

Längsschnittliche Analysen zur Entwicklung der Zufriedenheit im höheren Lebensalter

Inauguraldissertation

zur Erlangung des Grades eines Dr. phil. an der Fakultät für Verhaltens- und
Empirische Kulturwissenschaften der Ruprecht-Karls-Universität Heidelberg

von

Oliver Schilling

Gutachter:

Prof. Dr. Hans-Werner Wahl (Heidelberg)

Prof. Dr. Joachim Werner (Heidelberg)

Heidelberg, im September 2003

Danksagung:

Mein besonderer Dank gilt Prof. Dr. Hans-Werner Wahl. Als Betreuer dieser Dissertation und mein Dienstvorgesetzter während der Zeit, in der die nachfolgend dargestellten Untersuchungen durchgeführt wurden, hat er durch vielfältige Unterstützung und insbesondere durch gelegentliches moderates Drängen wesentlich dazu beigetragen, daß ihre Fertigstellung nicht nur in einem Prozeß unendlicher Überarbeitung bloß approximativ, sondern tatsächlich gelang.

Außerdem danke ich Prof. Dr. Joachim Werner für seine Bereitschaft zur Zweitbegutachtung dieser Arbeit, die doch immerhin etwas umfänglich geraten ist.

Und schließlich möchte ich noch den Kollegen am Deutschen Zentrum für Altersforschung, Heidelberg, danken – nämlich für das angenehme und die eigene Hirntätigkeit immer wieder anregende Arbeitsklima, welches dort herrscht und sozusagen den idealen Nährboden zur Abfassung derartiger Arbeiten bietet.

Die in dieser Dissertation verwendeten Daten des Sozio-oekonomischen Panels (SOEP) wurden vom Deutschen Institut für Wirtschaftsforschung (DIW), Berlin, bereitgestellt.

Hinweis zur Rechtschreibung:

Es bestand die ursprüngliche Absicht, bei der Niederschrift dieser Arbeit die Regeln der alten Rechtschreibung vor der Rechtschreibreform von 1998 zu befolgen, welche sich mir sozusagen über den größten Teil meiner Bildungskarriere hinweg eingeschliffen hatten. Leider scheint es aber so, daß infolge der besagten Reform meine Beherrschung dieser alten Regeln mehr und mehr verwässert wurde und sich zusehends diesbezügliche Unsicherheiten ausbilden, so daß sich im folgenden Text bisweilen alte und neue Rechtschreibung vermischt.

Ebenso ist zu befürchten, daß selbst intensivstes und aufmerksames Korrekturlesen nicht alle Tipp- und Flüchtigkeitsfehler aufspüren konnte.

Das Ziel, einen in orthographischer Hinsicht ganz und gar fehlerfreien Text zu produzieren, scheint ein kaum erreichbares, und in der Hoffnung, daß trotz alledem die Lesbarkeit des Textes gewährleistet ist, bitte ich um Nachsicht bezüglich aller möglicherweise auffindbaren Rechtschreibfehler.

Inhaltsverzeichnis

1	Theoretische Grundlagen und Ziele der Untersuchung	1
1.1	Zur Theorie des subjektiven Wohlbefindens	2
1.1.1	Grundbegriffe der Wohlbefindensforschung: Wohlbefinden, Glück, Zufriedenheit	3
1.1.2	Konzeptuelle Differenzierung des subjektiven Wohlbefindens: Af- fektive und kognitive Komponenten	5
1.1.3	Zum Konstrukt des habituellen affektiven Wohlbefindens	9
1.1.3.1	Die Affektfrequenztheorie von Diener, Sandvik & Pavot	11
1.1.3.2	Theorien unabhängiger Affektdimensionen	12
1.1.4	Zufriedenheit als Wohlbefindenskomponente	15
1.1.4.1	Zufriedenheit und objektive Lebensqualität	15
1.1.4.2	Messung von Zufriedenheit	18
1.1.4.3	Relativität von Zufriedenheit	19
1.1.4.4	Persönlichkeitseinflüsse auf das SWB	20
1.1.4.5	Affektive Inferenz	24
1.1.4.6	Das „judgement model of subjective well-being“ von Schwarz & Strack	25
1.1.4.7	Die Top-down/Bottom-up-Kontroverse	30
1.1.4.8	Fazit: Theoretische Standpunkte zum Zufriedenheitsbe- griff	35
1.2	Zufriedenheit im höheren Lebensalter: Stabilität trotz Verlust?	36
1.2.1	Das Zufriedenheitsparadox des höheren Lebensalters	40
1.2.2	Längsschnittliche Untersuchungen zum Zusammenhang von höher- em Lebensalter und Zufriedenheit	43
1.2.3	Zur Erklärung des Zufriedenheitsparadoxes: Mögliche Entwick- lungsdynamiken der Zufriedenheit im höheren Lebensalter	50
1.2.3.1	Überlagerung von Alters- und Kohorteneffekt?	50
1.2.3.2	Zufriedenheitsgewinner und Zufriedenheitsverlierer?	55
1.2.3.3	Lebenslange Stabilität von Zufriedenheit?	58
1.2.3.4	Erhöhte Stabilität der Zufriedenheit im Alter?	61
1.3	Hypothesen und Ziele der empirischen Untersuchung	66
2	Methodische Grundlagen: Statistische Modelle und Prozeduren	72
2.1	Analyse von Kohorteneffekten	72

2.1.1	Zeitvariable Effekte: Alter, Kohorte und Periode	73
2.1.2	Statistische Modellierung der Trennung von Alters- und Kohorteneffekten	78
2.2	Statistische Methoden der Analyse von Merkmalsstabilität	85
2.2.1	Typen und Kennwerte der Merkmalsstabilität	86
2.2.2	Längsschnittliche Strukturgleichungsmodelle	91
2.2.2.1	Das Quasi-Markov-Simplex-Modell	95
2.2.2.2	Das Modell latenter Differenzkomponenten	107
2.2.2.3	Das Modell latenter Differenzbeträge	113
2.2.2.4	Weitere längsschnittliche Veränderungsmodelle	118
2.3	Spezielle Probleme der Parameterschätzung	125
2.3.1	Behandlung von fehlenden Werten in längsschnittlichen Strukturgleichungsmodellen	125
2.3.1.1	Konventionelle Methoden zur Behandlung fehlender Werte	126
2.3.1.2	Stochastische Mechanismen des Datenausfalls	129
2.3.1.3	Neuere Methoden zur Behandlung fehlender Werte	131
2.3.1.4	Zur Anwendbarkeit des FIML-Schätzalgorithmus bei nicht-normalen Daten	138
2.3.2	Bewertung und Vergleich der Anpassungsgüte der Modelle	142
2.4	Fazit: Hinweise zur Auswahl statistischer Modelle und Prozeduren bei längsschnittlichen Datenanalysen	148
3	Empirische Untersuchungen zur Zufriedenheit im höheren Alter	151
3.1	Stichprobe und Untersuchungsvariablen	151
3.1.1	Das Sozio-oekonomische Panel (SOEP): Paneldesign, Substichproben, Weiterverfolgungskonzept	151
3.1.2	Zufriedenheitsbefragungen im SOEP	155
3.1.3	Untersuchungsstichprobe der Analysen zur Zufriedenheit im höheren Lebensalter	156
3.1.4	Fazit: Zur Eignung des SOEP für längsschnittliche Analysen zur Zufriedenheit im höheren Lebensalter	159
3.2	Analysen zur Entwicklung durchschnittlicher Zufriedenheitswerte	161
3.2.1	Deskriptive Analysen zum Zusammenhang von Alter und Zufriedenheit im höheren Erwachsenenalter	161
3.2.1.1	Querschnittliche Zusammenhänge von Alter und Zufriedenheit	163

3.2.1.2	Längsschnittliche Zusammenhänge von Alter und Zufriedenheit	168
3.2.1.3	Zusammenfassung und Diskussion der deskriptiven Analysen zum Zusammenhang von Alter und Zufriedenheit	181
3.2.2	Analysen zur Kohorteneffekthypothese	184
3.2.2.1	Ergebnisse der Analysen zur Kohorteneffekthypothese	185
3.2.2.2	Zusammenfassung und Diskussion der Analysen zur Kohorteneffekthypothese	189
3.3	Analysen zur Zufriedenheitsstabilität im höheren Lebensalter	194
3.3.1	Stabilität von Zufriedenheit in der Gesamtgruppe und bei älteren Befragungspersonen	199
3.3.1.1	Deskription der Analysetichproben für die allgemeinen Stabilitätsanalysen	199
3.3.1.2	Analysestrategie der allgemeinen Stabilitätsanalysen	206
3.3.1.3	Ergebnisse der allgemeinen Stabilitätsanalysen	210
3.3.1.4	Zusammenfassung und Diskussion der allgemeinen Stabilitätsanalysen	222
3.3.2	Entwicklung der Stabilität von Zufriedenheit in ausgewählten Geburtsjahrganggruppen älterer Befragungspersonen	229
3.3.2.1	Stichproben, Design und Strategie der altersspezifischen Stabilitätsanalysen	229
3.3.2.2	Ergebnisse der altersspezifischen Stabilitätsanalysen	240
3.3.2.3	Zusammenfassung und Diskussion der altersspezifischen Stabilitätsanalysen	246
4	Zusammenfassung und abschließende Bewertung	250
4.1	Zusammenfassung der Ergebnisse	253
4.1.1	Deskriptiven Analysen von Zufriedenheitsmittelwerten	253
4.1.2	Analysen zur Kohorteneffekthypothese	254
4.1.3	Analyse der intraindividuellen Merkmalsstabilität	257
4.2	Statistische Methoden der längsschnittlichen Untersuchungen	262
4.3	Abschließende Bewertung und Ausblick	265
Anhang A	Ergänzungen zum Kapitel 2	270
A.1	Zusammenhang zwischen Meßwertkorrelationen und Merkmalsstabilität	270
A.2	Identifikation des QMSM	273
A.3	Identifikation des LDCM	277

Anhang B Weitere Ergebnistabellen	279
Literatur	291

Abbildungsverzeichnis

1.1	Schematische Darstellung der Kohorteneffekthypothese	53
1.2	Schematische Darstellung der Gewinner-und-Verlierer-Hypothese	57
2.1	Schematische Darstellung des Modells linearer Alters-, Kohorten- und Perio- deneffekte	81
2.2	Typen der Merkmalsstabilität	86
2.3	Autoregressive Strukturgleichungsmodelle erster Ordnung	95
2.4	QMSM mit Phantomvariablen zur Restriktion der True-Score-Varianzen . . .	102
2.5	Latent-Difference-Components-Model (LDCM)	107
2.6	Latent-Difference-Scores-Model (LDSM)	113
2.7	LDSM mit Phantomvariablen zur Restriktion der True-Score-Varianzen . . .	115
2.8	Latent Growth Curve Model (LGCM)	119
3.1	Allgemeine Lebenszufriedenheit: Querschnittliche Mittelwertsverläufe über das Lebensalter	163
3.2	Einkommenszufriedenheit: Querschnittliche Mittelwertsverläufe über das Le- bensalter	165
3.3	Gesundheitszufriedenheit: Querschnittliche Mittelwertsverläufe über das Le- bensalter	165
3.4	Wohnzufriedenheit: Querschnittliche Mittelwertsverläufe über das Lebensalter	165
3.5	Freizeitzufriedenheit: Querschnittliche Mittelwertsverläufe über das Lebensal- ter	165
3.6	Allgemeine Lebenszufriedenheit: Längsschnittliche Mittelwertsverläufe der Ein-Jahres-Geburtsjahrganggruppen	169
3.7	Allgemeine Lebenszufriedenheit: Längsschnittliche Mittelwertsverläufe der Fünf-Jahres-Geburtskohorten	169
3.8	Bereichszufriedenheiten: Längsschnittliche Mittelwertsverläufe der Ein-Jahres- Geburtsjahrganggruppen	172
3.9	Einkommenszufriedenheit: Längsschnittliche Mittelwertsverläufe der Fünf- Jahres-Geburtskohorten	172
3.10	Gesundheitszufriedenheit: Längsschnittliche Mittelwertsverläufe der Fünf- Jahres-Geburtskohorten	172
3.11	Wohnzufriedenheit: Längsschnittliche Mittelwertsverläufe der Fünf-Jahres- Geburtskohorten	172
3.12	Freizeitzufriedenheit: Längsschnittliche Mittelwertsverläufe der Fünf-Jahres- Geburtskohorten	172

3.13	Allgemeine Lebenszufriedenheit: Mittlere Differenzen zwischen sukzessiven Lebensjahren	175
3.14	Einkommenszufriedenheit: Mittlere Differenzen zwischen sukzessiven Lebensjahren	178
3.15	Gesundheitszufriedenheit: Mittlere Differenzen zwischen sukzessiven Lebensjahren	178
3.16	Wohnzufriedenheit: Mittlere Differenzen zwischen sukzessiven Lebensjahren	178
3.17	Freizeitzufriedenheit: Mittlere Differenzen zwischen sukzessiven Lebensjahren	178
3.18	Analysestrategie der allgemeinen Stabilitätsanalysen	208
3.19	Analysestrategie der altersspezifischen Stabilitätsanalysen	238
3.20	Altersspezifische Stabilitätsanalysen (M1), allgemeine Lebenszufriedenheit: Geschätzte True-Score-Varianzen und Quotienten Q_{Δ}	245

Tabellenverzeichnis

1.1	Komponenten des SWB nach Diener et al. (1999, 277)	6
2.1	Schema: Geburtsjahrgänge, Lebensalter und Jahr der SOEP-Datenerhebung .	75
3.1	Zufriedenheitsbefragungen im SOEP (bis 1999)	154
3.2	Entwicklung der westdeutschen SOEP-Stichprobe auf Personenebene (bis 1999)	158
3.3	Todesfälle in der westdeutschen SOEP-Stichprobe (bis 1999)	159
3.4	Ergebnisse der Multilevel-Regressionsanalyse: Schätzung von Alters- und Kohorteneffekten	188
3.5	Multilevel-Regressionsanalyse: Alters- und Kohorteneffekte der ALZ bei den 1919-1934 und den 1914-1929-Geborenen	190
3.6	Stichprobenumfänge bei den allgemeinen Stabilitätsanalysen	200
3.7	Allgemeine Stabilitätsanalysen: Stichprobenmittelwerte der Meßwerte	201
3.8	Allgemeine Stabilitätsanalysen: Standardabweichungen der Meßwerte	203
3.9	Allgemeine Stabilitätsanalysen: Autokorrelationen erster Ordnung der Meßwerte	204
3.10	Allgemeine Stabilitätsanalysen: Durchschnittliche Autokorrelationen n -ter Ordnung (lag n) der Meßwerte	205
3.11	Allgemeine Stabilitätsanalysen: Anpassungsgüte der QMSM für Lebenszufriedenheit	211
3.12	Allgemeine Stabilitätsanalysen: Anpassungsgüte der QMSM für Einkommenszufriedenheit	212
3.13	Allgemeine Stabilitätsanalysen: Anpassungsgüte der QMSM für Gesundheitszufriedenheit	213
3.14	Allgemeine Stabilitätsanalysen: Anpassungsgüte der QMSM für Wohnzufriedenheit	214
3.15	Allgemeine Stabilitätsanalysen: Anpassungsgüte der QMSM für Freizeitzufriedenheit	215
3.16	Allgemeine Stabilitätsanalysen: Geschätzte True-Score-Varianzen	216
3.17	Allgemeine Stabilitätsanalysen: Geschätzte Stabilitätskoeffizienten und Quotienten Q_{Δ}	218
3.18	Allgemeine Stabilitätsanalysen (M1): Geschätzte Mittelwerte [†]	220
3.19	Allgemeine Stabilitätsanalysen (M1): Geschätzte Reliabilitäten [†]	222
3.20	Design der altersspezifischen Stabilitätsanalysen	234
3.21	Altersspezifische Stabilitätsanalysen: Anpassungsgüte der QMSM für allgemeine Lebenszufriedenheit	240

3.22	Altersspezifische Stabilitätsanalysen: Anpassungsgüte der QMSM für Einkommenszufriedenheit	241
3.23	Altersspezifische Stabilitätsanalysen: Anpassungsgüte der QMSM für Gesundheitszufriedenheit	242
3.24	Altersspezifische Stabilitätsanalysen: Anpassungsgüte der QMSM für Wohnzufriedenheit	242
3.25	Altersspezifische Stabilitätsanalysen: Anpassungsgüte der QMSM für Freizeitzufriedenheit	243
3.26	Altersspezifische Stabilitätsanalysen (M2): True-Score-Varianz, Stabilitätskoeffizienten und Q_{Δ}	244
B.1	Ergänzung Kapitel 3.1: Stichprobengrößen des SOEP nach Substichprobe und Befragungsjahr	279
B.2	Ergänzung Kapitel 3.1: Mittelwerte und Standardabweichungen der untersuchten Zufriedenheitsmessungen im SOEP (Westdeutsche)	280
B.3	Ergänzung Kapitel 3.2.1: ALZ – Mittelwerte der Ein-Jahres-Altersgruppen	281
B.4	Ergänzung Kapitel 3.2.1: EIZ – Mittelwerte der Ein-Jahres-Altersgruppen	282
B.5	Ergänzung Kapitel 3.2.1: GEZ – Mittelwerte der Ein-Jahres-Altersgruppen	283
B.6	Ergänzung Kapitel 3.2.1: WOZ – Mittelwerte der Ein-Jahres-Altersgruppen	284
B.7	Ergänzung Kapitel 3.2.1: FRZ – Mittelwerte der Ein-Jahres-Altersgruppen	285
B.8	Ergänzung Kapitel 3.2.1: ALZ – Anzahl der Werte in den Ein-Jahres-Altersgruppen	286
B.9	Ergänzung Kapitel 3.3.1: ALZ, univariate Normalverteilungseigenschaften	287
B.10	Ergänzung Kapitel 3.3.1: EIZ, univariate Normalverteilungseigenschaften	287
B.11	Ergänzung Kapitel 3.3.1: GEZ, univariate Normalverteilungseigenschaften	288
B.12	Ergänzung Kapitel 3.3.1: WOZ, univariate Normalverteilungseigenschaften	288
B.13	Ergänzung Kapitel 3.3.1: FRZ, univariate Normalverteilungseigenschaften	289
B.14	Ergänzung Kapitel 3.3.2: Modell 1, Vier-Jahres-Stabilitätskoeffizienten und True-Score-Varianzen	290

1 Theoretische Grundlagen und Ziele der Untersuchung

Die im folgenden dargestellte Untersuchung behandelt die Lebenszufriedenheit älterer Menschen. Sie ist damit „disziplinär“ in der psychogerontologischen Forschung und theoretisch im weiten Feld der Wohlbefindensforschung zu verorten. Konkret versucht sie, einen empirischen Beitrag zur gerontologischen Diskussion um die Stabilität der Lebenszufriedenheit im höheren Lebensalter zu leisten, welche in der Gerontologie v.a. durch Befunde angestoßen wurde, welche mit dem Begriff des Wohlbefindens- oder Zufriedenheitsparadoxes des höheren Alters bezeichnet wurden und knapp dadurch zu charakterisieren sind, daß ein angesichts vermehrter problematischer Bedingungen in den höheren Lebensaltersphasen für diese zu erwartender allgemeiner Rückgang von Lebenszufriedenheit nicht gefunden wurde. Dieser Befund wurde, wie im folgenden noch dargelegt werden wird, oft gleichgesetzt mit Stabilität von Lebenszufriedenheit im höheren Alter, so daß theoretische Erklärungen zumeist auf Faktoren abzielen, die diese Stabilität erzeugen. Die hier dargestellte Untersuchung geht demgegenüber gewissermaßen einen Schritt zurück, indem sie bei der methodischen Problematik ansetzt, die dem Stabilitätsbegriff dadurch innewohnt, daß er eine nur längsschnittlich erfaßbare Qualität des Merkmals Zufriedenheit benennt und deshalb auch längsschnittliche Untersuchungen erfordert. Es wird darum zunächst in Frage gestellt, ob die besagten Befunde, die überwiegend querschnittlichen Untersuchungen entstammen, überhaupt Stabilität anzeigen. Es werden verschiedene *mögliche* Dynamiken der Entwicklung von Lebenszufriedenheit über das höhere Lebensalter hinweg – d.h., mit anderen Worten, mögliche Charakteristika der Veränderung von Zufriedenheit während dieser Lebensphase – aufgezeigt und theoretisch diskutiert, welche diese Befunde bewirkt haben könnten. Die tatsächliche intraindividuelle Stabilität des Merkmals ist eine von diesen Möglichkeiten. Die eigentliche empirische Untersuchung wird dann sekundäranalytisch in den Zufriedenheitsmessungen eines längsschnittlichen Datensatzes – nämlich des deutschen Sozio-oekonomischen Panels (SOEP) – nach Belegen für diese möglichen Dynamiken suchen um so zu Aussagen darüber zu gelangen, ob und in welchem Ausmaß diese hypothetisch postulierten Charakteristika die tatsächliche alterstypische Veränderung der Lebenszufriedenheit beschreiben. Es geht also hier insbesondere *nicht* darum, das Gewicht möglicher konkreter psychischer, physischer oder sozialer Einflußfaktoren auf die Lebenszufriedenheit älterer Menschen zu erforschen, sondern v.a. um Klärung der Frage, inwieweit und in welcher Weise die Lebenszufriedenheiten im Alter tatsächlich stabil sind.

Zum Verständnis und zur Bewertung der besagten gerontologischen Diskussionsbeiträge über Stabilität von subjektivem Wohlbefinden bzw. Lebenszufriedenheit im höheren Alter und zur theoretischen Begründung der hier als Untersuchungshypothesen zugrundegelegten besagten möglichen Entwicklungsdynamiken der Zufriedenheit im Alter erscheint es zunächst notwendig, eine kurze Zusammenfassung der Theoriebildung und theoretischen

Diskussionslinien der wissenschaftlichen Auseinandersetzung mit diesen Konstrukten voranzustellen. Diese ist im folgenden Kapitel 1.1 enthalten und sie ist an dieser Stelle mit der einleitenden Einschränkung zu versehen, daß eine tatsächlich umfassende und erschöpfende Darstellung des Forschungsfeldes – insbesondere der Vielzahl empirischer Befunde zu Korrelaten von subjektivem Wohlbefinden bzw. Lebenszufriedenheit – in diesem Kapitel nicht angestrebt wurde, da dessen Umfang und Komplexität den durch die hier vorgegebenen Untersuchungsziele gesteckten Rahmen sprengen würde. Vielmehr sollen lediglich die in der gegenwärtigen Wohlbefindensliteratur hauptsächlich diskutierten theoretischen Konzepte über die Entstehung von Zufriedenheitsurteilen dargestellt werden, aus denen dann die eigentlich untersuchungsrelevanten Hypothesen über mögliche Entwicklungsdynamiken im höheren Alter begründbar sind.

Im Anschluß daran wird dann in Kapitel 1.2 die oben bereits in knappen Worten ange deutete gerontologische Diskussion um die Lebenszufriedenheit bzw. das Zufriedenheitsparadox im höheren Alter dargestellt und daraus die untersuchungsrelevanten Hypothesen begründet, welche zum Ende dieses Theorieteils noch einmal im Unterkapitel 1.3 explizit aufgelistet werden.

1.1 Zur Theorie des subjektiven Wohlbefindens

In Gottfried Benns Gedicht „Eure Etüden“ sind zwei Zeilen zu lesen, bei denen es sich um die wohl weithin bekannteste und am häufigsten zitierte Textstelle dieses Dichters handeln dürfte: „... *dumm sein und Arbeit haben: das ist das Glück*“. Daß dieser Satz, der ja einem literarischen Gesamtwerk entstammt, dessen Verbreitung ansonsten auf eine spezifisch interessierte intellektuelle Leserschaft beschränkt blieb, sich gewissermaßen verselbständigt und darüber hinaus verbreitet hat, ist kaum verwunderlich, denn er gibt immerhin eine Antwort auf die vielleicht elementarste Frage menschlicher Existenz, nämlich die nach dem Glück: Was ist bzw. worin erfüllt sich das, was wir als „Glück“ bezeichnen? Die Einprägsamkeit der obigen Worte dürfte wohl darauf beruhen, daß sie eine mögliche nackte Wahrheit zum Ausdruck bringen: Glück, so scheint der Satz in knappen Worten festzustellen, setzt eine gewisse intellektuelle Schlichtheit – eben das „dumm sein“ – voraus, welche nur jene einfachsten Grundbedürfnisse kennt, deren Befriedigung sozusagen den kleinsten gemeinsamen Nenner unter den vielen Zielen, um derentwillen Menschen irgendeine „Arbeit haben“, bildet – und es ist besonders bemerkenswert, was alles demnach Glück *nicht* ist! So ungefähr dürfte jedenfalls der Satz im Allgemeinen und losgelöst vom Kontext des Gedichts verstanden werden und dieses Verständnis enthält zwei Komponenten, die, wie noch gezeigt werden wird, auch die diesbezügliche wissenschaftliche Auseinandersetzung mit dem Thema Glück geprägt haben, wie sie v.a. in Psychologie und Soziologie stattgefunden hat und unvermindert stattfindet: Nämlich zum einen die primär materielle und objektive Bedürfnisbefriedigung und zum anderen die subjektive Zufriedenheit mit der eigenen objektiven Lebensrealität. Forschung zum „subjektiven Wohlbefinden“, zur „Lebenszufrie-

denheit“ bzw. „subjektiven Lebensqualität“, um einige der häufig gebrauchten Theoriebegriffe zu nennen, dreht sich häufig um diese beiden Komponenten. Sie fragt sozusagen nach den konkreten objektiven Lebensqualitäten, die zum subjektiv erlebten Glück beitragen – und die möglicherweise weit über die aus dem „Arbeit haben“ resultierenden Erfüllungen von Grundbedürfnissen hinaus reichen — und danach, wie und woraus sich individuelle Zufriedenheiten überhaupt konstituieren, d.h. nach der Psychologie der Zufriedenheit, die vielleicht eine intrapsychische Komplexität aufweist, der die Pointierung auf das „dumm sein“ nicht gerecht wird.

1.1.1 Grundbegriffe der Wohlbefindensforschung: Wohlbefinden, Glück, Zufriedenheit

Betritt man nun dieses wissenschaftliche Themenfeld, so fällt zunächst eine gewisse terminologische Beliebigkeit auf: „Das Chaos der Definitionen von Variablen subjektiven Wohlbefindens ist groß. So wird subjektives Wohlbefinden als Glück . . . , Glück als subjektives Wohlbefinden . . . , Glück als Lebensqualität . . . , Glück als Freude . . . , positive Stimmung als Glück . . . , Glück als Zufriedenheit . . . konzipiert“ (Mayring, 1991, 51). Und was hier mit Blick auf einschlägige psychologische Untersuchungen konstatiert wurde, gilt für die soziologischen Publikationen nicht minder: Es ist von subjektiver Lebensqualität, von subjektivem Wohlbefinden, von Glück („happiness“) und Zufriedenheit die Rede — und stets ist damit mehr oder weniger dasselbe gemeint. Daß wissenschaftliche Exaktheitsansprüche mit solchen begrifflichen Verschwommenheiten nicht gerade vorbildlich erfüllt sind, muß wohl nicht eigens betont werden — kaum ein Methodenlehrbuch der Sozial- und Verhaltenswissenschaften vergißt die Forderung nach exakten Begriffsdefinitionen als Kriterium für die Wissenschaftlichkeit theoretischer Aussagen. Jedoch resultiert das terminologische Durcheinander im Grunde aus der Vielfalt der Begriffe, die unsere Alltagssprache für diesen Bereich bereithält. Letztere aber ist Abbild der Komplexität dieses Bereichs selbst, verschiedene Begriffe benennen unterschiedliche Aspekte des menschlichen Wohlbefindens, Glücks — oder wie auch immer man es bezeichnen will.

Im folgenden jedenfalls wird der Begriff *subjektives Wohlbefinden* (SWB) zur Benennung des gesamten wissenschaftlichen Gegenstandsbereichs verwendet. Zumindest der Gebrauch dieses Terminus als Oberbegriff für das ganze Forschungsgebiet scheint sich mittlerweile – betrachtet man die einschlägigen Veröffentlichungen jüngerer Datums (z.B. das Review von Diener et al., 1999) – weitestgehend durchgesetzt zu haben. SWB bietet v.a. auch den inhaltlichen Vorteil, weniger werturteilsbehaftet zu sein, als das alltagssprachliche *Glück*, das sozusagen einen konnotativen Horizont besitzt, der durch die gesamte menschliche Kulturgeschichte und die darin stets thematischen Glücksvorstellungen aufgebläht ist.

In wissenschaftlichen Theoretisierungen des SWB spielt meist die *Zufriedenheit* eine gewichtige Rolle. Das den verschiedenen thematischen Arbeiten Gemeinsame in der Verwendung dieses Begriffes kann einstweilen so vorweggenommen werden: Zufrieden-

heit meint Bewertung, Beurteilung, subjektive Wertschätzung eigener objektiver Lebensrealitäten. In diesem Sinne wurden in einer Vielzahl v.a. sozialwissenschaftlicher Erhebungen Zufriedenheiten – mit bestimmten Bereichen oder Aspekten des eigenen Lebens oder auch mit dem Leben ganz allgemein – erfragt, die wissenschaftliche Relevanz der Ergebnisse solcher Befragungen hängt aber nicht zuletzt auch davon ab, *wie* Menschen überhaupt mehr oder weniger zufrieden werden und in welchem Verhältnis Zufriedenheit zu dem steht, was eingangs als Glück bezeichnet wurde. Zunächst also: Auf welche Weise mündet die Wahrnehmung eigener objektiver Lebensrealitäten in persönliche Zufriedenheit? Ist beispielsweise der Mensch ein „objektiver Beobachter“ seiner eigenen Situation und infolgedessen seine Zufriedenheit ein guter Indikator für die objektive Qualität derselben — ist derjenige zufrieden, dem es tatsächlich „gut geht“ bzw. geht es umgekehrt dem, der zufrieden ist, stets auch tatsächlich gut? Oder gibt es sozusagen gewisse intrapsychische Pufferungen, kognitive und/oder emotionale Mechanismen, die dem Zufriedenheitsurteil seine letztendliche Gestalt verleihen? Und weiter: Ist Zufriedenheit gleich Glück bzw. welchen Beitrag leistet Zufriedenheit zum SWB einer Person?

Derartige Fragen könnten nicht nur zur Beschäftigung in wissenschaftlichen Elfenbeintürmen beitragen: In Forschungsfeldern, für die eine auf Evaluation und Optimierung von (sozioökonomischen, gesundheitlichen, psychologischen) Lebensbedingungen ausgerichtete Anwendungsorientierung forschungsleitend ist, sind SWB und, als ein Aspekt desselben, insbesondere Zufriedenheit psychologische Variablen von zentraler Bedeutung, denn sie bilden ein mögliches Zielkriterium dieses Praxisbezugs. Ein solchermaßen auf Praxisrelevanz ausgerichtetes Forschungsinteresse führt über die Erfassung objektiv feststellbarer Kriterien von Lebensqualität hinaus geradewegs zur Subjektivität der Bewertung dieser Kriterien – es geht letztendlich dabei um Ursachen dafür, daß es den Menschen mehr oder weniger „gut geht“, und es entspricht wohl einer allgemeingültigen und jederzeit nachvollziehbaren Alltagserfahrung, daß sich Menschen in ihren diesbezüglichen Einschätzungen oft beträchtlich auch dann unterscheiden, wenn ihre objektiven Lebensumstände sich weitgehend gleichen. In diesem Sinne also bestünde die Anwendungsorientierung letztendlich in dem Ziel, SWB von Menschen zu verbessern, und man wird, um diesen Beitrag leisten zu können, auch stets fragen müssen, ob und weshalb Menschen mit ihren Lebensumständen zufrieden sind. Dies schließt aber die oben gestellten Fragen mit ein, nämlich Fragen danach *wie* – aufgrund welcher Bewertungskriterien und Beurteilungsregeln – überhaupt Zufriedenheitsbewertungen generiert werden.

Die zum Ende dieses Theorieteils im Kapitel 1.3 aufgelisteten Hypothesen beruhen auf Antworten auf diese Fragen bzw. auf theoretischen Annahmen über die Generierung von Zufriedenheitsurteilen. Damit solche aber rational begründbar sind, ist die Bedeutung des damit behandelten Konstrukts Zufriedenheit genau zu spezifizieren und das schließt dessen Verortung im gesamten Theoriefeld des SWB mit ein. Dazu wäre sozusagen zunächst einmal die Landkarte des Wohlbefindens zu erstellen, so daß dann die markanten Punkte darauf auch in ihren wechselseitigen Bezügen erkennbar werden.

1.1.2 Konzeptuelle Differenzierung des subjektiven Wohlbefindens: Affektive und kognitive Komponenten

Eine grundlegende Anforderung an eine psychologische Theorie des SWB besteht darin, die aspektivische und terminologische Vielfalt des Themenfelds logisch zu ordnen, um überhaupt wissenschaftliche Aussagen über eindeutig definierte und inhaltlich klar abgegrenzte Wohlbefindenskonstrukte machen zu können. Welches also sind die wissenschaftlich relevanten „Subkonstrukte“ des SWB?

Zur Beantwortung dieser Frage leistet das Review von Diener et al. (1999), das an ein älteres Review von Wilson (1967) anknüpfend eine sehr umfangliche Zusammenfassung von Ergebnissen und Entwicklungen in der SWB-Forschung seit den siebziger Jahren enthält, besonders wertvolle Dienste – die folgenden Ausführungen beziehen sich, wo nicht anders vermerkt, im Prinzip stets darauf. Dabei zeigt sich, daß der seitherige Fortschritt in der Untersuchung des SWB sich eben nicht nur in der empirischen Prüfung der von Wilson postulierten Prädiktoren des SWB erschöpfte, sondern auch bezüglich Definition und Theorie des Konstruktes selbst stattfand. Ganz allgemein läßt sich über diese Entwicklung sagen: „In the past, many researchers have treated SWB as a monolithic entity, but it is now clear that there are separable components that exhibit unique patterns of relations with different variables“ (Diener et al., 1999, 276).

Auch wenn diese Feststellung wohl bei keinem der involvierten Forscher auf Widerspruch stoßen dürfte und der Terminus SWB also allenfalls ein Dachkonstrukt benennt, welches unterschiedliche Komponenten einschließt, so scheint die diesbezügliche Klarheit nicht bei jedem Blick in die Wohlbefindensliteratur gegeben: Es ist häufig von SWB die Rede, ohne daß eindeutig geklärt wird, welche Komponente nun genau gemeint ist. Auch gibt es eine gewisse Vielfalt in der Benennung von Subkategorien des SWB, wie z.B. die in Tabelle 1.1 abgebildete tabellarische Auflistung von Diener et al. (1999, 277) zeigt (vgl. auch Mayring, 1991).

Der Eindruck einer vorherrschenden Beliebigkeit in der Konzeptualisierung dessen, was alles als Inhalt des SWB gelten kann, der aus solchen Auflistungen vielleicht entsteht, täuscht allerdings zumindest insofern, als den meisten der neueren Arbeiten zum SWB die grundsätzliche Unterscheidung zwischen einer *affektiven* und einer *kognitiven* Komponente des SWB gemeinsam ist. Es besteht ein breiter Konsens darin, daß das Bedeutungsfeld SWB einerseits das Erleben mehr oder weniger positiver Stimmungen bzw. Emotionen, andererseits aber auch die mehr oder weniger positive Selbstbeurteilung der eigenen Lebensumstände einschließt.

Bezüglich der affektiven Komponente werden die Dimensionen positiver und negativer Affektivität („pleasant“ vs. „unpleasant affect“ in der Terminologie von Diener et al. 1999) als zwei unabhängige Konstrukte behandelt: „Although the degree of independence between momentary pleasant and unpleasant affect is still debated, the separability of long-term dimensions is less controversial“ (ebd., 277). Dieses letzte Zitat verweist auch

TABELLE 1.1:
Komponenten des SWB nach Diener et al. (1999, 277)

<i>Components of Subjective Well-Being</i>			
<i>Pleasant affect</i>	<i>Unpleasant affect</i>	<i>Life satisfaction</i>	<i>Domain satisfaction</i>
<i>Joy</i>	<i>Guilt and shame</i>	<i>Desire to change life</i>	<i>Work</i>
<i>Elation</i>	<i>Sadness</i>	<i>Satisfaction with current life</i>	<i>Family</i>
<i>Contentment</i>	<i>Anxiety and Worry</i>	<i>Satisfaction with past</i>	<i>Leisure</i>
<i>Pride</i>	<i>Anger</i>	<i>Satisfaction with future</i>	<i>Health</i>
<i>Affection</i>	<i>Stress</i>	<i>Significant other's views of one's life</i>	<i>Finances</i>
<i>Happiness</i>	<i>Depression</i>		<i>Self</i>
<i>Ecstasy</i>	<i>Envy</i>		<i>One's group</i>

auf eine weitere Differenzierung, die bei einer Betrachtung affektiver SWB-Komponenten berücksichtigt werden muß: Das eigentliche Interesse der SWB-Forschung gilt den habituellen Aspekten der Befindlichkeit von Personen, der Begriff des SWB bezieht sich somit nicht (nur) auf momentane Reaktionen des Individuums auf jeweils aktuelle situative Bedingungen, sondern meint eine über den Augenblick hinaus in einem größeren Zeitrahmen beständige Qualität in diesen Reaktionen. Das Problem, wie eine solche Beständigkeit in den emotionalen Reaktionen theoretisch zu begreifen ist, wird im folgenden noch behandelt werden.

Auf der anderen Seite des Bedeutungsfeldes SWB – der sogenannten kognitiven Komponente – thematisiert die Forschung Zufriedenheit: Demnach bewerten Menschen sich selbst bzw. ihre jeweiligen Lebensverhältnisse und das Resultat solcher Selbstevaluationen sind mehr oder weniger hoch ausgeprägte Zufriedenheiten – wobei hier je nach Breite des bewerteten Spektrums differenziert werden kann: In der SWB-Forschung werden einerseits jeweils spezifische Bereichszufriedenheiten (z.B. mit Einkommen, Gesundheit, Beruf etc.) und andererseits die unspezifische allgemeine Lebenszufriedenheit betrachtet. Letztere wäre somit die eigentliche kognitive Komponente des SWB – eben die „kognitive“ Evaluation des eigenen Lebens insgesamt.

Jedenfalls kann somit die grundsätzliche Differenzierung des SWB in affektive und kognitive Anteile gewissermaßen zum „Stand der Dinge“ in der SWB-Forschung gezählt werden (zumindest ist mir keine einschlägige Veröffentlichung bekannt, die dieser Differenzierung explizit widerspricht). Darüber hinaus wurde eine Vielzahl spezifischer Wohlbefindenskonstrukte benannt, die das Feld weiter untergliedern (auf entsprechende Auflis-

stungen bei Diener et al., 1999, und Mayring, 1991, wurde oben bereits hingewiesen). Die diesbezügliche Entwicklung der SWB-Theorie beruht zu einem gewissen Teil auf Untersuchungen, in denen spezifische Wohlbefindenskonstrukte mittels explorativer Faktorenanalysen gesucht (und gefunden) wurden, d.h. es wurden vielfach Itempools von Wohlbefindensindikatoren größeren Probandenstichproben vorgegeben und faktorenanalytisch auf die ihnen zugrundeliegende dimensionale Struktur hin untersucht (vgl. den Überblick bei Mayring, 1987). Das Ergebnis solcher Analysen sind dann jeweils bestimmte Faktorenlösungen, die inhaltlich als Komponenten des SWB interpretiert wurden. Mayring (1987; 1990; 1991) kommt in der Zusammenfassung dieser Untersuchungen zu dem Schluß, daß die Forschungslage auf dem Gebiet psychologischer Wohlbefindensstheorie am besten durch einen „Vierkomponentenansatz“ beschrieben werden könne. Demnach läßt sich also Wohlbefinden grundsätzlich in vier inhaltlich verschiedene Konstrukte differenzieren, nämlich Belastungsfreiheit, Freude, Zufriedenheit und Glück. Es scheint mir hier durchaus erhellend, die Beschreibung dieser Komponenten im Mayringschen Wortlaut wiederzugeben:

Belastungsfreiheit: „Es scheint, als ob die Abwesenheit von subjektiven Belastungen, von Symptomen, von negativen Emotionen, allein schon ein Wohlbefindensfaktor sei“ (Mayring, 1990, 169).

Freude: „Freude meint den Wohlbefindensfaktor, der kurzfristige, situationsspezifische (aktuelle) positive Gefühle umfaßt“ (Mayring, 1991, 52).

Zufriedenheit: „Wohlbefinden besitzt auch eine kognitive Komponente, die in der subjektiven Einschätzung der eigenen Lebensbedingungen besteht. Der Einzelne wägt Positives und Negatives im eigenen Leben ab, überprüft für sich, inwieweit er die eigenen Vorstellungen vom Leben erfüllt hat, seine Ziele erreicht hat. Diese kognitive Wohlbefindenskomponente wird als Zufriedenheit bezeichnet“ (Mayring, 1990, 169).

Glück: „Glück wurde . . . zunächst als langfristige, umfassende emotionale Wohlbefindenskomponente konzipiert (. . .). Dies ist allerdings eine einseitige Glücksdefinition. Es muß hier zumindest zwischen einer ‚state‘-Komponente (aktuelles, intensivstes und tiefes, die ganze Persönlichkeit umfassendes emotionales Glückserleben) und einer ‚trait‘-Komponente (langfristiges, auf Glückserleben aufgebautes, im Lebenslauf entwickeltes Lebensglück) unterschieden werden“ (Mayring, 1991, 52).

Es handelt sich hierbei also um Deskriptionen faktoranalytisch gewonnener Dimensionen, d.h. letztendlich um die nachträgliche Beschreibung der inhaltlichen Gemeinsamkeiten verschiedener Wohlbefindensindikatoren, deren zunächst nur statistische Gemeinsamkeit in einer hohen Ladung auf dem jeweiligen Faktor bestand. Den Gewinn dieser Vorgehensweise könnte man darin sehen, daß auf diese Weise sozusagen das semantische Feld des Wohlbefindens vermessen wurde: Es offenbarten sich vier Bedeutungsaspekte, die in Items zum SWB jeweils unterschiedlich stark angesprochen werden und die somit relevante Gesichtspunkte der Beschreibung des Wohlbefindens darstellen. Der Theorie des SWB sollten diese

Bedeutungsaspekte die „strukturelle“ Dimensionalität des Themenfelds liefern, sollten also diejenigen inhaltlichen Kriterien benennen, nach denen das SWB-Feld logisch gegliedert werden kann – es sollte sich sozusagen, um im oben gebrauchten Bild zu bleiben, um die Himmelsrichtungen auf der Landkarte des SWB handeln.

In dieser Hinsicht lassen nun jedoch Beschreibungen wie die oben zitierten m.E. einiges zu wünschen übrig: Auch Mayring unterscheidet kognitive (wiederum „Zufriedenheit“) von eher affektiven Wohlbefindenskonstrukten („Freude“, „Glück“) und er spricht den Aspekt der zeitlichen Dauer an – inwieweit aber die Dichotomie aktuell-kurzfristig versus habituell-längerfristig zur Unterscheidung seiner spezifischen SWB-Komponenten beiträgt, bleibt in diesen Formulierungen unklar. Dieses verdeutlicht v.a. die Dimension Glück: Wie ist die besagte „state-Komponente“ eigentlich von dem abzugrenzen, was als „Freude“ beschrieben wurde? Und umfaßt die Beschreibung der „trait-Komponente“ des Glücks nicht im Grunde die restlichen Dimensionen? *Wie* wird denn das „langfristige Lebensglück“ „auf Glückserleben aufgebaut“ und „im Lebenslauf entwickelt“ und wie ist Lebensglück eigentlich genau definiert? Tatsächlich diskutiert Mayring (1990) nicht weniger als neun Definitionen, mit denen eigentlich jeweils unterschiedliche Glückskonzepte thematisiert sind, nämlich „Glück als intensiver, momentaner, positiver Affekt“, „... Spannungslösung, Lusterlebnis“, „... positive Affektbilanz“, „... langfristiges positives Gefühl“, „... Persönlichkeitseigenschaft (trait)“, „... Erfüllung von Lebenszielen“, „... Selbstaktualisierung“, „... Lebensbejahung, positive Lebensbewertung“, „... Harmonie, Einheit der Persönlichkeit“ – wie aber läßt sich angesichts dieser Auflistung Glück konzeptionell von Wohlbefinden einerseits und von den anderen der obengenannten Wohlbefindenskomponenten andererseits abgrenzen (immerhin beklagt gerade Mayring – vgl. seinen einleitend zitierten Satz – das begriffliche Chaos in diesem Themenfeld)?

Den Mayringschen Formulierungen wurde hier aber nur deswegen so viel Beachtung geschenkt, weil sie mir beispielhaft scheinen für eine in der SWB-Literatur immer wieder vorfindbare Schwierigkeit, das Bedeutungsfeld durch inhaltlich und logisch klar gegeneinander abgegrenzte Komponenten weitergehend zu strukturieren, als durch die bloße Unterscheidung von affektiver und kognitiver Komponente. Das Problem scheint v.a. hinsichtlich der Affektivität des SWB zu existieren und dieses ist für die Belange der hier vorgestellten Untersuchung insofern von Vorteil, als ihr Interesse eigentlich der sogenannten kognitiven Komponente gilt, welche, wie bereits gesagt, in der Regel als Zufriedenheit mit dem Leben allgemein oder mit speziellen Lebensumständen im Besonderen konzeptualisiert wird. Die Probleme bei der Konzeptualisierung der affektiven Wohlbefindenskomponente jedoch scheinen v.a. darin zu bestehen, die Habitualität dieses SWB-Konstruktes „affektiv“ zu entwerfen, d.h. SWB unter Zugrundlegung situationsübergreifend konstanter Einheiten affektiver Zustände zu konzipieren. Auch wenn das affektive Wohlbefinden nicht unmittelbarer Gegenstand der Fragestellungen dieser Untersuchung ist, so ist dessen Theorie relevant für die der Zufriedenheit. Denn es wäre überhaupt die Rede vom SWB und dessen Komponenten sinnlos, wenn man völlige konzeptuelle Unabhängigkeit der beiden Komponenten

annehmen würde, und so beinhaltet z.B. die im folgenden noch genauer darzustellende Annahme von affektiver Inferenz bei der Generierung von Zufriedenheitsurteilen die explizite konzeptuelle Verknüpfung der Komponenten.

Deshalb – und gewissermaßen der Vollständigkeit halber – soll noch vor der Darstellung wesentlicher theoretischer Aspekte des Zufriedenheitskonstruktes zunächst das Problem der Konzeptualisierung des habituellen affektiven Wohlbefindens noch etwas eingehender behandelt werden.

1.1.3 Zum Konstrukt des habituellen affektiven Wohlbefindens

Die oben aufgeworfene Frage lautet: Woraus genau besteht eigentlich habituelles affektives Wohlbefinden? Gemeint sind die situationsübergreifend konstanten Einheiten affektiver Zustände, auf die das Konstrukt einer affektiven Wohlbefindenskomponente konzeptuell zu begründen ist. Es ist jedoch zunächst besonders für diesen Themenpunkt der Hinweis zu wiederholen, daß mit den folgenden Darlegungen keine Vollständigkeit angestrebt wurde: Die erschöpfende Behandlung aller in der SWB-Literatur vorgeschlagenen affektiven Wohlbefindenskonstrukte würde wohl das eigentliche Thema dieser Untersuchung verfehlen. Es soll darum hier nur anhand einiger diesbezüglich theoretisch relevanter Arbeiten aufgezeigt werden, welche Probleme diese Konzeptualisierung bereitet und wie sie lösbar sind.

Besonders aufschlußreich in der zuletztgenannten Hinsicht – d.h. für die Konzeptualisierung der in der affektiven Wohlbefindenskomponente implizierten Habitualität – ist m.E. die Darstellung theoretischer Grundlagen von Becker (1991), die den von ihm mit herausgegebenen Sammelband zur Wohlbefindensforschung einleitet. Er strukturiert dort das Themenfeld durch den Vorschlag, SWB nach den Dichotomien aktuell versus habituell und physisch versus psychisch zu differenzieren. Demnach ist also prinzipiell zu unterscheiden zwischen dem aktuellen Wohlbefinden – das ist der „Oberbegriff zur Charakterisierung des *momentanen* Erlebens einer Person, der positiv getönte Gefühle, Stimmungen und körperliche Empfindungen sowie das Fehlen von Beschwerden umfaßt“ (Becker, 1991, 13, Hervorhebung vom Autor) – und dem habituellen Wohlbefinden – nämlich „Aussagen über das für eine Person typische Wohlbefinden“ (ebd., 15). Somit gibt es in diesem Strukturmodell durch Modalität und Zeitbezug klar gegeneinander abgegrenzte „Regionen“ des Wohlbefindens, die von der Forschung betrachtet werden können, nämlich die des aktuell psychischen und des habituell psychischen, sowie des aktuell physischen bzw. des habituell physischen Wohlbefindens.

Psychisches Wohlbefinden wird hierbei als gänzlich affektives Konstrukt behandelt: Das affektive Erleben ist in diesem Modell sozusagen das Material, aus dem Wohlbefinden aufgebaut ist. So jedenfalls kommt es in zum Ausdruck den Kennzeichnungen des aktuell psychischen Wohlbefindens durch das Vorhandensein positiver und das Fehlen negativer Gefühle und Stimmungen, und des habituell psychischen Wohlbefindens durch die Häufig-

keit positiver bzw. Seltenheit negativer Gefühle/Stimmungen. Allerdings ist auch in Beckers Text oft von Zufriedenheit die Rede, jedoch wird der Begriff wiederum – darin im Einklang mit vielen anderen themenbezogenen Veröffentlichungen – eher synonym zu Wohlbefinden gebraucht und wird er nicht weiter spezifiziert. Man kann also Beckers Darlegungen als Theorie der Affektivität bzw. der sogenannten affektiven Komponente des SWB auffassen.

Dieses bedeutet aber nicht, daß „Kognitivität“ in diesem Theorieentwurf keine Rolle spielt: Habituelles (psychisches) Wohlbefinden ist als das Resultat kognitiver Verarbeitung affektiver Erfahrungen konzipiert, es handelt sich um „Urteile über aggregierte emotionale Erfahrungen. Durch das Wort ‚Urteile‘ soll zum Ausdruck gebracht werden, daß Aussagen über HW (z.B. Aussagen zur allgemeinen Lebenszufriedenheit) primär durch kognitive Prozesse zustande kommen“ (Becker 1991, 15, Hervorhebung vom Autor). Betrachtet man die Beckersche Definition noch einmal genau, so fällt auf, daß die Formulierung das habituelle Wohlbefinden („HW“) selbst mit *Aussagen* gleichsetzt. Demnach handelte es sich dabei also genaugenommen nicht um das für eine Person typische Wohlbefinden, sondern die Aussage darüber. Diese Redeweise kann eigentlich nur so verstanden werden, daß das personentypische Wohlbefinden nicht als unmittelbar erfahrbares, psychisch reales Merkmal – etwa im Sinne eines positiv getönten allgemeinen Lebensgefühls oder einer zeitlich überdauernden Grundstimmung – aufgefaßt wird, sondern als kognitive Konstruktion, die in der Aussage, im bilanzierenden Urteil über die erlebten Affektzustände, überhaupt erst entsteht. Beckers knappe Charakterisierung habituellen (psychischen) Wohlbefindens impliziert also, daß der habituelle Aspekt des SWB ganz und gar an kognitive Verarbeitungen geknüpft wird, daß sozusagen die Habitualität des Wohlbefindens selbst als kognitives Konstrukt konzeptualisiert ist. Daß dieses in den oben zitierten Formulierungen so deutlich zum Ausdruck kommt, ist der Grund dafür, daß ausgerechnet diese Arbeit hier so ausgiebig zitiert wird.¹

Diese Kognitivität des habituellen Wohlbefindens wirft m.E. nun jedoch auch Fragen nach der Art und Weise der stattfindenden Verarbeitungsprozesse, d.h. danach, *wie* diese Urteile zustande kommen, auf. Folgt man hier Beckers Darstellung weiterhin, so erscheint die Sache zunächst klar und einfach: Je mehr positive und je weniger negative Affekte, desto größer das habituelle psychische Wohlbefinden. Jedoch wird man dann fragen müssen: Findet lediglich eine Art einfachster Auszählung hedonisch getönter Erlebnisse statt? Spielt dabei eine Art von Gewichtung dieser Erlebnisse – z.B. hinsichtlich Intensität und/oder Dauer – eine Rolle? Und weiter: Wie grenzen sich Gefühle und Stimmungen, die doch, um als Einheiten einer solchen Verrechnung zu taugen, sozusagen als abzählbare Elementarereignisse gegeben sein müssen, jeweils gegeneinander zeitlich und inhaltlich ab?

¹ Allerdings erscheint es mir durchaus fraglich, inwieweit Becker diese Implikation tatsächlich intendiert hat: Immerhin öffnet das Wort „primär“ im oben wiedergegebenen Satz sozusagen ein Hintertürchen für Beiträge zum habituellen Wohlbefinden, die nicht aus kognitiver Verarbeitung stammen. Verwandte theoretische Konzeptionen, wie die von Mayring (1990; 1991), schlagen jedenfalls auch emotionale Komponenten des habituellen Wohlbefindens (neben der kognitiven) vor.

1.1.3.1 Die Affektfrequenztheorie von Diener, Sandvik & Pavot

Einige mögliche – und empirisch belegte – Antworten auf die somit aufgeworfenen Fragen bezüglich der Verarbeitung des affektiven Erlebens zu habituellem Wohlbefinden geben Diener et al. (1991), welche bereits im Titel ihrer Veröffentlichung eine komprimierte Zusammenfassung ihres Inhalts liefern: „Happiness is the frequency, not the intensity, of positive versus negative affect.“ Die im hier interessierenden Zusammenhang theoretisch relevanten Aspekte dieser Arbeit könnte man wie folgt zusammenfassen: (Habituelles) Wohlbefinden beruht auf der Frequenz des affektiven Erlebens, hohes Wohlbefinden ist das Resultat häufiger positiver und seltener negativer Emotionen, die emotionale Intensität spielt dabei kaum eine Rolle. Die Frequenz positiver bzw. negativer Emotionen ist aber konzeptualisiert als relativer Zeitanteil in einem bestimmten Erlebenszeitraum, in dem das affektive Erleben der Person positiv (und nicht negativ) getönt ist. Diese Konzeptualisierung wird mit dem von Brandstätter (1987) entlehnten Argument einer Dichotomie der affektiven Tönung augenblicklichen Erlebens begründet: „... there is a natural ‚point of indifference‘ in emotion, above which people feel positive and below which they feel negative“ (Diener et al., 1991, 120). Es wird angenommen, daß Personen jederzeit (in wechselnder Intensität) Emotionen durchleben, daß also sozusagen das emotionale System einer Person niemals gänzlich inaktiv ist und in jedem Augenblick die Summe der jeweiligen Emotionen eine positive oder negative hedonische Tönung ergibt. Das oben hinsichtlich Beckers Konzept habituellen Wohlbefindens aufgeworfene Problem der Abgrenzung elementarer Gefühls-einheiten, deren Frequenzen zu einem Gesamturteil verrechnet werden, wäre damit elegant gelöst: Im Grunde handelt es sich nicht um die absolute Häufigkeit einzelner positiver Affekte, sondern um die relative Dauer des insgesamt positiv getönten Erlebens im Verhältnis zur Dauer des insgesamt negativ getönten – und da Personen stets entweder positiv oder negativ gestimmt (und auch zu entsprechenden Selbstauskünften in der Lage) sind, ist eine solch eindeutige dichotome Zeitaufteilung möglich.

In ihrer Arbeit liefern die Autoren empirische Belege für diese Bedeutung der Affektfrequenz: In mehreren Untersuchungen fanden sie, daß die solchermaßen konzipierte Frequenz, mit der positive Affekte erlebt werden, substantielle Beiträge zur Varianzaufklärung gebräuchlicher Indikatoren des SWB (der Begriff „happiness“ wird synonym zu „subjective well-being“ gebraucht, gemeint ist ausdrücklich das Wohlbefinden über einen längeren Zeitraum hinweg) leistet, während die entsprechenden Effekte der Intensität des positiven Affekterlebens gering ausfielen. Darüber hinaus zeigten sich in den Extremgruppen der Versuchspersonen mit sehr hoher bzw. sehr niedriger positiver Affektfrequenz ausschließlich sehr hohe bzw. sehr niedrige Werte der Wohlbefindensindikatoren, was die Autoren als empirischen Hinweis darauf werten, daß diese Frequenz den Charakter einer notwendigen und hinreichenden Bedingung für SWB habe. Diese Ergebnisse zeigten sich nicht nur hinsichtlich solcher „Glücksindikatoren“, die auf Selbstauskunft der Probanden beruhen, sondern auch bei solchen, die aus Fremdeinschätzungen (z.B. Expertenratings auf der Basis strukturierter Interviews) resultieren. Empirische Belege für einen „natürlichen Indifferenzpunkt“

des emotionalen Erlebens bzw. für die Eindeutigkeit der hedonischen Tönung jedes erlebten Augenblicks lieferten beispielsweise Diener & Iran-Nejad (1986).

Theoretisch wäre eine solche Bedeutung dieser positiven Affektfrequenz für das habituelle Wohlbefinden damit zu begründen, daß diese Information von Individuen besonders leicht verarbeitet werden kann: Die einfache dichotome Information kann leichter gespeichert und abgerufen werden, als etwa Information über die Intensität von Emotionen. So haben Hasher & Zacks (1979; 1984) zeigen können, daß zur normalen Funktionsweise menschlicher Informationsverarbeitung ein sehr genauer Abruf von Häufigkeitsinformationen aller Art gehört, sie begründen dies mit dem Argument einer biologischen Veranlagung zur Aufmerksamkeit gegenüber und Speicherung solcher Information. Thomas & Diener (1990) untersuchten die Genauigkeit, mit der Personen affektrelevante Informationen erinnern, und fanden wiederum, daß ihre Probanden den prozentualen Anteil glücklicher Stimmung während der Tageszeit überraschend genau einschätzen konnten. Bezüglich der nachträglichen Schätzung von Gefühlsintensitäten fanden sich dagegen keine derartigen Effekte.

Es ließen sich somit Fragen, die durch die oben zitierten Beckerschen Formulierungen aufgeworfen wurden, sehr gut durch die Hypothesen von Diener et al. (1991) – und die empirischen Belege, die dafür geliefert wurden – beantworten und das durch jene Formulierungen beschriebene Modell habituellen affektiven Wohlbefindens sozusagen zu einer erklärungs mächtigen psychologischen Theorie ergänzen, die im weiteren der Einfachheit halber als *Affektfrequenztheorie* des SWB bezeichnet wird: Habituelles affektives Wohlbefinden reflektiert das affektive Erleben einer Person, es entsteht durch kognitive Verrechnung dergestalt, daß es v.a. mit der relativen Häufigkeit positiver versus negativer Affekte bzw. dem Anteil der in positiver Tönung erlebten an der insgesamt erlebten Zeit ansteigt, wobei die Intensität der emotionalen Tönung eine eher nachgeordnete (wenn auch nicht prinzipiell auszuschließende) Rolle spielt.

Damit wäre nun insbesondere konkretisiert, auf welche Weise im SWB Affektivität überhaupt habituell konstituiert sein kann. Daß auf dieses Theoriefundament noch ein Gebäude gesetzt werden müßte, welches die Spezifizierung verschiedener inhaltlicher Aspekte – z.B. die Frage nach der Konstituierung des relevanten Zeitabschnitts, über den affektives Erleben aggregiert wird, – enthält, versteht sich von selbst.

1.1.3.2 Theorien unabhängiger Affektdimensionen

Bradburn (1969) bzw. Bradburn & Caplovitz (1965) begründen eine Theorie habituellen affektiven Wohlbefindens auf einer Art von Affektbilanz, allerdings unter Voraussetzung relativ unabhängiger Dimensionen positiven und negativen Affekterlebens: Gemäß dieser Theorie wäre es sozusagen die subjektiv wahrgenommene Bilanz aus erlebten positiven versus negativen Affekten – hier jedoch nicht als Affektfrequenz des Anteils der auf einem Positiv-Negativ-Kontinuum auf der Positivseite erlebten Zeit –, die der Einzelne zum

jeweiligen Zeitpunkt zieht, welche habituelles affektives Wohlbefinden begründet. Auch Diener et al. (1999) zählen, wie oben bereits erwähnt, die Differenzierung der affektiven SWB-Komponenten in voneinander weitgehend unabhängige Dimensionen positiver und negativer Affektivität zum Stand der Dinge in der SWB-Forschung. Tatsächlich liegt die konzeptuelle Unterscheidung der relativ unabhängigen Konstrukte negativer bzw. positiver Affekt einer Vielzahl von Untersuchungen zugrunde, in denen explizit affektives Wohlbefinden untersucht wurde (z.B. Kunzmann et al., 2000; Headey & Wearing, 1991; Costa et al., 1987; Costa & McCrae, 1984).

Viele Arbeiten – u.a. wiederum der Forschergruppe um Ed Diener – (Diener & Emmons, 1984; Diener & Larson, 1993; Diener et al., 1995; Lawton et al., 1992) belegten immer wieder die relative Unabhängigkeit der Häufigkeit, mit der positive Emotionen erlebt werden, von der Häufigkeit negativer Emotionen. Mit Blick auf die sozusagen von Ed Diener in Erstautorenschaft verbreitete Affektfrequenztheorie wäre hier nun auch zu fragen, wie diese mit der Unabhängigkeit der positiver und negativer Affektdimensionen vereinbar ist: Wenn ein „natürlicher Indifferenzpunkt“ des affektiven Erlebens angenommen wird, so daß selbiges stets *entweder* positiv *oder* negativ und niemals neutral getönt ist, so folgt daraus die wechselseitige Abhängigkeit positiv und negativ getönter Zeitanteile. Der dadurch zunächst aufscheinende Widerspruch zur Konzeption unabhängiger Affektdimensionen löst sich auf bei genauer Betrachtung dessen, was damit in den entsprechenden Veröffentlichungen gemeint ist: Es handelt sich um Häufigkeiten, mit denen diskrete und thematisch definierte Emotionen wie Freude, Liebe, Ärger, Furcht erlebt werden. Diese Emotionen sind jeweils durch die positive oder negative Tönung des dabei stattfindenden Affekterlebnisses charakterisierbar und Diener et al. (1995) fanden substantielle korrelative Zusammenhänge der Häufigkeiten diskreter Emotionen desselben hedonischen Typus, so daß insgesamt die statistischen Abhängigkeiten zwischen den Emotionen durch ein Zwei-Faktoren-Modell abgebildet werden konnten, in welchem „positiver Affekt“ und „negativer Affekt“ die Rolle übergeordneter latenter Konstrukte spielen – welche auf lediglich mittlerem Niveau negativ korrelierten und demnach jeweils große wechselseitig unabhängige Varianzanteile besaßen.

Die solchermaßen konzipierten beiden Affektdimensionen sind nun allerdings nicht gleichzusetzen mit relativen Zeitanteilen, die in positiver versus negativer Tönung durchlaufen werden, insbesondere besagt die Theorie eines „natürlichen Indifferenzpunktes“ affektiver Tönung nicht, daß Personen sozusagen zu jedem Zeitpunkt ihres Wachbewußtseins eine thematisch genau definierte, diskrete Emotion durchleben. Zudem zählt zu den Charakteristika von Emotionen wie den obengenannten m.E. auch eine spezifische Intensität des Affekterlebens, während, wie bereits erwähnt, in der Arbeit von Diener et al. (1991) diese Intensität sich nicht als relevanter Prädiktor von SWB erwies und folglich in der Affektfrequenztheorie keine Rolle spielt. Der hedonische Tonus, so könnte man also sagen, in dem sich gemäß der Affektfrequenztheorie eine Person stets befindet, erschöpft sich nicht in den grundlegenden und intensiven Emotionen, welche die beiden unabhängigen Dimensionen positiver und negativer Affektivität konstituieren, die in der Forschung postuliert wurden,

er kann darüber hinaus sowohl andere oder weniger eindeutig thematisch umgrenzte emotionale Befindlichkeiten, als auch relativ „emotionslose“ Zustände geringer affektiver Intensität umfassen. Hinsichtlich korrelativer Zusammenhänge zwischen den beiden Affektdimensionen sind daraus gewissermaßen diejenigen Erwartungen ableitbar, welche durch die oben zitierten Untersuchungen bestätigt wurden: Man kann zwar allgemein annehmen, daß eine Person, die beispielsweise sehr viel Zeit im positiv getönten Zustand verbringt, auch häufiger die besagten diskreten positiven und seltener die negativen Emotionen erlebt, jedoch sollte es sich dabei um lediglich unvollständige Zusammenhänge handeln. Aus dem Umstand, daß gemäß der Affektfrequenztheorie per definitionem ein vollständiger negativer Zusammenhang zwischen den in positiver und negativer Tönung verbrachten Zeitanteilen besteht, folgt *nicht*, daß entsprechendes für die Dimensionen positiven und negativen Affekts gilt, so daß auch ein lediglich geringerfügiger negativer Zusammenhang zwischen diesen Dimensionen im Einklang mit der Affektfrequenztheorie steht.

Somit erweisen sich also die Konzeption relativ unabhängiger Affektdimensionen und die Affektfrequenztheorie keineswegs als wechselseitig widersprüchlich. Es könnte jedoch gefragt werden, welches der beiden Konzepte sozusagen das wohlbefindensrelevantere ist. Mit Blick auf die Dimensionen positiver und negativer Affektivität könnte m.E. auch argumentiert werden, daß diese konzeptuell gar nicht unmittelbar auf das Wohlbefindenskonstrukt als solches bezogen und im Grunde nicht als Subkategorien des affektiven SWB aufzufassen sind: Vielmehr handelt es sich bei diesen Dimensionen um Konstrukte einer psychologischen Emotionstheorie, wie besonders deutlich die von Diener et al. (1995) diskutierten Erklärungen ihrer Ergebnisse zeigen: Danach könnten beispielsweise die unabhängigen Affektdimensionen die Aktivität zweier unterschiedlicher zentraler Hirnfunktionen anzeigen, welche Annäherungs- („approach-positive“) bzw. Vermeidungsreaktionen („avoidance-negative“) erzeugen, die durch nachgeordnete erlernte Bewertungen jeweils zu den bekannten thematisch bestimmten positiven oder negativen diskreten Emotionen werden. Oder, umgekehrt, könnten verschiedene intensive diskrete Emotionen auf der Aktivität jeweils dafür spezifischer Hirnstrukturen beruhen, wobei der Zusammenhang von Emotionen derselben hedonischen Valenz durch dichtere neuronale Vernetzung der entsprechenden Aktivitäten zustande käme. Mit derartigen Theorien wird Emotion überhaupt und insbesondere auch die positive oder negative hedonische Qualität derselben aus den Aktivitäten zentraler Hirnareale erklärt. Die Affektfrequenztheorie kann demgegenüber als nachgeordnet bezeichnet werden, sie erklärt den Zusammenhang zwischen Emotion und SWB und darf nicht mit der eigentlichen Emotionstheorie, in der wechselseitig unabhängige emotionale Subsysteme der positiven und negativen Affektivität konzipiert sind, verwechselt werden. Die Affektfrequenztheorie besagt, daß letztendlich die pure Frequenz positiven affektiven Erlebens, die aus dem gleichzeitigen Funktionieren dieser beiden Subsysteme resultiert, die wohlbefindensrelevante Größe ist, auf welcher die affektive Komponente des SWB beruht.

Die Affektfrequenztheorie wäre also durchaus mit den empirischen Befunden relativer Unabhängigkeit der beiden Affektdimensionen vereinbar. Sie ist, das sollten auch die

hier nur sehr knappen und ausschnitthaften Darlegungen zur Theorie der affektiven Wohlbefindenskomponente gezeigt haben, keinesfalls *die* Theorie dieser Komponente, die die diesbezügliche Forschungslandschaft bestimmt – sondern nur eine m.E. sozusagen besonders attraktive Möglichkeit der Konzeptualisierung habituellen affektiven Wohlbefindens, weil sie die durch die Habitualität aufgeworfenen Fragen besonders klar und eindeutig beantwortet.

1.1.4 Zufriedenheit als Wohlbefindenskomponente

Im Vergleich zu den Fragen, die die theoretische Konzeption einer affektiven Wohlbefindenskomponente aufwirft, erscheint die einer kognitiven Komponente des SWB zunächst relativ unproblematisch: Diese wird, wie bereits gesagt, in der einschlägigen Forschung im Allgemeinen als „Zufriedenheit“ konzeptualisiert. Zufriedenheiten sind demnach subjektive Bewertungen eigener Lebensumstände und als solche ein Aspekt des SWB einer Person, der von deren Selbstwahrnehmung bzw. -beurteilung eigener affektiver Gemütslagen grundsätzlich unterschieden werden kann. Was aber – darauf sei im Vorgriff auf noch folgende Argumentationen hingewiesen – nicht zwangsläufig bedeutet, daß Zufriedenheitsurteile gewissermaßen im affektfreien Raum gebildet werden: Die Unterscheidung von affektiver und kognitiver Komponente meint im Grunde inhaltlich verschiedene Gegenstände subjektiven Erlebens, die dasjenige bestimmen, was insgesamt mit dem Begriff SWB benannt werden kann, nämlich einmal „objektive“ Realitäten, wie z.B. die materielle Situation oder den gesundheitlichen Zustand der Person, und zum anderen – siehe die Affektfrequenztheorie – die Emotionen, die diese Person durchlebt. Daß nun allerdings Zufriedenheit aus einer gewissermaßen emotionslosen, unter vollständiger Ausblendung affektiver Befindlichkeit sich vollziehenden Bewertung solcher objektiven Umstände resultiert, ist damit keinesfalls gesagt.

1.1.4.1 Zufriedenheit und objektive Lebensqualität

Die solchermaßen als Zufriedenheit konzipierte kognitive Komponente des SWB spielt in der sozialwissenschaftlichen Forschung eine gewichtige Rolle, nämlich v.a. als Aspekt menschlicher Lebensqualität, welche zum einen hinsichtlich objektiv feststellbarer (sozialer, ökonomischer, aber auch physischer u.a.), zum anderen aber hinsichtlich subjektiver – und d.h. interindividuell variierender – Kriterien dargestellt werden kann (objektive versus subjektive Lebensqualität, vgl. Veenhoven, 1996). Subjektive Lebensqualität meint demnach Zufriedenheit mit objektiven Lebensumständen: „Satisfaction is a central concept in research on subjective quality-of-life. All quality-of-life surveys involve questions on the matter“ (Veenhoven, 1996, 1). Dabei werden die Begriffe subjektive Lebensqualität, Zufriedenheit und SWB in sozialwissenschaftlichen Arbeiten oft synonym verwendet, so daß gewissermaßen ein großer – nämlich weitestgehend der soziologische – Teil der wissen-

schaftlichen Forschung zum SWB im Grunde Zufriedenheitsforschung ist, nämlich in dem Sinne, daß darin wie auch immer operationalisierte Zufriedenheiten diejenigen Größen sind, die mit solcher Forschung erklärt werden sollen. Die Entwicklung dieser Zufriedenheitsforschung stellt wiederum ein komplexes und weitverzweigtes empirisches und theoretisches Problemfeld dar, das nur schwer zu überschauen ist. Wie schon im Falle der allgemeinen SWB-Forschung das Review von Diener et al. (1999), so bietet sich hier der vielzitierte Überblicksartikel von Veenhoven (1996) gewissermaßen als Führer durch dieses Gebiet an.

Ein gewisser „Forschungsmainstream“ sozialwissenschaftlicher Untersuchungen zum SWB könnte also durch die Frage charakterisiert werden, welche objektiven Lebensbedingungen in welchem Ausmaß Einfluß auf Indikatoren des SWB, nämlich v.a. auf Zufriedenheiten, ausüben. Nun kann aber diese Frage auch so verstanden werden, daß darin nicht nur die Identifizierung relevanter objektiver Prädiktoren des SWB, sondern insbesondere auch diese Relevanz als solche angesprochen ist, und tatsächlich stand bereits am Beginn derartiger Forschung die Erkenntnis teilweise überraschend geringer empirischer Zusammenhänge zwischen objektiven und subjektiven Indikatoren der Lebensqualität (Veenhoven, 1996). Damit aber sind naive Vorstellungen einer weitestgehenden, quasi-deterministischen Verursachung subjektiver durch objektive Lebensqualität zunächst diskreditiert und es zeigt sich die Notwendigkeit einer differenzierteren Konzeptualisierung von Zufriedenheit und der kausalen Bezüge, innerhalb derer sie sich vollzieht.

Ein wesentlicher Bestandteil des sozialwissenschaftlichen Zufriedenheitskonzepts wurde die Differenzierung anhand des inhaltlichen Fokus der Bewertung: Neben spezifischen Zufriedenheiten mit mehr oder weniger eng gefaßten Lebensbereichen – wie z.B. mit der Gesundheit, dem Einkommen, der Freizeit, den Wohnverhältnissen u.dgl. – wurde die allgemeine Lebenszufriedenheit als „Globalmaß subjektiven Wohlbefindens“ (Buhlman, 1996, 79), also als zusammenfassende und unspezifische subjektive Bewertung der eigenen Lebenssituation, Gegenstand der Forschung. Es besteht auch bezüglich dieser Differenzierung eine gewisse begriffliche Uneinheitlichkeit in der SWB-Literatur: Bisweilen beschränkt sich die Behandlung von Zufriedenheit als Komponente des SWB auf unspezifische Zufriedenheit mit dem Leben im allgemeinen, bisweilen jedoch ist Zufriedenheit überhaupt und sind also auch die spezifischen Bereichszufriedenheiten gemeint. So unterscheiden z.B. Headey et al. (1991) bei ihrer Untersuchung zu der sogenannten Top-down/Bottom-up-Kontroverse begrifflich und konzeptionell zwischen „domain satisfactions“ und SWB, wobei „life satisfaction“ als dessen Komponente benannt wird. Andererseits formulieren beispielsweise Diener et al. (1999, 277): „Subjective well-being is a broad category of phenomena that includes people’s emotional responses, domain satisfactions, and global judgements of life satisfaction.“ Das Problem aber scheint mir kein derart grundlegendes, daß es unbedingter Klärung bedürfte: Ob man nun die Bereichszufriedenheiten unter dem SWB-Begriff subsummieren will, oder nicht, ist m.E. eine bloß terminologische Frage ohne grundsätzliche Bedeutung für die Theorie des SWB – diese besteht hier vielmehr in der Frage nach (kausalen) Zusammenhängen zwischen Bereichszufriedenheiten und den son-

stigen SWB-Komponenten, wie sie v.a. in der besagten Top-down/Bottom-up-Kontroverse gestellt wird (vgl. Kapitel 1.1.4.7). Es bleibt also vorerst lediglich festzuhalten, daß meist dann, wenn in einschlägigen Veröffentlichungen von SWB die Rede und Zufriedenheit gemeint ist, es genau genommen um allgemeine Lebenszufriedenheit geht, – und daß zur wissenschaftlichen Thematisierung von Zufriedenheit die Fokussierung auf unterschiedliche Ebenen inhaltlicher Differenzierung des Phänomens gehört.

Anfänglich beruhte diese Differenzierung auf einer *Bottom-up-Theorie* (Diener, 1984) der Zufriedenheit, nämlich auf der Annahme, daß die Zufriedenheit von Personen sich sozusagen von unten nach oben entwickelt, wobei das „Unten“ die Basis der objektiven Lebensumstände meint, das „Oben“ dagegen die Zufriedenheit mit dem Leben als Ganzes (vgl. z.B. Campbell et al., 1976). Dazwischen liegen die Bereichszufriedenheiten: Menschen beurteilen demnach ihre jeweilige objektive Lage in verschiedenen wichtigen Lebensbereichen, mit dem Ergebnis spezifischer Bereichszufriedenheiten, welche dann zu einer globalen Gesamtzufriedenheit verrechnet werden. Eine einfache Darstellung dieses Modells liefern z.B. Smith et al. (1996, 499: Abbildung 1). In einem solchen Bottom-up-Zufriedenheitsmodell ist also die Erwartung impliziert, daß Zufriedenheit (welche auch immer) die objektiven Lebensrealitäten eines Menschen widerspiegelt: „Gute“ Wohnverhältnisse führen zu hoher Wohnzufriedenheit, „gute“ materielle Ausstattung zur hoher Einkommenszufriedenheit und die allgemeine Lebenszufriedenheit wird umso besser sein, je mehr solcher „guten“ Lebensbedingungen realisiert sind. Zu erforschen bleibt dabei, welchen speziellen objektiven Merkmalen jeweils subjektive Relevanz bei der Entstehung von Zufriedenheit zukommt.

Wie jedoch oben bereits erwähnt, zeigten sich in empirischen Studien die korrelativen Zusammenhänge zwischen objektiven sozioökonomischen Indikatoren einerseits und den erfragten Zufriedenheiten andererseits nicht in der gemäß des Bottom-up-Ansatzes zu erwartenden Höhe. So berichtet Veenhoven (1996, 23) mit Blick auf die Bereichszufriedenheiten, für die solche Korrelationen besonders hoch ausfallen müßten: „The great surprise was that only a part of the variance in satisfaction can be explained that way“ – gemeint ist die Varianzaufklärung durch objektive Indikatoren. Es ist also hinsichtlich spezifischer Zufriedenheiten zu konstatieren, daß die hier erwarteten sehr hohen Korrelationen mit Maßen der objektiven Lebensqualität in empirischen Studien i.d.R. nicht erreicht werden (Veenhoven, 1996). Dasselbe gilt für die allgemeine Lebenszufriedenheit, für die, als global aggregierte Gesamtbewertung, zwar geringere bivariate Korrelationen mit einzelnen Merkmalen objektiver Lebensqualität zu erwarten wären, jedoch bei weitestgehender Verursachung durch Bottom-up-Effekte eine insgesamt hohe multiple Korrelation mit Gesamtmengen objektiver Indikatoren. Veenhoven (1996, 26) stellt z.B. für den Zusammenhang allgemeiner Lebenszufriedenheit mit fundamentalen sozioökonomischen Variablen (Alter, Geschlecht, Einkommen, Bildung, Berufstätigkeit) fest: „Together, the above socio-demographic variables explain mostly no more than 10% of the variance in individual life-satisfaction, at least in rich nations.“ Dieses alles aber, darauf sei ausdrücklich hingewiesen, bedeutet nicht, daß

überhaupt keine erwartungskonformen Korrelationen gefunden wurden. Tatsächlich zeigten sich durchaus die Zusammenhänge gemäß der Erwartungen des Bottom-up-Ansatzes, wenngleich auch *in der Höhe* nicht diesen Erwartungen entsprechend (konkrete und relativ aktuelle Werte für Deutschland liefert z.B. Bulmahn, 1996; eine nahezu erschöpfende weltweite Erfassung empirischer Befunde zu Korrelaten von Zufriedenheit liefert Veenhoven, 2003).

Gefundene Zusammenhänge belegen also zwar die Wirksamkeit von Bottom-up-Effekten, nicht jedoch ein einfaches Bottom-up-Modell der Zufriedenheit. Die Erklärung niedriger empirischer Zusammenhänge zwischen objektiven Lebensbedingungen und subjektiver Lebensqualität hat sich nun sozusagen als kreativitätsförderndes Element der Zufriedenheitsforschung erwiesen: Es wurden Gründe dafür gesucht, daß die subjektiven Beurteilungen objektiver Lebensumstände offenbar zu einem nicht unerheblichen Teil gar nicht von letzteren beeinflußt werden. Sie wurden gefunden in differenzierteren theoretischen Annahmen über die Generierung von Zufriedenheitsurteilen, die im folgenden unter den Gesichtspunkten der Relativität bzw. Adaptivität von Zufriedenheit, Persönlichkeitseinflüsse auf das SWB und affektiver Inferenz dargestellt und diskutiert werden.

1.1.4.2 Messung von Zufriedenheit

Zunächst aber ist noch ein weiterer Gesichtspunkt zu erwähnen, der nicht die psychologische Theorie der Zufriedenheit unmittelbar betrifft, jedoch auch geringe Korrelationen zwischen objektiven Lebensbedingungen und Zufriedenheitsmessungen erklären könnte, nämlich die möglicherweise mangelnde Reliabilität letzterer. Damit ist die Frage der adäquaten Operationalisierung von Zufriedenheit in empirischen Untersuchungen angesprochen.

Es existiert eine Vielzahl von Verfahren zur Erfassung von SWB, insbesondere zur Messung von Zufriedenheiten (zum Überblick siehe: Mayring, 1991). In der sozialwissenschaftlichen Umfrageforschung haben sich inzwischen Ein-Item-Messungen von Zufriedenheit etabliert, meist schlicht Fragen der Art „Wie zufrieden sind sie mit ...?“, die von den Probanden mittels mehrfach abgestufter Ratingskalen beantwortet werden können. Ein Beispiel hierfür sind die entsprechenden 11-Punkte-Skalen des deutschen Sozio-oekonomischen Panels (vgl. Kapitel 3.1.2), mit deren Meßwerten die nachfolgend dargestellten eigenen empirischen Analysen durchgeführt wurden. Mit Blick auf seine Auflistung von Mehr-Item-Skalen zur Erfassung von Lebenszufriedenheit befindet Mayring (1991, 60): „Solche Skalen sind in jedem Fall der Ein-Item-Form vorzuziehen“ – er bleibt aber jede weitere Begründung bzw. empirische Belege für dieses Verdikt schuldig. Im Review von Veenhoven (1996) dagegen werden die meßtheoretischen Qualitäten der Ein-Item-Ratings nicht grundsätzlich schlechter bewertet, als die Mehr-Item-Skalen: Es scheint so, als sei die höhere Reliabilität von Mehr-Item-Skalen empirisch nicht wirklich bestätigt.

Für die nachfolgende Untersuchung ist in diesem Zusammenhang wichtig, daß Ein-

Item-Skalen aufgrund vorliegender Befunde zumindest nicht als unbrauchbar zu gelten haben (vgl. z.B. Scherpenzeel, 1995). Speziell für die Zufriedenheitsmessungen des sozio-ökonomischen Panels läßt sich diese Einschätzung auch aufgrund der Analysen von Landua (1993) und Schräpler (1995) bekräftigen, in denen jeweils durchaus befriedigende Reliabilitäten für diese Messungen geschätzt wurden.

Ganz abgesehen von solchen vorhandenen empirischen Belegen für die ausreichende Güte der Ein-Item-Messungen wurden in der nachfolgenden Untersuchung u.a. auch Strukturgleichungsmodelle berechnet, die eine explizite Schätzung von Reliabilitäten und meßfehlerbereinigten True-Score-Zusammenhängen beinhalten (vgl. die Darstellung dieser Modelle in Kapitel 2.2.2).

1.1.4.3 Relativität von Zufriedenheit

Auf inhaltlicher Ebene werden zur Erklärung der besagten niedrigen Korrelationen häufig Vorstellungen über die internale, psychologische Dynamik subjektiver Lebensqualität herangezogen, die in Anlehnung an Veenhoven (1991; 1996) unter dem Begriff der Relativität von Zufriedenheit zusammengefaßt werden können. Demnach beruhen Zufriedenheitsurteile auf einem Vergleich der tatsächlichen Lebensbedingungen mit subjektiven Standards – sozusagen internalen Repräsentationen von Idealzuständen (des Einkommens, der Gesundheit, des Wohnens usw.). Die Setzung dieser subjektiven Standards, so wird weiterhin angenommen, orientiert sich an der Wahrnehmung der Möglichkeiten, die in der jeweiligen Lebenslage als realisierbar erscheinen, welche sich v.a. auch sozialer Vergleichsprozesse bedient. Ein Beispiel, in dem diese Vorstellungen explizit formuliert wurden, ist die „multiple discrepancies theory“ von Michalos (1985; 1986). Die Relativität von Zufriedenheit ergibt sich dann aus der Hypothese einer beständigen Adjustierung der subjektiven Standards (Adaptationsniveautheorie): „At the individual level, this theory predicts that satisfaction is a short-lived phenomenon. We would be satisfied when life comes close to ideal, but in coming closer to ideal, we would tend to set higher demands and end up as equally dissatisfied as before. Likewise, social comparison would impede lasting satisfaction. When we have surpassed the Jones, reference drifts upwards to the Petersons, and we feel dissatisfied again“ (Veenhoven, 1996, 31). Es wäre also somit zu erwarten, daß über die Gesellschaft hinweg betrachtet Personen in objektiv guten Lebenslagen nicht zufriedener sind, als Personen in schlechten Lebenslagen, und folglich ist auch nicht mit hohen Korrelationen zwischen Zufriedenheiten und Indikatoren objektiver Lebensqualität zu rechnen.

Einen detaillierteren Überblick über psychologische Theorien, die solchermaßen die Relativität von Zufriedenheit postulieren, liefern Diener et al. (1999). Insbesondere bezüglich der erwähnten Theorie sozialer Vergleiche lautet das dort gezogene Fazit: „The question is no longer whether or not social comparison can affect SWB. ... The interesting questions now concern when and where social comparison produces which types of effects. The

choice of comparison, the use of upward or downward comparison, and the directionality of effects on SWB are flexible and, to some degree, dependent on personality“ (Diener et al., 1999, 283).

Adaptation, d.h. herabgesetzte Reaktivität auf wiederholte oder andauernde Stimuli, ist zu einem wichtigen Konzept neuerer psychologischer Theorien zum SWB geworden (Diener et al., 1999; Staudinger, 2000). Die Adjustierung subjektiver Standards bezüglich objektiver Lebensbedingungen bedeutet Adaptation in diesem Sinne und im Resümee der empirischen Befundlage zu solchen Adaptationseffekten kommen Diener et al. (1999) zu einem ähnlichen Ergebnis, wie hinsichtlich der Theorie sozialer Vergleiche: Die teilweise Wirksamkeit von Adaptationseffekten auf SWB ist empirisch ebenso belegt, wie deren teilweise Unwirksamkeit. Untersuchungen, die Adaptationseffekte zeigten, stehen solchen gegenüber, in denen keine Adaptation an veränderte Lebensumstände gefunden wurde, so daß hier letztlich wiederum eine Differenzierung gefordert ist: „A complete theory of subjective well-being must explain the effects of the temporal context of events; when adaptation occurs, what processes are responsible for adaptation, and any limits to individuals' abilities to adapt“ (Diener et al., 1999, 288).

Veenhoven (1996; 1991) bewertet in seiner Zusammenschau vorliegender empirischer Befunde die „Relativitätstheorie“ der Zufriedenheit negativ: Auf einen einfachen Nenner gebracht, sprechen substantielle Korrelationen von Zufriedenheiten mit objektiven Indikatoren, welche ja, wie gesagt, immer wieder gefunden wurden (wenn auch nicht in erwarteter Höhe), gegen die Relativität von Zufriedenheitsurteilen. Folgt man dem Veenhovenschen Fazit, so muß man zu dem Schluß kommen, daß Relativierungsprozesse – Adjustierungen von Vergleichsstandards – beim Zustandekommen von Zufriedenheit eine Rolle spielen können, daß es sich dabei aber sozusagen eher um eine Nebenrolle handeln dürfte. Er befindet sich damit nicht wirklich im Widerspruch zu den Schlußfolgerungen von Diener et al. (1999), der Unterschied der beiden Resümees bezüglich der Relativität von Zufriedenheit scheint eher in einer wechselseitigen Akzentverschiebung zu liegen: Veenhoven legt deutlicheres Gewicht auf die Feststellung, daß Zufriedenheitsurteile nicht wirklich vollkommen relativ und damit unabhängig von tatsächlichen Lebensrealitäten gefällt werden. Er betont damit im Grunde eine Selbstverständlichkeit: Zufriedenheitsurteile sind Urteile über die gegebenen Lebensumstände und als solche reflektieren sie selbstverständlich die objektive Lebensqualität des Individuums – jedoch, so ließe sich im Sinne des Dienerschen Resümees hinzufügen, nur in einem Ausmaß, dessen Grenzen z.B. durch die selektiven Aufmerksamkeiten, Gewichtungen und Bedeutungszuweisungen gesteckt sind, welche die objektive Information im Zuge des subjektiven Bewertungsprozesses durchläuft.

1.1.4.4 Persönlichkeitseinflüsse auf das SWB

Eine weitere Erklärung der besagten niedrigen Zusammenhänge zwischen Lebenszufriedenheit und objektiver Lebensqualität liefert ein *Trait-Ansatz*: SWB bzw. seine verschie-

denen Komponenten, so wurde postuliert (z.B. Costa & McCrae, 1980; 1984; Costa et al., 1983; 1987a; vgl. auch die Reviews von DeNeve & Cooper, 1998; Diener & Lucas, 1999; Diener et al., 2003) beruht demnach teilweise auf zeitlich stabilen Persönlichkeitsmerkmalen. Eine gewissermaßen radikale Variante des Trait-Ansatzes bestünde in der Annahme einer *weitestgehenden* Trait-Abhängigkeit und infolgedessen Unabhängigkeit von objektiven Lebensbedingungen: Die allgemeine Lebenszufriedenheit erschiene darin als intraindividuell invariante Bewertungstendenz – sozusagen als positiv verklärter oder negativ getrübler Blick – welche dann durch einen sogenannten Top-Down-Effekt (vgl. Kapitel 1.1.4.7) „von Oben nach Unten“ auch die spezifischen Zufriedenheiten determiniert. Eine soziologische Variante dieser Theorie bezeichnet Veenhoven (1996) als „folklore-theory“: Demnach bestehen kulturelle Dispositionen zu optimistischen bzw. pessimistischen Sichtweisen, die sich als Top-down-Effekte bei der Generierung von Zufriedenheiten auswirken.

Die empirische Befundlage zum Trait-Ansatz wurde von Veenhoven (1994) metaanalytisch untersucht und aufgrund der gefundenen Ergebnisse ablehnend beurteilt. Allerdings begründet Veenhoven diese Ablehnung auf eine strikte Definition: Demnach ist SWB dann als Trait aufzufassen, wenn es (1) zeitlich stabil, (2) transsituational konsistent und (3) internal verursacht ist. Die Trait-Theorie, so seine Argumentation, sei leicht damit zu widerlegen, daß empirisch gezeigt wird, daß SWB weder intraindividuell stabil im Zeitverlauf ist, noch insensitiv gegenüber Veränderungen objektiver Lebensbedingungen. Zur Abschätzung der Stabilität des SWB wurden von Veenhoven alle bis zum damaligen Zeitpunkt vorliegenden Veröffentlichungen, in denen längsschnittliche Korrelationen von Wohlbefindensindikatoren berichtet wurden, berücksichtigt – wovon allerdings nur ein kleinerer Teil von ihm als valide befunden wurde („I found some hundred studies, of which 26 appeared to have used acceptable indicators of happiness“, ebd., 108). Es zeigten sich überwiegend positive längsschnittliche Korrelationen von niederen bis mittlerem Niveau (im Mittel ungefähr um 0.35), wobei erwartungsgemäß diese Korrelation mit wachsendem Zeitintervall zwischen den Wiederholungsmessungen fiel, so daß Veenhoven folgern konnte, daß in der Extrapolation dieses Zusammenhangs bei einer Zeitspanne von 20 Jahren praktisch kein Zusammenhang zwischen den SWB-Werten mehr bestehen würde. Die insgesamt eher bescheidenen Meßwiederholungskorrelationen wertet Veenhoven als eindeutige Widerlegung des Trait-Ansatzes.

