.807	.902	.836	.796	.670	.928	.960	.970	.924	.957
			MEAN	.858	.931	.872	.862	.730	.944	.923	.974	.960	.973
		.6	1%	.906	.891	.708	.687	.650	.923	.870	.933	.984	.988
			5%	.900	.929	.752	.739	.542	.891	.917	.937	.945	.957
			10%	.890	.908	.723	.713	.580	.885	.900	.943	.958	.983
			MEAN	.899	.909	.727	.713	.591	.900	.896	.937	.962	.976
		MEAN		.882	.935	.850	.835	.705	.931	.919	.963	.958	.967
	300	.2	1%	.933	.957	.945	.950	.847	.961	.953	.982	.980	.980
			5%	.932	.975	.959	.952	.818	.966	.930	.986	.973	.973
			10%	.875	.954	.945	.941	.768	.931	.940	.965	.972	.979
			MEAN	.913	.962	.950	.947	.811	.953	.941	.978	.975	.977
		.4	1%	.941	.956	.897	.907	.795	.963	.898	.988	.990	.988
			5%	.883	.909	.830	.842	.714	.898	.976	.944	.961	.973
			10%	.852	.903	.834	.846	.646	.899	.935	.950	.955	.978
			MEAN	.892	.923	.853	.865	.718	.920	.937	.961	.969	.980
		.6	1%	.854	.856	.673	.670	.636	.893	.896	.890	.985	.988
			5%	.868	.893	.710	.715	.539	.884	.931	.951	.974	.991
			10%	.821	.885	.716	.701	.439	.864	.894	.918	.950	.966
			MEAN	.848	.878	.700	.695	.538	.880	.907	.920	.970	.981
		MEAN		.884	.921	.834	.836	.689	.918	.928	.953	.971	.979
	500	.2	1%	.922	.967	.947	.958	.859	.964	.930	.996	.995	.995
			5%	.904	.955	.946	.951	.809	.955	.964	.986	.979	.981
			10%	.880	.958	.941	.927	.726	.945	.962	.983	.975	.985
			MEAN	.902	.960	.945	.946	.798	.954	.952	.988	.983	.987
		.4	1%	.934	.942	.884	.894	.739	.943	.928	.979	.984	.984
			5%	.903	.947	.889	.897	.723	.934	.941	.977	.985	.994
			10%	.843	.925	.873	.871	.634	.897	.927	.957	.977	.985
			MEAN	.893	.938	.882	.887	.699	.925	.932	.971	.982	.987
		.6	1%	.909	.898	.723	.734	.616	.909	.941	.913	.979	.982
			5%	.821	.833	.651	.665	.507	.867	.914	.879	.949	.974
			10%	.792	.873	.684	.689	.520	.868	.919	.917	.942	.997
			MEAN	.841	.868	.686	.696	.548	.881	.925	.903	.957	.984
		MEAN		.879	.922	.838	.843	.681	.920	.936	.954	.974	.986
	MEAN			.882	.926	.841	.838	.692	.923	.928	.957	.968	.978

FISHER-R-ABWEICHUNGSINDEX : RELATIVE EFFIZIENZ DER VERFAHREN

MISSING DATA PROZESS	SAMPLE SIZE	MEAN INTER CORR	ANTEIL MD	LIST	PAIR	MEAN	CELL	DECK	REGR	HAZD	DEAR	MULT	ILS
M A R C	100	.2	1%	.703	.960	.957	.944	.840	.954	.937	.965	.960	.962
			5%	.653	.962	.962	.936	.794	.942	.937	.972	.936	.943
			10%	.591	.949	.955	.920	.737	.938	.932	.963	.887	.911
			MEAN	.649	.957	.958	.933	.790	.945	.935	.967	.928	.939
		.4	1%	.651	.948	.954	.911	.837	.956	.960	.963	.915	.922
			5%	.629	.926	.906	.911	.729	.910	.901	.970	.941	.938
			10%	.592	.936	.911	.897	.713	.892	.878	.986	.902	.947
			MEAN	.624	.937	.924	.906	.760	.919	.913	.973	.919	.936
		.6	1%	.669	.941	.942	.938	.806	.932	.908	.950	.939	.942
			5%	.531	.924	.895	.870	.729	.856	.860	.969	.932	.915
			10%	.561	.891	.866	.848	.672	.861	.849	.993	.910	.929
			MEAN	.587	.919	.901	.886	.736	.883	.872	.971	.927	.929
		MEAN		.620	.937	.928	.908	.762	.916	.907	.970	.925	.934
	300	.2	1%	.686	.972	.967	.969	.826	.972	.956	.982	.980	.980
			5%	.586	.967	.956	.964	.806	.937	.941	.983	.951	.972
			10%	.639	.965	.961	.950	.738	.943	.943	.975	.957	.976
			MEAN	.637	.968	.961	.961	.790	.950	.947	.980	.963	.976
		.4	1%	.701	.963	.954	.959	.827	.979	.979	.994	.988	.992
			5%	.639	.946	.929	.928	.726	.931	.931	.986	.989	.983
			10%	.607	.959	.953	.953	.694	.897	.891	.964	.937	.951
			MEAN	.649	.956	.945	.947	.749	.936	.934	.981	.971	.975
		.6	1%	.643	.957	.929	.938	.775	.939	.935	.981	.969	.970
			5%	.588	.964	.945	.942	.733	.868	.863	.975	.969	.971
			10%	.606	.911	.861	.871	.583	.822	.824	.982	.926	.954
			MEAN	.612	.944	.912	.917	.697	.876	.874	.979	.955	.965
		MEAN		.633	.956	.939	.942	.745	.921	.918	.980	.963	.972
	500	.2	1%	.725	.975	.973	.980	.855	.975	.972	.979	.986	.982
			5%	.627	.958	.952	.961	.806	.942	.945	.983	.967	.975
			10%	.624	.973	.968	.966	.731	.947	.943	.975	.969	.979
			MEAN	.659	.969	.964	.969	.797	.955	.953	.979	.974	.979
		.4	1%	.701	.965	.963	.952	.833	.979	.980	.968	.973	.973
			5%	.659	.944	.917	.931	.757	.905	.905	.988	.973	.979
			10%	.622	.947	.921	.923	.675	.887	.884	.990	.968	.966
			MEAN	.661	.952	.933	.935	.755	.924	.923	.982	.971	.973
		.6	1%	.640	.974	.975	.969	.806	.935	.937	.975	.972	.972
			5%	.583	.945	.922	.948	.736	.886	.888	.995	.975	.976
			10%	.539	.947	.880	.892	.541	.875	.881	.995	.936	.965
			MEAN	.587	.955	.926	.936	.694	.899	.902	.988	.961	.971
		MEAN		.636	.959	.941	.947	.749	.926	.926	.983	.969	.974
	MEAN			.629	.951	.936	.932	.752	.921	.917	.978	.952	.960
	TOTAL			.691	.933	.879	.875	.713	.914	.914	.973	.951	.969

EXAKTE SCHÄTZUNG (HIT) : RELATIVE EFFIZIENZ DER VERFAHREN

MISSING DATA PROZESS	SAMPLE SIZE	MEAN INTER CORR	ANTEIL MD	DECK	REGR	HAZD	DEAR	MULT	ILS
M A R	100	.2	1%	.407	.770	.643	.517	.643	.648
			5%	.731	.873	.886	.879	.800	.843
			10%	.612	.898	.898	.951	.810	.847
			MEAN	.583	.847	.809	.782	.751	.780
		.4	1%	.392	.637	.693	.787	.673	.710
			5%	.607	.869	.814	.878	.864	.871
			10%	.612	.807	.867	.987	.795	.884
			MEAN	.537	.771	.792	.884	.778	.822
		.6	1%	.550	.550	.733	.775	.821	.745
			5%	.526	.778	.763	.970	.841	.907
			10%	.555	.834	.824	.996	.797	.918
			MEAN	.544	.721	.773	.914	.819	.857
		MEAN		.555	.779	.791	.860	.783	.819
	300	.2	1%	.561	.789	.786	.897	.873	.890
			5%	.644	.909	.917	.929	.922	.935
			10%	.730	.941	.952	.988	.901	.943
			MEAN	.645	.880	.885	.938	.899	.923
		.4	1%	.555	.786	.812	.890	.828	.839
			5%	.628	.873	.874	.966	.903	.922
			10%	.673	.900	.906	.967	.957	.979
			MEAN	.619	.853	.864	.941	.896	.913
		.6	1%	.443	.806	.840	.925	.892	.864
			5%	.535	.795	.781	.979	.945	.961
			10%	.522	.808	.817	.983	.933	.965
			MEAN	.500	.803	.813	.962	.923	.930
		MEAN		.588	.845	.854	.947	.906	.922
	500	.2	1%	.537	.825	.787	.930	.898	.925
			5%	.668	.960	.953	.946	.957	.969
			10%	.702	.966	.964	.973	.963	.961
			MEAN	.636	.917	.901	.950	.939	.952
		.4	1%	.639	.816	.801	.920	.936	.918
			5%	.601	.873	.879	.971	.985	.983
			10%	.634	.898	.905	.980	.974	.982
			MEAN	.625	.862	.862	.957	.965	.961
		.6	1%	.517	.880	.793	.956	.933	.939
			5%	.516	.829	.828	.981	.968	.978
			10%	.546	.857	.852	.997	.968	.971
			MEAN	.526	.855	.824	.978	.957	.963
		MEAN		.596	.878	.862	.962	.954	.959
	MEAN			.579	.834	.836	.923	.881	.900

EXAKTE SCHÄTZUNG (HIT) : RELATIVE EFFIZIENZ DER VERFAHREN

MISSING DATA PROZESS	SAMPLE SIZE	MEAN INTER CORR	ANTEIL MD	DECK	REGR	HAZD	DEAR	MULT	ILS
M A T	100	.2	1%	.569	.770	.454	.683	.749	.774
			5%	.622	.835	.941	.876	.865	.884
			10%	.634	.890	.880	.945	.905	.896
			MEAN	.608	.832	.759	.835	.840	.851
		.4	1%	.295	.697	.762	.592	.748	.748
			5%	.535	.793	.891	.886	.875	.812
			10%	.554	.806	.896	.833	.885	.946
			MEAN	.461	.765	.849	.770	.836	.835
		.6	1%	.297	.562	.723	.671	.813	.853
			5%	.408	.754	.890	.720	.932	.921
			10%	.476	.750	.878	.717	.876	.980
			MEAN	.394	.689	.830	.703	.874	.918
		MEAN		.488	.762	.813	.769	.850	.868
	300	.2	1%	.615	.772	.804	.772	.810	.805
			5%	.662	.914	.896	.912	.941	.951
			10%	.668	.914	.953	.947	.935	.937
			MEAN	.648	.867	.884	.877	.895	.898
		.4	1%	.495	.705	.784	.647	.834	.873
			5%	.534	.803	.943	.848	.934	.925
			10%	.650	.813	.951	.901	.948	.965
			MEAN	.560	.774	.893	.799	.905	.921
		.6	1%	.293	.747	.812	.512	.929	.925
			5%	.424	.706	.893	.681	.967	.987
			10%	.452	.724	.921	.738	.948	.988
			MEAN	.390	.726	.876	.644	.948	.967
		MEAN		.532	.789	.884	.773	.916	.928
	500	.2	1%	.545	.827	.864	.956	.919	.912
			5%	.749	.906	.935	.940	.962	.969
			10%	.680	.931	.949	.960	.964	.984
			MEAN	.658	.888	.916	.952	.948	.955
		.4	1%	.396	.615	.882	.660	.877	.880
			5%	.561	.812	.969	.847	.934	.950
			10%	.583	.820	.937	.901	.951	.997
			MEAN	.513	.749	.929	.803	.921	.942
		.6	1%	.314	.671	.841	.554	.915	.919
			5%	.389	.744	.948	.683	.951	.968
			10%	.456	.701	.926	.742	.942	.997
			MEAN	.386	.705	.905	.660	.936	.962
		MEAN		.519	.781	.917	.805	.935	.953
	MEAN			.513	.777	.871	.782	.900	.917

EXAKTE SCHÄTZUNG (HIT) : RELATIVE EFFIZIENZ DER VERFAHREN

MISSING DATA PROZESS	SAMPLE SIZE	MEAN INTER CORR	ANTEIL MD	DECK	REGR	HAZD	DEAR	MULT	ILS
M A R C	100	.2	1%	.698	.845	.770	.750	.680	.605
			5%	.655	.838	.819	.908	.882	.864
			10%	.588	.885	.883	.939	.864	.889
			MEAN	.647	.856	.824	.866	.809	.786
		.4	1%	.542	.855	.785	.760	.672	.772
			5%	.571	.846	.874	.909	.882	.839
			10%	.749	.835	.836	.942	.908	.895
			MEAN	.621	.845	.832	.870	.821	.835
		.6	1%	.656	.559	.695	.744	.764	.837
			5%	.710	.776	.802	.891	.823	.779
			10%	.695	.823	.824	.991	.899	.892
			MEAN	.687	.719	.774	.875	.829	.836
		MEAN		.652	.807	.810	.870	.819	.819
	300	.2	1%	.691	.848	.829	.900	.820	.828
			5%	.672	.899	.890	.961	.931	.928
			10%	.679	.963	.959	.953	.947	.956
			MEAN	.681	.903	.893	.938	.899	.904
		.4	1%	.565	.817	.823	.901	.777	.808
			5%	.708	.921	.889	.974	.943	.947
			10%	.696	.926	.893	.983	.922	.904
			MEAN	.656	.888	.868	.953	.881	.886
		.6	1%	.559	.772	.775	.833	.867	.877
			5%	.641	.825	.825	.958	.976	.961
			10%	.621	.833	.856	.984	.929	.944
			MEAN	.607	.810	.819	.925	.924	.927
		MEAN		.648	.867	.860	.939	.901	.906
	500	.2	1%	.651	.910	.905	.841	.913	.909
			5%	.662	.939	.947	.960	.980	.951
			10%	.743	.975	.963	.967	.970	.958
			MEAN	.685	.941	.938	.923	.954	.939
		.4	1%	.587	.891	.881	.906	.931	.909
			5%	.699	.926	.905	.957	.944	.960
			10%	.654	.919	.895	.984	.971	.974
			MEAN	.647	.912	.894	.949	.949	.947
		.6	1%	.700	.813	.739	.911	.844	.853
			5%	.695	.867	.866	.993	.957	.938
			10%	.652	.841	.844	.998	.949	.956
			MEAN	.682	.840	.816	.967	.917	.916
		MEAN		.671	.898	.883	.946	.940	.934
	MEAN			.657	.857	.851	.918	.887	.886
	TOTAL			.583	.823	.853	.875	.889	.901