Das Veenhovensche Verdikt kann m.E. jedoch auch angezweifelt werden. Denn seine zugrundegelegte Trait-Definition trifft lediglich für eine theoretische Konzeption zu, die SWB selbst als Persönlichkeitskonstrukt auffaßt, nicht dagegen für die Annahme, daß SWB-Komponenten *teilweise* von zeitlich stabilen Persönlichkeitsmerkmalen abhängen und in diesem Sinne persönliche Dispositionen für mehr oder weniger SWB bestehen. Letzteres würde lediglich ein gewisses Maß an Stabilität und Konsistenz und d.h. eine teilweise Unabhängigkeit des SWB von situationalen Einflüssen bedingen, wie sie durch verschiedene längsschnittliche empirische Befunde gestützt wird, welche von Diener et al. (1999) referiert und zu folgenden Fazit zusammengefaßt werden: „Based on the large body of evidence

on the correlates of SWB, Veenhoven's claim that happiness itself is not a trait appears to be correct. However, stable personality traits can influence SWB, and thus SWB has both trait-like and state-like properties" (ebd., 280).

Letztere Formulierung, wie auch die Veenhovensche Metaanalyse, bezieht sich allerdings auf SWB insgesamt, und im Kontext der Darstellung der Theorie der Lebenszufriedenheit wäre hier natürlich v.a. zu fragen, ob alle diese Aussagen für die verschiedenen Wohlbefindenskomponenten in gleicher Weise gelten bzw. wie nun der Trait-Ansatz für die Zufriedenheit zu bewerten ist. In der Untersuchung von Veenhoven (1994) – die einen guten Überblick über die bis zu ihrem Veröffentlichungszeitpunkt vorhandenen Längsschnittstudien zum SWB liefert – bezogen sich beispielsweise nur 6 der einbezogenen 65 Korrelationswerte auf die kognitive Komponente („contentment“) und die Mehrzahl auf Maße der „overall happiness“ oder des affektiven Wohlbefindens. Insgesamt erlauben die vorliegenden Befunde m.E. keine abschließende Bewertung der „Trait-Stabilität“ von Lebenszufriedenheit, da die aus den einzelnen Studien ableitbaren Aussagen zur längsschnittlichen Stabilität meist auf unterschiedlichen Erhebungsintervallen zwischen den Zufriedenheitsmessungen und v.a. auch auf unterschiedlichen Operationalisierungen von SWB – d.h. Meßinstrumenten unterschiedlicher Reliabilität und Validität – begründet werden.

Abgesehen davon ist darauf hinzuweisen, daß für die Erforschung persönlichkeitspezifischer Anteile an Zufriedenheit bzw. SWB der Nachweis der Stabilität allein nicht ausreicht. Oft wird mehr oder weniger explizit die intraindividuelle Stabilität von SWB-Konstrukten mit deren Trait-Abhängigkeit gleichgesetzt bzw. als gänzlich auf Trait-Einflüsse zurückführbar betrachtet (z.B. Ferring & Filipp, 1997). Es lassen sich aber leicht Gründe anführen, um dem zu widersprechen: Zunächst einmal könnte natürlich Stabilität von Zufriedenheit auch schlicht aus entsprechend stabilen objektiven Lebensbedingungen resultieren, so daß immer dann, wenn aus empirischen Zufriedenheitsmessungen errechnete Stabilität als Nachweis von Trait-Einflüssen gewertet wird, zu fragen wäre, inwieweit das Ausmaß der Veränderungen in relevanten objektiven Variablen dabei statistisch berücksichtigt wurde (ein durchaus schwieriges Unterfangen, denn es setzt wiederum entsprechende „objektive“ Messungen voraus, und zwar möglichst zu allen jeweils möglicherweise relevanten Lebensbedingungen). Darüber hinaus kann aber Stabilität auch aus der Relativität von Zufriedenheit begründet werden: Veränderung objektiver Lebensbedingungen führt zu entsprechender Adjustierung subjektiver Standards und in der Folge zur Adaptation der Zufriedenheit in Richtung des Ausgangsniveaus, welches vor der Veränderung bestanden hat. Trait-Einflüsse stellen somit eine zwar hinreichende, jedoch nicht notwendige Bedingung für Stabilität von Zufriedenheit dar. Das Ausmaß intraindividuelle Stabilität von Zufriedenheit ist deshalb ein Forschungsgegenstand per se, der losgelöst von Trait-Theorien des SWB untersucht werden sollte – als solcher wird er auch in den im folgenden dargestellten eigenen empirischen Untersuchungen behandelt, bei denen die Möglichkeit hoher Zufriedenheitsstabilität hypothetisch zugrundegelegt wurde (vgl. die Kapitel 1.2.3.3 und 1.2.3.4).

Empirische Befunde zur längsschnittlichen Stabilität erlauben also kein abschließen-

des Urteil über die relative Stärke persönlichkeitspezifischer Anteile an den Zufriedenheitsurteilen. Darüber hinaus existiert eine Vielzahl von Befunden, die substantielle korrelative Zusammenhänge zwischen verschiedenen Persönlichkeitsindikatoren und Wohlbefindensmaßen zeigen – eine Zusammenfassung liefern DeNeve & Cooper (1998) (neuere Arbeiten: Hayes & Joseph, 2003; Creed & Evans, 2002; Steel & Ones, 2002; Schimmack et al., 2002). Die dabei herausragenden „Kandidaten“ wohlbefindenskorrelierter Persönlichkeitsdimensionen scheinen allerdings eher affektrelevant: „... many researchers have suggested that extraversion and neuroticism provide the primary links between personality and SWB“ (Diener et al., 2003, 407). Tatsächlich kann v.a. ein starker Zusammenhang zwischen Extraversion und positivem Affekt, sowie Neurotizismus und negativem Affekt als mittlerweile empirisch gesichert gelten (DeNeve & Cooper, 1998; Diener et al., 2002) und verschiedene theoretische Entwürfe betrachten Extraversion und Neurotizismus als diejenigen Persönlichkeitsdimensionen, welche maßgeblich das Erleben positiver und negativer Affekte regulieren und somit in hohem Maße relevant für affektives SWB sind. Ein Einfluß dieser Traits auf Zufriedenheitsurteile könnte demnach mittelbar dadurch bestehen, daß in solche Bewertungen eigener objektiver Lebensumstände auch affektives Erleben einfließt (vgl. Schimmack et al., 2002), wie es in dem nachfolgend diskutierten Aspekt der affektiven Inferenz thematisiert wird. In ihrer metaanalytischen Untersuchung fanden DeNeve & Cooper (1998) zudem einen Zusammenhang zwischen Neurotizismus und Lebenszufriedenheit.

Unter den sonstigen Persönlichkeitskonstrukten, deren Wohlbefindensrelevanz diskutiert wurde, sind wenige, deren möglicher direkter Einfluß eher die kognitive Wohlbefindenskomponente betrifft. So wurde von Scheier & Carver (1985) ein Optimismus-Trait postuliert, nämlich als dispositionelle Erfolgsoptimismus, welche verstärkten Einsatz zur Erreichung persönlicher Ziele und damit auch nach Art einer self-fulfilling-prophecy tatsächlich mehr Erfolg bedingt. Demnach sollten optimistischere Personen deshalb zufriedener sein, weil sie eben tatsächlich ihre Lebensumstände vorteilhafter gestalten konnten. In ähnlicher Weise wurden generalisierte Kontrollüberzeugungen zur Erklärung von SWB bzw. Zufriedenheit herangezogen (Lachman & Weaver, 1998; Kozma & Stones, 1978). Allerdings warnen Diener et al. (1999, 281) bezüglich dieser Theorien: „It is difficult to disentangle, however, whether the cognitive processes associated with optimism and expectancy for control are the cause or the result of higher well-being.“ Insgesamt, so läßt sich wohl die Lage zusammenfassen, liefert die vorhandene Theorie zur „Traithaftigkeit“ des SWB gute und deutlich belegte Gründe für die Annahme gewichtiger Trait-Einflüsse auf die affektive Wohlbefindenskomponente, während die Abhängigkeit der kognitiven Komponente von zeitlich stabilen Persönlichkeitsanteilen zumindest weniger theoretisches Interesse gefunden zu haben scheint. Es kann also, nun schon eher im Sinne des Veenhovenschen Fazits, gesagt werden, daß die gegenwärtige SWB-Theorie nur wenige Gründe zur Annahme *direkter* Trait-Einflüsse auf Zufriedenheit liefert. Gleichwohl erscheinen solche nicht ausgeschlossen und v.a. ist auch mit indirekten Trait-Einflüssen zu rechnen, die aus dem Einfluß

des stärker persönlichkeitsabhängigen affektiven Wohlbefindens auf Zufriedenheitsurteile resultieren.

1.1.4.5 Affektive Inferenz

Das einfache Bottom-up-Modell der Zufriedenheiten enthält (zumeist implizit) Annahmen über den psychologischen Verarbeitungsmodus bei der Entstehung von Zufriedenheit – es handelt sich um ein *kognitives* Modell der Informationsverarbeitung, das davon ausgeht, daß zur Bewertung der eigenen Lebensbedingungen die vorhandene Information über dieselben mittels eines Vergleichs von Ist- und Soll-Zustand kognitiv verrechnet wird. Die oben aufgeführten „Relativitätstheorien“ der Zufriedenheit bilden eine weitere Variante eines solchermaßen kognitiven Modells, welche auf der postulierten Instabilität der Soll-Zustände gründet. Eine Alternative zur „kognitiven Sichtweise“ bietet ein affektives Zufriedenheitsmodell: Zufriedenheit kann nicht nur aus Verarbeitung objektiver Information über eigene Lebensqualität, sondern bis zu einem gewissen Grade auch aus dem emotionalen Erleben des Individuums resultieren. Mit anderen und einfacheren Worten: Wenn wir uns wohl *fühlen*, dann sind wir auch zufrieden. Veenhoven (1996) bezeichnet dieses als *affektive Inferenz*.

Es scheint mir an dieser Stelle angebracht, noch einmal die Aufmerksamkeit auf terminologische Fragen zu richten. In Theorie und Forschung zum SWB ist, wie im Bisherigen ausreichend dargestellt sein sollte, eine begriffliche Dichotomie „kognitiv“ versus „affektiv“ durch die Unterscheidung entsprechend benannter Wohlbefindenskomponenten zu besonderer Bedeutung gelangt und es wurde bereits in Kapitel 1.1.3 darauf hingewiesen, daß dieses auch einer gewissen Ungenauigkeit Vorschub leisten kann. Denn die Konzeption einer Komponente *habituellen* affektiven Wohlbefindens impliziert Kognitivität in der Verarbeitung bzw. Bilanzierung affektiven Erlebens zu einem entsprechenden Urteil. Die Bezeichnung dieser Wohlbefindenskomponente als affektiv betrifft also den Inhalt der Informationsverarbeitung. Auch die sogenannte kognitive Komponente ist m.E. durch den inhaltlichen Gegenstand der damit gemeinten Beurteilungen definiert und gegen die affektive abgegrenzt: Nämlich eben durch die objektiven Lebensumstände, auf die sich die Zufriedenheitsurteile beziehen sollen. Deshalb ist hier die Bezeichnung als „kognitiv“ in der Weise mißverständlich, daß dieser Begriff nicht den Inhalt, sondern einen psychologischen Modus benennt, der dieser Komponente nicht exklusiv zukommt. Da sich aber diese beiden Bezeichnungen in der SWB-Forschung eingebürgert haben, sollen sie hier auch weiterhin verwendet werden. Wenn nun jedoch zu alledem noch von affektiver Inferenz die Rede ist, dann könnte das zu weiterer begrifflicher Verwirrung beitragen, und deshalb sei hier darauf hingewiesen, daß dieser Begriff im folgenden ausschließlich auf Zufriedenheitsurteile bezogen wird: Affektive Inferenz meint den Einfluß affektiver Wohlbefindensaspekte auf Zufriedenheitsurteile.

Die theoretische Begründung der affektiven Inferenz beruht auf einer informations-

verarbeitungstheoretischen Sichtweise des Urteilsprozesses, der zur Bildung von Zufriedenheitsurteilen führt. Die Konstruktion eines solchen Urteils wird als komplexe Aufgabe betrachtet, die effiziente Lösungsstrategien erfordert. Affektive Inferenz dient in diesem Zusammenhang als Heuristik, die solche raschen und einfachen Lösungen ermöglicht. Grundlegend für diese Betrachtungsweise war das vielzitierte „judgement model of subjective well-being“ von Schwarz & Strack (1991), welches aufgrund der besonderen Bedeutung, die es für die weitere Theoriebildung zum SWB gewonnen hat, im folgenden Unterkapitel ausführlicher dargestellt wird.

Folgt man nun der Vorstellung, daß globale, unspezifische Lebenszufriedenheit in starkem Maße durch die affektive Befindlichkeit der Person geprägt ist, so bedeutet dieses nicht die Unabhängigkeit dieser Zufriedenheitsurteile von der objektiven Ebene menschlicher Lebensrealität. So benutzt Veenhoven (1996) gerade die Affektivität der allgemeinen Lebenszufriedenheit als Argument für deren Brauchbarkeit als Sozialindikator bei Vergleichen auf der Macroebene, d.h. zwischen Ländern, Kulturen, Regionen u.ä.. Diese Brauchbarkeit nämlich ist in Zweifel gezogen durch den stets in „rein-kognitiven“ Theorien der sozialwissenschaftlichen Zufriedenheitsforschung schlummernden Relativitätsverdacht, daß sich individuelle Bewertungsstandards prinzipiell an dem allgemeinen materiellen Niveau der Gesellschaft ausrichten, so daß sich in jeder Gesellschaft unabhängig von deren absolutem Niveau objektiver Lebensbedingungen eine Nivellierung individueller Zufriedenheiten auf einem mittleren Durchschnittswert ergeben müßte. Demnach sollte also in einem schwach entwickelten Drittweltland mit schlechten materiellen Lebensbedingungen dieselbe durchschnittliche Lebenszufriedenheit beobachtet werden, wie in einem hochentwickelten Industriestaat mit besten sozialen und ökonomischen Konditionen. Dagegen argumentiert Veenhoven (1996, 10, Hervorhebung von OS): „Unlike conscious comparisons between ideal and reality, affects are *largely unreasoned experiences that probably signal the degree to which basic needs are met*. If so, satisfaction ratings reflect something universal, that can be meaningfully compared cross-culturally.“ Der erste dieser beiden Sätze enthält sozusagen eine soziologisch relevante Quintessenz der Theorie affektiver Inferenz: Affektives Erleben wird hier gewissermaßen als Vermittlungsinstanz betrachtet, durch die objektive Lebensqualität subjektiv erlebbar wird – nämlich in der Weise, daß objektiv „gute“ Bedingungen zu besserer Befriedigung elementarer und universeller menschlicher Bedürfnisse („needs“) und deshalb zu einem Mehr an positiv-affektiver Erfahrung führen.

1.1.4.6 Das „judgement model of subjective well-being“ von Schwarz & Strack

Schwarz & Strack (1991) berichten von einer ganzen Serie sozialpsychologischer Experimente, in denen durch die experimentelle Manipulationen sehr deutliche Effekte auf die Zufriedenheitsurteile der Probanden bewirkt werden konnten. Die besondere Relevanz dieser Ergebnisse besteht jedoch in den ihnen zugrundeliegenden theoretischen Leitmotiven, welche als Bausteine einer Theorie der Zufriedenheitsbewertung, die auch widersprüchli-

che empirische Befunde erklären kann, dienen können.

Die Basishypothese besteht darin, daß die Generierung von Zufriedenheitsurteilen einer Art von Sparsamkeitsprinzip folgt, indem nämlich stets die einfachstmögliche Strategie gewählt wird, die zu einem hinreichend sicheren Urteil führt. Ein wesentlicher Aspekt dieser „Einfachststrategie“ besteht darin, daß zur Urteilsbildung auf leicht verfügbare Information zugegriffen wird, die allerdings auch informellen Wert („informal value“) besitzen, d.h. als relevant für die Bewertungsaufgabe eingestuft werden muß. Die Verfügbarkeit von Information wächst mit der Häufigkeit und/oder dem geringen zeitlichen Abstand ihrer vorherigen Verwendung, sie ist am größten dann, wenn die Information in der Beurteilungssituation selbst bereits aktualisiert ist, z.B. dadurch, daß die Art der Zufriedenheitsbefragung bereits Hinweisreize auf diese Information enthält.

Aus der Annahme einer Einfachststrategie zur Bildung von Zufriedenheitsurteilen ist die Hypothese affektiver Inferenzen ableitbar, v.a. für Urteile allgemeiner, unspezifischer Lebenszufriedenheit: Im diesem Falle nämlich wäre die kognitive Verrechnung aller relevanten objektiven Information über die eigene Lebensrealität – viele Feststellungen aktueller Ist-Zustände, Adjustierungen von Soll-Zuständen und Vergleiche der beiden miteinander, danach Gewichtung jedes dieser Vergleiche und Verrechnung aller Ergebnisse zu einem Gesamturteil – eine komplizierte und ineffektive Strategie zur Generierung des Urteils, viel einfacher ist die affektive Inferenz vom jeweils aktuellen Affektzustand, welcher gewissermaßen jederzeit psychisch real und somit unmittelbar verfügbar ist. Somit wäre zu vermuten, daß zur Bewertung allgemeiner Lebenszufriedenheit in starkem Maße affektive Inferenzen eingesetzt werden. Ist jedoch in der Urteilssituation der informelle Wert des Affektzustandes in Frage gestellt, oder sind relevante objektive Informationen leicht verfügbar, so werden eher „kognitive“ Inferenzstrategien gewählt. Für die Beurteilung spezifischer Bereichszufriedenheiten wäre dementsprechend eine geringere Wahrscheinlichkeit affektiver Inferenzen anzunehmen: „Evaluations of specific life domains, on the other hand, are often less complex. In contrast to judgements of general life satisfaction, information is usually available for judgements of specific life domains and comparison criteria for evaluation are well-defined. . . . Moreover, one's affective state is not considered relevant information in evaluating many domains“ (Schwarz & Strack, 1991, 39).

Aus dieser informationsverarbeitungstheoretischen Perspektive folgt auch, daß Zufriedenheitsurteile gezielt beeinflußt werden können, indem beispielsweise durch entsprechende Manipulationen der Urteilssituation die Aktivierung, Akzentuierung oder auch Diskreditierung bestimmter Informationen bewirkt wird. Dazu hier nur zwei Beispiele aus der Vielzahl entsprechender Befunde: Es konnte experimentell gezeigt werden, daß Probanden, die vor dem Urteil über die allgemeine Lebenszufriedenheit gebeten wurden, drei besonders positive bzw. (in einer weiteren Experimentalbedingung) negative Ereignisse aus ihrer jüngsten Vergangenheit aufzuschreiben, gegenüber einer Kontrollgruppe ohne entsprechende Instruktion im Durchschnitt erhöhte bzw. verminderte Zufriedenheiten berichten (Strack et al., 1985). Aus Sicht der Autoren wurde von den Probanden bei der Generierung des

Zufriedenheitsurteils auf eben die Information zurückgriffen, die durch die vorangegangene Beschreibungsaufgabe ins Gedächtnis gerufen und dadurch leicht zugänglich wurde. Die Beeinflussungsrichtung kehrt sich jedoch um, wenn die Instruktion lautet, zeitlich weit zurückliegende positive bzw. negative Ereignisse zu schildern – die Beschreibung positiver Ereignisse bewirkt dann verminderte Zufriedenheiten und umgekehrt. Wiederum wurde hier – so die Erklärung der Autoren – die zuvor „aktualisierte“ Information verwendet, welche nun jedoch infolge der zeitlichen Distanz nicht mehr als repräsentativ für die momentane Lebenssituation betrachtet werden konnte, sondern stattdessen zur Adjustierung eines idealen Soll-Zustandes nach oben bzw. unten und damit zu den gefundenen Zufriedenheitsumkehrungen führte.² Einen geradezu plakativen Beleg für den Einfluß der momentanen Stimmung auf Urteile der allgemeinen Lebenszufriedenheit liefert der folgende Befund (Schwarz & Clore, 1983): Bei einer Probandengruppe, die bei sonnigem Wetter telefonisch um ein Zufriedenheitsurteil gebeten wurde, waren die Zufriedenheiten signifikant erhöht gegenüber einer Gruppe, die bei Regenwetter befragt wurde. Interessanterweise verschwand dieser Effekt, wenn der Interviewer vorgab, von außerhalb anzurufen und die Probanden vor dem Zufriedenheitsrating fragte, wie bei ihnen das Wetter sei. Mit dieser Manipulation sollte den Probanden sozusagen das momentane Wetter als Ursache ihrer Stimmung ins Bewußtsein gerufen und dadurch die Relevanz der Stimmung für das Urteil der allgemeinen Lebenszufriedenheit unterminiert werden.

Das Urteilsmodell von Schwarz & Strack (1991) hat, wie bereits erwähnt, großen Einfluß auf die weitere Theoriebildung genommen, der Verweis auf diese Veröffentlichung scheint in kaum einer neueren Veröffentlichung, in der die psychologische Theorie der Generierung von Lebenszufriedenheitsurteilen thematisiert ist, zu fehlen. Angesichts dieser Bedeutung des Modells erscheint es geboten, auch ihre möglicherweise kritisierbaren Aspekte eingehender zu betrachten.

Denn insgesamt hinterläßt die von Schwarz & Strack (1991) präsentierte Fülle von Effekten der Stimmung auf Zufriedenheitsurteile den Eindruck, es handle sich bei der allgemeinen Lebenszufriedenheit um ein derart flüchtiges, ganz und gar von variierenden kontextualen Bedingungen abhängiges Konstrukt, daß es gar nicht mehr reliabel erfaßbar ist. Welchen Sinn hat es, dieses Konstrukt überhaupt noch zu verwenden, wenn Zufriedenheitsäußerungen beispielsweise vom Wetter und von zufälligen situationalen Hinweisreizen zur Selektion von Information dominiert werden? Zur Beantwortung dieser Frage muß auf einige Schwächen dieser Arbeit hingewiesen werden, welche m.E. zunächst einmal – trivialerweise – in der mangelnden externen Validität der experimentellen Befunde zu vermuten sind. Die Quintessenz des psychologischen Experiments besteht ja in der Maximierung interner Validität zuungunsten der externen „Situations- und Populationsvalidität“ (Hager & Westermann, 1983), die möglichst optimale Isolierung einzelner Effekte führt gewisser-

²Man kann gegen die von den Autoren vorgebrachten Erklärungen dieser Effekte allerdings einwenden, daß zumindest die Frage nach noch nicht weit zurückliegenden „positiven“ bzw. „negativen“ Erlebnissen eher erlebte Affekte aktualisiert, als kognitive Vergleichsprozesse.

maßen auch zu deren Überakzentuierung und wirft die Frage nach deren relativer Bedeutung in realweltlichen Lebenskontexten auf. Diese aber ist in der Arbeit von Schwarz & Strack (1991) noch nicht beantwortet, auch wenn es angesichts der dargestellten Ergebnisse schwerfiele, die Existenz derartiger Effekte prinzipiell in Zweifel zu ziehen.

Im Grunde widerspricht die – möglicherweise unbeabsichtigte – Überakzentuierung zeitlich fluktuativer, kontextueller Einflüsse auf den Beurteilungsprozeß dem postulierten Prinzip des einfachstmöglichen Urteils: Eine besonders leichte Strategie könnte doch darin bestehen, einfach auf bereits gefällte Urteile zum SWB zurückzugreifen. Denn wieso sollte man annehmen, daß solch eine Beurteilung in jedem Augenblick gänzlich neu vorgenommen werden muß? Zieht man in Betracht, daß die subjektive Bewertung des eigenen Lebens von existentieller Bedeutung für den Menschen – für die Aufrechterhaltung von Identität und Selbstwert, für die notwendige Selbstverortung im sozialen Gefüge, für das Erkennen möglicher Gefährdungen usw. – ist, so scheint die Annahme zwingend, daß Individuen solche Bewertungen fortgesetzt durchführen und folglich dann, wenn sie um ein explizites Urteil gebeten werden, bereits über ein solches verfügen. Situationale Hinweisreize sollten somit, falls sie weder extreme Auffälligkeit besitzen (wie z.B. die in experimentellen Untersuchungssituationen gezielt manipulierten), noch die bisherige Bewertung grundsätzlich in Frage stellen (z.B. durch die unverhoffte Mitteilung eines hohen Lotteriegewinns), allenfalls moderierende Wirkung haben. Wenn also Menschen stets wissen, wie zufrieden sie ungefähr sind, so wären umfassende Neubewertungen schlicht unnötig und zu umständlich.

Des weiteren könnte man gegen das „judgement model“ einwenden, daß darin affektive Inferenz ausschließlich als Einfluß momentaner Stimmung konzipiert ist: Soll man etwa annehmen, daß das affektive Erleben eines Menschen in der Aneinanderreihung flüchtiger Launen besteht, die jeweils im nächsten Moment wieder gelöscht und aus dem Gedächtnis getilgt werden? Wenn man davon ausgeht, daß Personen Urteile über *habituelles* affektives Wohlbefinden bilden, so erscheint es nicht konsequent, das affektive Erleben des Individuums ausschließlich auf den Aspekt der momentanen Stimmung zu reduzieren.

Veenhoven (1996, 34) formuliert unter Bezug auf die Theorie von Schwarz & Strack: „A typical heuristic seems to involve departing from the mood of the moment, which can be read quite vividly, and next considering how representative that mood is for general affective experience.“ Diese Sichtweise, welche mir in der Originalarbeit von Schwarz & Strack (1991) nicht unbedingt erkennbar scheint, wendet gewissermaßen das Prinzip des Zugriffs auf die einfachst verfügbare Information auf die Generierung des Urteils über habituelles affektives Wohlbefinden an, es ist dabei allerdings nicht explizit gesagt, daß das einmal konstruierte affektive Wohlbefinden dann wiederum zur Generierung des Zufriedenheitsurteils verwendet wird – genau dieses aber scheint Veenhoven zu meinen, wenn man die im vorherigen Unterkapitel 1.1.4.5 dargelegte Richtung der Argumentation, die er in seinem Artikel hinsichtlich der affektiven Inferenz vertritt, zugrundelegt und nicht mißinterpretiert. Wenn aber die momentane Stimmung die einfachst zugängliche Information ist, wieso sollte dann bei der Generierung des Zufriedenheitsurteils nicht direkt auf diese zugegriffen

werden, sondern sie nur indirekt über den „Umweg“ des habituellen affektiven Wohlbefindens einfließen? Eine Antwort im Sinne von Schwarz & Strack könnte in der Berücksichtigung des informellen Werts dieser Informationen liegen: Die Annahme scheint plausibel, daß der habituelle Aspekt des affektiven Erlebens größeren Informationsgehalt zur Bildung von Zufriedenheitsurteilen besitzt, als die bloß augenblickliche Stimmungslage. Allerdings beinhaltet diese Sichtweise die Vorstellung, daß zur einfachstmöglichen Bildung eines Zufriedenheitsurteils auf ein Urteil über habituelles Wohlbefinden zugegriffen wird, welches ja auch erst generiert oder zumindest wieder abgerufen werden muß, und daß es sich somit bei dieser „vorgeschalteten“ Generierung bzw. diesem Abruf immer noch um einen weniger komplexen Informationsverarbeitungsprozeß handelt, als bei der kognitiven Verrechnung objektiver Information. Dieses könnte tatsächlich aus der oben dargestellten Affektfrequenztheorie begründet werden, gemäß der ja habituelles affektives Wohlbefinden auf einfachste Weise ausschließlich aufgrund leicht verfügbarer Häufigkeitsinformation entsteht.

Es scheint diesbezüglich aber auch eine Sichtweise möglich, die im Rahmen des von Schwarz & Strack vorgelegten Modells Selbsteinschätzungen des habituellen affektiven Wohlbefindens als mögliche Quelle affektiver Inferenzen berücksichtigt, ohne daß dieser Theorierahmen noch durch die explizite Zusatzannahme eines „Umwegs“ in der Verarbeitung momentaner Stimmungsinformation ergänzt werden müßte: Der Stimmungsbegriff kann auch so verstanden werden, daß damit sozusagen die jeweils momentan aktuelle Selbstwahrnehmung eigenen Affekterlebens gemeint ist, welche dann jeweils mehr oder weniger stark von dem über einen längeren Zeitabschnitt hinweg aggregierten habituellen Affekterleben abweichen kann. Möglicherweise können aktuell durchlebte spezifische Emotionen hoher Intensität, die nur durch bestimmte situationale Reizkonfigurationen induziert werden, das momentane Affekterleben gänzlich dominieren, während in Situationen „geringer Emotionalität“ die Person Selbstauskünfte bzw. Selbstwahrnehmungen des eigenen Affekterlebens eher auf dem habituellen Aspekt begründet. Dieses würde dann bedeuten, daß z.B. eine Person am Ende eines „guten Tages“ auch bei geringfügiger momentaner Verstimmung noch durchaus zu einer insgesamt positiven Einschätzung ihrer Befindlichkeit gelangen kann, wogegen eine unverhoffte intensive Verstimmung den Blick auf die vorherigen positiven Erlebnisse trübt bzw. soviel Selbstaufmerksamkeit beansprucht, daß vorherige positive Affekte für den Augenblick nicht mehr wahrgenommen werden. „Stimmung“ könnte somit also auch habituelles affektives Wohlbefinden enthalten.

Wie auch immer man den Stimmungsbegriff konzeptualisiert, der darin enthaltene Fokus darf nicht mit affektiver Instabilität bzw. mit den instabilen Anteilen an den Affektreaktionen eines Individuums verwechselt werden — dieses ist mit Blick auf die oben umrissene trait-orientierte Thematisierung intraindividuelle Stabilität gerade der affektiven Wohlbefindenskomponente ausdrücklich hervorzuheben. In dem Maße, in dem Affektreaktionen auf zeitlich invarianten Persönlichkeitskomponenten beruhen, wird sich in diesen auch Stabilität zeigen und werden folglich auch die momentanen Stimmungen einer Person relativ stabil sein.

Ungeachtet der konzeptuellen Fragen hinsichtlich des inhaltlichen Umfangs der von Schwarz & Strack (1991) gemeinten Stimmungseinflüsse besitzt deren Modell seine besondere Relevanz als Fundament für eine Theorie affektiver Inferenz in Zufriedenheitsurteilen und hat als solches, wie bereits gesagt, auch großen Einfluß auf die weitere Zufriedenheitsforschung genommen. Dieser besteht zum einen in dem, was oben als als Prinzip des einfachstmöglichen Urteils bezeichnet wurde, und zum anderen in der daraus abgeleiteten Annahme, daß Urteile zur allgemeinen Lebenszufriedenheit v.a. aufgrund affektiver Inferenzen gebildet werden, während Bereichszufriedenheiten eher durch kognitive Verrechnung von Information über objektive Lebensumstände gebildet werden dürften. Ganz allgemein läßt sich aus dem Modell von Schwarz & Strack ableiten, daß bei der Beurteilung von (bereichsspezifischer oder allgemeiner) Zufriedenheit sowohl affektive Inferenzen, als auch kognitive Verrechnungen objektiver Informationen beteiligt sind und daß sich das Mischungsverhältnis der beiden Elemente aus der jeweiligen Verfügbarkeit und dem informellen Wert affektiver bzw. objektiver Vergleichsinformation ergibt (vgl. diesbezüglich auch: Fox & Kahneman, 1991).

1.1.4.7 Die Top-down/Bottom-up-Kontroverse

In den bislang aufgeführten Gesichtspunkten – v.a. der Relativität (Adjustierung subjektiver Standards), der Trait-Abhängigkeit und der affektiven Inferenz – sind wesentliche Entwicklungslinien umrissen, die die Zufriedenheitsforschung seit den achtziger Jahren beschritten hat. Wie bereits dargelegt gab den Anstoß hierzu der mangelhafte korrelative Zusammenhang zwischen Indikatoren objektiver Lebensqualität und Zufriedenheitswerten, der in entsprechenden empirischen Daten immer wieder gefunden wurde, und somit könnten die behandelten Gesichtspunkte gewissermaßen als theoretische Standpunkte in der Debatte über Ausmaß und Richtung kausaler Zusammenhänge zwischen objektiver und subjektiver Lebensqualität angesehen werden.

Damit aber stehen sie in engem Bezug zu einer weiteren für die Theorieentwicklung maßgeblichen Fragestellung, die in der SWB-Forschung unter dem Begriff der *Top-down/Bottom-up-Kontroverse* behandelt wird: „In the top-down approach to happiness, global features of personality are thought to influence the way a person reacts to events. . . . In the bottom-up approach, a person should develop a sunny disposition and sanguine outlook as positive experiences accumulate in the person’s life“ (Diener, 1984, 565). Diener bezieht also die beiden Begriffe auf die Frage nach zeitlich stabilen Persönlichkeits- versus zeitlich instabilen situationalen Einflüssen auf SWB. Headey et al. (1991, 81) definieren dagegen die Begriffe wie folgt: „Bottom-up causation is where particular variables cause SWB and top-down causation is where SWB produces certain outcomes.“ Diese Formulierung weist den Begriffen ein sehr breites Bedeutungsspektrum zu: „top-down“ meint ganz allgemein die Verursachung von SWB, während „bottom-up“ sich auf die Wirkung des SWB bezieht. Die Unterschiedlichkeit in der Konzeptualisierung der beiden Begriffe

in den zitierten Arbeiten wird vollends daraus deutlich, daß Headey et al. (1991) in ihrem statistischen Modell die Top-down-Effekte der Lebenszufriedenheit unter expliziter Auspartialisierung möglicher Einflüsse von Persönlichkeitsvariablen schätzen. Es ließen sich noch weitere wörtliche Zitate aufführen, die wiederum andere Bedeutungsnuancen der beiden Begriffe aufzeigen, so daß hier wieder einmal Gelegenheit bestünde, begriffliche Uneindeutigkeiten in der SWB-Literatur zu beklagen. Tatsächlich jedoch wurde die Top-down/Bottom-up-Kontroverse meist in diesbezüglichen Veröffentlichungen (z.B. Headey et al., 1991; Lance et al., 1989; 1995; Mallard et al., 1997; Scherpenzeel & Saris, 1996) als Frage nach der kausalen Richtung der Zusammenhänge zwischen spezifischen Bereichszufriedenheiten und allgemeiner Lebenszufriedenheit geführt:³ Wird allgemeine Lebenszufriedenheit „bottom-up“ aus der Kombination einzelner subjektiver Bewertungen spezieller Bereiche objektiver Lebensqualität (Bereichszufriedenheiten) errechnet, oder entsteht das Lebenszufriedenheitsurteil weitgehend unabhängig von solchen Bottom-up-Einflüssen aufgrund anderer psychischer Ursachen und beeinflußt umgekehrt die Bewertungen spezifischer objektiver Lebensbereiche?

Ein sozusagen gedanklicher Grundbaustein, der in dieser Kontroverse mehr oder weniger implizit verarbeitet ist, besteht in der Vorstellung einer größeren Realitätsnähe der Bereichszufriedenheiten gegenüber allgemeiner Lebenszufriedenheit: Da der inhaltliche Fokus spezifischer Bereichszufriedenheiten einen reduzierten und leichter überschaubaren Ausschnitt der gesamten „objektiven“ Lebenswirklichkeit der Person umfaßt, und solch ein spezifisches Urteil einfacher zu fällen sein sollte, weil weniger und hinsichtlich ihrer Relevanz leichter einzuschätzende Informationen verarbeitet werden müssen, wird angenommen, daß Bereichszufriedenheiten gewissermaßen „objektiver“ sind in dem Sinne, daß sie tatsächlich die objektiven Verhältnisse in ihrem thematischen Bereich gut reflektieren. Diese Annahme läßt sich, wie bereits dargelegt, z.B. aus der Informationsverarbeitungstheorie von Schwarz & Strack (1991) begründen.

Dagegen ist angesichts des uneingeschränkten inhaltlichen Fokus allgemeiner Lebenszufriedenheit prinzipiell davon auszugehen, daß spezifische objektive Informationen, die für eine bestimmte Bereichszufriedenheit besonders relevant sein mögen, für dieses globale Urteil geringeres Gewicht besitzen, so daß geringere Zusammenhänge zwischen Lebenszufriedenheit und einzelnen spezifischen Indikatoren objektiver Lebensqualität von vornherein zu erwarten sind – die Frage ist, ob man entsprechende empirische Ergebnisse allein dadurch erklären möchte (bzw. kann). Die dargestellten Trait-Theorien liefern dazu ebenso eine Alternativerklärung, wie die aus der Theorie von Schwarz & Strack (1991) gefolgerte Annahme weitgehender affektiver Inferenz in Urteilen zur allgemeinen Lebenszufriedenheit. Wenn man aber die Lebenszufriedenheit als ein v.a. dispositionelle Faktoren und/oder

³Eine Ausnahme bilden Feist et al. (1995), die unter dem Label der Top-down/Bottom-up-Kontroverse die kausale Richtung zwischen globalem, über affektive und kognitive Komponenten aggregiertem SWB einerseits und verschiedenen psychologischen Konstrukten (subjektive Gesundheit, „daily hassles“, „world assumptions“, „constructive thinking“) andererseits untersuchen.

affektives Erleben reflektierendes Urteil betrachtet, so kann man konsequenterweise auch annehmen, daß dieses Urteil spezifische subjektive Bewertungen objektiver Lebensqualität eher beeinflusst, als daß es von solchen beeinflusst würde. Die Top-down-Hypothese behauptet demnach, daß Zufriedenheiten mehr (Lebenszufriedenheit) oder weniger (Bereichszufriedenheiten) unbeeinflusst von objektiver Lebensqualität entstehen, während dagegen die Bottom-up-Theorie die gewissermaßen unverfälschte Konzeption der Zufriedenheiten als „kognitiven“ Bewertungen objektiver Lebensqualität bedeutet.

Nun zeigen die in den vorherigen Unterkapiteln umrissenen forschungstheoretischen Entwicklungslinien und die jeweiligen empirischen Belege, die dafür geliefert wurden, daß bei der Generierung von Zufriedenheitsurteilen verschiedene Inferenzstrategien, die unterschiedliche Informationsquellen „bottom-up“ oder „top-down“ nutzen, beteiligt sein können, und somit gewissermaßen eine theoretische Komplexität dieser Urteilsbildung. Hinsichtlich der Frage der kausalen Richtung zwischen Bereichszufriedenheiten und allgemeiner Lebenszufriedenheit wurde dieser Komplexität zunächst dadurch Rechnung getragen, daß die prinzipielle Reziprozität der kausalen Beeinflussung und diesbezügliche Unterschiedlichkeit bei verschiedenen Bereichszufriedenheiten angenommen wurde. Demnach wäre also davon auszugehen, daß wechselseitige kausale Einflüsse zwischen Bereichszufriedenheiten und Lebenszufriedenheit möglich sind und daß bei manchen Bereichszufriedenheiten der Bottom-up-Einfluß dominiert, während andere v.a. „top-down“ von der Lebenszufriedenheit beeinflusst werden. Denn in verschiedenen empirischen Studien (Headey et al., 1991; Lance et al., 1989; 1995; Mallard et al., 1997) waren für den Zusammenhang von Lebenszufriedenheit und verschiedenen Bereichszufriedenheiten unterschiedlich starke Top-down- bzw. Bottom-up-Effekte geschätzt worden, woraus gefolgert wurde, daß die relative Bedeutung der Top-down- versus Bottom-up-Inferenzen über die verschiedenen Bereiche hinweg variiert, daß also somit für manche Lebensbereiche die entsprechende Zufriedenheit v.a. unter Verarbeitung objektiver Information gebildet wird, während bei anderen Bereichen solche objektive Information kaum verwendet und die spezielle Zufriedenheit eher top-down aus der allgemeinen gefolgert wird.

Alle diese Untersuchungen sind m.E. mit einem schwerwiegenden methodischen Problem belastet, denn es geht dabei um die Aufdeckung *kausaler* Beziehungen, die in der einen oder anderen Richtung zwischen spezifischer und allgemeiner Zufriedenheit wirken. Der empirische Nachweis von Kausalität aber stellt ein Grundproblem erfahrungswissenschaftlicher Forschungsmethodologie dar, denn er muß letztendlich lediglich aus der Beobachtung statistischer Assoziation zwischen Variablen begründbar sein, welche zudem noch, zumindest im Falle der Sozial- und Verhaltenswissenschaften, in der Regel eine unvollständige ist. In den Untersuchungen zur Top-down/Bottom-up-Kontroverse wurden typischerweise Strukturgleichungsmodelle berechnet, in denen die beiden fraglichen Effekte als jeweils gerichtete reziproke Pfadkoeffizienten zwischen Bereichszufriedenheiten und Lebenszufriedenheit modelliert und aufgrund ihres geschätzten Werts und statistischer Signifikanz hinsichtlich ihrer jeweiligen Bedeutung beurteilt wurden. Dieser methodische

Ansatz scheint zwar durchaus angemessen zur statistischen Modellierung kausaler Ursache-Wirkungs-Gefüge, jedoch liefert er – wie beispielsweise Bollen (1989) in seinem Standardwerk zur Strukturgleichungsmethodologie sehr ausführlich darlegt – keine „verifizierenden“ Beweise für Kausalität.

Angesichts der Schwierigkeiten, kausale Beziehungen zwischen Bereichszufriedenheiten einerseits und allgemeiner Lebenszufriedenheit andererseits überhaupt eindeutig empirisch zu bestätigen, erscheint die Frage angebracht, ob sich die entsprechende Forschung überhaupt auf dem richtigen Weg befindet, wenn sie die Top-down/Bottom-up-Kontroverse in dieser Weise als Problem kausaler Interdependenzen behandelt. Der eigentliche theoretische Kern der Kontroverse besteht in der Frage, inwieweit bei der Bildung von Zufriedenheitsurteilen momentane objektive Lebensbedingungen „bottom-up“ verrechnet werden bzw. in welchem Ausmaß Top-down-Effekte – als welche zeitlich-stabile Persönlichkeitsmerkmale ebenso in Frage kämen, wie affektive Inferenzen, – Zufriedenheitsurteile bestimmen. Daß die Generierung einzelner Bereichszufriedenheiten gewissermaßen den Zwischenschritt bei der Wirkung von Bottom-up-Einflüssen auf allgemeine Lebenszufriedenheit bildet, ist ebensowenig eine zwingende Annahme, wie umgekehrt die, daß Top-down-Effekte bei der Bildung von Bereichszufriedenheiten nur dadurch wirken, daß auf ein Urteil zur allgemeinen Lebenszufriedenheit zurückgegriffen wird, welches zuvor z.B. aufgrund stabiler Persönlichkeitsmerkmale gefällt wurde. Es wird durch diese Sichtweise ein grundsätzlicher konzeptioneller Unterschied zwischen Bereichs- und Lebenszufriedenheit gesetzt, der m.E. durchaus in Zweifel gezogen werden kann: Es ließe sich einwenden, daß es sich dabei nicht um Grundverschiedenes, sondern um Urteile von derselben Art und Weise handelt. Dieses zeigt sich auch in dem Umstand, daß die in den verschiedenen Forschungsarbeiten erfragten spezifischen Bereichszufriedenheiten sich wiederum teils sehr stark im inhaltlichen Umfang des thematisch zu bewertenden Bereichs unterscheiden. Beispielsweise wurde im deutschen sozio-oekonomischen Panel in den meisten Befragungsjahren eine eher allgemeine Zufriedenheit mit der Freizeit erfragt, während in den Jahren 1990 und 1995 viel spezifischer nach „Zufriedenheit mit dem Umfang der Freizeit“ und „Zufriedenheit mit der Freizeitverwendung“ gefragt wurde (vgl. Kapitel 3.1.2).

Im Grunde stellt die Frage nach allgemeiner Lebenszufriedenheit lediglich den umfanglichsten inhaltlichen Fokus von Zufriedenheit dar, der in sozial- und verhaltenswissenschaftlicher Forschung erfragt wird, und bildet insofern gewissermaßen den Endpunkt einer Dimension inhaltlicher Spezifität der Zufriedenheitsbeurteilung, die in die entgegengesetzte Richtung in beinahe beliebig eng umschriebene Bereiche der Bewertung ausgedehnt werden kann. Dieses bedeutet, daß im Prinzip die Aufgabencharakteristik bei der Bildung von Zufriedenheitsurteilen stets dieselbe ist, auch wenn sich die Aufgaben dieses Typus hinsichtlich des jeweiligen inhaltlichen Bewertungsfokus unterscheiden, was – z.B. im Sinne der Theorie von Schwarz & Strack (1991) – zu Unterschieden in der Verwendung möglicher Lösungsstrategien führt, welche jedoch prinzipiell für jegliche Zufriedenheitsaufgabe zur Verfügung stehen und eingesetzt werden können. Top-down- oder Bottom-up-Inferenzen

wären in diesem Sinne als unterschiedliche Lösungsstrategien bzw. Heuristiken anzusehen und die Top-down/Bottom-up-Kontroverse auf die Frage auszurichten, in welchem Maße das jeweilige Zufriedenheitsurteil top-down versus bottom-up gebildet wird.

Explizit wurde diese Sichtweise von Fox & Kahneman (1992) vertreten: In den Korrelationen zwischen verschiedenen Zufriedenheiten zeigt sich nicht kausale Abhängigkeit, sondern die Verwendung derselben Heuristik zur Generierung des jeweiligen Zufriedenheitsurteils. Einen dementsprechenden, als „konstruktionistisch“ bezeichneten Ansatz in der Top-down/Bottom-up-Kontroverse verfolgen Leonardi et al. (1999, 190): „In fact, some heuristics used by people involve top-down processes because they derive satisfaction judgements from more general and superordinate aspects, whereas other heuristics involve bottom-up processes because they derive satisfaction judgements from more specific and subordinate aspects.“ Der Bottom-up-Begriff meint hier also den Rückgriff auf die für das jeweilige Zufriedenheitsurteil spezifische relevante Information über konkrete objektive Sachverhalte im zu beurteilenden Lebensbereich, während top-down bedeutet, daß Zufriedenheit aus Gesichtspunkten abgeleitet wird, die insofern „übergeordnet“ sind, als ihre Zufriedenheitsrelevanz gewissermaßen für die urteilende Person allgemeingültig und nicht auf einen thematisch begrenzten Zufriedenheitsbereich beschränkt ist. Konkret wird von den Autoren das affektive Erleben als solchermaßen übergeordneter Aspekt benannt, so daß die in obigem Zitat angesprochenen Top-down-Prozesse schlicht als affektive Inferenzen identifiziert werden können. Auf der anderen Seite sind hier diejenigen Mechanismen, die oben unter dem Stichwort der Relativität von Zufriedenheit behandelt wurden, den Bottom-up-Prozessen zuzuordnen: Die Adjustierung subjektiver Standards für bestimmte Merkmale objektiver Lebensqualität, wie z.B. Einkommen und Gesundheitszustand, betrifft eindeutig die Verrechnung spezifischer objektiver Information zu einem Zufriedenheitsurteil.

Zuletzt könnte allerdings auch die Notwendigkeit einer Top-down/Bottom-up-Kontroverse überhaupt in Zweifel gezogen werden. Denn eigentlich geht es dabei einzig darum, das Zustandekommen von Zufriedenheit bzw. insbesondere die mehr oder weniger ausgeprägten Zusammenhänge zwischen objektiver und subjektiver Lebensqualität zu erklären, und man könnte fragen, ob dazu eine Top-down/Bottom-up-Terminologie überhaupt vonnöten und hilfreich ist. Adaptive Mechanismen, affektive Inferenz und die Berücksichtigung von Trait-Einflüssen erscheinen als wesentliche Beiträge zur Erklärung von Zufriedenheit, die aus der diesbezüglichen Forschung resultieren – die Top-down/Bottom-up-Kontroverse, wie auch immer sie konzipiert wurde, greift auf diese Erklärungsansätze zurück und geht nur insofern darüber hinaus, als sie im Grunde nur die relative Bedeutung dieser Ansätze für die verschiedenen bereichsspezifischen Zufriedenheiten und die allgemeine Lebenszufriedenheit thematisiert. Denn letzten Endes beruhen alle gegebenen Antworten auf die Frage, ob, warum und in welcher Richtung vorhandene Korrelationen zwischen Bereichszufriedenheiten und Lebenszufriedenheit kausal interpretierbar sind, auf Annahmen zur unterschiedlichen Gültigkeit dieser Erklärungsmöglichkeiten für die verschiedenen Zufriedenheiten und es sind diese Annahmen, in denen sich eine differenzierte

theoretische Sicht auf Zufriedenheit realisiert.

Die wissenschaftliche Relevanz der Frage nach Top-down- versus Bottom-up-Effekten könnte somit lediglich darin gesehen werden, daß die Top-down/Bottom-up-Terminologie in der Wohlbefindensliteratur nun einmal in sehr starkem Maße verwendet wird und die besagte Kontroverse eine große Anzahl diesbezüglicher sozialwissenschaftlicher Untersuchungen hervorgebracht hat. Die Top-down/Bottom-up-Kontroverse erscheint deshalb als Forschungsgegenstand, der in einer Zusammenfassung des theoretischen Stands in der Wohlbefindensforschung einfach nicht unerwähnt bleiben darf – und v.a. darum hier in einem eigenen Unterkapitel dargestellt wurde. Der m.E. einzig erkennbare Gewinn aus dieser Kontroverse scheint in der „konstruktionistischen“ Auflösung zu bestehen, nämlich in der darin akzentuierten Aufhebung eines grundsätzlichen konzeptuellen Unterschieds von spezifischen Bereichszufriedenheiten auf der einen und allgemeiner Lebenszufriedenheit auf der anderen Seite. Diese Aufhebung ist allerdings bereits in der vielzitierten Theorie von Schwarz & Strack (1994) impliziert, so daß auch diese Sichtweise nicht unbedingt der Top-down/Bottom-up-Kontroverse bedurft hätte.

1.1.4.8 Fazit: Theoretische Standpunkte zum Zufriedenheitsbegriff

Damit sind wichtigste Diskussionslinien der theoretischen Auseinandersetzung mit Zufriedenheit benannt und dargestellt. Daß angesichts der hohen Verzweigtheit und des großen Umfangs des Themenfelds bei dieser Darstellung Schwerpunkte gesetzt werden mußten, versteht sich von selbst, und daß deshalb kein Anspruch auf Vollständigkeit erhoben werden kann, wurde bereits vorne gesagt. Hinzuzufügen ist, daß im Grunde keine der besagten Diskussionslinien bislang am Punkt der endgültigen Klärung angelangt ist und die gezeigten Konzepte nach wie vor diskutiert werden, weshalb es sich gewissermaßen um eine Aufzählung theoretischer Standpunkte zu durchaus noch strittigen Fragen handelt.

Selbstverständlich benötigt Forschung über Zufriedenheit solche theoretischen Standpunkte, von denen aus hypothetische Aussagen abgeleitet werden können, die dann zu überprüfen sind. Mit den folgenden Punkten werden darum die besprochenen Entwicklungen und Ergebnisse der wissenschaftlichen Auseinandersetzung mit Zufriedenheit zusammengefaßt:

- ▶ Der Begriff Zufriedenheit bezeichnet subjektive Bewertungen eigener objektiver Lebensumstände. Informationsverarbeitung zur Generierung derartiger Zufriedenheitsurteile erfolgt v.a. über zwei verschiedenartige Prozesse: In Bottom-up-Prozessen werden Informationen über objektive Sachverhalte in dem zu bewertenden Bereich verrechnet, in Top-down-Prozessen findet affektive Inferenz statt, d.h. Zufriedenheit wird aus dem hedonischen Tonus des jeweiligen Affekterlebens abgeleitet.
- ▶ Die Verrechnung von Information über objektive Sachverhalte in Bottom-up-Prozessen geschieht v.a. auf der Grundlage des Vergleichs von Ist- und Soll-Zustand, d.h. es wird ermittelt, inwieweit objektiv feststellbare Ausprägungen beurteilter Variablen subjektive

Standards erfüllen. Mit anderen Worten: Objektive Lebensqualität wird an subjektiven Anspruchsniveaus relativiert.

- ▶ Die subjektiven Standards bei der Informationsverarbeitung in Bottom-up-Prozessen sind nicht statisch festgelegt, sondern werden adaptiv adjustiert, d.h. sie werden an die für die jeweilige Lebenssituation erkannten Ausprägungsmöglichkeiten (der beurteilten Variablen) angepaßt. Eine wichtige Informationsquelle zur Einschätzung jeweils vorhandener Ausprägungsmöglichkeiten relevanter Variablen objektiver Lebensqualität stellen soziale Vergleiche dar. Es kann deshalb bei Veränderung objektiver Lebensbedingungen eine Adaptation entsprechender Zufriedenheit in der Weise stattfinden, daß auf (überdauernde) Verbesserungen bzw. Verschlechterungen objektiver Lebensbedingungen mit gehobenen bzw. reduzierten diesbezüglichen Standards reagiert wird, woraus die Nivellierung entsprechender Zufriedenheit in Richtung des vor der Veränderung bestehenden Ausgangsniveaus resultiert.
- ▶ Im „Material“ affektiver Inferenz in Top-down-Prozessen, d.h. im jeweiligen Affekterleben der Person, können prinzipiell aktuelle von habituellen Bestandteilen unterschieden werden: Zufriedenheitsurteile könnten demnach sowohl aus den über gewisse Zeiträume hinweg aggregierenden Selbstbeurteilungen des eigenen affektiven Wohlbefindens abgeleitet, als auch durch die momentane Stimmung der Person beeinflußt werden.
- ▶ Zeitlich stabile Persönlichkeitseigenschaften (Traits) beeinflussen Zufriedenheit. Solche Einflüsse dürften v.a. indirekt über affektive Inferenz vermittelt wirksam sein: Eine starke Abhängigkeit des affektiven Wohlbefindens einer Person von stabilen Persönlichkeitsdispositionen (v.a. Extraversion, Neurotizismus) kann als gesichert gelten.

Mit diesen Punkten sollten also grundlegende theoretische Standpunkte zum Begriff der Zufriedenheit aufgelistet sein, auf denen die folgende Untersuchung aufbaut. Insbesondere wurde das Zufriedenheitskonstrukt selbst, welches hier untersucht werden soll, definiert und konzeptionell im Bereich des allgemeineren Themenfelds SWB verortet: Es sollte nun geklärt sein, was Lebenszufriedenheit überhaupt ist bzw. was zumindest der Begriff im hier folgenden bedeutet. Es wurden außerdem grundlegende psychologische „Mechanismen“ dargestellt und diskutiert, die bei der Entstehung von Zufriedenheit beteiligt sein können. Daraus können nun auch Annahmen über Entwicklungstendenzen der Lebenszufriedenheit im höheren Erwachsenenalter abgeleitet werden – genau dieses wird im nächsten Kapitel versucht, an dessen Ende dann solche Annahmen als die der eigenen empirischen Untersuchung zugrundegelegten Hypothesen stehen sollen.

1.2 Zufriedenheit im höheren Lebensalter: Stabilität trotz Verlust?

Nachdem im vorherigen Kapitel allgemeine Aspekte der Theorie von Lebenszufriedenheit abgehandelt wurden, sollen im nun folgenden die grundlegenden Fragestellungen der Un-

tersuchung theoretisch begründet werden: Es geht, um es vorerst noch allgemein zu sagen, um die Entwicklung der Lebenszufriedenheit im höheren Lebensalter – nämlich um die Beantwortung von Fragen, die aus der gerontologischen Wohlbefindensforschung der letzten beiden Dekaden heraus entstanden sind und unter dem im folgenden noch zu erläuternden Begriff des Paradoxes des SWB bzw. der Zufriedenheit zusammengefaßt werden können.

Das Interesse der Gerontologie an Lebenszufriedenheit und SWB wurde v.a. durch die Entstehung von Theorien des „erfolgreichen Alterns“ begründet (vgl. Lehr, 1991), welche Anfang der sechziger Jahre (des zwanzigsten Jahrhunderts) formuliert wurden (Havighurst, 1961) und dann einen langanhaltenden und prägenden Einfluß auf die psycho- und sozialgerontologische Forschung ausübten (zu dieser und anderen theoretischen Entwicklungen in der Gerontologie vgl. Wahl, 2003; Wahl & Heyl, 2003). Es handelt sich dabei um eine soziologisch geprägte Ausrichtung psychogerontologischer Theoriebildung an den typischen Charakteristika der (sozialen, gesundheitlichen) Situation des älteren Menschen. Diese Situation wurde in vielerlei Hinsicht als defizitär eingeschätzt, und im Vordergrund des theoretischen Interesses stand die Anpassung an diese problematischen Bedingungen, die im höheren Lebensalter häufig durchlebt werden müssen. Für die durch dieses inhaltliche Interesse geprägten Theorien jener frühen sechziger Jahre – das sind v.a. die über lange Zeit sehr einflußreichen „Antagonisten“ Aktivitätstheorie (z.B. Tobin & Neugarten, 1961) und Disengagementtheorie (Cumming & Henry, 1961) – beschreibt der Begriff des erfolgreichen Alterns eine gelungene Anpassung und als Kriterium für eine solche bzw. für erfolgreiches Altern wurden die diesbezüglichen subjektiven Einschätzungen der Betroffenen selbst – nämlich deren Lebenszufriedenheit bzw. SWB angesehen.

Zwar wurde – folgt man der Darstellung von Lehr (1991) – im Laufe der weiteren Entwicklung der Theorien des erfolgreichen Alterns die Fokussierung auf subjektive Erfolgskriterien zunehmend als Makel empfunden und durch Vorschläge von objektiven Erfolgsmaßen bereichert (eine Entwicklung, die Lehr zumindest mit einem deutlichen Fragezeichen versieht), doch kann wohl festgehalten werden, daß infolge einer im Sinne dieser Theorien pragmatischen Orientierung gerontologischer Forschung ein besonderes Interesse an den Zufriedenheiten älterer Menschen besteht. Daß es nicht nur aus der Perspektive von Theorien des erfolgreichen Alterns heraus begründet ist, versteht sich von selbst. Eine auf die Entwicklungspsychologie der Lebensspanne ausgerichtete Orientierung psychogerontologischer Theoriebildung (vgl. wieder Wahl, 2003) kann an SWB gerade auch deshalb interessiert sein, weil es als Indikator mehr oder weniger gelungener Umweltsanpassung des Individuums angesehen werden kann und somit z.B. der Verlauf von Zufriedenheiten über die Lebensspanne hinweg in gewisser Weise die Entwicklungsprozesse markiert, die aus der lebenslänglichen Auseinandersetzung des Individuums mit seiner Umwelt resultieren.

Das Zufriedenheitskonstrukt spielt also gerade auch in der Gerontologie eine nicht unbedeutende Rolle und somit bezeichnet die Frage nach möglichen Zusammenhängen zwischen (höherem) Lebensalter und Zufriedenheit einen Forschungsgegenstand von großer gerontologischer Relevanz. Dies betrifft ein breites Spektrum möglicher Fragestellungen

(siehe z.B. Okun, 1995): Was macht, einfach ausgedrückt, alte Menschen zufrieden bzw. unzufrieden und gibt es diesbezügliche Unterschiede zu Jüngeren? In welcher Weise entwickeln sich Zufriedenheiten im höheren Erwachsenenalter und welches sind die „Moderatoren“ solcher Entwicklung? Verändern sich die Bedeutungen verschiedener Einflußfaktoren auf Zufriedenheit im Lebenslauf – aufgrund altersspezifischer Interessen- und Bedürfnislagen und möglicherweise auch infolge altersabhängiger Veränderungen von Wahrnehmungs- und Informationsverarbeitungsprozessen? Die folgende Untersuchung behandelt gewissermaßen einen bestimmten Ausschnitt aus diesem Interessenspektrum: Es wird versucht, einen Beitrag zur Diskussion um die Stabilität der Zufriedenheit im höheren Alter zu leisten, welche durch die im nächsten Kapitel genauer beschriebenen Befunde des sogenannten Zufriedenheitsparadoxes angestoßen wurde. Dieser Beitrag soll aus einer quantitativen Analyse der längsschnittlichen Entwicklungsdynamiken, die sich hinter diesen besagten Befunden verbergen – und insbesondere in einer Untersuchung der tatsächlich vorhandenen Merkmalsstabilität und ihrer Veränderung über das höhere Lebensalter – bestehen. Denn gerade angesichts vieler Veröffentlichungen zu diesem Themenfeld, in denen eine gewisse Stabilität der Lebenszufriedenheit bei älteren Menschen immer wieder vorausgesetzt oder diskutiert wurde (vgl. die im folgenden aufgeführten Beispiele), erscheint die bisherige empirische Untersuchung dieser Stabilität eher dürftig: Es soll noch gezeigt werden, daß insgesamt die Entwicklung der Zufriedenheit über die Lebensphase der Höher- bzw. Hochaltrigkeit hinweg bei weitem noch nicht erschöpfend *längsschnittlich* untersucht ist – und längsschnittlicher Analysen bedarf es, wenn tatsächlich intraindividuelle Veränderung oder Stabilität eines Merkmals erfaßt werden soll.

Das Untersuchungsziel ist somit sozusagen eher deskriptiver, als erklärender Natur: Es wird, vereinfacht ausgedrückt, nicht untersucht, *was* Zufriedenheit im höheren Lebensalter beeinflußt und verändert, sondern *ob* und *wie* sie sich typischerweise verändert. Polemisierend könnte man sagen, daß dagegen manche Veröffentlichung diese Veränderung bzw. Stabilität erklärt, ohne sie überhaupt festgestellt zu haben. Jedenfalls erscheint es als durchaus lohnendes Ziel, die Entwicklungsdynamiken der Zufriedenheit im höheren Alter „an sich“ zu analysieren, der Gewinn dieser Analysen sollte dann gewissermaßen in einer Klarstellung liegen, nämlich derjenigen Effekte, die die Entwicklung von Zufriedenheit im höheren Lebensalter charakterisieren und folglich überhaupt theorierelevant sind. Eine sorgfältige Suche setzt allerdings voraus, daß man weiß, *wonach* man sucht – und d.h. hier, daß die Untersuchung begründeten Vermutungen über solche Effekte folgen kann. Diese Begründung von Hypothesen bezüglich möglicher entwicklungs-dynamischer Effekte auf die Zufriedenheiten älterer Menschen ist der Gegenstand der noch folgenden theoretischen Unterkapitel: Es werden die Befunde zum Zufriedenheitsparadox kurz dargestellt und dann Effekte diskutiert, die diese Befunde hervorgebracht haben könnten. Zur Orientierung des Lesers seien diese bereits hier genannt: Überlagerung negativer Alters- und Kohorteneffekte, längsschnittliche Differenzierung der Zufriedenheiten (wachsende Varianz bei invarian-ten Durchschnittswerten), lebenslang hohe sowie im höheren Alter wachsende Merkmals-

stabilität von Zufriedenheit.

Zunächst aber noch zwei grundsätzliche Anmerkungen:

- ▶ Es ist hier und im folgenden immer wieder von „höherem“ bzw. „hohem“ Lebensalter, von „alten Menschen“ bzw. „den Alten“ etc. die Rede, ohne daß dieses durch konkrete Altersangaben präzisiert würde. Tatsächlich ist diese Präzisierung auch nicht so ohne weiteres möglich bzw. sinnvoll: Die weitverbreitete Binsenweisheit, daß jemand so alt ist, wie er sich fühlt, verweist bereits darauf, daß das Alt-Sein oder die Hochaltrigkeit einen psychologisch und sozial konstruierten Abschnitt individueller Biographie darstellt, der durch eine Vielfalt möglicher altersnormativer Variablen vom mittleren Erwachsenenalter abgegrenzt werden kann. Altern ist ein lebenslanger Prozeß und es existiert keine Art „biologischer Sprung“, der im Leben eines jeden Einzelnen den Ausbruch der Hochaltrigkeit markiert (vielleicht bedeutet für viele Menschen des Ende der Berufstätigkeit einen derartigen „sozialen Sprung“) – der Übergang ins höhere Alter ist ein fließender und interindividuell variierender. Deshalb haftet präzisen definitorischen Festlegungen der Hochaltrigkeit durch Abgrenzung eines entsprechenden Bereichs des chronologischen Lebensalters eine gewisse Künstlichkeit und Willkürlichkeit an. Gleichwohl besteht innerhalb der Gesellschaft zweifelsohne breiter Konsens darüber, welcher Altersbereich *ungefähr* das „Alt-Sein“ umfaßt, und eine zumindest grobe Präzisierung der Rede von „höherem Alter“ u. dgl. könnte sich daran orientieren: Ich befinde mich wohl im Einklang mit allgemein geteilten Konnotationen des entsprechenden Altersbegriffs, wenn ich ihn auf diejenigen Personen anwende, die das sechste Lebensjahrzehnt vollendet haben.
- ▶ In der Regel ist die allgemeine Lebenszufriedenheit gemeint, wenn in gerontologischen Veröffentlichungen Zufriedenheit im Alter oder Alterseffekte auf Zufriedenheit erörtert werden. Wenn hier aber oft nur der Begriff Zufriedenheit gebraucht wird, so als Konsequenz aus den in den Kapiteln 1.1.4.6 und 1.1.4.7 vorgestellten theoretischen Sichtweisen: Aus dem Modell von Schwarz & Strack (1991) und der daraus ableitbaren konstruktionistischen Sichtweise der Top-down/Bottom-up-Kontroverse kann gefolgert werden, daß die im folgenden im Zusammenhang des Zufriedenheitsparadoxes dargestellten Überlegungen in mehr oder weniger starkem Maße für jegliche, auch für bereichsspezifischere Zufriedenheitsurteile gelten können, da jedes derartige Urteil, gleichwie weit oder eng sein inhaltlicher Fokus gefaßt sein mag, im Prinzip aufgrund derselben Inferenzstrategien zustandekommen kann. Deshalb erscheint es sinnvoll, diese Erörterungen allgemein auf Zufriedenheit im Alter zu beziehen. Sie sind jedoch gewissermaßen immer zuerst mit Blick auf Lebenszufriedenheit zu sehen und bewerten und sie mögen für spezifischere Zufriedenheiten in dem Maße gültig sein, wie diese durch dieselben Urteilsstrategien geprägt sind, wie erstere.

1.2.1 Das Zufriedenheitsparadox des höheren Lebensalters

Betrachtet man Forschungsbefunde zum „quantitativen“ Zusammenhang zwischen Alter und Zufriedenheit, so zeigt sich ein überraschendes Bild: In nahezu allen Studien der letzten drei Dekaden wurden bis weit ins hohe Erwachsenenalter hinein *keine* eindeutigen und deutlichen Zusammenhänge zwischen Zufriedenheitsindikatoren und Lebensalter bzw., wenn überhaupt, eine leichte Zunahme in den höheren Altersgruppen im Vergleich zu Jüngeren gefunden (siehe z.B. Perrig-Chiello et al., 1996; Smith et al., 1996; Horley & Lavery, 1995; Herzog & Rodgers, 1981; Larson, 1978; siehe v.a. auch Ergebnisse großangelegter internationaler Surveys bei: Diener & Suh, 1997; Inglehart, 1990; Butt & Beiser, 1987; Veenhoven, 1984). Dagegen konstatierte noch Wilson (1967) aufgrund der Ergebnisse älterer Untersuchungen eine alterskorrelierte Abnahme von Lebenszufriedenheit, die seither nur noch vereinzelt und wenn, dann nur als sehr schwacher Zusammenhang gefunden wurde (z.B. Doyle & Forehand, 1984).

Es kann also gesagt werden, daß – auch unter unterschiedlichen sozialen und kulturellen Bedingungen – zumindest während der letzten Dekaden alte Menschen sich im Durchschnitt nicht unzufriedener mit ihrem Leben zeigten, als Menschen aus jüngeren Altersgruppen, und dieses Ergebnis entsprach nicht den Erwartungen, die vorwissenschaftlich und zum Teil auch in der Alternsforschung bezüglich dieses Zusammenhangs bestanden. Demnach nämlich sollte ganz allgemein das SWB älterer Menschen sich auf niedrigerem Niveau befinden, als dasjenige Jüngerer, da das hohe Alter sowohl gemäß eines weitverbreiteten negativen Stereotyps, als auch erfahrungswissenschaftlich belegt eine krisenbehaftete Lebensphase ist, die mit erhöhter Morbidität und in der Regel mit Funktionseinbußen und Verlusten im sozialen und psychologischen Bereich verbunden ist (entsprechende empirische Befunde z.B. bei: Mayer & Baltes, 1996) – und dieses im Kontrast zum vorangegangenen mittleren Erwachsenenalter, für welches eine besonders günstige Konstellation psychosozialer, ökonomischer und gesundheitlicher Bedingungen charakteristisch scheint (Mirowski & Ross, 1999). Zudem scheint die in den letzten Jahrzehnten erreichte Verbesserung der gesundheitlichen Versorgung (zumindest in den entwickelten Industrienationen) zu einer zunehmenden Verschiebung des Krankheitsgeschehens ins höhere Erwachsenenalter, d.h. zu einer zunehmenden Konzentration schwerer Erkrankungen auf die späteren Lebensphasen geführt zu haben, so daß Hochaltrigkeit mehr denn je durch eine sozusagen relative Häufung gesundheitlicher Probleme (im Vergleich zum mittleren Erwachsenenalter) gekennzeichnet ist (Fries, 1991; Albaredo et al., 1993, 183-217: Kapitel IV). So ist für Brandtstädter & Rothermund (1998) Höheraltrigkeit v.a. durch den Verlust von gesundheitlichen (höhere Inzidenz chronisch-degenerativer Erkrankungen, höhere Multimorbidität), sozialen (Rollenverluste, Verluste von Interaktionspartnern und Interaktionsgelegenheiten) und lebenszeitlichen (weniger verfügbare Restlebenszeit) Handlungsressourcen geprägt, d.h. durch eine von den Älteren häufig erlebte und berichtete Diskrepanz zwischen Handlungszielen und Handlungsmöglichkeiten: „Solche Diskrepanzerlebnisse verweisen darauf, daß die Einschränkung von Handlungsressourcen als zentrales Problem des Alterns erlebt

wird“ (ebd., 224).

Alle diese Umstände sprechen also dafür, daß eigentlich bei älteren Menschen eine Reduktion des SWB stattfinden und folglich ein Abfall der Zufriedenheitswerte über das höhere Lebensalter hinweg beobachtbar sein sollte. Zudem wurde verschiedentlich argumentiert, daß zudem das dementsprechende Altersbild selbst hier verstärkend in diese Richtung in der Art einer self-fulfilling prophecy wirken sollte: „Das allgemeine Stereotyp des hohen Alters ist jedoch negativ. Dies legt die Vermutung nahe, daß es für alte Menschen sozial akzeptabler ist, eingeschränktes Wohlbefinden anzugeben, und daß möglicherweise der soziale Druck der Norm ‚mein Leben ist schön‘ zu entsprechen, nicht so stark wie im jüngeren Alter empfunden wird“ (Smith et al., 1996, 500). Zwar widersprechen neuere Studien der hier behaupteten Dominanz eines einseitig negativen Stereotyps (zum Überblick vgl. Filipp & Mayer, 1999) bzw. scheinen sie zu belegen, daß es dann, wenn es als soziale Repräsentation existiert, oft individuell nicht übernommen wird (z.B. Heckhausen & Brim, 1997). Jedoch könnte mit Blick auf die hier eigentlich angesprochene soziale Erwünschtheit des Antwortverhaltens immer noch argumentiert werden, daß die Vermutung eines allgemein vorhandenen, wenn auch selbst nicht geteilten negativen Altersbildes eine in der oben zitierten Weise von positiven normativen Verzerrungen befreite Angabe von Zufriedenheit erlaubt.

Befunde zu affektiven Wohlbefindenskomponenten im höheren Lebensalter ähneln denen zur Lebenszufriedenheit nur teilweise: Diener & Suh (1997) fanden im Altersgruppenvergleich stabile Durchschnittswerte in den negativen, nicht jedoch in den positiven Affekten, bei denen sich erwartungsgemäß mit höherem Alter niedrigere Durchschnittswerte zeigten. Dieses Ergebnis konnten Kunzmann et al. (2000) mit Daten aus der Berliner Altersstudie (BASE, vgl. Mayer & Baltes, 1996) bestätigen. Beide Untersuchungen beruhen auf den Konstrukten zweier unabhängiger Affektdimensionen, die vorne bereits diskutiert wurden (Kapitel 1.1): Es wurde dort bezüglich der wohlbefindenstheoretischen Bedeutung dieser Dimensionen der Standpunkt vertreten, daß sie nicht prinzipiell mit habituellem affektiven Wohlbefinden gleichzusetzen sind. Denn die Konstruktion der Habitualität setzt irgendeine Verrechnung des affektiven Erlebens voraus und es sind divergierende theoretische Standpunkte über den Modus dieser Verrechnung möglich. Beispielsweise ist gemäß der vorne dargestellten Affektfrequenztheorie des habituellen affektiven Wohlbefindens nur eine teilweise inhaltliche Überschneidung desselben mit den beiden Affektdimensionen anzunehmen, da letztere als wie auch immer erfragte Frequenzen bestimmter diskret wahrgenommener und intensiver Emotionen konzipiert sind, während die Affektfrequenztheorie letztlich auf die Verrechnung über die Zeit hinweg kontinuierlicher affektiver Tönung, unabhängig von Inhalt und Intensität, begründet ist. Konsequenterweise wäre dann aus den berichteten Ergebnissen auch kein eindeutiges Fazit abzuleiten was die Stabilität des nach der Affektfrequenztheorie konzipierten durchschnittlichen habituellen affektiven Wohlbefindens in höheren Lebensaltersgruppen anbelangt. Solange die konzeptionelle Beziehung zwischen diesen Affektdimensionen und dem, was als habituelles affektives SWB bezeich-

net werden kann, nicht ganz eindeutig geklärt ist, kann m.E. also keine Schlußfolgerung über die Stabilität des letzteren aus Befunden zur Stabilität ersterer abgeleitet werden und sollten diese Befunde als das benannt werden was sie sind: Nämlich solche zur Stabilität der Frequenz negativer bzw. positiver Emotionen.

Der Umstand, daß entgegen entsprechender Erwartungen Menschen in höheren Altersgruppen im Durchschnitt mit ihrem Leben nicht unzufriedener sind (und sich möglicherweise auch nicht deutlich schlechter fühlen), als Jüngere, wird in der gerontologischen Literatur vielfach als Zufriedenheits- oder Wohlbefindensparadox des höheren Lebensalters bezeichnet (z.B. Baltes & Baltes, 1990; Brandtstädter & Greve, 1992; 1994; Kunzmann et al., 2000; Staudinger, 2000). Ob diese Begriffswahl die Sache zutreffend bezeichnet, kann bezweifelt werden, denn „paradox“ im Sinne einer logischen Widersprüchlichkeit ist die Nichtübereinstimmung der für das höhere Alter typischen Entwicklungen der objektiven Lebensqualität und des SWB eigentlich nicht. Jedoch zeigt der Begriff sozusagen das gerontologische Erstaunen über dieses Phänomen: Warum nur sind alte Leute zufrieden, wenn sie guten Grund zur Unzufriedenheit hätten? Die Beantwortung dieser Frage ist, wie vorne schon gesagt wurde, von großer gerontologischer Relevanz: Würde das SWB älterer Menschen weitestgehend „reaktiv“ den gesundheitlichen und sozialen Veränderungen folgen, die im höheren Alter in der Regel stattfinden, so wäre der Rahmen dessen, was gerontologische Forschung überhaupt praktisch bewirken kann, eher eng gesteckt. Denn es müßte dann ja eine Verschlechterung des SWB in dieser Lebensphase prinzipiell vorausgesetzt werden, da nun einmal gewisse gesundheitliche, soziale und auch psychische Einbußen im hohen Alter unvermeidlich sind. Die Suche nach positiven Interventionsmöglichkeiten, mit denen Alternsprozesse verbessert werden könnten, erhielte damit sozusagen einen defensiven Charakter – es könnte dabei eigentlich nur um eine zumindest teilweise Vermeidung oder Hinauszögerung altersbedingter Verluste gehen und es wäre dadurch schlußendlich ein Altersmodell impliziert, welches Altern in erster Linie mit Abbau und herannahendem Tod gleichsetzt und eigentlich nur als Phase objektiv *und* subjektiv schwindender Lebensqualität begreifen kann. Das sogenannte Wohlbefindensparadox enthält somit gewissermaßen eine frohe Botschaft: Alte Menschen können sich möglicherweise auch dann noch (subjektiv) wohl befinden, wenn sich ihre objektiven Lebensumstände verschlechtern. Der Gerontologie selbst wurde damit ein interessantes Forschungsfeld zur Erklärung des Paradoxes eröffnet: Das Wissen darum, weshalb und unter welchen Bedingungen alte Menschen Lebenszufriedenheit (und möglicherweise auch affektives Wohlbefinden) bei alterstypischen Verlusten aufrechterhalten können, kann z.B. die Grundlage für gerontologische Interventionsstrategien sein, die auf eine Stärkung dieser „Zufriedenheitskompetenz“ im höheren Lebensalter abzielen.

Häufig wird nun in den Veröffentlichungen zu diesem Thema das Zufriedenheitsparadox begrifflich und konzeptuell mit Stabilität der Zufriedenheit im höheren Alter gleichgesetzt, beispielsweise auch durch die Bezeichnung als „stability-despite-loss paradox“ (z.B. Kunzmann et al., 2000). Tatsächlich zeigen die oben beschriebenen Befunde im – meist

querschnittlichen – Altersgruppenvergleich durchschnittlicher Zufriedenheitsmaße ja eine große Stabilität derselben bei den Höheraltrigen. Es kann jedoch daraus nicht ohne weiteres eine intraindividuelle Merkmalsstabilität von Zufriedenheit bei alten Menschen abgeleitet werden, vielmehr ist letztere nur eine von mehreren *Möglichkeiten* zur Erklärung solcher „querschnittlicher“ Mittelwertstabilität. Eine eingehendere Untersuchung der Frage, inwieweit SWB im höheren Alter „stabil aufrechterhalten“ wird, setzt zwingend die Erfassung intraindividuelle Veränderung des fraglichen Merkmals und d.h. längsschnittliche Untersuchungsdesigns voraus. Solche längsschnittlichen Untersuchungen jedoch sind eher rar gesät – ein Mangel, der bereits von Kozma & Stones (1978) und jüngerer Vergangenheit auch in den Reviews von Diener & Suh (1997) und Diener et al. (1999) beklagt wurde.

Die längsschnittliche Untersuchung des Ausmaßes intraindividuelle Veränderung von Zufriedenheit im höheren Alter kann also als Grundanforderung jeglicher wissenschaftlicher Auseinandersetzung mit dem sogenannten Zufriedenheitsparadox angesehen werden und es ist zu konstatieren, daß diese Anforderung zumindest auf den ersten Blick bislang nur bruchstückhaft erfüllt scheint. Ein zweiter Blick aber erscheint angesichts der Relevanz solcher längsschnittlicher Befunde gerade auch für die hier im folgenden noch dargestellte Untersuchung geboten, nämlich v.a. auf die wenigen existierenden gerontologischen Längsschnittstudien bzw. auf die dort möglicherweise auch nur am Rande berichteten diesbezüglichen Ergebnisse.

1.2.2 Längsschnittliche Untersuchungen zum Zusammenhang von höherem Lebensalter und Zufriedenheit

Schaie & Hofer (2001) listen in ihrem Überblick immerhin 27 gerontologische Längsschnittstudien auf (vgl. auch das Review von Frazier et al., 1993; sowie die mit diesbezüglichem Themenschwerpunkt erschienene Ausgabe der Zeitschrift für Gerontologie, 1993, 26(3)). Die Suche in den Veröffentlichungen aus diesen Projekten fördert allerdings erstaunlich wenige längsschnittliche Befunde zur Lebenszufriedenheit zutage. Viele dieser Studien waren in erster Linie an der Entwicklung kognitiver oder physischer Funktionen im höheren Alter interessiert und widmeten, falls überhaupt Maße des SWB untersucht wurden, sich kaum deren längsschnittlicher Veränderung. Im folgenden werden die diesbezüglichen „Fundstücke“ kurz beschrieben.

Explizit wurde die intraindividuelle Veränderung bzw. Stabilität der Lebenszufriedenheit Älterer in der DUKE-II-Studie (Überblick z.B. Busse, 1993) von Palmore & Kivett (1977; 1985; beide Veröffentlichungen sind inhaltlich identisch) thematisiert. Bei 378 Probanden im Alter von 46 bis 70 Jahren zum ersten Meßzeitpunkt 1968 wurden drei Datenerhebungen in Zwei-Jahres-Abständen durchgeführt. Für die Lebenszufriedenheit (gemessen mit der sogenannten „Cantril Ladder“, vgl. Beschreibung in der referenzierten Arbeit) fanden sich keine längsschnittlichen Mittelwertsveränderungen, auch nicht bei Betrachtung

enger gefaßter Altersgruppen über die 3 Meßzeitpunkte hinweg. In einer Serie von Regressionsanalysen wurden darüber hinaus gewisse Belege für eine hohe intraindividuelle Stabilität der Zufriedenheit gefunden: Für die Vorhersage der Zufriedenheiten zum zweiten bzw. dritten Meßzeitpunkt erwiesen sich sonstige Prädiktoren jeweils als relativ bedeutungslos im Vergleich zu den „vorherigen“ Zufriedenheiten zum ersten bzw. zweiten Meßzeitpunkt. Die frühere Zufriedenheit war also jeweils der wichtigste Prädiktor der späteren. Die Lebenszufriedenheit zum dritten Meßzeitpunkt korrelierte mit der bei der zweiten Messung erhobenen zu 0.56 und mit der des ersten Meßzeitpunkts zu 0.40 – nimmt man solche seriellen Korrelationen als Maß für die Stabilität des Merkmals (vgl. Kapitel 2.2.1), so könnten diese Werte als Beleg für eine zwar durchaus substantiell ausgeprägte moderate Stabilität gewertet werden, jedoch zeigen sie auch, daß die Varianz der späteren Werte bei weitem nicht ausschließlich auf die der zeitlich früheren zurückführbar ist und somit große Anteile von „Veränderungsvarianz“ enthält. Eine Einschätzung dieser Ergebnisse in Bezug auf die tatsächliche Stabilität der Lebenszufriedenheit fällt m.E. auch darum schwer, weil zu berücksichtigen ist, daß solche seriellen Meßwertkorrelationen durch die Reliabilität der Messungen beeinflusst werden und in der Regel unter den Korrelationen der „wahren Werte“ liegen (vgl. die Darlegungen in Kapitel 2.2.2, insbesondere Formel 2.23).

Weitere längsschnittliche Befunde zu Maßen des SWB liefern verschiedene Veröffentlichungen aus der Baltimore Longitudinal Study of Aging (Costa & McCrae, 1993): Das Interesse der Forschergruppe um P.T. Costa und R.R. McCrae galt der Stabilität von Persönlichkeitsdimensionen und dem Einfluß derselben auf SWB (Costa & McCrae, 1984) – und somit indirekt auch der Stabilität des SWB. Der Schwerpunkt lag dabei allerdings eher bei affektiven Wohlbefindensmaßen, für welche eine weitestgehende Stabilität der Durchschnittswerte über die zwei Meßzeitpunkte (1971-75 bzw. 1981-84) mit einem Zeitintervall von immerhin neun Jahren berichtet wird (befragt wurden 4942 Personen mit Interviews zu beiden Messungen, im Alter von 25 bis 74 zum ersten Zeitpunkt, vgl. Costa et al., 1987b). Allerdings zeigen die Werte durchaus für die (zum ersten Zeitpunkt) höheren Altersgruppen gewisse Verschlechterungen des Wohlbefindens, die von den Autoren allerdings als nicht bedeutsam eingeschätzt werden. Als „9-year retest stability“ des verwendeten SWB-Gesamtscores wurde 0.48 angegeben, was angesichts des sehr langen Zeitintervalls wohl als sehr hoher Wert einzuschätzen ist. Costa & McCrae (1984) berichten Retest-Stabilitäten von 0.47 bis 0.63 für die verschiedenen erhobenen SWB-Maße, darunter auch ein über die erfragten Zufriedenheiten in 14 Lebensbereichen gebildeter „Satisfaction-Index“. Insgesamt bleibt die Bedeutung dieser Ergebnisse für die Stabilität der Lebenszufriedenheit allein schon darum fraglich, weil eben diese nicht explizit operationalisiert wurde, die zitierten Arbeiten erscheinen so als Beitrag zur Stabilität des affektiven Wohlbefindens.

Neben den Veröffentlichungen aus der Baltimore-Studie berichten Costa & McCrae (1980) auch Ergebnisse aus der Normative Aging Study (allgemeiner Überblick z.B. bei Bell et al., 1972), einer weiteren vielzitierten gerontologischen Längsschnittstudie. Allerdings liefern sie hier keine Auswertungen zu Zeitreihen von Zufriedenheitsmessungen, son-

dern analysieren lediglich Zusammenhänge verschiedener Maße des SWB untereinander und mit Persönlichkeitsfaktoren im Querschnitt (des Befragungsjahres 1976). Auch Ekerdt et al. (1985) analysieren Lebenszufriedenheitsmessungen der Normative Aging Study wiederum nur querschnittlich, hier gilt das Augenmerk dem Zusammenhang von SWB und zeitlichem Abstand seit der Pensionierung. Mroczek & Spiro (2003) präsentieren eine interessante längsschnittliche Analyse zur intraindividuellen Stabilität von Extraversion und Neurotizismus mittels Daten der Normative Aging Study, im Zusammenhang der hier im folgenden präsentierten Untersuchung erscheint dieses jedoch nur aufgrund der statistischen Modellierung, die der in Kapitel 2.1 dargestellten sehr ähnelt, relevant – Daten zum SWB wurden leider nicht in ähnlicher Weise analysiert.

Eine im Vergleich zu den vorgenannten Studien etwas gewichtigere Rolle spielte die Lebenszufriedenheit als Indikator des erfolgreichen Alterns in der Bonner Gerontologischen Längsschnittstudie (BOLSA) (Thomae, 1987; zum knappen Überblick über BOLSA siehe Thomae, 1993). In den von Grombach (1976; 1987; beide Arbeiten sind weitgehend inhaltsgleich, ersteres ist die englischsprachige, letzteres die deutsche Version) vorgelegten Analysen zu Konstanz und Veränderlichkeit verschiedener BOLSA-Variablen wurde somit auch die „Zufriedenheit mit der gegenwärtigen Situation“ miteinbezogen. Geschätzt wurden u.a. Stabilitätskoeffizienten als meßfehlerbereinigte Korrelationen zwischen den aufeinanderfolgenden Erhebungen (drei Meßzeitpunkte: 1965/66, 1967/68, 1969/70; bei der Ersterhebung 222 Personen aus den Geburtsjahrgangskohorten 1900-1905 und 1890-1895), und zwar mittels des längsschnittlichen Strukturgleichungsmodells von Heise (1969). Dabei handelt es sich um die früheste Version des in der nachfolgenden Untersuchung verwendeten Quasi-Markov-Simplex-Modells (vgl. die Darstellung in Kapitel 2.2.2.1), welche allerdings wegen der ihr innewohnenden Restriktionen (konstante Reliabilitäten und Varianzen der „wahren Werte“ über die Zeitpunkte hinweg) kritisiert und weiterentwickelt wurde (vgl. Rudinger & Rietz, 1993). Diese kritisierten Beschränkungen des statistischen Modells dürften m.E. dafür verantwortlich sein, daß die Ergebnisse der Analyse für die Zufriedenheitswerte eher Zweifel an ihrer Gültigkeit bzw. Interpretierbarkeit nähren: Der Stabilitätskoeffizient im ersten Zeitintervall wurde mit -0.26, der im zweiten mit -1.00 und der Stabilitätskoeffizient im Gesamtintervall vom ersten zum dritten Meßzeitpunkt mit 0.90 geschätzt, dieses bei einem im Grund inakzeptablen Reliabilitätskoeffizienten von -0.15. Sowohl die negativen Vorzeichen bei den meisten dieser Koeffizienten (d.h. negative Zusammenhänge zwischen den Zufriedenheiten sukzessiver Erhebungszeitpunkte), als auch der Gegensatz von in ihrem Absolutbetrag teils extrem hohen Stabilitätskoeffizienten und sehr niedriger Reliabilität dürfte hier wohl eher ein Modellartefakt zeigen, das eben aufgrund der unrealistischen Restriktivität des Modells erzeugt wurde – jedenfalls erscheinen die Zahlen kaum sinnvoll interpretierbar.

Weitere Angaben über Veränderungen von Zufriedenheitswerten aus BOLSA finden sich bei Schmitz-Scherzer & Thomae (1983): Hier werden verschiedene Zusammenhänge mit ausgewählten Prädiktoren dieser Veränderung erwähnt, die aber für die hier relevante

Frage des statistischen Zusammenhangs von Alter und Lebenszufriedenheit nicht unmittelbar bedeutsam scheinen.

Unter den Veröffentlichungen von Ergebnissen aus der Berliner Altersstudie (BASE, allgemeiner kurzer Überblick bei Smith et al., 2002) findet sich die bereits oben erwähnte Arbeit von Kunzmann et al. (2000), in der explizit das Wohlbefindensparadox bereits im Titel adressiert und längsschnittlich untersucht wird. Allerdings beschränkt sich diese Studie, wie ebenfalls bereits dargelegt, auf affektive Wohlbefindenskomponenten, zur intraindividuellen Stabilität von Lebenszufriedenheit macht sie keine Aussagen. Dagegen untersuchen Smith et al. (1996) anhand der BASE-Daten den Zusammenhang von höherem Lebensalter und Lebenszufriedenheit, dieses allerdings nur querschnittlich anhand der Daten der ersten BASE-Erhebung. Smith & Freund (2002) untersuchen die Dynamiken von Selbstbildentwürfen im hohen Alter, u.a. auch in Abhängigkeit von Veränderungen der Lebenszufriedenheit. In diesem Zusammenhang wird ein standardisierter mittlerer Veränderungswert (durchschnittliche Veränderung in T-Score-Einheiten pro Jahr) aus zwei Meßzeitpunkten (1990-93, 1995/96) der Lebenszufriedenheitsmessungen angegeben, nämlich -0.54 (berechnet für 119 Personen im Alter von 70 bis 79 und 87 Personen 80 Jahre und älter zum ersten Meßzeitpunkt) – bei einer Standardabweichung von 2.53 zeigt dieser Wert also einen insgesamt leichten Rückgang der Lebenszufriedenheit bei den Hochaltrigen. Bei manchen der in dieser Studie untersuchten Subgruppen wurden aber auch durchschnittliche Zufriedenheitsgewinne gefunden.

Als weitere besonders „prominente“ gerontologische Längsschnittstudien wären z.B. die Seattle Longitudinal Study (Schaie, 1983), die Swedish Adoption/Twin Study of Aging (SATSA, Pedersen et al., 1991), die Victoria Longitudinal Study (Hultsch et al., 1998), die Australian Longitudinal Study of Aging (ALSA, vgl. z.B. Luszcz, 1999), sowie die Interdisziplinäre Längsschnittstudie des Erwachsenenalters (ILSE, siehe z.B. Martin et al., 2000) zu nennen. Die Suche in den Veröffentlichungen aus diesen Projekten förderte keine spezifischen Resultate zum längsschnittlichen Zusammenhang von Lebenszufriedenheit und Alter zutage. Lediglich am Rande noch erwähnenswert scheint hier eine neuere Veröffentlichung aus der Victoria Longitudinal Study zur Stabilität und Veränderung von Persönlichkeitsmerkmalen (Small et al., 2003), dieses allerdings nur wegen der dort gewählten und diskutierten methodischen Vorgehensweise, welche hinsichtlich der in Kapitel 2.2.2 dargestellten längsschnittlichen Strukturgleichungsmodellierungen zur Stabilitätsanalyse interessant scheint – SWB wird auch hier nicht untersucht. Darüber hinaus scheint es nur wenige vereinzelte Untersuchungen der längsschnittlichen Veränderung von Lebenszufriedenheit im höheren Lebensalter gegeben zu haben.