E. Realisierung der MD-Techniken mit SPSS-X

Als Hinweise für den praktischen Einsatz der MD-Verfahren sollen hier einige Möglichkeiten der Realisierung innerhalb des Standardpaketes SPSS-X, das sich zumindest in der Anwendung durch Sozialwissenschaftler ebenso wie die Vorläuferprogramme als das fast ausschließlich genutzte Paket durchsetzen dürfte, aufgezeigt werden. Mit nur leichten Veränderungen gelten diese Hinweise ebenso für andere Pakete, wie z.B. BMDP und SAS, wobei aufgrund der hervorragenden Möglichkeiten sowohl durch Macros als auch durch die vielen "Filemanagementutilities" die Realisierung in SAS sehr viel leichter als in den anderen Paketen ist.
Die Beispiele sollen zeigen, das ein Großteil der MD-Techniken auch dann mit einem Standardpaket durchgeführt werden kann, wenn das Programmpaket keine speziellen Prozeduren zur Analyse von MD- Strukturen oder Möglichkeiten der MD-Ersetzung bietet. Die aufgezeigten Möglichkeiten sind keineswegs ein erschöpfender Katalog, so sind beispielsweise selbst Ansätze der "multiple imputation"- Technik mit SPSS-X realisierbar. Da einige Techniken allerdings, wenn auch elementare, Programme auf der Ebene des Betriebssystems erfordern, werden solche Möglichkeiten hier nicht weiter verfolgt. Andererseits wird die Mittelwertersetzung demonstriert, die als Option in einigen Prozeduren verfügbar ist. Da diese Ersetzungen in den Prozeduren jedoch nicht den Datensatz verändert, wird die Ersetzung im Datensatz gezeigt.
Wie schon erwähnt, sind einige MD-Techniken in Standardpaketen nur durch die Benutzung externer Prozeduren oder durch die Programmierung innerhalb des Statistiksystems realisierbar[8/]. So lassen sich die iterativen Techniken, also z.B. iteriertes Verfahren nach DEAR oder BUCK, der EM-Algorithmus oder die unkorregierte Version der "iterated least squares", nur durch intensive Benutzung der in SAS vorhandenen PROC MATRIX bzw. der für SPSS-X angekündigten "Prozedur" MATRIX durchführen. Da selbst die Adaption der Prozeduren für ein gegebenes Problem Programmierkenntnisse verlangt, die über das bei Sozialwissenschaftlern in der Regel vorhandene Maß hinausgehen[9/] werden solche Verfahren nicht erörtert.
Zunächst sollen einige einfache Ersetzungstechniken demonstriert werden, unter anderem das Hot-Deck, dessen Implementierbarkeit in Standardpaketen gelegentlich in der Literatur bezweifelt wird. Dann soll eine Plot-Technik für MD-Strukturen gezeigt werden, die einen BMDPAM-ähnlichen Ausdruck der MD-Muster erlaubt.
Abschließend werden einige Hinweise zur Durchführung von Probit- Analysen und logistischen Regressionen mit PROBIT gegeben, die für die in Kapitel 4 erörterten Korrekturverfahren erforderlich sind.

---

8. Nahezu alle Möglichkeiten sind in BMDPAM einfach verfügbar.
9. Vgl. BÖHNING/WILKE (1980:291-292).

E.1 MD-Ersetzungstechniken

Die Beschreibung der Möglichkeiten erfolgt hier unabhängig von der Bewertung der Verfahren; ob eine gegebene Technik bei einem gegebenen Problem angewendet werden sollte, liegt im Ermessen des Anwenders. Insbesondere ist auf die äußerst kritische Bestimmung der Imputationklassen hinzuweisen, die theoretisch erfolgen sollte und nicht rein mechanisch.
Wie das Beispiel zeigt, ist die Ersetzung durch den Mittelwert der vorhandenen Fälle recht einfach, wenn Systemdateien benutzt werden. Die MD-behafteten Variablen V1-V5 werden lediglich aggregiert, wobei eine Konstante als "BREAK"-Variable verwendet wird. Die Ersetzung erfolgt direkt in die Ausgangsvariablen.

```
COMMENT ******************************************
COMMENT *  M E A N  S U B S T I T U T I O N       *
COMMENT ******************************************
COMMENT
FILE HANDLE REIN /NAME = 'DEMO SYS A'
GET FILE=REIN
COMPUTE DUMMY=1
FILE HANDLE SCRATCH/NAME= 'SCRATCH FILE A'
SAVE OUTFILE =SCRATCH /
AGGREGATE OUTFILE= *
 /BREAK=DUMMY
 /M1 TO M5 = MEAN (V1 TO V5)
MATCH FILES TABLE=* / FILE = SCRATCH / BY DUMMY/
DO REPEAT I = V1 TO V5 / M = M1 TO M5/
IF (MISSING(I)) I = M
END REPEAT
```

Sehr ähnlich läßt sich ein Zellenmittelwertverfahren durchführen. Zusätzlich zu den MD-behafteten Variablen V1-V5 wird eine vollständige Gruppierungsvariable ("GROUP") angenommen, die eventuell erst aus mehreren Gruppierungsvariablen, z.B. einer Kreuztabellierung berechnet werden muß. Durch die Aggregierung werden die Zellenmittelwerte entsprechend der Gruppierungsvariablen berechnet und den Hilfsvariablen M1-M5 zugewiesen. Die Mittelwerte der Zellen werden zunächst jedem Fall entsprechend seiner Zelle zugeschrieben. Abschließend werden die fehlenden Daten auf den Variablen V1-V5 durch die Mittelwerte der Zelle des jeweiligen Falles ersetzt, die Variablen M1-M5 können dann von der aktiven Datei gelöscht werden.

```
COMMENT ******************************************
COMMENT * C E L L M E A N  S U B S T I T U T I O N *
COMMENT ******************************************
COMMENT
FILE HANDLE REIN /NAME = 'DEMO SYS A'
GET FILE=REIN
SORT CASES BY GROUP
FILE HANDLE SCRATCH/NAME= 'SCRATCH FILE A'
SAVE OUTFILE =SCRATCH /
AGGREGATE OUTFILE= *
 /BREAK=GROUP
 /M1 TO M5 = MEAN (V1 TO V5)
MATCH FILES TABLE=* / FILE = SCRATCH / BY GROUP/
DO REPEAT I = V1 TO V5 / M = M1 TO M5/
IF (MISSING(I)) I = M
END REPEAT
PRINT FORMATS ALL (F3.0)
LIST
```

Ein Hot-Deck-Verfahren läßt sich mit etwas Mühe ebenfalls in SPSS-X durchführen. Wiederum werden die MD-behafteten Variablen V1-V5 und eine Gruppierungsvariable GROUP angenommen. Die Realisation des HOT-DECKS umfaßt zwei Probleme. Einmal müssen die Ersetzungszellen initialisiert werden, also bei jedem Wechsel der Variablen GROUP mit gültigen Werten gefüllt werden. Angenommen wird dabei hier, daß für jede Gruppe (Zelle) zumindest ein vollständiger Fall auf den Variablen V1-V5 existiert. Das zweite Problem besteht in der Übertragung der gültigen Werte des jeweils vorhergehenden Falls auf den aktuellen Fall.
Das erste Problem wird dadurch gelöst, daß die Datei nach GROUP (und MISS) sortiert und aggregiert wird, wobei die Hilfsvariablen M1-M5 die Variablenwerte V1-V5 des jeweils ersten vollständigen Falls in der Zelle enthalten (die zusätzliche Sortierung nach MISS garantiert, daß der erste Fall innerhalb jeder Gruppe auf den Variablen V1-V5 vollständig ist). Die Hilfsvariablen werden jedem Fall zunächst zugeschrieben. Die Variablen V1-V5 werden bei vollständigen Fällen lediglich in die Ausgabevariablen O1-O5 kopiert.
Das zweite Problem muß in zwei Fälle gegliedert werden. Zunächst der Fall, daß der erste Fall in der Datei unvollständig ist. In diesem Fall erhält die Hilfsvariablen M1-M5 als Ausgabevariablen O1-O5 zugewiesen, damit die Werte des ersten vollständigen Falls der Gruppe. Exakt dieselbe Initialisierung erfolgt bei jedem Wechsel der Gruppe.
Allen anderen unvollständigen Fälle werden einfach über eine LAG- Anweisung die stets gültigen Werte des vorhergehenden Falles (notwendigerweise derselben Zelle) zugewiesen.
Damit enthalten alle Fälle in den Variablen O1-O5 gültige Werte des vorhergehenden Falls bzw. des ersten gültigen der Gruppe. Innerhalb der Zellen arbeitet dieses HOT-DECK somit rein sequentiell, eventuelle Sequenzeffekte bleiben erhalten.
Wichtig ist die Beachtung der Ersetzung der Werte eines gesamten Falles, falls auch nur eine Variable fehlt: Der vollständige Variablensatz V1-V5 wird durch das HOT-DECK ersetzt. Wird lediglich variablenweise Ersetzung gewünscht, muß das Programm leicht modifiziert werden.

```
COMMENT ****************************************
COMMENT * H O T  D E C K                        *
COMMENT ****************************************
COMMENT
FILE HANDLE REIN /NAME = 'DEMO SYS A'
GET FILE=REIN
COMPUTE MISS=NMISS (V1 TO V5)
IF (MISS NE 0) MISS= 99
MISSING VALUES MISS (99)
SORT CASES BY GROUP MISS
AGGREGATE OUTFILE=* /
 BREAK= GROUP /
 M1 TO M5 = FIRST (V1 TO V5)/
MATCH FILE TABLE=* / FILE = REIN/ BY=GROUP /
DO IF (NMISS (V1 TO V5) = 0)
DO REPEAT I=V1 TO V5 /O=O1 TO O5/
COMPUTE O = I
END REPEAT
ELSE IF ((NMISS (V1 TO V5) NE 0) AND (($CASENUM EQ 1) OR
         (LAG (GROUP,1) NE GROUP)))
DO REPEAT I=M1 TO M5 /O=O1 TO O5/
COMPUTE O=I
END REPEAT
ELSE
DO REPEAT I=O1 TO O5/
COMPUTE I = LAG(I,1)
END REPEAT
END IF
PRINT FORMATS V1 TO O5 (F3.0)
LIST
```

Regressionstechniken lassen sich in vielen Varianten in SPSS-X realisieren. Das Problem bei einer nahezu "automatischen" Ersetzung besteht lediglich darin, REGRESSION dazu zu veranlassen bei der Berechnung der Vorhersagen für die unvollständigen Fälle diese Fälle mit in die Berechnung einzubeziehen. Eine Möglichkeit besteht darin, über die SELECT-Anweisung (innerhalb der Prozedur) nur vollständige Fälle zur Berechnung der Regressionsgleichung zu benutzen, aber MEANSUBSTITUTION als MISSING-Methode anzugeben. Damit wird die Gleichung aus den vollständigen Fällen berechnet, aber die Schätzung für alle Fälle vorgenommen, wobei nur bei der Vorhersage für die fehlenden Daten der Mittelwert ersetzt wird. Nach der DO REPEAT-Schleife enthalten die Variablen V1-V5 nur noch gültige Werte, die Hilfsvariablen können gelöscht werden.
Das Verfahren kann natürlich auch gleich mit der Korrelationsmatrix nach der Mittelwertersetzung gestarted werden, ebenso kann STEPWISE verwendet werden, falls ein "Over-Fit" befürchtet wird. Ebenso können andere Prädiktoren, beispielsweise die Anzahl fehlender Werte, Verwendung finden. Schließlich kann das Verfahren auch innerhalb der Zellen einer Gruppierungsvariablen getrennt (SPLIT FILE) durchgeführt werden.

```
COMMENT ******************************************
COMMENT * M U L T I P L E  R E G R E S S I O N    *
COMMENT ******************************************
COMMENT
FILE HANDLE REIN /NAME = 'DEMO SYS A'
GET FILE=REIN
COMPUTE MISS=NMISS (V1 TO V5)
REGRESSION  SELECT = MISS EQ 0
 /MISSING=MEANSUBSTITUTION
 /VARS=V1 TO V5
 /DEP = V1 /ENTER
 /SAVE PRED(O1)
 /DEP = V2 /ENTER
 /SAVE PRED(O2)
 /DEP = V3 /ENTER
 /SAVE PRED(O3)
 /DEP = V4 /ENTER
 /SAVE PRED(O4)
 /DEP = V5 /ENTER
 /SAVE PRED(O5)
DO REPEAT I=V1 TO V5 /O=O1 TO O5/
IF (NOT (MISSING(I))) O=I
END REPEAT
CONDESCRIPTIVE ALL
```