McNeil et al. (1986) befragten 408 ältere Personen mit einem Durchschnittsalter von 75 Jahren zu zwei Meßzeitpunkten (im 18-monatigem Abstand) und erhoben ein Globalmaß des SWB (sog. Memorial University of Newfoundland Scale of Happiness, vgl. im zitierten Text), sowie verschiedene bereichsspezifische Zufriedenheiten. Sie fanden einen minimalen, statistisch nichtsignifikanten durchschnittlichen Rückgang des SWB zwischen den

Meßzeitpunkten. Mit demselben Datensatz berechneten Kozma & Stones (1983) schrittweise Regressionsanalysen (für die Gesamtstichprobe und Substichproben urbaner, ländlicher und institutionalisierter Probanden), dabei erwies sich bei der Vorhersage des Zeitpunkt-2-SWB jeweils das Zeitpunkt-1-SWB als derjenige Prädiktor, der zuerst in das Regressionsmodell aufgenommen wurde und für sich bereits rund 50% Varianzaufklärung erzielte, während weitere Prädiktoren darüber hinaus kaum noch nennenswerte Zugewinne brachten. Nimmt man beide Untersuchungen zusammen, so zeigen sie Belege sowohl für eine durchaus hohe intraindividuelle Stabilität, als auch Mittelwertstabilität des untersuchten SWB-Maßes.

Baur & Okun (1983) untersuchten 91 Personen im Alter von 66 bis 70 Jahren (bei Erstbefragung) zu zwei Meßzeitpunkten (1977 und 1980). Für den verwendeten Lebenszufriedenheitsindex fanden sie eine lediglich minimale, statistisch insignifikante Mittelwertsdifferenz zwischen den Meßzeitpunkten, sowie einen geringfügigen Anstieg der Varianz dieser Scores, der von den Autoren als „borderline significant“ bezeichnet wird. Die serielle Korrelation der Zufriedenheiten betrug 0.61. Das Fazit der Autoren lautet: „Overall, despite a marginally significant difference between the variances of the two distributions, the LSIB [*das ist der verwendete Lebenszufriedenheitsindex, O.S.*] scores can be characterized as relatively stable“ (Baur & Okun, 1983, 262).

Ferring & Filipp (1997) untersuchten Personen im Alter von 65 bis 94 Jahren zu zwei Meßzeitpunkten im Halbjahresabstand (238 Personen mit vollständigen Daten) und berechneten mittels einer Strukturgleichungsmodellierung für verschiedene kognitive SWB-Indikatoren Konsistenzkoeffizienten als Anteile von über die Meßzeitpunkte stabiler, interindividueller Gesamtvarianz an der Varianz der Meßwerte. Es wurden u.a. Indikatoren der aktuellen Zufriedenheit mit dem gegenwärtigen, sowie der retrospektiven Zufriedenheit mit dem vergangenen und der prospektiven Zufriedenheit mit dem zukünftigen Leben erhoben. Für die Messung der aktuellen Lebenszufriedenheit wurden Konsistenzkoeffizienten zwischen 0.66 und 0.68 geschätzt – gegenüber einem Spezifitätskoeffizienten (Anteil situationsspezifischer Varianz) von jeweils 0.15. Die entsprechenden Konsistenzen der retrospektiven Lebenszufriedenheit waren demgegenüber etwas reduziert, die der prospektiven Lebenszufriedenheit fielen deutlich niedriger aus. Insgesamt zeigte sich hier also v.a. die Bewertung der aktuellen allgemeinen Lebenszufriedenheit in hohem Maße stabil im Sinne des verwendeten Modells der Varianzaufteilung (d.h. Abhängigkeit von personenspezifischen vs. von situationsspezifischen Varianzanteilen). Die Autoren werten die Ergebnisse als Beleg dafür, „daß mit zunehmendem Alter Erwartungen an die ‚Güte des eigenen Lebens‘ so adaptiv reguliert werden, daß sich ein stabilisierender Effekt auf die Höhe der Lebenszufriedenheit ergibt“ (ebd., 254).

Chen (2001) untersuchte zu zwei Meßzeitpunkten (1989, 1993) eine große Stichprobe taiwanesischer Personen im Alter ab 60 (3155 nahmen an beiden Interviews teil). Es wurde ein durchschnittlicher Rückgang der Lebenszufriedenheit nach dem 65. Lebensjahr gefunden, welcher durch verschiedene sozio-ökonomische Korrelate erklärt werden konnte. Nach

Auspartialisierung dieser Kovariaten resultierte ein tendenzieller Anstieg der Zufriedenheit mit dem Alter, welcher wiederum durch kohortenspezifische Lebenserfahrungen statistisch erklärt werden konnte.

Zwei nichtgerontologische Längsschnittstudien verdienen in diesem Zusammenhang noch Erwähnung. Nämlich zum einen die Victorian Quality of Life Panel Study, bei der im zweijährigem Abstand (1981, 1983, 1985, 1987) Personen im Alter von 18 bis 65 (zur Ersterhebung, 649 Personen nahmen über alle vier Meßzeitpunkte hinweg teil) untersucht wurden (Headey & Wearing, 1991). Die Autoren berichten einen Stabilitätskoeffizienten der Lebenszufriedenheit von 0.57, berechnet als Quotient der Sechs-Jahres-Follow-Up-Korrelation und des Reliabilitätskoeffizienten der Zufriedenheitsmessung (zur Erklärung dieser Berechnungsweise vgl. die Formel 2.23 im Kapitel 2.2.2). Es handelt sich somit um die Schätzung einer meßfehlerbereinigten seriellen Korrelation „wahrer Werte“, wodurch dieser Wert vergleichsweise niedrig erschien (z.B. im Vergleich zu den oben aufgeführten Werten von Ferring & Filipp, 1997), wenn nicht auch das immerhin sehr lange Zeitintervall bei der Bewertung berücksichtigt würde. Ein weiterer interessanter Befund, der hier mitgeteilt wird, ist ein im querschnittlichen Altersgruppenvergleich sich zeigender Zusammenhang zwischen Lebensalter und affektivem Wohlbefinden dergestalt, daß Jüngere sowohl mehr positiven, als auch mehr negativen Affekt erlebten, als Ältere.

Eine weitere erwähnenswerte nichtgerontologische Längsschnittuntersuchung liefert Schräpler (1995). Er untersucht die Zufriedenheitsdaten der ersten sieben Erhebungswellen des westdeutschen SOEP (vgl. die Beschreibung des Datensatzes in Kapitel 3.1) mittels des Strukturgleichungsmodells, welches in Kapitel 2.2.2.1 dargestellt und für die noch folgenden Stabilitätsanalysen benutzt wurde. Deshalb könnte man Schräplers Arbeit – die leider nur als „graue Literatur“ publiziert ist und gegenwärtig kaum noch erhältlich sein dürfte – als Vorläufer der hier noch gezeigten Untersuchung bezeichnen, die, bei Interesse v.a. an der Entwicklung der Reliabilitäten, meßfehlerbereinigte Schätzungen von Stabilitätskoeffizienten berichtet, welche allesamt sehr hoch im Bereich von um und über 0.9 liegen. Daß Schräplers methodischer Ansatz als hochbedeutsam eingeschätzt wurde, geht schon daraus hervor, daß er im folgenden für die eigenen Analysen weiterverfolgt wurde – allerdings ist die Entwicklung der Zufriedenheit nicht Schräplers Thema und somit untersucht er nicht gezielt Hypothesen über diese Entwicklung.

Zusammenfassend kann die „längsschnittliche Befundlage“ zum Zusammenhang von höherem Lebensalter und Lebenszufriedenheit bzw. zum Zufriedenheitsparadox des höheren Lebensalters als eher dürftig bewertet werden. Die aufgeführten Ergebnisse liefern mehr oder weniger deutliche Anhaltspunkte dafür, daß mit dem Übergang in die Lebensphase der Hochaltrigkeit tatsächlich kaum durchschnittliche Einbrüche der Lebenszufriedenheit zu beobachten sind und daß Lebenszufriedenheit bis zu einem gewissen Maß intraindividuell stabil aufrechterhalten wird, – was sie m.E. *nicht* liefern, sind abschließende, weitere Untersuchungen erübrigende Antworten auf die sich ergebenden Fragen zur Stabilität der Lebenszufriedenheit. Wie sind die vereinzelt berichteten leichten durchschnittlichen Zu-

friedenheitsverluste und Stabilitätskoeffizienten nun einzuschätzen – zeigen sie eine eher hohe, mittlere, niedrige Stabilität des Merkmals? Angesichts der geringen Anzahl solcher Untersuchungen, sowie der Unterschiede zwischen diesen in den darin jeweils untersuchten Zufriedenheitsmaßen, Personengruppen und Meßzeitpunkten bzw. Zeitintervallen, ist eine abschließende Eindeutigkeit der Befunde auch kaum möglich und bedarf die weitere wissenschaftliche Erklärung des Zufriedenheitsparadoxes weiterer Untersuchungen der längsschnittlichen Entwicklungsdynamik von Lebenszufriedenheit im höheren Alter.

Auch zeigt ein solcher zusammenfassender Blick auf die vorliegenden Befunde m.E. die Notwendigkeit einer forschungsmethodischen Weiterentwicklung bei der Untersuchung dieser Fragen. Damit soll nicht gesagt werden, daß die zitierten Arbeiten methodische Mängel aufwiesen – dafür ließen sich wohl kaum hinreichend Gründe finden –, aber es scheinen dort die methodischen Möglichkeiten längsschnittlicher Untersuchung, welche erst während der letzten Dekade mit der forcierten Verbreitung und Weiterentwicklung statistischer Strukturgleichungs- und Multilevel-Regressionsmodelle insbesondere für die längsschnittliche Datenanalyse geschaffen wurden (vgl. z.B. Collins & Sayer, 2001; Verbeke & Molenberghs, 2000), noch nicht hinreichend ausgeschöpft. Dieses gilt insbesondere für die Berechnung von Stabilitätsmaßen, welche mit dem Problem behaftet sind, daß in serielle Meßwertkorrelationen Stabilität eines Merkmals und Reliabilität der Messung desselben eingehen, so daß eine Einschätzung des „wahren“ Ausmaßes von Merkmalsstabilität im Grunde irgendeine Art von Meßfehlerbereinigung erfordern würde (vgl. die diesbezüglichen Erläuterungen in Kapitel 2.2.2), welche in konventionell regressionsanalytischen Methoden nicht enthalten ist. Ansätze zur Meßfehlerbereinigung bei Headey & Wearing (1991) und Grombach (1976; 1987) erweisen sich bei näherem Hinsehen als möglicherweise zu restriktiv in ihrer statistischen Modellierung (z.B. setzen Headey & Wearing über die Meßzeitpunkte gleichbleibend stabile Reliabilitäten voraus).

Ein weiteres und für alle aufgeführten Untersuchungen relevantes methodologisches Problem besteht in der Behandlung fehlender Werte, die natürlich gerade in längsschnittlichen Untersuchungen in großer Zahl auftreten (sogenannte Dropouts). Es wurde bei allen Studien ein fallweiser Ausschluß durchgeführt, d.h. Fälle mit teilweise fehlenden Werten (also alle Dropouts) wurden vor der Analyse eliminiert. Bei großen Dropoutraten kann diese Methode als suboptimal gelten, wie in Kapitel 2.3.1 noch genauer dargelegt werden wird, wo auch neuere Entwicklungen zur Behandlung von fehlenden Werten, welche unverzerrtere und effektivere statistische Schätzungen erlauben, gezeigt werden.

Ein weiterer Aspekt, der bei den aufgeführten Untersuchungen unberücksichtigt blieb, ist schließlich die längsschnittliche Entwicklung der Merkmalsstabilität, d.h. die Frage, ob sich die Stabilität von Zufriedenheitsurteilen über die Lebensspanne hinweg bzw. mit dem höheren Alter verändert. Zu dieser Untersuchung bräuchte es, wenn nicht die längsschnittlich berechneten Stabilitätskoeffizienten verschiedener Altersgruppen nur wiederum querschnittlich verglichen werden sollen, längsschnittliche Untersuchungsdesigns über längere Zeiträume mit möglichst vielen Meßwiederholungen, so daß die Entwicklung von Stabi-

litätskoeffizienten aus mehreren sukzessiven Meßintervallen analysiert werden kann.

Durch die jüngsten Entwicklungen der Multilevel- und Strukturgleichungsmethodologien sind die statistischen Modellierungsmöglichkeiten für längsschnittliche Datensätze enorm erweitert worden und es wäre darum ein lohnendes Forschungsziel per se, die Frage der Entwicklung von Zufriedenheit im höheren Alter sozusagen bewaffnet mit diesem erweiterten Methodenarsenal noch einmal anzugehen. Es sollte dabei den soeben angeführten Desideraten besser entsprochen werden können: Untersuchung der meßfehlerbereinigten Merkmalsveränderungen, sowie eine dem „Stand der Dinge“ entsprechende optimale Behandlung der längsschnittlichen Dropouts und möglichst auch die Erfassung der „Veränderung von Veränderung“, d.h. der Entwicklung von Stabilitätsmaßen über einen längeren Zeitraum hinweg. Die Erfüllung dieser Gesichtspunkte ist ein sozusagen zusätzlich zu den „inhaltlichen“ Fragestellungen bezüglich der Zufriedenheit im Alter gewähltes Forschungsziel der hier vorgelegten Arbeit und wird deswegen in einem eigenen ausführlichen Kapitel diskutiert werden (Kapitel 2).

1.2.3 Zur Erklärung des Zufriedenheitsparadoxes: Mögliche Entwicklungsdynamiken der Zufriedenheit im höheren Lebensalter

Es sind verschiedene Entwicklungsdynamiken zumindest denkbar, aus denen die querschnittliche Mittelwertstabilität des Zufriedenheitsparadoxes resultieren könnte – weitestgehende intraindividuelle Merkmalsstabilität der Zufriedenheit bei alten Menschen ist eine davon. Die Intention der vorliegenden Untersuchung besteht, wie nun bereits mehrfach gesagt wurde, in der Überprüfung dieser Möglichkeiten. Zunächst aber müssen diese noch ausreichend theoretisch begründet werden.

1.2.3.1 Überlagerung von Alters- und Kohorteneffekt?

Das Zufriedenheitsparadox des höheren Alters wurde v.a. in Untersuchungen gefunden, deren empirische Datenerhebungen im Zeitraum der letzten beiden Dekaden des zwanzigsten Jahrhunderts stattgefunden haben. Die zu diesen Erhebungszeitpunkten in den Industrienationen älteren Menschen haben – etwas plakativ formuliert – ihr Leben in schlechten Zeiten begonnen: Der Beginn der Lebensspannen der in den achtziger Jahren Älteren fiel in eine historische Periode, die praktisch durch eine kaum unterbrochene Aneinanderreihung sozialer und politischer Problemlagen (erster Weltkrieg, Weltwirtschaftskrise, zweiter Weltkrieg) sowie v.a. auch durch einen allgemeinen gesellschaftlichen Entwicklungsstand gekennzeichnet war, der im Vergleich zu den durch den raschen materiellen und technologischen Fortschritt gekennzeichneten Wohlstandsgesellschaften der achtziger Jahre nur als erheblich rückschrittlich bezeichnet werden kann. Es liegt darum m.E. auf der Hand, zu vermuten, daß Menschen, deren frühe Lebensjahre durch solche vergleichsweise reduzier-

ten sozialen und ökonomischen Bedingungen und in vielen Fällen sogar durch erhebliche Gefährdungen insbesondere während der durchlebten Kriege geprägt waren, ihr Leben unter vergleichsweise risikofreien und mit vielen, dem technologischen Fortschritt geschuldeten Annehmlichkeiten ausgestatteten Wohlstandsbedingungen insgesamt tendenziell besser bewerten, als solche Personen, die die besagten schlechten Zeiten gar nicht oder doch zumindest während weniger früher Lebensjahre erlebt haben.

Und somit könnte hier ein Kohorteneffekt vermutet werden: Es könnte sich bei den Älteren – wiederum etwas plakativ formuliert – um eine genügsamere Generation handeln, die im höheren Lebensalter trotz verstärkt problematischer Bedingungen nicht darum nicht unzufriedener als Jüngere sind, weil Zufriedenheit im Alter stabil aufrechterhalten wird, sondern aufgrund ihres kohortenspezifisch erhöhten Zufriedenheitslevels. Mit Blick auf die in Kapitel 1.1.4.3 dargestellten theoretischen Konzepte wäre anzunehmen, daß das Durchleben schlechter Zeiten bzw. die Erinnerung an diese die Adjustierung subjektiver internaler Standards bezüglich objektiver Lebensbedingungen beeinflusst, so daß die besagte Genügsamkeit in Form bescheidener internaler Soll-Zustände besteht, welche zum Vergleich mit den aktuell wahrgenommenen objektiven Lebensumständen herangezogen werden.

Ein negativer Effekt des *Alterns* auf Zufriedenheit (d.h. intraindividuelle Zufriedenheitsverluste mit zunehmendem Lebensalter) bliebe also bei der Betrachtung dieses Zusammenhangs zu nur einem bestimmten Zeitpunkt möglicherweise dann unentdeckt, wenn zusätzlich bei den heute Älteren ein derartiger Kohorteneffekt auf Zufriedenheit wirkte. In diesem Fall könnten sich Alterns- und Kohorteneffekte derart überlagern, daß Ältere mit mehr Zufriedenheitsverlusten bei kohortenspezifisch erhöhtem „Ausgangsniveau“ dieser Verluste sozusagen jederzeit bei derselben Zufriedenheit angelangt wären, wie die jeweils gleichzeitig Jüngeren mit weniger Verlusten auf niedrigerem Niveau.

Idealtypischerweise wäre dieses dann der Fall, wenn sowohl Alters-, wie auch Kohorteneffekt als jeweils über die Zeitdimensionen Lebensalter bzw. Geburtszeitpunkt hinweg negative lineare Effekte wirksam wären. Allerdings dürfte eine solche Linearität eines Kohorteneffekts einem verbreiteten Verständnis zuwiderlaufen, das den Begriff der Kohorte weniger hinsichtlich des Geburtszeitpunkts der Kohortenmitglieder, als vielmehr aufgrund inhaltlich relevanter historischer Zeitpunkte, die von den Kohortenmitgliedern gemeinsam durchlebt wurden, begreift. In diesem Verständnis ist also die Kohortenzugehörigkeit gewissermaßen eine dichotome (oder mehrfach abgestufte) kategoriale Variable und sind Kohorten Gruppen von Personen, die einen bestimmten ausgewählten Zeitpunkt oder Zeitraum erlebt haben – was mit der Annahme linearer Kohorteneffekte kaum in Einklang zu bringen ist. Hinsichtlich der oben dargelegten Vermutungen über den Einfluß der Erfahrung problematischer historischer Umstände ist eine solche kategoriale Auffassung des Kohortenbegriffs nur auf den ersten Blick naheliegend und unproblematisch, auf den zweiten Blick wäre beispielsweise zu fragen: Welches konkrete Datum bezeichnet genau das Ende der besagten schlechten Zeiten? Das Kriegsende z.B. markiert nicht den tatsächlichen Be-

ginn besserer Zeiten, sondern eben der Nachkriegszeit, die von den heute Älteren durchaus als eine durch Entbehrungen und Defizite geprägte geschildert wird, und die dort begonnene Entwicklung, die in die gegenwärtige Wohlstandsgesellschaft mündet, hat sich kontinuierlich, und nicht im abrupten Wechsel vollzogen. Und auch wenn ein Cutoff-Wert für die Zugehörigkeit zu der Kohorte, für die der besagte „Schlechte-Zeiten-Effekt“ wirksam sein sollte, festgelegt würde, so wäre darin eine enorme Vereinfachung und Ungenauigkeit enthalten, denn es würden sich dann ja auch die Kohortenmitglieder wiederum darin unterscheiden, daß sie unterschiedlich lange Lebensspannen in den schlechten Zeiten verbracht haben.

Derartige Probleme verschwinden bei konsequenter Verwendung des Kohortenbegriffs: „Cohort effects represent the impact of historical influences upon a group of individuals that share similar environmental circumstances at equivalent temporal points in their life course“ (Schaie & Hofer, 2001, 55). Diese Definition widerspricht zwar nicht der oben dargestellten verbreiteten Auffassung von an historischen Zeitpunkten orientierten Kohortendefinitionen, wenn man sie jedoch konsequent zuende denkt, dann führt dieses zu dem Geburtszeitpunkt der Personen als der eigentlichen „Kohortenvariablen“, denn der Einfluß historischer Ereignisse *zu demselben Zeitpunkt des Lebenslaufs* setzt zwingend voraus, als diese Personen zu ungefähr derselben Zeit geboren wurden. Tatsächlich ist es die Varianz der Geburtszeitpunkte, die die mit dem Begriff des Kohorteneffekts eigentlich gemeinte Variation erlebter historischer Zeitpunkte beinhaltet. Die sozusagen natürlichen Kohorten sind also die Gruppen gleichzeitig Geborener, jedoch ist auch die Messung des Geburtszeitpunkts nicht beliebig genau, sondern kann nur über die Zuordnung zu den durch die Einheiten der Geburtszeitpunktskala erzeugten Zeitintervallen erfolgen. Die gebräuchlichste Zeiteinheit der Geburtszeitmessung ist die Skala der Kalenderjahre, es mag aber auch für bestimmte Zwecke genügen, beispielsweise in Jahrzehnten zu messen, so daß daraus also Zehn-Jahres-Geburtsjahrgangskohorten entstünden, – die Frage der Definition des Kohortenbegriffs driftet hier sozusagen ins Technische ab und wird wieder in Kapitel 2.1, wo die Methode zur Untersuchung der hier zu erläuternden Hypothese dargestellt wird, aufgenommen. Wichtig ist hier vorerst nur, daß die Kohortenzugehörigkeit im Prinzip und sozusagen kontraintuitiv eine kontinuierlich ausgeprägte Variable ist, so daß es auch durchaus sinnvoll sein *kann*, einen Kohorteneffekt z.B. als linearen Effekt des Geburtsjahres zu konzipieren.

Genau diese Konzeption aber wird bei der in Kapitel 2.1 dargestellten Analysemethode zugrundegelegt. Die Linearität stellt hierbei eine sehr weitgehende Annahme dar, auch wenn damit natürlich nur eine näherungsweise Linearität im Sinne eines mit der Fortschaltung des Geburtsjahres einhergehenden *ungefähren* linearen Trends gemeint sein kann. Sie ist für die hier anstehende theoretische Diskussion so zu übersetzen: Es wird angenommen, daß die gesamte erste Hälfte des zwanzigsten Jahrhunderts bis in eine Phase allmählicher gesellschaftlicher Fortentwicklung nach dem Kriege hinein als im Vergleich zur Gegenwart schlechtere Zeiten im obengenannten Sinne aufzufassen sind und daß die Dauer der in diesen Zeiten verbrachten Zeit einen positiven Effekt auf die Zufriedenheiten ausübt. Dieses

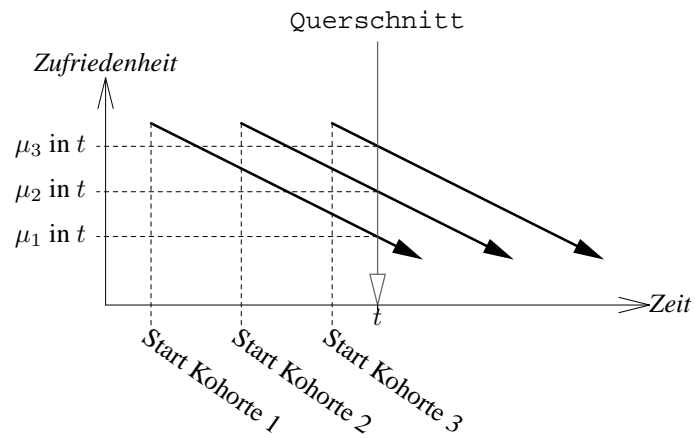
ist natürlich ein stark vereinfachtes Modell der eigentlich gemeinten Zusammenhänge, doch es erscheint unter den gegebenen Umständen als hinreichend genau – solange nämlich keine genauere Annahme über singuläre historische Ereignisse voraussetzbar ist, welche auf alle Mitglieder der jeweils untersuchten Population gleichermaßen einen besonders starken Einfluß auf spätere Zufriedenheitsbewertungen ausgeübt haben sollten. Zudem muß nicht angenommen werden, daß ein solcher linearer Trend zu über die Geburtszeitpunkte hinweg sinkender Zufriedenheit tatsächlich über das gesamte Spektrum der Geburtszeitpunkte hinweg besteht: Es wäre aus den obigen Überlegungen, die die Annahmen dieses Trends begründen, auch zu folgern, daß ein solcher Trend sozusagen zur Gegenwart hin abflaut: Den Nachkriegsgenerationen ist gemeinsam, daß sie praktisch die gesamte Lebenszeit unter Wohlstandsbedingungen verbracht haben, zwischen den Geburtsjahrganggruppen aus diesem Spektrum variiert sozusagen der eigentliche Prädiktor des Kohorteneffekts – die unter vergleichsweise rückschrittlichen sozialen Bedingungen verbrachte Zeit – nicht mehr.

Wenn also – zumindest bei den in der ersten Hälfte des zwanzigsten Jahrhunderts Geborenen – ein solcher Trend zu mit steigendem *Geburtsjahr* sinkender Zufriedenheit bestünde, so könnte dessen Überlagerung mit einem Trend zu mit wachsendem *Lebensalter* sinkender Zufriedenheit dazu führen, daß in querschnittlichen Vergleichen von Altersgruppen deren mittlere Zufriedenheiten sich kaum noch unterscheiden. Schematisch ist dieses in Abbildung 1.1 dargestellt. Dort repräsentieren die fettgedruckten Pfeile die Verläufe der durchschnittlichen Zufriedenheiten dreier fiktiver Geburtskohorten über die Zeit hinweg. In Schema A ist ein negativer Alterseffekt ohne Kohorteneffekt abgebildet: Die Mittelwerte aller drei Kohorten sinken über die Lebenszeit kontinuierlich, es gibt aber keine Unterschiede in deren Zufriedenheitsniveau (sie starten alle auf demselben Zufriedenheitslevel – inhaltlich wäre es natürlich wenig sinnvoll, diesen Startpunkt beim Geburtsjahr der Kohorte festzulegen, für die hier angestellten Überlegungen könnte man beispielsweise annehmen, der Startpunkt sei eine festgesetzte Altersschwelle im mittleren Erwachsenenalter). Würden zu irgendeinem Zeitpunkt t die durchschnittlichen Zufriedenheiten (μ_1 , μ_2 und μ_3) der Gruppen verglichen, so würde der Alterseffekt entdeckt (allerdings wäre ohne weitere Information nicht klar, ob die Mittelwertsunterschiede nicht durch einen Kohorteneffekt zustande gekommen sind, da die drei Gruppen sich sowohl im Alter, als auch im Geburtsjahr unterscheiden). In Schema B ist entsprechend ein negativer Kohorteneffekt ohne Alterseffekt abgebildet: Je höher das Geburtsjahr der Gruppe, desto niedriger die durchschnittliche Zufriedenheit, jedoch verändern sich die Durchschnittswerte der Gruppen über die Zeit, d.h. mit deren Altern, nicht. Beim querschnittlichen Vergleich würden wiederum die Unterschiede entdeckt, die entweder korrekt als Kohorteneffekt oder fälschlich als (positiver) Alterseffekt interpretiert werden könnten. Schema C schließlich zeigt die oben erörterte Möglichkeit einer Überlagerung von negativem Alters- und Kohorteneffekt: Je höher das Geburtsjahr der Gruppe, desto niedriger ist deren allgemeines Zufriedenheitsniveau, d.h. die durchschnittliche Zufriedenheit zu irgendeinem bestimmten Lebensalter (z.B. beim hier abgebildeten Startpunkt). Gleichzeitig fallen die Verläufe infolge eines Alterseffekts, so

ABBILDUNG 1.1:
Schematische Darstellung der Kohorteneffekthypothese

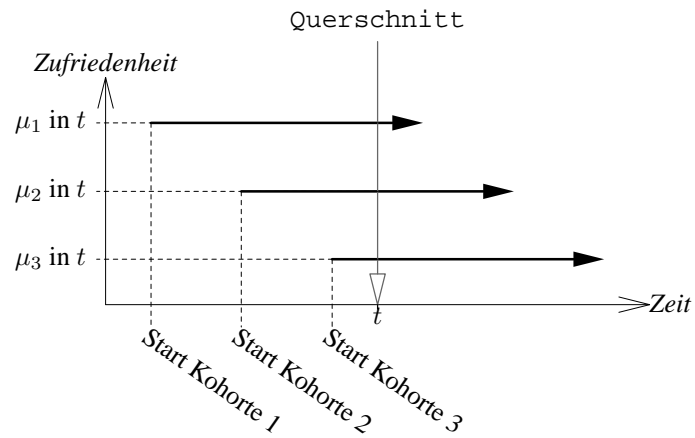
Schema A:

Negativer Alterseffekt,
kein Kohorteneffekt.



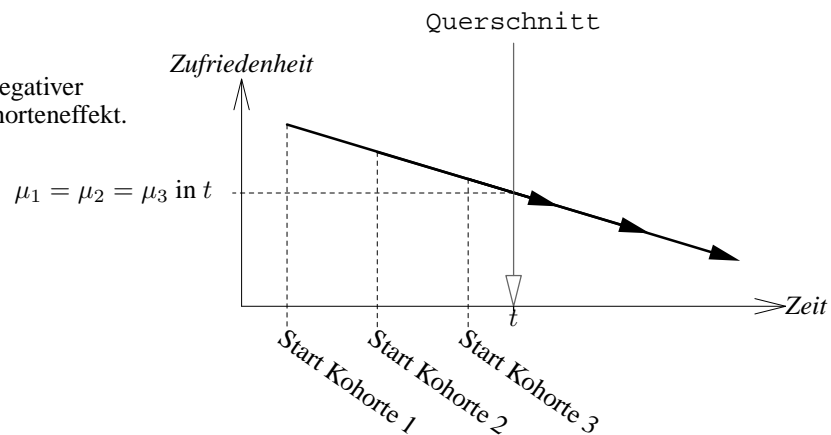
Schema B:

Negativer Kohorteneffekt,
kein Alterseffekt.



Schema C:

Überlagerung negativer
Alters- und Kohorteneffekt.



daß es zu einer Überlagerung kommt und die Gruppenmittelwerte zum Zeitpunkt t eines querschnittlichen Vergleichs sich nicht unterscheiden – es würde weder der Alters-, noch der Kohorteneffekt entdeckt, sondern der Eindruck entstehen, daß die durchschnittlichen Zufriedenheiten von derartigen über die Zeit variablen Effekten unbeeinflusst sind.

Allerdings erscheint es immerhin merkwürdig, daß eine solche Vermutung von Kohorteneffekten allem Anschein nach bislang kaum explizit formuliert und untersucht wurde: Auch eine intensive diesbezügliche Literatursuche förderte kaum entsprechende Ergebnisse zutage. Immerhin findet sich bei Lehr (1991, 296) ein Hinweis in diese Richtung und erklärten in einer taiwanesischen Längsschnittstudie kohortenspezifische Lebenserfahrungen einen Anstieg der Lebenszufriedenheit nach Ausparialisierung von Prädiktoren von Zufriedenheitsverlusten (z.B. alterstypischer Einkommensverluste) in einer Art und Weise, die der oben dargelegten Argumentation entspräche (Chen, 2001). Abgesehen davon muß die Vermutung von Kohorteneffekten nicht darum falsch sein, weil sie bislang wenig beachtet wurde, sondern begründet im Gegenteil dieser Umstand einen Forschungsbedarf nach der Analyse möglicher Kohorteneffekte.

1.2.3.2 Zufriedenheitsgewinner und Zufriedenheitsverlierer?

Von der Möglichkeit abgesehen, daß die Befunde stabiler durchschnittlicher Zufriedenheiten alter Menschen auf Kohorteneffekte zurückzuführen sind, wäre zu fragen, ob diese Stabilität auf derjenigen des Merkmals Zufriedenheit im höheren Lebensalter beruht. Mit anderen Worten: Besteht bei alten Menschen eine geringe intraindividuelle Varianz der Lebenszufriedenheit im zeitlichen Verlauf, d.h. wird ein jeweils individuell spezifisches Zufriedenheitsniveau weitestgehend aufrechterhalten auch angesichts alterstypischer Verlust Erfahrungen und Probleme? Es geht hierbei um jeweils intraindividuelle *Veränderung*, die in den Durchschnittswerten von Altersgruppen im Grunde nicht vollständig erfaßt wird, und zwar nicht nur nicht in querschnittlichen Vergleichen, sondern auch in längsschnittlichen dann nicht, wenn dort lediglich die Entwicklung der Mittelwerte einer Personengruppe über einen gewissen Zeitverlauf hinweg beobachtet wird. Möglicherweise nämlich beinhalten derartige Ergebnisse eine gewisse Vielfalt unterschiedlicher Veränderungsverläufe, möglicherweise verdeckt die Stabilität der Durchschnittswerte im Altersverlauf einen bedeutsamen Anteil von Fällen, bei denen tatsächlich eine Verschlechterung der Lebenszufriedenheit stattgefunden hat. Immerhin aber müßten, wenn dem so wäre, diese Verschlechterungen durch viele „Gewinner“ mit verbesserter Zufriedenheit so weit nivelliert werden, daß insgesamt im Vergleich zu den Werten jüngerer Personen der altersspezifische Durchschnittswert stabil bleibt. Somit wäre, wenn die Altersgruppenmittelwerte auch im längsschnittlichen Verlauf stabil wären, entweder anzunehmen, daß Zufriedenheit im höheren Alter weitestgehend intraindividuell stabil aufrechterhalten wird, oder aber daß Altersverläufe auch zu einem nicht unerheblichen Teil durch verbessertes Wohlbefinden – zumindest was die Selbstbeurteilung der eigenen Lage anbelangt – gekennzeichnet sind.

Theoretische Gründe für die letztere Annahme finden sich in der gerontologischen Literatur zum SWB älterer Menschen. Die Diskussion um das „stability-despite-loss paradox“ hat verschiedentlich zur Hypothese einer mit dem höheren Lebensalter erhöhten Kompetenz zu selbstregulatorischer Aufrechterhaltung bzw. Optimierung affektiver Wohlbefindenskomponenten geführt, die auch durch entsprechende empirische Befunde gestützt werden konnte (Lawton et al., 1992; Labouvie-Vief, 1998; Kunzmann et al., 2000). Alles in allem und von möglichen negativen Lebensereignissen abgesehen, wäre höheres Alter demnach sogar ein positiver Prädiktor für verbessertes Affekterleben: „What might be the gains that come with old age and potentially enhance subjective well-being? Richer knowledge about our own emotions and those of the people around us, deeper understanding of the situations and factors that are relevant to our emotional lives, and the accumulation and elaboration of strategies to regulate emotional experience are promising candidates for future research investigations. If it were possible to become older without encountering health problems and limitations in physical functioning, then subjective well-being would likely improve as we gain more and more emotional competence“ (Kunzmann et al., 2000, 524). Der letzte Satz enthält gewissermaßen auch eine wichtige theoretische Setzung bezüglich der kognitiven Wohlbefindenskomponente im Alter: Eine im Alter verbesserte Kompetenz zu affektivem Wohlbefinden müßte infolge affektiver Inferenz auch tendenziell verbesserte Zufriedenheiten bewirken. Es bedeutet demnach eine starke Simplifizierung, wenn Höheralterigkeit lediglich als problematische Bedingung für Lebenszufriedenheit angesehen wird, vielmehr stellen Verlusterlebnisse und Verschlechterungen objektiver Lebensumstände, die im hohen Alter gehäuft — aber nicht in jedem Fall und nicht während dieses gesamten Lebensabschnittes mit gleichermaßen hoher Frequenz — auftreten, gewissermaßen nur die eine, negative Seite der Medaille dar. Auf der anderen, positiven Seite könnte fortgeschrittenes Lebensalter per se ein Prädiktor für verbessertes affektives Wohlbefinden sein, so daß auch für Lebenszufriedenheit ein positiver Alterseffekt zu erwarten wäre, der sozusagen in Konkurrenz zu erwartbaren negativen Effekten alterstypischer Einbußen stünde und der umso stärker wirken könnte, je weniger solcher Einbußen tatsächlich stattfinden.

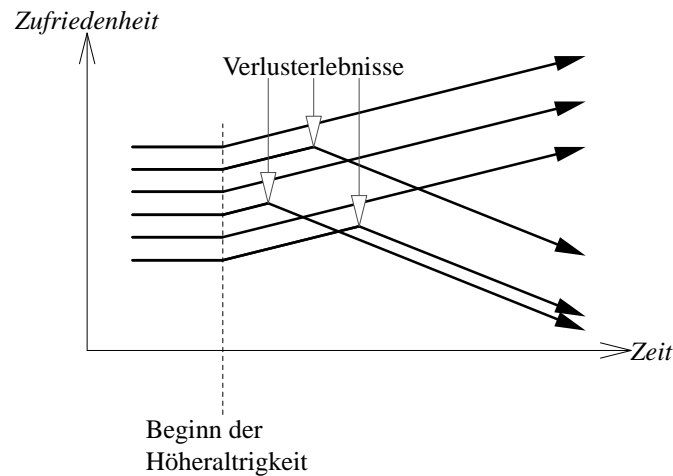
Von den besagten Entwicklungen „emotionaler Kompetenz“ abgesehen, ist auch ein „kognitiver“ Effekt des höheren Lebensalters denkbar, der Zufriedenheitsurteile positiv beeinflusst: Wiederum unter Rekurs auf das in Kapitel 1.1.4.3 dargestellte Konzept der Adjustierung internaler Standards für den Vergleich mit objektiven Lebensbedingungen könnte vermutet werden, daß negative Altersstereotype bzw. ganz allgemein Erwartungen alterstypischer Verluste bei den Betroffenen selbst eine wichtige Informationsquelle für die Festlegung der internalen Soll-Zustände zur Bewertung der eigenen Lebenslage darstellen. Mit anderen Worten: Wir alle erwarten gewisse Verschlechterungen unseres Lebens mit dem Älterwerden und möglicherweise sind diese Erwartungen übertrieben, so daß für viele von uns ein teilweises Ausbleiben dessen, was wir befürchten, zur positiven Überraschung werden wird – während andere, die die erwarteten Verschlechterungen erleben, nicht negativ, sondern überhaupt nicht überrascht sein werden. Es könnte also gerade auch aus den theo-

retischen Annahmen, die in Kapitel 1.1.4.3 unter dem Begriff der Relativität von Zufriedenheitsurteilen zusammengefaßt wurden, ein positiver Alterseffekt dergestalt abgeleitet werden, daß im höheren Alter negative Erwartungen und/oder Altersstereotype zu entsprechend niedrig angesetzten Soll-Zuständen bezüglich relevanter Lebensbereiche führen, die dann eher über- als unterboten werden können. Zudem könnte es noch zu einer stärkeren Gewichtung derjenigen Lebensumstände, in denen eine solche positive Überraschung festgestellt wird, kommen. Schlechtere Zufriedenheit wäre demnach nur für diejenigen Personen zu erwarten, die von alterstypischen Verlusten in besonders starkem, überdurchschnittlichem Maße betroffen sind.

Tatsächlich zeigen empirische Befunde zu impliziten Entwicklungstheorien eine allgemein vorhandene Erwartung eines mit dem Lebensalter sich verschlechternden Verhältnisses von Verlusten und Gewinnen (Heckhausen et al., 1989). Solche stereotypen Erwartungen wirken handlungsregulierend in dem Sinne, daß sie die persönliche Lebensplanung und Lebensführung beeinflussen (Brandtstädter, 1990) und bilden darüber hinaus „normative Entwicklungs- und Lebensvorstellungen, vor deren Hintergrund wahrgenommene oder antizipierte Entwicklungsergebnisse erst bewertet und als ‚Gewinne‘ oder ‚Verluste‘ verbucht werden können“ (ebd., 339). Wenn man also eine solchermaßen grundsätzliche Entwicklungsrelevanz von Altersnormen und impliziten Entwicklungstheorien bezüglich des höheren Lebensalters voraussetzt, so ist geradezu zwangsläufig anzunehmen, daß diese Bedeutung sich in der oben vermuteten Art und Weise, d.h. bei der Adjustierung interner Soll-Zustände, auch auf Zufriedenheitsurteile erstreckt: Aus der stereotypen Erwartung einer negativen Gewinn-Verlust-Bilanz resultieren „bescheidene“ Entwicklungsziele für das höhere Alter, welche als entsprechend niedrige Soll-Zustände bei der Generierung von Zufriedenheitsurteilen fungieren oder zumindest die Festsetzung derselben beeinflussen können.

Aus solchen Überlegungen resultiert also ein differenziertes Bild der Lebenszufriedenheit im höheren Alter, das nicht auf negative Aspekte fixiert bleibt, sondern v.a. die Varianz von Alternsprozessen berücksichtigt: Es ist durchaus anzunehmen, daß viele ältere Menschen unter negativen Entwicklungen ihrer Lebensumstände „leiden“, daß sie sich schlecht *fühlen*, wenn beispielsweise ihr Lebenspartner stirbt oder sie mit schwerwiegenden gesundheitlichen Einschränkungen konfrontiert werden, und daß sie dann, wenn sie sich schlecht *fühlen*, auch nicht allzu zufrieden mit ihrem Leben sein werden. Es ist auch anzunehmen, daß alte Menschen die aus solchen Ereignissen resultierenden schwerwiegenden Veränderungen ihres Alltagslebens bei „vernünftiger“ und gewissermaßen emotionsloser Einschätzung negativ bewerten und entsprechend unzufriedener werden. Jedoch sind nicht alle gleichermaßen von derartigen Negativerfahrungen betroffen und ist auch nicht die gesamte Spanne des höheren Alters eine ununterbrochene Abfolge derselben. Ältere Menschen mit relativ (zu ihrer Altersgruppe) durchschnittlicher oder gar überdurchschnittlicher *objektiver* Lebensqualität fühlen sich möglicherweise besser, als sie sich in jüngerem Alter in derselben Lage gefühlt hätten.

ABBILDUNG 1.2:
Schematische Darstellung der Gewinner-und-Verlierer-Hypothese



Das daraus resultierende Bild differentieller Zufriedenheitsverläufe mit Gewinnern und Verlierern ist in starker Vereinfachung schematisch in Abbildung 1.2 dargestellt: Mit dem Übergang in die Höheraltrigkeit steigt die Zufriedenheit (Gewinner) – solange, bis Verlust erlebt wird, dann aber fällt sie. Es ist also immerhin denkbar, daß es im höheren Lebensalter hinsichtlich der Lebenszufriedenheit nicht nur Verlierer, sondern auch Gewinner geben könnte, so daß auch die Möglichkeit bestünde, daß die solchermaßen unterschiedlichen Zufriedenheitsverläufe sich zu jedem Zeitpunkt auf einem gleichbleibenden Niveau ausmitteln – dieses würde dann zu den „paradox“ stabilen Zufriedenheitsmittelwerten im Altersgruppenvergleich führen, ohne daß die Zufriedenheit selbst entsprechend stabil wäre.

1.2.3.3 Lebenslange Stabilität von Zufriedenheit?

In Veröffentlichungen zum Zufriedenheitsparadox wird meist explizit oder implizit die Mittelwertstabilität mit der Merkmalsstabilität gleichgesetzt bzw. von ersterer auf letztere geschlossen. Diese Schlußfolgerung kann zwar die zitierten Befunde erklären, umgekehrt jedoch kann sie keineswegs als allein durch dieselben empirisch gesichert gelten. Vielmehr bedarf sie noch ausreichender empirischer Bestätigung: Es muß untersucht werden, ob über einen großen Teil der hochaltrigen Lebensspanne hinweg tatsächlich eine hohe *intraindividuelle* Merkmalsstabilität von Zufriedenheit besteht. Ein solcher Befund kann, wie vorne bereits gesagt wurde, nur in längsschnittlicher Untersuchung und nicht allein auf der Basis des Vergleichs altersspezifischer Durchschnittswerte gewonnen werden. In Kapitel 1.2.2 wurden entsprechende Befunde kurz dargestellt bzw. daraus das Fazit abgeleitet, daß die tatsächliche intraindividuelle Merkmalsstabilität von Zufriedenheit bislang nur unzurei-

chend empirisch erforscht ist.

Theoretische Gründe für die Annahme einer relativ hohen intraindividuellen Stabilität von SWB *nicht nur* bei hochaltrigen Menschen wurden bereits gegeben: Die trait-theoretischen Ansätze in der Wohlbefindenstheorie postulieren diese Stabilität geradezu als Wesensmerkmal des SWB (vgl. Kapitel 1.1.4.4). Es wurde bereits dargelegt, daß aus dem „Stand der Dinge“ in der Wohlbefindensforschung v.a. bezüglich affektiver Wohlbefindenskomponenten ein starker Einfluß stabiler Persönlichkeitsmerkmale anzunehmen ist, während eine solche Trait-Abhängigkeit der Lebenszufriedenheit in der Zusammenschau diesbezüglicher Veröffentlichungen nicht ebenso deutlich erscheint. Immerhin aber impliziert die Annahme affektiver Inferenzen bei der Bildung von Zufriedenheitsurteilen auch eine mittelbare stabilisierende Wirkung derjenigen Persönlichkeitsdimensionen, von denen affektives Wohlbefinden in starkem Maße beeinflußt scheint.

Zudem sollten auch die in Kapitel 1.1.4.3 diskutierten kognitiven Relativierungsmechanismen bei der Generierung von Zufriedenheitsurteilen zwangsläufig eine gewisse Stabilität bewirken: Diese bedeuten Adaptation dergestalt, daß infolge – negativer oder positiver – Veränderungen objektiver Lebensumstände eine entsprechende Angleichung subjektiver Bewertungsstandards stattfindet, die den Effekt objektiver Verluste oder Gewinne auf SWB gewissermaßen abfedert oder ausgleicht. Die Theorie des SWB liefert somit starke Gründe für die Annahme, daß Zufriedenheiten bis zu einem gewissen Maße über das gesamte Erwachsenenalter hinweg intraindividuell stabil sind und auch bleiben, wenn es am Ende dieser Lebensspanne vermehrt zu negativen Erfahrungen kommt.

Wenn das Zufriedenheitsparadox des höheren Alters durch Adaptationseffekte erklärt werden soll, dann ist v.a. auch zu berücksichtigen, daß die alterstypischen Verluste bzw. Verschlechterungen objektiver Lebensumstände oft nicht als langsam und allmählich ablaufende Entwicklung, sondern als kurzfristig sich vollziehende und somit intensive Veränderung der alltäglichen Lebenssituation erlebt werden. Typische Beispiele wie der Austritt aus dem Erwerbsleben, der Tod eines nahen Angehörigen oder der Ausbruch einer schweren Krankheit bedeuten für die betroffene Person gravierende Einschnitte, die deren Leben gewissermaßen „von einem Tag auf den anderen“ grundlegend verändern. D.h. also, daß viele alterstypische Verluste aus kritischen Lebensereignissen resultieren, und es ist kaum anzunehmen, daß SWB dadurch nicht beeinflußt würde. Vielmehr ist zu fragen, wie solche Ereignisse verarbeitet bzw. bewältigt werden können, und wie sich SWB in solchen Verarbeitungsprozessen entwickelt. Es geht also dabei nicht nur um die unmittelbaren Effekte der kurzfristigen problematischen Veränderungen, die im höheren Alter häufig stattfinden, sondern v.a. um deren längerfristige, dauerhafte Auswirkungen auf das Wohlbefinden der davon Betroffenen, – um Adaptation in der Folge von Verlusterlebnissen, die einen vorübergehenden Einbruch des SWB bewirken können.

Adaptivität des SWB und insbesondere der Lebenszufriedenheit wurde vielfach theoretisch begründet und empirisch gezeigt (vgl. Kapitel 1.1.4.3). Es erscheint darum geradezu als Selbstverständlichkeit, daß Menschen jeglichen Alters auf Verschlechterungen ihrer Le-

bensumstände – ob sie nun abrupt und krisenhaft eintreten, oder als langwierige, allmähliche Veränderungen – mit einem gewissen Maß an Adaptation reagieren, welches Einbußen ihres SWB entgegen- bzw. dessen zumindest teilweise Aufrechterhaltung bewirkt. In der Gerontologie bzw. der Entwicklungspsychologie der Lebensspanne wurde dieses ausgiebig unter dem begrifflichen Label der „Resilienz“ thematisiert und mit entsprechenden Resilienzmodellen beantwortet. Gemeint ist psychologische Widerstandskraft, die in gravierenden Krisensituation zur Aufrechterhaltung bzw. Wiederherstellung eines (positiven) SWB befähigt, und das Interesse gilt hier den psychologischen Ressourcen der Person, die hierfür funktional sein können. Die oben zitierte emotionale Kompetenz älterer Menschen kann als ein Beispiel für eine derartige psychische Ressource zur Stützung der affektiven Wohlbefindenskomponente angesehen werden. Es soll hier kein umfassender Überblick über die Resilienzforschung gegeben werden (siehe z.B. Staudinger et al., 1995; Tesch-Römer et al., 1997), sondern lediglich auf die explizite theoretische Begründung einer sozusagen zielgerichtet protektiven Adaptivität des SWB hingewiesen werden, die in diesen Resilienzmodelle enthalten ist: Es ist grundsätzlich anzunehmen, daß kurzfristige Einbußen des SWB infolge objektiver Verluste durchaus adaptiv reguliert bzw. zurückgewonnen werden können – wenn auch, je nach den jeweiligen persönlichen Ressourcen der Betroffenen einerseits und natürlich der tatsächlichen Schwere der zu bewältigenden Veränderung andererseits, in jeweils unterschiedlichem Maße.

Wenn man also annimmt, daß im hohen Alter häufig abrupte Verschlechterungen der objektiven Lebensqualität eintreten, so kann auch durchaus davon ausgegangen werden, daß ältere Menschen sich häufig subjektiv unwohl befinden, d.h. während solcher akuter kritischer Phasen sowohl verstärkt negative Emotionen durchleben, als auch ihr Leben eher negativ bewerten. Es liegt jedoch auf der Hand, daß die Befunde einer angesichts solcher häufigen negativen Ausschläge des SWB „paradox“ erscheinenden hohen Durchschnittszufriedenheit älterer Menschen auch mit deren Resilienz, d.h. mit einer protektiven Adaptation der subjektiven Bewertungen bei alterstypischen Verlusterlebnissen erklärbar sein könnten: Längerfristig würden dadurch die negativen Effekte vieler Verlusterlebnisse gewissermaßen abgemildert bzw. gänzlich annulliert, so daß sie in Durchschnittswerten von Stichproben höheraltriger Personen nur vermittels derjenigen Fälle zu Buche schlagen, die zum Befragungszeitpunkt aktuell von solchen Negativereignissen betroffen sind. Der Anteil dieser Fälle dürfte aber jeweils eher gering sein, sofern der Modus der Stichprobengenerierung nicht eine disproportional hohe Wahrscheinlichkeit der Auswahl von Personen in solchen kritischen Phasen bedingt (z.B. durch das Sampling von Personen, die sich in stationärer Krankenhausbehandlung befinden). Wenn überhaupt, so können für Zufallsstichproben Älterer aus nicht weiter selektierten Grundgesamtheiten wohl eher positive Selektivitätseffekte vermutet werden, insofern gerade die aktuelle Konfrontation mit Negativereignissen eine Ursache für Teilnahmeverweigerung sein könnte.

Insgesamt also liefern sowohl die trait-theoretischen Sichtweisen des SWB, als auch die Annahmen hoher Adaptivität von Zufriedenheit gute Gründe dafür, Zufriedenheit als

ein Merkmal anzusehen, welches in hohem Maße intraindividuell stabil aufrechterhalten wird. Diese Gründe gelten prinzipiell für alle Altersgruppen, d.h. es kann deshalb vermutet werden, daß Zufriedenheit sozusagen *lebenslang* hoch stabil ist. Diese lebenslang gleichermaßen hohe Stabilität könnte ein Ausmaß an Unabhängigkeit der kognitiven SWB-Komponente von objektiven Umständen erzeugen, welches in den erwähnten Befunden zu diesem Zusammenhang im Alter nur besonders deutlich zutage treten kann, weil dort eine gewisse Divergenz zwischen der Instabilität sich häufig abrupt verschlechternder Lebensumstände und der Stabilität ihrer subjektiven Bewertungen besteht.

1.2.3.4 Erhöhte Stabilität der Zufriedenheit im Alter?

Falls das Zufriedenheitsparadox des höheren Lebensalters auf intraindividuelle Merkmalsstabilität zurückzuführen ist, so wäre zu fragen, ob sich die Höheraltrigen in dieser Hinsicht von den Jüngeren unterscheiden, d.h. ob Zufriedenheit im Alter *stabiler* wird. Die Annahme stabilisierender Einflüsse von Persönlichkeitsfaktoren (vgl. Kapitel 1.1.4.4) und die Annahme adaptiver Stabilisierung von Zufriedenheit gelten prinzipiell für alle Altersgruppen und sind ohne weiteres nicht als Argumente für eine im Alter erhöhte Merkmalsstabilität zu gebrauchen.

Ein Aspekt, der in dieser Hinsicht jedoch bedeutsam sein könnte und m.E. im Zusammenhang mit der Lebenszufriedenheit im höheren Lebensalter wenig beachtet wurde (vgl. z.B. die Zusammenfassung von Staudinger, 2000), könnte die möglicherweise sich verändernde subjektive zeitliche Orientierung bzw. Zeitperspektive der alternden Personen sein. Es erscheint durchaus plausibel, daß Menschen, wenn sie solche Bewertungen vornehmen, nicht nur über die Summe ihrer gegenwärtigen Lebensumstände hinweg generalisieren, sondern sozusagen auch retrospektiv über die Zeitachse, so daß ihr Urteil im Extremfall eine subjektive Bilanz ihrer gesamten Lebensspanne, und nicht nur der aktuellen Episode enthält. Diese retrospektive Sichtweise könnte gerade für ältere Menschen besonders wichtig sein, die ja gewissermaßen den größeren Teil ihrer Lebenszeit bereits verbraucht haben und denen nur noch wenig Zeit zum Erreichen persönlicher Selbstverwirklichungs- oder Entwicklungsziele bleibt. Deshalb könnte für die Älteren die Frage, was in der Vergangenheit verwirklicht oder verpaßt wurde, mehr Gewicht haben, als bei Jüngeren, welche zumindest subjektiv noch über mehr Zukunft verfügen und folglich bei der Bewertung ihres Lebens weniger eindeutig auf tatsächlich realisierte Qualitäten verwiesen sind, sondern hierbei auch die aktuelle Situation unter dem Aspekt der darin enthaltenen Möglichkeiten für zukünftige Lebensqualität in Betracht ziehen können. Für eine retrospektive Bilanz über längere Lebenszeiträume hinweg ist aber möglicherweise selbst eine aktuelle Katastrophe nur von untergeordneter Bedeutung: Beispielsweise wird vielleicht ein Mensch, der auf ein erfolgreiches und erfülltes Leben, in dem ihm wenig Schlimmes widerfahren ist, zurückblickt, dieses insgesamt auch dann noch positiv bewerten, wenn er aktuell von schwerer Krankheit oder dem Tod des Lebenspartners betroffen ist. Die Annahme einer solchen „Re-

trospektivität“ älterer Menschen führt direkt zu der Vermutung, daß mit dem Lebensalter die Merkmalsstabilität der Lebenszufriedenheit wächst, weil gewissermaßen das Objekt der Bewertung bei Älteren – nämlich eben v.a. das Leben, das sie gelebt *haben* – zum größten Teil feststeht und in der Rückschau durch aktuelle, kurzfristige Ereignisse kaum noch verändert wird. Wenn aber das Merkmal selbst bei den Älteren weitgehend stabil bzw. nur gering veränderlich ist, so folgt daraus auch Stabilität in den Durchschnittswerten von Altersgruppen aus dem Segment der Höheraltrigen.

Befunde zur subjektiven zeitlichen Orientierung bzw. Zeitperspektive liefern nicht immer Hinweise darauf, daß tatsächlich ältere Menschen in stärkerem Maße vergangenheitsorientiert sind, als jüngere (Fingerman & Perlmutter, 1995; Shmotkin 1991). Eine gewisse Uneindeutigkeit der Befundlage könnte hier schlicht das Resultat unterschiedlicher Operationalisierungen zeitlicher Orientierung in den ohnehin nicht sehr zahlreichen Untersuchungen sein. So scheint es mir zumindest fraglich, ob beispielsweise Selbsteinschätzungen der Häufigkeit vergangenheits- oder zukunftsbezogener Gedanken, wie sie von Fingerman & Perlmutter (1995) verwendet wurden, dasselbe erfassen, wie die in einem Satzergänzungsverfahren inhaltsanalytisch gewonnenen Häufigkeiten solcher lebenszeitbezogener Kognitionen in der Untersuchung von Dittmann-Kohli (1995). Letztere fand zum Teil sehr deutliche Unterschiede in der Frequenz lebenszeitbezogener Kognitionen beim Vergleich zweier Stichproben junger (16 bis 34) und alter (60 bis 90) Erwachsener, die für die Jungen eine sehr starke Zukunfts- und eine geringe Vergangenheitsorientierung, für die Alten dagegen genau das gegenteilige Schema, d.h. einen weitgehenden kognitiven Vergangenheitsbezug bei geringer ausgeprägter Zukunftsperspektive, anzeigen. Hinsichtlich der oben dargestellten Überlegungen zur Bedeutung von „Retrospektivität“ für die Bildung von Urteilen der Lebenszufriedenheit liefert allerdings die bloße Häufigkeit, mit der an Vergangenes gedacht wird, im Grunde nur Anhaltspunkte, es geht dabei eigentlich um die Frage des Gewichts, das solche Inhalte für die Selbstbeurteilung des eigenen Lebens besitzen.⁴

Eine direkte Untersuchung dieser Frage versuchte Shmotkin (1991) mittels Strukturgleichungsmodellen, in denen Lebenszufriedenheit als latenter Faktor modelliert wurde, der drei beobachteten Indikatoren zugrundeliegt, nämlich der retrospektiven Bewertung des eigenen Lebens vor fünf Jahren, der Bewertung des eigenen Lebens zum gegenwärtigen Zeitpunkt und der prospektiven Bewertung des eigenen Lebens in fünf Jahren. Durch den Vergleich der drei Faktorladungen kann dann die relative Wichtigkeit jedes Indikators (eigentlich: die jeweilige Korrelation des Indikators mit dem Lebenszufriedenheitsfaktor)

⁴Stock et al. (1986) trugen dieser Frage dadurch Rechnung, daß sie eine konzeptionelle Differenzierung des Konstrukts SWB hinsichtlich des zeitlichen Fokus entsprechender Urteile vorschlugen: Das Subkonstrukt „life satisfaction“ wäre demnach exklusiv Selbstbeurteilungen der Vergangenheit vorbehalten, während „happiness“ auf die Gegenwart und „morale“ auf die Zukunft bezogene Urteile über SWB umfaßt. Dieser interessante Vorschlag deckt sich begrifflich – v.a. in der Benennung des vergangenheitsfokussierten SWB als Lebenszufriedenheit – allerdings nur unzureichend mit dem in der SWB-Forschung insgesamt vorherrschenden Zufriedenheitskonstrukt, welches im Allgemeinen den temporalen Bezug vernachlässigt oder zumindest nicht explizit mitberücksichtigt bzw. implizit Lebenszufriedenheit als Urteil über die aktuelle Situation auffaßt.

beurteilt werden. Auf diese Weise wurde also untersucht, wie eng eine allgemeine, „zeitlich indifferente“ Lebenszufriedenheit, als „common factor“ retrospektiver, aktueller und prospektiver Bewertungen, an eben diese jeweils gekoppelt ist, und es wurden in dieser Hinsicht Altersgruppen (querschnittlich) verglichen: Die größte Bedeutung besaß in allen Altersgruppen die Bewertung des gegenwärtigen Lebens, dagegen war die Wichtigkeit retrospektiver Bewertungen bei den jüngeren Gruppen relativ gering, jedoch stieg sie über die Altersgruppen hinweg fast linear an und übertraf schließlich sogar in der höchstaltigen Gruppe die Bedeutung der prospektiven Bewertungen, welche genauso, wie die der gegenwartsbezogenen, über die Altersgruppen hinweg weder bedeutsam zu- noch abnahm. Damit also existiert zumindest ein empirischer Beleg für eine mit dem Alter wachsende Relevanz retrospektiver Bewertungen für die allgemeine Lebenszufriedenheit.⁵

Explizit wurde die *entwicklungspsychologische* Bedeutung subjektiver zeitlicher Orientierung bzw. Zeitperspektive von der Forschergruppe um Laura Carstensen mit der Formulierung einer Theorie sozioemotionaler Selektivität thematisiert (Carstensen, 1993; 1998; Carstensen et al., 1999): Soziales Verhalten wird als zielreguliertes Handeln erklärt, wobei grundsätzlich zwei funktionale Kategorien handlungsmotivierender Ziele angenommen werden, nämlich zum einen informations- bzw. wissensbezogene und zum anderen emotionsbezogene Ziele. Es kann also demnach motiviert sein durch Bestrebungen nach Informationsgewinn und/oder danach, emotionale Zustände durch den Kontakt mit anderen möglichst vorteilhaft zu regulieren. Für die jeweilige Gewichtung informations- versus emotionsbezogener Ziele ist nun aber die Zeitwahrnehmung des Individuums von entscheidender Bedeutung: Die Informationsgewinnung dient primär der „preparedness“, d.h. es wird Wissen gesucht, das zur Bewältigung zukünftiger, antizipierter Aufgaben befähigt, während die emotionale Befriedigung, die vermittelt sozialer Kontakte gewonnen werden kann, unmittelbar und gegenwärtig stattfindet. Die beiden Typen sozialer Motivation sind also auch durch den Gegensatz ihrer längerfristigen versus kurzfristigen zeitlichen Perspektive gekennzeichnet und eine der Kernaussagen dieser Theorie besteht somit darin, daß die Wichtigkeit informationsbezogener Ziele in dem Maße wächst, wie zukünftige Zeit antizipiert wird, während die Wahrnehmung eines nahen Endes (nicht notwendigerweise des gesamten Lebens, sondern z.B. auch bestimmter sozialer Beziehungen) zu einer größeren Relevanz kurzfristiger emotionaler Ziele führt. Die Theorie sozioemotionaler Selektivität

⁵Die Aussagekraft dieses Belegs ist allerdings zum einen dadurch eingeschränkt, daß es sich wiederum um einen „nur“ querschnittlich untersuchten Alterseffekt handelt, und zum anderen dadurch, daß hier auch die Frage der Generalisierbarkeit der Ergebnisse gestellt werden muß: Es handelt sich um eine israelische Untersuchung und der Autor selbst verweist auf möglicherweise gerade in der Grundgesamtheit der israelischen Alten besonders stark wirksame Kohorteneffekte, da es sich bei dieser Generation zu einem sehr großen Teil um europäische Juden handelt, die die versuchte Vernichtung durch die nationalsozialistische Gewaltherrschaft traumatisch er- und überlebt haben, sowie auch als Pioniere der israelischen Staatsgründung in ihrem Lebenslauf besonderen historischen Belastungen ausgesetzt waren. Es wurde deshalb weiter vorne auch darauf verzichtet, das in dieser Studie gefundene Absinken durchschnittlicher Zufriedenheitswerte mit dem Alter als empirischen Befund gegen das Zufriedenheitsparadox aufzuzählen.

postuliert darum für Menschen im höheren Lebensalter einen stärkeren Gegenwartsbezug in ihren sozialen Motiven, als für die (normal gesunden) Jüngeren, denen mehr zukunftsorientiertes Handeln unterstellt wird — und es werden dafür auch eindrucksvolle empirische Belege geliefert (übersichtlich dargestellt bei Carstensen et al., 1999). Sie widerspricht damit den oben dargelegten Überlegungen insofern zumindest teilweise, als hier gerade die Alten als diejenigen erscheinen, die „in der Gegenwart leben“, so daß damit eher vermutet werden könnte, daß auch die Lebenszufriedenheit der Älteren stärker durch diesen Gegenwartsbezug gekennzeichnet ist. Insbesondere wird von den Vertretern der Theorie der Annahme widersprochen, ältere Menschen zeichneten sich durch verstärkten Vergangenheitsbezug aus (vgl. Carstensen et al., 1999, 168: Fußnote 1)

Die oben erwähnten empirischen Befunde (Dittmann-Kohli, 1995; Shmotkin, 1991) zeigen allerdings ein dazu konträres Bild: Von verstärktem Gegenwartsbezug ausgerechnet älterer Menschen kann dort nicht die Rede sein. Diese Diskrepanz ist möglicherweise damit zu erklären, daß in jenen Untersuchungen die zeitliche Orientierung anders konzeptualisiert ist, als in der sozioemotionalen Selektivitätstheorie: Nämlich in ersteren letztlich als bloße Prävalenz von Gedanken an Vergangenes, Gegenwärtiges oder Zukünftiges gegenüber der zeitlichen Ausrichtung sozialer Motive in letzterer. Der Zeitbezug handlungsregulierender Ziele aber ist a priori auf Gegenwart und Zukunft beschränkt und die eigentliche Unterscheidung, die dem zugrundeliegt und in der sozioemotionalen Selektivitätstheorie gemeint ist, ist die zwischen kurz- und langfristig zu realisierenden Zielen – wenn also Ältere motivational sehr gegenwartsbezogen sind, so muß das nicht bedeuten, daß sie nicht häufig an Vergangenes denken und dieses bei ihren Bewertungen z.B. der Lebenszufriedenheit berücksichtigen. Die sozioemotionale Selektivitätstheorie enthält darum m.E. keinen echten Widerspruch zu der Annahme, daß retrospektive Bewertungsaspekte im Alter zunehmend wichtiger werden.

Allerdings könnte diese Theorie in anderer Hinsicht zur Erklärung des Zufriedenheitsparadoxes beitragen: Sie postuliert eine im Alter steigende Wichtigkeit emotionsbezogener Ziele sozialen Handelns. Wenn man damit also annimmt, daß ältere Menschen in stärkerem Maße durch das Streben nach positivem affektivem Erleben motiviert sind, als jüngere, so könnte auch gefolgert werden, daß sie ihren gegenwärtigen Zustand mehr danach bewerten, ob diese Ziele erreicht werden. Dieses würde bedeuten, daß im Alter affektive Inferenzen die Zufriedenheitsurteile stärker beeinflussen. Wenn somit Zufriedenheit im Alter „affektiver“ wird, so könnte sie darum auch intraindividuell stabiler werden, wenn, wie vorne bereits dargelegt (vgl. Kapitel 1.1.4.4), eine starke Abhängigkeit der affektiven Wohlbefindenskomponente von stabilen Persönlichkeitsmerkmalen besteht. Es wurde allerdings auch bereits darauf hingewiesen, daß empirische Befunde zu affektivem Wohlbefinden kein ebenso klares Bild „paradoxe“ Mittelwertsstabilität in Altersgruppenvergleichen zeigen, wie diejenigen zur Lebenszufriedenheit (wobei jedoch auch hier wiederum gilt, daß querschnittliche Befunde allenfalls einen Hinweis auf entsprechende Merkmalsstabilität liefern könnten). Es ist also nicht klar, ob bzw. in welchem Ausmaß affektives Wohlbefinden im

Alter überhaupt intraindividuell stabil ist, so daß auch die Hypothese einer im höheren Alter aufgrund „größerer Affektivität“ gesteigerter Stabilität von Zufriedenheiten einstweilen auf unsicherem Grund stehen bleiben muß. Es handelt sich immerhin um eine Möglichkeit, die zukünftig vielleicht noch speziell erforscht werden könnte.

Von den noch wenig untersuchten Vermutungen möglicher „stabilisierender“ Veränderungen der zeitlichen Perspektive bzw. Affektbezogenheit von Zufriedenheitsurteilen abgesehen, könnte mit Blick auf die Adaptivität von Zufriedenheitsurteilen auch gefragt werden, ob ältere Menschen nicht sogar „mehr“ bzw. „besser“ adaptieren, als jüngere: Vielleicht werden mit fortschreitendem Alter Urteile häufiger und schneller relativiert. Auch wenn eine explizite Erörterung entsprechender Alterseffekte z.B. bezüglich Häufigkeit und Geschwindigkeit der Adjustierung internaler Soll-Zustände oder der Selektion bewertungsrelevanter Information m.E. in den vorliegenden Veröffentlichungen zum Thema nicht stattfand, so lassen sich doch Argumente finden, die zumindest in diese Richtung weisen. So könnte ganz allgemein vermutet werden, daß gerade für diejenigen Urteilsheuristiken, die vorne im Zusammenhang mit der Relativität von Zufriedenheitsurteilen besprochen wurden (Kapitel 1.1.4.3), auch Übungseffekte relevant sein können. Möglicherweise „funktioniert“ die Adaptation immer besser, je älter die Person wird und je mehr Veränderung sie erlebt hat, welche Neubewertungen und entsprechende Adaptationen hervorrief. Der Gedanke, daß Adaptation gewissermaßen im Lebenslauf geübt wird, erscheint v.a. auch dann plausibel, wenn man, wie bereits dargestellt, diese Adaptation unter dem Gesichtspunkt der Resilienz als zielgerichtet protektive Funktion betrachtet: Wenn Adaptation in dieser Weise für das Individuum sinnvoll und wichtig ist, so macht auch die Annahme Sinn, daß diese Funktion beständig verbessert und, mit anderen Worten, die adaptive Kompetenz gestärkt wird.

Ein Beispiel für eine derartige Argumentation gibt Staudinger (2000), wenn sie vorschlägt, das Entwicklungsmodell der selektiven Optimierung mit Kompensation (SOK-Modell; Baltes & Baltes, 1989, 1990) auf die Genese des SWB anzuwenden. Das SOK-Modell wurde von seinen Autoren als prototypische Beschreibung einer erfolgreichen Alternsstrategie entwickelt, es impliziert die Konzentration des alternden Menschen auf lebensweltliche Bereiche hoher Priorität (Selektion), in denen beständig Verbesserung angestrebt (Optimierung) wird, wobei dann, wenn bestimmte Wege zur Optimierung nicht mehr offen sind (d.h. Wegfall von Handlungsressourcen), auf andere Mittel zur Zielerreichung ausgewichen wird (Kompensation). Bezogen auf Zufriedenheit könnte dieses bedeuten, daß das Individuum bestrebt ist, positive Zufriedenheit zu erlangen durch möglichst günstige Selektion zufriedenheitsrelevanter Information, Bemühung um tatsächliche Optimierung in den ausgewählten Bereichen und Kompensation ungünstiger Information z.B. durch Neuadjustierung von Soll-Zuständen. Die dargestellten Möglichkeiten relativierender Urteilsstrategien und auch die affektive Inferenz wären gemäß dieser Sichtweise also sozusagen als Instrumente selektiver Optimierung mit Kompensation von Zufriedenheit im Alter anzusehen und Staudinger (2000) argumentiert, daß es möglicherweise für das Gelin-

gen dieses SOK-Prozesses (d.h. für die Herstellung hoher Zufriedenheit) von Vorteil sein könnte, „auf das mit Weisheit verbundene Wissens- und Urteilssystem der fundamentalen Pragmatik des Lebens zurückgreifen zu können“ (ebd., 195). Damit aber ist die Möglichkeit einer im Alter verbesserten Kompetenz zur Adaptation von Zufriedenheitsurteilen zumindest angedacht. In dieselbe Richtung weist auch das Modell entwicklungsregulativen Handelns, das Brandtstädter (1990, vgl. dort: Abb. 3, S. 341) vorschlägt: Die dort aufgezeigten „akkomodativen Bewältigungsprozesse“ beinhalten im Grunde die adaptive Bewertung der persönlichen Entwicklungssituation, wenn diese ursprünglich nicht zufriedenstellend ist und infolge fehlender Handlungs- und Kontrollmöglichkeiten primär auch nicht verändert werden kann. Sind nun ältere Menschen häufig von einschneidenden Verlusterlebnissen betroffen, so dürfte – zieht man die für das höhere Alter typische Einschränkung von Handlungsressourcen in Betracht (vgl. Brandtstädter & Rothermund, 1998) – meist genau dieser Fall das Resultat sein und somit häufiger der akkomodative Modus gewählt werden. Dieses würde bedeuten, daß im höheren Lebensalter zwangsläufig „mehr adaptiert“ wird.

Es könnte somit sein, daß ältere Menschen aufgrund einer über die Lebensspanne verbesserten adaptiven Kompetenz und/oder verringerter regulativer Handlungsressourcen beim Eintreten von Veränderungen ihrer objektiven Lebensumstände schlicht mehr Adaptation in ihren Zufriedenheitsurteilen zeigen, als jüngere bzw. als sie selbst noch in ihren jüngeren Jahren in diesen Fällen gezeigt hätten. Das Resultat wäre eine im Alter ausgeprägtere Merkmalsstabilität von Zufriedenheit über längere Zeitintervalle hinweg, durch welche die Befunde „paradox“ stabiler Zufriedenheitsmittelwerte erklärbar wären.

1.3 Hypothesen und Ziele der empirischen Untersuchung

In den Kapiteln 1.2.3.1 bis 1.2.3.4 wurden verschiedene Möglichkeiten entwicklungs-dynamischer Effekte für die Zufriedenheit im höheren Alter diskutiert und theoretisch begründet – es sind dieses die Hypothesen, die mit der nachfolgend dargestellten empirischen Untersuchung überprüft werden sollen:

- ▶ **Kohorteneffekthypothese** (vgl. Kapitel 1.2.3.1): Möglicherweise sind die heute älteren Menschen „mit weniger zufrieden“. Alterskorrelierte *intraindividuelle* Zufriedenheitsverluste könnten dann die Regel sein, jedoch in querschnittlichen Vergleichen von Altersgruppen lediglich dazu führen, daß keine Mittelwertsdifferenzen zwischen den Altersgruppen sichtbar werden, da jüngere Gruppen sozusagen auf niedrigerem Zufriedenheitsniveau gestartet sind, jedoch noch geringere Verluste erlitten haben. Mit anderen Worten: Kohorteneffekte und alterskorrelierte Zufriedenheitsverluste könnten sich in querschnittlichen Vergleichen wechselseitig in der Weise neutralisieren, wie es schematisch in Abbildung 1.1 dargestellt wurde.
- ▶ **Gewinner-und-Verlierer-Hypothese** (vgl. Kapitel 1.2.3.2): Möglicherweise stellt das Erreichen der Hochaltrigkeit per se einen positiven Prädiktor für Zufriedenheit dar. Erwartungen vielfältiger Einbußen und Verschlechterungen objektiver Lebensumstände

könnten häufig zu einer Art von positivem „Überraschungseffekt“ führen. D.h. aufgrund negativer Altersstereotype könnte es zu entsprechend niedrigen Adjustierungen interner Soll-Zustände kommen, die von der Realität eher überboten werden, so lange Verschlechterungen nicht in dem erwarteten Ausmaß eintreten. Im letzteren Falle aber käme es dann doch zu Zufriedenheitsverlusten, so daß es in der Gruppe Hochaltriger sozusagen häufig Gewinner *und* Verlierer der Zufriedenheit gäbe, die sich hinsichtlich der zentralen Tendenz wechselseitig neutralisieren, was wiederum die Mittelwertsstabilität in Altersgruppenvergleichen erklären würde. Allerdings müßte auf diese Weise auch die Streuung von Zufriedenheitswerten mit dem Alter steigen (vgl. Abbildung 1.2).

- ▶ **Hypothese allgemeiner Merkmalsstabilität** (vgl. Kapitel 1.2.3.3): Möglicherweise sind die kognitiven Komponenten des SWB ganz allgemein (d.h. in jedem Lebensalter) intraindividuell so stabil, daß im höheren Lebensalter gehäufte Verlusterlebnisse nicht zu dramatischen Zufriedenheitsverlusten führen. Theoretisch begründbar ist dieses mit einer Trait-Abhängigkeit und v.a. mit einer allgemeinen Adaptivität von Zufriedenheitsurteilen, welche sozusagen intraindividuelle Stabilität „auf lange Sicht“ bedingt, ungeachtet kurzfristiger Ausschläge in Reaktion auf abrupte Veränderungen objektiver Lebensumstände.
- ▶ **Hypothese erhöhter Merkmalsstabilität** (vgl. Kapitel 1.2.3.4): Möglicherweise werden Zufriedenheitsurteile im höheren Lebensalter intraindividuell stabiler. Dieses könnte z.B. für weitgehend bereichsunspezifische, globalere Zufriedenheitsbewertungen infolge einer stärker retrospektiven Perspektive, die das Urteil weniger an aktuelle Entwicklungen knüpft, oder aufgrund einer mit wachsender Lebenserfahrung verstärkten Kompetenz zur adaptiven Regulierung kritischer Lebensereignisse der Fall sein.

Diese Hypothesen sind logisch nicht strikt miteinander verknüpft in dem Sinne, daß es sich dabei um wechselseitig ausschließliche Alternativerklärungen des Phänomens einer häufig vorgefundenen paradoxen Mittelwertsstabilität der Lebenszufriedenheit im höheren Erwachsenenalter handelt. Vielmehr wurden sie gewissermaßen als Filtrat theoretischer Überlegungen zu möglichen, auf die längsschnittliche Entwicklungsdynamik wirksamen Effekten präsentiert, die prinzipiell auch simultan vorhanden sein und sich überlagern könnten.

Eine gewisse wechselseitige Unvereinbarkeit besteht m.E. zwischen der Gewinner- und Verlierer-Hypothese einerseits und den Stabilitätshypothesen andererseits. Denn erstere impliziert zumindest keine allzu hohe Merkmalsstabilität im höheren Erwachsenenalter: Da bei der theoretischen Begründung dieses Effekts von einer prinzipiell eher positiven Veränderungstendenz beim Nichteintreten massiver „äußerer“ Verlusterlebnisse ausgegangen und angenommen wurde, daß negative Veränderungen vom Auftreten solcher Verluste abhängen, ist darin auch die Annahme weitestgehender Unkorreliertheit des Ausgangsniveaus von Zufriedenheit und den im höheren Alter eintretenden Zufriedenheitsveränderungen bzw. der Zufriedenheiten zu verschiedenen Zeitpunkten impliziert. Somit bedeutet dieser Effekt vielerlei Veränderung der Zufriedenheit im höheren Alter und würde auch bei gleichzeitiger Überlagerung mit stabilisierenden Effekten hohe Stabilitätskoeffizienten

(d.h. serielle Korrelationen zwischen Zufriedenheiten zu verschiedenen Zeitpunkten, vgl. Kapitel 2.2.1) verhindern bzw. in der Phase des höheren Lebensalters eine alterskorrelierte Abnahme der Stabilität von Zufriedenheit bewirken – nämlich nicht nur der strikten Stabilität im Sinne *absolut* gleichbleibender Zufriedenheitswerte, sondern auch der „monotonen“ Stabilität im Sinne der Beibehaltung der *relativen* Merkmalsausprägung und hoher serieller Korrelationen der Zufriedenheitswerte über verschiedene Meßzeitpunkte hinweg (zur Unterscheidung verschiedener Stabilitätstypen siehe Kapitel 2.2.1).⁶ Stattdessen würde dieser Effekt v.a. in wachsenden Varianzen im längsschnittlichen Verlauf, hervorgerufen durch viele Zufriedenheitsveränderungen in beide Richtungen, manifest.

Die beiden Stabilitätshypothesen behaupten stabilisierende Effekte auf Zufriedenheitsurteile, im einen Fall solche, die lebenslang (zumindest über die gesamte Spanne des Erwachsenenalters hinweg) wirken, und im anderen solche, die im Lebenslauf sich entwickeln und im höheren Alter immer stärkeren Einfluß entfalten. Dieses widerspricht sich nicht wechselseitig, es wäre durchaus möglich, daß beiderlei stabilisierende Einflüsse wirksam sind und somit Zufriedenheit lebenslang und im Alter noch wachsend stabil ist.

Das logische Verhältnis der Kohorteneffekthypothese zu den anderen Hypothesen erscheint komplexer: Im Kern enthält diese die Annahme eines generellen negativen Alterseffekts auf Zufriedenheitsurteile, sie impliziert somit eine Tendenz durchschnittlicher Zufriedenheitsverluste über die Phase der Höheraltrigkeit hinweg. Dies widerspräche der Annahme strikter Stabilität im Sinne weitestgehender Aufrechterhaltung des *absoluten* Zufriedenheitswerts, es widerspräche nicht einer Annahme von Stabilität der relativen Zufriedenheitsausprägungen: Es wäre denkbar, daß allgemeine alterstypische Verlusterlebnisse, die von der Mehrzahl der Älteren erlebt werden, zu Zufriedenheitsverlusten führen (welche als durchschnittlicher Rückgang sichtbar würde), daß aber gleichzeitig stabilisierende Einflüsse diese Verluste teilweise abfedern und sich v.a. in der Weise auswirken, daß die Personen auch bei einem durchschnittlichen, bei fast allen Mitgliedern der eigenen Altersgruppe stattfindenden Rückgang ihre relative „Zufriedenheitsposition“ im Vergleich zu den anderen beibehalten. Die Kohorteneffekthypothese verweist auf eine gewisse Unabhängigkeit der Aspekte der Stabilität von Mittelwerten und der vermittels Stabilitätskoeffizienten erfaßten Stabilität relativer Merkmalsausprägungen (es sei hier noch einmal auf Kapitel 2.2.1 verwiesen, wo verschiedene Stabilitätstypen definiert werden). Theoretisch scheint eine Überlagerung alterstypisch zufriedenheitsmindernder Effekte, die einen negativen Alterseffekt erzeugen würden, und stabilisierender Effekte möglich, dieses müßte dann zu dieser Kombination von durchschnittlichen alterskorrelierten Zufriedenheitsverlusten und hohen Stabilitätskoeffizienten (bei niedriger strikter Stabilität) führen.

⁶Die Hypothese könnte allerdings so ergänzt bzw. modifiziert werden, daß sie monotone Stabilität erlaubt: Wenn v.a. Individuen mit eher niedrig ausgeprägten Zufriedenheiten *in der Folge* häufig von gravierenden Verlusten betroffen sind, dann könnte das Resultat eine wachsende Varianz der Zufriedenheitswerte bei gleichzeitig substantieller Korrelation der Veränderungen mit dem Ausgangsniveau sein. Ein solcher Zusammenhang aber bliebe noch eigens theoretisch zu begründen.

Kohorteneffekt- und Gewinner-und-Verlierer-Hypothese widersprechen sich auf den ersten Blick, denn erstere impliziert einen negativen Alterseffekt, während letztere ja, wie dargelegt, von einer prinzipiell positiven Entwicklung der Zufriedenheit im höheren Alter bei ausbleibenden Verlusterlebnissen ausgeht. Allerdings beruht die Gewinner-und-Verlierer-Hypothese auch auf der Annahme, daß es bei dramatischen Verlusterlebnissen zu Zufriedenheitsverlusten kommt, und somit wäre mit dieser Hypothese auch ein zumindest leichter negativer Alterseffekt denkbar, wenn nämlich sozusagen die Verluste die Gewinne in Anteil und Ausmaß übertreffen. Letztendlich beinhalten die Überlegungen zur Gewinner-und-Verlierer-Hypothese keine Annahmen über das quantitative Verhältnis von Gewinn und Verlust und somit wäre diese Hypothese mit der Kohorteneffekthypothese nicht wirklich unvereinbar, sondern im Gegenteil beide Hypothesen logisch unabhängig voneinander.

Betrachtet man das logische Verhältnis der Kohorteneffekthypothese zu den anderen Hypothesen zusammenfassend, so wird besonders deutlich, daß diese Hypothese gewissermaßen eine Sonderrolle spielt. Denn sie enthält eigentlich einen grundsätzlichen Widerspruch zu den Befunden des Zufriedenheitsparadoxes, welches gemäß der Kohorteneffekthypothese als methodisches Artefakt erschiene: Es gäbe demnach durchaus einen alterskorrelierten Rückgang der Zufriedenheitsmittelwerte, der nur nicht erkannt wird, solange in entsprechenden Mittelwertvergleichen eine Konfundierung durch den Kohorteneffekt enthalten ist. Die anderen drei Hypothesen sind dagegen im Grunde Hypothesen zum Ausmaß der Merkmalsstabilität von Zufriedenheit im höheren Alter. Die Kohorteneffekthypothese bezieht sich sozusagen auf die Mittelwertsstruktur einer Zeitreihe von Zufriedenheitswerten, während die anderen Hypothesen v.a. deren Kovarianzstruktur betreffen – dieses wird im folgenden bei den Erläuterungen zur statistischen Vorgehensweise noch genauer gezeigt.

Bei dieser komplexen logischen Beziehungsstruktur der obengenannten Hypothesen kann das Analyseziel der folgenden empirischen Untersuchungen nicht durch eine Analysestrategie erreicht werden, die einen strikten Entscheidungsalgorithmus zur Auswahl der „richtigen“ und zum Ausschluß der „falschen“ Hypothesen anwendet, wie er z.B. experimentellen Forschungsdesigns zugrundeliegt. Sondern es geht hier v.a. darum, in den vorhandenen Daten nach Hinweisen auf die graduelle Gültigkeit jeder einzelnen dieser Hypothesen, unabhängig von den jeweils anderen, zu suchen, – wobei hier die Zaghaftheit der Formulierung auch dem Umstand Rechnung tragen soll, daß es sich um eine Sekundäranalyse vorhandener Paneldaten handelt, die nicht spezifisch zum Zwecke dieser Fragestellung erhoben wurden und deshalb naturgemäß nur in beschränktem Umfang spezifische Analysen dazu erlauben. Die Hypothesen beinhalten mögliche Charakteristika der längsschnittlichen Entwicklung von Zufriedenheiten bei höheraltrigen Menschen und es soll schlicht untersucht werden, ob und wie stark diese Charakteristika in dem empirischen Datensatz auffindbar sind. Damit soll ein Beitrag zu der m.E. noch keineswegs ausreichend untersuchten Frage (vgl. Kapitel 1.2.2) geleistet werden, wie sich Zufriedenheiten im höheren Alter tatsächlich entwickeln und *ob* es überhaupt für diesen Lebensaltersabschnitt typische Entwicklungsverläufe gibt.

Ein Problem dieser Zielsetzung besteht in der Verfügbarkeit geeigneter empirischer Daten. Da eine allgemeine Analyse der Zufriedenheit „der Alten“ angestrebt wird, sollte es sich um einen Datensatz handeln, der eine entsprechend unspezifische und große Zufallsstichprobe aus der Altenbevölkerung erfaßt. Außerdem werden, wie bereits in Kapitel 1.2.2 dargelegt wurde, Längsschnittdaten über einen möglichst langen Zeitraum mit vielen Meßzeitpunkten benötigt: Der Datensatz muß eine sozusagen doppelte Veränderungsmessung, nämlich zum einen die Erfassung intraindividuelle Stabilität innerhalb kürzerer Zeitintervalle, und zum zweiten die Veränderung dieser Stabilität über größere Zeitspannen hinweg ermöglichen. *Idealerweise* könnten für derartige Untersuchungen über mehrere Lebensjahrzehnte und zu vielen Meßzeitpunkten mit jeweils kurzen Intervallen erhobene Zeitreihendaten erwachsener Personen analysiert werden – allerdings sind Längsschnittdaten, die diesen hohen Ansprüchen genügen, kaum verfügbar. Ein Datenerhebungsprojekt, das allen Ansprüchen genügt, wäre mit einem erheblichen ökonomischen Aufwand verbunden, der allein durch das Interesse an der Stabilität von SWB über die Lebensspanne nicht zu rechtfertigen ist. Immerhin aber finden in vielen Ländern sozialwissenschaftliche Längsschnittbefragungen großer Stichproben statt, die oft schon über mehr als ein Jahrzehnt mit meist jährlichen Erhebungswellen laufen und u.a. auch Zufriedenheitsurteile abfragen. In Deutschland ist dies das Sozio-oekonomische Panel (SOEP), welches 1984 gestartet wurde und somit inzwischen genügend Meßzeitpunkte liefert, um nicht nur Veränderung zwischen aufeinanderfolgenden Erhebungen zu erfassen, sondern auch die „Veränderung der Veränderung“ über einen langen Zeitraum hinweg (vgl. die Beschreibung des SOEP in Kapitel 3.1).

Zudem enthält die Zielsetzung der folgenden empirischen Analysen auch hohe forschungsmethodische Ansprüche: Es wurde in Kapitel 1.2.2 bereits dargelegt, daß gewissermaßen die Anwendung der statistischen Verfahren hier ein Forschungsziel per se darstellt, insofern nämlich die jüngsten Entwicklungen längsschnittlicher Methoden eine genauere und tiefgreifende Erfassung von Stabilität und Veränderung des zu untersuchenden Merkmals versprechen, als es in den dort aufgelisteten Untersuchungen möglich war. Zwei grundlegende Anforderungen an die statistische Methodologie wurden bereits genannt: Untersuchung meßfehlerbereinigter Merkmalsveränderungen sowie „state of the art“ bei der Behandlung fehlender Werte. Die Erfüllung solcher Gesichtspunkte ist im Rahmen allgemein bekannter Standardmethoden längsschnittlicher Datenanalyse, wie z.B. der ANOVA mit Meßwiederholung, kaum lösbar. Bereits dann, wenn die Veränderung von Varianzen untersucht werden soll (vgl. Gewinner-und-Verlierer-Hypothese), ist beispielsweise davon auszugehen, daß selbst bei Annahme des einfachen Meßmodells der klassischen Testtheorie für die untersuchten Meßwerte – infolge der in den Meßwertvarianzen enthaltenen Anteile von Meßfehlervarianz – deren Veränderungen die Veränderungen der „wahren“ Merkmalsvarianzen nur unzureichend wiedergeben. Erst recht stellt die Modellierung „wahrer“ intraindividuelle Veränderung bzw. der „Veränderung der Veränderung“ gewisse Ansprüche an die statistische Methodik, die auf einfachem Wege nicht optimal erfüllt werden können.

Die statistischen Methoden wurden darum als zusätzlicher thematischer Schwerpunkt der vorliegenden Studie betrachtet: Es soll für die angestrebten Untersuchungen intraindividuelle Stabilität und Veränderung ein möglichst aktueller diesbezüglicher „state of the art“ statistischer Analyse festgehalten und angewendet werden. Im folgenden wird darum auch der Darstellung der statistischen Methoden mehr Raum gegeben, als dies bei einer nur auf das inhaltliche Thema fokussierten Untersuchung üblich wäre (vgl. Kapitel 2).

Die obengenannten Hypothesen beziehen sich primär auf allgemeine Lebenszufriedenheit als *die* kognitive Wohlbefindenskomponente und es werden im folgenden Messungen allgemeiner Lebenszufriedenheit analysiert. Folgt man theoretischen Vorstellungen zur Generierung von Zufriedenheitsurteilen wie dem konstruktionistischen Ansatz (vgl. Kapitel 1.1.4.7) bzw. der informationsverarbeitungstheoretischen Sichtweise von Schwarz & Strack (1991; vgl. Kapitel 1.1.4.6), wonach auch bei Fokussierung der Beurteilung auf eng umgrenzte Bereiche des eigenen Lebens Zufriedenheitsurteile zumindest teilweise von unspezifischen Bewertungsaspekten gewissermaßen infiziert sein können, so *können* spezifische Bereichszufriedenheiten Ähnlichkeiten zur allgemeinen Lebenszufriedenheit aufweisen, sie müssen es aber nicht. Die grundlegende Annahme ist hierbei, daß solche Zusammenhänge aufgrund von Ähnlichkeiten der Urteilsgenerierung – nämlich beim Rückgriff auf Information, auf die das Urteil begründet wird – entstehen. Es ist deshalb von zusätzlichem, wenn auch gegenüber den hier untersuchten Hypothesen nachgeordnetem Interesse, welche Bereichszufriedenheiten in welchem Ausmaß strukturelle Ähnlichkeiten zur Lebenszufriedenheit aufweisen. Diesbezügliche Ergebnisse könnten weitere und möglicherweise zukünftige Zufriedenheitsforschung stimulierende Vermutungen darüber liefern, welche Lebensbereiche aufgrund derselben Informationen beurteilt werden, wie das Leben insgesamt. So wäre es beispielsweise interessant herauszufinden, bezüglich welcher Lebensbereiche die subjektiven Beurteilungen *nicht* stabil sind und möglicherweise kaum adaptiv reguliert werden, sondern unmittelbar objektive Verhältnisse reflektieren, als es bei der Globalbeurteilung der Fall zu sein scheint. Es sollen darum grundsätzlich auch die im SOEP mit vollständigen Zeitreihen vorhandenen Bereichszufriedenheitsmessungen stets mit analysiert werden – auch auf die Gefahr hin, daß damit ein „information overload“ in der Ergebnisdarstellung erzeugt wird. Diese Ergebnisse für die Bereichszufriedenheiten werden sozusagen zur Ergänzung berichtet und kommentiert, sie sollten nicht die Aufmerksamkeit von der Lebenszufriedenheit ablenken.

2 Methodische Grundlagen: Statistische Modelle und Prozeduren

Es wurde in Kapitel 1.1 dargelegt, daß die angestrebten Analysen zu Kohorteneffekten und zu Stabilität von Zufriedenheit einige Anforderungen an die Auswahl dazu geeigneter statistischer Methoden stellen und daß diese Auswahl – d.h. eine Auswertung entsprechend einem gegenwärtigen „Stand der Dinge“ in der Methodologie längsschnittlicher Datenanalyse – gewissermaßen ein Untersuchungsziel per se darstellt. Deshalb wird der Auswahl der statistischen Methoden hier ein eigenes Kapitel gewidmet, in dem dieser zugrundegelegte Anspruch, die möglichst aktuellsten Entwicklungen zur Behandlung der typischen Probleme bei der anstehenden längsschnittlichen Untersuchung miteinzubeziehen, verwirklicht oder zumindest seine Verwirklichung angestrebt werden soll. Dazu erscheint es notwendig, v.a. folgende Fragen sorgfältig zu diskutieren und zu beantworten: Wie können die möglichen Kohorteneffekte auf die Zufriedenheitsurteile Älterer erfaßt und von Alterseffekten statistisch isoliert werden? In welcher Weise kann die „wahre“, meßfehlerbereinigte Merkmalsstabilität der Zufriedenheit untersucht werden? Und wie soll mit der im Panelverlauf wachsenden Zahl von Stichprobenausfällen umgegangen werden?

Im folgenden wird zunächst die Problematik der Analyse von Kohorteneffekten diskutiert und ein Modell zur Untersuchung der Kohorteneffekthypothese vorgestellt. Danach folgt eine ausführliche Erörterung der Methoden zur Stabilitätsanalyse: Zunächst werden die Zielparameter einer solchen Analyse dargestellt, danach werden die „prominentesten“ längsschnittlichen Strukturgleichungsmodelle kurz vorgestellt und hinsichtlich ihrer Eignung für die hier angestrebten Analysen diskutiert. Schließlich werden noch spezifische Probleme dieser Analysen behandelt, nämlich die Behandlung fehlender Werte und die Kriterien zur Annahme oder Ablehnung von Parameterschätzungen aufgrund der Modellanpassungsgüte der Strukturgleichungsmodelle.

2.1 Analyse von Kohorteneffekten

Zur Untersuchung der Kohorteneffekthypothese wird ein statistisches Modell zur Trennung von Alters- und Kohorteneffekten benötigt. Dazu ist zunächst einmal der Begriff des Kohorteneffekts im Vergleich zu dem des Alterseffekts zu präzisieren bzw. festzuhalten, welche in der Zeit variierenden bzw. – wie sie im folgenden bezeichnet werden – zeitvariablen Effekte prinzipiell in den längsschnittlichen Datenreihen wirksam sein könnten. Es sind dieses: Alters-, Kohorten- und Periodeneffekte (van der Kamp & Bijleveld, 1998).

2.1.1 Zeitvariable Effekte: Alter, Kohorte und Periode

Der Begriff des Alterseffekts ist im Grunde selbsterklärend und wurde als solcher in den vorangegangenen Darstellungen auch häufig und ohne weitere Definition gebraucht. Er bezieht sich auf intraindividuelle Veränderungen von Merkmalen, die unabhängig von der jeweils historischen Lebenszeitspanne – und d.h. unabhängig vom Geburtszeitpunkt des Merkmalsträgers und von zeitlich begrenzten historischen Einflüssen – stattfinden. Bei eingehender Betrachtung erscheint jedoch auch der Begriff des Alterseffekts nicht problemfrei, zumindest für die sozialwissenschaftlichen und psychologischen Bereiche dann, wenn man ihn „inhaltlich-theoretisch“ im Wortsinne als von historischen Umständen unabhängigen, reinen Effekt des Lebensalters auf ein Merkmal auffaßt: Ist, so könnte man dann fragen, überhaupt intraindividuelle Entwicklung unabhängig von historischen Einflüssen denkbar? Impliziert eine solchermaßen postulierte Unabhängigkeit des Entwicklungsgeschehens nicht zwangsläufig einen biologischen Reifungsbegriff, welcher zur Erklärung von Veränderung in komplexeren psychischen oder sozialen Merkmalen im Grunde niemals ausreicht, und ist also nicht alle Entwicklung in diesen Bereichen nicht stets mitangestoßen durch soziale Umweltgegebenheiten, welche per se nicht als ahistorisch und zeitlich invariant betrachtet werden können? Das weite Feld, welches mit derartigen Fragen betreten wird, soll hier nun sofort wieder verlassen werden, denn ein solcher Ausflug wäre eine Abschweifung vom Thema der Untersuchung. Zudem sind diese Fragen leicht durch eher „statistische“ Auffassung des Begriffs zu vermeiden: Ein Alterseffekt bedeutet demnach schlicht intraindividuelle Veränderung, die für Individuen unterschiedlichster Geburtskohortenzugehörigkeit – und d.h. heißt auch in unterschiedlichen historische Zeitspannen – gleichermaßen über das Lebensalter hinweg stattfindet. Solchermaßen bezeichnet der Begriff durchschnittliche Veränderungsverläufe über viele Individuen zu vielen historischen Lebenszeitspannen hinweg, idealerweise die durchschnittliche alterskorrelierte Entwicklung aller Individuen zu allen Zeiten, pragmatischerweise natürlich nur aller Individuen einer auch zeitlich enger definierten Population.

Eine von Schaie & Hofer (2001) formulierte Definition des Kohorteneffekts wurde vorne schon genannt (S.52), sie enthält wiederum eine eher inhaltliche und weniger technische Auffassung: Demnach meint der Begriff Einflüsse auf das zu untersuchende Merkmal, die aus spezifischen Umweltbedingungen, denen eine Gruppe von Merkmalsträgern zu einem bestimmten historischen Zeitpunkt ausgesetzt war, resultieren. Für pragmatische Forschungszwecke erscheint eine solche Definition solange unproblematisch, wie damit der konkrete Effekt eines bestimmten und zeitlich sehr eng begrenzten historischen Ereignisses gemeint ist: Wenn z.B. der Einfluß des zweiten Weltkriegs auf die Bewertung von Lebensumständen untersucht werden sollte, so können leicht Personen, die diese Zeit erlebt haben, mit den später Geborenen verglichen werden. Jedoch wäre dabei schon zu fragen, welcher Geburtszeitpunkt als diesbezüglicher Cutoff-Wert gewählt werden soll – wenn man tatsächlich das Datum des Kriegsendes auswählen würde, so wäre dabei impliziert, daß tatsächlich auch alle während des Krieges Geborenen noch durch das Kriegserleben geprägt

wurden. Auch würde man annehmen, daß der Einfluß des Krieges auf Personen in jedem Lebensalter derselbe war, daß es also dabei keinerlei Interaktion von Kohorten- und Alters-effekt gab. Erst recht würde die Sache schwierig, wenn es, wie in der hier zu untersuchenden Kohorteneffekthypothese, nicht um den Einfluß eines eng umgrenzten historischen Ereignisses geht, sondern um einen langen Zeitraum, der durch eine kaum unterbrochene Abfolge von krisenhaften und belastenden historischen Entwicklungen, sowie durch insgesamt gegenüber der Friedens- und Wohlstandsphase der letzten Jahrzehnte deutlich schlechtere soziale Bedingungen gekennzeichnet ist, und der auch nicht durch ein bestimmtes historisches Ereignis zu einem bestimmten Datum, sondern durch eine längere Entwicklungsphase nach dem Krieg allmählich beendet wurde. Wie bzw. durch welchen Geburtszeitpunkt sollte hier die Gruppe derer, die diesen Einflüssen ausgesetzt waren, von der Gruppe derer, die ihnen nicht ausgesetzt waren, unterschieden werden? Und ist nicht auch zu berücksichtigen, wie vielen dieser Einflüsse die Personen ausgesetzt waren bzw. wie lange sie in diesen Zeiten gelebt haben? Die Schwierigkeiten, die sich hier entdecken lassen, vermeiden wiederum eine eher technische Festlegung des Begriffs Kohorteneffekt: Demnach bezeichnet dieser das Ausmaß, in dem die Merkmalsausprägungen vom Geburtszeitpunkt abhängen, d.h. die Variation des Merkmals, die unabhängig vom Alter der Personen *und* vom Zeitpunkt, an dem die Merkmalsausprägungen beobachtet werden, auf die Unterschiedlichkeit des Geburtszeitpunkts statistisch zurückführbar ist. Genaugenommen sind es ja auch die Personen mit gleichem Geburtszeitpunkt, die alle dieselben historischen Ereignisse in demselben Lebensalter erlebt haben. Somit muß also über die Kohortenzugehörigkeit nicht aufgrund schwieriger sozialgeschichtlicher und psychologischer Plausibilitätsüberlegungen ad hoc entschieden werden, sondern es steht die Kohortenzugehörigkeit immer schon fest: Kohorte kann mit Geburtszeitpunkt übersetzt werden und genaugenommen ist nur noch festzulegen, wie meßgenau dieser Zeitpunkt erhoben wird – pragmatischerweise bietet sich hier das Geburtsjahr als Einheit an, die eine immer noch sehr enge Unterteilung des Kontinuums möglicher Geburtszeitpunkte gewährleistet.

Periodeneffekte – in englischsprachigen Publikationen oft auch als „time-of-measurement effects“ bezeichnet – sind solche, die speziell während des Zeitpunkts bzw. der Periode der Datenerhebung bei allen Personen der untersuchten Population wirksam waren: „... time-of-measurement effects represent common historical exposures that influence all members of a population regardless of cohort membership“ (Schaie & Hofer, 2001, 56). Es handelt sich also bei den hiermit gemeinten Einflüssen wiederum um Umweltbedingungen „außerhalb“ der Personen, die zu einem bestimmten historischen Zeitpunkt – nämlich dem der Datenerhebung – bestanden. Beispielsweise könnte bei einer längsschnittlichen Untersuchung mit zwei Meßzeitpunkten aufgrund von ausgerechnet in diesem Zeitraum stattfindenden gesellschaftlichen Entwicklungen die Ausprägung des untersuchten Merkmals in der Population allgemein, d.h. im Durchschnitt, angestiegen sein, so daß dann dieser bei der Untersuchung auftretende Anstieg zu anderen Datenerhebungszeitpunkten gar nicht gefunden würde und jedenfalls *nicht* einen Alterseffekt anzeigt.

TABELLE 2.1:
 Schema: Geburtsjahrgänge, Lebensalter und Jahr der SOEP-Datenerhebung

	...	57	58	59	60	61	62	63	64	...
							1999			
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	1999		
1930	...	1987	1988	1989	1990	1991	1992	⋮	⋮	
1929	...	1986	1987	1988	1989	1990	1991	1992	...	1999
1928	...	1985	1986	1987	1988	1989	1990	1991	1992	...
1927		1984	1985	1986	1987	1988	1989	1990	1991	...
1926			1984	1985	1986	1987	1988	1989	1990	...
1925				1984	1985	1986	1987	1988	1989	...
1924					1984	1985	1986	1987	1988	...
1923						1984	1985	1986	1987	...
1922							1984	1985	1986	...
1921								1984	1985	...
⋮									⋮	⋮

Die Entwicklung sogenannter ACT-Modelle zur statistischen Trennung dieser zeitvariablen Effekte (das A steht hier für „age“, also den Effekt des Lebensalters; das C für „cohort“, d.h. den auf das Geburtsjahr zurückführbaren Kohorteneffekt; das T für „time“, den Effekt des Meßzeitpunktes) wurde wesentlich angestoßen und beeinflusst durch die Arbeiten von Schaie (1965, 1977) und Baltes (1968, 1973). Deren Anspruch allerdings, mit den vorgeschlagenen varianzanalytischen Modellierungen diese vollständige Trennung bewirken zu können, wurde heftig kritisiert. So beklagen Donaldson & Horn (1992), daß die prinzipielle Möglichkeit von Periodeneffekten bei der Erforschung von Alters- oder Kohorteneffekten oft und möglicherweise absichtlich ignoriert würde. Denn tatsächlich bewirkt diese Möglichkeit ein schier unlösbares Problem für die Untersuchung von Kohorten- und Alterseffekten: Wie auch immer Variabilität von Merkmalsausprägungen über eine in Zeiteinheiten unterteilte Dimension hinweg betrachtet wird, es ist dabei stets die Konfundierung von zwei der drei genannten Effekte möglich.

Zur Veranschaulichung dieser Problematik betrachte man Tabelle 2.1, die sozusagen einen Ausschnitt aus dem nach Alter und Geburtsjahr der Befragungspersonen differenzierten Schema der SOEP-Befragungen (vgl. Kapitel 3.1) zeigt und hier stellvertretend für alle möglichen Paneldesigns mit mehreren Wiederholungsmessungen steht. Die Zellen der Tabelle enthalten die SOEP-Erhebungsjahre, die Spalten repräsentieren die Lebensaltersjahre und die Zeilen die Geburtsjahre der Befragungspersonen. Wenn anhand der Daten aus diesem Panel ein möglicher Alterseffekt – z.B. in den Meßwerten der Lebenszufriedenheit – untersucht werden soll, so böte es sich zunächst einmal an, die Werte über die fortschreitenden Altersjahre hinweg, d.h. entlang der Zeilen der Tabelle 2.1, zu vergleichen: Pro Zeile, d.h. pro Lebensaltersgruppe und damit bei konstantgehaltener Kohortenzugehörigkeit, könnten z.B. die sukzessiven Mittelwerte verglichen werden. Der Alterseffekt würde dann

durch die – wie auch immer zu berechnende – statistische Aggregierung der in den einzelnen Geburtsjahresgruppen gefundenen sukzessiven Differenzen zu einer durchschnittlichen Entwicklung über die Altersdimension hinweg ermittelt. Es ist jedoch aus Tabelle 2.1 auf Anhieb zu erkennen, daß bei diesen zeilenweisen Vergleichen nicht nur durch einen Alterseffekt, sondern auch durch Periodeneffekte „infiziert“ sein können: Jede Fortschaltung des Lebensalters um eine Einheit geht dabei mit einer entsprechenden Fortschaltung des Erhebungszeitpunkts einher, und somit sind Veränderungen entlang der Zeilen nicht eindeutig auf eine der beiden Fortschaltungen zurückführbar. Allerdings wäre ein kurzfristiger Periodeneffekt, z.B. ein Anstieg oder Abfall aller Durchschnittswerte zu nur einem bestimmten Erhebungsjahr, erkennbar, da dieser dann in den zeilenweisen Verläufen der einzelnen Geburtsjahrganggruppen jeweils um ein Lebensjahr versetzt auftaucht. Genauso wäre umgekehrt ein kurzer, nichtlinearer Alterseffekt dergestalt, daß es typischerweise während einer kurzen Altersspanne zu bestimmten Veränderungen kommt, erkennbar, denn diese Veränderung würde in den zeilenweisen Verläufen jeweils bei denselben Spalten auftauchen. Ein annähernd linearer Periodeneffekt über die gesamte Erhebungsspanne hinweg, der von jedem Erhebungszeitpunkt zum nächsten einen jeweils annähernd gleichen durchschnittlichen Veränderungsbetrag in allen Zellen bewirkt, wäre jedoch von einem entsprechend annähernd linearen Alterseffekt beim zeilenweisen Vergleich nicht zu unterscheiden und es wäre dann, wenn sich ein solcher annähernd linearer Trend in allen kohortenweisen Verläufen gleichermaßen zeigt, nicht entscheidbar, ob dieser auf einen Alters- oder Periodeneffekt, oder auf die Überlagerung dieser beiden Effekte zurückzuführen ist.

Es zeigt sich somit hier der bemerkenswerte Umstand, daß die mögliche Linearität von Effekten, welche ansonsten bei vielerlei Arten statistischer Modellierung ja häufig als eine die Berechnung und Interpretation vereinfachende Restriktion vorausgesetzt wird, hier die Analyse sogar erschwert. Nichtlineare Alters- und Periodeneffekte können z.B. in der graphischen Abbildung der zeilenweisen Verlaufskurven als solche identifiziert werden, weil die dadurch erzeugten nichtlinearen Muster von „Veränderung der Veränderung“ (d.h. Trendwechsel in den Verlaufskurven) dann, wenn sie jeweils mit demselben Erhebungsjahr auftreten, als Periodeneffekte erkennbar sind und dementsprechend als Alterseffekte dann, wenn sie jeweils mit demselben Altersjahr auftreten (sofern man Interaktionen zwischen den drei Effekttypen ausschließt, vgl. Fußnote 7, S. 78). Allerdings mag auch dieses zwar im Prinzip gelten, praktisch jedoch wird der Datenanalytiker häufig auch dann, wenn ein sozusagen geringes Ausmaß solcher Nichtlinearität nicht zu besonders deutlichen Ausschlägen, sondern nur zu moderaten Trendveränderungen in den Verläufen führt, nur eine näherungsweise Linearität von Verläufen erkennen, welche dann wie gezeigt schwer interpretierbar ist. V.a. aber wird sein Interesse häufig längerfristigen globalen Entwicklungstrends gelten, die im Grunde die Annahme näherungsweise Linearität implizieren und er wird, selbst wenn er kurzfristige deutliche Ausschläge in den längsschnittlichen Verläufen erkennt, welche eindeutig dem Alter oder dem Meßzeitpunkt zuordenbar sind, immer noch vor der eigentlich zu beantwortenden Frage stehen, in welcher Weise längerfri-

stige, annähernd lineare Verlaufstrends „jenseits“ dieser Ausschläge von Alters- und/oder Periodeneffekten geformt wurden.

Diese Überlegungen zur zeilenweisen Analyse das in Tabelle 2.1 gezeigten Paneldesigns sind noch längst nicht erschöpfend, sie sollen hier nur andeuten, welche Schwierigkeit bei dieser „Analyserichtung“ die Konfundierung von Alter und Meßzeitpunkt bereitet. Genauso aber verhält es sich bei allen anderen Analysemöglichkeiten, die das Design bietet: Eine weitere Möglichkeit, den Alterseffekt zu untersuchen, bestünde in der Betrachtung entlang der Diagonalen. Dieses wäre schlicht die Analyse der einzelnen querschnittlichen Verläufe und bedeutet Konstanthaltung des Beobachtungszeitpunkts und somit Ausschluß eines möglichen Periodeneffekts. Daß hierbei jedoch eine Konfundierung von Alters und Geburtskohorte und folglich eine Überlagerung der beiden entsprechenden Effekte vorliegt, ist offensichtlich: Jede Fortschaltung des Alters um eine Einheit geht mit einer entsprechenden Rückschaltung des Geburtsjahres einher. Es gelten im Prinzip die Überlegungen, die oben für die zeilenweisen Vergleiche und die Identifikation von Alters- versus Periodeneffekten dargelegt wurden, in gleicher Weise für diese Diagonalvergleiche und die Identifikation von Alters- versus Kohorteneffekten. Auch hierbei ist diese Identifikation praktisch dann nicht mehr möglich, wenn beide Effekte über die Alters- bzw. Geburtsjahresdimension hinweg linear sind.

Eine weitere Möglichkeit des Paneldesigns gemäß des Schemas in Tabelle 2.1 besteht in spaltenweisen Analysen, d.h. z.B. im Vergleich der Mittelwerte der Gruppen jeweils Gleichaltriger in den sukzessiven Erhebungsjahren: Dabei gehen die Fortschaltung des Erhebungsjahres und des Geburtsjahres um jeweils eine Einheit miteinander einher und es ist bei diesen Analysen mit der Überlagerung von Kohorten- und Periodeneffekten zu rechnen – mit den bekannten Konsequenzen bezüglich der Identifizierbarkeit dieser beiden.

Es erweist sich als unmöglich, alle drei Typen von zeitvariablen Effekten in ACT-Modellen (Schaie, 1968) analytisch strikt zu trennen, da zwischen den unabhängigen Variablen dieses Modells (Lebensjahr, Geburtsjahr und Beobachtungszeitpunkt) perfekte Multikollinearität besteht und also keine der drei von den jeweils anderen beiden unabhängige Varianzanteile besitzt. Es gilt: Beobachtungszeitpunkt = Geburtsjahr + Lebensalter – und deshalb: „... age, cohort, and time constitute a muddle. They are redundant quantities that cannot be independently varied to produce unique contributions to a dependend variable. Assumptions must be made before quantitative analyses can be sensibly interpreted“ (Donaldson & Horn, 1992, 213).⁷

⁷Genaugenommen sind deshalb auch nichtlineare Veränderungen in den zeilenweisen, spaltenweisen oder diagonalen Verläufen der Tabelle 2.1 niemals eindeutig einem Alters-, Kohorten- oder Periodeneffekt zuordenbar. Dieses kann man an einem einfachen Beispiel demonstrieren: Wenn t den jeweiligen Meßzeitpunkt, A_{ti} das Alter einer Person i zum Zeitpunkt t und G_i deren Geburtsjahr bezeichnet, so wäre ein einfaches Regressionsmodell linearer Alters- und Kohorteneffekte, jedoch eines nichtlinearen, exponentiellen Periodeneffekts für eine abhängige Variable X_{ti} wie folgt zu schreiben:

$$X_{ti} = \beta + \alpha A_{ti} + \gamma G_i + \tau t^2 + \epsilon_i$$

2.1.2 *Statistische Modellierung der Trennung von Alters- und Kohorteneffekten*

Zur Untersuchung der Alters- und Kohorteneffekte in den Zufriedenheitsmessungen des SOEP wurde im folgenden eine statistische Modellierung gewählt, die dem Typus der sogenannten „multilevel models“ (auch „mixed models“, „random coefficient models“, „hierarchical models“) zur Analyse hierarchisch strukturierter Daten zuzuordnen ist. Obwohl diese Modelle eine weit zurückreichende Vorgeschichte haben, zählen sie, mit diesbezüglich grundlegenden Veröffentlichungen Ende der achtziger und Anfang der neunziger Jahre, doch zu den neueren Entwicklungen statistischer Analysemethoden (Bryk & Raudenbush, 1987; 1992; Goldstein, 1986; 1995). Hierarchisch strukturiert sind Daten dann, wenn die in einem statistischen Modell enthaltenen Variablen nicht allesamt für dieselben Einheiten, die die Träger der variierenden Merkmalsausprägungen sind, definiert sind, sondern wenn es sich dabei um eine Struktur hierarchisch ineinander geschachtelter Einheiten handelt. Diese etwas kryptische Formulierung bedarf wohl des anschaulichen Beispiels: Wenn beispielsweise in einer Pflegeheimstudie Variablen, die Eigenschaften der Heiminsassen beschreiben, und solche, die Eigenschaften des Pflegeheims beschreiben, erhoben wurden, dann liegt eine solche hierarchische Struktur der Daten vor: Die Einheiten des niedrigeren Levels „Personen“ sind in die Einheiten des höheren Levels „Institutionen“ geschachtelt. Wenn diese Studie z.B. noch international zum Vergleich über mehrere Länder hinweg durchgeführt würde, dann käme noch ein weiteres höheres Level „Länder“ hinzu, in das die Einheiten des Levels „Institutionen“ geschachtelt sind. Die Multilevel-Modelle sind Regressionsmodelle, die solche Levelhierarchien in den Daten – genauer: in den Prädiktoren der Modelle – explizit berücksichtigen. Dieses bedeutet, daß die Effekte (Regressionsgewichte) der Prädiktorvariablen eines jeweils niedrigeren Levels nicht mehr als fixe, d.h. für alle Einheiten dieses Levels gleichermaßen gültige, sondern als über die Einheiten des nächsthöheren Levels variierende „random effects“ modelliert sind.

Solche Multilevel-Regressionsmodelle haben insbesondere als Methode längsschnittlicher Analyse Bedeutung und Verbreitung erlangt, nämlich durch die Entwicklung der „growth curve models“ zur Analyse von Zeitreihendaten. Deren hierarchische Struktur besteht in der Schachtelung der einzelnen Meßzeitpunkte (Level 1) in die untersuchten Personen (oder sonstige untersuchte Einheiten, Level 2): Im längsschnittlichen Untersuchungsde-

Da nun aber $t = A_{ti} + G_i$, kann diese Gleichung wie folgt aufgelöst werden:

$$\begin{aligned} X_{ti} &= \beta + \alpha A_{ti} + \gamma G_i + \tau t^2 + \epsilon_i \\ &= \beta + \alpha A_{ti} + \gamma G_i + \tau(A_{ti} + G_i) + \epsilon_i \\ &= \beta + \alpha A_{ti} + \gamma G_i + \tau(A_{ti}^2 + 2A_{ti}G_i + G_i^2) + \epsilon_i \end{aligned}$$

Die letzte obige Gleichungszeile macht deutlich, daß der nichtlineare Periodeneffekt τ auch als Effekt einer komplexen Interaktion von Alter und Kohorte aufgefaßt werden kann.

sign „besitzt“ jede Person mehrere Merkmalsausprägungen zu unterschiedlichen Meßzeitpunkten, die Modellierung der Growth-Curve-Models beruht nun darauf, die Abhängigkeit der zu untersuchenden Variablen vom Level-1-Prädiktor Meßzeitpunkt in einer Regressionsgleichung zu spezifizieren, deren Regressionskonstante und Regressionskoeffizient (des Meßzeitpunkts) über die Personen hinweg variiert. Die Regressionskonstante (hier meist mit dem Begriff „level“ bezeichnet) und der Meßzeitpunktskoeffizient („slope“) beschreiben dann einen linearen Veränderungsverlauf über die Zeit und ihre Variation über die Personen hinweg bedeutet also, daß mit diesem Modell die Personen als in ihrer intraindividuellen Veränderung variierend modelliert werden. Es können entsprechend auch Modelle nichtlinearer, z.B. exponentieller, Veränderungsverläufe spezifiziert werden und die besondere Attraktivität der Methode besteht darin, daß beispielsweise Variablen der Personenebene oder auch eines noch höheren Levels als Prädiktoren der Veränderungskennwerte in das Modell aufgenommen und somit die Abhängigkeit der Veränderungscharakteristika von solchen Variablen untersucht werden kann. Prinzipiell sind solche Growth-Models auch äquivalent als Strukturgleichungsmodelle spezifizierbar, als solche werden sie noch im Kapitel 2.2.2.4 genauer dargestellt.

Zur weitergehenden Einführung bzw. Vertiefung der gesamten Multilevel-Methodologie sei hier auf gut lesbare neuere Darstellungen verwiesen (Hox, 2002; Snijders & Bosker, 1999; van der Leeden, 1998).

Das zur Analyse der simultanen Wirkung von Alters- und Kohorteneffekten auf die Zufriedenheitswerte spezifizierte Multilevel-Modell ähnelt sehr dem typischen einfachen Growth-Curve-Modell linearer Veränderung über die Zeit, der Unterschied besteht in der Modellierung der Level-1-Zeitdimension: Diese wird nicht durch die Meßzeitpunkte konstituiert, sondern durch das Lebensalter der Befragten.

Dieses bedarf der genaueren formalen Darstellung und Erläuterung. Es sei X_{ti} der Meßwert (Zufriedenheit) der Person i zum Meßzeitpunkt t (Jahr der Datenerhebung), A_{ti} das Lebensalter der Person i zum Zeitpunkt t und G_i das Geburtsjahr der Person i . Dann kann das folgende einfache, (noch) *nicht* hierarchische Regressionsmodell linearer Alters-, Kohorten- und Periodeneffekte spezifiziert werden:

$$X_{ti} = \beta + \alpha A_{ti} + \gamma G_i + \tau t + \epsilon_{ti} \quad (2.1)$$

Hierbei ist also β die übliche Regressionskonstante und α der Regressionskoeffizient des Prädiktors Lebensalter, d.h. der eigentliche Alterseffekt. Entsprechend ist γ der lineare Kohorteneffekt des Geburtsjahres und τ der Periodeneffekt des Jahres, in dem X_{ti} gemessen wurde. ϵ_{ti} bezeichnet wie üblich die Zufallskomponente des Modells.

Vor der weiteren Entwicklung des durch Gleichung 2.1 spezifizierten Modells zunächst noch einige Überlegungen zur Skalierung der darin enthaltenen Prädiktoren, denn diese ist von besonderer Bedeutung für die Interpretation von Regressionskoeffizientenwerten. Im vorliegenden Fall wären die unstandardisierten Absolutbeträge der Koeffizienten α , γ und τ

direkt miteinander vergleichbar, weil die dazugehörigen Prädiktoren Lebensalter, Geburtsjahr und Meßzeitpunkt in derselben Einheit – nämlich in Jahren – „gemessen“ werden: α ist somit der durchschnittliche Veränderungsbetrag, der gemäß der linearen Modellierung mit jedem zusätzlichen Lebensjahr einhergeht, und γ und τ sind entsprechend die durchschnittlichen Veränderungen von jedem Geburts- bzw. Erhebungsjahr zum darauffolgenden. Allerdings zöge die Verwendung der tatsächlichen Lebensalter- und Geburtsjahreswerte für die Analyse eine gewisse Unhandlichkeit für die Interpretation der Regressionskonstanten β nach sich: Das Intercept markiert eigentlich den Durchschnittswert der Kriteriumsvariablen bei „Nullausprägung“ der Prädiktoren – im obigen Modell 2.1 wäre dieses bei untransformierten Geburtsjahres- und Lebensalterswerten sozusagen die aus der rückwärtigen Fortschreibung der linearen Verläufe weit über die untersuchten Bereiche von Geburts- und Lebensjahren hinaus resultierende durchschnittliche Zufriedenheit eines im Jahre Null geborenen Säuglings bei seiner Geburt – und somit ein Wert, der selbst keine große Aussagekraft besitzt und vermutlich auch weit außerhalb des eigentlich nur möglichen Wertebereichs der Zufriedenheitsskala liegen dürfte. Auch wenn dieses für die Berechnung des Modells, sowie für die Interpretation der Koeffizienten α , γ und τ kein Problem darstellt, so kann doch durch eine Verschiebung des Nullpunkts der drei Jahresskalen die Bedeutung von β so verändert werden, daß er mehr anschauliche Aussagekraft besitzt und die Interpretation der Modellergebnisse insgesamt erleichtert: Die Nullpunkte der Lebensalter- und Geburtsjahresskala können so verschoben werden, daß sie eine Zelle des Paneldesigns, die innerhalb des untersuchten Spektrums von Alter \times Geburtsjahr \times Meßzeitpunkt liegt, bezeichnen, welche gewissermaßen als Referenzpunkt für Vergleiche entlang der Dimensionen dieses Designs und zur Berechnung von durch das Modell geschätzten Zellenmittelwerten dienen kann – β würde dann den Mittelwert in dieser Zelle bezeichnen. Im vorliegenden Fall können Alters-, Geburtsjahres- und Meßzeitpunktskalen z.B. wie folgt verschoben werden:

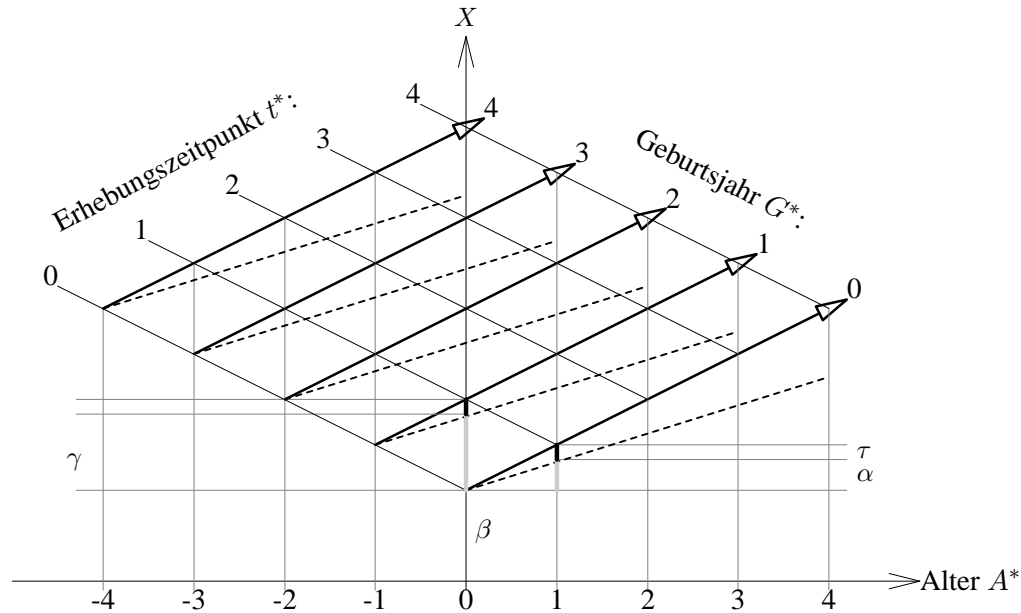
$$A_{ti}^* = A_{ti} - 60 \quad (2.2)$$

$$G_i^* = G_i - 1924 \quad (2.3)$$

$$t^* = t - 1984 \quad (2.4)$$

Durch die Transformation 2.2 von A_{ti} in A_{ti}^* wird also der Nullpunkt der neuen Lebensalterskala beim faktischen Lebensalter 60 festgelegt, der transformierte Alterswert der 61-Jährigen beträgt 1, der der 59-Jährigen -1 usw.. Entsprechend wird durch Gleichung 2.3 der Nullpunkt der neuen Geburtsjahresskala auf das kalendarische Jahr 1924 gelegt, während die Transformation 2.4 das Kalenderjahr 1984 zum neuen „Jahr Null“ der Datenerhebungen macht. Der Grund für ausgerechnet diese Verschiebungen ist ein pragmatischer: Das Lebensalter 60 hat bei den vorliegenden Untersuchungen eine gewisse Bedeutung, weil es gemäß der vorne (vgl. S. 39) gemachten Festlegung den Eintritt in die Phase der Höheralterigkeit markieren soll, es erscheint darum sinnvoll, diesen „Beginn“ der höheren Lebensalters sozusagen als transformiertes Lebensjahr Null zu betrachten. Die Verschiebung der

ABBILDUNG 2.1:
Schematische Darstellung des Modells linearer Alters-, Kohorten- und Periodeneffekte



Geburtsjahresskala bedeutet, daß den zum Beginn der untersuchten Zeitreihe am Beginn der Höheraltrigkeit Stehenden, also denjenigen, die im Jahre 1984 das sechzigste Lebensjahr vollendet haben und somit 1924 geboren wurden, der transformierte Geburtsjahreswert Null zugewiesen wurde. Daß schließlich der Nullpunkt der Meßzeitpunkte auf den Beginn der Messungen gelegt wird, bedarf wohl keiner weiteren Begründung.

Somit bezeichnet bei Verwendung der transformierten Prädiktoren A_{ti}^* , G_{ti}^* und t^* in Gleichung 2.1 die Regressionskonstante β den Mittelwert der 1924 Geborenen bei der Ersterhebung 1984 – in Tabelle 2.1 ist diese Zelle durch Umrandung hervorgehoben. Der aus Alter, Geburtsjahr und Meßzeitpunkt vorhersagbare Mittelwert irgendeiner anderen Zelle des Modells ergibt sich dann einfach durch Addition bzw. Subtraktion des Effekts α für jede Spalte, um die diese jeweilige Zelle gegenüber der „Referenzzelle“ nach rechts bzw. links verschoben ist, sowie Addition bzw. Subtraktion des Effekts γ für jede Zeile, um die diese Zelle gegenüber der umrandeten nach oben bzw. unten verschoben ist, und Addition von τ für jede Diagonale, um die diese Zelle nach rechts verschoben ist.

Zur Veranschaulichung des somit durch Gleichung 2.1 spezifizierten linearen Modells mit den nach den Gleichungen 2.2, 2.3 und 2.4 transformierten Prädiktoren soll Abbildung 2.1 dienen: Die fettgedruckten Pfeile symbolisieren den linearen Verlauf der Durchschnittswerte der abhängigen Variablen X_{ti} über das transformierte Lebensalter A_{ti}^* für die

verschiedenen Geburtsjahrgangsgruppen G_i^* : Jede Fortschaltung des Geburtsjahres um 1 bedeutet eine Verschiebung des Alters zum Beginn des ersten Meßzeitpunkts ($t^* = 0$) um -1 nach links auf der Altersachse. Die transformierten Meßzeitpunkte t^* ergeben sich hier aus der Kombination von A_{ti}^* und G_i^* , die eigentliche „historische“ Zeitdimension ist aus dem Koordinatensystem eliminiert, sie läuft über die als Diagonale eingezeichneten Erhebungswellen hinweg. Die gestrichelten Linien symbolisieren den allein durch einen linearen Alterseffekt vorherbestimmten längsschnittlichen Verlauf der Geburtsjahrgangsgruppenmittelwerte, d.h. die Geraden mit der Steigung α . Zu diesem Alterseffekt, d.h. zur Steigung α , ist dann ein Periodeneffekt als Steigungsbetrag τ hinzu zu addieren, so daß sich die in den fettgedruckten Pfeilen abgebildeten Verläufe ergeben. Der Kohorteneffekt γ wird in dieser Abbildung durch den vertikalen Abstand zwischen den gestrichelten Alterseffektgeraden der Geburtsjahrgänge repräsentiert: Beim Vergleich der Durchschnittswerte „benachbarter“ Geburtsjahrgänge zum jeweils gleichen Lebensalter ist eine Fortschaltung des Meßzeitpunktes impliziert, so daß diese Mittelwertsdifferenz den Periodeneffekt enthält – die um den Periodeneffekt reduzierte Mittelwertsdifferenz ist dann der Kohorteneffekt. Die Linearität von Alters- und Periodeneffekt ist durch die Geradenform der Verläufe abgebildet, die Linearität des Kohorteneffekts dadurch, daß die sukzessiven vertikalen Abstände der Verläufe nicht variieren. Der Ursprung der Gerade des Geburtsjahrgangs $G_i^* = 0$ auf der vertikalen Achse bezeichnet die Regressionskonstante β , nämlich den Mittelwert dieser Gruppe zum ersten Meßzeitpunkt bzw. im Alter $A_{ti}^* = 0$.

Es besteht nun die folgende Abhängigkeit zwischen den Prädiktoren der Modellgleichung 2.1:

$$G_i + A_{ti} = t$$

Und genauso:

$$G_i^* + A_{ti}^* = t^* \quad (2.5)$$

Das durch Gleichung 2.1 spezifizierte Modell wäre somit nicht schätzbar, weil zwischen den Prädiktoren perfekte Multikollinearität besteht. Um Schätzbarkeit zu erreichen muß einer der drei Prädiktoren eliminiert werden – im vorliegenden Fall interessieren Alters- und Kohorteneffekte, so daß es sich anbietet, den Meßzeitpunkt zu eliminieren. Dazu werden in Gleichung 2.1 die untransformierten durch die nach den Gleichungen 2.2 bis 2.4 transformierten Prädiktoren ersetzt und dann wird entsprechend Gleichung 2.5 umgeformt:

$$\begin{aligned} X_{ti} &= \beta + \alpha A_{ti}^* + \gamma G_i^* + \tau(G_i^* + A_{ti}^*) + \epsilon_{ti} \\ &= \beta + (\alpha + \tau)A_{ti}^* + (\gamma + \tau)G_i^* + \epsilon_{ti} \\ &= \beta + \alpha_{[\tau]}A_{ti}^* + \gamma_{[\tau]}G_i^* + \epsilon_{ti} \end{aligned} \quad (2.6)$$

Mit:

$$\alpha_{[\tau]} = \alpha + \tau \quad ; \quad \gamma_{[\tau]} = \gamma + \tau \quad (2.7)$$

Durch die Umformung wurde in Gleichung 2.6 ein Regressionsmodell erzeugt, das nur noch die beiden Prädiktoren Alter und Geburtsjahr enthält und somit im längsschnittlichen Untersuchungsdesign prinzipiell schätzbar ist, weil pro Person mehrere Messungen und somit auch pro Geburtsjahreswert mehrere Alterswerte vorliegen, so daß die beiden Prädiktoren nach Elimination von t^* unabhängige Varianz aufweisen. Wie aus der Umformung ersichtlich und in Gleichung 2.7 nochmal explizit formuliert, ist ein möglicher Periodeneffekt τ noch gewissermaßen unsichtbar im neuen Alters- und im neuen Kohorteneffekt des von t^* befreiten Modells enthalten, in diesen mit $\alpha_{[\tau]}$ und $\gamma_{[\tau]}$ benannten Effekten steckt jeweils der Periodeneffekt als additiver Anteil zusätzlich zum eigentlichen Alters- bzw. Kohorteneffekt.

In Abbildung 2.1 ist auch dieses veranschaulicht: Wenn der Periodeneffekt nicht berücksichtigt wird, dann würde dort gewissermaßen die gesamte Steigung der fettgedruckten Geraden als Alterseffekt $\alpha_{[\tau]}$ berechnet und entsprechend der Kohorteneffekt $\gamma_{[\tau]}$ als vertikaler Abstand zwischen *diesen* Geraden, in dem τ enthalten ist. Damit sind Schätzungen dieser beiden Koeffizienten, wenn das um t^* reduzierte Modell berechnet wird, stets mit der Unsicherheit behaftet, welchen Anteil der Periodeneffekt zu ihrem Betrag beisteuert. Immerhin aber wäre es jeweils der gleiche Betrag, der als Periodeneffektanteil in den Koeffizienten $\alpha_{[\tau]}$ und $\gamma_{[\tau]}$ enthalten ist, so daß bei Vergleichen dieser beiden Werte Unterschiede ganz und gar als Unterschiede zwischen Kohorten- und Alterseffekt erkannt werden können. Genau einen solchen Vergleich aber legt die Kohorteneffekthypothese nahe: Beide Effekte müßten negativ *und* in ihrem Absolutbetrag annähernd gleich groß sein, damit eine Überlagerung nach dem in Abbildung 1.1 gezeigten Schema zustande kommt. Wenn also annähernd gleich große Koeffizientenwerte $\alpha_{[\tau]}$ und $\gamma_{[\tau]}$ geschätzt würden, so wäre damit garantiert, daß Alters- und Kohorteneffekt gleich groß sind, sofern es sie überhaupt gibt und nicht beide Werte weitestgehend nur den Periodeneffekt τ enthalten.

Bei der Schätzung des Regressionsmodells 2.6 ist nun zu berücksichtigen, daß hierbei infolge der Meßwiederholung jede Person mehrfach mit ihren Meßwerten X_{ti} zu verschiedenen Meß- bzw. Lebensalterszeitpunkten vertreten ist. Somit liegt für die beiden Prädiktoren des Modells eine hierarchische Struktur vor: Die Variable G_i^* variiert über die Personen, während die Variable A_{ti}^* über die und innerhalb der Personen variiert. Bei der herkömmlichen Regressionsanalyse mit Kleinstquadrateschätzung würde eine Disaggregation der Personenlevelvariablen G_i^* dergestalt stattfinden, daß alle G^* -Werte des Datensatzes pro Person auf die Anzahl deren vorhandener A^* -Werte vervielfältigt werden, was zu einer künstlichen Vergrößerung der G^* -Fallzahlen und in der Folge zu Verzerrungen der Standardfehlerschätzung und der Signifikanztests für den Kohorteneffekt führen würde. Von anderer Warte aus betrachtet bestünde aber das ernsthafteste statistische Problem bei herkömmlicher Kleinstquadrat-Regression in einer Verletzung der Modellvoraussetzung unabhängiger Fehlerkomponenten (vgl. z.B. Werner, 1997, 196ff; Schilling, 1998, 209): Da jede Person sozusagen als mehrfacher Fall in die Berechnung miteingeht, ist mit Autokorrelationen der Fehlerkomponenten „innerhalb“ der Person zu rechnen. Denn

es ist ja davon auszugehen, daß die Zufriedenheitswerte der Personen im Zeitverlauf bis zu einem gewissen Grade intraindividuell stabil sind und also zwischen den Meßzeitpunkten korrelieren – genau diese Stabilität bzw. die sogenannte serielle Korrelation der Wiederholungsmessungen ist der Hauptgegenstand dieser Untersuchung in den noch folgenden Stabilitätsanalysen. Für das Modell 2.6 bedeutet dieses: Intraindividuell stabile Personeneigenschaften beeinflussen die Zufriedenheitswerte und sind deshalb als Teil der in der Fehlerkomponente zusammengefaßten Einflüsse anzusehen, für den serielle Abhängigkeiten bestehen. Somit führt die herkömmliche Regressionsrechnung hier zu verzerrten Schätzungen.

Die Lösung für dieses Problem liefern, wie bereits gesagt, die speziell für solche hierarchische Datenstrukturen entwickelten Multilevel-Regressionsmodelle. Für den vorliegenden Fall ist ein solches Modell dadurch zu spezifizieren, daß zunächst eine Level-1-Regressionsgleichung für die niedere Hierarchieebene der Meßzeitpunkte, mit dem Level-1-Prädiktor A_{ti}^* , spezifiziert wird:

$$X_{ti} = \beta_{i0} + \beta_{i1}A_{ti}^* + \epsilon_{ti} \quad (2.8)$$

Man beachte den Personenindex i bei den Regressionskoeffizienten der Gleichung 2.8: Dieser weist die Koeffizienten als jeweils für die Person i spezifische aus. β_{i0} ist der X -Wert der Person i im Alter $A_{ti}^* = 0$, β_{i1} ist der Betrag, um den der X -Wert von i mit jedem zusätzlichen Lebensjahr sich verändert. Die hierarchische Modellierung wird dadurch erzeugt, daß diese Effekte des Level-1-Modells 2.8 als über die Level-2-Einheiten, d.h. Personen, variierend spezifiziert werden. Durch die folgenden Gleichungen wird die Variation der Level-1-Koeffizienten über die Personen in Abhängigkeit vom Level-2-Prädiktor G_i^* modelliert:

$$\beta_0 = \beta_{00} + \beta_{01}G_i^* + v_{0i} \quad (2.9)$$

$$\beta_1 = \beta_{10} + \beta_{11}G_i^* + v_{1i} \quad (2.10)$$

Setzt man nun die Level-2-Gleichungen 2.9 und 2.10 in die Level-1-Gleichung 2.8 ein, so resultiert das folgende Multilevel-Regressionsmodell:

$$X_{ti} = \beta_{00} + \beta_{10}A_{ti}^* + \beta_{01}G_i^* + \beta_{11}A_{ti}^*G_i^* + v_{0i} + v_{1i}A_{ti}^* + \epsilon_{ti} \quad (2.11)$$

Dieses Modell enthält die Effekte des Modells 2.6, lediglich in anderer Benennung, die wie folgt zu übersetzen ist: $\beta_{00} = \beta$, $\beta_{10} = \alpha_{[\tau]}$, $\beta_{01} = \gamma_{[\tau]}$. Darüber hinaus enthält das Modell noch den Interaktion der Prädiktoren, welche aber ohne Weiteres entfernt werden könnte, wenn dies angebracht erschiene, und welche auch in Gleichung 2.6 bereits hätte aufgenommen werden können. Somit unterscheidet sich das Modell also in diesem systematischen Teil mit seinen fixen, nicht mehr über die Personen variablen Effekten β_{00} , β_{10} ,

β_{01} und β_{11} nicht von einem herkömmlichen Regressionsmodell. Der statistisch bedeutsame Unterschied besteht hier im nichtsystematischen Teil, nämlich darin, daß Modell 2.11 drei verschiedene Zufallskomponenten enthält, die zur nicht durch die fixen Effekte aufgeklärten Varianz der Kriteriumswerte X_{ti} beitragen. Demnach weichen die Meßwerte X_{ti} von den durch die fixen Effekte bestimmten Zellenmittelwerten des Paneldesigns ab, weil *erstens* die Personen nicht nur abhängig von ihrem Geburtsjahr, sondern darüber hinaus noch „zufällig“ im allgemeinen Basislevel ihrer X -Ausprägung – nämlich ihrem X -Wert im Alter $A_{ti}^* = 0$ – variieren (Zufallskomponente v_{0i}), weil sie *zweitens* auch noch nicht nur abhängig vom Geburtsjahr, sondern darüber hinaus noch zufällig in der „Veränderungsgeschwindigkeit“ des Merkmals X_{ti} über das Alter, d.h. in der Steigung ihrer individuellen X -Veränderungsverläufe, variieren (Zufallskomponente $v_{1i}A_{ti}^*$) und schließlich noch *drittens*, weil über diese interindividuellen Zufallsvariationen hinaus die Meßwerte bei den einzelnen Messungen nicht näher spezifizierten situativen Einflüssen ausgesetzt sind, welche eine weitere Quelle von Zufallsvarianz bilden (Komponente ϵ_{ti}). ϵ_{ti} enthält somit nicht mehr diejenigen Zufallseinflüsse auf die Meßwerte, die aus den im Modell nicht spezifizierten Eigenschaften der Level-2-Einheiten – der Personen – basieren und die daraus für das herkömmliche Regressionsmodell 2.6 resultierenden Autokorrelationen sind entfernt.

Das somit spezifizierte Multilevel-Modell 2.11 wurde also zur Untersuchung der Kohorteneffekthypothese verwendet. Zur Berechnung wurde die SAS-Prozedur MIXED in der von Singer (1998) gezeigten Weise benutzt. Die eigentliche Modellschätzung erfolgt über einen Maximum-Likelihood-Algorithmus („restricted maximum likelihood“: REML), der beispielsweise von Verbeke & Molenberghs (2000, 43ff) genauer dargestellt wird. Mit Blick auf die im folgenden noch eingehender diskutierte Problematik vieler Fälle mit fehlenden Werten infolge des unvermeidbaren längsschnittlichen Dropouts ist zu erwähnen, daß bei dieser Analyse Personen mit teilweise fehlenden Werten, also auch die Dropouts, nicht aus der Analyse ausgeschlossen, sondern mit ihren vorhandenen Datenwerten einbezogen werden. Es handelt sich damit um eines derjenigen Verfahren, die auch bei Datenausfällen, die die Bedingung „missing at random“ erfüllen, zu unverzerrten Schätzungen führt, wie Rubin (1976) bewiesen hat – dieses wird in Kapitel 2.3.1 genauer dargestellt.

2.2 Statistische Methoden der Analyse von Merkmalsstabilität

Im folgenden soll die statistische Methodologie für die angestrebte Untersuchung zur Stabilität von Zufriedenheit „erarbeitet“ werden. Dieses bedeutet v.a. zweierlei: Zum einen muß zunächst einmal der „technisch-statistische“ Stabilitätsbegriff, der den Stabilitätsanalysen zugrundegelegt wird, konkretisiert werden – d.h. es sind die statistischen „Zielparameter“, in denen die theoretisch gemeinte intraindividuelle Stabilität eines Merkmals zum Ausdruck kommt, zu benennen bzw. zu definieren. Zum zweiten muß die Auswahl eines Analysemo-

dells, mit dem diese Zielparameter bestmöglichst geschätzt werden können, begründet und müssen dessen Eigenschaften ausreichend dargestellt werden.

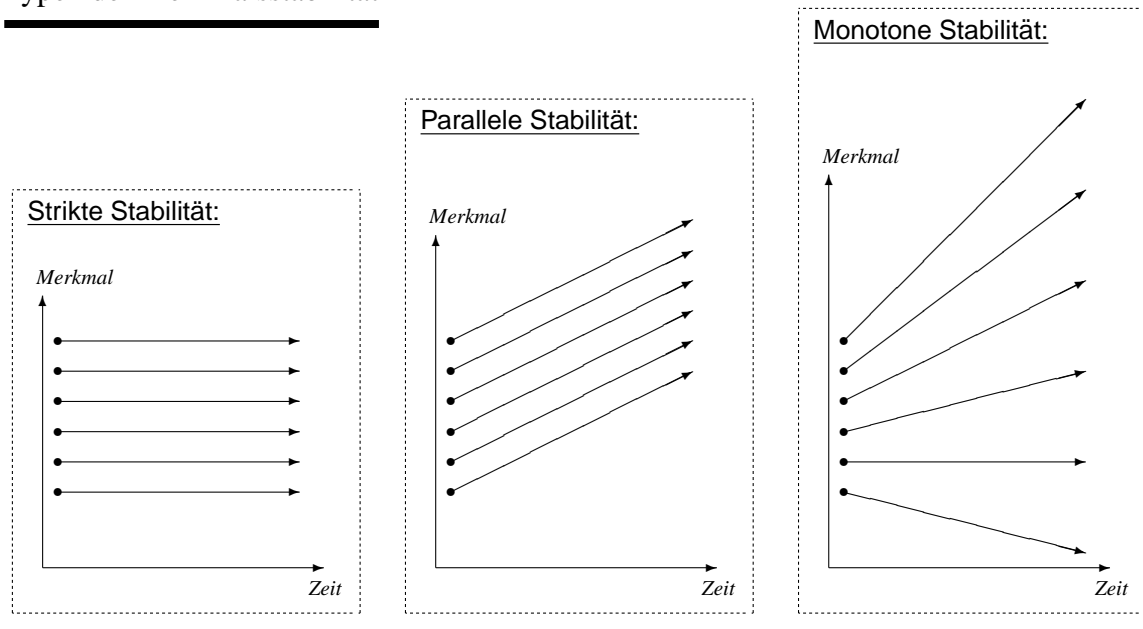
2.2.1 Typen und Kennwerte der Merkmalsstabilität

Verändern sich die Ausprägungen eines Merkmals über den Zeitverlauf hinweg „innerhalb“ der Individuen nur geringfügig und/oder nur bei wenigen Individuen in stärkerem Maße, so hätte dieses zur Folge, daß sowohl die Verteilungen dieses Merkmals zu verschiedenen Zeitpunkten sich kaum unterscheiden, als auch die Merkmalsausprägungen der Individuen zwischen verschiedenen Zeitpunkten hoch korrelieren. Dieser letztere Aspekt ist im Grunde der entscheidende, wenn Aussagen zum Ausmaß vorhandener Merkmalsstabilität gemacht werden sollen, denn wesentliche Verteilungsparameter wie Mittelwert und Varianz des Merkmals können im zeitlichen Verlauf auch unverändert bleiben, wenn die Merkmalsausprägungen große intraindividuelle Veränderungen durchlaufen (d.h. ohne daß das Merkmal selbst stabil ist). Dagegen begründet die intraindividuelle Invarianz der auf eine Population bezogenen *relativen* Merkmalsausprägung, die in korrelativen Zusammenhängen zwischen verschiedenen Zeitpunkten zum Ausdruck kommt, per se schon einen eingeschränkten Begriff von Stabilität: Stabil in diesem Sinne wäre ein Merkmal dann, wenn die Merkmalsträger durch alle Veränderungen hindurch sozusagen stets ihre Position in der Verteilung der Merkmalswerte halten.

Doch auch bei solchermaßen bloß relativer Stabilität sind doch zwei Fälle unterscheidbar. Wenn nämlich Zufriedenheitsveränderungen stets interindividuell invariant für alle Personen in gleichem Maße stattfänden, dann hätte das zur Folge, daß im Zeitverlauf nicht nur die relative, auf die Merkmalsverteilung in der Population bezogene Position der Merkmalsausprägung bei jedem Individuum invariant bliebe, sondern auch interindividuelle Abstände der Merkmalsausprägungen und damit die Varianz des Merkmals in der Population. Dagegen läge dann, wenn lediglich noch die relativen Positionen der Merkmalsausprägungen im Zeitverlauf invariant blieben, eine weniger strenge, allgemeinere Form relativer Stabilität vor: Die in einer bestimmten Zeitspanne stattfindenden Merkmalsveränderungen wären dann nicht mehr für alle Individuen dieselben, aber es bestünde immer noch ein hoher Zusammenhang zwischen den Merkmalsausprägungen zu verschiedenen Zeitpunkten bzw. zwischen der Merkmalsausprägungen zum Beginn einer bestimmten Zeitspanne und den Merkmalsveränderungen während derselben.

Es sind somit drei Typen von Merkmalsstabilität unterscheidbar, die von Tisak & Meredith (1990) als „strict“, „parallel“ und „monotonic“ bezeichnet wurden. Abbildung 2.2 illustriert diese Bezeichnungen und die gemeinten Stabilitätstypen anhand der schematischen Darstellung des jeweils idealtypischen Musters individueller Merkmalsverläufe über die Zeit. Unterscheidungskriterium dieser Typologie ist die zeitliche Invarianz verschiedener Aspekte in der Verteilung individueller Merkmalsausprägungen: Sowohl zentrale Tendenz, als auch Streuung, als auch die Rangwerte (Positionen) der Merkmalsausprägungen

ABBILDUNG 2.2:
Typen der Merkmalsstabilität



können invariant bleiben. Monotone Stabilität, als allgemeinsten und schwächsten Stabilitätstypus, besteht somit bei Invarianz lediglich der Positionen, während parallele Stabilität zusätzlich Invarianz der Varianzen und schließlich strikte Stabilität dazu noch Invarianz der Mittelwerte voraussetzt.

Wenn also die Stabilität eines Merkmals M während eines Zeitintervalls zwischen zwei Erhebungszeitpunkten t_1 und t_2 untersucht werden soll, so kann zunächst grundlegend formuliert werden:⁸

$$M_{2i} = M_{1i} + \Delta_i \quad (2.12)$$

Gleichung 2.12 bedeutet lediglich, daß der Wert eines Individuums i zum zweiten Zeitpunkt M_{2i} als Summe des Werts zum ersten Zeitpunkt M_{1i} und einer zwischenzeitlich stattgefundenen Veränderung Δ_i ausgedrückt wird.

Die zeitliche Invarianz der Positionen, als notwendige Bedingung irgendeiner Art von Merkmalsstabilität, kann durch die Korrelation der Merkmalswerte M_{1i} und M_{2i} (der Merkmalsträger i) zu beiden Zeitpunkten erfaßt werden – diese Korrelation ρ_{12} kann als *Stabilitätskoeffizient* verstanden werden, der Auskunft über das Ausmaß monotoner Stabilität gibt. Zur genaueren Typisierung der durch ρ_{12} erfaßten Merkmalsstabilität können dann

⁸Bei den formalen Ausdrücken bezeichnen $\sigma_1^2, \sigma_2^2, \sigma_{\Delta}^2$ die Varianzen der Merkmalswerte M_{1i}, M_{2i} und der Veränderungswerte Δ_i ; $\sigma_{12}, \sigma_{1\Delta}, \sigma_{\Delta 2}$ die Kovarianzen zwischen M_{1i} und M_{2i} , M_{1i} und Δ_i , Δ_i und M_{2i} ; sowie $\rho_{12}, \rho_{1\Delta}, \rho_{\Delta 2}$ die dementsprechenden Korrelationen.

entweder die beiden Verteilungen von M zu t_1 und t_2 verglichen und die Veränderung der Varianzen σ_1^2 und σ_2^2 und der Mittelwerte μ_1 und μ_2 untersucht werden. Oder es kann Ausmaß und interindividuelle Variabilität der Merkmalsveränderungen direkt untersucht werden: Dazu wären die Merkmalswerte M_{1i} und die nachfolgenden Veränderungen derselben, d.h. die Differenzwerte Δ_i zu analysieren.

Unter Anwendung bekannter Formeln (z.B. Hays 1994; Schilling, 1998) gilt dann:

$$\mu_2 = \mu_1 + \mu_\Delta \quad (2.13)$$

$$\sigma_2^2 = \sigma_1^2 + \sigma_\Delta^2 + 2\sigma_{1\Delta} \quad (2.14)$$

$$\sigma_{12} = \sigma_1^2 + \sigma_{1\Delta} \quad (2.15)$$

$$\sigma_\Delta^2 = \sigma_1^2 + \sigma_2^2 - 2\sigma_{12} \quad (2.16)$$

Und:

$$\rho_{12} = \frac{\sigma_{12}}{\sigma_1\sigma_2} = \frac{\sigma_1^2 + \sigma_{1\Delta}}{\sqrt{\sigma_1^2(\sigma_1^2 + \sigma_\Delta^2 + 2\sigma_{1\Delta})}} \quad (2.17)$$

Die Fälle strikter und paralleler Stabilität beschreiben restriktive Eigenschaften der Veränderung des Merkmals, welche hier nun als entsprechende Annahmen über Verteilungskennwerte der Veränderungsbeträge Δ_i übersetzt werden können, nämlich durch Nullsetzung von Mittelwert bzw. Varianz von Δ_i :

$$\mu_\Delta = 0 \iff \mu_1 = \mu_2 \quad (2.18)$$

$$\sigma_\Delta^2 = 0 \iff \sigma_{1\Delta} = 0 \quad \underline{\text{und}} \quad \sigma_1^2 = \sigma_2^2 \quad \underline{\text{und}} \quad \rho_{12} = 1 \quad (2.19)$$

Gleichung 2.19 bezeichnet den Typus paralleler Stabilität: Bei jedem Merkmalsträger findet dieselbe Veränderung statt, d.h. die Merkmalsveränderungen variieren nicht interindividuell, weshalb natürlich auch die Veränderungen Δ_i nicht mit den Ausgangswerten M_{1i} kovariieren. Folglich haben die Merkmalswerte M_{1i} und M_{2i} dieselbe Varianz (vgl. Gleichung 2.14) und sind perfekt positiv korreliert (vgl. Gleichung 2.17). Genaugenommen gilt in Gleichung 2.19 die Folgerung von rechts nach links immer dann, wenn mindestens zwei der drei rechtsseitigen Bedingungen erfüllt sind.

Zusammen mit Gleichung 2.18 beschreibt 2.19 den Typus strikter Stabilität: Die Veränderung streut nicht nur nicht interindividuell, sondern findet gar nicht statt.

Von den Korrelationen der Ausgangs- mit den Veränderungswerten kann nicht ohne Weiteres auf die zwischen den Merkmalswerten verschiedener Zeitpunkte geschlossen werden und v.a. sind erstere kein Ausdruck vorhandener Merkmalsstabilität, sondern erweist sich die Bewertung der Informationen, die ein gegebener Wert von $\rho_{1\Delta}$ hinsichtlich der Stabilität des Merkmals bietet, als durchaus vielschichtige und komplexe Aufgabe. Deshalb ist

der Stabilitätskoeffizient ρ_{12} hinsichtlich der Frage vorhandener Stabilität eines Merkmals der informativere Korrelationswert, denn er beinhaltet die eindeutige Information darüber, in welchem Maße die Merkmalsträger zumindest ihre relativen Merkmalsausprägungen über die Zeit hinweg beibehalten haben.

Ähnlich verhält es sich auch mit den dargestellten Varianzen: Es versteht sich von selbst, daß in der Veränderungsvarianz σ_{Δ}^2 prinzipiell etwas anderes zum Ausdruck kommt, als in der Varianz der veränderten Werte σ_2^2 , und dementsprechend besitzen beide Varianzen auch unterschiedlichen Informationsgehalt für die Analyse von Stabilität: σ_2^2 ist gewissermaßen überhaupt nur aussagekräftig im Zusammenhang mit σ_1^2 und ρ_{12} – wie dargelegt zeigt bei $\rho_{12} = 1$ die Gleichheit von σ_1^2 und σ_2^2 an, daß es sich um den Typus paralleler Stabilität handelt. Dagegen ist diese Anzeige durch $\sigma_{\Delta}^2 = 0$ direkt und ohne weitere Kennwerte möglich, denn der Begriff der parallelen Stabilität benennt ja genau diesen Umstand, daß die Veränderung nicht zwischen den Merkmalsträgern variiert.

In „realen“ längsschnittlichen Daten sind nun jedoch normalerweise keine perfekten Stabilitäten im Sinne der oben beschriebenen Typen zu erwarten. Die Frage lautet dann nicht: Ist das Merkmal (monoton/parallel/strikt) stabil? Sondern: *Wie* (monoton/parallel/strikt) stabil ist das Merkmal? Wenn also beurteilt werden soll, ob die mit einem vorliegenden ρ_{12} -Wert ermittelte Stabilität nicht bloß eine monotone, sondern vom parallelen Typus ist, so muß ein Weg gefunden werden, diese Frage in quantifizierender Weise zu beantworten. Denn da $\sigma_{\Delta}^2 = 0$, d.h. perfekte Parallelität, stets $\rho_{12} = 1$ impliziert (vgl. Gleichung 2.19), kann es umgekehrt bei $\rho_{12} < 1$ im strengen Sinne parallele Veränderungen gar nicht geben. Jedoch wäre zu fragen, ob man ein gewisses Maß von „Nicht-Parallelität“ noch so auffassen kann, daß die Veränderung als „annähernd parallele“ zu charakterisieren und die Veränderungsvarianz als relativ geringfügig zu bewerten ist. Dabei ist zu berücksichtigen, daß die fragliche Parallelität eng an die Voraussetzung hoher monotoner Stabilität der Merkmalsverläufe geknüpft ist: Je geringer die Kovarianz σ_{12} , desto höher wird die Veränderungsvarianz σ_{Δ}^2 (vgl. Gleichung 2.16) – je mehr also ρ_{12} vom Maximalwert 1 abweicht, desto weniger parallel können die Verläufe sein.

Ein Weg, der zur Bewertung der Parallelität der Merkmalsverläufe eingeschlagen werden könnte, ist der altbekannte der Signifikanztestung: Wenn der Stichprobenwert der Veränderungsvarianz nicht signifikant von Null abweicht, so könnte man dieses als Nachweis paralleler Stabilität werten. Die Nullhypothese dieses Tests lautet also $\sigma_{\Delta}^2 = 0$ und impliziert damit auch $\rho_{12} = 1$, so daß der Test schlicht überprüft, ob die vorhandenen Daten mit einem restriktiven Veränderungsmodell *perfekter* Parallelität vereinbar sind, bei dem sozusagen lediglich ein zufälliges Flimmern in den intraindividuellen Merkmalsverläufen um einen für alle Merkmalsträger gültigen, durchschnittlichen Merkmalsverlauf herum Veränderungsvarianz erzeugt und ρ_{12} mindert. Allerdings wäre diese Vorgehensweise mit zwei bekannten und in den meisten Statistiklehrbüchern neueren Datums (z.B. Schilling, 1998, 142ff) diskutierten Problemen der Signifikanztestung verbunden: Zum einen wird dabei die Annahme paralleler Stabilität an die Akzeptanz der Nullhypothese $\sigma_{\Delta}^2 = 0$ ge-

knüpft, so daß die üblichen Signifikanzniveaus (z.B. $p \leq 0.05$) kaum eine kritische Prüfung dieser Annahme bedeuten, zum zweiten bestünde natürlich auch dabei die Abhängigkeit der Teststärke von der Stichprobengröße, so daß im Grunde letztere darüber entscheidet, ob ein bestimmter Veränderungsvarianzwert als Indikator paralleler Stabilität gewertet wird, oder nicht.

Um zu einem quantifizierenden Urteil über das Ausmaß näherungsweise Parallelität der Verläufe bei vorhandener Veränderungsvarianz $\sigma_{\Delta}^2 > 0$ und nicht perfekter Stabilität $\rho_{12} < 1$ zu gelangen, könnte der Veränderungsvarianzwert auf irgendeine Art und Weise anhand der „sonstigen“, zeitlich invarianten Merkmalsvarianz relativiert werden. Also könnte σ_{Δ}^2 z.B. zu σ_1^2 oder σ_2^2 ins Verhältnis gesetzt werden – m.E. eignet sich σ_1^2 hierfür am besten, weil σ_2^2 ja bereits die Veränderungsvarianz enthält (vgl. Gleichung 2.14) und durch diese infolge der möglichen Kovarianz $\sigma_{1\Delta}$ sowohl vergrößert, als auch verringert werden kann. Dagegen bezeichnet σ_1^2 gewissermaßen den über die beiden Zeitpunkte hinweg invarianten, Anteil „unveränderter“ Merkmalsvarianz. Somit könnte der folgende Quotient zur Beurteilung dienen, ob σ_{Δ}^2 als hinreichend geringfügig einzuschätzen ist, um noch eine näherungsweise Parallelität der Verläufe anzunehmen:

$$Q_{\Delta} = \frac{\sigma_{\Delta}^2}{\sigma_1^2} \quad (2.20)$$

Der Wert von Q_{Δ} gibt also schlicht an, wie hoch die Veränderungsvarianz im Vergleich zur Varianz der Ausgangswerte ist. Q_{Δ} kann auch dann, wenn die Veränderungswerte Δ_i bzw. deren Veränderungsvarianz gar nicht direkt analysiert werden, berechnet werden. Ersetzt man σ_{Δ}^2 in Gleichung 2.20 entsprechend Gleichung 2.16, so resultiert:

$$Q_{\Delta} = \frac{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}}{\sigma_1^2} = 1 + \frac{\sigma_2^2 - 2\sigma_{12}}{\sigma_1^2} \quad (2.21)$$

Es könnte nun beispielsweise ein Schwellenwert für Q_{Δ} festgelegt werden, bis zu dem noch die Veränderungen als näherungsweise parallel eingestuft werden, z.B. $Q_{\Delta} \leq 0.1$.

Wenn über die Einschätzung näherungsweise paralleler Stabilität hinaus auch noch gefragt wird, ob ein Merkmal annähernd *strikt* stabil ist, so ist damit nun keine weitere Komplizierung über die dargestellte Problematik der Einschätzung näherungsweise Parallelität hinaus verbunden. Wie nach allen bisherigen Darlegungen unmittelbar einsehbar sein sollte, bedeutet strikte Stabilität schlicht parallele Stabilität *ohne* Mittelwertsverschiebung. Zur Einschätzung annähernd strikter Stabilität muß deshalb v.a. die näherungsweise Parallelität untersucht und dann zusätzlich sichergestellt werden, daß die Merkmalsmittelwerte μ_1 und μ_2 sich nicht substantiell unterscheiden oder daß der Mittelwert der Veränderungsbeträge nicht wesentlich von Null abweicht. Hierfür stehen die allseits bekannten statistischen Methoden zur Testung von Mittelwertsdifferenzen (z.B. die bekannten *t*-Tests) zur Verfügung, welche allerdings wiederum mit den soeben erwähnten Problemen der mangelhaften Kontrolle des β -Fehlerrisikos bei Akzeptanz der Nullhypothese und der Abhängig-

keit der Teststärke von der Stichprobengröße behaftet sind. Vorhandene Mittelwertsabweichungen können auch schlicht aufgrund ihrer praktischen Bedeutsamkeit oder durch Rückgriff auf Effektstärkeindikatoren (z.B. Cohen, 1992) eingeschätzt werden.

Somit lassen sich die Ziele statistischer Analyse von Merkmalsstabilität wie folgt zusammenfassen:

- ▶ Wichtigster „Zielparameter“ der Analyse von Merkmalsstabilität ist der Stabilitätskoeffizient ρ_{12} als quantifizierendes Maß der Stabilität: Hohe positive Korrelation der Merkmalswerte M_{1i} und M_{2i} bedeutet hohe (monotone) Stabilität.
- ▶ Parallelität der Merkmalsveränderungen bzw. der Typus paralleler Stabilität beruht auf der interindividuellen Varianz der Veränderungswerte σ_{Δ}^2 : Parallelität im strengen Sinne liegt dann vor, wenn $\sigma_{\Delta}^2 = 0$, bei geringfügigen Beträgen der Veränderungsvarianz liegt näherungsweise Parallelität vor. Zur Einschätzung eines Ausmaßes näherungsweise Parallelität könnte der Quotient Q_{Δ} entsprechend der obigen Darlegungen untersucht werden.
- ▶ Strikt stabil ist ein Merkmal dann, wenn keine Veränderung stattfindet, so daß $\rho_{12} = 1$, $\sigma_{\Delta}^2 = 0$ und $\mu_{\Delta} = 0$. Praktisch bedeutet dies: Parallele Stabilität plus Invarianz der Merkmalsmittelwerte im Zeitverlauf, entsprechend ist eine näherungsweise strikte Stabilität dadurch zu zeigen, daß die jeweils näherungsweise Parallelität der Veränderungsverläufe und Invarianz der Mittelwerte gezeigt wird.

2.2.2 Längsschnittliche Strukturgleichungsmodelle

Bei Analysen der Merkmalsstabilität spielt, wie in Kapitel 2.2.1 dargelegt wurde, der Stabilitätskoeffizient ρ_{12} , d.h. die Korrelation der Merkmalsausprägungen M_{1i} und M_{2i} zu den Endpunkten eines jeweils betrachteten Zeitintervalls $t_1 \mapsto t_2$, eine zentrale Rolle als Maß des tatsächlichen Zusammenhangs zwischen Merkmalswerten über die Zeit hinweg. Bei der Schätzung desselben ist nun allerdings die *Reliabilität* der Messungen zu berücksichtigen. Wenn im hier vorliegenden Fall zum Zwecke der angestrebten Stabilitätsanalysen Zufriedenheitsratings aus dem SOEP untersucht werden, dann können die Korrelationen zwischen diesen wiederholten Zufriedenheitsmessungen nicht ohne weiteres mit Stabilitäten von Zufriedenheit gleichgesetzt werden: Die tatsächliche, durch die Korrelation der meßfehlerbereinigten „wahren“ Merkmalsausprägungen zu unterschiedlichen Zeitpunkten ausgedrückte (monotone) Stabilität dürfte generell über solchen Meßwertkorrelationen liegen, wobei bei gegebener Korrelation der Meßwertreihen niedrigere Reliabilitätsschätzungen zu höheren Schätzungen des „wahren“ Stabilitätskoeffizienten (d.h. der Korrelationen zwischen den „wahren“ Werten) führen, und umgekehrt höhere Reliabilitäten zu niedrigeren Stabilitäten. Denn die (klassische) Testtheorie konzipiert Reliabilität bekanntlich als Anteil der „wahren Varianz“ des gemessenen Merkmals an der Meßwertvarianz (Testpunktwerte, Skalenwerte etc.), welche zudem durch Meßfehler, d.h. außer der tatsächlichen Merkmalsausprägung noch wirksame sonstige Einflüsse auf die Meßwerte, vergrößert wird.

Wenn also bisher, in Kapitel 2.2.1, allgemein von Merkmalswerten M_{ti} die Rede war, so sollte nun zunächst einmal genauer spezifiziert werden, ob damit jeweils die fehlerbehafteten Meßwerte, oder die „wahren“ Merkmalsausprägungen gemeint sind. Im folgenden werden die Meßwerte (des Merkmalsträgers i zum Meßzeitpunkt t) mit X_{ti} und die „wahren“ Werte entsprechend englischsprachiger Benennungskonventionen mit T_{ti} (für True-Scores) bezeichnet. Grundsätzlich läßt sich dann das testtheoretische Modell, das der Messung zugrundeliegt, in folgender Gleichung ausdrücken:

$$X_{ti} = T_{ti} + E_{ti} \quad (2.22)$$

E_{ti} bezeichnet hier die Meßfehlerbeträge und Gleichung 2.22 besagt somit nichts weiter, als daß ein Meßwert X_{ti} sich zusammensetzt aus dem True-Score, d.h. der tatsächlichen Ausprägung des gemessenen Merkmals beim Merkmalsträger i zum Meßzeitpunkt t , und eben einer Abweichung von diesem True-Score aufgrund der Fehlerbehaftetheit der Messung.

Aus 2.22 ergibt sich nun eine Zweideutigkeit hinsichtlich des bisher, in Kapitel 2.2.1, Dargestellten: Die Variablenwerte M_{1i} und M_{2i} , sowie die statistischen Kennwerte – und entsprechend ihre wechselseitigen Zusammenhänge –, die dort betrachtet wurden, haben gewissermaßen stets zwei Bedeutungen. Zum einen können damit Meßwerte gemeint sein, so daß beispielsweise ρ_{12} die Korrelation zwischen den Meßwerten X_{1i} und X_{2i} und σ_{Δ}^2 die Varianz der Veränderungen in den Meßwerten meint. Zum anderen aber können damit die True-Scores gemeint sein, so daß die beiden zuletzt genannten Kennwerte nun der True-Score-Korrelation bzw. der Varianz der True-Score-Veränderungen entsprechen. Die in Kapitel 2.2.1 gezeigten Ableitungen und Zusammenhänge bleiben unter beiden Bedeutungen gleich, die untersuchungsrelevante Frage ist lediglich, welche Bedeutung durch das jeweilige theoretische Interesse impliziert ist, d.h. ob man die Stabilität der Meßwerte oder der True-Scores untersuchen will. Im letzteren Fall müßten die betreffenden Kennwerte, z.B. der Stabilitätskoeffizient ρ_{12} oder die Veränderungsvarianz σ_{Δ}^2 jeweils meßfehlerbereinigt als Kennwerte der Verteilung der True-Scores T_{ti} geschätzt werden, was eine zusätzliche statistische Komplikation darstellt. Eine Untersuchung, deren theoretisches Interesse auf „inhaltlichen“ Hypothesen über die Stabilität eines Konstrukts Zufriedenheit beruht, meint damit natürlich die „wahren“ Werte, so daß im vorliegenden Fall diese Komplikation nicht gescheut werden darf.

Um Mißverständnisse zu vermeiden, die sich aus der besagten Zweideutigkeit der in Kapitel 2.2.1 dargestellten Kennwerte ergeben könnten, bezeichne ich im weiteren die Korrelation der True-Scores T_{1i} und T_{2i} mit $\rho_{T_1T_2}$ und belasse den Ausdruck ρ_{12} zur Bezeichnung der Meßwertkorrelation. Auch bei den anderen in 2.2.1 eingeführten formalen Bezeichnungen ergänze ich jeweils die Indizierung um ein T oder E , wenn die entsprechenden „rechtsseitigen“ Komponenten des Modells der Gleichung 2.22 gemeint sind. Insbesondere bezeichnet also $\rho_{T_1T_2}$ den „wahren“ Stabilitätskoeffizienten, der bei den nachfolgenden Stabilitätsanalysen interessiert.

Die ursprünglichste Form des Meßmodells für Gleichung 2.22 betrachtet die Meßfehler als reine Zufallskomponenten, die mit den wahren Werten und bei Meßwiederholung auch untereinander unkorreliert sind. Unter diesen Voraussetzungen aber ergibt sich bei Meßwiederholungen die Korrelation der Meßwerte aus den beiden Messungen als Produkt der Korrelation der wahren Werte mit dem geometrischen Mittel der Reliabilitäten der beiden Messungen:

$$\rho_{12} = \rho_{T_1T_2} \sqrt{Rel_1 Rel_2} \quad (2.23)$$

Da die Reliabilitäten im Wertebereich von 0 bis 1 liegen, ist damit die Meßwertkorrelation nur dann gleich der Korrelation der wahren Werte, wenn die Reliabilitäten gleich 1 sind (d.h. eine fehlerfreie Messung vorliegt) und liegt ansonsten unter der Korrelation der wahren Werte. Formal und detaillierter sind diese Zusammenhänge im Anhang A.1 dargestellt, wo auch eine Herleitung der Gleichung 2.23 gezeigt wird.

Die Reliabilitäten der Zufriedenheitsratings beinhalten demnach sozusagen einen möglichen Störfaktor für die Analyse intraindividuelle Stabilität der Zufriedenheit, der im folgenden zu berücksichtigen ist. Deshalb ist die Schätzung meßfehlerbereinigter Stabilitätskoeffizienten eine Anforderung an die statistischen Modelle für diese Analysen, d.h. es sollten in diesen Modellen Meßfehleranteile an den gemessenen Zufriedenheitswerten identifizier- und Reliabilitäten schätzbar sein.

Die statistische Modellierung längsschnittlicher Stabilitätsanalysen muß also explizit zwischen dem Modell der Messung „wahrer“ Zufriedenheitswerte und dem Modell der längsschnittlichen strukturellen Beziehungen zwischen diesen wahren Werten unterscheiden. Genau diese Differenzierung ist eine der „Spezialitäten“ von Strukturgleichungsmodellen (häufig mit dem aus dem englischen Begriff „structural equation models“ abgeleiteten Kürzel SEM bezeichnet): Das prinzipielle Rationale dieses Typus statistischer Verfahren besteht darin, die „strukturellen“ theoretischen Konstrukte, d.h. diejenigen, die durch die jeweilige inhaltlich-theoretische Fragestellung zueinander in Beziehung gesetzt werden, als nicht direkt beobachtete, latente Variablen zu behandeln, welche durch beobachtete Indikatoren, d.h. die im analysierten Datensatz erhobenen Variablen, jeweils nur fehlerbehaftet gemessen werden. Die Strukturgleichungsmodelle leisten somit die simultane Schätzung von Meßmodell (Beziehungen latenter True-Scores zu ihren beobachteten Indikatoren) und Strukturmodell (Beziehungen latenter True-Scores untereinander) – und sind deshalb erste Wahl als statistische Methode für die nachfolgenden Stabilitätsanalysen. Die ausführliche allgemeine Darstellung der Strukturgleichungsmodelle kann im Rahmen der hier vorgelegten Arbeit nicht gegeben werden, zudem haben sie v.a. während der letzten Dekade durchaus eine gewisse Verbreitung erfahren und können allmählich gewissermaßen zu den Standards quantitativer multivariater Analysemethoden in den Sozial- und Verhaltenswissenschaften gezählt werden, so daß sie hier nicht eigens im Detail dargestellt werden müssen. Stattdessen kann auf die inzwischen große Zahl entsprechender Darstellungen verwiesen werden, unter denen z.B. der „Klassiker“ von Bollen (1989) und die Monographie von

Hoyle (1995) hervorzuheben sind (kürzere einführende Darstellungen liefern z.B. Long, 1983; Rietz et al., 1996; Hox & Bechger, 1998).

In der spezifischen Literatur sind Strukturgleichungsmodelle zur längsschnittlichen Datenanalyse ein durchaus intensiv behandeltes Thema (zum Überblick: McArdle & Aber, 1990; Bijleveld et al., 1998; Rudinger & Rietz, 2001). Die längsschnittlichen Strukturgleichungsmodelle können wohl derzeit zum „state of the art“ fortgeschrittener längsschnittlicher statistischer Methoden gezählt werden und diese wachsende Bedeutung verdankt sich wohl v.a. der großen Flexibilität, die die Strukturgleichungsmethode bei der Modellierung von Veränderung über die Zeit bietet: „The main benefit of SEM comes because it allows us to formalize different kinds of change concepts. SEM makes it easier to represent and test change concepts, and this makes it easy to communicate, think about, and compare different change ideas. This makes it possible, for example, to use SEM to contrast standard ‘change in parameters’ models with ‘parameters of change’ models“ (McArdle & Aber, 1990, 210). In der Vielzahl entwickelter längsschnittlicher Strukturgleichungsmodelle lassen sich bestimmte Typen solcher Modelle erkennen: McArdle & Aber (1990) unterscheiden beispielsweise vier Typen, nämlich das „autoregressive change model“, das „difference components change model“, das „growth curve change model“ und das „factor analysis change model“. Rudinger & Rietz (2001) dagegen unterscheiden prinzipiell autoregressive von faktoranalytischen Modellierungen der Stabilität bzw. Variabilität in Zeitreihen von Variablenwerten, in ihrem Sinne kann man die letzten drei von McArdle & Aber behandelten Modelle als faktoranalytische bezeichnen.

Die Datengrundlage aller Längsschnittmodelle ist eine Zeitreihe von beobachteten, wiederholt gemessenen Variablen, die Meßwerte zu unterschiedlichen Meßzeitpunkten korrelieren untereinander in Abhängigkeit von der Stabilität der gemessenen wahren Werte und der Reliabilitäten der Messungen (siehe oben bzw. Anhang A.1) – diese längsschnittlichen Korrelationen (bzw. Kovarianzen) werden als Autokorrelationen (bzw. Autokovarianzen) bezeichnet. Man kann Autokorrelationen unterschiedlicher Ordnung aufgrund des „time-lag“ zwischen den beiden Meßwertreihen, d.h. der Distanz der beiden Meßzeitpunkte, unterscheiden: Die Korrelation der Meßwerte eines Meßzeitpunkts t mit denen des Meßzeitpunkts $t + k$ kann als Autokorrelation k -ter Ordnung bezeichnet werden.

Autoregressive und faktoranalytische Modelle unterscheiden sich nun darin, *wie* sie die Autokovarianzen zwischen den Meßzeitpunkten modellieren. In autoregressiven Modellen werden die Autokovarianzen durch die Regression der Variablenwerte eines jeden Meßzeitpunkts (außer dem ersten) auf die Variablenwerte früherer Meßzeitpunkte spezifiziert. In faktoranalytischen Längsschnittmodellen werden die längsschnittlichen Kovarianzen sozusagen auf theoretisch übergeordnete Varianzquellen zurückgeführt, also auf Faktoren im Sinne der konfirmatorischen Faktorenanalyse, welche – v.a. mit den diesbezüglichen Arbeiten von Jöreskog (1967; 1969) – den Ursprung der Entwicklung von Strukturgleichungsmodellen mit latenten Konstrukten markiert (vgl. auch: Schönemann & Borg, 1996). In einführenden Anwendungsbeispielen von (explorativen und/oder konfirmatorischen) Fak-

torenanalysen (z.B. bei Stevens, 1996, 362ff; Lewis-Beck, 1994; Arminger, 1979) handelt es sich meist um querschnittliche Analysen und haben die Faktoren in der Regel eine theoretische Bedeutung als theoretisch beschriebene und inhaltlich valide Konstrukte, d.h. es handelt sich z.B. um psychologische, soziokulturelle, gesundheitliche Merkmale der Untersuchungsobjekte. Die Faktoren in faktoranalytischen Längsschnittmodellen haben dagegen meist eine eher „technische“ Bedeutung dergestalt, daß sie lediglich der Modellierung von Variabilität und Stabilität von Merkmals- oder Meßwerten im zeitlichen Verlauf dienen, d.h. es werden hier theoretische Veränderungsmodelle so in Strukturgleichungsmodelle übersetzt, daß die latenten Variablen zeitlich stabile oder veränderliche Anteile an den Merkmals- oder Meßwerten enthalten.

Der Einfachheit halber wird im folgenden stets davon ausgegangen, daß zwischen den Meßwiederholungen jeweils ein Zeitintervall gleicher Länge liegt: Dieses vereinfacht die Darstellung der Modelle und ist im SOEP – mit *jährlichen* Wiederholungsbefragungen – ungefähr der Fall (allgemeinere Modellspezifikationen für ungleiche Zeitintervalle sind in der referenzierten Literatur zu finden). Außerdem wird die Darstellung auf Ein-Indikator-Modelle, d.h. auf Strukturgleichungsmodelle für Zeitreihen von jeweils nur *einem* Meßwert für die zu untersuchende wahre Merkmalsausprägung pro Meßzeitpunkt beschränkt. Ist für jeden Meßzeitpunkt mehr als ein Indikator vorhanden, durch den das Merkmal gemessen wird, so können die dargestellten Modelle ohne weiteres zu Mehr-Indikatoren-Modellen erweitert werden – die jeweilige Spezifikation der True-Score-Veränderung bleibt davon im Grunde unberührt. Allerdings sind Mehr-Indikatoren-Modelle leichter identifizierbar, da ja in solchen mehr „beobachtete Information“ zur Schätzung der Modellparameter vorhanden ist, so daß manche der nachfolgend noch dargestellten, zur Modellidentifikation notwendigen Restriktionen bei solchen Modellen nicht mehr notwendig sind.

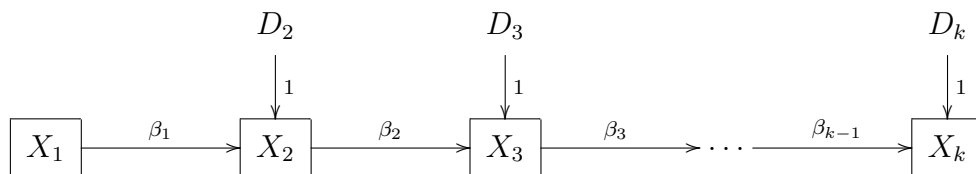
2.2.2.1 Das Quasi-Markov-Simplex-Modell

Die wohl gebräuchlichste Form autoregressiver Modelle sind solche erster Ordnung, die auch als Markov-Modelle bezeichnet werden: Dabei wird jeder Meßzeitpunkt der Zeitreihe lediglich auf den jeweils vorangegangenen Meßzeitpunkt regrediert. In einem autoregressiven Modell zweiter Ordnung würde jeder Meßzeitpunkt in der Regression auf die letzten beiden vorangegangenen Zeitpunkte modelliert usw.. Man könnte solche Modelle direkt für die beobachteten Variablen, d.h. ohne latente Konstrukte und damit ohne explizite Berücksichtigung der Reliabilitäten, spezifizieren – ein solches autoregressives Modell erster Ordnung zeigt Abbildung 2.3.A.⁹ Jöreskog (1970) bezeichnet dieses Modell

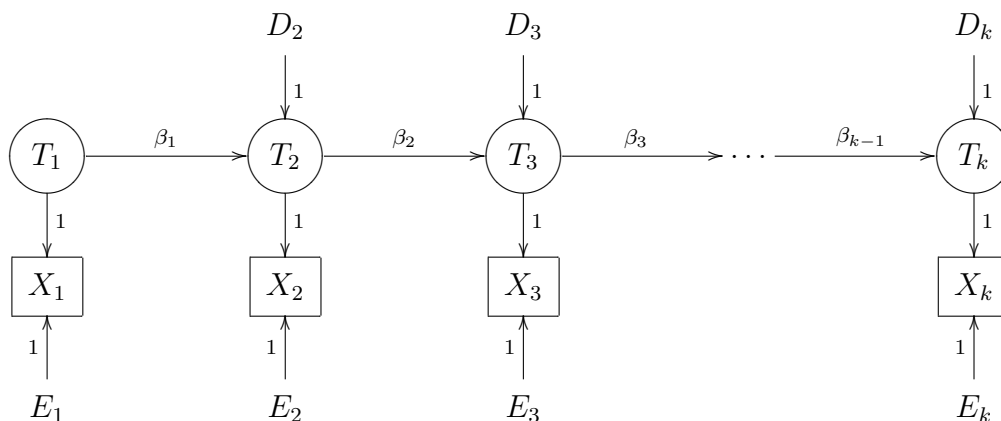
⁹Bei graphischen Darstellungen von Strukturgleichungsmodellen werden die üblichen Konventionen befolgt, wie sie z.B. bei Bollen (1989) eingehalten werden. D.h. Rechtecke stellen beobachtete und Kreise latente Variablen dar, die Random- bzw. Fehlerkomponenten der Strukturgleichungen werden ohne Umrandung abgebildet. Pfeile bezeichnen Pfadkoeffizienten, gebogene Doppelpfeile zwischen verschiedenen Modellkomponenten Kovarianzen. Varianzen werden bei dieser gängigen Darstellungsweise meist nicht abgebildet, prinzipiell sind die Varianzen aller exogenen Variablen (d.h. aller Variablen, auf die kein Pfeil zeigt)

ABBILDUNG 2.3:
Autoregressive Strukturgleichungsmodelle erster Ordnung

A Markov-Simplex-Modell:



B Quasi-Markov-Simplex-Modell (QMSM):



als Markov-Simplex-Modell, da es eine Simplex-Struktur der Autokorrelationen impliziert: „The typical property in a simplex correlation structure . . . is that the correlations decrease as one moves away from the main diagonal“ (ebd., 121) – einfach ausgedrückt bedeutet dieses für die Autokorrelationen, daß sie praktisch mit wachsendem time-lag abnehmen.