E.2 Analyse von MD-Mustern

Das Beispiel zeigt die Ausgabe einer BMDPAM-ähnlichen Darstellung des Musters fehlender Werte. Im Gegensatz zu BMDPAM werden nur unterschiedliche Muster ausgegeben, diese aber mit ihrer Häufigkeit und der Anzahl fehlender Variablen pro Muster.
Die DO-REPEAT-Schleife erstellt nur die MD-Indikatorvariablen, nach denen dann sortiert und aggregiert wird. AGGREGATE ermittelt die Anzahl der Fälle innerhalb eines MD-Musters und kopiert die Indikatorvariablen in die neue Arbeitsdatei.
Für jedes Muster wird die Anzahl fehlender Variablen berechnet (MISSV). Dann werden fünf neue String-Variablen erzeugt, die lediglich den Output lesbarer machen sollen, indem nicht Zahlen, sondern die Symbole "M" für fehlende Werte und Blanks für vorhandene Werte ausgegeben werden. Wie die PRINT-Anweisungen zeigen, müssen diese Anweisungen sehr stark an die jeweilige Aufgabe angepaßt werden. Die Druckausgabe kann natürlich verändert werden, indem z.B. noch einmal nach MISSV sortiert wird.

```
COMMENT ******************************************
COMMENT * M D - P A T T E R N  A N A L Y S E     *
COMMENT ******************************************
COMMENT
FILE HANDLE REIN /NAME = 'DEMO SYS A'
GET FILE=REIN
DO REPEAT I =V1 TO V5 /O = M1 TO M5/
COMPUTE O=MISSING (I)
END REPEAT
SORT CASES BY M1 TO M5
AGGREGATE OUTFILE=* /
 BREAK=M1 TO M5/
```

```
 MPN=NU(M1)/
 O1 TO O5 = LAST (M1 TO M5)/
COMPUTE MISSV=SUM (M1 TO M5)
STRING MD1 TO MD5 (A3)
DO REPEAT I = M1 TO M5/ O = MD1 TO MD5/
DO IF (I=1)
COMPUTE O = ' M '
ELSE
COMPUTE O = '   '
END IF
END REPEAT
DO IF $CASENUM EQ 1
PRINT EJECT /'******** M D - P A T T E R N  A N A L Y S I S
********'
PRINT SPACE
PRINT / 'PATTERN NR.' 1 'N OF CASES' 15
 '         V1  V2  V3  V4  V5'
    'NMISS' 60
END IF
PRINT / $CASENUM 1-4 MPN 15-18  MD1 TO MD5 (15X,5(A3,1X)) MISSV
60-62
EXECUTE
```

Abbildung E-1: Druckausgabe eines MD-Pattern mit SPSS-X

```
******** M D - P A T T E R N  A N A L Y S I S ********

PATTERN NR.   N OF CASES         V1  V2  V3  V4  V5        NMISS
   1            8                                           0
   2            1                                   M       1
   3            2                               M           1
   4            1                               M   M       2
   5            2                       M                   1
   6            3                   M                       1
   7            2               M                           1
   8            1               M           M               2
```

E.3 Hazard-Modelle

Seit der Version 2.0 steht in SPSS-X die Prozedur PROBIT zur Verfügung, die die Berechnung von Probit- und Logit-Modellen erlaubt, die für die in Kapitel 4 erwähnten Korrekturverfahren erforderlich sind. Die Implementierung der PROBIT-Prozedur weist allerdings neben einigen Programmfehlern, die sich in nicht vom Programm abgefangenen Zusammenbrüchen ("underflow-error" mit FORTRAN-CPU-Registerdump) zeigen, einige Besonderheiten auf, die die Bedienung etwas komplizieren und daher im Interesse einer praktischen Anwendung kurz erläutert werden sollen.
Zunächst einmal empfiehlt sich, nicht mit Daten auf Fallbasis zu arbeiten, sondern mit einer aggregierten Datei. PROBIT produziert zwar korrekte Parameterschätzungen, falls die Daten in Rohform an das Programm übergeben werden, aber der Chi-Quadrat- Test wird falsch berechnet, daneben wird der Output extrem umfangreich, sofern die Voreinstellungen benutzt werden.
Für ein Hazard-Modell kann z.B. folgendermaßen vorgegangen werden: Angenommen werden fehlende Werte ausschließlich auf der Variablen V1; die Variablen V2 und V3 sollen für die substantielle Gleichung verwendet werden, die Variablen V4 und V5 als Prädiktoren für die Selektion.

```
COMMENT ******************************************************
COMMENT * H A Z A R D - M O D E L L : Logistische Regression *
COMMENT ******************************************************
COMMENT
FILE HANDLE REIN  / NAME='DEMO SYS A'
GET FILE=REIN
SORT CASES BY V4 V5
AGGREGATE OUTFILE=*
 /BREAK= V4 V5
 /TOTAL= N (V4)
 /MISS = NMISS(V1)
PROBIT NMISS OF TOTAL WITH V4 V5
 /MODEL= LOGIT
 /LOG= NONE
 /PRINT=FREQ
```

Nach der Durchführung der Aggregierung stellt die aktuelle Datei eine Aggregatdatei mit den Variablen NMISS, TOTAL, V4 und V5 dar. PROBIT führt dann eine logistische Regression durch, um die Wahrscheinlichkeiten für das Fehlen auf V1 (NMISS OF TOTAL) durch V4 und V5 zu erklären. Wichtig hierbei ist die Unterdrückung der Voreinstellung der logarithmischen Transformation der Prädiktoren und die Prüfung auf Konvergenz, denn PROBIT berechnet auch dann Parameterschätzungen, wenn die Schätzungen nicht konvergieren. Ebenso sollte der Chi-Quadrat-Test auf Homogenität der Residuen nicht signifikant sein, da ein hoher Chi-Quadrat-Wert hier z.B. auf ein falsch spezifiziertes Reaktionsmodell hinweisen kann.
Die im Output zu findenen Regressionsparameter führen noch nicht direkt zu den gewünschten Wahrscheinlichkeiten: Obwohl sie über PRINT=FREQ ausgegeben werden, können sie durch PROBIT nicht der Systemdatei zugefügt werden.
Um die Ergebnisse der logistischen Regression leichter mit den Ergebnissen einer Probit-Analyse vergleichbar zu machen, berechnet PROBIT logistische Regressionen mit einer leicht veränderten Transformationsregel. Sollen die nach PRINT= FREQ zu findenden Wahrscheinlichkeiten den Fällen der ursprünglichen Datei als Variablen zugeschrieben werden, muß diese Transformation wieder aufgehoben werden. Angenommen, es hätte sich eine Konstante von 6.9 ergeben, die Regressionsparameter wären -.27 für V4 und -.22 für V5.

```
COMMENT ******************************************************
COMMENT * H A Z A R D - M O D E L L : Logistische Regression *
COMMENT *     Vorhersage der Nonresponsewahrscheinlichkeiten *
COMMENT ******************************************************
COMMENT
FILE HANDLE REIN  / NAME='DEMO SYS A'
GET FILE=REIN
COMPUTE EST=  ( 1.9  - .27  *  V4  -  .22 *  V5 ) * 2.0
COMPUTE PRED= EXP (EST) / (  1 + EXP (EST) )
REGRESSION VARS= V1 V2 V3 PRED
 /DEP= V1 /ENTER
```

Von den Regressionsvorhersagen muß zunächst die Konstante 5.0 subtrahiert werden (dies geht auf eine übliche Standardisierung bei der Probit-Analyse zurück (vgl. FINNEY 1971) und besitzt keine inhaltliche Bedeutung). Dann muß die Schätzung mit der Konstanten 2.0 multipliziert werden, um die Standardisierung von PROBIT zu beseitigen. Schließlich werden die Logits in die Schätzungen der Wahrscheinlichkeiten (PRED) umgerechnet und in die substantielle Gleichung mit einbezogen.

Soll ein Probit-Modell gerechnet werden, so liegt der wesentliche Unterschied nur in der Umrechnung der Regressionvorhersage. Die Wahrscheinlichkeiten werden durch die Berechnung der Regression mit den Parameterschätzungen des Probit-Modells durchgeführt, die entsprechende Variable, z.B. PRO, von der wiederum 5.0 subtrahiert werden muß, wird dann mit der COMPUTE-Funktion CDFNORM(PRO) umgerechnet. Für die Hazardfunktion wird dann noch die Dichtefunktion der Standardnormalverteilung benötigt, die nicht als SPSS-X intrinsische Funktion zur Verfügung steht. Diese kann aber über COMPUTE D= .398942 * EXP (-1.0 * PRO**2 / 2) berechnet werden.

F. Literaturverzeichnis

AFIFI,A.A., ELASHOFF,R.M. (1966): Missing Observations in Multivariate Statistics. I. Review of the Literature, in: JASA, 61, S.595-604

AFIFI,A.A., ELASHOFF,R.M. (1967): Missing Observations in Multivariate Statistics. II. Point Estimation in Simple Linear Regression, in: JASA, 62, S.10-29

AFIFI,A.A., ELASHOFF,R.M. (1969): Missing Observations in Multivariate Statistics. III. Large Sample Analysis of Simple Linear Regression, in: JASA, 64, S.337-358

AFIFI,A.A., ELASHOFF,R.M. (1969): Missing Observations in Multivariate Statistics. IV. A Note on Simple Linear Regression, in: JASA, 64, S.359-365

AIKEN,L.R. (1981): Proportions of Returns in Survey Research, in: Educational and Psychological Measurement, 41, 4, S.1033-1038

AITKIN,M. (1981): A Note on the Regression Analysis of Censored Data, in: Technometrics, 23, 2, S.161-163

ALBERT,J.H. (1982): A Bayesian Treatment of Nonresponse When Sampling from a Dichotomous Population, in: ASA Proceedings of the Section on Survey Research Methods, S.314-317

ALEMANN,H.v. (1977): Der Forschungsprozeß, Stuttgart

ALLERBECK,K. (1972): Datenverarbeitung in der empirischen Sozialforschung, Stuttgart

ALLERBECK,K. (1977): Computerunterstützte Datenaufbereitung und Datenanalyse, in: KOOLWIJK,J.v./WIEKEN-MAYSER,M.(Hrsg.): Techniken der empirischen Sozialforschung, S.170-188, München

ALLIGER,G.M., ALEXANDER,R.A. (1984): Correcting for Multivariate Range Restriction: Two Computer Programs, in: Educational and Psychological Measurement, 44, 677-678

AMEMIYA,T. (1973): Regression Analysis When the Dependent Variable Is Truncated Normal, in: Econometrica, 41, 6, S.997-1016

AMEMIYA,T. (1974): Multivariate Regression and Simultaneous Equation Models When the Dependent Variables Are Truncated Normal, in: Econometrica, 42, 6, S.999-1012

ANDERSON,A.B., BASILEVSKY,A., HUM,D.J. (1983): Missing Data: A Review of the Literature, in: ROSSI,P.H./WRIGHT,J.D./ANDERSON, A.B. (Hrsg.): Handbook of Survey Research, S.415-493, New York

ANDERSON,H. (1979): On Nonresponse Bias and Response Probabilities, in: Scandinavian Journal of Statistics, 6, S.107-112

ANDERSON,J.I. (1976): A Comparison of three Methods of Handling Missing Data in MANOVA, Ph.D.Thesis, Arizona State University

ANDERSON,T.W. (1957): Maximum Likelihood Estimates for a Multivariate Normal Distribution When Some Observations Are Missing, in: JASA, 52, S.200-203

ANSCOMBE,F.J. (1972): Outliers, in: SILLS,D.L. (Hrsg.): International Encyclopedia of the Social Sciences, Vol.15, S.178-181, New York

ARMINGER,G. (1976): Anlage und Auswertung von Paneluntersuchungen, in: HOLM,K. (Hrsg.): Die Befragung Bd.4, S.134-235, München

ATKINSON,J.W. (1975): Einführung in die Motivationsforschung, Stuttgart

BAILAR,B.A., BAILAR,J.C. (1983): Comparison of the Biases of the Hot-Deck Imputation Procedure with an "Equal-Weights" Imputation Procedure, in: MADOW,W.G./OLKIN,I. (Hrsg.): Incomplete Data in Sample Surveys, Vol.3, S.299-311, New York

BAILAR,J.C., BAILAR,B.A. (1978): Comparison of Two Procedures for Imputing Missing Survey Values, in: ASA Proceedings of the Section on Survey Research Methods, S.462-467

BAKER,R.J., NELDER,J.A. (1978): The GLIM System, Release 3, Generalized Linear Interactive Modelling, Oxford

BALESTRA,P., KMENTA,J. (1984): Missing Measurements in a Regression Problem with no Auxiliary Relations. Working paper R-113.84, Department of Economics, University of Michigan

BANISTER,J. (1980): Use and Abuse of Census Editing and Imputation, in: Asian and Pacific Census Forum, 6, 3, S.1-2,16-18,20

BARNETT,V. (1983): Principles and Methods for Handling Outliers in Data Sets, in: WRIGHT,T. (Hrsg.): Statistical Methods and the Improvement of Data Quality, S.131-166, Orlando

BARNOW,B.S., CAIN,G.C., GOLDBERGER,A.S. (1980): Issues in the Analysis of Selectivity Bias, in: STROMSDORFER,E.W./FARKAS,G. (Hrsg.): Evaluation Studies Review Annual, 5, S.43-59, Beverly Hills