Aus den oben (und in Anhang A.1) dargelegten Gründen wäre für eine unverzerrte Analyse von Merkmalsstabilität ein autoregressives Modell der wahren Werte notwendig – Abbildung 2.3.B zeigt ein solches. Jöreskog (1970) bezeichnet dieses Modell, in dem zwar durch die Autoregressivität erster Ordnung für die Autokorrelationen der beobachteten Werte wiederum die Simplex-Struktur impliziert ist, welche jedoch durch die Meßfehlerkomponente sozusagen aufgeweicht sein kann (genaugenommen ist hier die Simplex-Struktur für die Autokorrelationen der wahren Werte impliziert), als Quasi-Markov-Simplex-Modell (QMSM). Wie bereits erwähnt, könnte ein solches Markov-Modell für die True-Scores auch

Modellparameter – wenn sie graphisch sichtbar gemacht werden sollen, dann durch einen gebogenen Doppelpfeil, dessen Enden beide auf die entsprechende Variable weisen. Dreiecke bezeichnen Konstanten (in der Regel den Wert 1), die in Strukturgleichungsmodellen nur dann als Prädiktor eingesetzt werden, wenn eine Mittelwertsstruktur geschätzt werden soll.

entsprechend für mehrere Indikatoren pro Meßzeitpunkt spezifiziert werden. Formal läßt sich dieses QMSM für die in einer Stichprobe der Größe n erhobenen Werte einer Variablen X relativ einfach durch die folgenden beiden Gleichungen darstellen:

$$\begin{aligned} X_{ti} &= T_{ti} + E_{ti} \\ T_{ti} &= \beta_{t-1}T_{(t-1)i} + D_{ti} \quad (t = 2, \dots, k) \end{aligned} \quad (2.24)$$

Die erste dieser beiden Gleichungen wurde oben bereits als Gleichung 2.22 spezifiziert – es handelt sich um die Meßmodelle für die T_{ti} , welche somit Teil des QMSM sind. Gleichung 2.24 spezifiziert die Abhängigkeit der True-Scores des Zeitpunkts t von denen des vorangegangenen Zeitpunkts $t - 1$ in einem linearen Regressionsmodell, d.h. D_{ti} bezeichnet die Residuen dieses Regressionsmodells. Zu erwähnen ist, daß die beiden Gleichungen, insbesondere 2.24 keine Regressionskonstante enthalten: Dieses entspricht der normalen Vorgehensweise der Berechnung von Strukturgleichungsmodellen als reine Kovarianzstrukturanalyse ohne Schätzung von Mittelwerten. Es werden in einem solchen Modell also lediglich Varianzen und Kovarianzen durch eine Struktur linearer Abhängigkeiten parametrisiert bzw. die entsprechenden Modellparameter durch Anpassung an die empirisch ermittelte Varianz-Kovarianz-Matrix der beobachteten Variablen geschätzt. Variablenmittelwerte sind in der ursprünglichen Form der Strukturgleichungsmodelle schlicht nicht enthalten und folglich benötigen die Modellgleichungen keine Konstanten, welche ja letztlich nur Mittelwertsverschiebungen zwischen Kriteriums- und Prädiktorvariablen, „jenseits“ des linearen Zusammenhangs beinhalten. Tatsächlich bleiben somit bei einem solchen Modell die Mittelwerte schlicht unberücksichtigt, sie sind für die Zusammenhänge, die in den Pfadkoeffizienten abgebildet werden, ebenso bedeutungslos, wie für die Varianzen der Modellkomponenten. Besteht ein Interesse an den Mittelwerten, so können diese ohne Weiteres in das Modell aufgenommen werden, d.h. es werden dann die Modellgleichungen um entsprechende Regressionskonstanten erweitert und der Modellanpassung wird nicht nur die (empirische) Varianz-Kovarianz-Matrix, sondern zusätzlich der Mittelwertsvektor der beobachteten Variablen zugrundegelegt.

In dem durch Abbildung 2.3.B bzw. die Gleichungen 2.22 und 2.24 dargestellten Modell entsprechen die standardisierten Werte der Autoregressionskoeffizienten β_t den Autokorrelation erster Ordnung zwischen den wahren Werten der jeweils aufeinanderfolgenden Meßzeitpunkte, d.h. den „reliabilitätsbereinigten“ Stabilitätskoeffizienten im jeweiligen Zeitintervall:

$$\rho_t = \rho_{T_t T_{(t+1)}} = \frac{\beta_t \sigma_{T_t}^2}{\sqrt{\sigma_{T_t}^2 \sigma_{T_{(t+1)}}^2}} = \frac{\beta_t \sigma_{T_t}}{\sigma_{T_{(t+1)}}} = \beta_t^s \quad (2.25)$$

Der Ausdruck β_t^s soll hier die standardisierten Pfadkoeffizienten bezeichnen, Gleichung 2.25 wiederholt lediglich bekannte Formeln zur Standardisierung eines Regressions-

koeffizienten in bivariaten Regressionsmodellen (z.B. Schilling, 1998, 216) und wendet sie auf die Autoregressionskoeffizienten des QMSM an.

Es stellt sich nun die Frage nach der Identifizierbarkeit des QMSM. Tatsächlich wären in der durch Abbildung 2.3.B dargestellten Form nicht alle Modellparameter schätzbar. Jöreskog (1970) und Werts et al. (1971) zeigten, daß in dem Modell die „inneren“ Parameter, d.h. alle Reliabilitäten und Stabilitäten außer denen an den Rändern der Zeitreihe zu $t = 1$ und $t = k$, identifiziert sind, die „äußeren“ Parameter müssen dann durch Restriktionen schätzbar gemacht werden. Für die Analyse einer längeren Zeitreihe mit mehreren Meßzeitpunkten (z.B. $k > 3$) kann dann ein solcher Verzicht auf die Ränder durchaus hinnehmbar sein, da immer noch genügend Reliabilitäten und Stabilitäten zu schätzen bleiben. Da das QMSM in den folgenden Analysen noch eine zentrale Rolle bei der Stabilitätsanalyse spielen wird, wird die Identifikation der Modellparameter detaillierter in Anhang A.2 dargestellt.

Wenn unbedingt alle Parameter einschließlich derer an den Rändern der Zeitreihe berechnet werden sollen, so müssen letztere über Restriktionen identifizierbar gemacht werden. Wie im Anhang A.2 gezeigt wird, sind β_1 und σ_{E1}^2 nur zusammen identifiziert (das Produkt $\beta_1 \sigma_{E1}^2$ kann geschätzt werden) und es müßte deshalb einer von beiden Werten durch eine Restriktion berechenbar gemacht werden, damit dann auch der andere identifiziert werden kann. Ähnlich verhält es sich für die beiden Varianzausdrücke am Ende der Kette σ_{Ek}^2 und σ_{Dk}^2 .

Eine solche Restriktion, die das Modell als ganzes identifizierbar macht, wäre die gleicher Fehlervarianzen $\sigma_{Et}^2 = \sigma_E^2$ für alle Messungen, die in dem von Heise (1969) und von Wiley & Wiley (1970) erstmals präsentierten Modell zugrundegelegt wurde: Damit würden also auch die beiden „äußeren“ Meßfehlervarianzen σ_{E1}^2 und σ_{Ek}^2 dadurch identifiziert, daß sie den identifizierten „inneren“ $\sigma_{E2}^2 \dots \sigma_{E(k-1)}^2$ gleichgesetzt werden. Diese Annahme homogener Meßfehlervarianzen ist allerdings nicht zu verwechseln mit der gleicher Reliabilitäten: Letzteres wäre bei homogenen Fehlervarianzen nur dann der Fall, wenn ebenso die True-Score-Varianzen über die Zeitreihe hinweg invariant blieben (vgl. wieder Formel A.3 im Anhang A.1).

Man könnte allerdings diesen Fall als die einzig realistische Bedingung für Homogenität der Reliabilitäten ansehen, welche ansonsten einen sehr spezifischen Zusammenhang der Veränderung von True-Score- und Meßfehlervarianzen voraussetzen würde: „Both true score variance and error variance may change such that the ratio given in Equation (3) [entspricht den Gleichungen A.3, O.S.] remains constant. This appears to be unlikely“ (Wiley & Wiley, 170, 112/3). Die realistischere Annahme homogener Reliabilität aufgrund der Homogenität von Fehler- und True-Score-Varianzen könnte aber mit dem Modell von Wiley & Wiley nicht ohne weiteres überprüft werden, da dieses eben eine zusätzliche Gleichheitsrestriktion für die Varianzen σ_{Tt}^2 erfordern würde, welche, wie im folgenden noch gezeigt wird, zumindest nicht auf einfachem Wege zu erzeugen ist. Man kann, wiederum mit Wiley & Wiley, auch argumentieren, daß überhaupt die Annahme homogener Reliabilitäten auf

sehr weitgehenden theoretischen Implikationen basiert: „Error variance is best conceived as a property of the measuring instrument itself and not of the population to which it is administered. On the other hand, the true score variance is more realistically considered as a property of the population. Thus the specification of stable reliability will normally require assumptions about populations as well as assumptions about the measuring instrument“ (ebd., 112). So besehen, würde eine Gleichheitsrestriktion für die Meßfehlervarianzen eine nur „schwache“ inhaltliche Beschränkung des QMSM darstellen: Es wird dann nur angenommen, daß die bei der Messung wirksamen Eigenschaften des Meßinstruments bei allen Messungen dieselben sind, während bezüglich der auf die Messung einwirkenden Eigenschaften der Population (der Merkmalsträger) solches nicht von vorneherein angenommen wird.

Von Restriktionen, die zur Schätzung der „Randparameter“ notwendig wären, abgesehen, stellt das QMSM ein weitgehend unrestruktives Modell der Merkmalsveränderung über die „inneren“ Meßzeitpunkte $t = 2, \dots, k - 1$ dar, welches keinerlei implizite Beschränkungen bezüglich des Typus der Merkmalsstabilität, der Reliabilitäten, Stabilitäten und aller „direkten“ Modellparameter, die den zeitlichen Veränderungsprozeß beschreiben, enthält. Es eignet sich somit bestens für eine Analyse von Merkmalsstabilität, die durch keine theoretische Festlegung auf ein spezifisches Veränderungsmodell „voreingenommen“ ist. Die weitestgehende Uneingeschränktheit der durch das QMSM modellierbaren Merkmalsverläufe ist m.E. der eigentliche Vorteil, den dieses Modell gegenüber den im folgenden noch zu besprechenden bietet.

Ein Nachteil des QMSM besteht allerdings in gewissen „praktischen“ Schwierigkeiten, die dabei Restriktionen für True-Score-Varianzen, Reliabilitäten und Stabilitätskoeffizienten bereiten. Denn alle diese Kennwerte sind insofern „indirekte“ Modellparameter, als es sich nicht um Elemente des zu schätzenden Parametervektors des Modells handelt bzw. sie nicht in den Ausdrücken der implizierten Kovarianzmatrix des Modells enthalten sind (vgl. wieder Anhang A.2). Die „direkten“ Modellparameter des QMSM sind:

- ▶ Die Varianz der True-Scores bei der ersten Messung σ_{T1}^2 .
- ▶ Die *unstandardisierten* Autoregressionskoeffizienten β_t .
- ▶ Die Meßfehlervarianzen σ_{Et}^2 .
- ▶ Die „Random-Varianzanteile“ der True-Scores ab der zweiten Messung σ_{Dt}^2 .

Die True-Score-Varianzen σ_{T2}^2 bis σ_{Tk}^2 , sowie die Reliabilitäten und die Stabilitätskoeffizienten ρ_t sind aus den „direkten“ Parametern zu berechnen. Für ρ_t wurde diese Berechnung bereits mit Gleichung 2.25 gezeigt – in dieser sind rechtsseitig neben dem „direkten“ Parametern β_t noch die Standardabweichungen oder Varianzen der True-Scores enthalten, welche mit Ausnahme derjenigen des ersten Meßzeitpunkts wiederum nur indirekt geschätzt werden. Für die True-Score-Varianz eines Zeitpunkts $t > 1$ gilt:

$$\sigma_{Tt}^2 = \beta_{t-1}^2 \sigma_{T(t-1)}^2 + \sigma_{Dt}^2 \quad (2.26a)$$

$$\begin{aligned} &= \beta_{t-1}^2 (\beta_{t-2}^2 \sigma_{T(t-2)}^2 + \sigma_{D(t-1)}^2) + \sigma_{Dt}^2 \\ &= (\beta_{t-1}^2 \beta_{t-2}^2 \cdots \beta_1^2) \sigma_{T1}^2 + (\beta_{t-1}^2 \beta_{t-2}^2 \cdots \beta_2^2) \sigma_{D2}^2 + (\beta_{t-1}^2 \beta_{t-2}^2 \cdots \beta_3^2) \sigma_{D3}^2 + \cdots + \sigma_{Dt}^2 \\ &= \left(\prod_{i=1}^{t-1} \beta_i^2 \right) \sigma_{T1}^2 + \sum_{j=2}^{t-1} \left(\prod_{k=j}^{t-1} \beta_k^2 \right) \sigma_{Dj}^2 + \sigma_{Dt}^2 \end{aligned} \quad (2.26b)$$

$$\text{(für: } \sum_2^{t-1} = 0 \text{ wenn } t-1 < 2)$$

Es wird deutlich, daß mit wachsendem Meßzeitpunkt t der rechtsseitige Ausdruck in Gleichung 2.26b immer länger und komplexer wird und daß σ_{Tt}^2 in keiner linearen Beziehung zu den „direkten“ Modellparametern steht, durch die es berechnet wird. Damit aber würde die Festlegung von Restriktionen für diese True-Score-Varianzen zumindest unhandlich.

Solche Restriktionen werden aber dann notwendig, wenn inhaltliche Hypothesen über die zeitliche Invarianz von Meßfehler- und/oder True-Score-Varianzen überprüft werden sollen. Soll beispielsweise mittels des QMSM getestet werden, ob für die wahren Merkmalswerte über die Zeitreihe hinweg Varianzhomogenität besteht, oder ob sich die True-Score-Varianzen im Zeitverlauf verändern – wie es z.B. aus der Gewinner-und-Verlierer-Hypothese zu folgern wäre (vgl. Kapitel 1.2.3.2) –, so müßte ein (hinsichtlich der True-Score-Varianzen) unrestringiertes QMSM mit einem solchen verglichen werden, in dem vorab die Restriktion $\sigma_{T1}^2 = \sigma_{T2}^2 = \cdots = \sigma_{Tk}^2$ festgelegt wird, und es wären hierbei anstelle der $\sigma_{T2}^2, \dots, \sigma_{Tk}^2$ die Ausdrücke einzusetzen, die sich aus Gleichung 2.26b ergeben. Nun würde die Spezifikation einer solchen Gleichheitsrestriktion nicht nur mit wachsender Zahl von Meßzeitpunkten immer komplexer, sondern bei den meisten der gegenwärtig verfügbaren Computerprogramme zur Berechnung von Strukturgleichungsmodellen schlicht unmöglich, da diese keine nichtlinearen Restriktion erlauben.¹⁰

Dasselbe Problem besteht für Restriktionen der Stabilitätskoeffizienten und Reliabilitäten, in deren Berechnung jeweils ebenfalls die True-Score-Varianzen eingehen (vgl. Gleichung 2.25 und A.3 im Anhang A.1). Wenn also die Invarianz der Stabilität selbst im zeitlichen Verlauf (d.h., in der Terminologie der Zeitreihenanalyse, die Stationarität des Markov-Prozesses) überprüft werden soll, dann müßte die Gleichheit der standardisierten Autokorrelationskoeffizienten $\rho_t = \rho$ festgelegt werden. Beispielsweise erhielte man bei einer Zeitreihe von $k = 4$ Messungen aus den Gleichungen 2.25 und 2.26b die folgende Gleichheitsrestriktion für die drei Stabilitätskoeffizienten:

¹⁰Zum Zeitpunkt der Berechnung der vorgelegten Analysen war die Festlegung nichtlinearer Restriktionen in der Prozedur CALIS, die Bestandteil des SAS-Statistiksoftwarepakets ist, sowie in der 2001 neu erschienenen Programmversion LISREL 8.51 möglich.

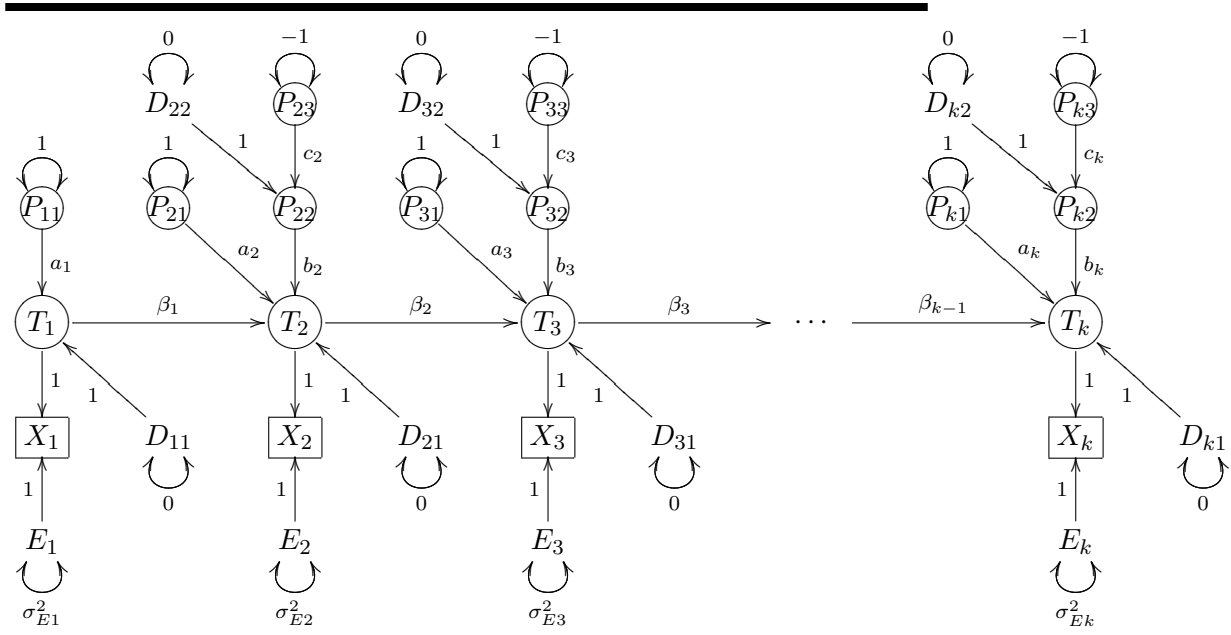
$$\begin{aligned} \frac{\beta_1 \sigma_{T1}}{\sqrt{\beta_1^2 \sigma_{T1}^2 + \sigma_{D2}^2}} &= \frac{\beta_2 \sqrt{\beta_1^2 \sigma_{T1}^2 + \sigma_{D2}^2}}{\sqrt{\beta_2^2 \beta_1^2 \sigma_{T1}^2 + \beta_2^2 \sigma_{D2}^2 + \sigma_{D3}^2}} \\ &= \frac{\beta_3 \sqrt{\beta_2^2 \beta_1^2 \sigma_{T1}^2 + \beta_2^2 \sigma_{D2}^2 + \sigma_{D3}^2}}{\sqrt{\beta_3^2 \beta_2^2 \beta_1^2 \sigma_{T1}^2 + \beta_3^2 \beta_2^2 \sigma_{D2}^2 + \beta_3^2 \sigma_{D3}^2 + \sigma_{D4}^2}} \end{aligned}$$

Im QMSM ist also die Gleichheit von Stabilitätskoeffizienten durch nichtlineare Restriktionen zu erreichen, die mit wachsender Anzahl von Meßzeitpunkten immer komplexer werden – die obigen Ausdrücke für die True-Score-Korrelationen bei $k = 4$ geben eine Vorstellung vom Umfang, den die entsprechenden formalen Ausdrücke für die hier angestrebten Stabilitätsanalysen für Zufriedenheitsratings aus 16 SOEP-Wellen annehmen würden.

Um der Schwierigkeit der Festlegung solcher Gleichheitsrestriktionen für Reliabilitäten, True-Score-Varianzen und Stabilitätskoeffizienten abzuwehren – bzw. sie für die meisten gebräuchlichen Softwarepakete überhaupt zu ermöglichen – präsentierten Rudinger & Rietz (1993; 1998; 2001) ein QMSM, das eine Gleichheitsrestriktion der True-Score-Varianzen *ohne* Setzung nichtlinearer Restriktionen für die „direkten“ Modellparameter erlaubt. Unter der Voraussetzung der Invarianz der True-Score-Varianzen sind dann auch die Invarianzen der Stabilitätskoeffizienten (durch einfache Gleichheitsrestriktion der Autoregressionskoeffizienten $\beta_t = \beta$, vgl. Gleichung 2.25) und der Reliabilitäten (durch Gleichheitsrestriktion der Meßfehlervarianzen $\sigma_{E_t}^2 = \sigma_E^2$, vgl. Gleichung A.3 im Anhang A.1) einfach festzulegen, so daß mit dem Modell eine sehr weitgehende Überprüfung von Homogenitätsannahmen möglich wird.

Der „Trick“, der dabei angewendet wird, ist eine Technik zur Setzung von Restriktionen für die Varianzen der Random-Komponenten in Strukturgleichungsmodellen, die von Rindskopf (1983; 1984) vorgestellt und im zweiten der genannten Artikel mit dem Begriff der *Phantomvariablen* bezeichnet wurde: Solche Phantomvariablen sind schlicht zusätzliche latente Variablen, die gewissermaßen als Prädiktoren derjenigen Variablen, deren Varianzen dadurch restringiert werden sollen, eingesetzt werden, so daß im Endeffekt diese Varianzen durch einfache Restriktionen für die Phantomvariablen kontrolliert werden können. Auf diese Weise können dann auch die Varianzen der abhängigen True-Scores im QMSM restringiert werden: „Die Grundidee, mit der diese Modellierbarkeit gewährleistet werden kann, besteht darin, zu einem Meßzeitpunkt $t + 1$ die Varianz der latenten Variablen des Meßzeitpunkts t abzuziehen und der latenten Variablen zum Meßzeitpunkt $t + 1$ eine ‚neue‘ – kontrollierte – Varianz ... zu geben“ (Rudinger & Rietz, 1993, 67). Abbildung 2.4 stellt das resultierende Phantomvariablenmodell dar, wobei hier um der besseren Anschaulichkeit (und der Übereinstimmung mit der bisherigen Darstellung) willen die Bezeichnungen der einzelnen Bestandteile von der Rudinger-Rietz-Vorgabe (und wie auch bisher schon von der häufig gewählten LISREL-Notation) abweichen: Die neu hinzugekommenen Phantomvariablen werden mit $P_{...}$ bezeichnet, die neuen Pfade des Modells

ABBILDUNG 2.4:
QMSM mit Phantomvariablen zur Restriktion der True-Score-Varianzen



mit $a_{...}$, $b_{...}$, $c_{...}$. Auch sind hier die unabhängigen Varianzen des Modells (in Form kleiner runder Doppelpfeile) abgebildet, so daß die wichtigen Restriktionen bezüglich dieser Varianzen sichtbar gemacht werden können. Sozusagen der Ordnung und Übereinstimmung mit Rudinger & Rietz (1993; 1998; 2001) halber sind auch Random-Komponenten für die abhängigen latenten Konstrukte eingezeichnet (D_{t1} für die True-Scores T_t und D_{t2} für die Phantomvariablen P_{t2}) – deren Varianzen sind aber jeweils auf Null restringiert, die Rolle dieser Zufallskomponenten übernehmen in diesem Modell gewissermaßen die Phantomvariablen, durch die hier jedem abhängigen T_t Varianzkomponenten zugewiesen werden, welche unabhängig von der jeweiligen Varianz des vorherigen True-Scores $\sigma_{T(t-1)}^2$ sind.

Die Varianzen der Phantomvariablen sind durch Gleichheitsrestriktionen festgelegt und dabei kann nun auffallen, daß die Varianzen $\sigma_{P_{t3}}^2$ jeweils auf den Wert -1 fixiert sind – also auf einen Negativwert, den Varianzen per definitionem gar nicht annehmen können. Diese scheinbar sinnlose Festlegung ist hier möglich, da zum einen auf „inhaltlicher Ebene“ die zugehörigen Phantomvariablen keinerlei Bedeutung haben, es handelt sich sozusagen lediglich um einen technischen Bestandteil in den Strukturgleichungen des Modells, der zum Zwecke der Subtraktion von Varianzanteilen eingefügt wurde (siehe im folgenden), und zum anderen in „numerischer Hinsicht“ diese Negativwerte keine prinzipielle Unlösbarkeit der Modellschätzung implizieren.¹¹ Die Warnungen, welche gebräuchliche

¹¹Die Anpassung der modellimplizierten an die empirisch geschätzte Kovarianzmatrix der beobachteten Variablen ist von negativen „Phantomvarianzen“ nicht grundsätzlich betroffen. Letztlich gelten auch in diesem Modell die Formeln A.8 und A.9 (Anhang A.2, wenn dort $\sigma_{T_1}^2$ durch $a_1^2 \sigma_{P_{11}}^2 = a_1^2$ und die Ausdrücke $\sigma_{D_t}^2$

Softwareprogramme bei negativen Varianzwerten produzieren, können in diesem Fall also ausnahmsweise ignoriert werden.

Für das Modell sind folgende Restriktionen festzusetzen (vgl. Rudinger & Rietz, 1998):

$$\sigma_{D11}^2 = \dots = \sigma_{D16\cdot1}^2 = \sigma_{D22}^2 = \dots = \sigma_{D16\cdot2}^2 = 0 \quad \text{bzw.} \quad \sigma_{Dtj}^2 = 0 \quad (2.27a)$$

$$\sigma_{P11}^2 = \dots = \sigma_{P16\cdot1}^2 = 1 \quad \text{bzw.} \quad \sigma_{Pt1}^2 = 1 \quad (2.27b)$$

$$\sigma_{P23}^2 = \dots = \sigma_{P16\cdot3}^2 = -1 \quad \text{bzw.} \quad \sigma_{Pt3}^2 = -1 \quad (2.27c)$$

$$\beta_1 = b_2, \beta_2 = b_3, \dots, \beta_{15} = b_{16} \quad \text{bzw.} \quad \beta_t = b_{t+1} \quad (2.27d)$$

$$a_1 = c_2, a_2 = c_3, \dots, a_{15} = c_{16} \quad \text{bzw.} \quad a_t = c_{t+1} \quad (2.27e)$$

$$a_1 = a_2 = \dots = a_{16} \quad \text{bzw.} \quad a_t = a \quad (2.27f)$$

Zur Veranschaulichung der Auswirkung dieser Restriktionen, welche möglicherweise nicht auf den ersten Blick erkennbar sind, wird diese für σ_{T1}^2 und σ_{T2}^2 gezeigt. Aufgrund der in Abbildung 2.4 dargestellten Abhängigkeiten gilt:

$$\begin{aligned} \sigma_{T1}^2 &= a_1^2 \sigma_{P11}^2 \\ &\Downarrow \quad (\text{vgl. Restriktion 2.27b}) \\ &= a_1^2 \\ \sigma_{T2}^2 &= \beta_1^2 \sigma_{T1}^2 + a_2^2 \sigma_{P21}^2 + b_2^2 (\sigma_{D22}^2 + c_2^2 \sigma_{P23}^2) + \sigma_{D21}^2 \\ &\Downarrow \quad (\text{vgl. Restriktionen 2.27a, 2.27b, 2.27c}) \\ &= \beta_1^2 a_1^2 + a_2^2 - b_2^2 c_2^2 \\ &\Downarrow \quad (\text{vgl. Restriktionen 2.27e, 2.27f}) \\ &= \beta_1^2 a_1^2 + a_1^2 - \beta_1^2 a_1^2 \\ &= a_1^2 \end{aligned}$$

Es ist leicht zu erkennen, daß nach dem für σ_{T2}^2 gezeigten Schema auch für die weiteren True-Score-Varianzen $\sigma_{T3}^2, \dots, \sigma_{Tk}^2$ aufgelöst werden kann und jeweils a_1^2 als Wert resultiert. Somit wäre also durch die Restriktionen 2.27a bis 2.27f die Homogenität der True-Score-Varianzen über die Zeitreihe hinweg festgelegt und durch die zusätzliche Restriktion $\beta_1 = \beta_2 = \dots = \beta_{k-1}$ könnte dann noch zusätzlich die Invarianz der Stabilitätskoeffizienten bzw. durch $\sigma_{E1}^2 = \sigma_{E2}^2 = \dots = \sigma_{Ek}^2$ die der Reliabilitäten spezifiziert werden.

Es bleibt die Schwierigkeit, z.B. invariante Stabilitätskoeffizienten bei frei variierenden Varianzen festzulegen. Dennoch kann festgehalten werden, daß durch die von Rudinger

jeweils durch $a_t^2 \sigma_{Pt1}^2 + b_t^2 c_t^2 \sigma_{Pt3}^2 = a_t^2 - b_t^2 c_t^2$ ersetzt werden. Problematisch würde das Ganze nur, wenn dieser letzte Ausdruck negativ würde, was aber durch die zusätzlichen Restriktionen für die Pfadkoeffizienten a_t, b_t und c_t ausgeschlossen wird.

& Rietz gezeigte Vorgehensweise die „Unhandlichkeit“ von Restriktionen über True-Score-Varianzen, Stabilitätskoeffizienten und Reliabilitäten größtenteils vermieden werden kann, so daß dadurch die Bandbreite der durch das QMSM relativ unkompliziert modellier- und testbaren Charakteristika von Veränderung und Stabilität eines Merkmals, enorm erweitert wird. Dieses, zusammen mit der oben bereits konstatierten prinzipiellen Uneingeschränktheit dieses Veränderungsmodells, macht das QMSM zu einer guten Wahl, wenn es darum geht, die tatsächliche meßfehlerbereinigte Stabilität, die in einer Zeitreihe von Meßwerten eines Merkmals zum Ausdruck kommt, ohne vorherige theoretische Festlegung auf spezifische Verlaufsmodelle zu analysieren.

Es wurde bereits vorne darauf hingewiesen, daß dieses QMSM solchermaßen erweitert werden kann, daß es auch eine Modellierung der Entwicklung der Mittelwerte über die Zeitreihe enthält. In der bisher dargelegten Form handelt es sich um ein Modell zur Analyse der Varianz-Kovarianz-Matrix der beobachteten Variablen X_{ti} bzw. um eine reine Kovarianzstrukturanalyse, in der die Varianzen der und Abhängigkeiten zwischen den Modellvariablen parametrisiert und alle Mittelwert auf Null gesetzt sind. Da Varianzen und Kovarianzen (bzw. Pfadkoeffizienten und Korrelationen) von konstanten Skalentransformationen (durch Addition der Variablenwerte mit einem konstanten Wert) unberührt bleiben, sind für die Analyse der Kovarianzstruktur die Mittelwerte ohne Belang, so daß sie mit „mean deviation scores“, bei denen die Skala auf den Mittelwert Null „verschoben“ wurde, durchgeführt werden kann. Hinsichtlich der in Kapitel 2.2.1 dargestellten Stabilitätstypen aber sind Mittelwertsveränderungen im zeitlichen Verlauf ein zusätzlich interessanter Aspekt der Analyse: Während die Typen monotoner und paralleler Stabilität ausschließlich auf Eigenschaften der Kovarianzstruktur beruhen, bezieht sich der Typus strikter Stabilität auf die Invarianz der Mittelwerte zusätzlich zu den Merkmalen paralleler Stabilität. Für eine vollständige Stabilitätsanalyse mit dem QMSM wäre es darum wünschenswert, das Modell um eine „Mittelwertsstruktur“ zu erweitern, so daß es eine simultane Analyse von Stabilitätskoeffizienten, True-Score-Varianzen und Mittelwertsveränderungen enthält.

Bei der reinen Kovarianzstrukturanalyse wird durch die Spezifikation eines Strukturgleichungsmodells letztendlich eine theoretische Struktur der Varianzen und Kovarianzen aller Variablen im Modell postuliert und den empirischen Stichprobendaten angepaßt: Das wesentliche Rationale der Vorgehensweise liegt darin, daß Varianzen und Kovarianzen der beobachteten Variablen aus den im spezifizierten Modell postulierten Beziehungen theoretisch erklärt werden, so daß sie als Funktionen der zu schätzenden Modellparameter (Pfadkoeffizienten usw.) ausgedrückt werden können. Die Parameterschätzung besteht dann in der Auswahl optimaler Parameterwerte dergestalt, daß die durch diese Schätzwerte implizierte Varianz-Kovarianz-Matrix der beobachteten Variablen möglichst gering von der aus den empirischen Daten geschätzten abweicht (zur Veranschaulichung siehe Anhang A.2, insbesondere z.B. die Gleichungen A.8a bis A.9f). Eine Mittelwertsstruktur kann dem nun dadurch hinzugefügt werden, daß die Modellanpassung auch den Mittelwertsvektor mit einbezieht, daß also zusätzlich die modellimplizierten den empirisch geschätzten Mittelwerten

der beobachteten Variablen angepaßt werden. Dieses wird technisch dadurch erreicht, daß das Modell um eine Konstante (normalerweise mit dem Wert 1) erweitert und anstelle der Varianz-Kovarianz- die Kreuzproduktmatrix der beobachteten Variablen *und* dieser Konstanten analysiert wird (sogenannte „augmented moment matrix“ – McArdle & Aber, 1990, 160f, liefern eine knappe und erhellende Beschreibung der verschiedenen Typen von Matrizen, die bei der SEM-Analyse verwendet werden können). Mit Blick auf die Modellgleichungen bedeutet dieses die Aufnahme von Intercepts, also von Regressionskonstanten, wie sie aus der herkömmlichen Regressionsanalyse bekannt sind.

Sollen nun im QMSM auch die Veränderungen der Mittelwerte über die Zeit modelliert werden, so würde das zunächst logischerweise bedeuten, daß die Gleichung 2.24 um ein (hier durch ein α bezeichnetes) Intercept zu erweitern wäre:

$$T_{ti} = \alpha_t + \beta_{t-1}T_{(t-1)i} + D_{ti} \quad (t = 1 \dots k; \beta_0 = 0) \quad (2.28)$$

Daraus resultieren dann die True-Score-Mittelwerte wie folgt:

$$\begin{aligned} \mu_1 &= \alpha_1 \\ \mu_2 &= \alpha_2 + \beta_1\mu_1 \\ \mu_3 &= \alpha_3 + \beta_2\mu_2 = \alpha_3 + \beta_2\alpha_2 + \beta_2\beta_1\alpha_1 \\ &\vdots \\ \mu_t &= \alpha_t + \beta_{t-1}\mu_{t-1} = \alpha_t + \sum_{j=1}^{t-1} \left(\prod_{k=1}^{t-1} \beta_k \alpha_j \right) \end{aligned} \quad (2.29)$$

Bei Zugrundelegung des mit Gleichung 2.22 beschriebenen einfachen Modells der klassischen Testtheorie sind die Mittelwerte von True-Scores und Meßwerten identisch, da die Koeffizienten des Meßmodells auf Eins fixiert und die Meßfehler als Zufallskomponente mit dem Mittelwert Null konzipiert sind (vgl. Gleichung A.2 in Anhang A.1). Für die Gleichungen des Meßmodells ist darum kein zusätzliches Intercept nötig – ein solches würde lediglich eine „Skalenverschiebung“ der True-Score-Mittelwerte gegenüber dem jeweiligen Mittelwert der Meßskala bezeichnen.

Wie nun die obigen Ableitungen der aus 2.28 resultierenden True-Score-Mittelwerte zeigen (Gleichung 2.29), würden hier wiederum Restriktionen z.B. der Gleichheit von Mittelwerten verschiedener Meßzeitpunkte ähnlich komplex, wie entsprechende Gleichheitsrestriktionen für Varianzen und Stabilitätskoeffizienten. Da es sich aber bei der Kovarianz- und der Mittelwertsstruktur im Grunde um zwei voneinander unabhängige Eigenschaften des Modells handelt und eine Fixierung der obigen Intercepts α_t auf welchen Wert auch immer keine Auswirkung auf die Schätzung der Parameter der Kovarianzstruktur (also auch der Stabilitätskoeffizienten, True-Score-Varianzen und Reliabilitäten) hat, können Restriktionen für die Mittelwerte auch auf einem einfacher zu handhabendem Wege spezifiziert

werden – nämlich dadurch, daß die True-Score-Mittelwerte auf Null restringiert bleiben (d.h. $\alpha_t = 0$) bzw. für das autoregressive Strukturmodell der latenten Variablen T_{ti} die Gleichung 2.24 *ohne* Intercepts beibehalten wird und stattdessen die Meßgleichungen 2.22 mit Regressionskonstanten erweitert werden:

$$X_{ti} = \mu_t + T_{ti} + E_{ti} \quad (2.30)$$

Die Intercepts wurden hier der Einfachheit halber direkt mit μ_t bezeichnet, weil nach Gleichung 2.30 dann, wenn die Mittelwerte von T_{ti} und E_{ti} jeweils gleich Null sind, der jeweilige Mittelwert der Meßwerte X_{ti} gleich dieser Konstante ist.

Eine solche Modellierung mag auf den ersten Blick vielleicht kontraintuitiv erscheinen, da es doch bei der Frage nach dem Ausmaß strikter Merkmalsstabilität darum geht, Hypothesen über die Veränderung von True-Score-Mittelwerten – und nicht von Meßwertmittelwerten – zu überprüfen. Dazu ist jedoch zu bedenken, daß nach der intuitiv plausibleren und „korrekteren“ Modellierung mit den Gleichungen 2.22 und 2.28 die Gleichheit von True-Score- und Meßwertmittelwerten festgelegt wäre: Es wären somit auch bei dieser Modellierungsvariante nur k Mittelwerte ($\mu_1, \mu_2, \dots, \mu_k$) zu schätzen. Also wäre bei der Modellierung von True-Score-Mittelwerten nach 2.28 bzw. 2.29 jede Restriktion der True-Score-Mittelwerte gleichermaßen auch für die jeweiligen Meßwertmittelwerte gültig und würde die Schätzung der Parameter der theoretisch postulierten Mittelwertsstruktur (d.h. dann der Intercepts α_t) durch Anpassung an den empirischen Mittelwertsvektor der beobachteten Werte erfolgen.

Letzteres aber gilt genauso für die Modellierung durch die Gleichungen 2.30 und 2.24: Im Modell sind wiederum k Mittelwerte zu schätzen, dazu wird die theoretisch postulierte Struktur dieser k Mittelwerte dem empirischen Mittelwertsvektor der beobachteten Werte angepaßt. Würde also beispielsweise in der ersteren Modellierungsvariante (2.22 und 2.28) die Gleichheit der k True-Score-Mittelwerte vorab festgelegt (durch Gleichsetzung der aus 2.29 resultierenden komplexen Ausdrücke), so würde dieses auch die Gleichsetzung der k Meßwertmittelwerte bedeuten und würde unter dieser Restriktion dann eine optimale Modellanpassung möglich, wenn auch die empirisch vorgefundenen k Stichprobenmittelwerte der Variablen X_{ti} identisch wären. In der zweiten Modellierungsvariante (2.30 und 2.24) könnte dementsprechend direkt die Gleichheit der Meßwertmittelwerte (d.h. der Intercepts μ_t in den Modellgleichungen 2.30) festgelegt werden. In beiden Fällen würde äquivalent die Annahme überprüft, daß die Meßwertmittelwerte invariant sind. Praktisch bedeutet die Aufnahme der Intercepts in die Meßgleichungen 2.30 lediglich, daß in dem Modell eine Verschiebung der True-Score-Skala um den jeweiligen Mittelwert vorgenommen wird, so daß bei dieser Modellierung die latenten Variablen des Strukturmodells in „klassischer“ Weise als Mittelwertsabweichungswerte, die beobachteten Meßwerte dagegen als untransformierte „Originalwerte“ genommen werden. Somit enthält diese Modellierung zwar eine implizite Skalentransformation der True-Scores, durch die deren Mittelwerte gewissermaßen eliminiert werden, dennoch bleibt die Mittelwertsstruktur der *untransformierten* True-

Scores aufgrund ihrer Identität mit derjenigen der Meßwerte modellierbar – nämlich eben durch die Modellierung der Meßwertmittelwerte.

Da aber in dieser Modellierung mittels der Gleichungen 2.30 und 2.24 die Mittelwerte μ_t direkt zu schätzende Modellparameter sind und deshalb Restriktionen für diese Mittelwerte einfacher spezifiziert werden können, ist sie der Modellierung durch die Gleichungen 2.22 und 2.28 vorzuziehen. Da die Schätzung der Kovarianzstrukturparameter von der Aufnahme der Intercepts μ_t in die Meßgleichungen unberührt bleibt, behalten die zuvor dargestellten Modelleigenschaften der „mittelwertslosen“ Modellierung unverändert Gültigkeit, wenn Gleichung 2.22 durch 2.30 ersetzt wird.

2.2.2.2 Das Modell latenter Differenzkomponenten

In einem „difference components change model“ (McArdle & Aber, 1990) werden gewissermaßen die wahren Merkmalsausprägungen, die den Messungen der Zeitreihe zugrundeliegen, als Summe sukzessiver Differenzen der True-Scores dekomponiert, indem jede dieser Differenzkomponenten als latentes Konstrukt spezifiziert wird. Dieses wird wohl am ehesten am Beispiel des in Abbildung 2.5 dargestellten Modells deutlich. Formal beschreibt die folgende Gleichung dieses Modell, das in Anlehnung an die Charakterisierung von McArdle & Aber als Latent-Difference-Components-Model (LDCM) bzw. als Modell latenter Differenzkomponenten bezeichnet werden kann:

$$X_{ti} = \sum_{j=0}^{t-1} \Delta_{(t-j)i} + E_{ti} \quad (i = 1 \dots n, t = 1 \dots k) \quad (2.31)$$

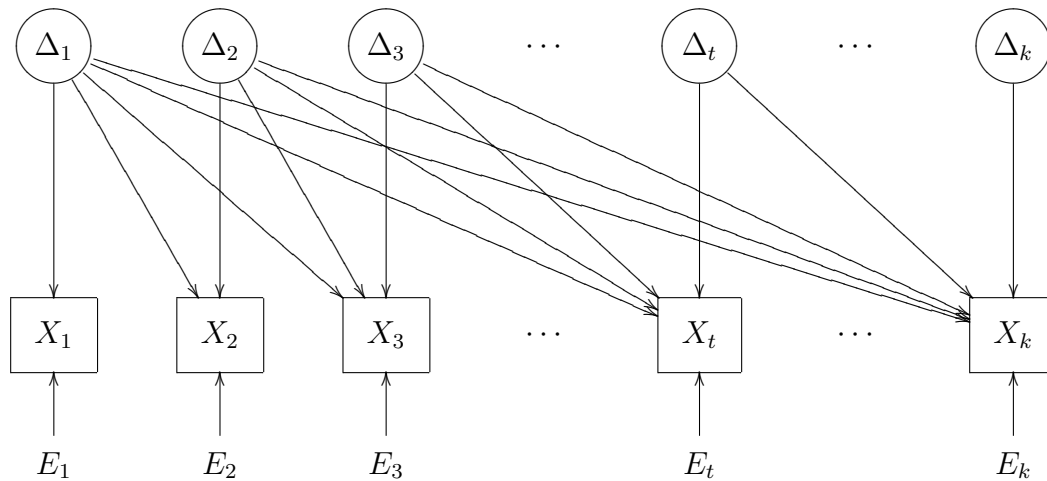
Die Bedeutung der Komponenten Δ_t wird dann deutlich, wenn man 2.31 mit Gleichung 2.22 kontrastiert, welche ja das allgemeine testtheoretische Meßmodell für die Meßwerte X_{ti} spezifiziert (Aufteilung in True-Score und Meßfehler). Für die ersten drei Meßzeitpunkte ($t = 1, 2, 3$) ergeben sich so z.B. die folgenden Auflösungen:

$$\begin{aligned} X_{1i} &= \Delta_{1i} + E_{1i} \\ &= T_{1i} + E_{1i} && \implies \Delta_{1i} = T_{1i} \\ X_{2i} &= \Delta_{1i} + \Delta_{2i} + E_{2i} \\ &= T_{2i} + E_{2i} && \implies \Delta_{1i} + \Delta_{2i} = T_{2i} && \implies \Delta_{2i} = T_{2i} - T_{1i} \\ X_{3i} &= \Delta_{1i} + \Delta_{2i} + \Delta_{3i} + E_{3i} \\ &= T_{3i} + E_{3i} && \implies \Delta_{1i} + \Delta_{2i} + \Delta_{3i} = T_{3i} && \implies \Delta_{3i} = T_{3i} - T_{2i} \end{aligned}$$

Diese Auflösungen können wie folgt verallgemeinert werden:

$$\sum_{j=0}^{t-1} \Delta_{(t-j)i} = T_{ti} \implies \Delta_{ti} = T_{ti} - \sum_{j=1}^{t-1} \Delta_{(t-j)i} = T_{ti} - T_{(t-1)i} \quad (2.32)$$

ABBILDUNG 2.5:
Latent-Difference-Components-Model (LDCM)



Anmerkung: Alle Pfade = 1 !

Die latente Differenzkomponente Δ_{1i} entspricht also dem True-Score der ersten Messung, die latente Komponente Δ_{2i} der Differenz des True-Scores bei der zweiten zu dem der ersten Messung, d.h. der Veränderung der True-Scores vom ersten zum zweiten Meßzeitpunkt. Δ_{3i} ist dann die True-Score-Veränderung vom zweiten zum dritten Meßzeitpunkt usw. – bei diesem Modell werden somit die wahren Merkmalsausprägungen zu einem Meßzeitpunkt $t > 1$ als Summe des „Basiswerts“ der wahren Merkmalsausprägung bei der ersten Messung und der Veränderungen, die bei den darauffolgenden Meßzeitpunkten stattgefunden haben, modelliert.

Also ist Δ_{1i} der True-Score des Merkmals zum ersten Meßzeitpunkt und sind die Δ_{ti} für $t \geq 2$ die „wahren“ Merkmalsveränderungen: Das LDCM beschreitet gewissermaßen den zum QMSM alternativen Weg der Stabilitätsanalyse – während letzteres die Modellierung und Untersuchung der sukzessiven True-Score-Verteilungen zum Inhalt hat, sind im LDCM die Verteilungen der Veränderungsscores Gegenstand der Analyse. Folglich gelten hier prinzipiell die im Kapitel 2.2.1 spezifizierten formalen Zusammenhänge für die Δ -Werte (abgesehen davon, daß dort noch nicht zwischen True-Score und Meßwert unterschieden wurde) – wobei allerdings, wie im folgenden gezeigt wird, das LDCM wesentlich durch implizite Restriktionen charakterisiert ist, die in den dortigen Ausdrücken nicht vorausgesetzt wurden.

Es handelt sich auch beim LDCM um ein Quasi-Simplex-Modell, denn ebenso wie das QMSM impliziert es eine durch die Meßfehlerkomponente „aufgeweichte“ Simplex-Struktur der Autokorrelationen.¹² Jöreskog (1970) bezeichnet es als Quasi-Wiener-Simplex-

¹²Jöreskog (1970) zeigt, daß die meßfehlerfreien Versionen der beiden Modelle sogar äquivalente Korrela-

Modell, da es sich um die Modellierung eines stochastischen Prozesses handelt, der als Wiener-Prozeß bezeichnet wird – diese Bezeichnung wird jedoch hier nicht übernommen, da nicht alle Differenzkomponentenmodelle die Eigenschaften eines Wiener-Prozesses erfüllen.¹³

In der gezeigten Form ist das LDCM nicht vollständig identifiziert: Wie Jöreskog (1970) zeigt, sind die jeweils letzten Varianzen der Zeitreihe, d.h. bei k Meßzeitpunkten die Varianz der letzten Differenzkomponente Δ_{ki} und des letzten Meßfehlers E_{ki} , nicht einzeln identifiziert und ist lediglich deren Summe schätzbar – auch dieses wird im Anhang A.3 detaillierter dargestellt. Um eine vollständige Schätzbarkeit des Modells zu erreichen, muß mindestens einer der beiden unidentifizierten Parameter $\sigma_{\Delta k}^2$ oder $\sigma_{E k}^2$ durch eine Restriktion identifizierbar gemacht werden. Normalerweise wird dieses durch die Annahme gleichbleibender Meßfehlervarianzen erreicht (Jöreskog, 1970; McArdle & Aber, 1990):

$$\sigma_{E t}^2 = \sigma_E^2 \quad \text{für alle } t = 1, \dots, k \quad (2.33)$$

Wie auch schon beim QMSM bedeutet dieses *nicht* die Gleichsetzung der Reliabilitäten für alle Meßzeitpunkte: Aus dem durch die Gleichung 2.22 spezifizierten Meßmodell ergibt sich die Varianz der Meßwerte zum Zeitpunkt t aus der Summe von True-Score- und Meßfehlervarianz und folglich wächst bzw. schrumpft bei gleichbleibender Meßfehlervarianz die Reliabilität dann, wenn die True-Score-Varianz wächst bzw. schrumpft (vgl. auch Gleichung A.3, Anhang A.1).

Damit aber wird in dem durch die Restriktion gleicher Meßfehlervarianzen vollständig identifizierten LDCM die Invarianz der Reliabilität über die Meßzeitpunkte an sehr spezielle Bedingungen geknüpft, denn es sind auch gleiche True-Score-Varianzen zu verschiedenen Meßzeitpunkten nur dann möglich, wenn die Veränderungen der True-Scores nicht variieren. Dieses ist die Konsequenz grundlegender Restriktionen des LDCM, nämlich der wechselseitigen Unkorreliertheit der Differenzkomponenten.¹⁴ Die True-Score-Varianzen $\sigma_{T t}^2$ setzen sich dann nämlich additiv aus den Varianzen der Differenzkomponenten zusammen:

$$\sigma_{T t}^2 = \sum_{j=1}^t \sigma_{\Delta j}^2 = \sigma_{T(t-1)}^2 + \sigma_{\Delta t}^2 \quad (2.34)$$

Und somit gilt:

tionsmatrizen implizieren.

¹³Beispielsweise gilt für das „moving average model“ von McArdle & Aber (1990; siehe auch Rovine & von Eye, 1991, 56f) Jöreskogs Formel (34) nicht.

¹⁴Durch diese Unkorreliertheit der Differenzkomponenten handelt es sich, wie bereits erwähnt, bei dem modellierten stochastischen Prozeß um einen sogenannten Wiener-Prozeß, weswegen dieses Modell von Jöreskog (1970) als Quasi-Wiener-Simplex-Modell bezeichnet wird.

$$\sigma_{T(t-1)}^2 = \sigma_{Tt}^2 \iff \sigma_{\Delta t}^2 = 0 \quad (2.35)$$

D.h. nur dann, wenn von einem Zeitpunkt $t - 1$ zum nächsten Zeitpunkt t die Veränderung der wahren Merkmalsausprägungen für alle Merkmalsträger exakt dieselbe war, so daß die Varianz dieser Veränderungen bzw. der Differenzkomponente Δ_{ti} gleich Null ist, sind die True-Score-Varianzen zu den Zeitpunkten $t - 1$ und t gleich. Damit aber ist der Fall paralleler Stabilität beschrieben: Identische Veränderung für alle Merkmalsträger und invariante True-Score-Varianz (vgl. auch Gleichung 2.19).

Eine Alternative zur vollständigen Identifikation des Modells über gleiche Meßfehlervarianzen wäre die Restriktion $\sigma_{\Delta t}^2 = \sigma_{\Delta}^2$ für $t = 2, \dots, k$ – d.h. also, daß True-Score-Veränderungen zwischen den sukzessiven Meßzeitpunkten jeweils mit derselben Varianz stattfinden. Dieses bedeutet eine spezifische „inhaltliche“ Annahme über den Veränderungsprozeß: Veränderungen sind demnach nicht nur unabhängig von bereits bestehenden Merkmalsausprägungen, es werden auch keine zeitspezifischen Einflüsse auf den Prozeß dergestalt angenommen, daß sich die Merkmalsträger während mancher Zeitintervalle sozusagen gleichmäßiger verändern, als während anderer. Diese Restriktion bedingt (vgl. Gleichung 2.34) einen linearen Anstieg der True-Score-Varianzen über die Meßzeitpunkte, gleichbleibende Reliabilitäten sind dann durch entsprechend wachsende Meßfehlervarianzen (nämlich: $\sigma_{Et}^2 = \sigma_{E(t-1)}^2(1 + \sigma_{\Delta}^2/\sigma_{T(t-1)}^2)$) möglich.

Durch die Annahme unkorrelierter Differenzkomponenten erweist sich also dieses Modell als sehr restriktiv hinsichtlich der Möglichkeit zeitlicher True-Score-Varianzhomogenität, v.a. aber impliziert es auch einen sehr rigiden Stabilitätsbegriff: Es „erlaubt“ nämlich lediglich eine mindestens parallele und keine bloß monotone Stabilität der wahren Merkmalsausprägungen. Denn für die True-Score-Kovarianzen $\sigma_{T(t-1)Tt}$ gilt (für alle Zeitpunkte $t > 1$ – vgl. auch Formel 2.15 in Kapitel 2.2.1, die allgemein die Kovarianz zwischen zwei Merkmalswerten einer Zeitreihe ausdrückt und entsprechend auch für die True-Score-Kovarianzen gilt):

$$\sigma_{T(t-1)Tt} = \sigma_{T(t-1)}^2 \quad (2.36)$$

Daraus resultiert die True-Score-Korrelation wie folgt (vgl. wiederum die entsprechende Gleichung 2.17 aus Kapitel 2.2.1):

$$\rho_{T(t-1)Tt} = \frac{\sigma_{T(t-1)Tt}}{\sqrt{\sigma_{T(t-1)}^2 \sigma_{Tt}^2}} = \frac{\sigma_{T(t-1)}^2}{\sqrt{\sigma_{T(t-1)}^4 + \sigma_{T(t-1)}^2 \sigma_{\Delta t}^2}} \quad (2.37)$$

Wie man auf der rechten Seite der Gleichung 2.37 erkennen kann, kann die True-Score-Korrelation nur dann den Wert 1 annehmen, wenn $\sigma_{\Delta t}^2$, d.h. die Varianz der Veränderungen von $t - 1$ zu t , gleich 0 ist, und es gilt bei Annäherung an die „perfekte“ True-Score-Korrelation: $\rho_{T(t-1)Tt} \rightarrow 1 \implies \sigma_{\Delta t}^2 \rightarrow 0$. Vollkommen monoton stabil kann somit in

diesem Modell ein Merkmal nur sein, wenn die Veränderungsbeträge zwischen Merkmals-trägern nicht variieren und alle ein und dieselbe Veränderung durchlaufen, – d.h. es handelt sich dann um parallele Stabilität. Denn bei vorhandener Varianz der Veränderungsbeträge käme eine vollkommene monotone Stabilität dann zustande, wenn die Veränderungen „perfekt“ mit den Ausgangswerten (positiv oder negativ) korreliert wären.¹⁵

Die bloß monotone Stabilität des Merkmals würde also in diesem Differenzkomponentenmodell dann möglich, wenn die Restriktionen bezüglich der Kovarianzen der Differenzkomponenten aufgehoben würden, d.h. wenn diese frei schätzbare Parameter wären. Damit aber wäre das Modell auf jeden Fall unteridentifiziert, da bereits die erste notwendige Bedingung für die Schätzbarkeit der Modellparameter, nämlich die von Bollen (1989) sogenannte „t-Regel“, daß die Anzahl der zu schätzenden Parameter nicht die Anzahl non-redundanter Elemente in der Varianz-Kovarianz-Matrix der beobachteten Variablen überschreiten darf, nicht erfüllt wäre: Bei insgesamt k Meßzeitpunkten und invarianten Meßfehlervarianzen wäre diese Meßfehlervarianz, sowie k Varianzen der Differenzkomponenten und $(k^2 - k)/2$ Kovarianzen zwischen den Differenzkomponenten, d.h. also insgesamt $1 + (k^2 + k)/2$ Parameter zu schätzen, bei nur $(k^2 + k)/2$ nichtredundanten Elementen der Varianz-Kovarianz-Matrix der beobachteten Variablen. Sieht man von den Möglichkeiten ab, die Identifikation des Modells durch weitere Restriktionen (z.B. Gleichsetzung *aller* Differenzkomponentenvarianzen) zu erreichen, so bleibt das Modell also auf die hinsichtlich des implizierten Veränderungsprozesses sehr restriktive Variante unkorrelierter Differenzkomponenten beschränkt.

In dieser Beschränktheit des LDCM liegt ein m.E. entscheidender Nachteil gegenüber dem QMSM, welches wie gezeigt bei der Modellierung einer Meßwertreihe mit jeweils einer Meßwertvariablen X_{ti} pro Meßzeitpunkt eine weitgehend unrestringierte Modellierung verschiedener Stabilitätscharakteristika erlaubt. Allerdings wäre ein LDCM mit frei schätzbaren Kovarianzen zwischen den Differenzkomponenten dem QMSM gleichwertig,

¹⁵Wenn die Restriktion unabhängiger Differenzkomponenten entfällt, ändert sich Formel 2.37 wie folgt:

$$\rho_{T(t-1)Tt} = \frac{\sigma_{T(t-1)}^2 + \sigma_{T(t-1)\Delta t}}{\sqrt{\sigma_{T(t-1)}^4 + \sigma_{T(t-1)}^2 \sigma_{\Delta t}^2 + 2\sigma_{T(t-1)}^2 \sigma_{T(t-1)\Delta t}}}$$

Im Falle von $\rho_{T(t-1)Tt} = 1$ gilt dann:

$$\begin{aligned} (\sigma_{T(t-1)}^2 + \sigma_{T(t-1)\Delta t})^2 &= \sigma_{T(t-1)}^4 + \sigma_{T(t-1)}^2 \sigma_{\Delta t}^2 + 2\sigma_{T(t-1)}^2 \sigma_{T(t-1)\Delta t} \\ \sigma_{T(t-1)}^4 + 2\sigma_{T(t-1)}^2 \sigma_{T(t-1)\Delta t} + \sigma_{T(t-1)\Delta t}^2 &= \sigma_{T(t-1)}^4 + \sigma_{T(t-1)}^2 \sigma_{\Delta t}^2 + 2\sigma_{T(t-1)}^2 \sigma_{T(t-1)\Delta t} \\ \sigma_{T(t-1)\Delta t}^2 &= \sigma_{T(t-1)}^2 \sigma_{\Delta t}^2 \end{aligned}$$

Für die Korrelation von $T_{(t-1)i}$ mit Δ_{ti} ergibt dieses:

$$\rho_{T(t-1)\Delta t} = \frac{\sqrt{\sigma_{T(t-1)\Delta t}^2}}{\sqrt{\sigma_{T(t-1)}^2 \sigma_{\Delta t}^2}} = \pm 1$$

wie die obigen Erläuterungen gezeigt haben sollten: Das LDCM wäre dann ein unrestringiertes Modell zur Analyse der Veränderungsbeträge in den True-Scores und es würden die im Kapitel 2.2.1 für die Δ -Komponenten spezifizierten Ausdrücke berechenbar.

Ein vollständig identifiziertes Modell korrelierter Differenzkomponenten Δ_{ti} kann spezifiziert werden, wenn für jeden Meßzeitpunkt mehr als eine Messung des Merkmals vorhanden ist. So stellen z.B. Steyer et al. (1997) ein LDCM mit jeweils zwei Meßwerten, die zu jedem Meßzeitpunkt als Indikatoren für die latenten Differenzkomponenten Δ_{ti} zur Verfügung stehen, vor und Steyer et al. (2000) zeigen ein solches mit drei Indikatoren pro Meßzeitpunkt (allerdings werden in beiden Veröffentlichungen diese Modelle nicht als Differenzkomponentenmodell bezeichnet). Ein solches Modell, das müßte aus den bisherigen Darlegungen deutlich geworden sein, wäre eine gleichwertige Alternative zu einem entsprechenden QMSM mit zwei/drei Indikatoren pro Meßzeitpunkt.

Auch das LDCM kann natürlich zur Modellierung von Veränderungen in den True-Score-Mittelwerten um eine Mittelwertsstruktur erweitert werden. Dieses könnte genauso, wie beim QMSM dargestellt und diskutiert, auf die Weise geschehen, daß lediglich die Meßwerte X_{ti} mit Mittelwerten ungleich Null modelliert werden und die latenten Differenzkomponenten mit dem Mittelwert Null (d.h. als „mean deviation scores“) modelliert bleiben – dazu müßte die Modellgleichung 2.31 wiederum mit einem Intercept μ_t ergänzt werden, welches den jeweiligen Meßwertmittelwert zum Zeitpunkt t bezeichnet. Allerdings entspräche m.E. hier eine „direkte“ Modellierung von Mittelwerten der Differenzkomponenten mehr der Modellogik der Dekomponierung der True-Scores in sukzessive Veränderungsbeträge. Denn die Mittelwerte der Differenzkomponenten Δ_{ti} wären dann die durchschnittlichen True-Score-Veränderungen vom Zeitpunkt $t-1$ zum Zeitpunkt t , welche somit direkt geschätzt würden, so daß theoretische Annahmen über die im Durchschnitt stattfindenden Veränderungen leicht in entsprechende Restriktionen der Δ -Mittelwerte umgesetzt werden könnten. Für eine solche Modellierung müßte zusätzlich zur Modellgleichung 2.31 noch eine Gleichung zur Modellierung dieser der Differenzkomponentenmittelwerte z.B. folgendermaßen spezifiziert werden:

$$\Delta_{ti} = \alpha_t + \Delta_{ti}^* \quad (2.38)$$

Durch Gleichung 2.38 werden die Differenzkomponenten Δ_{ti} , welche bislang exogene, durch keine Modellgleichung abhängige Variablen des Modells waren, sozusagen zu endogenen gemacht, deren einziger „Prädiktor“ jedoch die jeweilige Konstante α_t ist – welche dann den Mittelwert von Δ_{ti} bezeichnet. Die Komponente Δ_{ti}^* bezeichnet nun die jeweilige Abweichung des individuellen Differenzwerts von diesem Mittelwert, die somit – als nach wie vor modellexogene Zufallskomponente – identisch mit den Δ_{ti} bei einer Modellierung ohne Mittelwerte ist. Wie aus den Gleichungen 2.31 und 2.38 leicht zu ersehen ist, ergibt sich der jeweilige Meßwertmittelwert zum Zeitpunkt t als Summe aller Differenzkomponentenmittelwerte bis zu diesem Zeitpunkt:

$$\mu_t = \sum_{i=1}^t \alpha_i$$

Da durch das Meßmodell der klassischen Testtheorie die Identität von True-Score- und Meßwertmittelwerten eines jeden Zeitpunkts t festgelegt ist, bezeichnet α_1 den True-Score- bzw. Meßwertmittelwert des ersten Meßzeitpunkts μ_1 und ergibt sich jeder weitere Mittelwert μ_t als Summe von μ_1 und den darauffolgenden mittleren Veränderungen $\alpha_2, \dots, \alpha_t$. Zur Festlegung z.B. invarianter True-Score-Mittelwerte müßten also alle durchschnittlichen Veränderungen auf Null restringiert werden (d.h. $\alpha_t = 0$ für alle $t > 1$).

2.2.2.3 Das Modell latenter Differenzbeträge

Abbildung 2.6 zeigt ein Modell, welches gewissermaßen Eigenschaften von QMSM und LDCM vereint. Das Meßmodell für die True-Scores T_{ti} ist dabei wiederum durch Gleichung 2.22 beschrieben, die Modellgleichungen lauten:

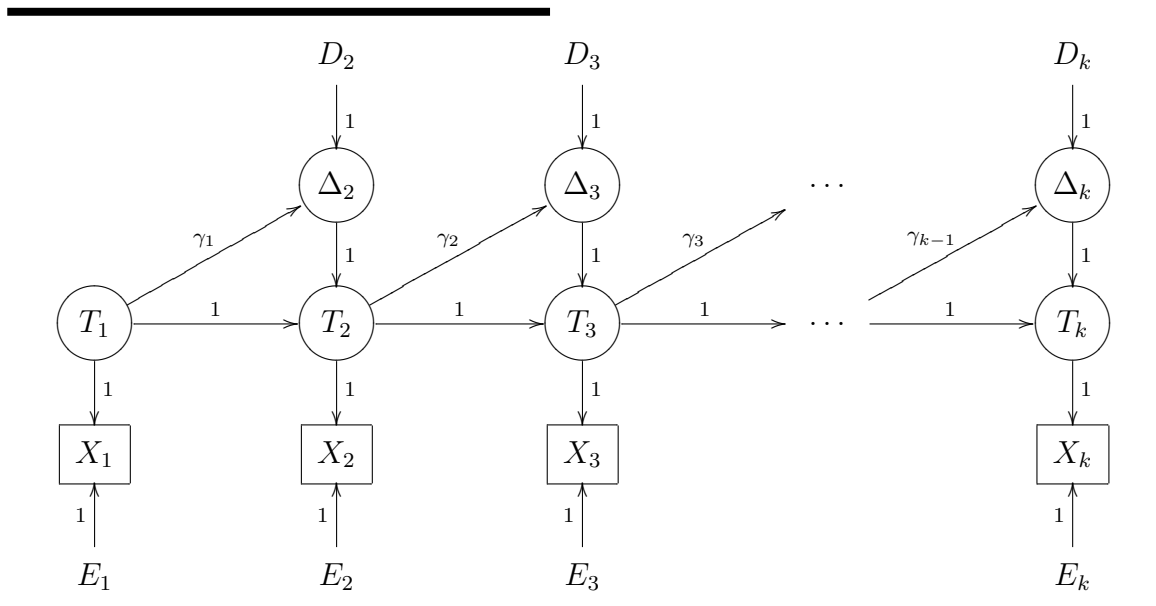
$$X_{ti} = T_{ti} + E_{ti}$$

$$T_{ti} = T_{(t-1)i} + \Delta_{ti} \quad (t = 2, \dots, k) \quad (2.39)$$

$$\Delta_{ti} = \gamma_{t-1} T_{(t-1)i} + D_{ti} \quad (t = 2, \dots, k) \quad (2.40)$$

Die Modellgleichung 2.39 modelliert also den True-Score zum Zeitpunkt $t (> 1)$ als Summe des „vorherigen“ True-Scores zum Zeitpunkt $t - 1$ und der True-Score-Veränderung Δ_{ti} , die im vorangegangenen Zeitintervall stattgefunden hat. Damit haben die Komponenten Δ_{ti} dieselbe Bedeutung, wie im LDCM (vgl. Gleichung 2.32). Im Unterschied zum LDCM sind die True-Score-Veränderungen Δ_{ti} jedoch hier nicht als exogene Variablen modelliert: Das Charakteristikum dieses Modells, welches in Gleichung 2.40 zum Ausdruck kommt, besteht darin, daß die sukzessiven Veränderungen der True-Scores zu jedem Zeitpunkt $t (> 1)$ abhängig vom True-Score des vorherigen Zeitpunkts $t - 1$ modelliert sind. Damit, so könnte man sagen, stellt das Modell gewissermaßen eine direktere und allgemeinere Beschreibung eines in k diskreten Erhebungszeitpunkten (mit jeweils gleichen Zeitintervallen dazwischen) beobachteten Veränderungsprozesses in der Sprache der Strukturgleichungsmodelle dar, als alle zuvor besprochenen Modelle: Hier ist die in einem der zugrundegelegten Zeitintervalle stattfindende Veränderung in der jeweiligen Δ -Komponente „direkt“ modelliert und kann es kann sowohl vom jeweiligen Ausgangslevel zum Beginn diese Modells abhängen, wieviel Veränderung stattfindet – der jeweilige Koeffizient γ_t ist das Maß der Stärke dieser Abhängigkeit –, als auch von nicht spezifizierten Faktoren außerhalb des Modells, welche durch die modellexogenen Zufallskomponenten D_{ti} beschrieben sind. Die beim LDCM diskutierte Problematik unkorrelierter Differenzkomponenten Δ_{ti} entfällt hier also.

ABBILDUNG 2.6:
Latent-Difference-Scores-Model (LDSM)

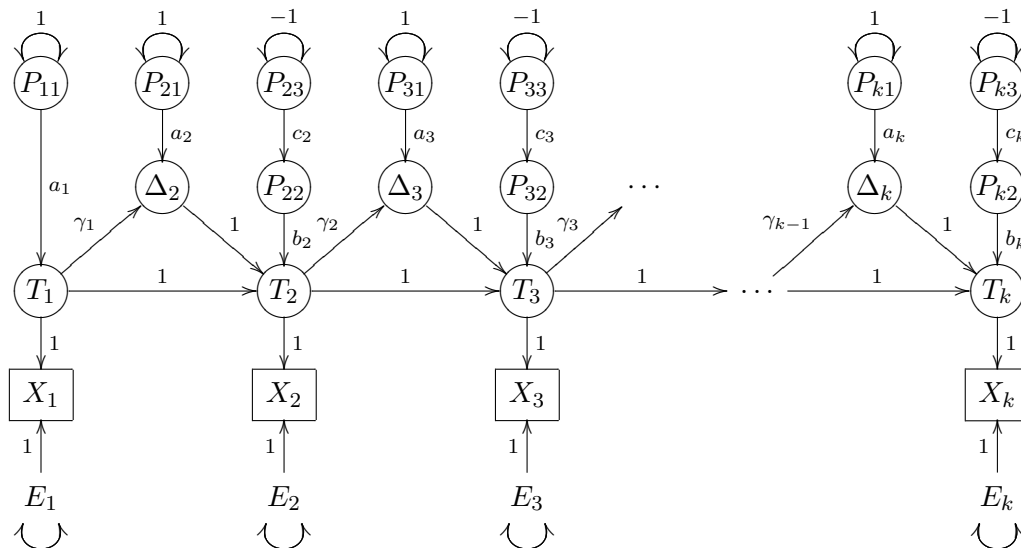


Die längsschnittliche Modellierung der Δ_{ti} durch die Gleichungen 2.39 und 2.40 entspricht der von McArdle mit verschiedenen Koautoren vorgestellten Modellierung von sogenannten „latent difference scores“ (z.B. McArdle & Nesselroade, 1994; McArdle & Hamagami, 2001) und kann deshalb als Latent-Difference-Scores-Model bzw. als Modell latenter Differenzbeträge bezeichnet werden – im folgenden wird in Anlehnung an die weitverbreitete englische Bezeichnung die Abkürzung LDSM verwendet. Mit seiner Modellierung von in zwei additive Komponenten ($\gamma_{t-1}T_{(t-1)i}$ und D_{ti}) dekomponierten latenten Veränderungsbeträgen ähnelt es auf den ersten Blick besonders dem von McArdle & Hamagami (2001) vorgestellten „dual change scores model“. Auf den zweiten Blick erweist sich letztes allerdings ähnlich restriktiv in der Modellierung des Veränderungsprozesses, wie das im folgenden noch vorgestellte Latent-Growth-Curve-Model, da es ebenfalls mit nur zwei modellexogenen interindividuellen „Varianzquellen“ für die True-Scores zu allen Meßzeitpunkten auskommt (vgl. das folgende Unterkapitel 2.2.2.4).

Dagegen offenbart der Vergleich des LDSM mit dem QMSM, daß beide Modellierungen leicht ineinander überführt werden können:

$$\begin{aligned}
 T_{ti} &= \beta_{t-1}T_{(t-1)i} + D_{ti} && \text{[QMSM]} \\
 &= T_{(t-1)i} + \Delta_{ti} = (1 + \gamma_{t-1})T_{(t-1)i} + D_{ti} && \text{[LDSM]} \\
 &\Downarrow \\
 \beta_{t-1} &= 1 + \gamma_{t-1} && (2.41)
 \end{aligned}$$

ABBILDUNG 2.7:
LDSM mit Phantomvariablen zur Restriktion der True-Score-Varianzen



In beiden Modellen handelt es sich bei der Komponente D_{ti} um dasselbe, nämlich die modellexogenen, von den True-Score-Werten $T_{(t-1)i}$ unabhängigen Veränderungsbeiträge (weshalb sie auch in den Modellgleichungen 2.24 bzw. 2.40 gleich bezeichnet wurden). Deshalb besteht der einzige prinzipielle Unterschied zwischen beiden Modellen sozusagen in der Parametrisierung der autoregressiven Komponente, wie er in der obigen Gleichung 2.41 zum Ausdruck kommt: Wenn man im QMSM die unstandardisierten Stabilitätskoeffizienten β_t durch die Ausdrücke $1 + \gamma_t$ ersetzt, so hat man bereits das QMSM in das LDSM überführt – und umgekehrt. Da zudem die sonstigen zu schätzenden Modellparameter in QMSM und LDSM dieselben sind (nämlich die Varianzen σ_{T1}^2 , σ_{Dt}^2 , σ_{Et}^2), handelt es sich bei beiden Modellen also im Grunde nur um Varianten ein und desselben autoregressiven Modells, deren Unterschied in der Art und Weise der Modellierung modellendogener True-Score-Veränderungen besteht.

Die Variante des LDSM bietet den praktischen Vorteil, daß darin die Veränderungsbeiträge in den Modellkomponenten Δ_{ti} modelliert sind, so daß die Veränderung gewissermaßen direkter untersucht werden kann. Beispielsweise könnte auf der Basis dieses Modells der Einfluß bestimmter – zeitlich variierender oder invarianter – Variablen auf die Veränderung untersucht werden, indem diese Variablen als zusätzliche Prädiktoren der Δ_{ti} eingesetzt würden.

Bezüglich der Berechnung von Stabilitätskennwerten allerdings erscheint das LDSM etwas unhandlicher, als das QMSM, denn es impliziert eine zusätzliche Komplexität der Berechnung von Stabilitätskoeffizienten, da ja im LDSM der Einfluß von T_{ti} auf $T_{(t+1)i}$

aus der Summe des direkten und eines indirekten Einflusses über die Δ -Komponente modelliert ist, so daß die Korrelation $\rho_{T_t T_{(t+1)_i}}$ nicht mehr einfach dem standardisierten Wert des Koeffizienten β_t entspricht. Zwar sind auch im QMSM, wie oben bereits dargestellt, die True-Score-Korrelationen bzw. standardisierten Werte der β -Koeffizienten nur indirekt zu schätzende Parameter, deren Berechnung auf den True-Score-Varianzen beruht, welche wiederum selbst nur indirekt geschätzt werden. Jedoch geben die gebräuchlichen SEM-Softwarepakete in der Regel standardisierte Koeffizientenwerte aus, so daß diese Berechnung nicht „von Hand“ durchgeführt werden muß. Im LDSM werden die Stabilitätskoeffizienten, da es sich nicht um den standardisierten Wert des direkten Pfadkoeffizienten von T_{ti} auf $T_{(t+1)_i}$ handelt, nicht direkt ausgegeben. Allerdings zeigt der zweite Blick, daß ihre Berechnung kaum komplizierter ist, denn tatsächlich entspricht die Summe der standardisierten direkten und indirekten Einflüsse von T_{ti} auf $T_{(t+1)_i}$ dem standardisierten Gesamteinfluß und somit, gemäß Gleichung 2.41, dem standardisierten Wert des Koeffizienten β_t im QMSM.¹⁶

Da es sich bei dem LDSM lediglich um eine Modellvariante des QMSM handelt, kann auch die dort vorgestellte Möglichkeit der Restriktion gleicher True-Score-Varianzen, Reliabilitäten und Stabilitätskoeffizienten mithilfe von Phantomvariablen gleichermaßen für das LDSM angewandt werden – dieses zeigt Abbildung 2.7 (wobei hier, anders als in Abbildung 2.4, überflüssige D -Zufallskomponenten weggelassen wurden). Die dazu notwendigen Restriktionen für die Pfadkoeffizienten lauten:

$$b_2 = \gamma_1 + 1, b_3 = \gamma_2 + 1, \dots, b_k = \gamma_{k-1} + 1 \quad \text{bzw.} \quad b_t = \gamma_{t-1} + 1 \quad (2.42a)$$

$$a_1 = c_2, a_2 = c_3, \dots, a_{k-1} = c_k \quad \text{bzw.} \quad a_t = c_{t+1} \quad (2.42b)$$

$$a_1 = a_2 = a_3 = \dots = a_k \quad \text{bzw.} \quad a_t = a \quad (2.42c)$$

¹⁶Dieses ist leicht zu zeigen. Dazu werden die beteiligten Pfadkoeffizienten, die in Abbildung 2.6 nicht eigens benannt wurden (da ihre unstandardisierten Werte auf 1 festgesetzt sind) behelfsweise benannt – nämlich der Pfadkoeffizient von T_{ti} auf $T_{(t+1)_i}$ mit b_t und der von Δ_{ti} auf $T_{(t+1)_i}$ mit c_t .

Für das dargestellte LDSM gelten demnach die Restriktionen $b_t = c_t = 1$ und die standardisierten Koeffizientenwerte können wie folgt berechnet werden:

$$\gamma_t^s = \frac{\gamma_t \sigma_{Tt}}{\sigma_{\Delta t}} \quad ; \quad b_t^s = \frac{b_t \sigma_{Tt}}{\sigma_{T(t+1)}} = \frac{\sigma_{Tt}}{\sigma_{T(t+1)}} \quad ; \quad c_t^s = \frac{c_t \sigma_{\Delta t}}{\sigma_{T(t+1)}} = \frac{\sigma_{\Delta t}}{\sigma_{T(t+1)}}$$

Die Summe aller standardisierten Einflüsse von T_{ti} auf $T_{(t+1)_i}$ beträgt demnach:

$$\begin{aligned} b_{1t}^s + \gamma_t^s c_t^s &= \frac{\sigma_{Tt}}{\sigma_{T(t+1)}} + \frac{\gamma_t \sigma_{Tt}}{\sigma_{\Delta t}} \frac{\sigma_{\Delta t}}{\sigma_{T(t+1)}} \\ &= (1 + \gamma_t) \frac{\sigma_{Tt}}{\sigma_{T(t+1)}} = \beta_t \frac{\sigma_{Tt}}{\sigma_{T(t+1)}} = \beta_t^s \end{aligned}$$

Daraus resultieren, wenn man die entsprechenden Varianzausdrücke berechnet, wiederum gleiche True-Score-Varianzen $\sigma_{T_1}^2 = \sigma_{T_2}^2 = \sigma_{T_3}^2 = a_1^2$. Invariante Stabilitätskoeffizienten wären durch die zusätzliche Restriktion $\gamma_1 = \gamma_2$ bzw. $\gamma_t = \gamma$ zu erzwingen.

Auch bezüglich der Modellidentifikation gilt hier die weitgehende Äquivalenz von LDSM und QMSM: Da sich beide Modelle letztendlich nur darin unterscheiden, daß anstatt der Koeffizienten β_t die Koeffizienten $\gamma_t = \beta_t - 1$ zu schätzen sind, gilt für die Identifikation der Parameter des LDSM dasselbe, wie beim QMSM (vgl. auch Anhang A.2).

Die Einbeziehung einer Mittelwertsstruktur in das LDSM könnte wiederum genauso gehandhabt werden, wie beim QMSM vorgeschlagen, also über die Aufnahme von Intercepts in die Meßmodellgleichungen, d.h. durch Ersetzen der Gleichungen 2.24 durch 2.30. Jedoch würde damit der Vorteil der ‘‘direkten‘‘ Modellierung der Veränderungsbeträge Δ_{ti} gewissermaßen verschenkt: Der eigentliche Sinn der Analyse der Mittelwertsstruktur besteht ja in der Untersuchung durchschnittlicher Veränderungen, und da die Veränderungsbeträge selbst Modellkomponenten des LDSM sind, liegt es hier natürlich nahe, daß auch deren Mittelwerte ‘‘direkt‘‘ zu modellieren. Mit anderen Worten: Es erscheint der Logik des LDSM zu entsprechen, die Modellgleichungen für die Δ -Komponenten um ein Intercept zu erweitern. Gleichung 2.40 würde also wie folgt ergänzt:

$$\Delta_{ti} = \alpha_t + \gamma_{t-1}T_{(t-1)i} + D_{ti} \quad (2.43)$$

Die α_t in Gleichung 2.43 entsprechen dann den jeweiligen Mittelwerten der Δ -Komponenten, d.h. es handelt sich um die sukzessiven True-Score-Mittelwertsdifferenzen. Allerdings fehlt in 2.43 noch der Mittelwert des ersten Meßzeitpunkts, welcher zusätzlich ins Modell aufgenommen werden müßte. Formal geschieht dieses durch folgende weitere Modellgleichung:

$$T_{1i} = \alpha_1 + D_{1i} \quad (2.44)$$

In Gleichung 2.44 entspricht α_1 dem Mittelwert von T_{1i} und D_{1i} den Abweichungsbeträge der individuellen True-Scores von diesem Mittelwert – das Erscheinen dieser zusätzlichen Komponente mag etwas verwirren, es ist die Konsequenz der vorne gewählten einfachen formalen Darstellung, die sozusagen lediglich den Erfordernissen der reinen Kovarianzstrukturanalyse angepaßt ist. Man kann diese Konsequenz in einfachen Worten wie folgt umschreiben: Da jetzt die Variable T_{ti} durch ihren Mittelwert vorhergesagt wird, ist sie von einer ursprünglich exogenen zur endogenen geworden, deren Varianz aus einer exogenen Varianzquelle D_{ti} gespeist wird. Tatsächlich war diese eigentlich exogene Varianzquelle gewissermaßen immer schon mitgedacht und wurde z.B. in den Abbildungen 2.3 und 2.6 nur nicht explizit dargestellt. Die Aufnahme von D_{1i} hat nur formalen Charakter, praktisch bedeutet die Gleichung 2.44 lediglich, daß damit auch der Mittelwert des ersten True-Scores zum frei schätzbaren Modellparameter wird.

LDSM und QMSM sind also weitgehend äquivalente autoregressive Modelle, deren

Unterschiede eigentlich nur nach pragmatischen Gesichtspunkten relevant scheinen: Zielt das Forschungsinteresse mehr auf die Stabilitätskoeffizienten, so bietet das QMSM den etwas einfacheren Weg zur Schätzung derselben, während dann, wenn eher die direkte Analyse der Veränderungsbeträge vorrangig ist, die „direkte“ Modellierung derselben im LDSM von Vorteil ist.

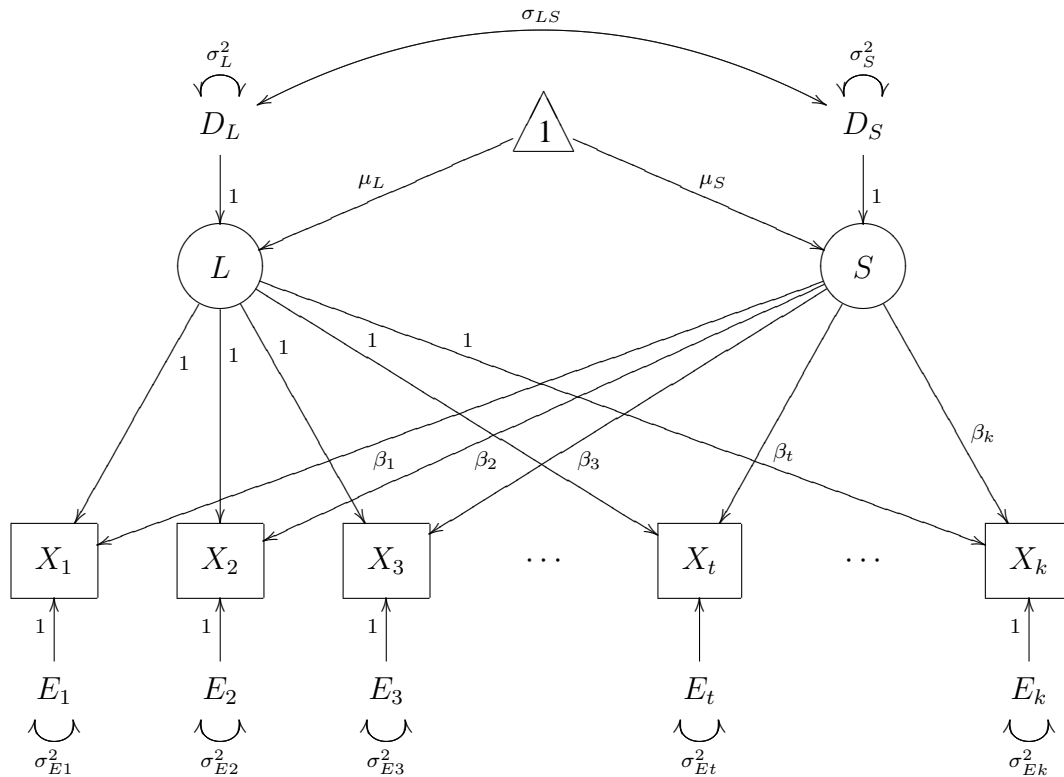
2.2.2.4 Weitere längsschnittliche Veränderungsmodelle

Es existiert, wie bereits gesagt, eine gewisse Fülle an Veröffentlichungen zu längsschnittlichen Strukturgleichungsmodellen und somit auch an vorgeschlagenen Modellvarianten. Die besprochenen QMSM und LDCM sind sozusagen zwei Urtypen solcher Ein-Indikator-Modelle, d.h. Veränderungsmodelle für Zeitreihen mit jeweils nur einer Meßwertvariablen X_{ti} pro Meßzeitpunkt. Von Jöreskog (1970) als Modelle zur Analyse von Zeitreihen von Meßwerten mit Simplex-Struktur der Autokovarianzmatrix und vorhandenem Meßfehler vorgestellt, handelt es sich gewissermaßen um die grundlegende autoregressive bzw. faktoranalytische Modellvariante eines Quasi-Simplex-Modells. Hinsichtlich der Typisierung verschiedener Modelltypen von McArdle & Aber (1990) stellen QMSM – und damit auch LDSM, welches ja wie dargelegt lediglich eine Modellvariation des QMSM ist, – und LDCM diejenigen Versionen eines autoregressiven Modells und eines Differenzkomponentenmodells dar, die für die hier angestrebten Stabilitätsanalysen mit Ein-Item-Zufriedenheitsratings des SOEP von Interesse wären.

Abgesehen von diesen bisher behandelten Modellen existiert eine Vielfalt weiterer längsschnittlicher Modellierungen. Die gewissermaßen „prominentesten“ und in neueren Veröffentlichungen über Strukturgleichungsmodelle für Längsschnittdaten dem Anschein nach meistbehandelten Modelle des faktoranalytischen Typus sind die Latent Growth Curve Models (LGCM), welche in den Jahren seit den ersten (expliziten) diesbezüglichen Veröffentlichungen von McArdle (1986) und McArdle & Epstein (1987) eine weite Verbreitung erfahren haben (eine umfassende Darstellung liefert McArdle, 1998).¹⁷ Abbildung 2.8 stellt die sozusagen klassische Variante des LGCM in allgemeiner Form dar, zur Veranschaulichung der Besonderheiten dieses Modelltyps sind dort (wie auch schon in Abbildung 2.4) alle zu schätzenden Parameter – inklusive der unabhängigen Varianzen und Kovarianzen – abgebildet. Auch enthält die Abbildung – anders als die bisherigen Darstellungen

¹⁷Die geläufigen Bezeichnungen dieses Modelltyps in den einschlägigen Veröffentlichungen variieren ein wenig: McArdle (1986) nennt sie „latent growth models“, weitere gebräuchliche Begriffe sind „latent curve analysis“ (z.B. bei Meredith & Tisak, 1990; Tisak & Tisak, 1996) und „growth curve models“ (z.B. bei Bijleveld et al., 1998). M.E. ist die Bezeichnung als „latent growth curve model“ am präzisesten: Einerseits zeigt der Begriff „latent“ an, daß es sich um ein Strukturgleichungsmodell mit als latenten Konstrukten modellierten Level- und Slope-Komponenten handelt – statistische Längsschnittmodelle dieses Typus können auch als hierarchische Multilevel-Regressionsmodelle spezifiziert werden (z.B. van der Leeden, 1998), wie bereits im Kapitel 2.1.2, erwähnt wurde. Andererseits aber macht der Begriff „curve“ deutlich, daß es sich nicht nur um unspezifische Veränderungs-, sondern um spezifische Verlaufsmodelle handelt, wie im folgenden noch dargelegt wird.

ABBILDUNG 2.8:
Latent Growth Curve Model (LGCM)



des QMSM, LDCM und LDSM (vgl. Abbildungen 2.3, 2.5, 2.6) – eine Mittelwertskomponente, denn im Gegensatz zu QMSM und LDCM umfaßt die ursprüngliche und allgemein gebräuchliche Spezifikation des LGCM die Modellierung von Autokovarianzen *und* Mittelwerten in der Zeitreihe der beobachteten Variablen.

Das Charakteristikum dieses LGCM ist die Modellierung der den beobachteten Werten X_{ti} zugrundeliegenden Veränderung durch eine sogenannte Level- und eine sogenannte Slope-Komponente, welche in Abbildung 2.8 als latente Variablen L und S dargestellt sind. Die Bedeutung dieser Komponenten wird erkennbar, wenn man die folgende grundlegende Modellgleichung für das LGCM, betrachtet:

$$X_{ti} = L_i + \beta_t S_i + E_{ti} \quad (2.45)$$

Demnach setzt sich also der zu einem Zeitpunkt t gemessene X -Wert eines Merkmalsträgers i zusammen aus einem individuellen, zeitlich invarianten Levelwert L_i , einer „systematischen“ Abweichung von diesem Levelwert zum Zeitpunkt t , welche durch den Ausdruck $\beta_t S_i$ genauer festgelegt ist, sowie einem jeweils zufälligen Anteil E_{ti} . Abgesehen

von einer Zufallsschwankung der Meßwerte zum jeweiligen Meßzeitpunkt t , die durch E_{ti} beschrieben ist, ist im LGCM somit Veränderung als Abweichung von einem zeitlich unveränderlichen Levelwert spezifiziert, welche als Produkt eines für jeden Merkmalsträger individuell spezifischen Slope-Scores S_i und eines für den jeweiligen Meßzeitpunkt spezifischen, jedoch über alle Merkmalsträger i hinweg invarianten Koeffizienten β_t festgelegt ist – dieses ist gewissermaßen das fundamentale Merkmal längsschnittlicher Modellierung durch das LGCM, durch welches die Besonderheiten dieses Modelltyps begründet sind, die im folgenden noch kurz dargestellt werden. Zunächst aber ist noch darauf hinzuweisen, daß auch in diesem Modell die E_{ti} – als wechselseitig unkorrelierte Zufallskomponenten an den beobachteten Werten X_{ti} – mit den Eigenschaften der Meßfehler des Modells der klassischen Testtheorie modelliert sind, so daß die True-Scores des Meßzeitpunktes t sich hier also wie folgt ergeben:

$$T_{ti} = L_i + \beta_t S_i \quad (2.46)$$

Somit sind die Varianzen $\sigma_{E_t}^2$ wiederum die Meßfehlervarianzen und die Reliabilitäten entsprechend mit $Rel_t = 1 - \sigma_{E_t}^2 / \sigma_t^2$ zu bestimmen (vgl. insbesondere Tisak & Tisak, 1996).

Vergleicht man nun die Gleichung 2.46 mit den Gleichungen 2.24 und 2.32, welche die für das QMSM bzw. LDCM spezifische Zusammensetzung der True-Scores zeigten, so wird ein wesentlicher Unterschied deutlich: Im LGCM bestimmen zu jedem Zeitpunkt t lediglich die zwei “individualspezifischen“ Werte L_i und S_i den True-Score T_{ti} , während in QMSM und LDCM mit fortschreitendem t auch die Zahl der Modellkomponenten steigt, aus denen sich die True-Scores zusammensetzen. Beim LDCM ist der T_{ti} in t Differenzkomponenten (True-Score der ersten Messung plus die $t - 1$ darauffolgenden True-Score-Veränderungen) dekomponiert und löst man in Gleichung 2.24 den Ausdruck $T_{(t-1)i}$ sukzessive bis zu T_{1i} auf, so erhält man wiederum t Bestimmungsstücke von T_{ti} , nämlich den True-Score zur ersten Messung, sowie die Random-Komponenten der $t - 1$ folgenden Autoregressionsgleichungen D_{2i} bis D_{ti} . In diesen Modellen ist also für jeden Zeitpunkt ein exogener, d.h. nicht aus den anderen Modellkomponenten vorhersagbarer und insofern zufälliger, Veränderungsbeitrag spezifiziert, die Veränderung kann deshalb von Meßzeitpunkt zu Meßzeitpunkt frei zwischen den Merkmalsträgern variieren und die Veränderung des True-Scores eines Merkmalsträgers i , die im Zeitintervall t bis $t + 1$ stattfindet, ist nicht völlig durch T_{ti} bzw. die vorherigen True-Score- und/oder Veränderungswerte festgelegt.

In dem durch Gleichung 2.45 spezifizierten LGCM dagegen ist nur eine einmalige interindividuelle Variation von Veränderungswerten möglich. Das Modell impliziert insofern eine vergleichsweise restriktive Modellierung längsschnittlicher Veränderung, als dadurch festgelegt ist, daß die über die k Messungen stattfindenden sukzessiven Veränderungen allesamt durch lediglich einen zwischen den Merkmalsträgern frei variierenden Veränderungswert vollständig charakterisiert werden können. Anschaulicher wird dieses, wenn man die zwei üblichen Restriktionen für die Basiskoeffizienten β_t hinzufügt, die zur Identifikation der Slope-Variablen notwendig sind, – nämlich die Nullsetzung eines der β_t und die Fest-

setzung eines weiteren auf einen Wert ungleich Null, normalerweise Eins. Beispielsweise (dies ist wohl die gebräuchlichste Vorgehensweise) könnte festgelegt werden:

$$\beta_1 = 0 \quad ; \quad \beta_2 = 1 \quad (2.47)$$

Damit erhalten L_i und S_i eine leicht einzusehende Bedeutung. Denn für die ersten beiden Meßzeitpunkte resultiert dann aus Gleichung 2.46:

$$T_{1i} = L_i \quad (2.48)$$

$$T_{2i} = L_i + S_i \implies S_i = T_{2i} - T_{1i} = \Delta_{2i} \quad (2.49)$$

Somit ist der Levelwert L_i der True-Score zu Beginn der untersuchten Zeitreihe und der Slope-Score S_i die Veränderung des True-Scores vom ersten zum zweiten Meßzeitpunkt – was der Differenzkomponente Δ_{2i} im LDCM entspräche. Wenn die Restriktionen 2.47 für andere β_t der Zeitreihe festgesetzt würden, dann ergäben sich diese Bedeutungen entsprechend: $\beta_t = 0$ legt den Zeitpunkt t sozusagen als Referenzzeitpunkt für die Levelkomponente fest, so daß $L_i = T_{ti}$, und $\beta_{t'} = 1$ legt t' als Referenzzeitpunkt zur Bestimmung der Slope-Komponente fest, so daß $S_i = T_{t'i} - T_{ti}$. Entscheidend ist nun, daß alle sukzessiven Veränderungen der beobachteten Zeitreihe stets, d.h. zu jedem Zeitpunkt t , ein für alle Merkmalsträger gleich proportionales Vielfaches des individuellen Slope-Scores S_i sind: Wird z.B. die Restriktion 2.47 für Gleichung 2.46 festgelegt, so beträgt zu jedem Zeitpunkt $t > 2$ bei jedem Merkmalsträger i die Veränderung des True-Scores T_{ti} gegenüber T_{1i} das β_t -fache der True-Score-Differenz vom ersten zum zweiten Meßzeitpunkt. Mit anderen Worten: Das LGCM beruht auf der Annahme eines für alle Merkmalsträger identischen funktionalen Ablaufs der True-Score-Veränderung, die β_t beschreiben eine interindividuell invariante Verlaufsfunktion, die auf die interindividuell variierenden Level- und Slope-Werte anzuwenden ist. Merkmalsträger unterscheiden sich also in der – durch Restriktionen in der Art von 2.47 wie auch immer konkret festgelegten – Level- und Slope-Scores, nicht jedoch im Verlauf ihrer Veränderung relativ zu diesen beiden Werten. Das LGCM kann deshalb, im Gegensatz zu QMSM und LDCM, als *Verlaufmodell* bezeichnet werden.

Die meistgebrauchte Spezifikation des LGCM ist die eines *linearen* Verlaufmodells: Dazu wird vermittels weiterer Restriktionen der Basiskoeffizienten β_t ein linearer Verlauf der True-Score-Veränderungen über die untersuchte Zeitspanne hinweg festgelegt. Beispielsweise könnte zusätzlich zu den Restriktionen 2.47 noch spezifiziert werden:

$$\beta_3 = 2; \beta_4 = 3; \dots; \beta_k = k - 1$$

Bzw., ganz allgemein:

$$\beta_t = t - 1 \quad (2.50)$$

Durch die Restriktion 2.50 wird festgelegt, daß in jedem Zeitintervall der untersuchten Zeitreihe der True-Score eines jeden Merkmalsträgers i , ausgehend vom Wert L_i bei der ersten Messung, um jeweils denselben Betrag S_i wächst (oder schrumpft), wie die Ausformulierung der unter Restriktion 2.50 resultierenden Gleichungen 2.46 zeigt:

$$\begin{aligned} T_{1i} &= L_i \\ T_{2i} &= L_i + S_i \\ T_{3i} &= L_i + 2S_i \\ T_{4i} &= L_i + 3S_i \\ &\vdots \\ T_{ki} &= L_i + (k - 1)S_i \end{aligned}$$

Bzw., wiederum allgemein:

$$T_{ti} = L_i + S_i(t - 1) \quad (2.51)$$

Aus den Restriktionen 2.50 resultiert also die Geradengleichung 2.51: Der Verlauf der True-Score-Werte eines jeden Merkmalsträgers i über die Zeitpunkte t kann als Gerade mit der Steigung S_i und dem Ordinatenabstand L_i graphisch dargestellt werden – die Begriffe „Level“ und „Slope“ erhalten dadurch anschauliche Bedeutung.

Insgesamt ist also das LGCM durch eine im Vergleich zu QMSM und LDCM restriktivere Modellierung der sukzessiven Veränderungen charakterisiert, welche natürlich auch „technische“ Beschränkungen für Stabilitätskoeffizienten und True-Score-Varianzen impliziert. Die entsprechenden formalen Ableitungen sind komplex und die diesbezügliche Restriktivität des Modells soll deshalb hier nur am Beispiel eines Sonderfalls demonstriert werden. Dazu formuliere ich zunächst die folgenden Gleichungen für die die True-Score-Varianzen und -Kovarianzen:

$$\sigma_{Tt}^2 = \sigma_L^2 + \beta_t^2 \sigma_S^2 + 2\beta_t \sigma_{LS} \quad (2.52)$$

$$\sigma_{TtTt'} = \sigma_L^2 + \beta_t \beta_{t'} \sigma_S^2 + (\beta_t + \beta_{t'}) \sigma_{LS} \quad (2.53)$$

Daraus resultiert unter den Restriktionen 2.47 für die Korrelationen des ersten True-Scores T_{1i} mit den folgenden True-Scores T_t ($t > 1$):

$$\rho_{T_1 T_t} = \frac{\sigma_{T_1 T_t}}{\sigma_{T_1} \sigma_{T_t}} = \frac{\sigma_L^2 + \beta_t \sigma_{LS}}{\sqrt{\sigma_L^2 (\sigma_L^2 + \beta_t^2 \sigma_S^2 + 2\beta_t \sigma_{LS})}} \quad (2.54)$$

Der zu betrachtende Sonderfall ist der *einer* perfekten Korrelation $\rho_{T_1 T_t} = \pm 1$ zwischen erstem True-Score und dem True-Score zu irgendeinem Meßzeitpunkt t :

$$\begin{aligned}
\rho_{T_1 T_t} &= \pm 1 \\
&\Downarrow \\
\sigma_L^2 + \beta_t \sigma_{LS} &= \pm \sqrt{\sigma_L^2 (\sigma_L^2 + \beta_t^2 \sigma_S^2 + 2\beta_t \sigma_{LS})} \\
&\Downarrow \\
\sigma_L^4 + \beta_t^2 \sigma_{LS}^2 + 2\beta_t \sigma_L^2 \sigma_{LS} &= \sigma_L^4 + \beta_t^2 \sigma_S^2 \sigma_L^2 + 2\beta_t \sigma_L^2 \sigma_{LS} \\
&\Downarrow \\
\sigma_{LS}^2 &= \sigma_L^2 \sigma_S^2 \\
&\Downarrow \\
\rho_{LS} &= \frac{\sigma_{LS}}{\sigma_L \sigma_S} = \pm 1
\end{aligned} \tag{2.55}$$

Die Ableitung 2.55 zeigt, daß dann, wenn *irgendeine* True-Score-Korrelation in der Zeitreihe perfekt (positiv oder negativ) ist, auch die Level-Slope-Korrelation perfekt sein muß (allenfalls das Vorzeichen kann sich ändern, was v.a. von Größe und Vorzeichen des Basiskoeffizienten β_t abhängt). Da die Ableitung auch in umgekehrter Richtung gilt, ist des weiteren zu folgern, daß dann, wenn zwischen dem ersten True-Score T_{1i} und *irgendeinem* weiteren True-Score T_{ti} ein perfekter Zusammenhang besteht, auch zwischen T_{1i} und *allen* weiteren True-Scores ein perfekter Zusammenhang bestehen muß (wiederum von möglichen von Vorzeichenumkehrungen abgesehen). Da nun aber jeder beliebige Meßzeitpunkt t anstelle des ersten als Referenzzeitpunkt für die Level-Komponente L_i (durch die Restriktion $\beta_t = 0$) festgelegt werden kann, ohne daß dadurch die obige Ableitung grundsätzlich verändert würde, ist zu folgern, daß im LGCM entweder zwischen allen Meßzeitpunkten ein perfekter Zusammenhang der True-Scores besteht, oder zwischen keinen. Dies bedeutet eine Restriktion gegenüber QMSM und LDCM, in denen es jeweils möglich ist, daß nur in einzelnen und nicht in allen Zeitintervallen perfekte Stabilität besteht. Damit sollte, wie gesagt, lediglich beispielhaft gezeigt werden, daß das LGCM Beschränkungen möglicher Merkmalsstabilität impliziert, die in QMSM und LDCM nicht bestehen.

Die Stärke des LGCM liegt in seiner Sparsamkeit bei der Beschreibung von Veränderung durch nur wenige interindividuell variable Modellkomponenten, welche dann von Vorteil ist, wenn die längsschnittliche Analyse der Veränderung auf theoretischen Festlegungen auf ein interindividuell invariantes Verlaufsmodell beruht. Zur Analyse von Merkmalsstabilität, die hier für die Zufriedenheitsmessungen des SOEP durchgeführt werden soll, ist jedoch ein solch restriktives Verlaufsmodell nur bedingt tauglich. Dieses geht schon aus den theoretischen Überlegungen hervor, die vorne dargelegt wurden und die letztlich sehr stark auf der Frage beruhen, in welchem Ausmaß sich Zufriedenheit intraindividuell verändert aufgrund von Veränderungen objektiver Lebensumstände. Inzidenz und Intensität der objektiven Veränderungen aber muß als inter- und intraindividuell variierend angenommen werden, so daß dann, wenn diese auf Zufriedenheit wirken, keineswegs angenommen wer-

den kann, daß die intraindividuelle, längsschnittliche Variabilität letzterer durch ein interindividuell fixes Verlaufsmodell zu charakterisieren ist. Beispielsweise führt die diskutierte Annahme adaptiver Regulation durch Neuadjustierung internaler Soll-Zustände, anhand derer ein subjektives Vergleichsurteil gebildet wird, zur Annahme interindividuell variierender Zufriedenheitsverläufe, die jeweils durch eine individuell spezifische Abfolge von Phasen gleichbleibender bzw. sich verändernder Zufriedenheit – je nachdem, was der betreffenden Person widerfährt, – gekennzeichnet sind. Die angestrebte Stabilitätsanalyse soll deshalb in der Hinsicht „theoretisch unvoreingenommen“ sein, daß sie eben nicht a priori auf ein spezifisches Verlaufsmodell beschränkt bleibt.

Die LGCM wurden hier etwas ausführlicher dargestellt, weil die Verwendung von Modellen vom Typus des LGCM die derzeit bei weitem vorherrschende Vorgehensweise längsschnittlicher Datenanalyse mit Strukturgleichungsmodellen zu sein scheint, so daß sie gewissermaßen ganz von selbst in den Blick geraten, wenn überhaupt längsschnittliche Strukturgleichungsmodelle spezifiziert werden sollen. Die Darstellung längsschnittlicher Strukturgleichungsmodelltypen ist damit aber noch nicht vollständig und könnte noch um einiges ausgeweitet werden. Da aber diese sonstigen Modelle weder für die hier angestrebten Stabilitätsanalysen zusätzlichen Gewinn versprechen, noch eine ähnliche „Prominenz“ besitzen, wie das LGCM, möchte ich es hier noch bei einigen kurzen Hinweisen „der Vollständigkeit halber“ belassen.

Wie vorne bereits erwähnt listen McArdle & Aber (1990) längsschnittliche Faktorenanalysen als weiteren grundlegenden Modelltypus auf. Diese allerdings scheiden für Stabilitätsanalysen auf der Basis von Zeitreihen von Ein-Item-Zufriedenheitsmessungen schon deshalb aus, weil es sich dabei prinzipiell um ein Mehr-Indikatoren-Modell handelt, – nämlich, einfach gesagt, um eine konfirmatorische Faktorenanalyse über Gruppen wiederholt gemessener Indikatoren, bei denen der Meßzeitpunkt als Faktor erster Ordnung modelliert wird.

Prinzipiell ermöglicht die Strukturgleichungsmethode eine sehr große Flexibilität der Modellierung längsschnittlicher Daten, so daß die Auflistung von McArdle & Aber nicht mehr als eine Bestandsaufnahme zu ihrem Veröffentlichungszeitpunkt sein konnte und mittlerweile viele Fortentwicklungen aus diesen Grundtypen vorgeschlagen wurden – einen sehr guten Eindruck davon vermittelt der Sammelband von Collins & Sayer (2001). Unter den dort behandelten Modellen erscheint mir v.a. das „dual change score model“ besonders erwähnenswert, welches man hier knapp dadurch charakterisieren könnte, daß es sich zum LGCM ähnlich verhält, wie das LDSM zum QMSM: Es wird dabei gewissermaßen ein Verlaufsmodell, wie es einem LGCM entspräche, so modelliert, daß darin die latenten Veränderungsbeträge „direkte“ Modellkomponenten sind (vgl. McArdle & Hamagami, 2001). Hinsichtlich einer bezüglich der True-Score-Verläufe nicht festgelegten Stabilitätsanalyse ist damit für die „dual change score models“ dasselbe zu sagen, wie oben zum LGCM.

2.3 *Spezielle Probleme der Parameterschätzung*

Im folgenden werden noch einige statistische Probleme der geplanten Stabilitätsanalysen mit längsschnittlichen Strukturgleichungsmodellen behandelt: Zum einen muß geklärt werden, wie mit fehlenden Werten, die gerade in Längsschnittdatensätzen praktisch sicher erwartet werden können, umgegangen werden soll. Zum zweiten müssen vorab Regeln zur Bewertung der durch die Parameterschätzung erreichten Anpassungsgüte der zu analysierenden Modelle festgesetzt werden, insbesondere auch solche, die Vergleiche der Modellanpassung verschiedener Modellversionen mit unterschiedlichen Parameterrestriktionen, wie sie bei den Stabilitätsanalysen durchgeführt werden sollen, betreffen.

2.3.1 *Behandlung von fehlenden Werten in längsschnittlichen Strukturgleichungsmodellen*

Das Problem fehlender Werte ist ein grundsätzliches, welches sich praktisch bei jeder statistischen Analyse empirischer Daten stellt. Jedenfalls entspricht es sozusagen alltagspraktischer Datenauswertererfahrung, daß zumindest dann, wenn die Stichproben groß genug sind, um eine Anwendung von Strukturgleichungsmodellen zuzulassen, mit großer Wahrscheinlichkeit auch im Datensatz Werte fehlen, d.h. daß es für eine beliebige Variable X in einem Datensatz einzelne Fälle gibt, für die – aus welchen Gründen auch immer – kein X -Wert vorliegt. Besonders aber bei längsschnittlichen Untersuchungen ist das Problem fast unvermeidbar: Abgesehen von den Ausfallgründen, die auch in querschnittlichen Erhebungen zu fehlenden Werten führen können, kommt es hier in der Regel über die Wiederholungsmessungen hinweg zu Ausfällen derart, daß bei jeder neuen Datenerhebung ein Teil der Teilnehmer bei den vorherigen Erhebungen nun nicht mehr teilnimmt. Mögliche Ursachen für solche Ausfälle gibt es viele, in sozialwissenschaftlichen Panelstudien wie dem SOEP könnte man diesbezüglich grob Nichterreichbarkeit (z.B. infolge von Umzug zwischen zwei Meßzeitpunkten), Verweigerung und Mortalität der Teilnehmer unterscheiden. Das spezielle Muster fehlender Werte, das durch solche längsschnittlichen Ausfälle entsteht, wird im englischen Sprachraum als „drop-out“ oder „attrition“ bezeichnet (vgl. van der Kamp & Bijleveld, 1998, 11): Für einzelne Fälle liegen jeweils bis zu einem bestimmten Meßzeitpunkt alle Meßwerte vor, danach keine mehr.

Bei der Anwendung längsschnittlicher Strukturgleichungsmodelle stellt sich somit die Frage, wie dabei mit dem Problem fehlender Werte umgegangen werden soll. Die „klassischen“ SEM-Schätzalgorithmen „maximum likelihood“ (ML), „unweighted least squares“ (ULS), „generalized least squares“ (GLS) „weighted least squares“ (WLS) bzw. „asymptotically distribution free“ (ADF) (vgl. deren detaillierte Darstellungen z.B. bei Bollen, 1989, 107ff, 333ff, 425ff; Bentler, 1995, 207ff), gehen implizit von vollständigen Datensätzen der beobachteten Modellvariablen aus bzw. sind nur auf solche anwendbar.

2.3.1.1 Konventionelle Methoden zur Behandlung fehlender Werte

Es hat während der letzten beiden Jahrzehnte eine intensive wissenschaftliche Beschäftigung mit dem Problem fehlender Werte stattgefunden, die zur Entwicklung komplexer statistischer Verfahren zum Umgang mit denselben geführt hat – und die bislang nur wenig Eingang in die empiriewissenschaftliche Forschungspraxis gefunden zu haben scheint, wo allem Anschein nach bis zum gegenwärtigen Zeitpunkt einige allgemein bekannte ältere Vorgehensweisen beim Umgang mit fehlenden Werten – v.a. deren fallweiser Ausschluß – vorherrschen. Dieses mag zum Teil daran liegen, daß diese gebräuchlichen Verfahren sehr einfach anzuwenden und in vielen herkömmlichen Statistiksoftwareprogrammen bereits als Voreinstellung für die Behandlung unvollständiger Datensätze implementiert sind, zum Teil aber auch daran, daß die besagten neueren statistischen Abhandlungen zum Thema oft eine für den Nichtspezialisten nur schwer verständliche formale Komplexität aufweisen (vgl. z.B. die diesbezüglichen Kapitel bei Diggle et al., 2002; Verbeke & Molenberghs, 2000) bzw. daß die neueren Verfahren selbst komplexer sind und erst allmählich durch entsprechende Statistikprogramme einfacher handhabbar werden. Zumindest hinsichtlich des ersten dieser beiden Aspekte könnte dank ihrer m.E. sehr guten Lesbarkeit die erst in jüngster Zeit erschienene Bestandsaufnahme von Schafer & Graham (2002) Abhilfe schaffen.

Die besagten älteren und herkömmlichen Verfahren zum Umgang mit fehlenden Werten lassen sich in solche des Ausschlusses von Fällen und solche der einfachen Ersetzung fehlender Werte unterteilen. Die allgemein bekannten Ausschlußverfahren sind der fallweise Ausschluß („listwise deletion“, „complete-case analysis“) und der paarweise Ausschluß („pairwise deletion“, „available-case analysis“) von fehlenden Werten. Die bekanntesten Verfahren der einfachen Ersetzung fehlender Werte dürften die Ersetzung durch den Stichprobenmittelwert („(unconditional) mean imputation“) und die Ersetzung durch eine regressionsbasierte Schätzung („conditional mean imputation“) sein.

Beim fallweisen Ausschluß wird jede Observation i , die in einer der zu analysierenden Variablen fehlende Werte aufweist, vor der Durchführung der statistischen Analyse aus dem Datensatz eliminiert bzw. es werden nur diejenigen Fälle analysiert, die keinerlei fehlende Werte aufweisen. Auch wenn diese Methode die wohl am häufigsten und oft gewohnheitsmäßig zur Behandlung fehlender Werte angewandte ist, so ist sie doch mit offensichtlichen Nachteilen verbunden. Denn zum einen ist dabei zu fragen, ob und in welcher Weise der Ausschluß von Fällen die Ergebnisse der Analyse beeinflusst, – nämlich dann, wenn die Datenausfälle selbst in irgendeiner Art und Weise mit den untersuchten Variablen statistisch zusammenhängen, so daß die Verteilung dieser Analysevariablen in der durch den fallweisen Ausschluß bereinigten Stichprobe nicht mehr die Populationsverteilung repräsentiert. Zum zweiten aber reduziert der fallweise Ausschluß die den Analysen zugrundeliegende Stichprobengröße und damit auch die Effizienz der Parameterschätzungen. Gerade in Längsschnittstudien mit vielen Wiederholungsmessungen würde so infolge „attrition“ eine oft enorme Reduktion der Stichprobengröße bewirkt – so auch im SOEP,

(vgl. Tabelle 3.2 in Kapitel 3.1.3, sowie Tabelle B.1 in Anhang B).

Beim paarweisen Ausschluß werden nicht alle Fälle mit fehlenden Werten in einer Analysevariablen ausgeschlossen, sondern es werden bei der Berechnung von Statistiken (z.B. Mittelwerten, Varianzen, Korrelationen etc.) alle jeweils dafür vorhandenen Werte einbezogen. Wenn also bei einem Fall i in der Analysevariablen X_1 der Wert fehlt, jedoch in den Variablen X_2 und X_3 Meßwerte vorhanden sind, so würde beispielsweise dieser Fall in die Berechnung der Mittelwerte \bar{X}_2 und \bar{X}_3 und der Korrelation von X_2 und X_3 einbezogen und wäre von der Berechnung des Mittelwerts \bar{X}_1 und der Korrelationen von X_1 mit X_2 bzw. X_3 ausgeschlossen. Die Bezeichnung „available-case analysis“ beschreibt also im Grunde die Vorgehensweise genauer, während die bekanntere und weithin gebräuchlichere Bezeichnung „pairwise deletion“ eigentlich nur dem Verfahren bei der Berechnung von Kovarianzen bzw. Korrelationen entlehnt ist. Somit kommt es dabei nicht zu den Datenverlusten, die die Methode des fallweisen Ausschlusses kennzeichnen. Jedoch entstehen durch den paarweisen Ausschluß statistische Probleme komplexerer Natur dadurch, daß bei dieser Vorgehensweise unterschiedliche Statistiken auf unterschiedlichen Fällen beruhen, – für die Berechnung von Strukturgleichungsmodellen fallen zwei daraus resultierende Probleme besonders ins Gewicht: Zum einen kann die multivariate Wahrscheinlichkeitsverteilung der Varianz-Kovarianz-Matrix der beobachteten Modellvariablen, auf der die statistische Evaluation der Modellanpassung beruht, nicht mehr ohne weiteres bestimmt werden, wenn die Elemente dieser Matrix aus unterschiedlichen Fallzahlen errechnet wurden (vgl. Bollen, 1989, 370f; Wothke, 2000). Zum zweiten besteht ein sehr ernsthaftes Problem in der Möglichkeit gewisser Inkonsistenzen bei der Berechnung der Varianzen und Kovarianzen der beobachteten Modellvariablen aus sozusagen jeweils unterschiedlichen Anteilen des Datensatzes: Es können so Varianz-Kovarianz-Matrizen resultieren, die nicht positiv-definit sind, was ein Problem für die Anwendung der Modellanpassungsfunktionen (Schätzalgorithmen) bedeutet, da diese dann nicht mehr durch das theoretische Minimum Null (bei optimaler Anpassung) beschränkt sind (vgl. Wothke, 1993). Ein sehr einfaches und anschauliches Beispiel dafür zeigt Wothke (2000): Dabei resultiert aus der mittels des paarweisen Ausschlusses berechneten Varianz-Kovarianz-Matrix zweier Variablen ein Korrelationswert außerhalb des zulässigen Wertebereichs.

Angesichts der Probleme, die bei den Methoden des fallweisen oder paarweisen Ausschlusses daraus resultieren, daß entweder die Stichprobengröße insgesamt reduziert wird, oder aber verschiedene Statistiken auf der Basis unterschiedlicher Stichprobengrößen berechnet werden, erscheint es auf den ersten Blick vorteilhaft, jeglichen Ausschluß von Fällen dadurch zu umgehen, daß die fehlenden Werte geschätzt und jeweils durch einen Schätzwert ersetzt werden. Die wohl bekannteste und einfachste derartige Vorgehensweise ist die Ersetzung fehlender Werte durch den Stichprobenmittelwert, d.h. durch das arithmetische Mittel der Werte aller vorhandenen Werte in der jeweiligen Analysevariablen. Dieses Verfahren ist allerdings in mehrfacher Hinsicht problematisch: Sein Rationale beruht darauf, daß das arithmetische Stichprobenmittel einer Variablen X ein erwartungstreu-

er Schätzer des entsprechenden Populationsmittelwertes bzw. Erwartungswerts $E(X)$ ist, welcher wiederum der effektivste Schätzer für jeden individuellen Wert X_i aus der Populationsverteilung von X ist (solange keine zusätzliche spezifische Information über den zu schätzenden Wert verwendet wird). Damit aber ist auch klar, daß der Stichprobenmittelwert nur dann ein „guter“ Ersatz für die fehlenden X -Werte sein kann, wenn diese derselben Populationsverteilung entstammen, wie die vorhandenen bzw. wenn sich die Fälle mit fehlenden von denen mit vorhandenen X -Werten nicht systematisch unterscheiden. Damit ist wiederum die Möglichkeit gemeint, daß die Ausfälle nicht zufällig, sondern im Zusammenhang mit der Ausprägung des Merkmals, welches durch X gemessen wird, stattgefunden haben, wie es z.B. der Fall wäre, wenn Personen mit wachsender X -Ausprägung stärker dazu tendieren, die Antwort zu verweigern, so daß sozusagen $E(X)$ in der theoretischen Verteilung aller verweigerten Werte unter $E(X)$ in der theoretischen Verteilung aller nicht verweigerten Werte läge. Dann nämlich wäre der Stichprobenmittelwert der vorhandenen X -Werte sowohl ein verzerrter Schätzer jedes einzelnen nicht vorhandenen Meßwerts, als auch des Mittelwerts der gesamten Grundgesamtheit aller X -Werte. Ein weiteres Problem bei der Ersetzung fehlender Werte durch den Stichprobenmittelwert ist auch ohne entsprechende formale Ableitung offensichtlich: Das Verfahren reduziert die Stichprobenvarianzen bzw. Varianzschätzungen, z.B. im Vergleich zu denen, die bei paarweisem Ausschluß errechnet würden. In die Quadratsumme der Abweichungen vom Stichprobenmittelwert gehen dann sozusagen viele Nullsummanden mit ein, so daß diese Quadratsumme durch die Imputation der Werte nicht vergrößert wird, während jedoch Anzahl der Summanden und damit auch der scheinbaren Freiheitsgrade, durch die diese Quadratsumme zur Schätzung der Populationsvarianz zu teilen wäre, wächst. Die tatsächlichen Freiheitsgrade der Abweichungsquadratsumme bleiben natürlich durch die Imputation unberührt, aber dieses „weiß“ ein SEM-Computerprogramm nicht, wenn es über einen mittels Mittelwertsimputation behandelten Datensatz läuft, so daß ein negativer Bias bei den Varianzschätzungen im Grunde unvermeidbar ist. Entsprechendes gilt auch für Kovarianzen, so daß bei der Anwendung dieser Methode die Schätzung von Strukturgleichungsmodellen über die Modellanpassung an eine verzerrte Schätzung der Varianz-Kovarianz-Matrix der beobachteten Modellvariablen erfolgt. Dieses Problem allein wiegt schwer genug, um hier dem diesbezüglichen Fazit von Graham & Hofer (2000, 205) zuzustimmen: „This procedure should never be used.“

Ein weiteres bekanntes Verfahren der einfachen Ersetzung basiert auf der Regression der vorhandenen Werte einer Variablen X , deren fehlende Werte ersetzt werden sollen, auf andere Variablen im Datensatz (mit vorhandenen Werten), so daß dann mittels des so berechneten Regressionsmodells aus den Werten dieser Prädiktoren die fehlenden X -Werte geschätzt werden können. Dieses wird als „conditional mean imputation“ bezeichnet, da ja der so geschätzte Wert der auf der Basis des Regressionsmodells bedingte Mittelwert aller Fälle mit der jeweils gleichen Ausprägung in den Prädiktorvariablen ist. Letztendlich sind auch diese Mittelwertsschätzungen prinzipiell mit den Problemen verbunden, die für die Ersetzung durch den „unbedingten“ Stichprobenmittelwert skizziert wurden (wenn auch in

möglicherweise abgeschwächter Form), insbesondere hinsichtlich der Schätzung von Varianzen und Kovarianzen. Eine einleuchtende Verbesserung dieses Verfahrens der Ersetzung durch eine regressionsbasierte Schätzung besteht darin, zu dem so ermittelten Schätzwert einen zufällig gezogenen Residualwert aus der Fehlerverteilung des Regressionsmodells (also in der Regel aus einer Normalverteilung um den Mittelwert Null mit der geschätzten Fehlervarianz des Regressionsmodells) zu addieren. Mit dieser als „imputing from a conditional distribution“ bezeichneten Vorgehensweise kann die Verzerrung von Varianz- und Kovarianzschätzungen vermieden werden (vgl. Schafer & Graham, 2002, 159), allerdings ist die Durchführung dieses Verfahrens zumindest dann, wenn bei mehreren Variablen im zu analysierenden Datensatz fehlende Werte auftreten, genauso komplex, wie die im folgenden dargestellten neueren Verfahren, gegenüber denen es somit, angesichts der insgesamt überlegenen Qualitäten derselben, keinen Vorteil mehr böte.

Auch die sonstigen „älteren“ Methoden zum Umgang mit fehlenden Werten, die Schafer & Graham (2002) behandeln, sind gegenüber diesen neueren Verfahren prinzipiell im Nachteil, so daß in den neueren statistischen Abhandlungen nahezu einstimmig von der Verwendung der herkömmlichen einfachen Verfahren abgeraten wird bzw. als einziger Grund für diese Verwendung deren praktische Einfachheit bleibt, welcher dann zu rechtfertigen ist, wenn der zu behandelnde Datenausfall nur geringfügig oder im Sinne der im folgenden dargestellten Bedingung „missing completely at random“ ist (vgl. Schafer & Graham, 2002; Graham & Hofer, 2000; Verbeke & Molenberghs, 2000, 221-229).

2.3.1.2 Stochastische Mechanismen des Datenausfalls

Für die Entwicklung spezieller neuerer Verfahren zum Umgang mit fehlenden Werten war die von Rubin (1976) bzw. Little & Rubin (1987) eingeführte konzeptionelle und terminologische Unterscheidung von verschiedenartigen Mechanismen der Datenausfälle grundlegend – nämlich „missing completely at random“ (MCAR), „missing at random“ (MAR) und „missing not at random“ (MNAR).¹⁸ Bei dieser Unterscheidung wird das Fehlen von Werten als probabilistisches Phänomen betrachtet, welches durch entsprechende Zufallsvariablen dargestellt werden kann: Zu den zu analysierenden Variablen X_j ($j = 1, \dots, k$) werden die Variablen R_j definiert, die den Datenausfall bei den Messungen für X_j beschreiben (R_j sind also Kodiervariablen des Datenausfalls, z.B. mit $R_{ji} = 1$, wenn der Wert R_{ji} vorhanden ist und $R_{ji} = 0$, wenn dieser Wert fehlt). Die verschiedenartigen Mechanismen des Datenausfalls sind Unterscheidungen bezüglich der Wahrscheinlichkeitsverteilung der R_j . Schafer & Graham (2002) weisen auf das Mißverständnis hin, welches durch die dafür häufig gewählte Bezeichnung als „missingness mechanism“ (oder ähnlich) resultieren kann und in einer kausalen Interpretation dieser sogenannten Mechanismen besteht. Denn die

¹⁸Die Bezeichnungen variieren in der diesbezüglichen Fachliteratur: MAR wurde auch als „accessible missingness“ bezeichnet (z.B. Graham & Donaldson, 1993; Graham et al., 2001), MNAR als „missing informative“ (z.B. Diggle et al., 2000), „missing nonignorable“ (z.B. Wothke, 2000) oder „inaccessible missingness“ (z.B. Graham & Donaldson, 1993; Graham et al., 2001).

Unterscheidung bezieht sich nicht auf tatsächliche, realweltliche Ausfallursachen, sondern auf mathematische Eigenschaften der Wahrscheinlichkeiten des Fehlens von Werten. Diese können auf relativ einfache Weise dargestellt werden (vgl. Schafer & Graham, 2002, 151; vgl. auch die komplexeren formalen Definitionen z.B. bei Verbeke & Molenberghs, 2000, 215ff; eine daran angelehnte deutschsprachige und auf längsschnittliche „attrition“ bezogene Darstellung liefert Zimprich, 2002, 99ff): \mathbf{X}_i sei der vollständige Vektor der Werte eines Merkmalsträgers i (aus der untersuchten Stichprobe mit $i = 1, \dots, n$) in den Variablen X_j , welcher in die Vektoren der bei i beobachteten bzw. im Datensatz für i vorhandenen Werte $\mathbf{X}_i^{(v)}$ und der bei i nicht beobachteten bzw. im Datensatz für i fehlenden Werte $\mathbf{X}_i^{(f)}$ partitioniert werden kann, \mathbf{R}_i sei der Vektor der Werte des Merkmalsträgers i in den Variablen R_j . Dann bezeichnen die Bedingungen MCAR, MAR und MNAR Eigenschaften von $P(R_i|X_i)$, d.h. der bedingten Wahrscheinlichkeit von R_i (das Vorhandensein/Fehlen von X_j -Werten) bei gegebenen Werten X_i . Die Datenausfälle sind MAR, wenn gilt:

$$P(R_i|X_i) = P(R_i|X_i^{(v)}) \quad (2.56)$$

D.h. die Wahrscheinlichkeit für das Vorhandensein/Fehlen von X_j Werten hängt nicht von $X_i^{(f)}$, sondern lediglich von $X_i^{(v)}$ ab. Dieses bedeutet also, daß für jeden Merkmalsträger i die jeweiligen Wahrscheinlichkeiten für das Fehlen/Vorhandensein des Werts in den Variablen X_j lediglich von den Werten in denjenigen X_j -Variablen abhängen, die bei i auch beobachtet werden konnten und nicht von denjenigen X_j -Werten, die bei i fehlen. Es sind somit bei einem gegebenen Datensatz für n Merkmalsträger prinzipiell verschiedene Muster stochastischer Abhängigkeit des Datenausfalls möglich, die Gleichung 2.56 erfüllen, je nachdem, wieviele Muster vorhandener Werte es gibt. Bei z.B. $k = 3$ Variablen dürfte die Wahrscheinlichkeit für R_3 , also für einen Datenausfall in X_3 in der Subgruppe derjenigen Merkmalsträger mit $R_{1i} = 1, R_{2i} = 1$, d.h. mit vorhandenen X_1 - und X_2 -Werten, von X_1 und X_2 abhängen, in der Subgruppe aller Merkmalsträger mit $R_{1i} = 1, R_{2i} = 0$ dagegen nur von X_1 und in der Subgruppe mit $R_{1i} = 0, R_{2i} = 1$ nur von X_2 , entsprechendes müßte für die R_1 - und R_2 -Wahrscheinlichkeiten gelten.

Die MCAR-Bedingung ist demgegenüber einfacher zu verstehen:

$$P(R_i|X_i) = P(R_i) \quad (2.57)$$

Gleichung 2.57 bezeichnet schlicht die stochastische Unabhängigkeit von R_i und X_i (vgl. z.B. Schilling, 1998, 71). MCAR sind die Datenausfälle also dann, wenn ihre Wahrscheinlichkeit von keiner der Modellvariablen abhängt.

MNAR ist der Ausfall dann, wenn weder MCAR-, noch die MAR-Bedingungen erfüllt ist – d.h., wenn das Fehlen bzw. die Wahrscheinlichkeit eines Datenausfalls in einer Variablen X_j von X_j selbst bzw. wenn \mathbf{R}_i auch von $\mathbf{X}_i^{(f)}$ abhängt. Um Mißverständnissen vorzubeugen ist dazu anzumerken, daß der Zusammenhang von \mathbf{R}_i und $\mathbf{X}_i^{(f)}$ nach Ausparialisierung von $\mathbf{X}_i^{(v)}$ gemeint ist: Denn ansonsten würde, wenn die Variablen X_j unterein-

ander korreliert sind, jegliche stochastische Abhängigkeit des Datenausfalls von $\mathbf{X}_i^{(v)}$ auch automatisch eine Abhängigkeit von $\mathbf{X}_i^{(f)}$ nach sich ziehen, so daß die MAR-Bedingung nur dann erfüllt sein könnte, wenn die X_j -Variablen keinerlei Abhängigkeiten untereinander aufweisen. Dieses betrifft insbesondere auch den Datenausfall durch längsschnittliche „attrition“, wie er bei den folgenden Stabilitätsanalysen für die Mehrzahl aller Fälle mit unvollständigen Daten kennzeichnend ist: Eine hohe Korrelation der zeitlich aufeinanderfolgenden Zufriedenheitsvariablen ist ja die hypothetische Voraussetzung der Stabilitätsanalyse und wenn demnach die Zufriedenheit zu einem Zeitpunkt t hoch korreliert mit der Zufriedenheit zu einem späteren Zeitpunkt t' , so würde immer dann, wenn der Dropout zu t' mit der Zufriedenheit zu t zusammenhängt, dieser auch mit der Zufriedenheit zu t' zusammenhängen. MNAR wäre der Datenausfall hier jedoch nur dann, wenn ein Zusammenhang mit der Zufriedenheit zu t' auch nach der statistischen Auspartialisierung des Zusammenhangs mit der Zufriedenheit zu t noch besteht.

Fast alle der im Unterkapitel 2.3.1.1 aufgeführten Methoden zum Umgang mit fehlenden Werten führen lediglich unter der sehr restriktiven MCAR-Bedingung zu unverzerrten Parameterschätzungen (Schafer & Graham, 2002; Wothke, 2000; Little & Schenker, 1995). Die Ausnahme bildet hier lediglich das „imputation from a conditional distribution“, welches auch dann unverzerrte Schätzungen liefern kann, wenn die Datenausfälle lediglich MAR sind. Diese herkömmlichen Verfahren sind deshalb generell nur dann anwendbar, wenn entweder von MCAR-Datenausfällen ausgegangen werden kann, oder wenn damit nur eine sehr geringe Zahl von fehlenden Werten behandelt werden muß, denn natürlich wächst ein möglicher Bias mit der Anzahl der auszuschließenden Fälle oder zu ersetzenden Werte. Zu dieser Einschränkung ihrer Anwendbarkeit auf die MCAR-Bedingung kommt ein weiterer gravierender genereller Nachteil der Verfahren des Ausschlusses oder der einfachen Ersetzung, nämlich der ihrer Ineffizienz im Vergleich zu neueren Verfahren, die im folgenden behandelt werden. Es sind somit selbst dann, wenn die MCAR-Bedingung erfüllt ist, bei Ausschluß oder einfacher Ersetzung der fehlenden Werte Parameterschätzungen zu erwarten, die mit höheren Standardfehlern behaftet sind, als Schätzungen, die aus den neueren Verfahren resultieren. Sowohl diese Ineffektivität, wie auch die Verzerrungen von Schätzern unter MAR-Datenausfällen, konnte in einer Vielzahl von Simulationsstudien belegt werden (z.B. Schafer & Graham, 2002; Enders, 2001a; Enders & Bandalos, 2001; Wothke, 2000; Arbuckle, 1996; Brown, 1994; Muthén et al., 1987).

2.3.1.3 Neuere Methoden zur Behandlung fehlender Werte

Der eigentliche forschungspraktische Gewinn der konzeptionellen Unterscheidung der Datenausfallmechanismen MCAR, MAR und MNAR besteht darin, daß damit die strikte Dichotomie eines zufälligen versus systematischen Fehlens von Werten einer Datenmatrix aufgehoben und durch die Bedingung MAR gewissermaßen eine Zwischenstufe abgeschwächter Zufälligkeit des Datenausfalls definiert wurde, welche die Entwicklung neuerer,

an diese weniger restriktive Bedingung geknüpfter Verfahren zum Umgang mit den fehlenden Werten befördert hat. Es sind dies die folgenden Methoden (vgl. Schafer & Graham, 2002; Enders, 2001b; Graham & Hofer, 2000): Der „multiple-group approach“ (Allison, 1987; Muthén et al., 1987), der EM-Algorithmus („expectation maximization algorithm“) (Dempster et al., 1977), die FIML-Schätzung („full information maximum-likelihood estimation“) (Arbuckle, 1996) sowie die multiple Ersetzung („multiple imputation“) fehlender Datenwerte (Rubin, 1987). Es handelt sich bei den drei erstgenannten Verfahren um Maximum-Likelihood-Algorithmen, während die multiple Ersetzung eine „bayesianische“ Vorgehensweise bedeutet, d.h. es handelt sich um einen Algorithmus, der auf den Prinzipien der Bayes-Statistik beruht. Eine für den praktischen Anwender adressierte zusammenfassende und vergleichende Darstellung aller vier Verfahren liefert Enders (2001b), im folgenden wird lediglich die FIML-Schätzung genauer dargestellt, da m.E. diese das für die Anwendung in längsschnittlichen Strukturgleichungsmodellen attraktivste Verfahren ist. Die Verfahrenslogik der anderen genannten neueren Methoden zum Umgang mit fehlenden Werten wird unten nur knapp skizziert.

Der Wert dieser vier Methoden zum Umgang mit fehlenden Werten besteht, wie gesagt, in der Unverzerrtheit der aus ihnen resultierenden Parameterschätzungen bei Datenausfällen, die die im Vergleich zu MCAR weniger restriktive MAR-Bedingung erfüllen. Diese grundlegende Eigenschaft leitet sich aus den von Rubin (1976) formulierten Theoremen her, die häufig mit dem Begriff der „ignorability“ zitiert werden (z.B. Verbeke & Molenberghs, 2000, 217f; Diggle et al., 2002, 283f) und besagen, daß dann, wenn die Datenausfälle in den Modellvariablen eines statistischen Modells MCAR oder MAR sind, der Prozeß des Zustandekommens fehlender Werte bei Maximum-Likelihood-Parameterschätzungen für dieses Modell unberücksichtigt bleiben kann. Ausführliche formale Herleitungen dieses Theorems sind komplex (vgl. dazu auch Diggle et al., 2002, 283f; Verbeke & Molenberghs, 2000, 217f), um den Umfang der statistischen Darstellungen hier etwas zu begrenzen, wird die Bedeutung dieses Theorems sehr vereinfachend wie folgt erläutert: Für die Modellvariablen X_j wird ein durch die im Parametervektor θ zusammengefaßten Parameter definiertes statistisches Modell postuliert, d.h. es wird angenommen, daß die empirisch beobachteten Daten aus einer Population mit der Wahrscheinlichkeitsdichtefunktion $f(\mathbf{X}_i|\theta)$ stammen. Bei Datenausfällen im empirischen Datensatz stellen diese, wie oben bereits dargelegt, selbst wiederum Zufallsvariablen R_j dar, die sozusagen gemeinsam mit den X_j die empirische Datenmatrix bestimmen, so daß das statistische Modell um diese Datenausfälle erweitert werden muß: Das Submodell für die R_j , sei durch einen Parametervektor ψ (Parameter, die die Wahrscheinlichkeitsverteilung der Datenausfälle beschreiben) definiert, so daß das um den Datenausfall erweiterte Modell durch die Wahrscheinlichkeitsdichtefunktion $f(\mathbf{X}_i, \mathbf{R}_i|\theta, \psi)$ beschrieben werden kann. Die oben aufgeführten neueren Maximum-Likelihood-Methoden zum Umgang mit fehlenden Werten beruhen auf der Maximierung der Wahrscheinlichkeit der tatsächlich *beobachteten* Daten, d.h. der Maximierung der Likelihoodfunktion $\mathcal{L}(\theta, \psi|\mathbf{X}_i^{(v)}, \mathbf{R}_i)$ (da ja $\mathbf{X}_i^{(f)}$ nicht beobachtet wird).

Rubin (1976) hat nun gezeigt, daß dann, wenn der Datenausfall MCAR oder MAR ist, $\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\psi} | \mathbf{X}_i^{(v)}, \mathbf{R}_i)$ proportional zur Likelihood $\mathcal{L}(\boldsymbol{\theta} | \mathbf{X}_i^{(v)}, \mathbf{R}_i)$ ist – letzteres ist die Likelihoodfunktion für die Daten ohne die Parameter $\boldsymbol{\psi}$, d.h. unter Ignorierung des Prozesses des Zustandekommens der Datenausfälle.¹⁹ Dies bedeutet, daß dasjenige $\hat{\boldsymbol{\theta}}$, welches $\mathcal{L}(\boldsymbol{\theta} | \mathbf{X}_i^{(v)}, \mathbf{R}_i)$ maximiert, auch die $\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\psi} | \mathbf{X}_i^{(v)}, \mathbf{R}_i)$ maximiert, so daß die Schätzung der Parameter $\boldsymbol{\theta}$ unter Ignorierung des Prozesses des Zustandekommens der fehlenden Werte zu demselben Ergebnis führt, wie unter Einbeziehung desselben. Damit ist letztendlich gewährleistet, daß bei MAR-Datenausfällen Maximum-Likelihood-Parameterschätzungen mit den üblichen asymptotischen Qualitäten (Erwartungstreue und Effizienz) möglich sind.

Der von Muthén et al. (1987) bzw. Allison (1987) vorgestellte „multiple-group approach“ besteht darin, die Stichprobe in Substichproben solcher Fälle zu unterteilen, die jeweils dasselbe Muster vorhandener und fehlender Werte in den Variablen des Modells aufweisen: Mit diesen Substichproben könnte das Modell dann als sogenanntes Multisample-Modell gerechnet werden, wobei jedes Substichprobenmodell gewissermaßen um diejenigen Teile reduziert wird, für die in dieser Substichprobe die Werte fehlen. Für die zu schätzenden Modellparameter werden Gleichheitsrestriktionen zwischen den Gruppen (in deren Subgruppenmodell diese Parameter noch enthalten sind) festgesetzt. Diese Methode ist, wie z.B. Bentler (1995, 197) anmerkt, nur dann praktikabel, wenn es nur eine kleine Anzahl solcher Subgruppen mit gleichem Muster fehlender Werte gibt: „If there are dozens of patterns of missing data, with only a few subjects showing a given pattern of missing data, the modeling approach is useless because some of these samples may be too small to yield stable results . . . , and the method may be too computationally demanding to work with so many samples.“ Genau dieses ist aber der Fall bei den SOEP-Zufriedenheitsmessungen über alle 16 Erhebungswellen, so daß diese Vorgehensweise hier nicht anwendbar ist.

Der EM-Algorithmus (Dempster et al., 1977; siehe auch Verbeke & Molenberghs, 2000, 387-390; Enders, 2001b) kann als iterative Durchführung eines zweistufigen Vorgehens beschrieben werden: Gegeben sei ein durch den Parametervektor $\boldsymbol{\theta}$ definiertes statistisches Modell für die Analysevariablen (z.B: Regressionsmodell, SEM). Dann werden im ersten Schritt – dem sogenannten „E-step“ – die fehlenden Werte eines Falls i aus den für i vorhandenen und mittels der jeweils aktuellen Parameterschätzungen $\hat{\boldsymbol{\theta}}$ geschätzten Werte ersetzt (dies entspricht einer „conditional means imputation“, vgl. oben). Aus der so vervollständigten Datenmatrix werden dann im zweiten Schritt – dem sogenannten „M-step“ – neue Maximum-Likelihood-Schätzungen $\hat{\boldsymbol{\theta}}$ errechnet, welche die Basis für den nächsten Durchlauf der beiden Schritte bilden. Diese Prozedur wird solange wiederholt, bis bei den aufeinanderfolgenden Durchläufen die modellimplizierten Kovarianzmatrizen der Variablen sich gemäß eines vorab spezifizierten Konvergenzkriteriums nicht mehr bedeutsam verändern. Ein Nachteil dieses Verfahrens besteht in der dadurch unvermeidbaren

¹⁹Rubin (1976) zeigt auch in ähnlicher Weise die „ignorability“ von MAR-Datenausfällen für bayesstatische Inferenzen, so daß auch die Methode der multiplen Ersetzung äquivalent zu den (direkten) Maximum-Likelihood-Verfahren an die weniger restriktive MAR-Bedingung geknüpft werden kann.

Unterschätzung der Standardfehler für die Modellparameter, so daß daraus verzerrte Signifikanztests resultieren würden. Die Standardfehler können allerdings ad hoc durch sogenannte Resampling-Methoden (Bootstrap) geschätzt werden.

Das Verfahren der multiplen Ersetzung der fehlenden Werte wurde erstmals von Rubin (1987) vorgeschlagen und ist grob dadurch zu beschreiben, daß dabei anstatt eines einzelnen durch die Ersetzung fehlender Werte vervollständigten Datensatzes mehrere solcher vervollständigten Datensätze erzeugt werden, für die dann jeweils die Modellparameter θ mittels konventioneller Methoden für vollständige Daten geschätzt werden. Bei $m > 1$ durch jeweils einfache Imputation fehlender Werte vervollständigten Datensätzen resultieren also m Vektoren von Parameterschätzungen $\hat{\theta}_j$ ($j = 1, \dots, m$). Diese werden dann zu einer letztendlichen Schätzung $\hat{\theta}$ kombiniert. Die komplexen Details des Verfahrens betreffen die Generierung der zu imputierenden Werte und die Kombination der multiplen Parameterschätzungen – dazu sei auf ausführlichere Darstellungen z.B. von Schafer & Olsen (1998) verwiesen.

Ein direktes Maximum-Likelihood-Verfahren zur Schätzung von Strukturgleichungsmodellparametern beim Vorliegen fehlender Werte in der empirischen Datenmatrix der beobachteten Variablen ist der von Arbuckle (1996) vorgelegte FIML-Schätzalgorithmus. Die Bezeichnung „full information maximum likelihood“ bezieht sich auf den Umstand, daß mit diesem Verfahren alle in der empirischen Datenmatrix vorhandenen Daten sozusagen unvermittelt zur Schätzung der Parameter genutzt werden. Dieses geschieht dadurch, daß zunächst die Likelihood der (vorhandenen) Daten *fallweise* bestimmt und dann die Summe der logarithmierten fallweisen Likelihoods maximiert wird. Das Verfahren vollzieht also die folgenden Berechnungsschritte (vgl. Arbuckle, 1996, 248):

1. Unter der Annahme multivariater Normalverteilung der beobachteten Modellvariablen X_j ist die logarithmierte Likelihood für die Daten des i -ten Falles:

$$\log \mathcal{L}_i = K_i - \frac{1}{2} \log |\Sigma_i| - \frac{1}{2} (\mathbf{X}_i^{(v)} - \boldsymbol{\mu}_i)' \Sigma_i^{-1} (\mathbf{X}_i^{(v)} - \boldsymbol{\mu}_i) \quad (2.58)$$

Hierbei ist K_i eine für den Fall i spezifische Konstante die nur von der Anzahl der im Fall i vorhandenen Werte abhängt. $\mathbf{X}_i^{(v)}$ ist, wie bereits dargelegt, der Vektor der im Fall i vorhandenen Werte der Variablen X_j . Σ_i ist die Varianz-Kovarianz-Matrix derjenigen Variablen X_j , deren Werte im Fall i vorhanden sind, d.h. also diejenige Submatrix der allgemeinen Varianz-Kovarianz-Matrix aller beobachteten Variablen Σ , die nur die Varianzen bzw. Kovarianzen der in $\mathbf{X}_i^{(v)}$ vorhandenen Variablen X_j enthält. Entsprechend bezeichnet $\boldsymbol{\mu}_i$ den Vektor der Mittelwerte derjenigen Variablen X_j , deren Werte im Fall i vorhanden sind, also einen Subvektor von $\boldsymbol{\mu}$, dem Vektor der Mittelwerte von X_j . $|\dots|$ bezeichnet die Determinante einer Matrix.

2. Die logarithmierte Likelihood der Daten aller n Fälle der untersuchten Stichprobe ist dann die Summe der fallweise logarithmierten Likelihoods 2.58:

$$\log \mathcal{L}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \sum_{i=1}^n \log \mathcal{L}_i \quad (2.59)$$

3. Die Vorgehensweise bei der SEM-Parameterschätzung besteht dann darin, daß wie auch im „konventionellen“ Fall ohne fehlende Werte $\boldsymbol{\Sigma}$ und $\boldsymbol{\mu}$ als Funktionen eines Vektors der Modellparameter $\boldsymbol{\theta}$ – also $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$, $\boldsymbol{\mu} = \boldsymbol{\mu}(\boldsymbol{\theta})$ – spezifiziert und die durch 2.59 spezifizierte logarithmierte Likelihood $\log \mathcal{L}(\boldsymbol{\mu}(\boldsymbol{\theta}), \boldsymbol{\Sigma}(\boldsymbol{\theta}))$ durch die Auswahl der Schätzwerte $\hat{\boldsymbol{\theta}}$ maximiert wird. Dieses geschieht durch Minimierung der folgenden Funktion:

$$\begin{aligned} C(\boldsymbol{\theta}) &= -2 \log \mathcal{L}(\boldsymbol{\mu}(\boldsymbol{\theta}), \boldsymbol{\Sigma}(\boldsymbol{\theta})) + 2 \sum_{i=1}^n K_i \\ &= \sum_{i=1}^n \log |\boldsymbol{\Sigma}_i(\boldsymbol{\theta})| - \sum_{i=1}^n (\mathbf{X}_i^{(v)} - \boldsymbol{\mu}_i(\boldsymbol{\theta}))' \boldsymbol{\Sigma}_i(\boldsymbol{\theta})^{-1} (\mathbf{X}_i^{(v)} - \boldsymbol{\mu}_i(\boldsymbol{\theta})) \quad (2.60) \end{aligned}$$

Das FIML-Verfahren erlaubt also für Strukturgleichungsmodelle bei fehlenden Werten der beobachteten Modellvariablen eine Parameterschätzung mit den üblichen Gütekriterien der Maximum-Likelihood-Schätzer (asymptotische Erwartungstreue, Konsistenz, asymptotische Effizienz – vgl. Bollen, 1989, 108) dann, wenn der Datenausfall MCAR oder MAR ist. Gegenüber den obengenannten herkömmlichen älteren Methoden zum Umgang mit fehlenden Werten bedeutet dieses eine generelle Überlegenheit der FIML-Schätzung, die in einer Vielzahl von Studien belegt werden konnte (Arbuckle, 1996; Wothke, 2000; Enders, 2001a; Enders & Bandalos, 2001): Es wurde bereits darauf hingewiesen, daß die Ausschlußverfahren und die (meisten) Verfahren der einfachen Ersetzung fehlender Werte unter MAR-Bedingungen zu verzerrten und ineffizienten Schätzungen führen und auch unter MCAR-Bedingungen sich als ineffizient erwiesen – auch wurde auf weitere Schwierigkeiten dieser Methoden hingewiesen, so daß im Vergleich dazu die FIML-Schätzung von Strukturgleichungsmodellen mit unvollständigen Daten auf jeden Fall vorzuziehen ist.

Zudem zeigten in verschiedenen Studien Maximum-Likelihood-Schätzungen unter simulierten MNAR-Datenausfällen einen deutlich geringeren Bias, als die Resultate bei fall- oder paarweisem Ausschluß (Muthén et al., 1987; Little & Rubin, 1989; Schafer, 1997;), was als Hinweis darauf gewertet werden kann, daß zumindest auch unter moderaten Verletzungen der MAR-Bedingung Verzerrungen der FIML-Schätzungen tolerierbar sein dürften. Dieses erscheint in pragmatischer Hinsicht als ein durchaus relevanter Aspekt, denn natürlich stellt sich im konkreten Anwendungsfall der oben aufgeführten Methoden stets auch die Frage, ob die Datenausfälle überhaupt noch die MAR-Voraussetzung erfüllen bzw. wie Verzerrungen infolge einer irrtümlichen Annahme dieser Bedingung zu vermeiden sind.

In jüngster Zeit wurden Verfahren zum Umgang mit MNAR-Datenausfällen entwickelt, nämlich die „selection models“ (siehe z.B. Verbeke & Molenberghs, 2000, 231ff) und die „pattern-mixture models“ (siehe z.B. Verbeke & Molenberghs, 2000, 275ff). Für die An-

wendung mit den vorne dargestellten längsschnittlichen Strukturgleichungsmodellen scheinen diese Methoden allerdings noch kaum geeignet, weshalb sie hier – um den Umfang der methodologischen Ausführungen zumindest ein wenig zu begrenzen – nicht eingehender dargestellt werden: Für beide Methoden fehlen schlicht Implementierungen des jeweiligen Maximum-Likelihood-Schätzalgorithmus in den gegenwärtig erhältlichen SEM-Softwarepaketen. Zudem sind die „selection models“ mit dem Nachteil behaftet, daß sie sehr stark an nicht testbare Verteilungsannahmen bezüglich der Analysevariablen und der Datenausfälle gebunden sind, so daß z.B. Schafer & Graham (2002, 172) anmerken: „Many consider these models to be too unstable for scientific applications and to be more useful for raising questions than generating answers“ Dieser Nachteil betrifft auch die mögliche statistische Testung der MAR-Bedingung, die auf derselben Modellierung und Faktorisierung der zusammengefaßten Dichtefunktion der Variablen X_j und der Datenausfälle R_j beruhen, wie die „selection models“ und deshalb extrem sensibel gegenüber auch leichten Abweichungen von zugrundegelegten Verteilungsannahmen sind (siehe z.B. Kenward, 1998). Bezüglich der „pattern-mixture models“ ist anzumerken, daß diese zwar voraussetzungsärmer und eventuell auch mit Standard-SEM-Software durch die Setzung komplexer Restriktionen anwendbar wären, jedoch im vorliegenden Falle einer längsschnittlichen Analyse über 16 Meßzeitpunkte schon allein aus demselben Grunde problematisch wären, wie der oben genannte „multiple-group approach“, zu dem das Verfahren eine gewisse Verwandtschaft aufweist: Es ergeben sich bei so vielen Meßzeitpunkten zuviele und zu kleine Gruppen von Merkmalsträgern mit jeweils demselben Muster des Datenausfalls. Es scheint somit der Test der MAR-Bedingung wenig sinnvoll und im vorliegenden Fall geboten, die Anwendung von Verfahren für MNAR-Datenausfälle möglichst von vorneherein auszuschließen bzw. a priori eine Entscheidung für die Anwendung der FIML-Schätzung zu treffen, so daß hier eher die Frage zu stellen ist, wodurch bzw. unter welchen Umständen letztere gerechtfertigt ist.

Eine solche Rechtfertigung kann aus verschiedenen Gründen abgeleitet werden. Zunächst aus den erwähnten Hinweisen darauf, daß auch bei MNAR-Datenausfällen die Verzerrungen der FIML-Schätzungen gering ausfallen. Dieses dürfte wohl auch vom Ausmaß der Nichtzufälligkeit des Fehlens von Werten abhängen: MNAR bedeutet, wie oben dargelegt, stochastische Abhängigkeit des Vorhandenseins/Fehlens der Werte einer Variablen X_j von eben diesen Werten, diese Abhängigkeit kann aber mehr oder weniger stark ausgeprägt sein und es ist hier zu fragen, wie stark sie sein muß, um zu ernsthaften Verzerrungen bei Analysemethoden zu führen, die bei MAR zu unverzerrten Schätzungen führen. Zu dieser Frage ist m.E. nur wenig publiziert, Schafer & Graham (2002) argumentieren, daß dann, wenn die Variable mit den fehlenden Werten nicht selbst die Ursache der Datenausfälle darstellt, sondern mit dieser Ursache lediglich moderat korreliert, ein vernachlässigbarer Bias aus den auf die MAR-Bedingung begründeten Methoden resultieren sollte, und sie bezeichnen einen Korrelationswert von 0.4 noch als solchermaßen moderat. Außerdem ist daran zu erinnern, daß MNAR die bloß stochastische (und nicht kausale) Abhängigkeit

von X_j und R_j nach Auspartialisierung der anderen Modellvariablen meint: „One way to establish MAR processes is to include completely observed variables \mathbf{X} that are highly predictive of Y [meint die Variable mit dem Datenausfall, O.S.]. For instance, inasmuch as past behavior is an effective predictor of future behavior, initial (complete) measurements in longitudinal designs can be a good choice of \mathbf{X} “ (Wothke, 2000, 225). Es liegt auf der Hand, daß dieses im vorliegenden Fall der SOEP-Zufriedenheitswerte zumindest für all die Datenausfälle durch „attrition“ dann zutrifft, wenn Zufriedenheit in hohem Maße stabil ist.

Ein großer Teil der Dropouts von Zufriedenheitswerten v.a. auch der Hochaltrigen im SOEP ist durch den Tod von Untersuchungsteilnehmern verursacht und diese Fälle sind hinsichtlich der Frage der Gültigkeit der MAR-Bedingung speziell zu beurteilen. Denn wenn eine Versuchsperson vor einem Meßzeitpunkt t stirbt, so existiert für diese Person kein X_t -Wert, dem Fehlen dieses Wertes in der Datenmatrix kann nicht die Bedeutung unterlegt werden, wie in den obengenannten Definitionen von MCAR, MAR und MNAR, bei denen davon ausgegangen wurde, daß die fehlenden Werte tatsächlich existent sind und nur infolge eines wie auch immer bewirkten Datenausfalls nicht erhoben bzw. beobachtet wurden. Schafer & Graham (2002, 155) weisen darauf hin, daß generell das Fehlen von Werten in der Datenmatrix infolge tatsächlicher Nichtexistenz dieser Werte eine Situation darstellt, in der die Methoden, für die die „ignorability“ von MAR-Datenausfällen gilt, angewendet bzw. in der die Ausfälle so, als ob sie MAR wären, behandelt werden können. Dieses schließt prinzipiell die mortalitätsbedingten Dropouts mit ein, wobei hier allerdings einschränkend zu berücksichtigen ist, daß bei größeren Zeitintervallen zwischen den Befragungen möglicherweise rapide Verluste im physischen und kognitiven Status einer Person unmittelbar vor dem Tod unentdeckt bleiben und deshalb nicht adäquat in die Schätzung von Kennwerten längsschnittlicher Veränderung einfließen. Übertragen auf die längsschnittliche Analyse von Zufriedenheiten würde dieses bedeuten, daß z.B. geschätzte Stabilitätskoeffizienten gewissermaßen nur die „normale“ Stabilität des Merkmals anzeigen, wie sie *bis zu* solchen Sterbephasen besteht.

Es sind damit mehrere Aspekte genannt, welche zusammengenommen m.E. ausreichende Begründung dafür geben, die fehlenden Zufriedenheitswerte in den Modellen zur Stabilitätsanalyse a priori als solche zu behandeln, die MAR sind. Dieses geschähe dann sozusagen ganz im Einklang mit dem Fazit von Schafer & Graham (2002), welche die Anwendung der MAR-basierten Methoden zum Umgang mit fehlenden Werten als den für die meisten konkreten Anwendungsfälle angemessenen „state of the art“ benennen (zumindest solange MNAR-Methoden noch nicht in Standardsoftwareprogrammen implementiert sind). Für die hier angestrebten Stabilitätsanalysen vermittelt der Berechnung von Strukturgleichungsmodellen erscheint deshalb die FIML-Schätzung erste Wahl, zumal für diesen Algorithmus die infolge seiner Implementierung in aktuellen Versionen verschiedener SEM-Softwarepakete vergleichsweise einfache Anwendbarkeit spricht.

2.3.1.4 Zur Anwendbarkeit des FIML-Schätzalgorithmus bei nichtnormalen Daten

Die große Attraktivität des FIML-Verfahrens kontrastiert allerdings auch mit einer gewissen Problematik infolge der dabei – wie auch bei den anderen der oben dargestellten MAR-basierten Verfahren zum Umgang mit fehlenden Werten – für die Spezifizierung der Likelihoodfunktion zugrundegelegte Annahme multivariater Normalverteilung der beobachteten Modellvariablen. Konkret bestehen bei dem FIML-Verfahren einige Beschränkungen sowohl hinsichtlich der Überprüfung der Normalverteilungsannahme, als auch hinsichtlich der Verfahrensmöglichkeiten bei Verletzungen derselben. Die Überprüfung der Annahme multivariater Normalverteilung mittels der statistischen Tests multivariater Schiefe und Kurtosis von Mardia (1985; siehe auch die Darstellungen bei Bollen, 1989, 423 und Bentler, 1995, 214), die in den meisten gegenwärtig gebräuchlichen SEM-Programmpaketen angeboten wird, ist auf vollständige Datensätze ohne fehlende Werte beschränkt, sie könnte also im hier vorliegenden Falle jeweils nur für die durch fallweisen Ausschluß reduzierten Stichproben derjenigen Befragten, die bei allen 16 Erhebungen ein Zufriedenheitsurteil abgaben, durchgeführt werden und wäre somit angesichts einer erheblichen Anzahl von Dropouts gewissermaßen ein Muster von geringem Wert. Eine Alternative zur Testung der multivariaten Verteilungsannahme besteht in der jeweils univariaten Testung aller manifesten Modellvariablen auf Normalverteilung: Univariate Normalverteilung jeder Variablen ist eine notwendige Voraussetzung der multivariaten Verteilung aller Variablen, die jedoch nicht hinreichend ist, so daß die univariate Normalverteilung aller Variablen strenggenommen lediglich als starkes Indiz dafür gewertet werden könnte, daß die Annahme multivariater Normalverteilung nicht verletzt sein dürfte.

Im Falle der Verletzung der Normalverteilungsannahmen kann es bei „konventioneller“ Maximum-Likelihood-Schätzung von Strukturgleichungsmodellen (bei vollständiger Datenmatrix) – und dieses ist wohl auch auf die FIML-Schätzung übertragbar – zu Verfälschungen in den involvierten Signifikanztests kommen: Einfach ausgedrückt, ist dann die Variation der beobachteten Modellvariablen nicht mehr vollständig in der Stichprobenkovarianzmatrix „zusammengefaßt“, so daß zwar die Parameterschätzungen erwartungstreu bleiben, jedoch mit größeren Standardfehlern behaftet sind, als es aufgrund der reduzierten Variation in der Kovarianzmatrix erscheint. Eine Folge ist eine gewisse Liberalisierung der Signifikanztests einzelner Modellparameter (v.a. Regressionskoeffizienten), so daß „falsche Signifikanzen“ bzw. Unterschätzungen der tatsächlichen α -Fehlerwahrscheinlichkeit bei Zurückweisung der Nullhypothese möglich werden. Eine weitere Folge ist die Verzerrung des χ^2 -Tests der Modellanpassung – hier kann ein zu hoher Diskrepanzwert (χ^2 -Wert) zu fälschlicher Annahme der H_1 bzw. der Schlußfolgerung eines unzureichenden Model-Fit führen. Von letzterer Verzerrung sind dann auch die üblichen Fit-Indices, deren Berechnung ja jeweils auf die eine oder andere Art und Weise auf dem Diskrepanzwert basiert, betroffen, so daß generell bei Nichtnormalität der beobachteten Modellvariablen davon auszugehen ist, daß die vermittels der Maximum-Likelihood-Schätzung erzielten Model-Fit-Werte die tatsächliche Modellanpassung unterschätzen.

Es wurde eine Reihe von Vorgehensweisen zur Vermeidung oder Korrektur dieser Verzerrungen infolge nichtnormaler Daten vorgeschlagen, beispielsweise eine vorherige „normalisierende“ Transformation der Daten, die Berechnung von Bootstrap-Konfidenzintervallen, die Anwendung des asymptotisch verteilungsfreien Schätzers (ADF-Schätzer) von Browne (1984) oder der Scaled- χ^2 - bzw. Satorra-Bentler-Statistik und der darauf begründeten Berechnung robuster Standardfehler (siehe z.B. Bentler, 1995, 217f). Die Beantwortung der Frage, wann welche dieser Vorgehensweisen angewandt werden sollte, hängt v.a. von Art und Ausmaß der Verletzung der Normalverteilungsannahme, sowie von der Stichprobengröße ab – eine m.E. sehr gute Zusammenfassung des komplexen diesbezüglichen Forschungsstandes leisten West et al. (1995). Hinsichtlich der hier angestrebten Stabilitätsanalysen und angesichts der oben diskutierten Erfordernisse des Umgangs mit fehlenden Werten erübrigen sich jedenfalls von vornherein die Anwendung des ADF-Schätzers wie auch der Satorra-Bentler-Statistik, da diese Verfahren wiederum auf vollständige Datensätze ohne fehlende Werte beschränkt sind und also im vorliegenden Fall die inadäquate Methode des fallweisen Ausschlusses fehlender Werte voraussetzen würden.

Ein spezieller Fall von Nichtnormalität, der hier Beachtung verdient, resultiert aus der Verwendung nichtkontinuierlicher, ordinalskalierten Variablenwerte: Um solche handelt es sich genaugenommen bei den häufig – und auch im vorliegenden Fall – verwendeten Ratingskalen mit einer begrenzten Anzahl möglicher Antwortkategorien, die als ganzzahlige Skalenwerte kodiert und verrechnet werden und denen im Grunde nur „per fiat“ metrische Qualität unterstellt wird. Da es sich bei der Normalverteilung um ein *stetiges* Verteilungsmodell handelt, kann allein schon aufgrund der begrenzten Anzahl von Antwortkategorien die Verteilung solchermaßen skalierten Variablen zwangsläufig nur näherungsweise normal sein. Man könnte aber gegebenenfalls annehmen, daß es sich hierbei gewissermaßen nur um eine Grobheit des Meßinstruments handelt, welches eben nur eine begrenzte Anzahl diskreter Skalenwerte erlaubt, während bei Verwendung eines „unendlich feinen“ Meßinstruments bzw. einer „beliebig fein unterteilten“ Meßskala tatsächlich eine Normalverteilung der Meßwerte vorläge. Bollen (1989) spricht hier von „latent continuous indicators“ die den kategorialen Messungen zugrundeliegen, und häufig werden darum solche nichtkontinuierlichen ordinalen Variablen als grobkategorisierte Variablen („coarsely categorized variables“, z.B. West et al., 1995) bezeichnet. Es ist dann davon auszugehen, daß Kovarianzen zwischen oder mit diesen grobkategorisierten Meßvariablen geringer sind, als zwischen oder mit den entsprechenden zugrundeliegenden latenten kontinuierlichen Meßvariablen, und dieses ist deshalb von besonderem Belang, weil ja die Berechnung eines Strukturgleichungsmodells durch die Anpassung der modellimpliziten an die empirisch geschätzte Varianz-Kovarianz-Matrix der beobachteten Variablen erfolgt. Hier wäre nun aber die modellimplizierte Varianz-Kovarianz-Matrix die der latenten kontinuierlichen Meßvariablen, während die empirisch geschätzte die der grobkategorisierten Meßvariablen wäre, so daß sozusagen das Modell nicht an dasjenige angepaßt wird, woran es angepaßt werden soll und ein Misfit gewissermaßen vorprogrammiert ist. Eine weitere wahrscheinliche Konse-

quenz der Grobkategorisierung besteht darin, daß sich die Verteilungen von kategorisierter und kontinuierlicher Meßvariablen unterscheiden, so daß möglicherweise auch dann, wenn die latente kontinuierliche Meßvariable normalverteilt ist, dieses für die kategorisierte nicht zutrifft. Grobkategorisierte Variablen begründen also sozusagen prinzipiell einen Verdacht auf Verletzung der Normalverteilungsannahme und daraus resultierender Verzerrung der Ergebnisse.

Als Möglichkeit, den spezifischen Problemen nichtkontinuierlicher ordinaler beobachteter Variablen in Strukturgleichungsmodellen adäquat Rechnung zu tragen, wurde die Verwendung polychorischer und polyserieller Korrelationen vorgeschlagen, welche in der Korrelationsmatrix der beobachteten Variablen die Produkt-Moment-Korrelationen zwischen bzw. mit den kategorialen Variablen ersetzen, so daß dann die Modellanpassung der modellimplizierten an diese empirische *Korrelationsmatrix* durchgeführt werden kann (z.B. Lee et al., 1995; 1992; Muthén, 1984). Wiederum sehr vereinfacht dargestellt, bedeutet dieses sozusagen ein zweistufiges Vorgehen, bei dem die Beziehungen zwischen den kategorialen und ihren angenommenen zugrundeliegenden kontinuierlichen Meßvariablen und daraus dann die Assoziationen der kontinuierlichen Meßvariablen untereinander geschätzt werden. Abgesehen von auch gegenwärtig noch bestehenden Beschränkungen, die der Anwendung solcher Vorgehensweisen durch deren Implementation in vorhandene Softwarepakete (z.B. auf eine begrenzte Anzahl möglicher kategorialer Variablen) und die Speicherkapazitäten der üblicherweise verfügbaren Computer gesetzt sind,²⁰ sind sie für die hier angestrebten Analysen wiederum deshalb nicht anwendbar, weil sie im Grunde keines der MAR-basierten Verfahren zum Umgang mit fehlenden Werten erlauben: Zwar ist mit einigen der vorhandenen Softwarepakete (z.B. EQS 5.7) die Verknüpfung dieser Methode mit dem „multiple-group approach“ prinzipiell möglich, dieser aber ist, wie oben bereits dargelegt, im vorliegenden Fall nicht brauchbar. Eine der SEM-Analyse vorangestellte multiple Ersetzung der fehlenden Werte würde hier eine gewisse Zirkularität beinhalten, da bei dieser Vorgehensweise wiederum multivariate Normalverteilung vorauszusetzen wäre. Die Implementation dieses Schätzalgorithmus in den derzeit verfügbaren Softwarepaketen (EQS 5.7, LISREL 8.51) beschränkt die Anwendung wiederum auf Fälle mit vollständigen Werten in allen beobachteten Modellvariablen.

Angesichts der soeben aufgezählten Probleme, die für die Anwendung von Maximum-Likelihood-Schätzalgorithmen bei im Grunde nichtkontinuierlichen ordinalen Zufriedenheitsmessungen und erwartbaren Verletzungen der Normalverteilungsannahme zu berücksichtigen sind, stellt sich die Frage nach den möglichen *praktischen* Konsequenzen bzw. hier insbesondere nach der Robustheit des FIML-Schätzverfahrens unter den behandelten Voraussetzungssverletzungen.

²⁰Beispielsweise scheiterte eine probeweise versuchte Modellrechnung für das QMSM der Werte allgemeiner Lebenszufriedenheit (Stichprobe: Alle SOEP Teilnehmer mit vollständigen Werten, d.h. listwise deletion) mittels des entsprechenden „categorical data feature“ des Programms EQS 5.7, auch bei Beschränkung auf lediglich 4 Meßzeitpunkte daran, daß die Speicherkapazitäten des dazu verwendeten PCs erschöpft wurden.

Es existiert eine Vielzahl diesbezüglicher Untersuchungen zur Robustheit der „konventionellen“ Maximum-Likelihood-Schätzung von Strukturgleichungsmodellen bei Nichtnormalität der Daten, aus denen v.a. dank der Zusammenfassung von West et al. (1995) die folgenden beiden anwendungsrelevanten Gesichtspunkte abgeleitet werden können (vgl. auch Raykov & Widaman, 1995; Bollen, 1989, 415ff und 433ff):

- ▶ Die spezifische Problematik der kategorialen Messungen wird umso dringlicher, je weniger Ausprägungsmöglichkeiten vorhanden sind, und umso mehr vernachlässigbar, je größer die Anzahl möglicher Skalenwerte ist. Auch wenn kein allgemein akzeptierter Grenzwert angegeben wurde, so läßt sich aus den zitierten Veröffentlichungen ableiten, daß v.a. bei weniger als fünf Ausprägungskategorien eine explizite Berücksichtigung des Problems indiziert ist – und daß es für die immerhin 11-stufige Zufriedenheitsskala, die im SOEP verwendet wurde, im Grunde vernachlässigbar ist, so daß allein aus dem Umstand, daß solche Ratingskalen benutzt wurden, die Durchführung einer Maximum-Likelihood-Schätzung nicht in Frage gestellt ist.
- ▶ Die Problematik der Nichtnormalität von Daten wächst mit dem Ausmaß der Abweichung von der Normalität und ist v.a. bei kleinen Stichproben bedeutsam. Auch hierfür existieren keine allgemein akzeptierten Grenzwerte z.B. dafür, welche Abweichungen von der Normalität (in Schiefe und Exzeß) noch unbedenklich und welches kleine Stichprobengrößen sind, aber immerhin kann wohl zusammengefaßt werden, daß moderate Abweichungen von der Normalität einigermaßen unbedenklich erscheinen, zumal ihnen ja auch dadurch Rechnung getragen werden kann, daß ihre möglichen Konsequenzen bei der Interpretation der Ergebnisse berücksichtigt werden. So könnte z.B. für die Tests der Pfadkoeffizienten – deren Schätzwerte, wie bereits gesagt, auch bei Nichtnormalität „unbiased“ sind – ein strengeres Signifikanzniveau festgelegt werden, während hinsichtlich der Model-Fit-Kennwerte Nichtnormalität sich gewissermaßen als Verschärfung der Beurteilungskriterien (Unterschätzung der tatsächlichen Anpassungsgüte) auswirken würde. Es erscheint deshalb als durchaus rational zu rechtfertigende Anwendungsregel, Maximum-Likelihood-Schätzungen solange für praktikabel zu halten, wie in den univariaten Verteilungen der beobachteten Variablen keine extremen Schiefe- bzw. Kurtosiswerte vorhanden sind – beispielsweise bezeichnen West et al. (1995) einen Schiefewert ≥ 2 und Kurtosis ≥ 7 als substantielle Abweichung von der Normalität.

Zur speziellen Robustheit des FIML-Schätzers bei Verletzungen der Normalverteilungsannahme wurde m.E. bis zum gegenwärtigen Zeitpunkt lediglich eine Untersuchung publiziert: Enders (2001a) fand für das FIML-Verfahren bei MAR-Datenausfällen und leichten Abweichungen von der multivariaten Normalverteilung praktisch vernachlässigbare Verzerrungen der Parameterschätzungen ähnlich denen, die auch bei „konventioneller“ Maximum-Likelihood-Schätzung vollständiger Datensätze ohne fehlende Werte berichtet wurden. Es scheint deshalb so, daß die oben aufgelisteten Schlußfolgerungen und Regeln zur Robustheit der Maximum-Likelihood-Schätzungen von Strukturgleichungsmodellen auch auf die FIML-Schätzung übertragbar sind, was ja auch intuitiv einleuchtet, da dieser auf

dieselben Verteilungsannahmen bzw. auf dieselbe Wahrscheinlichkeitsdichtefunktion der beobachteten Modellvariablen, auf der die jeweils zu maximierende Likelihood beruht, begründet ist, wie erstere.

2.3.2 *Bewertung und Vergleich der Anpassungsgüte der Modelle*

Eine wichtige und bei den nachfolgenden Stabilitätsanalysen angewandte grundlegende Verfahrensweise der Verwendung von Strukturgleichungsmodellen ist der Vergleich von (im englischen Sprachraum) sogenannten „nested models“ (Bollen, 1989, 289ff; MacCallum, 1995, 34-35): Es werden jeweils paarweise verschiedene Modellvarianten eines Strukturgleichungsmodells miteinander verglichen, die sich darin unterscheiden, daß die frei schätzbaren, unrestringierten Parameter des einen Modells eine Teilmenge der frei schätzbaren Parameter des anderen Modells bilden. Beispielsweise könnte im QMSM getestet werden, ob die unstandardisierten Autoregressionskoeffizienten β_t über die Zeitreihe hinweg invariant sind: Dazu wäre ein Modell mit der Gleichheitsrestriktion $\beta_t = \beta$ gegen ein in dieser Hinsicht unrestringiertes Modell mit frei schätzbaren β_t zu vergleichen.

Die prinzipielle Logik solcher Modellvergleiche besteht darin, das „sparsamste“ (d.h. am stärksten restringierte) Modell auszuwählen, für das noch eine Modellanpassung an die empirischen Daten gefunden wird, die gegenüber dem unrestringiertesten Modell keinen substantiellen Verlust an sogenanntem Model-Fit bedeutet. Grundsätzlich ist diese Logik durch ein epistemologisches Objektivitätskriterium begründet, welches auf der Zurückweisbarkeit des Modells durch die in der Modellanpassung implizierte empirische Überprüfung beruht – die Zurückweisbarkeit des Modells aber wächst mit dessen Restriktivität bzw., technisch betrachtet, mit dessen Freiheitsgraden: „For a given set of data, models with excellent fit and more degrees of freedom are to be preferred, because such models are subjected to more conditions of disconfirmability“ (Mulaik & James, 1995, 132; diese Arbeit liefert eine umfassendere Darstellung des hier nur knapp wiedergegebenen wissenschaftstheoretischen Hintergrunds dieser Schlußfolgerung). Es wird also diejenige Modellvariante ausgewählt, die bestmöglich zu den empirischen Daten paßt und dabei mit möglichst wenigen frei zu schätzenden Kennwerten auskommt. Diese Vorgehensweise entspricht, wie im folgenden Kapitel noch gezeigt wird, sehr gut den inhaltlichen Fragestellungen der hier angestrebten Untersuchung der Merkmalsstabilität von Zufriedenheit, weil die zugrundegelegten Hypothesen in solche Restriktionen einzelner Modellparameter übersetzt werden können.

Zum Vergleich unterschiedlich restringierter Modellversionen müssen Entscheidungskriterien für Annahme oder Zurückweisung des jeweils sparsameren Modells festgelegt werden. Ein gewissermaßen „klassisches“ Kriterium liefert der $\Delta\chi^2$ -Test (andere geläufige Bezeichnungen sind χ^2 -Differenzen-Test, Likelihood-Ratio-Test). Dieser beruht auf der Differenz der Diskrepanzwerte (χ^2 -Werte der Modellanpassungsfunktion) des stärker restringierten und des schwächer restringierteren Modells, welche wiederum dann, wenn der

der Diskrepanzunterschied in der Population nicht vorhanden ist (Nullhypothese), als Realisation einer χ^2 -verteilten Zufallsvariable aufzufassen ist, deren Freiheitsgrade sich aus der Differenz der Freiheitsgrade des stärker und des schwächer restringierten Modells ergeben (vgl. z.B. Bollen, 1989, 292). $\Delta\chi^2$ (die Differenz der χ^2 -Werte) und Δdf (die Differenz der Freiheitsgrade) sind auf jeden Fall nichtnegativ, denn jede Hinzufügung einer Parameterrestriktion bedeutet einen zusätzlichen Freiheitsgrad des Modells und kann nur in einer Beibehaltung oder Erhöhung des Diskrepanzwerts resultieren. Somit kann also $\Delta\chi^2$ ebenso auf Signifikanz getestet werden, wie die χ^2 -Modellanpassungswerte selbst, das jeweils weniger restringierte Modell würde dann zurückgewiesen, wenn die χ^2 -Differenz, d.h. der aus den zusätzlichen Restriktionen resultierende „Misfit“, nicht signifikant ist. Allerdings ist m.E. damit der $\Delta\chi^2$ -Test zumindest im Prinzip mit denselben Problemen behaftet, wie auch der χ^2 -Test des Model-Fit selbst. Im folgenden wird darum zunächst diese Problematik der praktischen Bewertung der Anpassungsgüte einzelner Strukturgleichungsmodelle diskutiert, bevor dann die des Modellvergleichs betrachtet wird.

Der χ^2 -Test der durch eine SEM-Parameterschätzung für ein bestimmtes Modell erzielten Modellanpassung (allgemeine Darstellung siehe z.B. bei Hu & Bentler, 1995, 77-78; Bollen, 1989, 263) wurde im Laufe der wachsenden Entwicklung und Verbreitung der Strukturgleichungsmodelle zusehends kritisiert aufgrund seiner im Grunde unrealistischen Nullhypothese eines perfekten Model-Fit und der aus der Abhängigkeit der Teststärke von der Stichprobengröße resultierenden Tendenz zur Zurückweisung von „guten“ Modellen, wenn die Analyse auf großen Stichproben beruht, wie sie eigentlich angesichts der bloß asymptotischen Güte (Erwartungstreue, Effizienz) der SEM-Parameterschätzung wünschenswert sind. Durch den χ^2 -Test würde dann auch bei einer eigentlich guten Anpassung der modellimplizierten an die empirisch geschätzte Varianz-Kovarianz-Matrix oft ein Model-Misfit diagnostiziert. Bei Maximum-Likelihood-Schätzungen kann, wie oben bereits erwähnt, das Problem noch durch Nichtnormalität der Daten verschärft werden. Dieser Nachteil hat zur Entwicklung verschiedener Model-Fit-Indizes geführt, welche jeweils als Maßzahl für die mit dem Modell erzielte Anpassungsgüte brauchbar sein sollen. Die praktische Verwendung solcher Fit-Indizes ist allerdings auch nicht ganz problemfrei, da die verwirrende Vielzahl solchermaßen vorgeschlagener Kennwerte die Frage aufwirft, welche davon und wie diese zur Beurteilung der Anpassungsgüte einer Modellschätzung verwendet werden sollen. Während allem Anschein nach bei der praktischen Anwendung von Strukturgleichungsmodellen zu Forschungszwecken gewisse Gewohnheiten vorherrschen, sowohl was die häufige Verwendung einiger weniger besonders „populärer“ Indizes, als auch die Befolgung von Daumenregeln zur Festlegung von Kriterien der „guten“ Anpassung anbelangt, offenbart die jüngere diesbezügliche statistische Fachliteratur einen Eindruck, der angesichts der weiten Verbreitung und langjährigen intensiven Entwicklung der Strukturgleichungsmodelle doch überrascht: Es scheint so, als wisse die „Gemeinde der Strukturgleichungstheoretiker“ keine eindeutigen und allgemein akzeptierten Antworten auf die obige Frage.

Die Vielzahl von Fit-Indizes reflektiert eine Vielzahl unterschiedlicher Konzeptionen einer graduell ausgeprägten Anpassungsgüte von SEM-Modellanpassungen, charakterisierbar durch eine jeweils spezifische Akzentuierung und Gewichtung unterschiedlicher inhaltlicher Attribute der Anpassungsgüte, wie von Tanaka (1993) dargelegt wurde. Sollen die Indizes allein aufgrund der ihnen jeweils innewohnenden Logik beurteilt werden, so muß letztlich ein Urteil über den Wert dieser jeweiligen Gewichtung gefällt werden, und es scheint so, daß dafür ein allgemeingültiger theoretischer Referenzrahmen, aus dem sich dann der wichtigste Fit-Index quasi von selbst ergäbe, nicht existiert bzw. noch nicht gefunden wurde.

Eine Möglichkeit, die verschiedenen Fit-Indizes auf konkrete Anwendungsempfehlungen hin zu beurteilen, besteht darin, ihre Aussagekraft in Simulationsstudien zu testen. Dieser Weg wurde in einigen Untersuchungen besprochen (zum Überblick siehe z.B. Hu & Bentler, 1995; Marsh et al., 1996), jedoch ergeben diese zusammengenommen auch kein eindeutiges Bild etwa der generellen Überlegenheit eines bestimmten Kennwerts. Dieses kann im Grunde auch nicht überraschen, da m.E. gerade bei Simulationsstudien zu diesem Thema die jeweilige Generalisierbarkeit der Ergebnisse besonders schwierig zu beurteilen ist. Letztlich geht es dabei stets darum, verschiedene Indizes darin zu vergleichen, wie exakt sie richtige oder falsche Spezifikationen eines Strukturgleichungsmodells unter bestimmten experimentell variierten Bedingungen (z.B. Stichprobengröße, Schätzalgorithmus, Verletzung von Verteilungsannahmen etc.) anzeigen, und angesichts des sehr breiten Spektrums möglicher Spezifikationen und Anwendungsbedingungen von Strukturgleichungsmodellen ist hier die Generalisierbarkeit über die jeweils wenigen spezifizierten Bedingungen hinaus und insbesondere auch, da es ja gewissermaßen um die Diagnostik von Modellspezifikationen geht, über die jeweils untersuchten speziellen Modelle hinaus fraglich. Denn es muß in solchen Studien jeweils ein bestimmtes „wahres“ Strukturgleichungsmodell (oder einige wenige) der Datengenerierung zugrundegelegt werden und es können für die generierten Stichproben dann nur einige mehr oder weniger falsch spezifizierte Modelle berechnet werden, so daß nicht nur zu fragen ist, wie gut das jeweilige „wahre“ Modell die Gesamtheit aller möglichen Modelle repräsentiert, sondern darüber hinaus auch, wie gut die jeweils realisierten Mißspezifikationen dieses Modells die Gesamtheit der möglichen Mißspezifikationen dieses einen und zudem noch aller möglichen anderen „wahren“ Modelle repräsentieren. Außerdem unterscheiden sich diese Studien auch darin, wie sie jeweils die Sensibilität der Indizes auf die untersuchte Variation der Modellspezifikationen operationalisieren, was wiederum zu unterschiedlichen Schlußfolgerungen bezüglich dieser Sensibilitäten führen kann. Auch sind in den etwas älteren Studien neuere Entwicklungen, v.a. der in jüngster Zeit zusehends ins Zentrum des diesbezüglichen wissenschaftlichen Interesses gerückte RMSEA (siehe unten), noch gar nicht berücksichtigt. Und außerdem, mit Blick auf die der hier geplanten Analysen, bleibt zu fragen, ob die berichteten Ergebnisse zur Aussagekraft einzelner Indizes bei konventioneller Maximum-Likelihood-Schätzung ohne weiteres auf die Verwendung des FIML-Algorithmus übertragbar sind – immerhin wird der

χ^2 -Wert, der nicht nur im obengenannten Test verwendet wird, sondern auch der Berechnung der meisten Fit-Indizes zugrundeliegt, hier auf andere Weise berechnet, als bei der konventionellen Schätzmethode (vgl. Arbuckle & Wothke, 1999; Byrne, 2001, 293).