BATESON,N. (1984): Data Construction in Social Surveys, London

BEALE,E.M.L., LITTLE,R.J.A. (1975): Missing Values in Multivariate Analysis, in: Journal of the Royal Statistical Society, Series B, 37, S.129-146

BECKMANN,R.J., COOK,R.D. (1983): Outlier.....s, in: Technometrics, 25, 2, S.119-149

BELL,R. (1984): Item Nonresponse in Telephone Surveys: An Analysis of Who Fails to Report Income, in: Social Science Quarterly, S.207-215

BERGER,M.P.F. (1979): A FORTRAN IV Program for the Estimation of Missing Data, in: Behavior Research Methods and Instrumentation, 11,3, S.395-396

BERK,R.A., RAY,S.C. (1982): Selection Biases in Sociological Data, in: Social Science Research, 11, S.352-398
BERK,R.A. (1983): An Introduction to Sample Selection Bias in Sociological Data, in: American Sociological Review, 48, S.386- 398

BIEMER,P.P. (1980): A Survey Error Model which Includes Edit and Imputation Error, in: ASA Proceedings of the Section on Survey Research Methods, S.610-615

BIERVERT,B., WALZEL,A. (1969): Automatischer Split zur optimalen Kombination erklärender Faktoren, in: KZfSS, 21, S.304-323

BIRNBAUM,Z.W., SIRKEN,M.G. (1950): Bias Due to Non-Availability in Sampling Surveys, in: JASA, 45, S.98-111

BISCO,R.L. (1970): Handling Missing Data Problems, in: BISCO,R.L. (Hrsg.): Data Bases, Computers and the Social Sciences, S.173- 175, New York

BISHOP,G.F., OLDENDICK,R.W., TUCHFARBER,A.J. (1983): Effects of Filter Questions in Public Opinion Surveys, in: Public Opinion Quarterly, 47, S.528-546

BISHOP,G.F., OLDENDICK,R.W., TUCHFARBER,A.J., BENNETT,S.E.(1980): Pseudo-Opinions on Public Affairs, in: Public Opinion Quarterly, 44, S.198-209

BLAU,P.M., DUNCAN,O.D. (1967): The American Occupational Structure, New York

BOAS,J. (1967): A Note on the Estimation of the Covariance between Two Random Variables Using Extra Information on the Separate Variables, in: Statistica Neerlandica, 21, 3/4, S.291- 292

BÖHNING,D., WILKE,H. (1980): Statistische Analysepakete und ihre Einbeziehung in sozialwissenschaftliche Studiengänge, in: Zeitschrift für Soziologie, 9, 3, S.290-307

BOGART,L. (1967): No Opinion, Don't Know and Maybe No Answer, in: Public Opinion Quarterly, 31, S.331-345

BOGESTRÖM,B., LARSSON,M., LYBERG,L. (1983): Bibliography on Nonresponse and Related Topics, in: MADOW,W.G./OLKIN,I./RUBIN,D.B. (Hrsg.): Incomplete Data in Sample Surveys, Vol.2, S.479-567, New York

BOHANNON,T.R. (1976): Discriminant Analysis with Missing Data, Ph.D. Thesis, Texas A.& M. University

BOHANNON,T., SMITH,W.B. (1976): Classification Based on Incomplete Data Records, in: ASA Proceedings of the Social Statistics Section, S.214-218

BOLLEN,K.A., BARB,K.H. (1981): Pearson's r and Coarsely Categorized Measures, in: American Sociological Review, 46, S.232-239

BOLLINGER,G., HERRMANN,A., MÖNTMANN,V. (1983): BMDP Statistikprogramme für die Bio-, Human- und Sozialwissenschaften, Stuttgart

BORGATTA,E.F. (1965): Difficulty Factors and the Use of r-phi, in: The Journal of General Psychology, 73, S.321-337
BORTZ,J. (1977): Lehrbuch der Statistik, Berlin

BOYLES,R.A. (1983): On the Convergence of the EM Algorithm, in: Journal of the Royal Statistical Society, Series B, 45, 1, S.47- 50

BRADBURN,N.M. (1978): Respondent Burden, in: ASA Proceedings of the Section on Survey Research Methods, S.35-40

BRADBURN,N.M., SUDMAN,S., BLAIR,E., STOCKING,C. (1978): Question Threat and Response Bias, in: Public Opinion Quarterly, 42, S.221-234

BRADBURN,N.M. (1981): Improving Interview Method and Questionnaire Design, San Francisco

BRENNER,M. (1981): Patterns of Social Structure in the Research Interview, in: BRENNER,M. (Hrsg.): Social Method and Social Life, S.115-158, London

BRENNER,M. (1982): Response Effects of "Role-restricted" Characteristics of the Interviewer, in: DIJKSTRA,W./VAN DER ZOUWEN,J. (Hrsg.): Response Behaviour in the Survey-interview, S.131-165, London

BRONNER,A.E. (1983): A Way to Handle Nonresponse: Simulation of Missing Values, in: European Research, S.86-92

BROUWER,U., VIJN,P. (1980): A Program to Estimate the Correlation Coefficient in Incomplete Datasets, in: International Association for Statistical Computing: COMPSTAT 1980, S.194-200, Wien

BROUWER,U., VIJN,P. (1980b): Missing Data and Bayesian Estimators of the Correlation Coefficient, Technical Report, Technical Center University of Amsterdam

BROWN,C.H. (1983): Asymptotic Comparison of Missing Data Procedures for Estimating Factor Loadings, in: Psychometrika, 48, 2, S.269-291

BROWN,P.B. (1979): BEMSD: A General-purpose Program for Estimating Missing Data in a Sociomatrix, in: Behavior Research Methods and Instrumentation, 11, 4, S.465-466

BRYANT,N.D., GOKHALE,S. (1972): Correcting Correlations for Restrictions in Range Due to Selection on an Unmeasured Variable, in: Educational and Psychological Measurement, 32, S.305-310

BUCK,S.F. (1960): A Method of Estimation of Missing Values in Multivariate Data Suitable for Use with an Electronic Computer, in: Journal of the Royal Statistical Society, Series B, 22, 2, S.302-306

BUNGARD,W., LÜCK,H.E. (1974): Forschungsartefakte und nicht- reaktive Meßverfahren, Stuttgart

CHAN,L.S., DUNN,O.J. (1972): The Treatment of Missing Values in Discriminant Analysis - I. The Sampling Experiment, in: JASA, 67, 338, S.473-477

CHAN,L.S., GILMAN,J.A., DUNN,O.J. (1976): Alternative Approaches to Missing Values in Discriminant Analysis, in: JASA, 71, S.842- 844

CHAPMAN,D.W. (1976): A Survey of Nonresponse Imputation Procedures, in: ASA Proceedings of the Social Statistics Section, S.245-251

CHAPMAN,D.W. (1983): An Investigation of Nonresponse Imputation Procedures for the Health and Nutrition Examination Survey, in: MADOW,W.G./NISSELSON,H./OLKIN,I. (Hrsg.): Incomplete Data in Sample Surveys, Vol.1, S.435-483, New York

CHAPMAN,D.W. (1983): The Impact of Substitution on Survey Estimates, in: MADOW,W.G./OLKIN,I./RUBIN,D.B. (Hrsg.): Incomplete Data in Sample Surveys, Vol.2, S.45-61, New York

CHILTON,S.P. (1980): Discussion, in: ASA Proceedings of the Section on Survey Research Methods, S.257-258

COCHRAN,W.G. (1972): Stichprobenverfahren, Berlin

COHEN,A.C. (1955): Restriction and Selection in Samples from Bivariate Normal Distributions, in: JASA, 50, S.884-893

COHEN,J. (1968): Multiple Regression as a General Data-Analytic System, in: Psychological Bulletin, 70, S.426-443

COHEN,J., COHEN,P. (1975): Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences, Hillsdale

CONNIFFE,D. (1983): Comments on the Weighted Regression Approach to Missing Values, in: The Economic and Social Review, 14, 4, S.259-272

CONNIFFE,D. (1983b): Small-Sample Properties of Estimators of Regression Coefficients Given a Common Pattern of Missing Data, in: Review of Economic Studies, L, 111-120

CONVERSE,P.E. (1970): Attitudes and Non-Attitudes: Continuation of a Dialogue, in: TUFTE,E.R. (Hrsg.): The Quantitative Analysis of Social Problems, Reading,Mass.

COOLEY,W.W., LOHNES,P.R. (1971): Multivariate Data Analysis, New York

COOMBS,C.H., COOMBS,L.C. (1976): "Don't Know": Item Ambiguity or Respondent Uncertainty, in: Public Opinion Quarterly, 40, S.497- 514

COX,B.G. (1980): The Weighted Sequential Hot Deck Imputation Procedure, in: ASA Proceedings of the Section on Survey Research Methods, S.721-726

COX,B.G., FOLSOM,R.E. (1978): An Empirical Investigation of Alternative Item Nonresponse Adjustments, in: ASA Proceedings of the Section on Survey Research Methods, S.219-223

CRAIG,C.S., McCANN,J.M. (1978): Item Nonresponse in Mail Surveys: Extent and Correlates, in: Journal of Marketing Research, XV, S.285-289

CRAMER,E.M., NICEWANDER,W.A. (1979): Some Symmetric, Invariant Measures of Multivariate Association, in: Psychometrika, 44, 1, S.43-54

CRANO,W.D., BREWER,M.B. (1975): Einführung in die sozialpsychologische Forschung, Köln

DAGENAIS,M.G. (1971): Further Suggestions Concerning the Utilization of Incomplete Observations in Regression Analysis, in: JASA, 66, S.93-98

DAGENAIS,M.G. (1972): Asymptotic Behaviour and Small Sample Performance: Experiments on Regression Parameter Estimation with Incomplete Observations, in: European Economic Review, 3, S.389- 398

DAGENAIS,M.G. (1973): The Use of Incomplete Observations in Multiple Regression Analysis. A Generalized Least Squares Approach, in: Journal of Econometrics, 1, S.317-328

DAGENAIS,M.G. (1974): Multiple Regression Analysis with Incomplete Observations from a Bayesian Viewpoint, in: FIENBERG,S.E./ ZELLNER,A. (Hrsg.): Studies in Bayesian Econometrics and Statistics, S.274-287, Amsterdam

DAGENAIS,M.G. (1976): Incomplete Observations and Simultaneous- Equations Models, in: Journal of Econometrics 4, S.231-241

DALENIUS,T. (1962): Automatic Estimation of Missing Values in Censuses and Sample Surveys, in: Statistical Review, 11, 7, S.395-400

DALENIUS,T. (1977): Bibliography on Non-sampling Errors in Surveys, in: International Statistical Review,45. Teil I (A bis G) S.71-89, Teil II (H bis Q) S.181-197, Teil III (R bis Z) S.303- 317

DALENIUS,T.E. (1981): The Survey Statistician's Responsibility for Both Sampling and Measurement Errors, in: KREWSKI,D./ PLATEK,R./RAO,J.N.K. (Hrsg.): Current Topics in Survey Sampling, S.17-29, New York

DANIEL,W.W., SCHOTT,B., ATKINS,F.C., DAVIS,A. (1982): A Adjustment for Nonresponse in Sample Surveys, in: Educational and Psychological Measurement, 42, S.57-67

DE GOEDE,W.H.V. (1984): Incomplete Samples from Multivariate Normal Distributions with the Same, Known Covariance Matrix, in: VAN VARK,G.N./HOWELLS,W.W. (Hrsg.): Multivariate Statistical Methods in Physical Anthropology, S.37-48, Dordrecht

DEAR,R.E. (1959): A Principal Component Missing Data Method for Multiple Regression Models. Technical Report SP-86, System Development Corporation, Santa Monica, Calif.

DEININGER,R., KERLER,E. (1971): Maschinelle Datenprüfung in der amtlichen Statistik, in: Allgemeines Statistisches Archiv, S.290- 302

DEMING,W.E., TEPPING,B.J., GEOFFREY,L. (1942): Errors in Card Punching, in: JASA, 37, S.525-536

DEMPSTER,A.P. (1971): An Overview of Multivariate Data Analysis, in: Journal of Multivariate Analysis, 1, S.316-346

DEMPSTER,A.P., LAIRD,N.M., RUBIN,D.B. (1977): Maximum Likelihood from Incomplete Data via the EM Algorithm, in: Journal of the Royal Statistical Society, Series B, 39, S.1-22

DEMPSTER,A.P. (1978): Discussion, in: ASA Proceedings of the Section on Survey Research Methods, S.30-31

DILLMAN,D.D. (1978): Mail and Telephone Surveys: The Total Design Method, New York

DILLON,W.R., GOLDSTEIN,M. (1984): Multivariate Analysis. Methods and Applications, New York

DINGMAN,H.F. (1958): The Relation between Coefficients of Correlation and Difficulty Factors, in: The British Journal of Statistical Psychology, 11, 1, S.13-17

DIXON,W.J., BROWN,L., ENGELMAN,L., FRANE,J.W., HILL,M.A., JENNRICH,R.I., TOPOREK,J.D. (1981): BMDP Statistical Software 1981, Berkeley

DOMENCICH,T., McFADDEN,D. (1975): Urban Travel Demand: A Behavioral Analysis, Amsterdam

DONNER,A. (1982): The Relative Effectiveness of Procedures Commonly Used in Multiple Regression Analysis for Dealing with Missing Values, in: The American Statistician, 36, 4, S.378-381

DONNER,A., ROSNER,B. (1982): Missing Value Problems in Multiple Linear Regression with Two Independent Variables, in: Communications in Statistics A, Theory and Methods, 11, 2, S.127-140

DUNSMUIR,W., ROBINSON,P.M. (1981): Estimation of Time Series Models in the Presence of Missing Data, in: JASA, 76, S.560-568

DURAND,R.M., GUFFEY,H.J., PLANCHON,J.M. (1983): An Examination of the Random Versus Nonrandom Nature of Item Omissions, in: Journal of Marketing Research, XX, S.305-313

DWYER,J.H. (1983): Statistical Models for the Social and Behavioral Sciences, New York

DeLAMATER,J. (1982): Response-effects of Question Content, in: DIJKSTRA,W./VAN DER ZOUWEN,J.(Hrsg.): Response Behaviour in the Survey-interview, S.13-48, London

EDGETT,G.L. (1956): Multiple Regression with Missing Observations Among the Independent Variables, in: JASA, 51, S.122-131

EKLUND,G. (1959): Studies of Selection Bias in Applied Statistics, Uppsala

ELASHOFF,J.D., ELASHOFF,R.M. (1974): Two-Sample Problems for a Dichotomous Variable with Missing Data, in: Applied Statistics, 23, S.26-34

ELDER,S.R. (1979): One Approach to the Practice of Computer Editing. Diskussionspapier der Konferenz "Getting the Data Right" der Study Group on Computers in Survey Analysis, London, City University

ELSHOUT,J., ROE,R.A. (1973): Restriction of Range in the Population, in: Educational and Psychological Measurement, 33, S.53-62

ELSHOUT,J., OVERBEEK,H.Van, ROE,R., VIJN,P. (1979): Testing the Hypothesis That Rho = 0 in Selected Samples (Case 1), in: Educational and Psychological Measurement, 39, S.573-576

ENGEL,B. (1984): Zur Erhebbarkeit des Merkmals Einkommen in Repräsentativumfragen - Untersuchungen zur Struktur fehlender Werte und zur Konsistenz der Angaben im Arbeitnehmersurvey. Manuskript, Universität Frankfurt, SFB-3

ENGELMAN,L. (1982): An Efficient Algorithm for Computing Covariance Matrices from Data with Missing Values, in: Communications in Statistics, Simulation and Computation, B11, 1, S.113-121

ERDOS,P.L., MORGAN,A.J. (1970): Professional Mail Surveys, New York

ESSER,E. (1982): Ausländerinnen in der Bundesrepublik Deutschland. Eine soziologische Analyse des Eingliederungsverhaltens ausländischer Frauen, Frankfurt

ESSER,H. (1975): Soziale Regelmäßigkeiten des Befragtenverhaltens, Meisenheim am Glan

ESSER,H. (1979): Methodische Konsequenzen gesellschaftlicher Differenzierung, in: Zeitschrift für Soziologie, 8, 1, S.14-27

ESSER,H. (1983): Fehler bei der Datenerhebung, Kurseinheit 3: Datenerhebung als sozialer Prozeß (Studienbrief Fernuniversität Hagen)

ESSER,H. (1984): Determinanten des Interviewer- und Befragtenverhaltens: Probleme der theoretischen Erklärung und empirischen Untersuchung von Interviewereffekten, in: MAYER,U./SCHMIDT,P. (Hrsg.): Allgemeine Bevölkerungsumfrage der Sozialwissenschaften, S.26-71, Frankfurt

EVANS,R.W., COOLEY,P.C., PISERCHIA,P.V. (1979): A Test for Evaluating Missing Data Imputation Procedures, in: ASA Proceedings of the Social Statistics Section, S.469-474

FAHRMEIR,L., HAMERLE,A. (1984): Multivariate statistische Verfahren, Berlin

FAULKENBERRY,G.D., MASON,R. (1978): Characteristics of Nonopinion and No Opinion Response Groups, in: Public Opinion Quarterly, 42, S.533-543

FEDERSPIEL,C.F., MONROE,R.J., GREENBERG,B.G. (1959): An Investigation of Some Multiple Regression Methods for Incomplete Samples, University of North Carolina Institute of Statistics, Memeo Series #236

FELLEGI,I.P., HOLT,D. (1976): A Systematic Approach to Automatic Edit and Imputation, in: JASA, 71, 353, S.17-35

FERBER,R. (1966): Item Nonresponse in a Consumer Survey, in: Public Opinion Quarterly, 30, S.399-415

FIELDING,A. (1979): Binary Segmentation: The Automatic Interaction Detector and Related Techniques for Exploring Data Structure, in: O'MUIRCHEARTAIGH,C.A./PAYNE,C. (Hrsg.): The Analysis of Survey Data, Bd.1, S.221-257, Chichester

FIERING,M.B. (1962): On the Use of Correlation to Augment Data, in: JASA, March, S.20-32

FINKBEINER,C. (1979): Estimation for the Multiple Factor Model When Data Are Missing, in: Psychometrika, 44, S.409-420

FINNEY,D.J. (1971): Probit Analysis, Cambridge

FINNEY,D.J. (1974): Problems, Data, and Inference, in: Journal of the Royal Statistical Society, Series A, 137, S.1-19

FORD,B.L. (1976): Missing Data Procedures: A Comparative Study, in: ASA Proceedings of the Social Statistics Section, S.324-329

FORD,B.L. (1983): An Overview of Hot-Deck Procedures, in: MADOW,W.G/OLKIN,I./RUBIN,D.B. (Hrsg.): Incomplete Data in Sample Surveys, Vol.2, S.185-207, New York

FORD,B.L., KLEWENO,D.G., TORTORA,R.D. (1980): The Effects of Procedures Which Imputes for Missing Items: A Simulation Study Using an Agricultural Survey, in: ASA Proceedings of the Section on Survey Research Methods, S.251-256

FORSYTH,R.A. (1971): An Empirical Note on Corrrelation Coefficients Corrected for Restriction in Range, in: Educational and Psychological Measurement, 31, S.115-123

FRANCES,J.D., BUSCH,L. (1975): What We Don't Know about "I don't Knows", in: Public Opinion Quarterly, 39, S.207-218

FRANE,J.W. (1976): Some Simple Procedures for Handling Missing Data in Multivariate Analysis, in: Psychometrika, 41, 3, S.409- 415

FRANE,J. (1978): Missing Data and BMDP: Some Pragmatic Approaches, in: ASA Proceedings of the Statistical Computing Section, S.27-33

FRANKEL,M.R., BANKS,M.J. (1979): Adjusting for Nonresponse to Specific Questions, in: ANDERSON,R./KASPER,J./FRANKEL,M.R. (Hrsg.): Total Survey Error, S.97-103, San Francisco

FREUND,R.J., HARTLEY,H.O. (1967): A Procedure for Automatic Data Editing, JASA, 62, S.341-352

FREUND,R.J., MINTON,P.D. (1979): Regression Methods, New York

FREY,J.H. (1983): Survey Research by Telephone, Beverly Hills

FRIEDMAN,J.H., BENTLEY,J.L., FINKEL,R.A. (1977): An Algorithm for Finding Best Matches in Logarithmic Expected Time, in: ACM Transactions on Mathematical Software, 3, 3, S.209-226

FUCHS,C. (1982): Maximum Likelihood Estimation and Model Selection in Contingency Tables with Misssing Data, in: JASA, 77, 378, S.270-278

GAENSSLEN,H., SCHUBÖ,W. (1976): Einfache und komplexe statistische Analyse, München

GLASSER,M. (1964): Linear Regression Analysis with Missing Observations Among the Independent Variables, in: JASA, 59, S.834-844

GLEASON,T.C., STAELIN,R. (1975): A Proposal for Handling Missing Data, in: Psychometrika, 40, S.229-252

GOLDBERGER,A.S. (1981): Linear Regression after Selection, in: Journal of Econometrics, 15, S.357-366

GOODNIGHT,J.H. (1979): A Tutorial on the SWEEP Operator, in: The American Statistician, 33, 3, S.149-158

GOURIEROUX,C., MONFORT,A. (1981): On the Problem of Missing Data in Linear Models, in: Review of Economic Studies, S.579-586

GREENE,W.H. (1981): On the Asymptotic Bias of the Ordinary Least Squares Estimator of the Tobit Model, in: Econometrica, 49, 2, S.505-513

GREENER,J.M., OSBURN,H.G. (1979): An Empirical Study of the Accuracy of Corrrections for Restriction in Range Due to Explicit Selection, in: Applied Psychological Measurement, 3, 1, S.31-41

GREENER,J.M., OSBURN,H.G. (1980): Accuracy of Corrections for Restriction in Range Due to Explicit Selection in Heteroscedastic and Nonlinear Distributions, in: Educational and Psychological Measurement, 40, S.337-346

GREENHOUSE,S.W., GEISSER,S. (1959): On Methods in the Analysis of Profile Data, in: Psychometrika, 24, 2, S.95-112

GREENLEES,J.S., REECE,W.S., ZIESCHANG,K.D. (1982): Imputation of Missing Values When the Probability of Response Depends on the Variable Beeing Imputed, in: JASA, 77, 378, S.251-261

GRILICHES,Z., HALL,B.H., HAUSMAN,J.A. (1978): Missing Data and Self-selection in Large Panels, in: Annales de l'INSEE, S.30-31, S.137-176

GROSS,A.L. (1982): Relaxing the Assumptions Underlying Corrections for Restriction of Range, in: Educational and Psychological Measurement, 42, S.795-801

GROSS,A.L., FLEISCHMAN,L. (1983): Restriction of Range Corrections When Both Distribution and Selection Assumptions Are Violated, in: Applied Psychological Measurement, 7, 2, S.227-237

GROSS,A.L., KAGEN,E. (1983): Not Correcting for Restriction of Range Can Be Advantageous, in: Educational and Psychological Measurement, 43, S.389-396
GROSS,A.L., PERRY,P. (1983): Validating a Selection Test: A Predictive Probability Approach, in: Psychometrika, 48, 1, S.113-127

GROVES,R.M., KAHN,R.L. (1979): Surveys by Telephone. A National Comparison with Personal Interviews, New York

GROVES,R.M., MAGILAVY,L.J. (1981): Increasing Response Rates to Telephone Surveys: A Door in the Face for Foot-in-the-Door? in: Public Opinion Quarterly, 45, S.346-358

GUERTIN,W.H. (1968): Comparison of Three Methods of Handling Missing Observations, in: Psychological Reports, 22, S.896

GUILFORD,J.P. (1956): Fundamental Statistics in Psychology and Education, 3.Aufl., New York

GULLICKSON,A., HOPKINS,K. (1976): Interval Estimation of Correlation Coefficients Corrected for Restriction of Range, in: Educational and Psychological Measurement, 36, S.9-25

GULLIKSEN,H. (1950): Theory of Mental Tests, New York

HAGENAARS,J.A., HEINEN,T.G. (1982): Effects of Role-independent Interviewer Characteristics on Responses, in: DIJKSTRA,W./VAN DER ZOUWEN,J. (Hrsg.): Response Behaviour in the Survey-interview, S.91-130, London

HAITOVSKY,Y. (1966): Estimation of Multivariate Statistics from Grouped and Missing Data, Ph.D. Thesis, Harvard University

HAITOVSKY,Y. (1968): Missing Data in Regression Analysis, in: Journal of the Royal Statistical Society, Series B, 63, S.67-82

HAMILTON,M.A. (1975): How to Estimate Multiple Regression Coefficients When Some Data Are Missing. Technical Report 1-8-75, Statistical Laboratory, Department of Mathematics, Montana State University

HAMILTON,M.A. (1975b): Regression Analysis When There Are Missing Observations - A Survey and Bibliography. Technical Report 1-3- 75. Statistical Laboratory, Montana State University

HANUSHEK,E.A., JACKSON,J.E. (1977): Statistical Methods for Social Scientists, New York

HARMAN,H.H. (1976): Modern Factor Analysis, Chicago

HARTIGAN,J.A. (1975): Clustering Algorithms, New York

HARTLEY,H.O., HOCKING,R.R. (1971): The Analysis of Incomplete Data, in: Biometrics, 27, S.783-823

HASSELBLAD,V., CREASON,J.P., STEAD,A.G. (1983): Applications of the Missing-Information Principle, in: MADOW,W.G./OLKIN,I. (Hrsg.): Incomplete Data in Sample Surveys, Vol.3, S.183-202, New York

HATTIE,J. (1983): The Tendency to Omit Items: Another Deviant Response Characteristic, in: Educational and Psychological Measurement, 43, S.1041-1045

HAUCK,W.W., ANDERSON,S. (1984): A Survey Regarding the Reporting of Simulation Studies, in: The American Statistician, 38, 3, S.214-216

HAUSMAN,J.A., WISE,D.A. (1977): Social Experimentation, Truncated Distributions, and Efficient Estimation, in: Econometrica, 45, 4, S.919-938

HAWKINS,D.I, CONEY,K.A. (1981): Uninformed Response Error in Survey Research, in: Journal of Marketing Research, XVIII, S.370- 374

HECKMAN,J.J. (1976): The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models, in: Annals of Economic and Social Measurement, 5/4, S.475-492

HECKMAN,J.J. (1979): Sample Selection Bias as a Specification Error, in: Econometrica, 47, 1, S.153-161

HEIBERGER,R.M. (1977): Regression with the Pairwise-Present Covariance Matrix: A Dangerous Practice, in: ASA Proceedings of the Statistical Computing Section, S.38-47

HEIBERGER,R.M. (1978): Sample Indicators of Perturbation for Missing Value Regression Techniques, in: International Association for Statistical Computation: COMPSTAT 1978, S.88-94, Wien

HERTEL,B.R. (1976): Minimizing Error Variance Introduced by Missing Data Routines in Survey Analysis, in: Sociological Methods and Research, 4, 4, S.459-474