Diese Anmerkungen sollen lediglich andeuten, daß die Verwendung der Fit-Indizes ein komplexes Spezialthema der Strukturgleichungsmethodologie darstellt und genauegenommen zum gegenwärtigen Zeitpunkt kaum ein eindeutiger diesbezüglicher „state of the art“ ausgemacht werden kann. Eine einigermaßen erschöpfende Darstellung der bislang dazu vorgelegten Untersuchungen würde den Rahmen dieser Arbeit sprengen, ich verweise stattdessen auf die Buchkapitel von Hu & Bentler (1995) und Marsh et al. (1996), welches aber auf die sogenannten inkrementellen Fit-Indizes beschränkt ist, – beide Arbeiten beziehen allerdings die jüngsten Entwicklungen nicht mehr mit ein. Einen aktuelleren, wenn auch weniger detaillierten Überblick liefert Byrne (2001, 79ff).

Zu diesen jüngsten Entwicklungen zählt v.a. die wachsende „Prominenz“ des RMSEA (Kürzel für „root mean squared error of approximation“) in Forschung und Anwendung: Obwohl dieser Kennwert bereits 1980 vorgestellt wurde (vgl. Steiger, 2000), scheint man erst in den letzten Jahren, ausgelöst durch Browne & Cudeck (1993), auf seine zumindest theoretisch hervorragenden Qualitäten für die Beurteilung der Anpassungsgüte eines Modells aufmerksam geworden zu sein. Seine seitherige Verbreitung mag sich u.a. dem Umstand verdanken, daß dieser Kennwert im Kontext der Intervallschätzung benutzt werden kann: Der RMSEA ist konzipiert als Populationsparameter der absoluten Anpassungsgüte eines Modells in der Population: „Error of approximation is relevant to the question, How well would the model, with unknown, but optimally chosen, parameter values fit the population covariance matrix if it were available?“ (Browne & Cudeck, 1993, 137-8). Der jeweils errechnete Stichprobenkennwert ist die Schätzung dieses Parameters und im Gegensatz zu gebräuchlichen Kennwerten wie GFI, AGFI oder den inkrementellen Indizes NFI, RFI, IFI, TLI, CFI (dieses sind die für das hier verwendete Softwarepaket AMOS 4.0 gewählten Bezeichnungen, vgl. Byrne, 2001, 79ff) sind die Verteilungseigenschaften dieses Stichprobenkennwerts ableitbar, so daß z.B. für einen gegebenen Wert Konfidenzintervalle berechnet werden können oder dieser Wert gegen die Nullhypothese eines zuvor spezifizierten RMSEA-Parameterwerts getestet werden kann. Die häufige Verwendung des RMSEA – d.h. genauegenommen, des aus der Stichprobe geschätzten RMSEA-Werts – hat m.E. bislang kaum Widerspruch in der „Strukturgleichungsgemeinde“ hervorgerufen,²¹ was gewissermaßen als eine Art von unausgesprochenem Konsens bezüglich der prinzipiellen Relevanz dieses Index gewertet werden könnte. Eine Simulationsstudie, die die Bedeutung des RMSEA stützt, lieferten Fan et al. (1999): Dort zeigte dieser Index unter zehn untersuchten die deutlich größte Sensitivität gegenüber Modellmißspezifikationen und die geringste gegenüber Variationen der Stichprobengröße (es gelten auch hier die oben dargelegten Vor-

²¹ Eine Ausnahme bildet hier die Kritik von Hayduk & Glaser (2000), welche allerdings von Steiger (2000) in sehr überzeugender Art und Weise widerlegt wurde. Letztere Arbeit enthält auch eine sehr klare und empfehlenswerte formale Darstellung des RMSEA.

behalte bezüglich der Generalisierbarkeit solcher Ergebnisse).

Ein m.E. unbedingt erwähnenswerter, sehr interessanter Ansatz zu der Frage der praktischen Anwendung der Fit-Indizes liefern Hu & Bentler (1999), die die Kombination von Cut-Off-Werten des SRMR (Kürzel für „standardized root mean squared residual“, in AMOS als unstandardisierter Wert RMR ausgegeben) und eines Index aus einer Gruppe, in der u.a. auch RMSEA, TLI, IFI und CFI enthalten sind zur Entscheidung über Zurückweisung oder Annahme eines Modells vorschlagen. Grundlage dieses Vorschlags war die in Hu & Bentler (1998) berichtete Untersuchung, bei der diese Indizes sich als solche mit brauchbareren Sensitivitäten gegenüber Modellmißspezifikationen zeigten und der SRMR sich von der restlichen Gruppe der über verschiedene untersuchte Bedingungen hinweg ähnlich reagierenden Indizes unterschied. Ein Blick in die umfangreichen Ergebnistabellen zeigt z.B. die Kombination der Cut-Off-Werte $RMSEA \leq 0.05$ und $SRMR \leq 0.06$ als Kombination, die bei mittleren und großen Stichprobenumfängen (≥ 250) kaum noch Fehler (Zurückweisung eines korrekt spezifizierten, Annahme falsch spezifizierter Modelle) produziert – ähnlich erfolgreich zeigte sich nur die Kombination von $SRMR \leq 0.06$ und $TLI \geq 0.96$. Dieses könnten, wiederum mit dem einschränkenden Hinweis auf die Generalisierbarkeit solcher Ergebnisse versehen, vernünftig scheinende Entscheidungsregeln sein, die z.B. die immer häufiger zu beobachtende praktische Einschränkung auf die von Browne & Cudeck (1993) vorgegebene Daumenregel $RMSEA \leq 0.05$ als Kriterium für „guten“ (bzw. $RMSEA \leq 0.08$ für „akzeptablen“) Model-Fit aufheben bzw. die Entscheidung auf eine breitere Basis der Fit-Bewertung stellen würde. Jedoch konnte eine solche Regel für die eigene Untersuchung nicht angewandt werden, weil bei unvollständigen Datensätzen und Anwendung des FIML-Algorithmus der SRMR schlicht nicht berechenbar ist.

Die statistische Fachliteratur liefert somit der folgenden Untersuchung keine strikten und verbindlichen praktischen Regeln zur Bewertung der Anpassungsgüte, sondern allenfalls ein gewisses „business as usual“ in Gestalt der derzeit allgemein üblichen Praxis der Auflistung und ad hoc Interpretation verschiedener Fit-Werte. Dieses muß nicht unbedingt als Problem gesehen werden, denn immerhin ist aus den verschiedenen Studien sozusagen auch eine positive Botschaft herauszulesen: Eindeutige Empfehlungen zur Bevorzugung einzelner Indizes fallen darum schwer, weil im Grunde die meisten Kennwerte durchaus aussagekräftig und somit verwendbar sind. Und so lautet die schlichte allgemeine Empfehlung für die Anwender, die Bewertung des Model-Fit nicht ausschließlich auf einen Kennwert zu begründen, sondern möglichst mehrere Kennwerte unterschiedlicher Model-Fit-Typen zu berücksichtigen (Bollen & Long, 1993; Tanaka, 1993; Bollen, 1989, 281). Hoyle & Panter (1995) geben sogar explizite Empfehlungen darüber, welche Fit-Indizes berichtet werden sollten - jedoch sind auch dabei wiederum neuere Entwicklungen wie der RMSEA unberücksichtigt.

Zum gewohnheitsmäßigen Umgang mit den Fit-Indizes zählen auch bestimmte Cutoff-Werte zur Interpretation derselben: Für die inkrementellen Fit-Indizes wie NFI, TLI, ICI und CFI besagte eine ursprünglich weitverbreitete Daumenregel, daß Werte größer/gleich

0.9 als Ausdruck eines guten Model-Fits gelten könnten (z.B. Bentler & Bonnet, 1980), dieser Cutoff-Wert wurde aber inzwischen aufgrund vieler Studien nach oben korrigiert, so daß heute für diese Indizes 0.95 als Cutoff-Wert empfohlen wird (z.B. Hu & Bentler, 1999). Bei der Verwendung des RMSEA dominiert der wie schon erwähnt von Browne & Cudeck (1993) empfohlene Cutoff-Wert von 0.05. Hu & Bentler (1999) schlagen den etwas höheren Wert 0.06 vor, da in ihrer Simulation ein Cutoff von 0.05 zu vielen Zurückweisungen korrekt spezifizierter Modelle bei kleinen Stichprobenumfängen führte. Dagegen verweist Steiger (2000), der eigentliche Erfinder des RMSEA, auf den mit diesem Kennwert intendierten Fortschritt durch die Intervall-, anstatt Punktschätzung der Modellanpassungsgüte und empfiehlt, einen einzelnen Cutoff-Wert nicht zu ernst zu nehmen und stattdessen das Konfidenzintervall des RMSEA-Werts stärker in die Beurteilung miteinzubeziehen.

Für die folgenden Analysen werden aus dem hier nur ansatzweise wiedergegebenen Forschungsstand bezüglich der Bewertung der Modellanpassungsgüte von SEM-Parameterschätzungen pragmatische Regeln abgeleitet. Prinzipiell sollen die folgenden Kennwerte berichtet werden:

- ▶ Der χ^2 -Wert und die Freiheitsgrade der Modellanpassung, sowie die Signifikanz des χ^2 -Tests – dieses gewissermaßen der Vollständigkeit und Gewohnheit wegen. Aufgrund der dargelegten Problematik dieses Tests soll dem Signifikanztest keine allzu große Bedeutung beigemessen werden.
- ▶ Unter Berücksichtigung der Ergebnisse und Empfehlungen von Hu & Bentler (1998), Hoyle & Panter (1995) sowie Marsh et al. (1996) werden folgende inkrementelle Fit-Indizes berichtet: NFI, TLI, CFI und IFI. Für diese Werte wird der Cutoff-Wert von 0.95 für „guten“ Model-Fit zugrundegelegt. Es wird erwartet, daß sich diese kaum unterscheiden werden, wenn doch, so könnte aus den von Hu & Bentler (1999) und Marsh et al. (1996) berichteten Ergebnissen ein Hinweis auf Bevorzugung des TLI abgeleitet werden.
- ▶ Es wird der RMSEA berichtet und als vorrangiges Kriterium zur Beurteilung des Model-Fit herangezogen: Der konventionelle Cutoff-Wert von $RMSEA \leq 0.05$ wird berücksichtigt. Außerdem werden die von AMOS 4.0 ausgegebenen 90%-Konfidenzintervalle angegeben.

Dieses sind Regeln zur jeweils singulären Beurteilung der Modellanpassungsgüte der einzelnen Strukturgleichungsmodelle. Es wurde oben aber die Frage aufgeworfen, welche Regeln für die im folgenden angestrebten Vergleiche von „nested models“ zu befolgen sind. Die diesbezügliche statistische Fachliteratur behandelt m.E. kaum in ähnlich intensiver Weise Alternativen zum $\Delta\chi^2$ -Test, wie zum χ^2 -Test. Solche scheinen aber notwendig, denn es liegt auf der Hand, daß genauso, wie letzterer bei großen Stichprobenumfängen zur häufigen Zurückweisung gut, jedoch nicht perfekt, angepaßter Modelle führt, durch ersteren dann auch häufig praktisch irrelevante Differenzen der Anpassungsgüte von weniger und stärker restringierten Modelversionen signifikant werden und somit das sparsamere Modell zu oft zurückgewiesen wird. Diese Befürchtung wurde für die nachfolgenden Analysen auch durch erste „Erkundungsanalysen“ mit den SOEP-Zufriedenheitswerten und auch

sonstige Auswertungserfahrungen mit SOEP-Daten genährt: Bei Verwendung des FIML-Algorithmus wurde bei Modellvergleichen der $\Delta\chi^2$ -Test hier praktisch immer hochsignifikant mit $\alpha \leq 0.001$, auch dann, wenn in den Fit-Indizes praktisch kein Unterschied zwischen den Modellen festzustellen war. Übrigens ist bei Anwendung des FIML-Algorithmus der χ^2 -Test der Modellanpassung nichts anderes als ein $\Delta\chi^2$ -Test – nämlich der Differenz der Diskrepanzwerte des untersuchten Modells und des sogenannten saturierten, d.h. vollkommen restriktionsfreien Modells (Arbuckle & Wothke, 1999).²² Konsequenterweise sollte hier dann auch den $\Delta\chi^2$ -Tests bei geplanten Modellvergleichen keine größere Bedeutung beigemessen werden, als dem χ^2 -Modelltest.

Alternative Verfahren zur Signifikanztestung der aus Modellrestriktionen resultierenden Verschlechterung der Anpassungsgüte sind der Lagrange-Multiplier-Test und der Wald-Test (siehe z.B. Bollen, 1989, 293f; Bentler, 1995, 219-221). Für beide Tests wurde bislang keine für das FIML-Schätzverfahren adjustierte Version vorgestellt und sie stehen in AMOS 4.0 nicht zur Verfügung. Angesichts ihrer asymptotischen Äquivalenz zum $\Delta\chi^2$ -Test (Jöreskog, 1993) ist auch anzunehmen, daß sie in derselben Weise mit dem Problem zu hoher Teststärke bei großen Stichproben behaftet sind.

Somit besteht Bedarf nach vernünftigen Regeln, die unter Absehung vom $\Delta\chi^2$ -Testergebnis eine Annahme des sparsameren Modells dann erlauben, wenn der Unterschied in der tatsächlichen Anpassungsgüte minimal ist. Hierzu scheint besonders der RMSEA-Wert in seiner Eigenschaft als Maßzahl der „Diskrepanz pro Freiheitsgrad“ des Modells (vgl. Browne & Cudeck, 1993, 144) geeignet: Ein sparsameres Modell kann dann als einem weniger sparsamen gleichwertiges angesehen werden, wenn es nicht mehr Diskrepanz pro Freiheitsgrad aufweist, wie dieses. In Ermangelung konkreter diesbezüglicher Vorschläge in der statistischen Fachliteratur wurde für die nachfolgenden Modellvergleiche deshalb die folgende, vernünftig scheidende pragmatische Vorgehensweise festgelegt: Das jeweils sparsamere Modell wird dann nicht zurückgewiesen, wenn sein durch die oben ausgewählten Fit-Indizes, insbesondere den RMSEA-Wert, ausgedrückter Model-Fit genauso gut ist, wie der des weniger restringierten Modells. Als „genauso gut“ soll der Model-Fit dann gelten, wenn die Differenzen aller aufgeführten Fit-Indizes jeweils nicht größer als 0.01 sind und auch die RMSEA-Vertrauensintervalle sich überschneiden.

2.4 Fazit: Hinweise zur Auswahl statistischer Modelle und Prozeduren bei längsschnittlichen Datenanalysen

Nachdem in den vorangegangenen Unterkapiteln der Darstellung und Diskussion verschiedener statistischer Analysemodelle sehr viel Raum gegeben wurde, erscheint es an dieser Stelle angebracht, aus alledem ein Resultat in Form einer Art praktischer Handreichung ab-

²²Bei unvollständiger Datenmatrix produziert das saturierte Modell nicht zwangsläufig eine perfekte Anpassung.

zuleiten, welche nicht nur für die im folgenden dargestellten eigenen empirischen Analysen dienlich war, sondern auch einem Leser dieser Abhandlung für zukünftige längsschnittliche Analysen hilfreich sein könnte.

Im Grunde jedoch sind solche praktischen Empfehlungen in den vorherigen Darlegungen bereits enthalten, so daß hier nur noch die zusammenfassende Auflistung der darin m.E. wesentlichsten Punkte bleibt:

- ▶ Zur Analyse von Kohorteneffekten bzw. zur Trennung von Alters- und Kohorteneffekten in längsschnittlichen Daten eignet sich ein Multilevel-Regressionsmodell, wie es in Kapitel 2.1 dargestellt wurde. Der Nachteil dieser Modellierung liegt in den darin implizierten Linearitätsannahmen – welche allerdings durch die Einführung exponentieller Prädiktortermine für Alter und Kohorte modifiziert werden könnten –, sowie in der Möglichkeit eines durch dieses Modell unentdeckten Periodeneffekts, der hier zum geschätzten Alters- und Kohorteneffekt beitragen kann. Demgegenüber besitzt diese Modellierung viele Vorteile, die für die hier relevante Fragestellung gar nicht alle genutzt werden müssen: Es wird durch die Random-Komponente des Alters interindividuelle Variabilität der Wirksamkeit des Alterseffekts erfaßt (entsprechend der Slope-Varianz in den bekannteren Growth-Models), dazu besteht die prinzipielle Möglichkeit, weitere Prädiktoren auf dem „Personenlevel“ einzufügen und damit die Abhängigkeit alterskorrelierter Veränderungen von diesen Prädiktoren zu untersuchen, außerdem wird bei der üblichen Maximum-Likelihood-Parameterschätzung dieses Modells eine Modellanpassung für alle vorhandenen Meßwerte durchgeführt, so daß Merkmalsträger mit fehlenden Werten nicht ausgeschlossen und auch fehlende Werte nicht ersetzt werden müssen – das Verfahren erfüllt das im Kapitel 2.3.1 erläuterte Kriterium der „ignorability“ von MAR-Datenausfällen.
- ▶ Zur Analyse von Stabilität und Veränderung eines Merkmals ist aufgrund der Vermischung von Reliabilität und tatsächlicher Stabilität in Meßwertkovarianzen bzw. -korrelationen eine meßfehlerbereinigte Schätzung von Stabilitätsparametern wünschenswert, wie sie durch längsschnittliche Strukturgleichungsmodelle mit latenten True-Score-Komponenten ermöglicht wird. Das für die nachfolgenden Analysen ausgewählte QMSM ist dann zu empfehlen, wenn Merkmalsstabilität „unvoreingenommen“ durch bestimmte theoretisch vorab spezifizierte Verlaufsmodelle der Veränderung untersucht werden soll. LDCM und LDSM stellen im Grunde modifizierte Modellvarianten des QMSM dar, deren Besonderheit in der expliziten Modellierung von Veränderungswerten liegt, welche im ersteren Falle als modellexogene, im letzteren als jeweils vom vorherigen True-Score abhängige absolute Veränderungsbeträge modelliert sind. Solche Modellvarianten können v.a. dann nützlich sein, wenn weitere exogene Einflußgrößen auf die Veränderung einbezogen werden sollen, welche dann als Prädiktoren der Veränderungsbeträge modellierbar wären. Das weithin bekannte LGCM ist durch eine im Vergleich zu den vorgenannten Modellen erheblich restriktivere Modellierung der stattfindenden Veränderung gekennzeichnet: Wie gezeigt handelt es sich um ein Verlaufsmodell, durch welches gewissermaßen

die Form der zeitlichen Veränderung interindividuell invariant festgesetzt (z.B. im einfachsten Falle als lineare Veränderung) und Variation zwischen den Merkmalsträgern nur in den Parametern, die diese Verlaufsform beschreiben (z.B. die Steigung des linearen Verlaufs) zugelassen ist. Das LGCM wäre somit dann nicht geeignet, wenn solche Restriktivität der Fragestellung nicht angemessen scheint (wie im vorliegenden Fall), gibt es dagegen gute (theoretische) Gründe für die Annahme, daß die zu untersuchende Veränderung durch ein interindividuell invariantes Verlaufsmodell hinreichend abgebildet werden kann, so besitzt das LGCM den Vorteil der größeren Sparsamkeit und ist den vorgenannten Modellen vorzuziehen.

- ▶ Ein in längsschnittlichen Untersuchungen nahezu zwangsläufig sich ergebendes Problem sind große Anzahlen fehlender Werte infolge des Dropouts im Panelverlauf. Im Kapitel 2.3.1.3 wurden Methoden des Umgangs mit fehlenden Werten aufgeführt, die gegenüber weitverbreiteten Vorgehensweisen wie dem fallweisen Ausschluß erhebliche Vorteile aufweisen – insbesondere den, daß die Unverzerrtheit der daraus resultierenden Schätzungen nicht an die sehr restriktive Annahme der vollständigen Zufälligkeit der Datenausfälle (MCAR) geknüpft ist, sondern an die weniger restriktive Bedingung von MAR-Datenausfällen. Letztere ist leicht dadurch herzustellen, daß Prädiktoren des Datenausfalls ins statistische Modell einbezogen werden. Es ist also zu empfehlen, bei der Berechnung längsschnittlicher Strukturgleichungsmodelle mit größeren Dropoutraten eine der an die MAR-Bedingung geknüpften Methoden anzuwenden – besonders einfach ist hier die Anwendung der FIML-Schätzung. Von diesen Empfehlungen abgesehen besteht sozusagen die praktische Handreichung bezüglich des Problems fehlender Werte im Grunde in der Verfassung des Kapitels 2.3.1 selbst, denn vorhandene Texte zu diesem Thema erscheinen oft sehr wenig verständlich, weil sie entweder formal hochkomplex, oder aber im Gegenteil zu oberflächlich gehalten sind. Im obigen Kapitel wurde versucht, in dieser Hinsicht einen Mittelweg zu gehen – wofür dann auch eine etwas umfängliche Länge dieser Ausführungen in Kauf genommen wurde: Es schien hier durchaus lohnend, einmal den „Stand der Dinge“ einigermaßen ausführlich zu dokumentieren.
- ▶ Bei der Anwendung von Strukturgleichungsmodellen ist der Vergleich von „nested models“ eine häufig gewählte Vorgehensweise: Ihre Attraktivität besteht in der Möglichkeit, theoretische Annahmen in verschiedene Modellrestriktionen eines Modells zu übersetzen und durch diesen Vergleich zu testen. Dabei aber kann oft ein praktisches Problem in der zu großen Teststärke des $\Delta\chi^2$ -Test bei hinreichend großen Untersuchungsstichproben bestehen, wodurch dann auch geringfügige Unterschiede in der Modellanpassung statistisch signifikant werden. Die in Unterkapitel 2.3.2 aufgestellten Entscheidungsregeln für den Modellvergleich können hier als Beispiel dienen, wie man stattdessen vorgehen könnte. Sie sind aber, darauf sei nochmals hingewiesen, praktische Festlegungen für die nachfolgenden Analysen, welche nicht entsprechenden Veröffentlichungen in der SEM-Literatur entnommen wurden.

3 Empirische Untersuchungen zur Zufriedenheit im höheren Alter

Nach den umfänglichen statistischen Darlegungen des letzten Kapitels, mit denen sozusagen das „Methodenarsenal“ zur Untersuchung der im Kapitel 1.3 aufgelisteten Hypothesen der empirischen Untersuchung bereitgestellt wurde, sollen nun wieder die dort zusammengefaßten inhaltlichen Fragestellungen aufgegriffen und weiterverfolgt werden. Es folgt also die Darstellung eigener empirischer Analysen zu Kohorteneffekthypothese, Gewinner-und-Verlierer-Hypothese, Hypothese allgemeiner und erhöhter Merkmalsstabilität von Zufriedenheit. Daß dabei Daten des Sozio-oekonomischen Panels (SOEP) verwendet wurden, wurde ebenfalls bereits im Kapitel 1.1.4.8 angekündigt, im folgenden wird zunächst diese Datengrundlage kurz beschrieben, bevor dann die durchgeführten Datenanalysen – grob untergliedert in eher mittelwertsbezogene Analysen und eher kovarianzbezogene Stabilitätsanalysen – eingehend behandelt werden.

3.1 Stichprobe und Untersuchungsvariablen

Es wurde im Kapitel 1.3 bereits dargelegt, daß der empirische Nachweis der dort hypothetisch postulierten Effekte längsschnittliche Analysen erfordert, denn im Kern betreffen alle Hypothesen die Veränderung von Zufriedenheit im Altersverlauf, so daß sie nur anhand solcher Daten statistisch überprüfbar sind, in denen diese Veränderung auch erfaßt ist. Ein Datensatz längsschnittlicher Zufriedenheitsmessungen ist also die Grundvoraussetzung dieser Untersuchung, zudem sollte dieser Datensatz Stichprobengrößen bieten, die den diesbezüglichen Anforderungen der zur Abschätzung der postulierten Effekte notwendigen komplexen statistischen Modelle genügen. Auch wäre es angesichts des großen Grades an Allgemeinheit bzw. des nicht durch spezifische psychologische oder sozioökonomische Merkmale eingeschränkten Geltungsbereichs, der für diese Effekte angenommen wurde, wünschenswert, daß auch die Stichprobe, die diesen Datensatz liefert, bei ihrer Generierung möglichst wenig auf eine in ihren psychologischen oder sozioökonomischen Merkmalen spezifische Gruppe Erwachsener beschränkt wurde.

3.1.1 Das Sozio-oekonomische Panel (SOEP): Paneldesign, Substichproben, Weiterverfolgungskonzept

Ein Datensatz, der diese Anforderungen erfüllt, ist das deutsche Sozio-ökonomische Panel (SOEP), eine seit 1984 vom Deutschen Institut für Wirtschaftsforschung (Berlin) durchgeführte jährliche Erhebung repräsentativer Daten über Personen, Haushalte und Familien in der Bundesrepublik Deutschland (im ostdeutschen Beitrittsgebiet wurde die Erhebung im Jahr 1990 gestartet) (Burkhauser et al., 1997; SOEP-Group, 2000). Deren allgemeine

Zielsetzung besteht in der Erfassung von Stabilität und Veränderung von Lebensbedingungen, wobei eine microökonomische Perspektive die Grundlage bildet, welche mit soziologischen, psychologischen und politwissenschaftlichen Variablen ergänzt wurde, so daß das SOEP neben objektiven Indikatoren (z.B. Einkommen, Beschäftigung) auch viele subjektive Indikatoren zur Erfassung individueller Wahrnehmung objektiver Lebensbedingungen enthält.²³

Der gesamte SOEP-Datensatz setzt sich aus den Daten von fünf Teilstichproben zusammen, welche durch Zufallsauswahl aus der jeweils zugrundegelegten Grundgesamtheit von *Haushalten* gewonnen wurden, so daß es sich (zumindest bei der jeweiligen Erstbefragung) um für diese Grundgesamtheiten global repräsentative Haushaltsstichproben handelt. Diese fünf Stichproben sind:

1. Westdeutsche Haushalte: 1984 erstmals befragte Haushalte der alten westdeutschen Bundesländer (einschließlich Westberlin), deren Haushaltsvorstand Deutscher oder Angehöriger einer sonstigen nicht durch die folgende Substichprobe erfaßten Nationalität war. Diese Stichprobe umfaßte bei der Erstbefragung 4528 Haushalte mit 9076 befragten Personen.
2. Ausländerhaushalte: 1984 erstmals befragte Haushalte der alten westdeutschen Bundesländer (einschließlich Westberlin), deren Haushaltsvorstand Türke, Italiener, Spanier, Grieche oder Yugoslawe war. Diese Stichprobe umfaßte bei der Erstbefragung 1393 Haushalte mit 3169 befragten Personen.
3. Ostdeutsche Haushalte: 1990 erstmals befragte Haushalte der ehemaligen DDR, deren Haushaltsvorstand Staatsbürger der DDR gewesen war. Diese Stichprobe umfaßte bei der Erstbefragung 2179 Haushalte mit 4453 befragten Personen.
4. Immigrantenhaushalte: 1994 oder 1995 erstmals befragte Haushalte mit mindestens einem Mitglied, welches seit 1984 aus dem Ausland in die Bundesrepublik übersiedelt hatte. Diese Stichprobe wurde sozusagen in zwei Schritten 1994 und 1995 generiert, im Jahr 1995 umfaßte sie 522 Haushalte mit 1078 Befragungspersonen.
5. Auffrischungsstichprobe: Im Jahr 1998 wurden weitere 1067 Haushalte aus der Grundgesamtheit aller bundesdeutschen Haushalte gezogen. 1998 umfaßte diese Stichprobe 1932 Befragungspersonen.

Die für diese Stichproben zu den einzelnen Erhebungswellen realisierten Stichprobenumfänge werden in Tabelle B.1 im Anhang B gezeigt.

²³Die grundlegende Themenbereiche des SOEP, mit jeweils regelmäßigen jährlichen Erhebungen, wurden wie folgt benannt: „Demographie und Bevölkerung“, „Arbeitsmarkt und Beschäftigung“, „Einkommen, Steuern und Soziale Sicherung“, „Wohnen“, „Gesundheit“, „Leistungen privater Haushalte“, „Bildung und Qualifikation“, „Grundorientierungen, Partizipation und Integration“. Dazu kommen noch spezielle Themenschwerpunkte der jeweiligen Erhebungsjahre, die nicht regelmäßig über den gesamten Erhebungszeitraum hinweg berücksichtigt wurden (siehe Haisken-DeNew & Frick, 2000, 16: Table 1.1).

Der Grund für die spezifische Generierung einer Ausländer- und einer Immigrantenstichprobe lag in speziellen Forschungsinteressen an diesen Bevölkerungsgruppen, für die auf diese Weise ausreichende Fallzahlen gewährleistet werden sollten. Der historische Grund für eine getrennte Generierung west- und ostdeutscher Teilstichproben liegt auf der Hand, während das Auffrischungssample die im Panelverlauf unvermeidliche allmähliche Schrumpfung der Ausgangsstichproben ausgleichen soll. Das Verhältnis der fünf Stichprobenumfänge entspricht nicht dem Verhältnis der Größen der jeweils zugrundegelegten Populationen, d.h. die jeweiligen Ziehungswahrscheinlichkeiten bei der Stichprobengenerierung unterscheiden sich zum Teil erheblich, so daß in der Gesamtstichprobe v.a. die ostdeutschen (gegenüber den westdeutschen) und die Ausländerhaushalte (gegenüber den deutschen) überrepräsentiert sind.²⁴

In der Datenstruktur muß zwischen haushaltsbezogenen Daten, d.h. haushaltsspezifischer, vermittelt eines Haushaltsfragebogens erhobener Information, und den vermittelt eines Personenfragebogens erhobenen personenbezogenen Daten unterschieden werden. Welche Personen bei einer SOEP-Erhebungswelle befragt wurden, geht aus dem durch die folgenden Punkte beschriebenen *Weiterverfolgungskonzept* des SOEP hervor:

- ▶ Als SOEP-Teilnehmer werden alle Personen erfaßt, die in den SOEP-Haushalten leben.
- ▶ Weiterverfolgt werden alle Personen, die in der jeweils ersten Befragungswelle erfaßt wurden - sowohl dann, wenn sie nach wie vor dem ursprünglichen SOEP-Haushalt angehören, als auch nach Umzug in einen anderen Haushalt. In letzterem Fall wird dieser Haushalt mit allen Personen, die in ihm leben, als neuer SOEP-Haushalt erfaßt.
- ▶ Alle erfaßten Personen werden persönlich befragt, sobald sie das Alter von 16 Jahren erreicht haben.
- ▶ Personen, die in einen existenten SOEP-Haushalt eingezogen sind, werden erfaßt. Seit 1989 werden diese Personen auch weiterverfolgt, wenn sie diesen Haushalt wieder verlassen.
- ▶ Personen und Haushalte, die in einem Jahr nicht interviewt werden konnten, werden weiterverfolgt (temporäre Dropouts). Die Weiterverfolgung endet im Falle von zwei aufeinanderfolgenden temporären Dropouts bei allen Haushaltsmitgliedern oder bei finaler Verweigerung. Bei Interviews nach einem temporären Dropout werden grundlegende, für das Vorjahr fehlende Informationen retrospektiv erfragt.

TABELLE 3.1:
Zufriedenheitsbefragungen im SOEP (bis 1999)

Inhalt:	Erhebungsjahr:															
	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99
L-jetzt	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
L-retro	✓	✓	✓	✓												
L-pro1	✓	✓	✓	✓												
L-pro5								✓	✓	✓	✓	✓	✓	✓	✓	✓
Ges	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Hhein	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Hhttg	✓	✓	✓	✓	✓	✓	✓			✓	✓	✓	✓	✓	✓	✓
Arb	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Ausb							✓			✓						
Lbstd								✓	✓	✓		✓	✓	✓	✓	✓
Whg	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Whgg								✓			✓					✓
AWD								✓	✓		✓		✓		✓	✓
ÖNV															✓	
Frz	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓		✓	✓	✓	✓
Frzumf								✓				✓				
Frzttg								✓				✓				
SozSi				✓					✓					✓		
Uwzst									✓	✓	✓	✓	✓	✓	✓	✓
Kibetr														✓	✓	✓

Erläuterungen:

✓ = Zufriedenheit erfragt.

Erste Spalte: Zufriedenheitsurteile über ...

L-jetzt = Leben gegenwärtig; **L-retro** = Leben retrospektiv vor 1 Jahr; **L-pro1** = Leben prospektiv in 1 Jahr; **L-pro5** = Leben prospektiv in 5 Jahren; **Ges** = Gesundheit; **Hhein** = Haushaltseinkommen; **Hhttg** = Haushaltstätigkeit; **Arb** = Arbeit; **Ausb** = (eigene) schulische und berufliche Ausbildung; **Lbstd** = (eigener) Lebensstandard; **Whg** = Wohnung; **Whgg** = Wohngegend; **AWD** = Angebot an Waren und Dienstleistungen; **ÖNV** = Öffentlicher Personennahverkehr; **Frz** = Freizeit; **Frzumf** = Freizeitumfang; **Frzttg** = Freizeittätigkeit; **SozSi** = Netz der sozialen Sicherheit; **Uwzst** = Umweltzustand; **Kibetr** = Möglichkeiten der Kinderbetreuung.

Gültigkeit: Westdeutsche Stichprobe, ostdeutsche Stichprobe ab 1991 (vgl. Text).

3.1.2 Zufriedenheitsbefragungen im SOEP

Ein-Item-Beurteilungen der allgemeinen Lebenszufriedenheit und verschiedener spezifischer Bereichszufriedenheiten wurden bei den jährlichen SOEP-Erhebungen regelmäßig durchgeführt. Tabelle 3.1 gibt einen Überblick über sämtliche im SOEP enthaltenen Zufriedenheitsfragen. Die Befragungen zu allgemeiner (aktueller, retrospektiver und prospektiver) Lebenszufriedenheit stehen in den jeweiligen Personenfragebögen jeweils ganz am Ende, während die Items zu spezifischen Zufriedenheiten in der Regel – mit Ausnahme der Jahre 1990, 1991 und 1992 – die persönliche Befragung einleiten. 1991 und 1992 wurden auch die spezifischen Zufriedenheiten gegen Ende des Personenfragebogens erfragt, ebenso 1990 bei der in diesem Jahr erstmals befragten ostdeutschen Stichprobe, während in diesem Jahr diese Items bei den seit 1984 laufenden Weststichproben wie gewohnt am Anfang standen. Auch bildet die Befragung 1990 insofern eine Besonderheit, als hier die Angaben in Tabelle 3.1 lediglich für die Weststichproben gelten und bei den Ostdeutschen teilweise andere Zufriedenheiten erfragt wurden.²⁵

Wie die Tabelle 3.1 zeigt, sind vier Zufriedenheiten über den gesamten Zeitraum von sechzehn Jahren hinweg ununterbrochen erfragt worden, nämlich außer der allgemeinen Lebenszufriedenheit noch die mit der Arbeit, dem Haushaltseinkommen, der Gesundheit sowie die Wohnzufriedenheit. Eine weitere Zufriedenheit, die mit der Freizeit, wurde 1990 und 1995 nicht erfragt, stattdessen wurde bei diesen beiden Erhebungen detaillierter nach der Zufriedenheit mit dem Umfang der Freizeit und der Freizeittätigkeit gefragt. Bei vielen der nachfolgenden Analysen werden jeweils alle diese vollständigen Zufriedenheiten untersucht – mit Ausnahme der Arbeitszufriedenheit, weil die Mehrzahl der höheraltrigen Personen nicht mehr aktiv am Arbeitsleben teilnimmt. Um auch für die Freizeit Zufriedenheit eine vollständige Zeitreihe zu erhalten, wurde für die Jahre 1990 und 1995 ein Wert aus dem arithmetischen Mittel der Zufriedenheiten mit der Freizeittätigkeit und dem Freizeitumfang gebildet.

Tabelle B.2 im Anhang B listet die grundlegendsten Verteilungskennwerte – Mittelwerte und Standardabweichungen – der im folgenden untersuchten Zufriedenheitsmessungen in der westdeutschen SOEP-Teilstichprobe auf.

²⁴Die (gerundeten) Ziehungswahrscheinlichkeiten (Wahrscheinlichkeit für jedes Mitglied der Grundgesamtheit, ins SOEP gezogen zu werden) betragen: 0.0002 für die westdeutsche Stichprobe, 0.0008 für die Ausländerstichprobe, 0.0004 für die ostdeutsche Stichprobe, 0.0002 für die Immigrantstichprobe und 0.00003 für die Auffrischungstichprobe (Haisken-De New & Frick, 2000).

²⁵Im Osten wurden 1990 erfragt: Aktuelle Lebenszufriedenheit, Lebenszufriedenheit retrospektiv vor 5 Jahren, Lebenszufriedenheit prospektiv in 5 Jahren, sowie spezifische Zufriedenheiten mit Gesundheit, Haushaltseinkommen, Arbeit, Lebensstandard, Wohnung, Wohngegend, Angebot an Waren und Dienstleistungen, Umweltzustand.

3.1.3 Untersuchungsstichprobe der Analysen zur Zufriedenheit im höheren Lebensalter

Für die folgenden Analysen wurden lediglich die Daten der westdeutschen SOEP-Stichprobe genutzt. Die besondere Attraktivität des SOEP für die Untersuchung der Entwicklung der Zufriedenheit im höheren Erwachsenenalter liegt in der Möglichkeit, Zeitreihen von Zufriedenheitsmessungen über einen möglichst langen Zeitraum hinweg zu analysieren – die vollständige Kette von sechzehn jährlichen Meßwiederholungen liegt aber nur für die westdeutsche und die Ausländerstichprobe vor. Eine Einbeziehung letzterer ohne explizite Berücksichtigung der unterschiedlichen Ziehungswahrscheinlichkeiten aber würde das wahrscheinlichkeitstheoretische Zufallsmodell, das bei den anzuwendenden inferenzstatistischen Verfahren implizit zugrundegelegt ist, verletzen und müßte strenggenommen durch eine unterschiedliche Gewichtung der Fälle in der kombinierten Stichprobe ausgeglichen werden. Zwar werden entsprechende Gewichtungsfaktoren im SOEP geliefert, aber die damit verbundene praktische Umständlichkeit für die Analysen scheint schlicht unnötig, da es m.E. keinen zwingenden Grund dafür gibt, bei der Untersuchung der zum Ende des Theorierteils in Kapitel 1.3 aufgelisteten Hypothesen über mögliche längsschnittliche Entwicklungen der Zufriedenheit im höheren Lebensalter nicht nur die Population der Westdeutschen deutscher Staatsangehörigkeit, sondern unbedingt die Population der Westdeutschen *und* der in Westdeutschland lebenden ausländischen Staatsbürger zugrunde zu legen. Denn es geht bei den Analysen zunächst lediglich darum, die Hypothesen in einer möglichst breiten und unspezifischen Grundgesamtheit zu untersuchen. Aus demselben Grund erscheint es hier auch angemessen, die ostdeutsche Teilstichprobe nicht zu berücksichtigen: Weder wäre durch die Einbeziehung der Ostdeutschen eine globale Populationsvalidität erreicht, noch wurden Hypothesen formuliert, deren Überprüfung interkulturelle Vergleiche erfordert. Womit freilich nicht gesagt sein soll, daß solch ein Ost-West-Vergleich kein lohnendes Unterfangen sein könnte – er unterblieb hier eben nur im Sinne einer Beschränkung und Konzentration auf die in Kapitel 1.3 genannten Untersuchungsziele.

Tabelle 3.2 zeigt nochmals die Personenfallzahlen der westdeutschen Stichprobe bzw. deren Entwicklung im Verlauf der Erhebungszeitspanne von 1984 bis 1999. Wie zu erkennen ist bleiben, nach einem anfänglichen, v.a. von der ersten zur zweiten Welle, stärkeren Verlust von Fällen, die Fallzahlen relativ stabil, mit jährlichen Verlustraten, die meist deutlich unter 5% liegen. Allerdings entsprechen diese Verlustraten nicht den Dropouts, d.h. den Stichprobenausfällen, die in jeder Welle zu verbuchen sind, denn diese Dropouts werden teilweise durch „Dropins“, d.h. durch neu hinzugekommene Befragungspersonen kompensiert – nämlich durch Jugendliche, die das 16. Lebensjahr erreicht haben, sowie Personen, die in die SOEP-Befragungshaushalte gezogen sind, und Rückkehrer, d.h. temporäre Dropouts, die bei der vorherigen Befragung ausgefallen waren. Von der wiederum etwas erhöhten Anzahl von Ausfällen von der ersten zur zweiten Befragung abgesehen, bewegen sich die Dropoutstraten im Bereich zwischen fünf und zehn Prozent – insgesamt kann da-

mit dem SOEP ein durchaus niedriger Stichprobenausfall bescheinigt werden, wenn man in Rechnung stellt, daß solche Ausfälle in längsschnittlichen Befragungen praktisch unvermeidbar sind: Neben der Verweigerung, welche auch bei querschnittlichen Befragungen zur Reduktion der ursprünglich gezogenen Stichprobe führen kann, können hier bei den Wiederholungsbefragungen auch Ausfälle wegen der Nichterreichbarkeit (z.B. infolge von Umzug zwischen zwei Meßzeitpunkten) oder zwischenzeitlichen Mortalität der Teilnehmer entstehen.

V.a der letztgenannte Ausfallgrund, die Mortalität von Teilnehmern im Panelverlauf, ist für die nachfolgenden Analysen im Bereich der Höheraltrigkeit von grundsätzlichem Interesse, da in diesem Spektrum besonders viele Todesfälle zu erwarten sind. Diese Ausfälle sind „datentechnisch“ prinzipiell anders zu bewerten, als die sonstigen Dropouts, da es bei letzteren ja eigentlich diejenigen Werte der untersuchten Variablen, die infolge des Ausfalls *im Datensatz* fehlen, gibt, während dieses bei verstorbenen Personen nicht der Fall ist. Die statistischen Verfahren zum Umgang mit fehlenden Werten sind v.a. auf das Problem der Vermeidung von Ergebnisverfälschung durch die fehlenden *existenten* Werte ausgerichtet und wären deshalb hinsichtlich der der Stichprobenausfälle infolge Mortalität genaue genommen anders zu bewerten. Dieses wurde bereits im Kapitel 2.3.1 besprochen, hier geht es lediglich darum, bei der allgemeinen Darstellung des den weiteren Analysen zugrundegelegten Datensatzes auch über die darin auftretenden Todesfälle zu informieren.

Diese Information ist allerdings nicht problemfrei, da nicht alle Todesfälle als solche erfaßt sind: Diese Erfassung setzt zunächst einmal voraus, daß bei der erneuten Kontaktaufnahme mit dem SOEP-Befragungshaushalt die Information über das zwischenzeitliche Ableben eines Haushaltsmitglieds in irgendeiner Weise erhoben werden konnte. Bei völliger Kontaktverweigerung des gesamten Haushaltes oder wenn die Kontaktaufnahme gänzlich mißlingt, kann eine solche Information nicht erhoben werden und so ist damit zu rechnen, daß es bei den Verweigerern und „verlorenen“ Personen des SOEP noch einmal eine gewisse Anzahl von Todesfällen gibt. Gerade bei hochaltrigen Personen könnten Verschlechterungen der gesundheitlichen Situation, die dem Tod vorausgehen, ein bedeutsamer Grund für die weitere Teilnahmeverweigerung oder das Scheitern der Kontaktaufnahme (z.B. wegen längerer Krankenhausaufenthalte oder Umzug in Pflegeheime) sein. Zur Entdeckung der nicht erfaßten Todesfälle wurde 1992 eine Verbleibstudie durchgeführt: Durch eine Adress-Recherche bei den deutschen Einwohnermeldeämtern konnte der weitere Verbleib fast aller Ausfälle mit vorher unbekanntem Status aufgeklärt werden, so daß die Datensätze bis zum Erhebungsjahr 1992 entsprechend ergänzt werden konnten und weitestgehend vollständige Informationen über die Todesfälle enthalten. Eine zweite solche Verbleibstudie wurde im Jahr 2001 durchgeführt, ihre Ergebnisse sind aktuell publiziert (Infratest Sozialforschung, 2002), jedoch sind die daraus gewonnenen Informationen noch nicht in den hier analysierten Datensätzen ergänzt.

Über die Todesfälle in der westdeutschen SOEP-Stichprobe informiert Tabelle 3.3: Ab dem Jahr 1993 sind die Todesfälle bei den Stichprobenausfällen mit unbekanntem Verbleib

TABELLE 3.2:
Entwicklung der westdeutschen SOEP-Stichprobe auf Personenebene (bis 1999)

Jahr	Teilnehmer Vorjahr N	Ausfälle bei Teilnehmern im Vorjahr		Kompensation						Saldo	
		N	%	Kinder		Zugezogene		Rückkehrer		N	%
1985	9070	1079	11.9	179	2.0	172	1.9	29	0.3	-699	-7.7
1986	8371	817	9.8	177	2.1	167	2.0	111	1.3	-362	-4.3
1987	8008	561	7.0	148	1.8	143	1.8	129	1.6	-141	-1.8
1988	7868	690	8.8	135	1.7	104	1.3	64	0.8	-387	-4.9
1989	7481	615	8.2	115	1.5	115	1.5	105	1.4	-280	-3.7
1990	7201	495	6.9	88	1.2	124	1.7	120	1.7	-163	-2.3
1991	7036	401	5.7	83	1.2	102	1.4	153	2.2	-63	-0.9
1992	6974	431	6.2	74	1.1	109	1.6	95	1.4	-153	-2.2
1993	6821	397	5.8	74	1.1	104	1.5	145	2.1	-74	-1.1
1994	6747	432	6.4	70	1.0	112	1.7	140	2.1	-110	-1.6
1995	6637	381	5.7	87	1.3	102	1.5	122	1.8	-70	-1.1
1996	6567	389	5.9	86	1.3	95	1.4	95	1.4	-113	-1.7
1997	6454	383	5.9	80	1.2	94	1.5	133	2.1	-76	-1.2
1998	6378	460	7.2	91	1.4	85	1.3	90	1.4	-194	-3.0
1999	6184	442	7.1	83	1.3	97	1.6	123	2.0	-139	-2.2

Prozentangaben bezogen auf Anzahl Vorjahresteilnehmer.

Quelle: Infratest Sozialforschung (2001).

nicht enthalten, für die Jahre zuvor zeigt die Tabelle auch die entsprechende Aufschlüsselung nach der Informationsquelle. Es zeigt sich dabei eine jeweils nur sehr geringe Anzahl von nachträglich aus der Verbleibestudie aufgeklärten im Vergleich zu den aus der Befragung direkt aufgeklärten Todesfällen. Interessanterweise handelt es sich bei den aus der Befragung aufgeklärten Todesfällen in der Regel um solche, die jeweils im Jahr zuvor noch befragt werden konnten und deren Tod somit nicht mit einem vorherigen Stichprobenausfall verbunden war, während die aus der Verbleibestudie erfaßten Todesfälle zum Teil erst längere Zeit nach dem Stichprobenausfall eintraten (vgl. auch Infratest Sozialforschung, 2002). Tabelle 3.3 zeigt dazu die Gesamtzahl der jeweils aufgeklärten Todesfälle mit einer im Vorjahr erfolgreich durchgeführten Befragung an: Es zeigt sich, daß dieses beim größten Teil der Verstorbenen der Fall war und jeweils nur bei ungefähr zwischen fünf und fünfzehn Prozent aller Todesfälle ein Stichprobenausfall im Vorjahr stattfand.²⁶

²⁶Anzumerken ist, daß in der erwähnten neueren Verbleibestudie eine erheblich größere Anzahl von 720 Todesfällen bei den SOEP-Dropouts nachträglich festgestellt wurde, allerdings wurden hier sämtliche SOEP-Dropouts aus allen Teilstichproben und allen Erhebungsjahren 1984 bis 2001 nachverfolgt (Infratest Sozialforschung, 2000). In gewisser Weise illustriert dieses Ergebnis den Umstand, daß natürlich alle Dropouts irgendwann einmal sterben.

TABELLE 3.3:
Todesfälle in der westdeutschen SOEP-Stichprobe (bis 1999)

Jahr	Sterbefälle N	Informationsquelle				Befragung im Vorjahr		Sterbealter 60+	
		Befragung N	%	Verbleibestudie N	%	N	%	N	%
1985	71	63	88.7	8	11.3	68	95.8	56	78.9
1986	94	87	92.5	7	7.5	85	90.4	81	86.2
1987	90	81	90.0	9	10.0	75	83.3	69	76.7
1988	97	90	92.8	7	7.2	90	92.8	79	81.4
1989	91	82	90.1	9	9.9	79	86.8	78	85.7
1990	72	70	97.2	2	2.8	63	87.5	60	83.3
1991	85	82	96.5	3	3.5	74	87.1	73	85.9
1992	69	68	98.5	1	1.5	64	92.6	63	91.3
1993	81	81				72	88.9	66	81.5
1994	63	63				59	93.6	57	90.5
1995	98	98				82	83.7	78	79.6
1996	68	68				64	94.1	53	77.9
1997	56	56				47	83.9	49	87.5
1998	58	58				52	89.7	48	82.7
1999	57	57				53	93.0	51	89.5

Um der Vollständigkeit der Beschreibung des höheraltrigen Teils der untersuchten SOEP-Stichprobe willen zeigt die Tabelle 3.3 in den letzten beiden Spalten noch die Anzahlen bzw. prozentualen Anteile der Todesfälle höheraltriger Personen, d.h. solche, bei denen der Verstorbene mindestens 60 Jahre alt war (vgl. Kapitel 1.2, S. 39). Wie nicht anders zu erwarten, fällt der größte Teil aller Todesfälle in diesen Altersbereich.

3.1.4 Fazit: Zur Eignung des SOEP für längsschnittliche Analysen zur Zufriedenheit im höheren Lebensalter

Für die eigenen Untersuchungen eignet sich das SOEP v.a. aufgrund der darin vorhandenen vollständigen Zeitreihen von im Ein-Jahres-Abstand durchgeführten Zufriedenheitsbefragungen. Es wurde vorne (Kapitel 1.2.2 und 1.3) dargelegt, daß für die angestrebten Analysen eine möglichst lange Zeitreihe mit möglichst vielen Meßwiederholungen benötigt wird, um nicht nur Veränderung der Zufriedenheitswerte erfassen zu können, sondern auch „Veränderung der Veränderung“, d.h. die längsschnittliche Entwicklung z.B. der Stabilitätskoeffizienten Älterer (vgl. die Hypothese erhöhter Merkmalsstabilität). Zudem handelt es sich um eine „breitgestreute“ Stichprobe, die die Gesamtbevölkerung der Bundesrepublik Deutschland (bzw., bei der im folgenden verwendeten Substichprobe, des westdeutschen Teils) sehr gut repräsentieren dürfte – es wurde bei der Ersterhebung eine global repräsen-

tative Haushaltsstichprobe gezogen. Diese Breite hat für die folgenden Analysen nicht nur den Vorteil ausreichender Fallzahlen, sondern ermöglicht auch gegebenenfalls Vergleiche älterer mit jüngeren Befragungspersonen und verhindert eine allzu große Selektivität der in dieser Befragung einbezogenen Alten.

Mit diesen Gesichtspunkten sind m.E. Vorteile des SOEP beschrieben, die den Datensatz vor allen anderen verfügbaren und einigermaßen aktuellen deutschen Datensätzen auszeichnen. Gegenüber den bedeutenden gerontologischen deutschen Längsschnittstudien – BOLSA (Thomae, 1993), BASE (Mayer & Baltes, 1996), ILSE (Martin et al., 2000), sowie Alterssurvey (Hoff et al., 2003) – bietet es den Vorteil der langen Zeitreihe mit vielen Meßwiederholungen, bei dem allerdings selbstverständlichen Nachteil, daß es sich eben beim SOEP um keine spezifisch gerontologische Studie handelt und auch Fragen der Wohlbefindensforschung bei der Konzeption des SOEP sicherlich nicht zu den vordringlichen Interessenschwerpunkten gezählt wurden.

Dieses bedeutet schlicht, daß nicht viele psychogerontologisch und/oder zufriedenheitstheoretisch relevante Instrumente und Maße im SOEP enthalten sind. Bezüglich der Zufriedenheit sind dieses eben lediglich die oben aufgeführten Ein-Item-Messungen, es fehlen gebräuchliche Mehr-Item-Maße der Lebenszufriedenheit. Immerhin aber kann hier an im Kapitel 1.1.4.2 Erwähntes erinnert werden: In seinem Review verweist Veenhoven (1996) auf Befunde zur durchaus befriedigenden Reliabilität solcher Ein-Item-Messungen, für die Zufriedenheitsfragen des SOEP konnten Landua (1993) und Schräpler (1995) durchaus befriedigende Reliabilitäten für diese Messungen schätzen.

Andere, dem SOEP inhaltlich vergleichbare, breitangelegte sozialwissenschaftliche Umfragedatensätze mit wiederholten Befragungswellen sind in Deutschland z.B. die Allgemeine Bevölkerungsumfrage der Sozialwissenschaft (ALLBUS, zur Information vgl.: Braun & Mohler, 1991; oder die ALLBUS-Homepage, derzeit unter: <http://www.gesis.org/Dauerbeobachtung/Allbus/>) oder der Wohlfahrtssurvey (vgl. Zapf & Habich, 1996). Dabei aber handelt es sich nicht um echte Längsschnittdatensätze, sie eignen sich somit nicht für die angestrebten Analysen.

Zusammenfassend kann somit gesagt werden, daß das SOEP eine Datenbasis für längsschnittliche Analysen zur Entwicklung von Zufriedenheit im höheren Alter bietet, deren besondere Vorteile im langen Untersuchungszeitraum, der großen Anzahl von Wiederholungsmessungen und der breiten Personenstichprobe bestehen. Die Nachteile eines solchen sekundäranalytischen Rückgriffs auf diese Datenquelle bestehen in der geringen Spezifität der Daten für die eigene inhaltliche Fragestellung, d.h. in dem Umstand, daß Indikatoren des SWB bzw. der Lebenszufriedenheit nur in begrenztem Umfang erhoben wurden.

3.2 Analysen zur Entwicklung durchschnittlicher Zufriedenheitswerte

Im folgenden Kapitel werden Auswertungsschritte zusammengefaßt, bei denen der Verlauf bzw. die Entwicklung durchschnittlicher Zufriedenheitswerte über verschiedene mögliche Zeitdimensionen der Untersuchung – nämlich Meßzeitpunkt, Lebensalter oder Geburtszeitpunkt – hinweg untersucht werden. Diese können somit prinzipiell unterschieden werden von denjenigen Analysen, die auf der längsschnittlichen Modellierung intraindividuelle Merkmalsveränderung beruhen und welche später (im Kapitel 3.3) dargestellt werden.

Es werden zunächst die durchschnittlichen Zufriedenheiten von Lebensalters- und/oder Geburtsjahrganggruppen bei den einzelnen SOEP-Befragungen rein deskriptiv untersucht (Unterkapitel 3.2.1). Dabei soll der Befund des Zufriedenheitsparadoxes repliziert bzw. ganz allgemein ein differenziertes Bild des typischen Verlaufs durchschnittlicher Zufriedenheiten über die höhere Lebensaltersspanne hinweg gewonnen und möglicherweise bestehende Unterscheide zwischen quer- und längsschnittlichen Mittelwertsverläufen erkannt werden. Danach werden Analysen zur Kohorteneffekthypothese durchgeführt: Auch die in dieser Hypothese vermuteten Alters- und Kohorteneffekte betreffen die Entwicklung durchschnittlicher Zufriedenheiten über das Lebensalter bzw. die Geburtsjahrgänge hinweg.

3.2.1 Deskriptive Analysen zum Zusammenhang von Alter und Zufriedenheit im höheren Erwachsenenalter

Die empirischen Befunde, die als Zufriedenheitsparadox des höheren Lebensalters bezeichnet wurden und die im Kapitel 1.2 nachgezeichnete gerontologische Debatte angestoßen haben, betreffen, wie dort bereits dargelegt wurde, den Zusammenhang zwischen Lebensalter und Lebenszufriedenheit und d.h. die durchschnittlichen Zufriedenheiten von Altersgruppen gleichaltriger Personen: Das „Paradox“ besteht darin, daß die Zufriedenheitsmittelwerte pro Lebensalter – und d.h. in der Regel: innerhalb von durch Ein-Jahres-Altersintervalle definierten Personengruppen – im höheren Lebensalter stabil bleiben und nicht sinken. Es erscheint deshalb sinnvoll, noch vor allen Analysen zur Untersuchung der Hypothesen, die hier als Erklärungsmöglichkeiten für dieses Zufriedenheitsparadox formuliert wurden, den Befund desselben im verwendeten Datensatz zu replizieren bzw. einen zumindest deskriptiven Eindruck davon zu gewinnen, inwieweit auch hier die Zufriedenheitsmittelwerte über die höhere Lebensaltersphase hinweg invariant sind und wie sich der Zusammenhang von Mittelwerten und Lebensalter innerhalb dieser Phase entwickelt.

Denn die meisten der zitierten Untersuchungen vergleichen entweder für breite Altersintervalle definierte Altersgruppen (z.B. Doyle & Forehand, 1984; Inglehard 1990), oder erfassen den Zusammenhang zwischen Lebensalter und Zufriedenheit durch die Produkt-Moment-Korrelation über die gesamte Erwachsenenaltersspanne hinweg (z.B. Herzog & Rodgers, 1981). Bei beiden Vorgehensweisen können aber die Zusammenhänge zwischen

Alter und Zufriedenheit *innerhalb* der Phase des höheren Lebensalters kaum differenziert erfaßt werden. Denn bei einer breit definierten Gruppe Höheraltriger (z.B. 65+) sind, entsprechend der Verteilung des Alters in der Population, in der Regel relativ wenige „ältere Alte“ enthalten, so daß der Durchschnittswert der Gruppe eher die Zufriedenheit während der früheren Phase der Höheraltrigkeit widerspiegelt.²⁷ Bei der Analyse von Produkt-Moment-Korrelationswerten aber könnten aufgrund des dabei impliziten *linearen* Zusammenhangsmodells eventuelle Nichtlinearitäten zum Ende der Lebensspanne hin (z.B. Zufriedenheitsverluste im sehr hohen Alter) nicht erfaßt worden sein, zumal auch dabei wieder dieser Altersabschnitt in der Regel nur durch relativ wenige Fälle im Datensatz vertreten ist. Es ist deshalb m.E. aus den publizierten Untersuchungsergebnissen, bei aller Eindeutigkeit bezüglich des Befunds von noch bis ins höhere Alter hinein stabilen Mittelwerten, nicht eindeutig herauszulesen, ob die Stabilität altersspezifischer Durchschnittswerte bis ins höchste Alter hinein besteht, oder nur für eine gewisse Phase früher Hochaltrigkeit. Beispielsweise zeigen Befunde der Berliner Altersstudie durchaus einen gewissen, allerdings nicht statistisch signifikanten, Abfall durchschnittlicher Zufriedenheit in der Altersgruppe 75-79 (und der älterer Altersgruppen) gegenüber der Gruppe 70-74 (Smith et al., 1996). Durch die genaue Deskription durchschnittlicher SOEP-Zufriedenheitswerte soll auch ein differenzierteres Bild der vermeintlichen Mittelwertstabilität im höheren Alter gewonnen werden.

Ein solchermaßen genauestmögliches Bild ergibt sich bei Betrachtung der Mittelwerte der im SOEP „engstmöglichen“, d.h. der Ein-Jahres-Altersgruppen, was allerdings zu einer kaum noch überschaubaren Masse von Mittelwerten führt, die darum hier nur im Anhang B im Detail wiedergegeben wird: Die Mittelwerte der allgemeinen Lebenszufriedenheit aller Ein-Jahres-Altersgruppen ab dem fünfundvierzigsten Lebensjahr für jede SOEP-Befragung von 1984 bis 1999 sind dort in der Tabelle B.3 aufgelistet. Die entsprechenden Mittelwerte der Einkommens-, Gesundheits-, Wohn- und Freizeitzufriedenheit können in den Tabellen B.4 bis B.7 nachgesehen werden. Die Beschränkung auf die Altersgruppen ab 45 erfolgte zur Begrenzung der in diesen Tabellen sowieso schon ausufernden Informationsmenge, der Beginn mit diesem Lebensjahr erschien deshalb sinnvoll, weil die zum Panelbeginn 45-Jährigen die jüngsten Teilnehmer sind, die im Verlauf des untersuchten Zeitraums bis 1999 noch die Altersgrenze 60 erreicht haben, welche vorne (S. 39) als Kriterium der Höheraltrigkeit festgelegt wurde. Auch sind zur weiteren Begrenzung der Informationsflut in den Tabellen Mittelwerte dann nicht mehr aufgeführt, wenn sie aus weniger als zehn Werten berechnet wurden, wenn also in der entsprechende Zelle in einer der Tabellen nur noch höchstens neun Werte vorhanden waren. Dieses führt bei allen Zufriedenheiten dazu, daß für alle Altersgruppen über 90 kein Mittelwert mehr aufzulisten ist. Die leeren Zellen in den

²⁷Auch sollte prinzipiell daran gedacht werden, daß bei Vergleichen enger gefaßter Altersgruppen dann, wenn nicht disproportional geschichtete Stichproben dieser engen Altersgruppen gezogen wurden, die geringen Stichprobenumfänge bei den „älteren Alten“ zu geringerer Teststärke der entsprechenden Mittelwertvergleiche führt und möglicherweise deshalb auch größere Mittelwertsdifferenzen nicht statistisch signifikant werden.

Tabellen bedeuten also nicht, daß darin keine Werte mehr vorhanden waren, – sie waren im Gegenteil alle noch besetzt, nur eben mit weniger als zehn Fällen. Tabelle B.8 zeigt diese entsprechenden Zellhäufigkeiten der Werte allgemeiner Lebenszufriedenheit, für die anderen Zufriedenheiten ergeben sich allenfalls gelegentlich marginal abweichende Zellhäufigkeiten, so daß diese nicht eigens tabelliert wurden. Diese Häufigkeiten werden auch noch für die Planung der Stabilitätsanalysen von Belang sein.

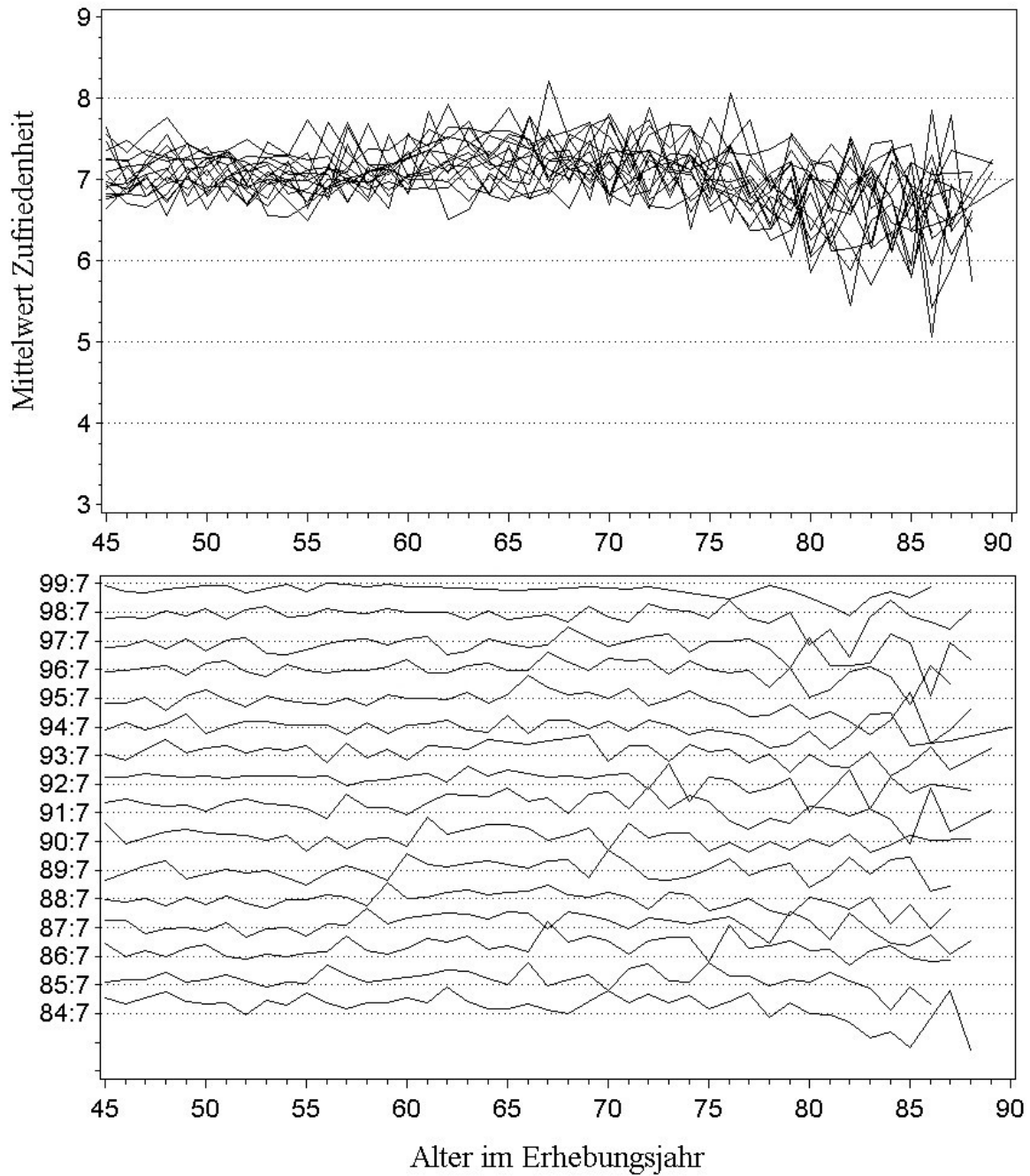
Die Tabellen B.3 bis B.7 wurden lediglich um der Vollständigkeit und Dokumentation willen erstellt, tatsächlich ist die bloße Auflistung dieser vielen Mittelwerte wohl kaum geeignet, das oben geforderte genaue Bild der Mittelwertstabilität im höheren Lebensalter zu vermitteln. Als diesbezüglich geeigneteres Mittel erscheinen graphische Aufbereitungen der in diesen Tabellen enthaltenen Informationen, welchen im folgenden der Vorzug gegeben wird. Insbesondere dabei auch auf jegliche „ritualisierte“ Signifikanztestung verzichtet (Sedlmeier, 1996), da solche m.E. hier kaum etwas zum eigentlichen Analyseziel der *Deskription* der zeitlichen Entwicklungen durchschnittlicher Zufriedenheitswerte beitragen und eher von den eigentlich interessierenden Effekten ablenken würden.

3.2.1.1 Querschnittliche Zusammenhänge von Alter und Zufriedenheit

Zunächst wären, im Sinne der Replikation von Befunden des Zufriedenheitsparadoxes, welche wie dargestellt in der Regel auf der querschnittlichen Untersuchung des Zusammenhangs von Lebensalter und Zufriedenheit beruhen, die Entwicklungen der Mittelwerte über die Ein-Jahres-Altersgruppen hinweg, die sich bei querschnittlicher Betrachtung der SOEP-Daten ergeben, darzustellen. Bei 16 SOEP-Erhebungen sind also 16 solche Querschnitte – über die jeweils 16 Spalten der Tabellen B.3 bis B.7 – möglich.

Eine graphische Darstellung der 16 querschnittlichen Verläufe der Mittelwerte allgemeiner Lebenszufriedenheit zeigt Abbildung 3.1: Im oberen Teil sind die 16 Verlaufskurven der Mittelwerte über die Altersgruppen hinweg der 16 SOEP-Wellen abgebildet (16 Polygonzüge über jede der Spalten in Tabelle B.3). Es entsteht sozusagen ein Knäuel sich überlagernder Verläufe, die einzeln kaum noch nachzuverfolgen sind. Dennoch besitzt dieses Knäuel einige Aussagekraft: Zunächst ist bis ungefähr zum fünfundsiebzigsten Lebensjahr keine Abwärtstendenz erkennbar, sondern sogar eher eine Tendenz geringfügigen Anstieges, mit einem kleinen Gipfel ungefähr beim fünfundsechzigsten Lebensjahr. Nach dem Alter 75 beginnt das Knäuel allerdings abzufallen, die Verläufe variieren hier auch stärker. Letzteres dürfte v.a. dem Umstand geschuldet sein, daß es sich hier zusehends um Mittelwerte handelt, die nur noch aus sehr wenigen Zufriedenheitswerten errechnet wurden (vgl. Tabelle B.8 im Anhang B): Die größere Variation der Mittelwerte entspricht hier dem infolge geringerer Stichprobenumfänge größeren Standardfehler derselben (als Schätzer der entsprechenden Populationswerte). Abgesehen von einigen Ausschlägen der Verlaufskurven v.a. zum Ende der Altersspanne bewegen sich die Mittelwerte innerhalb eines sehr engen Bereichs von ungefähr 6.7 bis 7.7. Im unteren Teil der Abbildung ist nun sozusagen das

ABBILDUNG 3.1:
Allgemeine Lebenszufriedenheit: Querschnittliche Mittelwertsverläufe über das Lebensalter



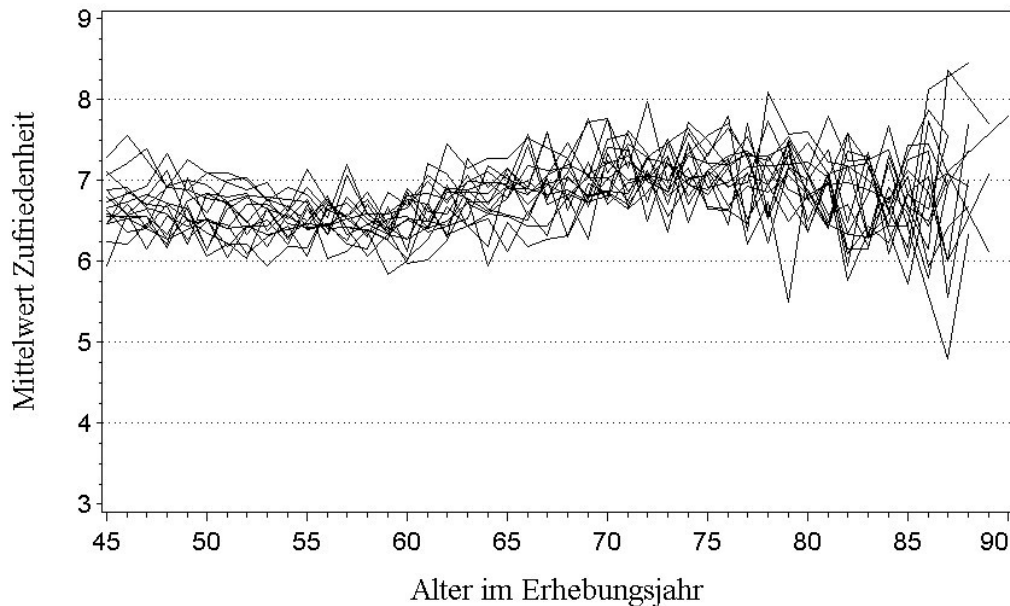
Knäuel entwirrt: Dazu wurden einfach die Mittelwerte jeder SOEP-Welle t um $t + 1$ nach oben verschoben, so daß also die unterste Verlaufskurve diejenige des Jahres 1984 ist, die nächste dann die um 1 nach oben verschobene des Jahres 1985 usw. bis zur obersten Kurve, die den Verlauf der Mittelwerte von 1999 abbildet. Die gepunkteten Linien markieren hier jeweils den Zufriedenheitsskalenwert 7 für die zugehörige Verlaufskurve. Die einzelnen Verlaufskurven zeigen keine optischen Auffälligkeiten und bestätigen den Eindruck, den das gesamte Knäuel vermittelt: Sie verlaufen stets mit einem einigermaßen waagerechten Trend bis über das Lebensjahr 75 hinaus und zeigen, wenn überhaupt, dann erst zum rechten Rand hin eine abfallende Tendenz mit Verläufen unter der jeweiligen gepunkteten Linie (die vertikale Auflösung ist in dieser Teilabbildung gering, so daß die Ausschläge in diese Richtung ebenfalls eher geringfügig erscheinen).

Insgesamt belegt die Abbildung 3.1 die Replikation des Zufriedenheitsparadox in den SOEP-Daten: Im jeweiligen Querschnitt zeigt sich Mittelwertsstabilität bis ins höhere Alter hinein. Allerdings nicht bis zum Ende der gesamten Altersspanne: Eine leichte Tendenz abfallender durchschnittlicher Lebenszufriedenheitswerte ist ungefähr ab dem fünfundsiebzigsten Lebensjahr erkennbar.

Es ist noch zu erwähnen, daß ein gemäß der Kohorteneffekthypothese bestehender Kohorteneffekt sich in Abbildung 3.1 in tendenziellen Abständen der Verlaufskurven voneinander zeigen müßte: Ein systematischer Effekt steigender Zufriedenheit mit sinkendem Geburtsjahr würde dazu führen, daß die Mittelwerte späterer Erhebungswellen unter den Mittelwerten der jeweils Gleichaltrigen zu früheren Erhebungen liegen (also müßte auch in den Werten der Tabelle B.3 im Anhang B jeweils über die Zeilen hinweg ein tendenzieller Abfall ablesbar sein). Deshalb sollten mit wachsendem Jahr der SOEP-Erhebung die Verlaufskurven insgesamt, in ihrem „geglätteten“ Trend, niedriger liegen. Dieses ist aus der Abbildung 3.1 natürlich so genau nicht herauszulesen, immerhin aber zeigt das Knäuel eine gewisse Breite, in der sich geringe Abstände zwischen den Verläufen andeuten und es ist zu bedenken, daß die inkrementellen Zufriedenheitsverluste *pro Geburtsjahr*, die bei einem über mehrere Geburtsjahrzehnte hinweg wirksamen Kohorteneffekt zu erwarten wären, ja nur sehr gering (in Relation zur Streuung der Meßwerte) sein könnten. Zur Untersuchung eines möglichen Kohorteneffekts reicht die Betrachtung graphischer Darstellungen nicht aus, Abbildung 3.1 zeigt keinen Kohorteneffekt, sie zeigt aber auch nicht, daß ein solcher nicht vorhanden ist.

Für die vier untersuchten Bereichszufriedenheiten werden nur die dem oberen Teil der Abbildung 3.1 entsprechenden Darstellungen der „Verlaufskurvenknäuel“ gezeigt, da es hier ja vornehmlich um die Replikation von Befunden zur Lebenszufriedenheit geht und die Bereichszufriedenheiten nur zum Vergleich mit letzterer bzw. zur Veranschaulichung möglicher Ähnlichkeiten mit derselben betrachtet werden. Die Verlaufskurven der Einkommenszufriedenheit sind in der Abbildung 3.2, die der Gesundheitszufriedenheit in Abbildung 3.3, die der Wohnzufriedenheit in Abbildung 3.4 und die der Freizeitzufriedenheit sind in Abbildung 3.5 dargestellt. Sie zeigen jeweils typische Verläufe, die, wenn auch

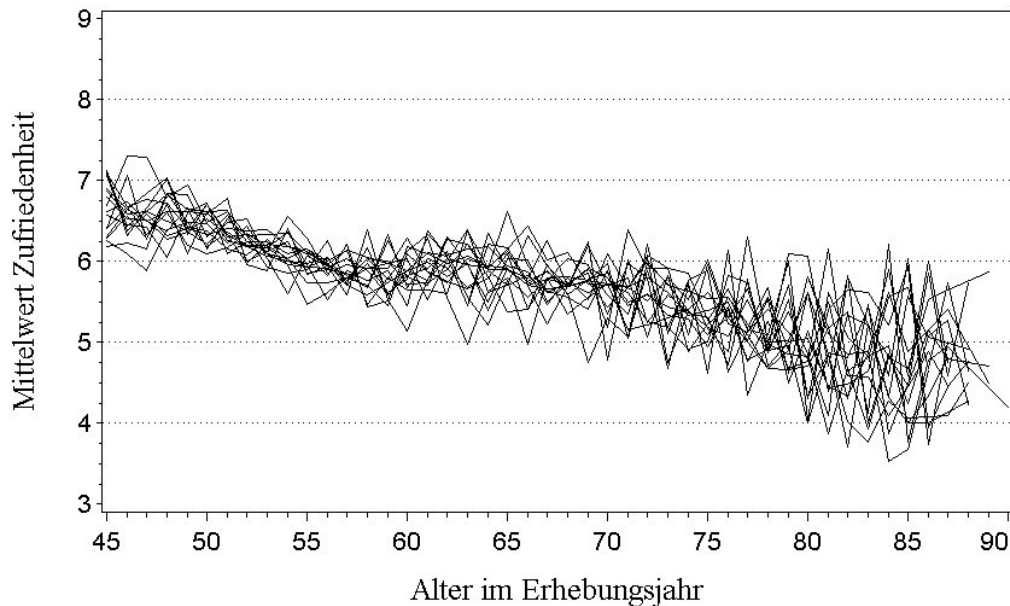
ABBILDUNG 3.2:
Einkommenszufriedenheit: Querschnittliche Mittelwertsverläufe über das Lebensalter



weniger für die hier fraglichen Untersuchungshypothesen zum Zufriedenheitsparadox, so doch immerhin von allgemeinem gerontologischem Interesse sein mögen. Für die Einkommenszufriedenheit (Abbildung 3.2) scheint v.a. ein Anstieg über das siebzigste Lebensjahrzehnt hinweg charakteristisch, nach einer zuvor eher fallenden Tendenz bis ungefähr zum sechzigsten Lebensjahr. Bei der Gesundheitszufriedenheit (Abbildung 3.3) zeigt sich ein wenig überraschender deutlicher Abfall in den späteren Jahren, beginnend ungefähr ab 65. Davor zeigt das Knäuel allerdings einige mittelwertsstabile Jahre, eine weitere Phase fallender Mittelwerte ist zum Beginn des abgebildeten Altersspanne bis ungefähr zum Alter 58 zu erkennen. Die Gesundheitszufriedenheit zeigt auch die im Vergleich zu den anderen Zufriedenheiten niedrigsten Mittelwerte, die ungefähr ab dem fünfundfünfzigsten Lebensjahr tendenziell stets unter 6 liegen. Die Wohnzufriedenheitsmittelwerte (Abbildung 3.4) steigen dagegen von Anfang der untersuchten Altersspanne bis wiederum ungefähr zum Lebensjahr 75 kontinuierlich leicht an und befinden sich insgesamt auf einem sehr hohen Niveau (die Mehrzahl der Werte > 8). Die Freizeitzufriedenheit (Abbildung 3.5) offenbart das vielleicht überraschendste Bild eines sehr deutlichen Anstiegs vom zuvor niederen Niveau (< 7) während des mittleren Erwachsenenalters auf ein sehr hohes Niveau (> 8) über die Spanne früher Höheraltrigkeit hinweg (ca. fünfundfünfzigstes bis siebzigstes Lebensjahr), danach erscheinen die Mittelwerte noch über ein gutes Jahrzehnt stabil.

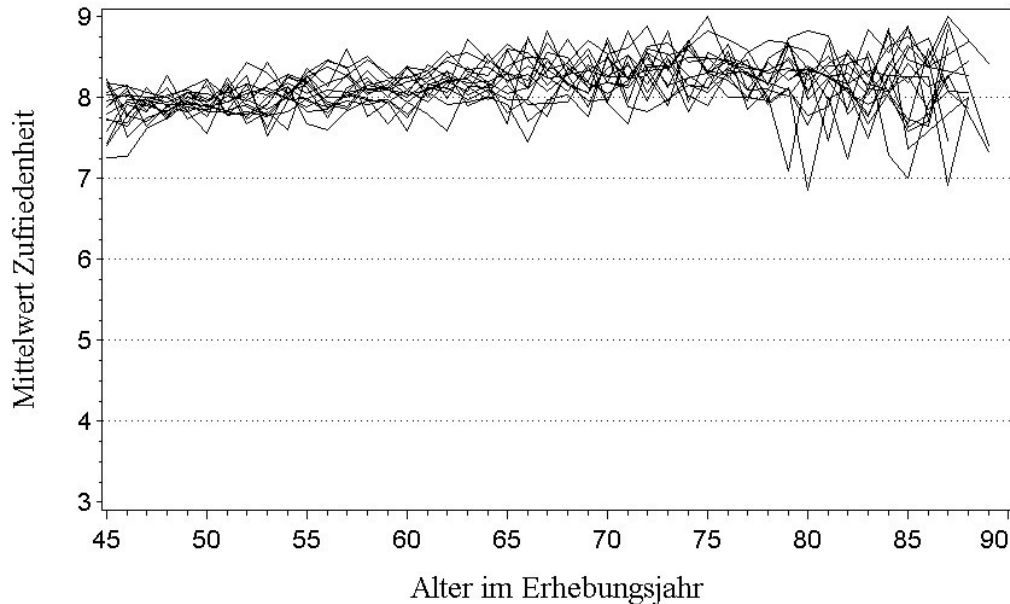
Sucht man nach Ähnlichkeiten in den Verlaufsformen dieser Knäuel, so findet sich v.a.

ABBILDUNG 3.3:
Gesundheitszufriedenheit: Querschnittliche Mittelwertsverläufe über das Lebensalter



der Befund von weit ins höhere Lebensalter hinein *nicht* fallenden Zufriedenheiten in allen Bereichen mit Ausnahme der Gesundheitszufriedenheit. In den Bereichen Einkommen, Wohnen und Freizeit scheint ein Knick nach unten sogar erst ab ungefähr dem achtzigsten Lebensjahr sichtbar, also in jenem Altersspektrum, in dem die entsprechenden Altersgruppenmittelwerte infolge reduzierter Fallzahlen die Verhältnisse in der Population immer unzuverlässiger (im Sinne eines erhöhten Standardfehlers) wiedergeben – womit auch die zum rechten Rand hin wachsende Breite der Knäuel, welche immer größere Sprünge der aufeinanderfolgenden Mittelwerte innerhalb der Verlaufskurven erkennen läßt, erklärt werden kann. Es kann also im Grunde nicht sicher behauptet werden, daß für diese Zufriedenheiten überhaupt ein relevanter Abfall während irgendeiner Phase des höheren Lebensalters gefunden wurde. Die Lebenszufriedenheit unterscheidet sich in dieser Hinsicht von den vorgenannten nur leicht dadurch, daß ein schwacher finaler Trend nach unten schon etwas früher ab ungefähr dem fünfundsiebzigsten Lebensjahr erkennbar wird. Die Zufriedenheit mit der Gesundheit bildet hier die Ausnahme: Der Knick nach unten beginnt früher und ist erscheint deutlicher ausgeprägt. Mit großer Spitzfindigkeit und sozusagen ausgeprägten Gestalterkennungsfähigkeiten könnte man die deutlichste Ähnlichkeit zum Knäuel der allgemeinen Lebenszufriedenheit bei der Einkommenszufriedenheit erkennen, nämlich in dem „Tief“ um das fünfundfünfzigste Lebensjahr und dem Umschwung nach unten ab ungefähr dem fünfundsiebzigsten, allerdings ist im Bereich Einkommen ersteres stärker und letzterer schwächer erkennbar, als bei der Lebenszufriedenheit.

ABBILDUNG 3.4:
Wohnzufriedenheit: Querschnittliche Mittelwertsverläufe über das Lebensalter

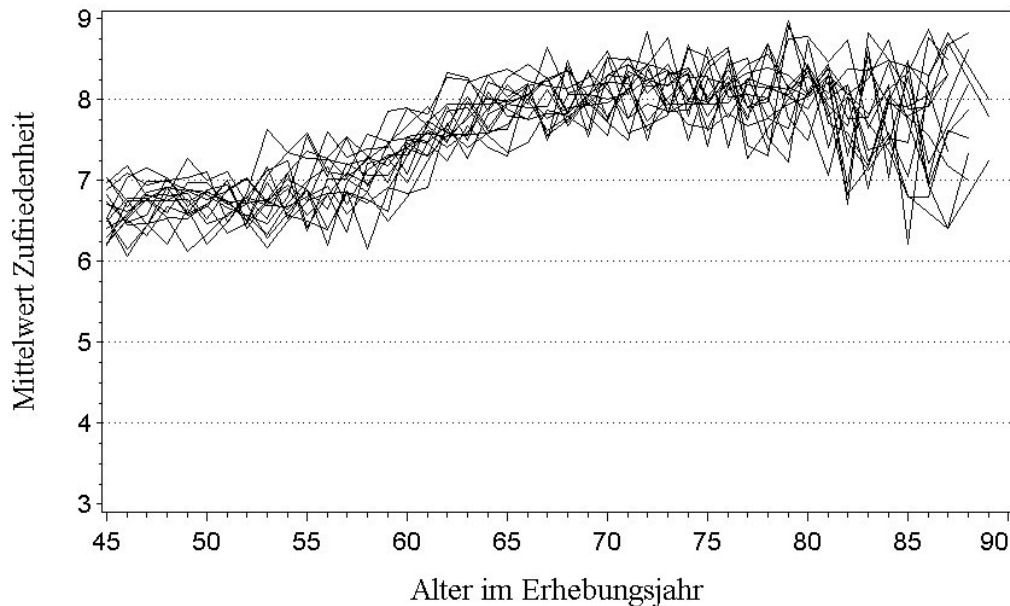


3.2.1.2 Längsschnittliche Zusammenhänge von Alter und Zufriedenheit

Die querschnittliche Betrachtung offenbarte somit, von der Gesundheitszufriedenheit abgesehen, kaum negative Alterseffekte auf die Zufriedenheitsmessungen und bestätigte den aus vielen Studien bekannten Befund von über die verschiedenen Altersgruppen hinweg bis ins hohe Alter hinein „paradox“ stabiler durchschnittlicher Lebenszufriedenheit. Wie im Theorieteil bereits überlegt, müßte sich ein solcher Alterseffekt beim querschnittlichen Vergleich gar nicht unbedingt zeigen, wenn er durch Kohorteneffekte entsprechend überlagert wird. Anders bei längsschnittlicher Betrachtung der Altersgruppenmittelwerte, d.h. beim Vergleich der jeweils 16 Mittelwerte aus den 16 SOEP-Erhebungen innerhalb jeder der Ein-Jahres-Altersgruppen des ersten Meßzeitpunkts: Die dabei zu beobachtenden Mittelwertsunterschiede können nicht mehr durch Kohorteneffekte beeinflußt sein, da ja nun genaue genommen die Mittelwertsentwicklungen innerhalb von Ein-Jahres-Geburtsjahrganggruppen mit konstanter Kohortenausprägung betrachtet werden. Allerdings sind dabei auch mögliche Periodeneffekte zu berücksichtigen, welche bei solchen längsschnittlichen Mittelwertsvergleichen gewissermaßen in Konkurrenz zum Alterseffekt vorgefundene Unterschiede erklären könnte (vgl. die Erläuterungen in Kapitel 2.1).

Die übersichtliche deskriptive Darstellung der längsschnittlichen Entwicklung der oben querschnittlich betrachteten Mittelwerte über die 16 Meßzeitpunkte hinweg fällt insofern etwas schwerer, als man es hier mit einer Vielzahl von Ein-Jahres-Geburtsjahrganggruppen

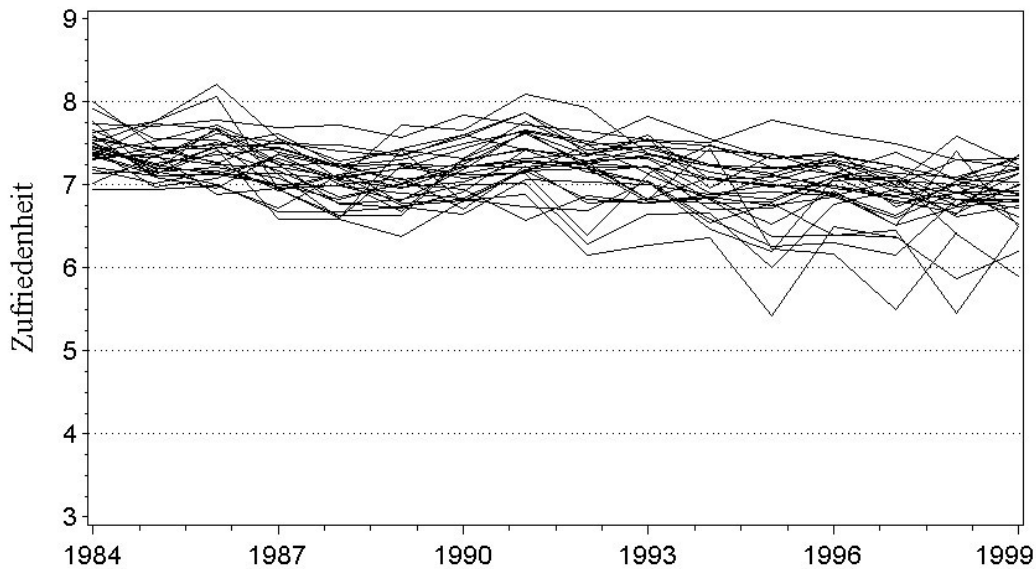
ABBILDUNG 3.5:
Freizeitzufriedenheit: Querschnittliche Mittelwertsverläufe über das Lebensalter



zu tun hat: Bei Beschränkung auf diejenigen Personen, die irgendwann während des untersuchten Zeitraums von 1984 bis 1999 ein Lebensalter von mindestens 60 Jahren erreicht haben (d.h. vor 1940 geboren wurden), sind in der Tabelle B.3 (Anhang B) alle Diagonalverläufe (von links oben nach rechts unten) unterhalb und einschließlich der Diagonale abzubilden, welche jeweils die Geburtsjahrgänge bis einschließlich 1939 repräsentieren. Wenn die Verläufe all dieser Geburtsjahrgänge über die 16 Wellen hinweg in ähnlicher Weise abgebildet werden, wie oben die querschnittlichen Verläufe über das Alter, so entstehen noch vollere und möglicherweise nicht mehr informative Knäuel – das solchermaßen für die Mittelwerte allgemeiner Lebenszufriedenheit resultierende wird in Abbildung 3.6 dargestellt. Eine insgesamt leicht fallende Tendenz dieser Verlaufskurven über den Untersuchungszeitraum hinweg ist dort deutlich erkennbar: Die Mehrzahl der Verläufe beginnt ungefähr knapp unter 7.5 und endet unter 7.0. Es scheint jedoch auch nicht wenige Verläufe zu geben, die keinen Abfall zeigen und nach den querschnittlichen Ergebnissen wäre zu vermuten, daß es sich dabei um diejenigen der jüngeren Geburtsjahrgänge handelt, die zum Ende des untersuchten Zeitraums z.B. noch keine 75 Jahre alt waren. Jedoch ist dieses aus Abbildung 3.6 nicht herauszulesen, da eine Beschriftung oder graphische Unterscheidung der einzelnen Verlaufskurven bei deren Vielzahl wohl kaum noch erkennbar wäre.