HERZOG,T.N., RUBIN,D.B. (1983): Using Multiple Imputations to Handle Nonresponse in Sample Surveys, in: MADOW,W.G./OLKIN,I./ RUBIN,D.B. (Hrsg.): Incomplete Data in Sample Surveys, Vol.2, S.209-245, New York

HESTER,D.D. (1976): Estimation from Incomplete Samples, Technical Report, University of Wisconsin

HILL,C.J. (1978): A Report on the Application of a Systematic Method of Automatic Edit and Imputation to the 1976 Canadian Census, in: ASA Proceedings of the Section on Survey Research Methods, S.474-479

HILL,M., DIXON,W.J. (1981): Missing Data: Search for Patterns, in: ASA Proceedings of the Statistical Computing Section, S.57-60

HINKINS,S. (1980): RFACTOR - A Program to Create Rubin's Factorization When There Are Incomplete Multivariate Data, in: The American Statistician, 34, 3, S.182

HINKINS,S.M. (1980b): RFACTOR - A Computer Program to Create Rubin's Table for Factoring a Multivariate Estimation Problem When Some Data Are Missing, Technical Report 1-12-76, rev. June 1980, Montana State University

HINKINS,S.M., HAMILTON,M.A. (1976): MISSMLE - A Computer Program for Calculating Maximum Likelihood Estimates When Some Data Are Missing, Technical Report 1-6-76, Statistical Center, Department of Mathematics, Monatana State University

HOAGLIN,D.C., ANDREWS,D.F. (1975): The Reporting of Computation- Based Results in Statistics, in: The American Statistician, 29, 3, S.122-126

HOCKING,R.R., HUDDLESTON,H.F., HUNT,H.H. (1974): A Procedure for Editing Survey Data, in: Applied Statistics, 23, S.121-133

HOLM,K. (1975): Ablochung der Daten und automatische Fehlersuche, in: HOLM,K. (Hrsg.): Die Befragung Bd.2, S.22-45, Stuttgart

HOLM,K. (1975b): Die Erstellung einer Korrelationsmatrix, in: HOLM,K. (Hrsg.): Die Befragung Bd.2, S.155-218, München

HOLM,K. (1979): Das allgemeine lineare Modell, in: HOLM,K. (Hrsg.): Die Befragung Bd.6, S.11-213, München

HUDDLESTON,H.F., HOCKING,R.R. (1978): Imputation in Agricultural Surveys, in: ASA Proceedings of the Section on Survey Research Methods, S.480-485

HUGHES,D.A. (1978): A Statistical Approach to the Analysis of Item Omission in Mail Surveys, in: ASA Proceedings of the Section on Survey Research Methods, S.594-598

HUGHES,D.A. (1979): The Correlates of Item Nonresponse in a Large Scale Mail Survey, in: ASA Proceedings of the Section on Survey Research Methods, S.363-365

HUNTER,J.E., SCHMIDT,F.L., JACKSON,G.B. (1982): Meta-analysis: Cumulating Research Findings Across Studies, Beverly Hills

HURST,R.L., KNOP,R.E. (1972): Generation of Random Correlated Normal Variables: Algorithm 425, in: Communications of the ACM, 15, 5, S.355-357

HUSEBY,J.R., SCHWERTMAN,N.C., ALLEN,D.M. (1980): Computation of the Mean Vector and Dispersion Matrix for Incomplete Multivariate Data, in: Communications in Statistics, Simulation and Computation, B9, 3, S.301-309

HUTCHESON,J.D., PRATHER,J.E. (1977): Assessing the Effects of Missing Data, in: ASA Proceedings of the Social Statistics Section, S.279-283

HUTCHESON,J.D., PRATHER,J.E. (1981): Interpreting the Effects of Missing Data in Survey Research, in: South-Eastern Political Review, 9, S.129-143

HUTCHESON,J.D., PRATHER,J.E., SNOW,R.E. (1979): The Effects of Interviewer Characteristics on Item Response, in: ASA Proceedings of the Section on Survey Research Methods, S.372-377

INNES,J.M. (1977): Extremity and "Don't Know" Sets in Questionnaire Response, in: British Journal of Social and Clinical Psychology, 16, S.9-12

JACKSON,E.C. (1968): Missing Values in Linear Multiple Discriminant Analysis, in: Biometrics, 24, S.835-844

JÄGER,R. (1974): Methoden zur Mittelung von Korrelationen, in: Psychologische Beiträge, 16, S.417-427

JENNRICH,R.J. (1970): An Asymptotic Test for the Equality of Two Correlation Matrices, in: JASA,65, S.904-912

JESSEN,R.J. (1978): Statistical Survey Techniques, New York

JÖRESKOG,K.G., SÖRBOM,D. (1984): LISREL VI - User's Guide, Department of Statistics, University of Uppsala

JOHNSON,D.R., CREECH,J.C. (1983): Ordinal Measures in Multiple Indicator Models: A Simulation Study of Categorization Error, in: American Sociological Review, 48, S.398-407

JOHNSON,N.L., KOTZ,S. (1970): Continuous Univariate Distributions, 2 Bde., Boston

JOHNSON,N.L., KOTZ,S. (1972): Continuous Multivariate Distributions, Boston

JORDAN,L.A., MARCUS,A.C., REEDER,L.G. (1980): Response Styles in Telephone and Household Interviewing: A Field Experiment, in: Public Opinion Quarterly, 44, S.210-222

JUDGE,G.G., GRIFFITHS,W.E., HILL,R.C., LEE,T.C. (1980): The Theory and Practice of Econometrics, New York

KAISER,J. (1983): The Effectiveness of Hot-Deck Procedures in Small Samples, in: ASA Proceedings of the Section on Survey Research Methods, S.523-528

KALSBEEK,W.D. (1980): A Conceptional Review of Survey Error Due to Nonresponse, in: ASA Proceedings of the Section on Survey Research Methods, S.131-136

KALTON,G. (1983): Compensating for Missing Survey Data, Survey Research Center, Institute for Social Research, University of Michigan

KALTON,G., KASPRZYK,D. (1982): Imputing for Missing Survey Responses, in: ASA Proceedings of the Section on Survey Research Methods, S.22-31

KALTON,G., KISH,L. (1981): Two Efficient Random Imputation Procedures, in: ASA Proceedings of the Section on Survey Research Methods, S.146-151

KALTON,G., SCHUMANN,H. (1980): The Effect of the Question on Survey Responses: A Review, in: ASA Proceedings of the Section on Survey Research Methods, S.30-39

KALTON,G., STOWELL,R. (1979): A Study of Coder Variablity, in: Applied Statistics, 28, S.276-289

KARWEIT,N., MEYERS,E.D. (1983): Computers in Survey Research, in: ROSSI,P.H./WRIGHT,J.D./ANDERSON,A.B. (Hrsg.): Handbook of Survey Research, S.379-414, New York

KAUFMAN,G.M., KING,B. (1973): A Bayesian Analysis of Nonresponse in Dichotomous Processes, in: JASA, 68, S.670-678

KELEJIAN,H.H. (1969): Missing Observations in Multivariate Regression: Efficiency of a First-order Method, in: JASA, 64, S.1609-1616

KENDALL,M., STUART,A. (1977): The Advanced Theory of Statistics, Vol. 1: Distribution Theory, London

KENNEDY,W.J., GENTLE,J.E. (1980): Statistical Computing, New York

KIM,J.O., CURRY,J. (1977): The Treatment of Missing Data in Multivariate Analysis, in: Sociological Methods and Research, 6, 2, S.215-239

KIM,J.O., NIE,N., VERBA,S. (1977): A Note on Factor Analyzing Dichotomous Variables: The Case of Political Participation, in: Political Methodology, 1, S.39-62

KISH,L. (1953): Selection of the Sample, in: FESTINGER,L./KATZ,D. (Hrsg.): Research Methods in the Behavioral Sciences, S.175-239, New York

KISH,L. (1983): Repeated Replication Imputation Procedure (RRIP), in: KALTON,G.: Compensating for Missing Survey Data, S.131-136, Survey Research Center, Institute for Social Research, The University of Michigan

KLEIJNEN,J.P.C. (1974): Statistical Techniques in Simulation, Part 1 & 2, New York

KMENTA,J. (1981): On the Problem of Missing Measurements in the Estimation of Economic Relationships, in: CHARATSIS,E.G. (Hrsg.): Proceedings of the Econometric Society European Meeting 1979, S.233-257, Amsterdam

KNUTH,D.E. (1973): The Art of Computer Programming, Vol.1: Fundamental Algorithms, 2.Aufl., Reading,Mass.

KOCH,G.G., IMREY,P.B., REINFURT,D.W. (1972): Linear Model Analysis of Categorical Data with Incomplete Response Vectors, in: Biometrics, 28, S.663-692

KOOPMAN,R.F. (1976): Fast Regression Estimates of Missing Data, in: Psychometrika, 41, 2, S.277

KOSOBUD,R. (1963): A Note on a Problem Caused by Assignment of Missing Data in Sample Surveys, in: Econometrica, 31, 3, S.562- 563

KRAUTH,J., LIENERT,G.A. (1973): Die Konfigurationsfrequenzanalyse (KFA), Freiburg

KREUTZ,H. (1972): Soziologie der empirischen Sozialforschung, Stuttgart

KRIZ,J. (1981): Methodenkritik empirischer Sozialforschung, Stuttgart

KÜCHLER,M. (1982): CATI - Die Umfrage-Methodologie der Zukunft?, in: ZUMA-Nachrichten, 11, S.3-8

LAGA,G., LAUFFER,B. (1980): Meinungslosigkeit im Interview, in: Angewandte Sozialforschung, 8, 3, S.227-236

LAIRD,N. (1985): Missing Information Principle, in: KOTZ,S./ JOHNSON,N.L. (Hrsg.): Encyclopedia of Statistical Sciences, Vol.5, S.548-552, New York

LANGENHEDER,W. (1975): Theorie menschlicher Entscheidungshandlungen, Stuttgart

LANSING,J.B., EAPEN,A.T. (1959): Dealing with Missing Information in Surveys, in: Journal of Marketing, S.21-27

LANSING,J.B., MORGAN,J.N. (1971): Economic Survey Methods, Ann Arbor

LAWLESS,J.F. (1982): Statistical Models and Methods for Lifetime Data, New York

LAWLEY,D.N. (1943): A Note on Karl Pearson's Selection Formulae, in: Proceedings of the Royal Society of Edinburgh, Section A, LXII, S.28-30

LEUTHOLD,J.H. (1981): Taxation and the Value of Nonmarket Time, in: Social Science Research, 10, S.267-281

LEVERKUS-BRÜNING,I. (1966): Die Meinungslosen. Die Bedeutung der Restkategorie in der empirischen Sozialforschung, Berlin

LIEBERMAN,B. (Hrsg.) (1971): Contemporary Problems in Statistics, New York

LIEPINS,G. (1983): Can Automatic Data Editing Be Justified? One Person's Opinion, in: WRIGHT,T. (Hrsg.): Statistical Methods and the Improvement of Data Quality, S.205-213, Orlando

LIEPINS,G.E., PACK,D.J. (1980): An Integrated Approach to Data Editing, in: ASA Proceedings of the Section on Survey Research Methods, S.777-781

LIGNY,C.L.de, NIEUWDORP,G.H.E., BREDERODE,W.K., HAMMERS,W.E., HOUWELINGEN,J.C.van (1981): An Application of Factor Analysis with Missing Data, in: Technometrics, 23, 1, S.91-95

LINN,R.L. (1968): Range Restriction Problems in the Use of Self- Selected Groups for Test Validation, in: Psychological Bulletin, 69, 1, S.69-73

LITTLE,R.J.A. (1979): Maximum Likelihood Inference for Multiple Regression with Missing Values: A Simulation Study, in: Journal of the Royal Statistical Society, Series B, 41, 1, S.76-87

LITTLE,R.J.A. (1981): Discussion, in: ASA Proceedings of the Statistical Computing Section, S.61-62

LITTLE,R.J.A. (1982): Models for Nonresponse in Sample Surveys, in: JASA, 77, 378, S.237-250

LITTLE,R.J.A. (1983): The Ignorable Case, in: MADOW,W.G./ OLKIN,I./RUBIN,D.B. (Hrsg.): Incomplete Data in Sample Surveys, Vol.2, S.341-382, New York

LITTLE,R.J.A. (1983b): The Nonignorable Case, in: MADOW,W.G./ OLKIN,I./RUBIN,D.B. (Hrsg.): Incomplete Data in Sample Surveys, Vol.2, S.383-413, New York

LITTLE,R.J.A., RUBIN,D.B. (1983): Incomplete Data, in: KOTZ,S./ JOHNSON,N.L. (Hrsg.): Encyclopedia of Statistical Sciences, Vol.4, S.46-53, New York

LITTLE,R.J.A., RUBIN,D.B. (1983b): Missing Data in Large Data Sets, in: WRIGHT,T. (Hrsg.): Statistical Methods and the Improvement of Data Quality, S.215-243, Orlando

LITTLE,R.J.A., RUBIN,D.B. (1983c): On Jointly Estimating Parameters and Missing Data by Maximizing the Complete-Data Likelihood, in: The American Statistician, 37, 3, S.218-220

LITTLE,R.J.A., SAMUHEL,M.E. (1983): Alternative Models for CPS Income Imputation, in: ASA Proceedings of the Section on Survey Research Methods, S.415-420

LITTLE,R.J.A., SMITH,P.J. (1983): Multivariate Edit and Imputation for Economic Data, in: ASA Proceedings of the Section on Survey Research Method, S.518-522

LÖSEL,F., WÜSTENDÖRFER,W. (1974): Zum Problem unvollständiger Datenmatrizen in der empirischen Sozialforschung, in: KZfSS, 26, S.342-357

LORD,F.M. (1955): Estimation of Parameters from Incomplete Data, in: JASA, 50, S.870-876

LORD,F.M. (1974): Estimation of Latent Ability and Item Parameters When There Are Omitted Responses, in: Psychometrika, 30, 2, S.247-264

LORD,F.M. (1980): Applications of Item Response Theory to Practical Testing Problems, Hillsdale,N.J.

LORD,F.M. (1983): Maximum Likelihood Estimation of Item Response Parameters When Some Responses Are Omitted, in: Psychometrika, 48, 3, S.477-482

LORD,F.M., NOVICK,M.R. (1968): Statistical Theories of Mental Test Scores, Reading,Mass.

LYNCH,C.J. (1972): A Method for Computing Regression Coefficients Utilizing Incomplete Observations, Ph.D.Thesis, The American University

MACRAE,D. (1970): Issues and Parties in Legislative Voting, Methods of statistical Analysis, New York

MANSKI,C.F. (1981): Structural Models for Discrete Data: The Analysis of Discrete Choice, in: S.LEINHARDT (Hrsg.): Sociological Methodology 1981, S.58-109, San Francisco

MARDIA,K.V., KENT,J.T., BIBBY,J.M. (1979): Multivariate Analysis, London

MARINI,M.M., OLSEN,A.R., RUBIN,D.B. (1980): Maximum-Likelihood Estimation in Panel Studies with Missing Data, in: SCHUESSLER,K.F. (Hrsg.): Sociological Methodology 1980, S.315- 357, San Francisco

MATTHAI,A. (1951): Estimation of Parameters from Incomplete Data with Application to Design of Sample Surveys, in: Sankhya, S.145- 152

MAURANA,C.A. (1976): An Analysis of Missing Data in Household Health Surveys, Ph.D. Thesis, Purdue University

MAYS,R. (1978): Correlation Coefficients Corrected for Missing Data, in: Behavior Research Methods and Instrumentation, 10, 3, S.420

MEHTA,J.S., SWAMY,P.A.V.B. (1974): Bayesian Analysis of a Bivariate Normal Distribution When Some Observations are Missing, in: FIENBERG,S.E./ZELLNER,A. (Hrsg.): Studies in Bayesian Econometrics and Statistics, S.289-309, Amsterdam

MELINO,A. (1982): Testing for Sample Selection Bias, in: Review of Economic Studies, 49, S.151-153

MESSMER,D.J., SEYMOUR,D.T. (1982): The Effects of Branching on Item Nonresponse, in: Public Opinion Quarterly, 46, S.270-277

MEULEMANN,H., REUBAND,K.-H. (Hrsg.) (1984): Soziale Realität im Interview, Frankfurt/Main

MEYERS,E.D. (1973): Time-sharing Computation in the Social Sciences, Englewood Cliffs

MINTON,G. (1969): Inspection and Correction Error in Data Processing, in: JASA,64, S.1256-1275

MÖNTMANN,V., BOLLINGER,G., HERRMANN,A. (1983): Tests auf Zufälligkeit von "Missing Data", in: WILKE,H.(Hrsg.): Statistik- Software in der Sozialforschung, S.87-101, Berlin

MOLENAAR,N.J. (1982): Response-effects of "Formal" Characteristics of Questions, in: DIJKSTRA.W./VAN DER ZOUWEN,J. (Hrsg.): Response Behaviour in the Survey-interview, S.49-89, London

MORRISON,D.F. (1976): Multivariate Statistical Methods, 2.Aufl., New York

MORRISSEY,E.R. (1974): Sources of Error in the Coding of Questionnaire Data, in: Sociological Methods and Research, 3, 2, S.209-231

MUSTAFI,C.K. (1981): Analysis of Opinion Polls in Presence of Non-Response, in: The Indian Political Science Review, XV, 2, S.145-150

McCOOL,J.I. (1982): Censored Data, in: KOTZ,S./JOHNSON,N.L. (Hrsg.): Encyclopedia of Statistical Sciences, Vol.1, S.389-396, New York

McFADDEN,D. (1974): Conditional Logit Analysis of Qualitative Choice Behavior, in: P.ZAREMBKA (Hrsg.): Frontiers in Econometrics, S.105-142, New York

NAUS,J.I. (1975): Data Quality Control and Editing, New York

NAUS,J.I., JOHNSON,T.G., MONTALVO,R. (1972): A Probabilistic Model for Identifying Errors in Data Editing, in: JASA, 67, 340, S.943-950

NELSON,F.D. (1977): Censored Regression Models with Unobserved, Stochastic Censoring Thresholds, in: Journal of Econometrics, 6, 1977, S.309-327

NORDBOTTEN,S. (1963): Automatic Editing of Individual Statistical Observations, Conference of European Statisticians, Statistical Standards and Studies No.2, United Nations, New York

NORDHEIM,E.V. (1978): Inference from Non-Randomly Missing Data, Ph.D.Thesis, University of Minnesota

NORDHEIM,E.V. (1978): Obtaining Information from Nonrandomly Missing Data, in: ASA Proceedings of the Statistical Computing Section, S.34-38

NUMERICAL ALGORITHM GROUP LIMITED (1983): NAGFLIB Manual, Mk-10, Oxford

O'BRIEN,R.G., KAISER,M.K. (1985): MANOVA Methods for Analyzing Repeated Measure Designs: An Extensive Primer, in: Psychological Bulletin, 97, 2, S.316-333

O'BRIEN,R.M. (1979): The Use of Pearson's r with Ordinal Data, in: American Sociological Review, 44, S.851-857

O'GRADY,K.E. (1982): Regression Estimation of Missing Data, in: Behavior Research Methods and Instrumentation, 14, 3, S.359-360

OH,H.L., SCHEUREN,F.J. (1983): Weighting Adjustment for Unit Nonresponse, in: MADOW,W.G./OLKIN,I./RUBIN,D.B.(Hrsg.): Incomplete Data in Sample Surveys, Vol.2, S.143-184, New York

OLSEN,R.J. (1980): Approximating a Truncated Normal Regression with the Method of Moments, in: Econometrica, 48, 5, S.1099-1105

OLSON,C.A., BECKER,B.E. (1983): A Proposed Technique for the Treatment of Restriction of Range in Selection Validation, in: Psychological Bulletin, 93, 1, S.137-148

OMURA,G.S. (1983): Correlates of Item Nonresponse, in: Journal of the Market Research Society, 25, 4, S.321-330

ONO,M., MILLER,H.P. (1969): Income Nonresponse in the Current Population Survey, in: ASA Proceedings of the Social Statistics Section, S.277-288

OPP,K.D. (1976): Methodologie der Sozialwisenschaften, Reinbek

OPP,K.D. (1979): Individualistische Sozialwissenschaft, Stuttgart

ORCHARD,T., WOODBURY,M.A. (1970): A Missing Information Principle: Theory and Practice, in: Proceedings of the 6th Berkeley Symposium on Mathematical Statistics and Probability, 1, S.697- 715

OVERALL,J.E., KLETT,C.J. (1972): Applied Multivariate Analysis, New York

PAGE,R.L. (1974): Remark on Algorithm 425, in: Communications of the ACM, 17, 6, S.325

PALM,F.C., NIJMAN,T.E. (1982): Missing Observations in the Dynamic Regression model, Manuskript, Economische Faculteit, Vrije Universiteit, Amsterdam

PAWLIK,K. (1971): Dimensionen des Verhaltens, 2.Aufl., Bern

PAYNE,C. (1977): The Preparation and Processing of Survey Data, in: O'MUIRCHEARTAIGH,C.O./PAYNE,C.:The Analysis of Survey Data. Vol. I: Exploring Data Structures, S.41-61, Chichester

PEKALA,R.J., KUMAR,V.K. (1985): A Short Program for Assessing for the Equality of Two Independent Correlation Matrices, in: Educational and Psychological Measurement,45, S.175-177

PINDYCK,R.S., RUBINFELD,D.L. (1981): Econometric Models and Economic Forecasts, New York

PLATEK,R., GRAY,G.B. (1983): Imputation Methodology: Total Survey Error, in: MADOW,W.G./OLKIN,I./RUBIN,D.B. (Hrsg.): Incomplete Data in Sample Surveys, Vol.2, S.249-333, New York

PLATEK,R., SINGH,M.P., TREMBLAY,V. (1978): Adjustment for Nonresponse in Surveys, in: NAMBOODIRI,N.K. (Hrsg.): Survey Sampling and Measurement, S.157-174, New York

POHL,N.F., BRUNO,A.V. (1978): Reducing Item-Specific Nonresponse Bias, in: Journal of Experimental Education, 46, 3, S.57-64

POIRIER,D.J., RUUD,P.A. (1983): Diagnostic Testing in Missing Data Models, in: International Economic Review, 24, 3, S.537-546

POPPER,K.R. (1976): Logik der Forschung, 6.Aufl., Tübingen

PREGIBON,D. (1976): Typical Survey Data: Estimation and Imputation, in: Survey Methodology (Statistics Canada), 2, 1, S.70-103

PREGIBON,D. (1978): Discussion, in: ASA Proceedings of the Section on Survey Research Methods, S.492-493

PRESS,S.J., SCOTT,A. (1974): Missing Variables in Bayesian Regression, in: FIENBERG,S.E./ZELLNER,A. (Hrsg.): Studies in Bayesian Econometrics and Statistics, S.259-272, Amsterdam

PRESS,S.J., SCOTT,A.J. (1976): Missing Variables in Bayesian Regression II, in: JASA, 71, 354, S.366-369

PRESS,S.J., YANG,C.F. (1974): A Bayesian Approach to Second Guessing "Undecided" Respondents, in: JASA, 69, 345, S.58-67

PROCTOR,C.H. (1974): What Proportion of Item Non-response Should Lead to Deleting the Case?, in: ASA Proceedings of the Social Statistics Section, S.404-408

PROCTOR,C.H. (1978): More on Imputing Versus Deleting When Estimating Scale Scores, in: ASA Proceedings of the Section on Survey Research Methods, S.209-211

RAMSAY,J.O., BERGE,J.T., STYAN,G.P.H. (1984): Matrix Correlation, in: Psychometrika, 49, 3, S.403-423

RAPOPORT,R.B. (1979): What They Don't Know Can Hurt You, in: American Journal of Political Science, 23, 4, S.805-815

RATKOWSKY,D.A. (1974): Maximum Likelihood in Small Incomplete Samples from the Bivariate Normal Distribution, in: Applied Statistics, 23, S.180-189

REUBAND,K.H. (1984): Dritte Personen beim Interview - Zuhörer, Adressaten oder Katalysatoren der Kommunikation?, in: MEULEMANN,H./REUBAND,K.H. (Hrsg.): Soziale Realität im Interview, S.117-156, Frankfurt/Main

RIZVI,H. (1983): An Empirical Investigation of Some Item Nonresponse Adjustment Procedures, in: MADOW,W.G./NISSELSON,H./ OLKIN,I. (Hrsg.): Incomplete Data in Sample Surveys, Vol.1, S.299-366, New York

ROCKWELL,R.C. (1975): An Investigation of Imputation and Differential Quality of Data in the 1970 Census, in: JASA, 70, 349, S.39-42

ROE,R.A. (1979): The Correction for Restriction of Range and the Difference between Intended and Actual Selection, in: Educational and Psychological Measurement, 39, S.551-559

ROE,R.A., ELSHOUT,J.J. (1972): Some New Formulas for the Correction for Restriction of Range, in: Nederlands Tijdschrift voor de Psychologie en haare Grensgebieden, 27, S.134-139

ROGERS,T.F. (1976): Interview by Telephone and in Person: Quality of Responses and Field Performances, in: Public Opinion Quarterly, 40, S.51-65

ROISTACHER,R.C. (1976): A General Consistency Check Procedure for Machine-readable Data, in: Sociological Methods and Research, 4, 3, S.301-320

ROSENBERG,M. (1969): The Logic of Survey Analysis, New York

ROSENTHAL,R., ROSNOW,R.L. (1975): The Volunteer Subject, New York

ROZEBOOM,W.W. (1979): Ridge Regression: Bonanza or Beguilement?, in: Psychological Bulletin, 86, 2, S.242-249

RUBIN,D.B. (1972): A Non-Iterative Algorithm for Least Squares Estimation of Missing Values in Any Analysis of Variance Design, in: Applied Statistics, 21, 2, S.136-141

RUBIN,D.B. (1974): Characterizing the Estimation of Parameters in Incomplete Data Problems, in: JASA, 69, S.467-474

RUBIN,D.B. (1974b): Inference and Missing Data, in: Biometrika, 63, S.581-592

RUBIN,D.B. (1976): Comparing Regressions When Some Predictor Values Are Missing, in: Technometrics, 18, 2, S.201-205

RUBIN,D.B. (1977): Formalizing Subjective Notions About the Effect of Nonrespondents in Sample Surveys, in: JASA, 72, 359, S.538-543

RUBIN,D.B. (1978): Multiple Imputations in Sample Surveys - A Phenomenological Bayesian Approach to Nonresponse, in: ASA Proceedings of the Section on Survey Research Methods, S.20-28

RUBIN,D.B. (1979): Illustrating the Use of Multiple Imputations to Handle Nonresponse in Sample Surveys, in: Bulletin of the International Statistical Institute, S.517-531

RUBIN,D.B. (1980): Handling Nonresponse in Sample Surveys by Multiple Imputation, Monographie, US Bureau of the Census

RUBIN,D.B. (1981): The Bayesian Bootstrap, in: The Annals of Statistics, 9, 1, S.130-134

RUMMEL,R.J. (1970): Applied Factor Analysis, Evanston

RYDBERG,S. (1963): Bias in Prediction, Stockholm

SANDE,G. (1979): Numerical Edit and Imputation, in: Bulletin of the International Statistical Institute, 42nd Session of the International Statistical Institute, S.455-463

SANDE,I.G. (1982): Imputation in Surveys: Coping with Reality, in: The American Statistician, 36, 3, 1982, S.145-152

SANDS,W.A., ALF,E.F., ABRAHAMS,N.M. (1978): Correction of Validity Coefficients for Direct Restriction in Range Occasioned by Univariate Selection, in: Journal of Applied Psychology, 63, 6, S.747-750

SANTOS,R.L. (1981): Effects of Imputation on Regression Coefficients, in: ASA Proceedings of the Section on Survey Research Methods, S.140-145

SANTOS,R.L. (1981b): Effects of Imputation on Complex Statistics. Survey Research Center, Institute for Social Research, University of Michigan

SAS Institute Inc. (1982): SAS User's Guide: Basics, 1982 Edition, Cary,N.C.

SAS Institute Inc. (1982): SAS User's Guide: Statistics, 1982 Edition, Cary,N.C.

SCHIEBER,S.J. (1978): A Comparison of Three Alternative Techniques for Allocating Unreported Social Security Income on the Survey of the Low-Income Aged and Disabled, in: ASA Proceedings of the Section on Survey Research Methods, S.212-218

SCHLITTGEN,R., STREITBERG,B.H.J. (1984): Zeitreihenanalyse, München

SCHLOSSER,O. (1976): Einführung in die sozialwissenschaftliche Zusammenhangsanalyse, Reinbek

SCHMEE,J., HAHN,G.J. (1979): A Simple Method for Regression Analysis with Censored Data, in: Technometrics, 21, 4, S.417-432

SCHMIDT,P. (1982): Subjective Status Identification, Class Consciousness, and Political Attitudes, in: RAUB,W. (Hrsg.): Theoretical Models and Empirical Analyses, S. 227-252, Utrecht

SCHUMAN,H., PRESSER,S. (1981): Questions and Answers in Attitude Surveys, New York

SCHWAB,B., SITTER,R. (1974): Economic Aspects of Computer Input- output Equipment, in: HOUSE,W.C. (Hrsg.): Data Base Management, New York

SCHWERTMAN,N.C., ALLEN,D.M. (1979): Smoothing an Indefinite Variance-Covariance Matrix, in: Journal of Statistical Computation and Simulation, 9, S.183-194

SEBER,G.A.F. (1984): Multivariate Observations, New York

SEIGLE,D. (1967): Some Aids in the Handling of Missing Data, in: Social Science Information, 6, 5, S.133-150

SHARP,L.M., FRANKEL,J. (1983): Respondent Burden: A Test of Some Common Assumptions, in: Public Opinion Quarterly, 47, S.36-53

SHIH,W.J. (1981): Case Analysis of Multiple Linear Regression with Incomplete Data, Ph.D. Thesis, University of Minnesota

SIEMIATYCKI,J. (1979): A Comparison of Mail, Telephone, and Home Interview Strategies for Household Health Surveys, in: American Journal of Public Health, 69, 3, S.238-245

SINGER,E. (1978): Informed Consent: Consequences for Response Rate and Response Quality in Social Surveys, in: American Sociological Review, 43, S.144-162

SINGER,E., KOHNKE-AGUIRRE,L. (1979): Interviewer Expectation Effects: A Replication and Extension, in: Public Opinion Quarterly, 43, S.245-260

SINGER,E., FRANKEL,M.R., GLASSMAN,M.B. (1983): The Effects of Interviewer Characteristics and Expectations on Response, in: Public Opinion Quarterly, 47, S.68-83

SKINNER,C.J. (1984): On the Geometric Approach to Multivariate Selection, in: Psychometrika, 49, 3, S.383-390

SLUIS,D.M. (1981): Statistical Programs and Missing Data, in: Kwantitatieve Methoden, 2, 2, S.56-65

SMITH,T.W. (1982): Educated Don't Knows: An Analysis of the Relationship between Education and Item Nonresponse, in: Political Methodology, 8, 3, S.47-57

SMOUSE,E.P. (1982): Bayesian Estimation of a Finite Population Total Using Auxiliary Information in the Presence of Nonresponse, in: JASA, 77, 377, S.97-102

SÖRBOM,D. (1979): A General Method for Studying Differences in Factor Means and Factor Structure between Groups, in: JÖRESKOG,K.G./SÖRBOM,D.: Advances in Factor Analysis and Structural Equation Models, S.207-217, Cambridge,Mass.

SONDERFORSCHUNGSBEREICH 3 (1985): Das Sozio-ökonomische Panel. Bericht über die Forschungstätigkeit 1983-1985, Johann Wolfgang Goethe-Universität Frankfurt

SONQUIST,J.A., DUNKELBERG,W.C. (1977): Survey and Opinion Research: Procedures for Processing and Analysis, Englewood Cliffs

SONQUIST,J.A., BAKER,E.L., MORGAN,J.N. (1971): Searching for Structure, Institute for Social Research, University of Michigan

SPECHT,D.A., WARREN,R.D. (1975): Comparing Causal Models, in: HEISE,D. (Hrsg.): Sociological Methodology 1976, S.46-82, San Franzisko

SPIERS,E.F., KNOTT,J.J. (1969): Computer Method to Process Missing Income and Work Experience Information in the Current Population Survey, in: ASA Proceedings of the Social Statistics Section, S.289-297

SPSS Inc. (1983): SPSS-X User's Guide, New York

SPSS Inc. (1985): SPSS-X Release 2.1: New Procedures and Facilities, Chicago

SRIVASTAVA,J.N., ZAATAR,M.K. (1973): A Monte Carlo Comparison of Four Estimators of the Dispersion Matrix of a Bivariate Normal Population, Using Incomplete Data, in: JASA, 68, 341, S.180-183

STANISH,W.M., GILLINGS,D.B., KOCH,G.G. (1977): Ratio Estimates for Multivariate Data When Some Observations are Missing, in: ASA Proceedings of the Statistical Computing Section, S.358-363

STEEH,C.G. (1981): Trends in Nonresponse Rates 1952-1979, in: Public Opinion Quarterly, 45, S.40-57

STEIGER,J.H. (1980): Tests for Comparing Elements of a Correlation Matrix, in: Psychological Bulletin, 87, S.245-251

STEINHAUSEN,D., LANGER,K. (1977): Clusteranalyse, Berlin

STEWART,D.W. (1982): Filling the Gap: A Review of the Missing Data Problem, in: WALKER,B.J. (Hrsg.): Assessment of Marketing Thought and Practice. Proceedings of the American Marketing Association, 48, S.395-399, Chicago

STUMPF,S.A. (1978): Dealing with Missing Data: Comparing Methods and Results,in: Catalog of Selected Documents in Psychology,Bd.8, S.2-3

SUDMAN,S. (1983): Response Effects to Behavior and Attitude Questions, in: WRIGHT,T. (Hrsg.): Statistical Methods and the Improvement of Data Quality, S.85-115, Orlando

SUDMAN,S., BRADBURN,N.M. (1974): Response Effects in Surveys, Chicago

SUDMAN,S., BRADBURN,N.M., BLAIR,E., STOCKING,C. (1977): Modest Expectations. The Effects of Interviewers' Prior Expectations on Responses, in: Sociological Methods and Research, 6, 2, S.171-182

SWAN,A.V. (1969): Maximum Likelihood Estimation from Grouped and Censored Normal Data: Algorithm AS 16, in: Applied Statistics, 18, S.110-114

SZAMEITAT,K., ZINDLER,H.J. (1965): The Reduction of Errors in Statistics by Automatic Corrections, in: Bulletin of the International Statistical Institute, Proceedings of the 35th Session, S.395-417

TABACHNICK,B.G., FIDELL,L.S. (1983): Using Multivariate Statistics, New York

TABONY,R.C. (1982): The Estimation of Missing Values in Highly Correlated Data, in: International Association for Statistical Computing: COMPSTAT 1982, S.425-430, Wien

THOMSEN,I. (1973): A Note on the Efficiency of Weighting Subclass
Means to Reduce the Effects of Non-response When Analyzing Survey Data, in: Statistical Review, 4, S.278-283

Thomsen,I. (1975): Evaluating the Efficiency of Two Weighting Procedures to Reduce Non-response Bias: An Application of Cochran's Non-response Stratum Approach, in: Bulletin of the International Statistical Institute, 4, S.396-401

TIMM,N.H. (1969): Estimating Variance-Covariance and Correlation Matrices from Incomplete Data, Ph.D.Thesis, University of California, Berkley,

TIMM,N.H. (1970): The Estimation of Variance-Covariance and Correlation Matrices from Incomplete Data, in: Psychometrika, 35, S.417-437

TOBIN,J. (1958): Estimation of Relationships for Limited Dependent Variables, in: Econometrica, 26, S.24-36

VACEK,P.M., ASHIKAGA,T. (1980): An Examination of the Nearest Neighbor Rule for Imputing Missing Values, in: ASA Proceedings of the Statistical Computing Section, S.326-331

VAN DER ZOUWEN,J., DIJKSTRA,W. (1982): Conclusions, Appendix A to Chapter 8, in: DIJKSTRA,W./VAN DER ZOUWEN,J. (Hrsg.): Response Behaviour in the Survey-interview, S.209-231, London

VAN GUILDER,M., AZEN,S. (1981): Conclusions Regarding Algorithms for Handling Incomplete Data, in: ASA Proceedings of the Statistical Computing Section, S.53-56

VARLEY,T.C. (1969): Data Input Error Detection and Correction Procedures, Ph.D. Thesis, Georg Washington University, Washington,D.C.

WAINER,H. (1978): Discussion, in: ASA Proceedings of the Section on Survey Research Methods, S.31

WALSH,J.E. (1961): A Computer-feasible Method for Handling Incomplete Data in Regression Analysis, in: Journal of the Association for Computing Machinery, 18, S.647-657

WARWICK,D.P., LININGER,C.A. (1975): The Sample Survey: Theory and Practice, New York

WEEDE,E., JAGODZINSKI,W. (1977): Einführung in die konfirmatorische Faktorenanalyse, in: Zeitschrift für Soziologie, 6, 3, S.315- 333

WELCH,W.P., FRANK,R.G., COSTELLO,A.J. (1983): Missing Data in Psychiatric Research: A Solution, in: Psychological Bulletin, 94, 1, S.177-180

WELNIAK,E.J., CODER,J.F. (1980): A Measure of the Bias in the March CPS Earnings Imputation System, in: ASA Proceedings of the Section on Survey Research Methods, S.421-425

WERTS,C.E., ROCK,D.A., GRANDY,J. (1979): Confirmatory Factor Analysis Applications: Missing Data Problems and Comparison of Path Models between Populations, in: Multivariate Behavioral Research, S.199-213

WIBERG,T. (1976): Computation of Principal Components When Data Are Missing, in: International Association for Statistical Computing: COMPSTAT 1976, S.229-236, Wien

WILKE,H. (1983): Datenmanagement mit statistischen Programmpaketen - Möglichkeiten und Grenzen, in: WILKE,H. (Hrsg.): Statistik- Software in der Sozialforschung., Berlin

WILKS,S.S. (1932): Moments and Distributions of Estimates of Population Parameters from Fragmentary Samples, in: Annals of Mathematical Statistics, 3, S.163-195

WINER,B.J. (1971): Statistical Principles in Experimental Design, 2.Aufl., Tokio

WISHART,D. (1978): Treatment of Missing Values in Cluster Analysis, in: International Association for Statistical Computing: COMPSTAT 1978, S.281-287, Wien

WOLYNETZ,M.S. (1979): Maximum Likelihood Estimation from Confined and Censored Normal Data: Algorithm AS 138, in: Applied Statistics, 28, 2, S.185-195

WOLYNETZ,M.S. (1979b): Maximum Likelihood Estimation in a Linear Model from Confined and Censored Normal Data: Algorithm AS 139, in: Applied Statistics, 28, 2, S.195-206

WOODBURY,M.A., HASSELBLAD,V. (1970): Maximum Likelihood Estimates of the Variance-Covariance Matrix from the Multivariate Normal, Annual Meeting of IBM Share, Denver

WOODBURY,M.A., SILER,W. (1966): Factor Analysis with Missing Data, in: Annals of the New York Academy of Science, 128, S.746- 754

YOLKINA,V.N., ZAGORUIKO,N.G. (1978): Some Classification Algorithms Developed at Novosibirsk, in: Revue Francaise d'automatique, d'informatique et de recherche operationelle - Informatique/Computer Science, 12, 1, S.37-46

ZAGORUIKO,N.G., YOLKINA,V.N. (1982): Inference and Data Tables with Missing Values, in: KRISHNAIAH,P.R./KANAL,L.N. (Hrsg.): Handbook of Statistics, Vol.2, S.493-500, New York

ZARKOVICH,S.S. (1966): Quality of Statistical Data, (Food and Agriculture Organization of the United Nations), Rom

ZEH,J. (1976): Der Verzerrungsfehler durch Ausfälle bei Meinungsbefragungen, Dissertation, Bonn

ZEISEL,H. (1970): Die Sprache der Zahlen, Köln

ZINDLER,H.J. (1967): Die Auswirkungen maschineller Korrekturverfahren, in: Metrika, S.206-229