Zur besseren Verdeutlichung von möglicherweise über die Geburtskohorten hinweg veränderten längsschnittlichen Verläufen (d.h. genaugenommen einer Interaktion zwischen

ABBILDUNG 3.6:
Allgemeine Lebenszufriedenheit: Längsschnittliche Mittelwertsverläufe der Ein-Jahres-Geburtsjahrganggruppen

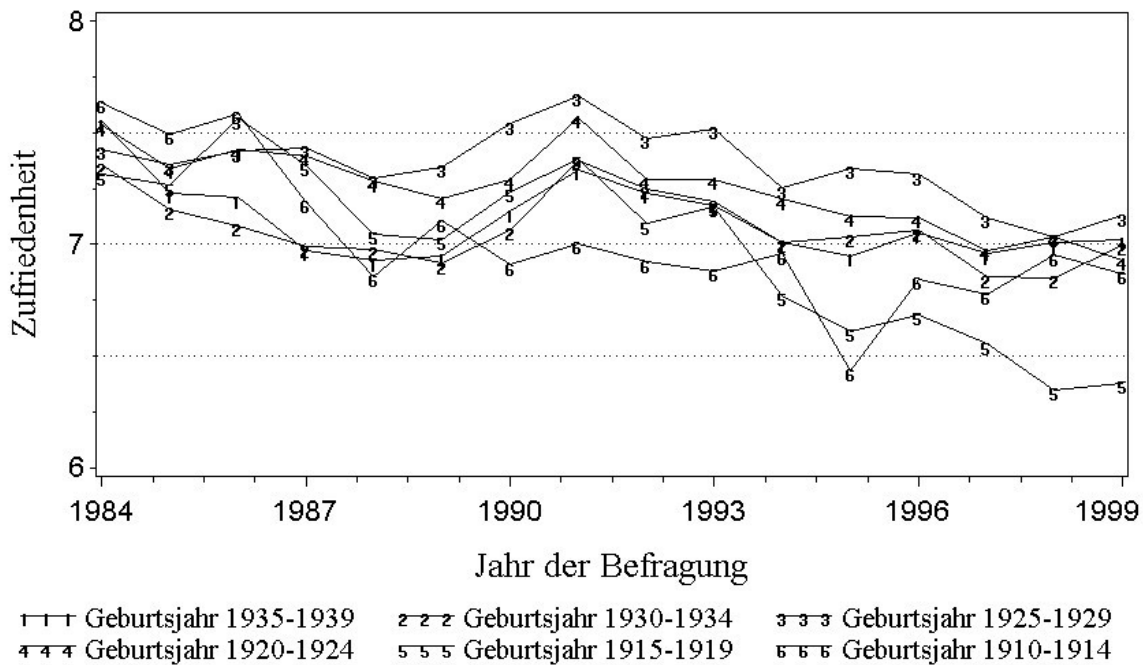


Alters- und Kohorteneffekt) wurden breitere Geburtsjahrganggruppen gebildet, so daß eine übersichtliche Zahl von Mittelwertsverläufen betrachtet werden kann: Abbildung 3.7 zeigt die längsschnittlichen Verläufe der Mittelwerte von sechs Fünf-Jahres-Geburtskohorten, nämlich der beim Panelbeginn 45- bis 49-Jährigen (Gruppe 1: Geburtsjahre 1935 bis 1939), 50- bis 54-Jährigen (Gruppe 2: Geburtsjahre 1930 bis 1934), 55- bis 59-Jährigen (Gruppe 3: Geburtsjahre der 1925 bis 1929), 60- bis 64-Jährigen (Gruppe 4: Geburtsjahre 1920 bis 1924), 65- bis 69-Jährigen (Gruppe 5: Geburtsjahre 1915 bis 1919) und 70- bis 74-Jährigen (Gruppe 6: Geburtsjahre der 1910 bis 1914). Die noch Älteren wurden nicht mehr berücksichtigt, da sich bei den entsprechenden Gruppen kaum noch zum Ende hin vollständige Zeitreihen von mit akzeptablen Fallzahlen berechneten Mittelwerten ergeben würden (vgl. Tabelle B.8 in Anhang B). Die Zahlen unter den Verlaufskurven verdeutlichen deren Zuordnung zu den obengenannten Gruppen.

Bei der zu beachtenden stark vergrößerten vertikalen Auflösung (im Vergleich zu Abbildung 3.6) zeigt sich, daß die Verläufe während der ersten Jahre wiederum stark „verknäuelte“ sind und insbesondere kein eindeutiger Trend dergestalt, daß die älteren Jahrgänge jeweils deutlich unter den jüngeren lägen, erkennbar wird. Nach 1990 lösen sich dann die Gruppen 5 und 6 mit jeweils steilerem Abfall etwas vom Rest. Allerdings steigt die Kurve der ältesten Geburtsjahrganggruppe bei den letzten beiden Jahren wieder an: Diese Werte der zum letzten untersuchten Zeitpunkt ja immerhin 85- bis 89-Jährigen sind in besonderem

ABBILDUNG 3.7:

Allgemeine Lebenszufriedenheit: Längsschnittliche Mittelwertsverläufe der Fünf-Jahres-Geburtskohorten



Maße vom Dropout betroffen und somit wieder mit einem besonders großen Standardfehler behaftet, womit dieser Anstieg als Resultat einer stärkeren „Zufallsschwankung“ dieser Mittelwerte erklärbar würde. Jedoch ist auch eine interessante Alternativerklärung denkbar: Möglicherweise geht der Dropout von Personen sehr hohen Alters mit einem intraindividuellen Absacken der Zufriedenheitswerte einher, so daß in der ältesten Geburtsjahrgangsguppe zum Ende der untersuchten Zeitreihe hin in wachsendem Maße die Mittelwerte von „überlebenden“ Personen mit intraindividuell stabilen, hohen Zufriedenheitswerten bestimmt werden und somit wieder ansteigen. Mit anderen Worten: Vielleicht deutet sich in diesem Anstieg auch der Effekt einer positiven Selektivität des Dropouts an. Allerdings erscheinen die diesbezüglichen Fallzahlen in der Tat zu gering (vgl. Tabelle B.8), um hier derartigen Schlußfolgerungen einen gewissen Nachdruck zu verleihen.

Für die jüngeren Geburtsjahrgangsguppen zeigt die Abbildung 3.7 tatsächlich einen sehr geringen, aber immerhin noch erkennbaren Abfall der Mittelwerte über die Zeit: Sie liegen 1999 jeweils um ungefähr 0.2 bis 0.4 Punkte unter dem Wert von 1984 – ob dieser Abfall geringfügig genug ist, um auch die längsschnittliche Entwicklung dieser Mittelwerte noch als „paradox stabil“ im Sinne der im Kapitel 1.2 dargestellten Diskussion zu be-

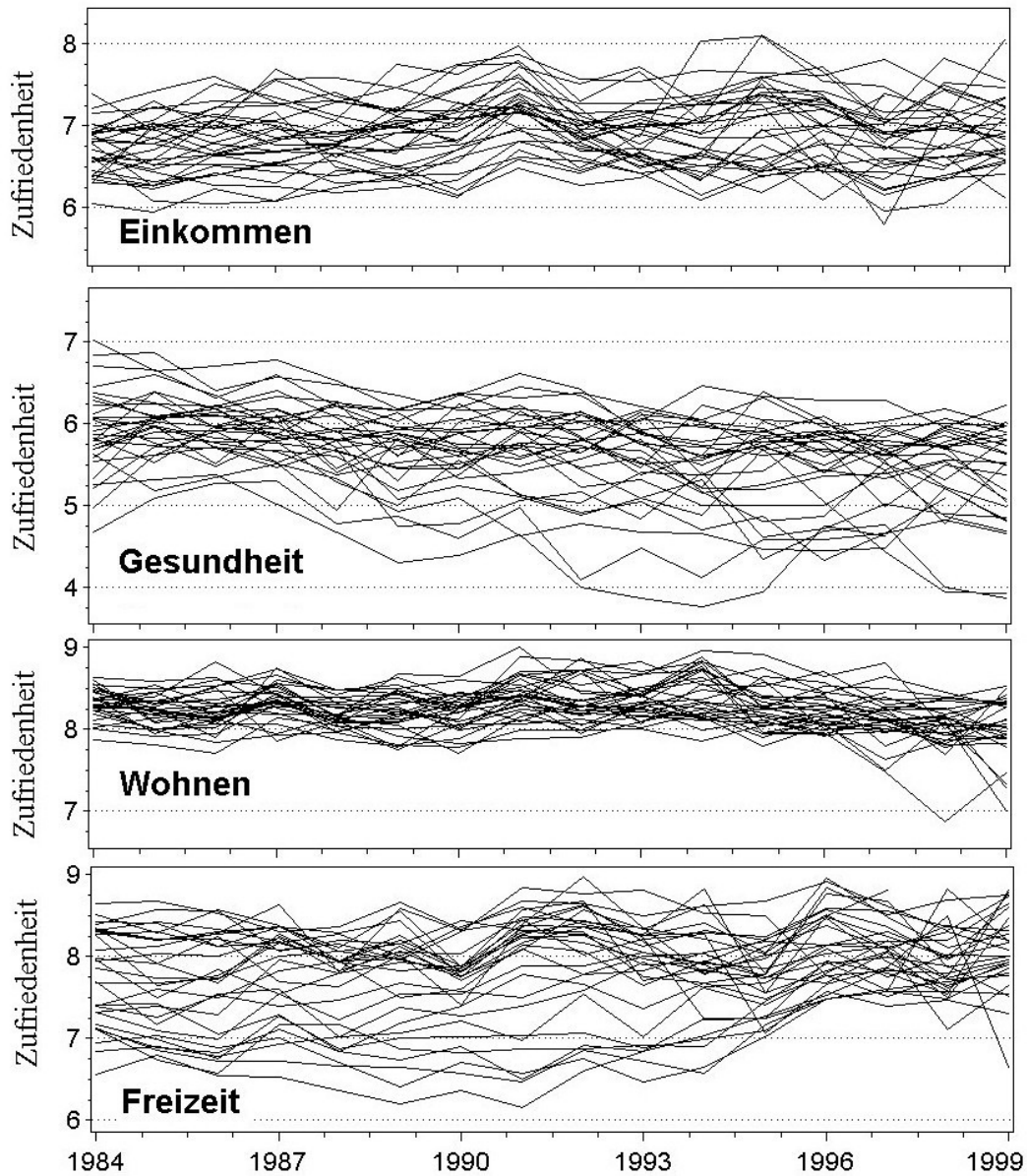
werten, kann an dieser Stelle nicht geklärt werden. Zwischen den Jahren 1984 und 1999 allerdings zeigen alle Verläufe ein etwas erstaunliches Auf und Ab, welches nicht ohne Weiteres erklärbar sein dürfte: Es scheint so, als würde die generelle schwache Abwärtstendenz in den Jahren um 1990 durch einen kurzfristigen Anstieg unterbrochen, um sich danach wieder fortzusetzen. Möglicherweise wäre dieses als Periodeneffekt einer allgemein während dieser Jahre gehobenen Zufriedenheit – z.B. infolge einer gewissen durch den historischen Umbruch der deutschen Wiedervereinigung erzeugten Euphorie – zu bewerten, jedoch erscheint auch diese Erklärung ohne zusätzliche spezifische diesbezügliche Untersuchungen allein aufgrund der hier beschriebenen Mittelwerte sehr vage. Ungeachtet dieser kurzfristigen Anstiege bleibt festzuhalten, daß die Mittelwerte der untersuchten Fünf-Jahres-Geburtskohorten eine leichte Tendenz des Abfalls mit zunehmendem Alter aufweisen.

Für die untersuchten Bereichszufriedenheiten fasse ich die zu Abbildung 3.6 analogen und eher verwirrende Darstellungen der längsschnittlichen Mittelwertverläufe der Ein-Jahres-Geburtsjahrganggruppen in Abbildung 3.8 zusammen und zeige in den Abbildungen 3.9, 3.10, 3.11 und 3.12 wiederum die Verläufskurven der sechs Fünf-Jahres-Geburtskohorten. Es werden zum Teil recht unterschiedliche Tendenzen in der Entwicklung der einzelnen durchschnittlichen Zufriedenheiten sichtbar:

Die Verläufe der Einkommenszufriedenheitsmittelwerte in Abbildung 3.9 sind durch weitestgehend stabilen Verlauf mit allenfalls geringen Aufwärts- oder Abwärtstendenzen über die gesamte Zeitreihe hinweg gekennzeichnet, dabei verlaufen anfangs die Kurven der jüngeren Kohorten unter denen der älteren, anfangs sind also die Ältesten auch die zufriedensten, was allerdings allmählich durch einen geringen Aufwärtstrend der beiden jüngeren und eine etwas stärkere Abwärtsentwicklung der beiden ältesten Kohorten aufgehoben wird. Berücksichtigt man die Lebensaltersspannen der einzelnen Gruppen zu den jeweiligen Messungen, so kann man in Abbildung 3.9 die teilweise Bestätigung des Befundes erkennen, den auch die querschnittlichen Verläufe in Abbildung 3.2 zeigten: Fallende Zufriedenheiten während des jeweils sechsten Lebensjahrzehnts sind nicht zu erkennen, jedoch ein leichter kontinuierlicher Anstieg über das siebte hinweg (bei den Gruppen 3 und 4 am Anfang der Verläufe, bei den Gruppen 1 und 2 zum Ende hin), ebenso zeigt sich der Abfall im sehr hohen Alter (Gruppen 5 und 6 zum Ende hin).

Die Verläufe der Gruppenmittelwerte in der Gesundheitszufriedenheit in Abbildung 3.10 bestätigen den bereits bei der querschnittlichen Betrachtung gewonnenen Eindruck einer allgemein im höheren Alter kontinuierlich fallenden Tendenz. Auch der dort (in Abbildung 3.3) sichtbare erste Abfall im Altersbereich um das fünfzigste Lebensjahr wird hier sichtbar (Gruppe 1, Gruppe 2 anfänglich), sowie bei genauem Hinsehen eine zwischenzeitliche „Erholungsphase“ um das sechzigste Lebensjahr herum (Gruppe 3 mit anfangs geringem Anstieg). Insgesamt aber erscheinen hier v.a. die Gruppen 2 und 3 über die gesamte Zeit hinweg relativ mittelwertstabil und zeigen sich somit die durchschnittlichen Gesundheitszufriedenheiten v.a. während der Altersspanne zwischen ungefähr dem fünfundfünfzigsten

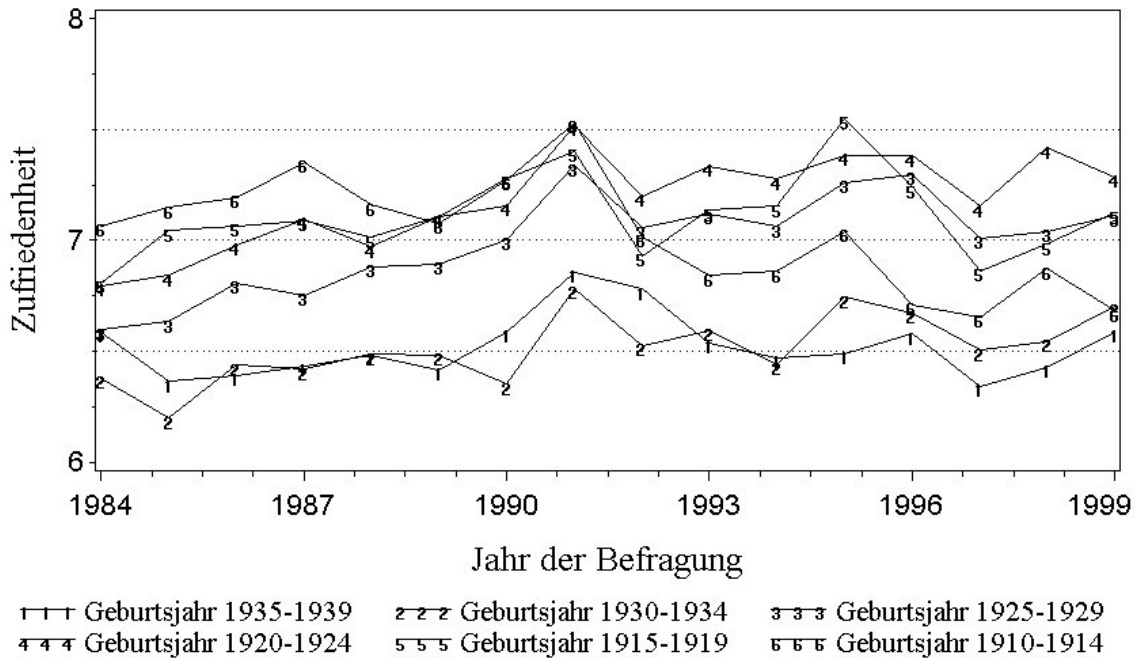
ABBILDUNG 3.8:
Bereichszufriedenheiten: Längsschnittliche Mittelwertsverläufe der Ein-Jahres-Geburtsjahrgangsgruppen



und fünfundsechzigsten Lebensjahr in allen Gruppen (auch in Gruppe 4 am Anfang und in Gruppe 1 zum Ende hin) im Großen und Ganzen kaum variabel.

Bei der Wohnzufriedenheit (Abbildung 3.11) zeigen sich geringe Unterschiede zwischen den Gruppen und wiederholt sich das bereits bekannte „querschnittliche“ Bild allge-

ABBILDUNG 3.9:
Einkommenszufriedenheit: Längsschnittliche Mittelwertsverläufe der Fünf-Jahres-Geburtskohorten

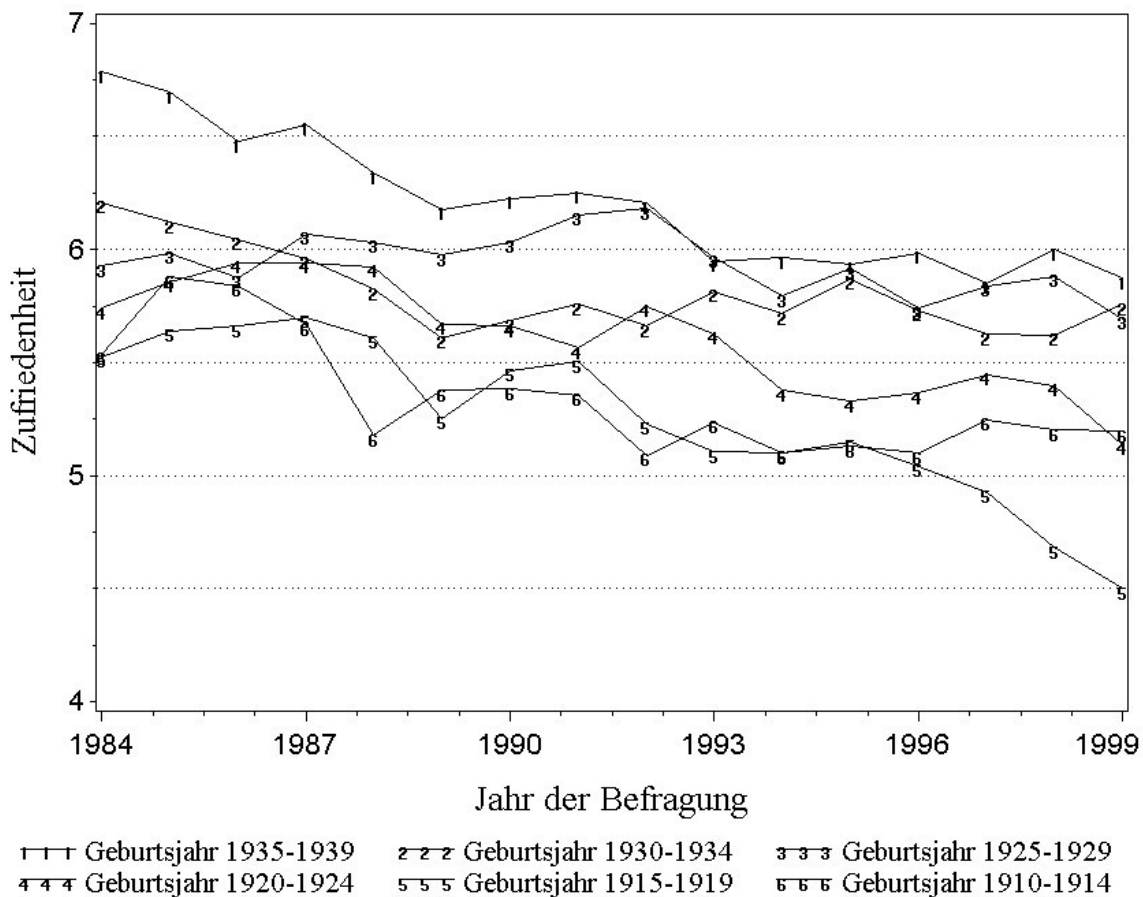


mein sehr hoher und in dieser Höhe kaum sich verändernder durchschnittlicher Zufriedenheiten. Eine Ausnahme bildet, mit Abfällen am rechten Ende der Kurven für die Gruppen 5 und 6, das sehr hohe Lebensalter, wobei hier allerdings wieder das Problem der infolge geringer Fallzahlen reduzierten Zuverlässigkeit dieser Werte zu berücksichtigen ist.

Auch bei der Freizeitzufriedenheit (Abbildung 3.12) bestätigt sich der schon aus Abbildung 3.5 bekannte Eindruck eines kontinuierlichen Anstiegs vor und während des Übergangs in die Phase der Höheraltrigkeit und anschließender Mittelwertsstabilität auf hohem Niveau.

Abschließend noch eine Art der Abbildung, die m.E. besonders gut geeignet ist, die aus den bisherigen Betrachtungen quer- und längsschnittlicher Mittelwertsverläufe gewonnenen Eindrücke bezüglich der Zusammenhänge zwischen Lebensalter und durchschnittlichen Zufriedenheiten besonders anschaulich und zusammenfassend darzustellen: Es wurden dazu für alle Personen der untersuchten Stichprobe ab dem fünfundvierzigsten bis zum neunzigsten Lebensjahr sukzessive Ein-Jahres-Differenzen der Zufriedenheiten errechnet und gemittelt. D.h. es wurde also für jede Person in der Stichprobe, die zu *irgendeinem* Meßzeitpunkt (1984 bis 1998) fünfundvierzig Jahre alt und auch im folgenden Jahr als 46-

ABBILDUNG 3.10:
Gesundheitszufriedenheit: Längsschnittliche Mittelwertsverläufe der Fünf-Jahres-Geburtskohorten



Jährige befragt worden war, die Differenz der beiden Zufriedenheitswerte gebildet, ebenso wurden alle in der Stichprobe vorhandenen intraindividuellen Zufriedenheitsdifferenzen zwischen den Lebensjahren 46 und 47, 47 und 48 usw. bis hin zur Differenz zwischen den Lebensjahren 89 und 90 gebildet. In den Abbildungen 3.13 bis 3.17 sind die Mittelwerte dieser sukzessiven intraindividuellen Ein-Jahres-Altersdifferenzen der Zufriedenheitsmeßwerte (wiederum in Form einer Verlaufskurve) dargestellt. Die Abbildung zeigen somit die insgesamt und unabhängig vom Meßzeitpunkt durchschnittliche Veränderung der Zufriedenheitswerte von jeweils einem zum nächsten Lebensjahr während der Altersspanne 45 bis 90 an, die in der Stichprobe stattgefunden hat. Da jeweils die Differenzen zwischen dem Wert des Folge- zu dem des Vorjahres berechnet wurden, bedeuten negative Werte einen Abfall und positive einen Anstieg der Zufriedenheit.

ABBILDUNG 3.11:
Wohnzufriedenheit: Längsschnittliche Mittelwertsverläufe der Fünf-Jahres-Geburtskohorten

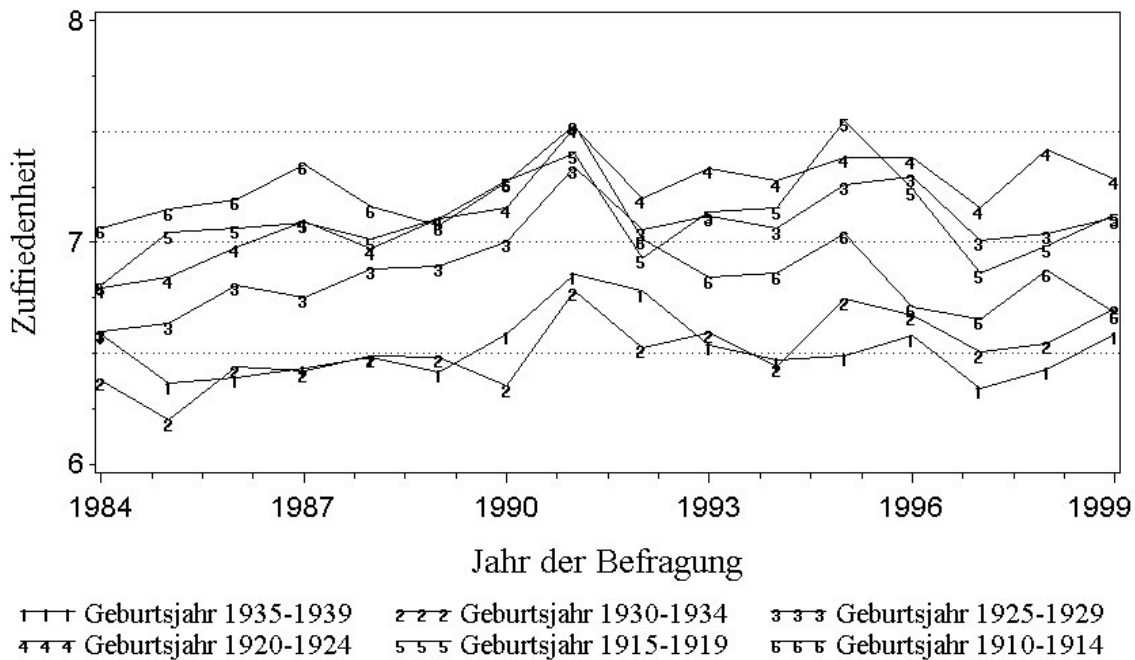
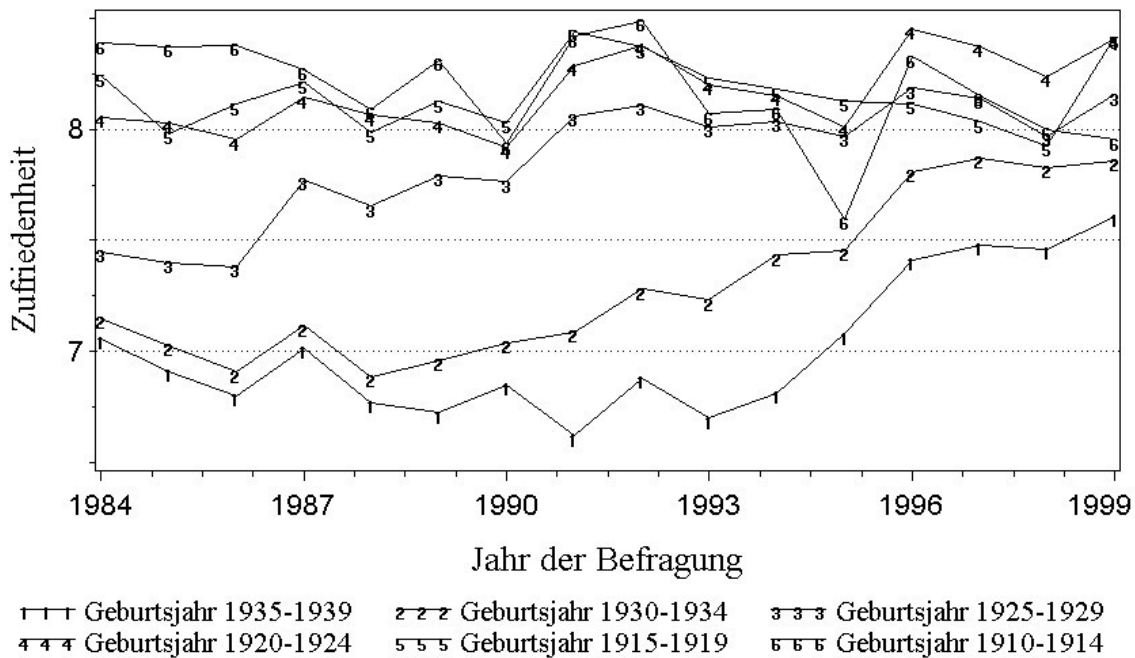


Abbildung 3.13 zeigt, daß insgesamt, über die gesamte Untersuchungsspanne hinweg, die Lebenszufriedenheitswerte der Befragten nach dem fünfundvierzigsten Lebensjahr im Durchschnitt fast kontinuierlich abnahmen – und zwar von Lebensjahr zu Lebensjahr bis ins hohe Alter hinein um einen jeweils nur sehr geringen, von gelegentlich stärkeren Ausschlägen abgesehen ungefähr zwischen 0 und -0.1 schwankenden Betrag, der dann zum Ende hin, ungefähr ab dem siebzigsten Lebensjahr, tendenziell zunimmt (die Werte für den sehr hohen Altersbereich sind wiederum aufgrund reduzierter Fallzahlen etwas weniger aussagekräftig). Diese abgebildeten durchschnittlichen Veränderungsbeträge mögen winzig erscheinen, man bedenke aber, daß z.B. eine ab dem fünfundvierzigsten Lebensjahr jährliche Abnahme um 0.05 sich bis ins Alter von 70 zu einem Verlust von 1.25 aufsummieren würde. Tatsächlich aber gibt es bei den einzelnen Personen keine derart geringen jährlichen Abnahmen, weil das Meßinstrument keine solchen geringen Ausprägungsintervalle erfaßt und nur Abnahmen um ganzzahlige Skalenwerte ermöglicht. Wenn also eine mittlere jährliche Veränderung z.B. -0.1 beträgt, so bedeutet dieses *wahrscheinlich*, daß hier die Mehrzahl der Personen unverändert geblieben sind und eine im Verhältnis zu diesen stabilen Personen kleine, jedoch im Verhältnis zu den Personen mit gestiegenem Zufrie-

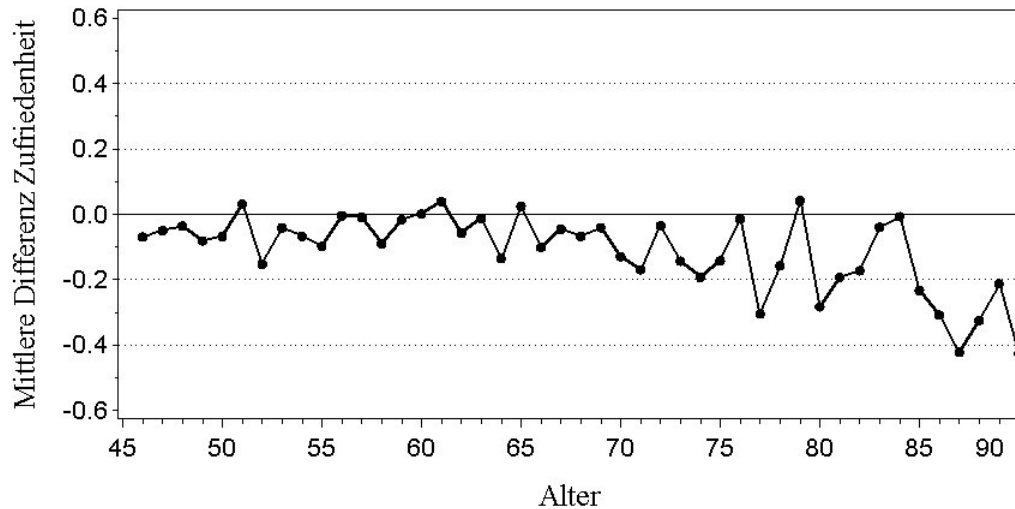
ABBILDUNG 3.12:
Freizeitzufriedenheit: Längsschnittliche Mittelwertsverläufe der Fünf-Jahres-Geburtskohorten



denheitswert große Zahl von Personen eine geringere Zufriedenheit angegeben hat, als im Vorjahr. Natürlich wären dabei prinzipiell auch ganz andere Schemata der stattgefundenen Veränderung möglich, die den Mittelwert knapp unter Null hervorgerufen haben könnten (z.B. könnten sich sehr viele Personen mit deutlicher Verbesserung und sehr viele Personen mit noch deutlicherer Verschlechterung wechselseitig „ausgemittelt“ haben) – die genaue Untersuchung solcher Veränderungscharakteristika wird Gegenstand von noch folgenden Analysen sein und geht über die hier thematische Deskription der Zusammenhänge von höherem Lebensalter und durchschnittlichen Zufriedenheiten hinaus.

Bezüglich letzterer ist jedenfalls festzuhalten, daß der nach Art von Abbildung 3.13 dargestellte Zusammenhang von möglichen Kohorteneffekten befreit ist, da ja jeweils die lebensaltersbezogenen Ein-Jahres-Differenzen aus allen Erhebungswellen – also über alle Geburtsjahrgänge hinweg – gemittelt wurden. Auch kurzfristige und nicht über den gesamten Untersuchungszeitraum hinweg kontinuierlich linear wirksame Periodeneffekte dürften hier die Mittelwerte kaum deutlich prägen, da diese sich dann sozusagen auf alle sukzessiven Werte verteilen und jeden einzelnen davon nur zu einem geringen Teil beeinflussen würden. Darum gibt diese Art der Auswertung bzw. Abbildung von Differenzmittelwerten

ABBILDUNG 3.13:
Allgemeine Lebenszufriedenheit: Mittlere Differenzen zwischen sukzessiven Lebensjahren



m.E. das klarste Bild vom möglichen Effekt des Lebensalters auf die durchschnittlichen Zufriedenheiten: Es zeigt immerhin eine kontinuierliche schwache Abnahme der durchschnittlichen Lebenszufriedenheit über das höhere Erwachsenenalter hinweg, welche im höchsten Lebensalter (> 70) noch zuzunehmen scheint.

Abbildung 3.13 widerlegt den Eindruck der Mittelwertsstabilität von Lebenszufriedenheit bis weit ins hohe Alter hinein, welcher bei querschnittlicher Betrachtung der Abbildungen 3.1 entsteht und bestätigt die Tendenz eines kontinuierlichen leichten Abfalls über die gesamte untersuchte Altersspanne hinweg, die bereits in Abbildung 3.6 bzw. 3.7 sichtbar wurde. Berücksichtigt man, daß in Abbildung 3.1 Alters- und Kohorteneffekte die Verläufe beeinflussen haben können, so läßt der Unterschied der Befunde aus quer- und längsschnittlicher Betrachtung bereits hier darauf schließen, daß der im Längsschnitt sichtbare Alterseffekt bei querschnittlichen Mittelwertsvergleichen durch einen Kohorteneffekt überlagert sein könnte, so wie es in Kapitel 1.2.3.1 diskutiert wurde (vgl. dort Abbildung 1.1). Das „Zwischenhoch“ um 1990 herum, welches in Abbildung 3.7 erschien, zeigt sich in Abbildung 3.13 in keiner Weise – und kann sich auch nicht zeigen, da hier über alle Meßzeitpunkte bzw. -intervalle hinweg die jeweiligen durchschnittlichen Ein-Jahres-Differenzen berechnet und somit auch, wie bereits gesagt, kurzfristige Periodeneffekte, die zur Erklärung dieses zwischenzeitlichen Anstiegs in allen Kohortengruppen in Frage kommen, „weggemittelt“ wurden.

Für die Einkommenszufriedenheit zeigt Abbildung 3.14 dagegen keinen solchen Abwärtstrend: Der Verlauf der mittleren Ein-Jahres-Differenzen windet sich sozusagen um die Nulllinie, ohne einen längeren Zeitraum deutlich darunter zu verlaufen. Schwach spiegelt

ABBILDUNG 3.14:
Einkommenszufriedenheit: Mittlere Differenzen zwischen sukzessiven Lebensjahren

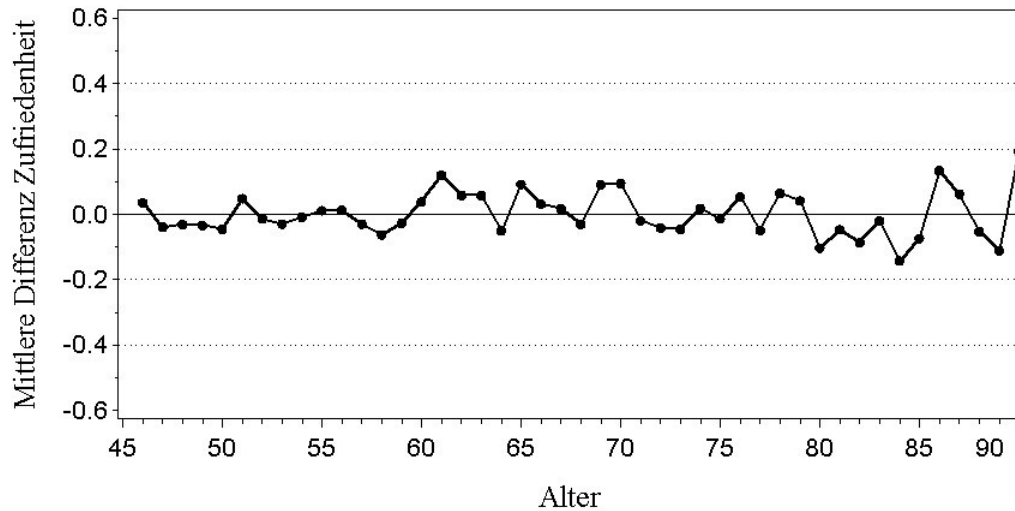
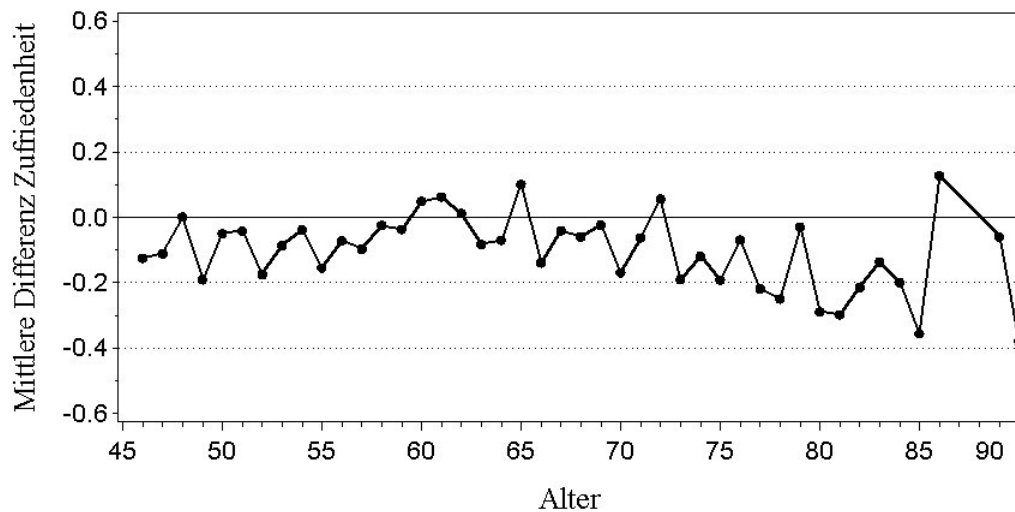


ABBILDUNG 3.15:
Gesundheitszufriedenheit: Mittlere Differenzen zwischen sukzessiven Lebensjahren



sich hier auch die bei Abbildung 3.9 gemachte Beobachtung von über das siebte Lebensjahrzehnt hinweg leicht ansteigenden Zufriedenheitsmittelwerten in den während dieses Zeitraums meist etwas deutlicher über der Nulllinie bei ca. 0.1 liegenden Werten.

Auch im Verlauf der mittleren Ein-Jahres-Differenzen der Gesundheitszufriedenheit

ABBILDUNG 3.16:
Wohnzufriedenheit: Mittlere Differenzen zwischen sukzessiven Lebensjahren

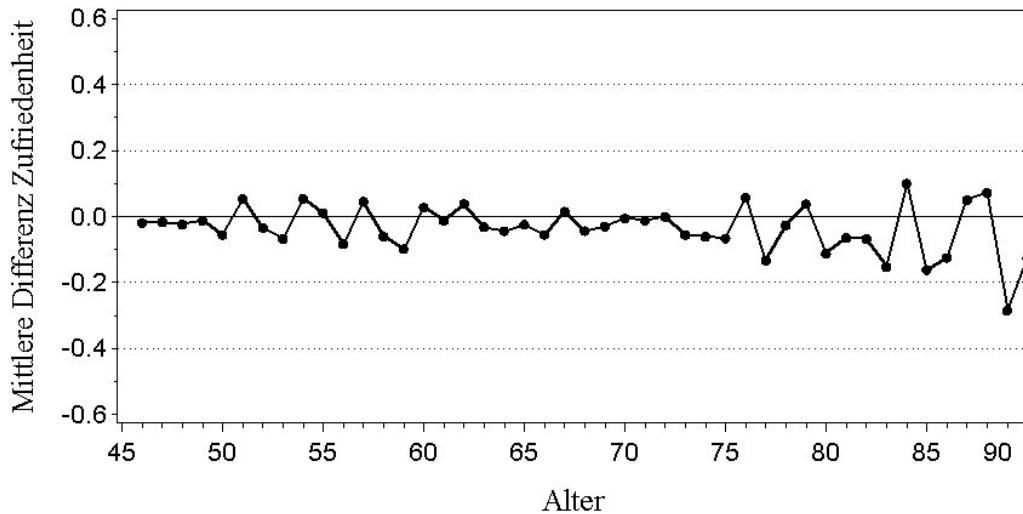
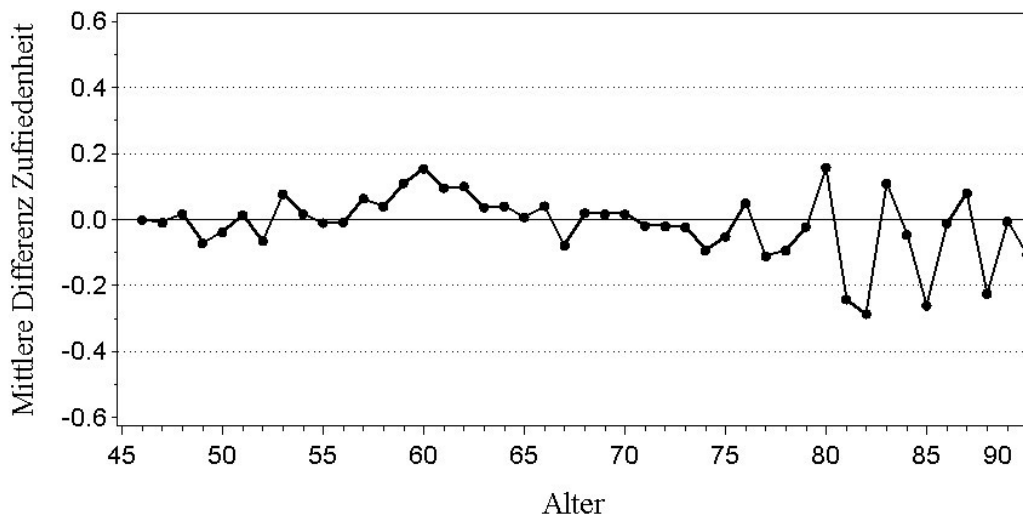


ABBILDUNG 3.17:
Freizeitzufriedenheit: Mittlere Differenzen zwischen sukzessiven Lebensjahren



in Abbildung 3.15 findet sich das wieder, was aus Abbildung 3.10 noch eher schwierig herauszulesen war, nämlich nicht nur ein fortgesetzter Abfall über das höhere Alter nach ungefähr dem fünfundsechzigsten Lebensjahr hinweg, sondern auch die erste Phase sinkender Zufriedenheit bis fast zum sechzigsten Lebensjahr hin und die leichte „Erholungsphase“

um dasselbe herum: Tatsächlich gibt es hier einige wenige Ausschläge der Kurve über die Nulllinie, die Zuwächse erweisen sich aber in dieser Darstellung als noch geringfügiger, als es bei Abbildung 3.10 noch erscheinen mochte: Insgesamt sind es während dieser Phase kaum mehr als drei solcher Lebensjahre, während derer die Gesundheitszufriedenheit im Durchschnitt in nennenswerter Weise leicht ansteigt, während ansonsten die durchschnittlichen Differenzen sich meist deutlich im negativen Bereich zwischen 0 und -0.2 bewegen und zum Ende hin bis fast -0.4 ausschlagen.

Abbildung 3.16 bestätigt die große Mittelwertstabilität der Wohnzufriedenheit mit einem Verlauf nahe der Nulllinie und um diese herum ohne längere Negativ- oder Positivphasen. Lediglich im sehr hohen Alter tendierten die Personen dann zu im Durchschnitt eher negativen Veränderungen ihrer Wohnzufriedenheit, so wie es ja auch bereits in Abbildung 3.11 offenbar wurde.

Abbildung 3.17 zeigt den Anstieg der durchschnittlichen Freizeitzufriedenheiten beim Übergang in die Phase des höheren Lebensalters als kleinen „Berg“ in der Verlaufskurve rund um des sechzigste Lebensjahr herum, bei weitestgehender Mittelwertstabilität vorher und nachher, die im sonstigen Kurvenverlauf nahe um die Nulllinie offenbar wird (von den Ausschlägen am rechten Ende wiederum abgesehen). Auch dieses bestätigt und präzisiert die Eindrücke, die aus Abbildung 3.12 gewonnen wurden.

3.2.1.3 Zusammenfassung und Diskussion der deskriptiven Analysen zum Zusammenhang von Alter und Zufriedenheit

Die somit vorgelegten Ergebnisse der deskriptiven Betrachtung von durchschnittlichen Zufriedenheiten über die Phase des mittleren bis höchsten Erwachsenenalters hinweg zeigten einige grundlegende Charakteristika, die nur zum Teil für die Fragestellungen dieser Untersuchung von Bedeutung sind, zum Teil aber hier sozusagen nur als Zugabe mitgeteilt werden, weil sie von allgemeinerontologischem Interesse sein dürften.

Zunächst konnte der Befund des Zufriedenheitsparadoxes auch in den hier untersuchten Daten repliziert werden: Dieser betrifft vornehmlich die allgemeine Lebenszufriedenheit und hier den Zusammenhang von Lebensalter und Zufriedenheit im Querschnitt. Tatsächlich zeigen bei der querschnittlichen Betrachtung dieses Zusammenhangs in den einzelnen SOEP-Erhebungsjahren die durchschnittlichen Lebenszufriedenheitswerte pro Lebensalter bis weit ins höhere Alter hinein keinen Abfall, so daß also von einer langanhaltenden Stabilität der Altersgruppenmittelwerte gesprochen werden könnte. Allerdings ist dieser Befund nicht uneingeschränkt und undifferenziert für die gesamte hochaltrige Lebensphase zu reklamieren, sondern kann hier eine in diesem Sinne stabile Phase früher Höheraltrigkeit von einer durch fortgesetzte durchschnittliche Zufriedenheitsverluste gekennzeichneten zweiten Phase später Höheraltrigkeit unterschieden werden: In allen 16 untersuchten querschnittlichen Verläufen zeigt sich ab ungefähr der Mitte des achten Lebensjahrzehnts ein moderater Abfall der Durchschnittswerte mit fortschreitendem Alter.

Die Betrachtung der längsschnittlichen Entwicklung der allgemeiner Lebenszufriedenheit ergab dagegen sehr wohl eine mit wachsendem Alter rückläufige Tendenz: Auch wenn in allen 16 Querschnitten die Altersgruppenmittelwerte über das Lebensalter hinweg bis ungefähr zum fünfundsiebzigsten Lebensjahr stabil bleiben, entwickeln sich doch die durchschnittlichen Lebenszufriedenheiten von Geburtsjahrganggruppen während deren Älterwerdens auch über frühere Altersspannen hinweg nach unten und es zeigt sich, daß während des gesamten Altersspektrums des Übergangs in die und Durchlaufens der Höheraltrigkeit die intraindividuellen Veränderungen der Lebenszufriedenheit von Lebensjahr zu Lebensjahr im Durchschnitt negativ sind. Somit liefern, wie bereits im Theorieteil dieser Arbeit (Kapitel 1.2.3.1) für möglich gehalten, quer- und längsschnittliche Untersuchung des Zusammenhangs von Alter und Lebenszufriedenheit divergierende Befunde, welche es wahrscheinlich erscheinen lassen, daß eine im Querschnitt mögliche Konfundierung von Alters- und Kohorteneffekten stattgefunden hat, die diese Divergenz erklären könnte. Es könnte also, wie durch die Kohorteneffekthypothese vermutet, in querschnittlichen Vergleichen der negative Alterseffekt durch einen positiven Effekt früherer Geburtsjahre neutralisiert worden sein. Allerdings ist auch im Längsschnitt eine Konfundierung möglich, nämlich die von Alters- und Periodeneffekten (vgl. die diesbezüglichen Erläuterungen im Kapitel 2.1), so daß im Prinzip auch ein über die gesamte Untersuchungsspanne hinweg wirksamer „epochaler“ Periodeneffekt sinkender Zufriedenheiten die besagte Divergenz der Befunde hervorgerufen haben könnte – daß also die längsschnittlichen Verläufe diesen ausgerechnet während des untersuchten Zeitraums stattgefundenen allgemeinen Zufriedenheitsschwund zeigen, statt einer lebensalterstypischen Entwicklung. Abgesehen von der Schwierigkeit, einen solchen „Epocheneffekt“ untersuchungstechnisch von Kohorten- oder Alterseffekten zu trennen und aufzuzeigen, erscheint er allerdings als keine gute Erklärung für die besagte Divergenz, solange keine theoretischen Anhaltspunkte gefunden werden, aus denen er sich hypothetisch begründen ließe.

Wenn man also keinen langanhaltenden Periodeneffekt allgemeiner Lebenszufriedenheitsverluste annimmt, so zeigt sich in den längsschnittlichen Betrachtungen durchaus ein negativer Zusammenhang von Alter und Zufriedenheit während der gesamten Phase der Höheraltrigkeit. Ob das in den Daten erscheinende Ausmaß dieser Zufriedenheitsverluste praktisch bedeutsam und ob also der Alterseffekt nicht nur ein geringfügiger ist, kann hier noch nicht beantwortet werden, denn es fehlen vorerst, bei der bloßen Deskription der Zusammenhänge, statistische und erst recht inhaltliche Kriterien für eine solche Beurteilung. Eine wie auch immer durchgeführte Testung des Alterseffekts auf statistische Signifikanz würde dieses Kriterium nicht liefern, da mit Signifikanztests ja, wie von Kritikern deren gewohnheitsmäßigen Gebrauchs immer wieder hervorgehoben wurde (vgl.z.B. Sedlmeier, 1996; Frick, 1996), gerade *nicht* zur eigentlichen Bewertung von Effektgröße und praktische Bedeutsamkeit taugen. Auch Maße der Effektstärke, wie z.B. das von Kraemer & Thiemann (1987, 47) vorgeschlagene Maß Δ für Varianzanalysen mit Meßwiederholung, würden hier kaum zur Klärung der aufgeworfenen Frage beitragen, da auch geläufige Dau-

menregeln zur Interpretation solcher Maße letztlich nur ad hoc auf sehr allgemeinen Plausibilitätskriterien begründet sind und die hier spezifischen Gesichtspunkte praktischer Relevanz auch sehr geringer Veränderungen kaum repräsentieren. Es bleibt hier also offen, ob der vorgefundene negative Alterseffekt für die Lebenszufriedenheit als stark genug einzuschätzen ist, um die Rede von „paradoxe“ Stabilität der durchschnittlichen Lebenszufriedenheit bis weit ins höhere Alter hinein zu widerlegen bzw. ob er nicht so geringfügig ist, daß noch eine weitestgehender oder näherungsweise Mittelwertstabilität konstatiert werden könnte. In den noch folgenden Stabilitätsanalysen wird auch versucht, diese Frage anhand von statistischen Kriterien zu beantworten.²⁸

Bezüglich der Durchschnittswerte der untersuchten Bereichszufriedenheiten ergab die Deskription Merkmale der längsschnittlichen Entwicklungen, die teilweise stark von denen der allgemeinen Lebenszufriedenheit abweichen. Das höhere Erwachsenenalter, so der Eindruck aus den vorangestellten Beschreibungen, ist eine hinsichtlich der Zufriedenheiten mit dem Einkommen, Wohnen und der Freizeit durchaus günstige Lebensphase und lediglich mit deutlichen Verlusten in der Zufriedenheit mit der Gesundheit verbunden. Die Einkommenszufriedenheit steigt während der frühen Höheraltrigkeit sogar beständig leicht an und ist bei älteren Personen eher besser, als bei jüngeren Personen im mittleren Erwachsenenalter. Erst recht erfährt die Freizeit Zufriedenheit beim Übergang in die Höheraltrigkeit einen deutlichen positiven „Wachstumsschub“ auf eine sehr hohes Niveau und kann damit geradezu als günstiges Merkmal des Alterns in dieser Lebensphase bezeichnet werden. Die Wohnzufriedenheit befindet sich in allen Altersphasen auf einem sehr hohen, stabilen Niveau und verbleibt bis weit ins hohe Lebensalter hinein auf demselben. *Das* negative Merkmal des Alterns im höheren Lebensalter ist hier die Zufriedenheit mit der Gesundheit: Sie fällt deutlich ab. Interessanterweise ist dieser Abfall aber bereits in recht frühen Lebensphasen schon vor dem Altersbereich, den man mit dem Begriff der Höheraltrigkeit belegen könnte (hier: 60+), zu beobachten und scheint er gerade während des Übergangs in die Höheraltrigkeit durch eine kurze Stabilitätsphase unterbrochen.

Diese Befunde zu den Bereichszufriedenheiten sollen hier nicht im einzelnen weiter diskutiert und interpretiert werden. Nimmt man sie als das, was sie sein sollen, nämlich als subjektive Bewertung der jeweiligen Lebensumstände im bezeichneten Bereich, so bietet sich aus der Kenntnis alterstypischer Entwicklungen in diesen Lebensumständen manche plausible Erklärung quasi von selbst an (z.B. wachsende Morbidität im höheren Alter zur Erklärung des negativen Alterseffekts bei der Gesundheitszufriedenheit, Austritt aus dem Arbeitsleben zur Erklärung des Anstiegs der Freizeit Zufriedenheit), während andererseits mancher Befund im Lichte dieser Kenntnisse auch überraschen mag (z.B.: Sollte die Freizeit Zufriedenheit nicht infolge von durch gesundheitliche Schwierigkeiten oder soziale Ausgrenzungen reduzierten Teilnahmemöglichkeiten eher schwinden?). Solche Überlegungen

²⁸Das Kriterium wird dort – im Kapitel 3.3 – die Anpassungsgüte (an die empirischen Daten) von statistischen Modellen, in denen die zeitliche Invarianz der Mittelwerte über die Zeitreihe hinweg festgelegt ist, im Vergleich zu solchen ohne diese Festlegung sein.

sind aber nicht das Thema der hier vorgestellten Untersuchung und werden deshalb nicht vertieft. Immerhin bemerkenswert erscheint es aber, daß in den gezeigten Ergebnissen sich das höhere Erwachsenenalter einmal mehr als Lebensphase zeigt, die nicht durch einfache Altersstereotype zu beschreiben ist: Während sich manches (Gesundheit) erwartungsgemäß negativ entwickelt, sind ältere Menschen mit manch anderen Lebensumständen im Durchschnitt deutlich zufriedener, als jüngere bzw. als sie selbst es in jüngeren Jahren waren.

3.2.2 Analysen zur Kohorteneffekthypothese

Zur Überprüfung der Kohorteneffekthypothese sollen die Zufriedenheitsmittelwerte der Höheraltrigen daraufhin analysiert werden, ob darin eine Überlagerung negativer Kohorten- und Alterseffekte dergestalt, wie in Kapitel 1.2.3.1 überlegt und dort in Abbildung 1.1 schematisch gezeigt, erkennbar ist. Gemäß der Kohorteneffekthypothese wäre zu vermuten, daß bei älteren Geburtskohorten die Lebenszufriedenheit beim Älterwerden gewissermaßen von einem höheren Niveau aus abfällt, als bei jüngeren. Für die vorkriegsgeborenen Generationen Deutscher, so könnte man die Überlegung beschreiben, kann gelten: Je weiter der Geburtszeitpunkt zurückliegt, desto mehr Lebenszeit wurde in krisenhaften Lebensumständen verbracht. Und in dem Maße, in dem in jungen Jahren „schlechte Zeiten“ durchlebt wurden, könnte die positive Bewertung der Wohlstandsbedingungen in den dann vergleichsweise guten bundesdeutschen Nachkriegszeiten wachsen.

Belege für die vermutete Überlappung von Kohorten- und Alterseffekten könnten bereits deskriptive Analysen wie die im vorherigen Kapitel vorgestellten liefern: Die Abbildungen 3.6 bis 3.12 zeigen jeweils längsschnittliche Verlaufskurven von Zufriedenheitsmittelwerten verschiedener Kohortengruppen, in denen negative Alterseffekte immerhin erkennbar sein müßten – wobei dann, wie in Kapitel 2.1 dargelegt, Abwärtstrends über die Zeitachse auch Periodeneffekten geschuldet sein könnten. Insbesondere aber sollten gemäß der Kohorteneffekthypothese *in diesen Abbildungen* die Verläufe der verschiedenen Kohortengruppen nicht nur kontinuierlich fallende Tendenzen aufweisen, sondern die Verläufe der Kohorten alle möglichst nahe beieinander liegen, so daß tatsächlich bei einem vertikalen Querschnitt durch die Verläufe nur geringe Gruppenunterschiede gefunden würden.

Vor allem aber wurde zum Zwecke dieser Analysen das in Kapitel 2.1 durch Gleichung 2.11 spezifizierte Multilevel-Regressionsmodell zur simultanen Schätzung von Alters- und Kohorteneffekten für die Meßwertreihen der allgemeinen Lebenszufriedenheit und für die vier Bereichszufriedenheiten berechnet. Die Untersuchungsgruppe wurde auf die in den Jahren 1924 bis 1939 geborenen westdeutschen SOEP-Befragungspersonen begrenzt: Dieses sind diejenigen Personen, deren sechzigstes Lebensjahr in die Zeitspanne 1984 bis 1999 fiel und die also im Befragungszeitraum in der transformierten Altersvariablen A_{ti}^* das Alter Null erreichten und somit sozusagen die Nulllinie des Alters kreuzten (vgl. Abbildung 2.1). Allerdings können dabei auch Personen sein, die noch vor Erreichen des sechzigsten Lebensjahres als SOEP-Befragungspersonen ausschieden.

Die Beschränkung auf diese kürzere Spanne von Geburtsjahrgängen erschien einerseits inhaltlich sinnvoll, da diese Gruppe für den Erhebungszeitraum ungefähr die Lebensphase des Übergangs in die und Durchlaufens der frühen Höheraltrigkeit abdeckt, welche für die Kohorteneffekthypothese eigentlich nur von Bedeutung ist, da ja das dadurch zu erklärende Zufriedenheitsparadox querschnittlicher Stabilität von Altersgruppenmittelwerten, wie die in Kapitel 3.2.1 dargestellten deskriptiven Untersuchungen ergaben, im sehr hohen Altersbereich zumindest bei der allgemeinen Lebenszufriedenheit gar nicht mehr vorgefunden wurde. Andererseits erschien diese Beschränkung v.a. aber auch deswegen notwendig, weil die Linearität der Effekte des Multilevel-Modells mit wachsender Breite der einbezogenen Geburtsjahres- und Altersspanne zusehends fraglicher erscheint: Daß über einen sehr langen Bereich von Geburtsjahrgängen hinweg tatsächlich ein näherungsweise linearer Effekt des Geburtsjahres auf die Zufriedenheiten besteht, ist eigentlich schwer anzunehmen, dagegen aber erscheint es durchaus denkbar, daß über die überschaubare Spanne der 15 Geburtsjahre von 1924 hinweg deren möglicher Effekt nicht derart ausgeprägte Trendveränderungen durchläuft, daß die Annahme auch nur näherungsweise Linearität ganz und gar unsinnig wird. Genausowenig ist anzunehmen, daß beispielsweise über die gesamte Lebensaltersspanne hinweg ein linearer Abfall oder Anstieg durchschnittlicher Altersgruppenmittelwerte der Zufriedenheiten stattfindet, in einem gewissen Abschnitt jedoch – hier von 30 Lebensaltersjahren zwischen 45 (die 1934 Geborenen im Jahr 1984) und 75 (die 1929 Geborenen im Jahr 1999) – mag eine näherungsweise Linearität noch annehmbar sein. Die Beschränkung des einbezogenen Spektrums von Geburtsjahrgängen erfolgte also v.a. zur Rechtfertigung der Linearitätsannahmen des Analysemodells, wie in der folgenden Diskussion der Ergebnisse noch einmal genau dargelegt werden wird.

3.2.2.1 Ergebnisse der Analysen zur Kohorteneffekthypothese

Zunächst kann zur deskriptiven Analyse möglicher Kohorteneffekte noch einmal auf die im vorherigen Kapitel gezeigten Abbildungen zurückgegriffen werden. Die graphischen Darstellungen der längsschnittlichen Mittelwertsverläufe in den besagten Abbildungen 3.6 und 3.7 zeigten deutliche Abwärtstendenzen der Durchschnittswerte von Lebenszufriedenheit, sowie in 3.10 diejenigen der Gesundheitszufriedenheit. Bei ersterer liegen die Verläufe auch recht nahe beieinander (angesichts der hohen vertikalen Auflösung auch in Abbildung 3.7), so daß es den Anschein hat, daß hier eine Überlagerung eines relativ schwachen Alterseffekts (leicht fallende Tendenz) und eines relativ schwachen Kohorteneffekts (geringe Abstände der Verlaufskurven) genauso, wie in der Kohorteneffekthypothese postuliert, zu beobachten ist.

Bei der Gesundheitszufriedenheit besteht, wenn man das Ein-Jahres-Geburtskohorten-Knäuel in Abbildung 3.8 betrachtet, nicht derselbe gleichlaufende Abwärtstrend in allen Verläufen, auch ist das Knäuel insgesamt breiter. Die Untersuchung der Abbildung 3.10, sowie auch der querschnittlichen Verläufe (Abbildung 3.3) und der durchschnittlichen Ein-

Lebensjahres-Mittelwertsdifferenzen (Abbildung 3.15) ergab hier, daß ungefähr während eines Lebensjahrzehnts um das sechzigste Lebensjahr herum stets eine gewisse Stabilität der sukzessiven Mittelwerte über das Alter hinweg zu beobachten ist, außerhalb dieses Altersbereichs ansonsten aber ein stetiger Abfall derselben. In Abbildung 3.8 ist darum der eigentlich deutliche Abfall der Mittelwerte über das Alter hinweg kaum zu erkennen, da dort viele Ein-Jahres-Geburtsjahrgangskurven bei einem anfänglichen Abfall allmählich in die Stabilitätsphase übergehen (jüngere Jahrgänge) und viele andere umgekehrt bei anfänglicher Stabilität in die Phase des Abfalls (Verläufe aus dem mittleren Bereich des abgebildeten Spektrums von Geburtsjahrgängen). Auch ist bei genauem Hinsehen erkennbar, daß der Eindruck der Breite des Knäuels v.a. durch einige wenige tieflaufende Kurven erzeugt wird. Es dürfte sich hierbei um diejenigen sehr alter Geburtsjahrgänge mit teilweise geringen Fallzahlen zur Errechnung der Mittelwerte handeln – dieses bestätigt auch der Blick in die Tabelle B.5 im Anhang B. Somit könnte trotz auf den ersten Blick dagegensprechender Eindrücke hier noch näherungsweise das der Kohorteneffekthypothese entsprechende Muster vorhanden sein, das auch deutlich sichtbar würde, wenn man die sehr alten Geburtsjahrgänge aus der Darstellung entfernt und anstatt der tatsächlichen Verlaufskurven deren linearen Trend über alle Jahre hinweg darstellt. Ersteres erschiene hier auch durchaus theoretisch begründbar: Denn tatsächlich scheint es so, daß im sehr hohen Alter ein beschleunigter Abfall der durchschnittlichen Gesundheitszufriedenheiten stattfindet, welcher kaum mehr durch eine entsprechende Verstärkung eines Kohorteneffekts so kompensiert werden könnte, daß auch bis in diesen Altersbereich hinein die querschnittliche Mittelwertstabilität besteht.

Bei allen anderen untersuchten Bereichszufriedenheiten zeigen die graphischen Darstellungen der längsschnittlichen Mittelwertverläufe keinerlei Hinweise auf ein Muster, das dem der Abbildung 1.1 und somit der Kohorteneffekthypothese entspräche. Insbesondere findet sich kein annähernd kontinuierlicher Abfall über das Lebensalter, wie er in dieser Hypothese impliziert ist, so daß scheinbar keine Rede davon sein, daß ein negativer Alterseffekt durch einen negativen Kohorteneffekt so überlagert wird, daß im querschnittlichen Alters- bzw. Geburtsjahresgruppenvergleich kaum noch Unterschiede auftreten. Jedoch besteht prinzipiell immerhin noch die Möglichkeit, daß ein negativer Alterseffekt bei diesen Zufriedenheiten deshalb nicht sichtbar wird, weil er durch einen entsprechend positiven Periodeneffekt überlagert und sozusagen neutralisiert wird, so wie natürlich auch die oben aufgezeigten durchschnittlichen Abfälle in den längsschnittlichen Verläufen der Lebens- und Gesundheitszufriedenheit nicht nur auf Alters-, sondern im Prinzip auch auf Periodeneffekten beruhen könnten. Die Möglichkeit solcher Periodeneffekte, das sei hier noch einmal gesagt, wird hier nicht deswegen erwogen, weil konkrete, auf die gesellschaftlichen und historischen Umstände der Datenerhebungsjahre bezogene, „inhaltliche“ Hypothesen Anlaß dazu geben, sondern nur deswegen, weil bei der Beobachtung vermeintlicher Alterseffekte in den längsschnittlichen Verläufen nicht logisch korrekt ausgeschlossen werden kann, daß diese Beobachtung durch Periodeneffekte verzerrt ist, und somit eine diesbezügliche Vor-

sicht bei Interpretationen des Beobachteten prinzipiell angebracht ist. Es erscheint darum, weil die graphische Auswertung sozusagen bei keiner der untersuchten Zufriedenheiten eine hundertprozentige Gewißheit über die zeitvariablen Effekte liefern kann, nicht ganz und gar unsinnig, der Vollständigkeit halber die regressionsanalytische Prozedur zur Schätzung von Alters- und Kohorteneffekten auch für die Bereichszufriedenheiten mitlaufen zu lassen. Allerdings kann mit Blick auf die graphischen Darstellungen festgehalten werden, daß diese deutliche Hinweise darauf liefern, daß die Kohorteneffekthypothese hier, in den Zufriedenheitsdaten des SOEP, für die Einkommens-, Wohn- und Freizeitzufriedenheiten keine Gültigkeit besitzt.

Die Ergebnisse der Berechnung des Multilevel-Regressionsmodells zeigt Tabelle 3.4: Für die allgemeine Lebenszufriedenheit (ALZ), Einkommenszufriedenheit (EIZ), Gesundheitszufriedenheit (GEZ), Wohnzufriedenheit (WOZ) und Freizeitzufriedenheit (FRZ) sind die Schätzungen für die fixen Effekte (Koeffizienten), sowie der Varianzen und Kovarianzen der Zufallskomponenten des Modells angegeben. Letztere wurden nur der möglichst vollständigen Dokumentation der Ergebnisse wegen aufgelistet, sie sind für die Bewertung der Kohorteneffekthypothese nicht relevant.²⁹ Bei den fixen Effekten sind die unstandardisierten Koeffizientenwerte, sowie die Ergebnisse der Signifikanztests für diese Koeffizienten abgebildet – im Grunde sind es nur diese wenigen Werte, die hier interessieren. Bei den Signifikanztests handelt es sich um approximative *t*-Tests (siehe z.B. bei Verbeke & Molenberghs, 2000, 56f).

„Stand-alone-Kennwerte“ zur Bewertung der Modellanpassung, wie sie bei Strukturgleichungsmodellen berechnet werden, wurden für diesen Typus statistischer Analyse nicht entwickelt, die von SAS Prozedur MIXED ausgegeben „Fit Statistics“ sind sinnvoll verwendbar nur zum Vergleich verschiedener Modelle (Singer, 1998), sie werden darum nicht berichtet. Der sogenannte Null-Model-Likelihood-Ratio χ^2 -Test, der für Multilevel-Modelle berechnet werden kann und von der Prozedur MIXED ausgegeben wird, entspricht noch am ehesten der Logik der Testung der durch die Maximum-Likelihood-Lösung erzielten Anpassungsgüte. Er testet praktisch die Likelihood des Modells gegen die des vergleichsweise restriktiven Regressionsmodells ohne zufällige (d.h. hier über die Personen variierende) Effekte und seine Signifikanz würde anzeigen, daß die restriktive Modellierung zu einer nicht ignorierbaren Verschlechterung der Anpassung führt und somit die Modellierung der zufälligen Effekt notwendig ist – Signifikanz würde also sozusagen *für* das Modell sprechen. Dieser Test war in allen fünf Analysen jeweils hochsignifikant mit $p \leq 0.0001$, allerdings gilt auch er als nur bedingt verlässlich und wird von seiner Verwendung eher abgeraten (Verbeke & Molenberghs, 2000, 106).

Bei den hier eigentlich interessierenden fixen Effekten zeigt sich nun für die allge-

²⁹ Auch ist hier darauf hinzuweisen, daß die Signifikanztests zu diesen Varianzkomponenten – sogenannte approximative Wald-Tests –, deren Ergebnisse in Tabelle 3.4 in „Sternchenform“ berichtet werden, nur bedingt aussagekräftig sind. Verbeke & Molenberghs (2000, 108) raten beispielsweise von deren Verwendung ganz ab.

TABELLE 3.4:
Ergebnisse der Multilevel-Regressionsanalyse: Schätzung von Alters- und Kohorteneffekten

	ALZ	EIZ	GEZ	WOZ	FRZ
Konstante β_{00}	7.401***	6.723***	5.929***	8.368***	7.738***
Alterseffekt $\beta_{10} = \alpha_{[\tau]}$	-0.033***	0.032***	-0.024***	-0.011	0.053***
Kohorteneffekt $\beta_{01} = \gamma_{[\tau]}$	-0.038***	-0.015	-0.014	-0.032***	-0.044***
Interaktion β_{11}	0.001	-0.002**	-0.003***	0.000	-0.001
Varianz σ_{v0}^2	1.812***	2.671***	2.911***	1.577***	2.022***
Varianz σ_{v1}^2	0.010***	0.014***	0.012***	0.010***	0.016***
Kovarianz σ_{v0v1}	0.021***	0.019**	0.010	0.017***	-0.010
Varianz σ_{ϵ}^2	1.783***	2.178***	2.498***	1.596***	2.754***

*** $p \leq 0.001$; ** $p \leq 0.01$

$n = 2380$

meine Lebenszufriedenheit in großer Deutlichkeit das gemäß der Kohorteneffekthypothese vermutete Ergebnis: Sowohl der Alters-, als auch der Kohorteneffekt sind mit jeweils negativem Vorzeichen statistisch signifikant, dazu noch unterscheiden sich ihre Beträge nur ganz gering. Dieses entspricht der formulierten Kohorteneffekthypothese in einem Maße, das fast schon überrascht: Die Werte sagen praktisch die querschnittliche Stabilität der Altersgruppenmittelwerte bei vorhandenen negativen Alters- und Kohorteneffekten voraus! Wenn man beispielsweise den Referenzwert der im Jahre 1984 60-Jährigen querschnittlich mit dem der in diesem Jahr 59-Jährigen (d.h. 1925 Geborenen) vergleicht, so ergibt sich die vorhergesagte Differenz $-\alpha_{[\tau]} + \gamma_{[\tau]} = -(-0.033) - 0.038 = -0.005$ und angesichts eines praktisch nicht vorhandenen Interaktionseffekts ist dieses der Vorhersagewert für sämtliche querschnittlichen Mittelwertvergleiche benachbarter Altersgruppen der von 1924 bis 1939 Geborenen zu allen 16 Erhebungsjahren. Für die 16 einbezogenen Geburtsjahrganggruppen ist eine Altersspanne von jeweils 16 Jahren in jedem Querschnitt abgedeckt – von 45 bis 60 im ersten Erhebungsjahr bis hin zu 60 bis 75 im Jahr 1999 – und über diese 16 verschiedenen 16-Jahres-Querschnitte hinweg betrüge die gesamte vorhergesagte Mittelwertdifferenz jeweils nur -0.075 . Dagegen betrüge die aus dem Modell allein aufgrund des Alterseffekts ohne den Kohorteneffekt vorhergesagte Differenz über 16 Altersjahre -0.48 , d.h. also über ungefähr eineinhalb Lebensjahrzehnte hinweg eine durchschnittliche Abnahme der Lebenszufriedenheit um einen halben Punktwert auf der Skala von 0 bis 10.

Bei den Bereichszufriedenheiten zeigt sich – nach der Untersuchung der längsschnittlichen graphischen Verlaufskurven erwartungsgemäß – keine derartige Überlappung annähernd gleich großer negativer Alters- und Kohorteneffekte. Allenfalls tendenziell ist dieses Muster bei der Gesundheitszufriedenheit zu erkennen, wo es ja auch nach den Eindrücken der graphischen Verlaufskurven in dieser abgeschwächten Form erwartet werden konnte. Allerdings ist hier der Kohorteneffekt statistisch nicht signifikant und auch schon deutlich kleiner, als der Alterseffekt. Abgesehen davon bestätigen hohe positive Alterseffekte

bei der Einkommens- und bei der Freizeitzufriedenheit nochmals den diesbezüglichen Eindruck, der aus den deskriptiven Analysen gewonnen wurde, nämlich daß es offenbar beim Übergang ins höhere Erwachsenenalter häufig zu einem Zuwachs der Zufriedenheit mit den Lebensumständen in diesen Bereichen kommt.

Da angesichts des Ergebnisses für die allgemeine Lebenszufriedenheit die Frage interessant scheint, ob und wie die gefundene Überlagerung von Alters- und Kohorteneffekt sich in den höheren Lebensaltersbereich hinein verändert, wurde das Multilevel-Modell nochmals für veränderte Geburtsjahresspannen berechnet, nämlich einmal für die 1919 bis 1934 Geborenen, d.h. im Panelstartjahr 50- bis 65-Jährigen und im letzten untersuchten Befragungsjahr 65- bis 80-Jährigen, sowie für die 1914 bis 1929 Geborenen und somit beim Panelstart bereits 55- bis 70-Jährigen. Dabei wurde jeweils wiederum der Nullpunkt der Geburtsjahresskala auf das jeweils früheste Geburtsjahr verschoben, also auf 1919 bzw. 1914 und der Nullpunkt der Altersskala auf das beim Panelstart älteste Lebensalter in der Untersuchungsgruppe, also auf 65 bzw. 70. Die Ergebnisse für die allgemeine Lebenszufriedenheit sind in Tabelle 3.5 abgebildet: Es zeigt sich bei ansteigendem Altersbereich und früherer Geburtsjahresspanne eine Zunahme des errechneten Alterseffekts und bei der ältesten Untersuchungsgruppe eine drastische Abnahme des Kohorteneffekts. Was die Zunahme der Alterseffekte anbelangt, so erschien diese erwartbar: Die vorherigen deskriptiven Untersuchungen deuteten bereits darauf hin, daß im höheren Alter eine Beschleunigung der durchschnittlichen Abnahme der Lebenszufriedenheitswerte stattfindet, so daß der Alterseffekt in diesem Bereich nicht mehr annähernd linear, sondern eher exponentiell nach unten verläuft. Bei den in Tabelle 3.5 gezeigten Analysen wurden gewissermaßen in zwei Schritten (gegenüber der ursprünglichen Analyse) Personen aus dem jüngeren Höheraltrigkeitsspektrum entfernt und solche aus einem älteren Altersspektrum hinzugefügt, der aus diesen Veränderungen der Altersstruktur der Gruppe resultierende Anstieg des Alterseffektes kann darum m.E. durchaus als Ausdruck einer solchen nichtlinearen Beschleunigung gewertet werden. Das Verschwinden des Kohorteneffekts dann, wenn in der Untersuchungsgruppe die Geburtsjahrgänge von 1930 bis 1939 entfernt und durch die Geburtsjahrgänge von 1914 bis 1923 ersetzt werden, ist m.E. ähnlich zu bewerten: Die in diesem frühen Geburtsjahreszeitraum erlebten zusätzlichen Lebensjahre, so scheint es, heben das durchschnittliche Niveau der Lebenszufriedenheit kaum mehr an, so daß im Bereich dieser frühesten Jahrgänge sozusagen eine Abschwächung des negativen Kohortentrends mit sinkender Geburtsjahrgangszahl stattfindet.

3.2.2.2 Zusammenfassung und Diskussion der Analysen zur Kohorteneffekthypothese

Es wurden in den SOEP-Daten gewisse Belege dafür gefunden, daß bei den Höheraltrigen – genauer: Während der Lebensaltersphase des Übergangs in die und Durchlaufens der „frühen Höheraltrigkeit“, d.h. ungefähr des sechzigsten und siebzigsten Lebensjahrzehnts

TABELLE 3.5:
Multilevel-Regressionsanalyse: Alters- und Kohorteneffekte der ALZ bei den 1919-1934 und den 1914-1929-Geborenen

	$G_i = 1919, \dots, 1934$ $n = 2042$	$G_i = 1914, \dots, 1929$ $n = 1768$
Konstante β_{00}	7.481***	7.269***
Alterseffekt $\beta_{10} = \alpha_{[\tau]}$	-0.065***	-0.095***
Kohorteneffekt $\beta_{01} = \gamma_{[\tau]}$	-0.040***	-0.017
Interaktion β_{11}	0.004***	-0.005***

*** $p \leq 0.001$; ** $p \leq 0.01$

– ein negativer Zusammenhang von Lebensalter und Lebenszufriedenheit besteht, jedoch durch einen ebenfalls negativen Zusammenhang von Geburtsjahr und Lebenszufriedenheit überlagert wird, so daß bei querschnittlicher Untersuchung ein Zusammenhang von Zufriedenheit und Alter in dieser Altersspanne praktisch nicht gefunden wird. Kurz gesagt: Die Kohorteneffekthypothese für die allgemeine Lebenszufriedenheit konnte durch die vorangegangenen Analysen gestützt werden.

Bei der Bewertung dieses Befunds ist jedoch vielerlei Vorsicht angebracht. Zunächst ist dabei die forschungsmethodische Problematik zu berücksichtigen, die die Untersuchung von Alters- und Kohorteneffekten mit sich bringt und die im Kapitel 2.1 ausführlich dargelegt wurde: Sofern man die prinzipielle Möglichkeit von Periodeneffekten miteinbezieht, kann man im Grunde niemals ganz sicher sein, *welche* Art von zeitvariablem Effekt man wie auch immer untersucht hat. Jegliche solche Untersuchung muß auf unbewiesene Vorabannahmen über einen oder alle dieser Effekte begründet werden und deshalb stets gewisse Ungenauigkeiten in Kauf nehmen. Im Falle des von berechneten Multilevel-Regressionsmodells bestehen m.E. diese Ungenauigkeiten v.a. in zweierlei Hinsicht: Zum einen bedeutet natürlich die angenommene Linearität der Alters- und Kohorteneffekte eine starke Restriktion, zum anderen sind die berechneten Koeffizienten nur unter der Annahme geringfügiger Periodeneffekte als unverzerrter Ausdruck von Alters- und Kohorteneffekten zu bewerten.

Die Linearitätsrestriktion für Alters- und Kohorteneffekte ist damit zu rechtfertigen, daß das Interesse hier nur einem globalen linearen Trend über eine eingeschränkte Spanne von Geburtsjahrgängen und Lebensaltern der während der SOEP-Erhebungsphase Höheraltrigen gilt. Wenn also beispielsweise über das gesamte Spektrum aller Geburtsjahrgänge der letzten hundert Jahre hinweg auch kein wirklich linearer Zuwachs oder Verlust durchschnittlicher Zufriedenheit pro Geburtsjahr stattfindet – und eine solche Annahme erschiene auch kaum vernünftig begründbar –, so mag es doch möglich sein, daß die über dieses gesamte Spektrum hinweg stattfindenden Trendwechsel im Zusammenhang von Geburtsjahr und Zufriedenheit, welche eine Nichtlinearität dieses Zusammenhangs bewirken, doch in kleineren Ausschnitten dieses Geburtsjahresspektrums nicht so gravierend sind, daß sie *darin* nicht noch mit einer linearer Trendanpassung „aufzufangen“ wären. Ähnlich kann

man es für den Zusammenhang von Lebensalter und Zufriedenheit überlegen, so daß also bezüglich der Problematik der Linearitätsannahme auf die den oben gezeigten Analysen zugrundegelegten Untersuchungsgruppen zu verweisen ist, welche durch ihr jeweils nicht allzu breites Spektrum einbezogener Geburtsjahrgänge eine gewisse Rechtfertigung für die Modellierung der Alters- und Kohorteneffekte als lineare Trends über die jeweilige Zeitdimension hinweg liefern.

Schwieriger zu bewerten erscheint allerdings der Umstand, daß dabei implizit auch die möglichen Periodeneffekte als linearer Trend über den Datenerhebungszeitraum hinweg modelliert wurden (vgl. Kapitel 2.1.2, Gleichung 2.1). Ein vorhandener, jedoch im statistischen Modell (Gleichung 2.11) nicht berücksichtigter Periodeneffekt wäre, wie gezeigt wurde, auf jeden Fall in den durch dieses Modell berechneten Alters- und Kohorteneffekten enthalten – und wenn er linear wäre, dann praktischerweise solchermaßen, daß er als jeweils additiver Beitrag gleichermaßen in den geschätzten Alters- und Kohorteneffektwerten enthalten ist, so daß die relative Größe des tatsächlichen Alters- im Vergleich zum tatsächlichen Kohorteneffekt dieser beiden Effekte in der Differenz der beiden geschätzten Effektwerte durchaus zum Ausdruck käme (vgl. die Gleichungen 2.6 und 2.7, Kapitel 2.1.2). Tatsächlich aber konnte die Annahme der Linearität möglicher Periodeneffekte ebensowenig rational begründet werden, wie die Annahme irgendeines konkreten nichtlinearen solchen Effekts während der Erhebungszeitspanne. Einzig eine in den längsschnittlichen Verläufen der durchschnittlichen Lebenszufriedenheiten verschiedener Geburtskohortengruppen gefundener leichter „Gipfel“ um das Jahr 1991 herum (vgl. Abbildungen 3.6 und 3.7) konnte als Hinweis auf einen kurzfristigen positiven Periodeneffekt bewertet werden, für den sich der zu diesem Zeitpunkt in Deutschland stattfindende historische Umbruch als nur allzu plausible ad hoc Erklärung anbot. Allerdings zeigte sich dieser kurze Ausschlag nicht so extrem, daß er nicht sozusagen einfach durch ein lineares Modell hinweggemittelt werden dürfte, ohne dabei deutliche Verzerrungen der berechneten groben linearen Trends befürchten zu müssen. Abgesehen davon bleibt festzuhalten, daß das, was in den berichteten Ergebnissen als Alterseffekt $\alpha_{[t]}$ bzw. Kohorteneffekt $\gamma_{[t]}$ bezeichnet wurde, möglicherweise weitestgehend ein in den Erhebungsjahren wirksamer Periodeneffekt sein könnte.

Bei der Bewertung der Relevanz dieser Ungenauigkeit für die Interpretation der vorgelegten Befunde ist allerdings zu bedenken, ob sich für eine Hypothese eines über den langen Zeitraum der 16 Erhebungsjahre hinweg vorhandenen Periodeneffekts ebenso theoretische Gründe finden ließen, wie für die Vermutung von Alters- und Kohorteneffekten, welche im Theorieteil dieser Arbeit begründet wurde. Die Analysen dienten der Überprüfung dieser begründeten Kohorteneffekthypothese, bei deren Gültigkeit ein Ergebnis wie das für die Lebenszufriedenheit gefundene auftreten sollte (abgesehen von der nun allerdings schon sehr abstrakten Möglichkeit, daß es durch entsprechende Interaktionen mit einem möglichen Periodeneffekt verschleiert würde) – wenn es also auftrat, so konnte diese Hypothese nicht widerlegt werden und die Aussage ihrer Bestätigung in den untersuchten Daten wird

durch die nicht ausgeschlossene Möglichkeit eines Periodeneffekts im Sinne einer Störhypothese (Hager & Westermann, 1983) lediglich eingeschränkt. Oder anders ausgedrückt: Die Kohorteneffekthypothese konnte nicht widerlegt werden. Gegen die Störhypothese eines Periodeneffekts wäre dazu aber auch noch ihre mangelnde theoretische Begründung ins Feld zu führen, so daß man sagen könnte: Unter der Voraussetzung eines negativen Alters- und Kohorteneffekts einerseits *oder* unter der Voraussetzung eines negativen Periodeneffekts andererseits wären die oben gezeigten Befunde gleichermaßen wahrscheinlich, umgekehrt erscheint aber unter der Voraussetzung der Befunde das Vorhandensein von negativem Alters- und Kohorteneffekt zumindest subjektiv wahrscheinlicher, als das eines negativen Periodeneffekts.

Zudem müßte ein linearer Periodeneffekt „in“ $\alpha_{[\tau]}$ und $\gamma_{[\tau]}$ dann auch bei Berechnung in durch unterschiedliche Geburtsjahresspannen definierten Gruppen stets derselbe sein, da bei diesen Analysen ja stets die „latente“ Prädiktorvariable Meßzeitpunkt in denselben Ausprägungen variiert. Nun ergab aber die Untersuchung der Gruppe der 1914 bis 1929 Geborenen einen niedrigen Kohorteneffektwert $\gamma_{[\tau]}$, welcher somit gewissermaßen als Obergrenze für den möglicherweise in den verschiedenen Koeffizientenschätzungen enthaltenen Periodeneffektwert τ aufgefaßt werden könnte – wenn man annimmt, daß keine Interaktion von Perioden- mit Alters- bzw. Kohorteneffekt vorlag.

Die Ergebnisse wären somit bei allen Einschränkungen als solche zu bewerten, die die Kohorteneffekthypothese stützen. Von methodischen Spitzfindigkeiten abgesehen, wäre allerdings auch noch zu diskutieren, welche Relevanz man ihnen unter dem Gesichtspunkt der absoluten Effektgröße beimessen würde. Für die Diskussion um das Zufriedenheitsparadox des höheren Lebensalters wäre zu fragen, ob der hier trotz weitgehender querschnittlicher Mittelwertstabilität gefundene negative Alterseffekt groß genug ist, um letztendlich diese ganze Diskussion in Frage zu stellen mit der simplen Feststellung, daß das sogenannte Paradox sich hier letztlich nur als Artefakt erwiesen habe und ein negativer Alterseffekt vorliegt, welcher alle Erwartungen eines alterskorrelierten Verlustes von Lebenszufriedenheit erfüllt und weitere Theoriedebatten über „stability despite loss“ erübrigt. Zu einer definitiven Beurteilung dieses Aspektes praktischer Bedeutsamkeit der gefundenen Alterseffektwerte fehlen m.E. rationalen Kriterien: Wann ist sozusagen der Abfall des Mittelwertverlaufs über das höhere Alter steil genug, um in diesem Sinne bedeutsam zu sein? Diese Frage ist mit den hier vorgelegten Ergebnissen nicht zu beantworten – als Ergebnis der vorliegenden Untersuchung ist lediglich festzuhalten, daß ein solcher Abfall festgestellt werden konnte. Als vager Anhaltspunkt für derartige Bewertungen, der aus den Daten gewonnen werden kann, könnte die Streuung der Meßwerte dienen: In der untersuchten Gruppe betrug die Standardabweichung der Lebenszufriedenheitswerte insgesamt, d.h. über alle Personen und Meßzeitpunkte hinweg (wobei also jede Person mehrere Meßwerte beiträgt), 1.90 (vgl. auch die Werte in Tabelle B.2 in Anhang B). Gemessen an diesem Wert betrüge die vorhergesagte Reduktion des Lebenszufriedenheitswertes pro Lebensjahrzehnt ungefähr 0.17 Standardabweichungen. Wie auch immer man diesen Wert bewerten möchte: Die Feststel-

lung eines signifikanten negativen Alterseffekts erscheint angesichts des weitverbreiteten Eindrucks von im höheren Alter weitestgehend stabiler, intraindividuell kaum veränderter Lebenszufriedenheit, der aus vielen querschnittlichen Studien erzeugt wurde, nicht ganz unwichtig.

Die Fragen nach Stabilität und Veränderung der Lebenszufriedenheit im höheren Alter aber, die durch die Diskussion um das Zufriedenheitsparadox nun einmal aufgeworfen wurden, sind, wie vorne theoretisch überlegt wurde und in Kapitel 2.2.1 noch eher formal gezeigt werden wird, nicht durch die Bewertung der Größe eines Alterseffekts, d.h. der systematischen über das Lebensalter hinweg stattfindenden *durchschnittlichen* Veränderung, zu beantworten. Sie betreffen die Varianz der im höheren Alter stattfindenden intraindividuellen Veränderungen genauso bzw. noch viel mehr, wie deren Durchschnittswert und werden erst in den noch folgenden Stabilitätsanalysen eingehender untersucht.

Daß die gemäß der Kohorteneffekthypothese vermutete simultane Überlagerung von Alters- und Kohorteneffekt hier nicht für die Bereichszufriedenheiten gefunden wurde, schmälert m.E. die Bedeutung des Befundes für die allgemeine Lebenszufriedenheit nicht. Es sei an das zum Ende des Kapitels 1.3 Gesagte erinnert: Ergebnisse für die Bereichszufriedenheiten werden ergänzend berichtet, um Anhaltspunkte für mögliche strukturelle Ähnlichkeiten einzelner Bereichsbeurteilungen zur Globalbeurteilung zu gewinnen. Mit Blick auf die Ergebnisse des Multilevel-Regressionsmodells erscheinen sie nun allesamt der allgemeinen Lebenszufriedenheit eher strukturell unähnlich. Im Vergleich zur Lebenszufriedenheit zeigt sich folgendes: Ein deutlich geringerer und insignifikanter Kohorteneffekt der Einkommens- und Gesundheitszufriedenheit, jedoch ein ähnlicher und statistisch signifikanter Kohorteneffekt für Wohn- und Freizeitzufriedenheit, ein niedrigerer und insignifikanter Alterseffekt bei der Wohnzufriedenheit, während Einkommens- und Freizeitzufriedenheit statistisch signifikante *positive* Alterseffekte aufweisen. Mögliche ad hoc Erklärungen dafür könnten sein: Eine größere „Realitätsnähe“ der Gesundheitszufriedenheit: Möglicherweise wirken sich gesundheitliche Einschränkungen im Alter unmittelbarer auf die entsprechende Bewertung aus und ist letztere weniger durch ein aus der gesamten Lebenserfahrung heraus ausgebildetes zeitlich stabiles Anspruchsniveau relativiert. Dieses könnte deshalb der Fall sein, weil gesundheitliche Einschränkungen unmittelbarer körperlich erlebt werden und affektive Reaktionen hervorrufen (ganz im Sinne der bereits zitierten Unterscheidung von „needs“ und „wants“, die Veenhoven, 1996, gegen die Relativität von Zufriedenheitsurteilen ins Feld führt) und/oder weil der Gesundheitsbereich derjenige ist, in dem sich im Leben alter Menschen am meisten ereignet, so daß hier die Aufmerksamkeit für objektive Information heraufgesetzt ist. Möglicherweise sind Bewertungen des Wohnens und der Freizeit besonders „anfällig“ für Kohorteneffekte, weil einerseits hier die objektiven Ausstattungen in der gegenwärtigen Wohlstandsgesellschaft deutlich mit durchschnittlichen Vor- und Nachkriegsbedingungen kontrastieren und andererseits sich der Bereich Wohnen im Alter nur wenig bzw. der Freizeitbereich sich zum Positiven verändert (zumindest was das Ausmaß freier Zeit nach der Verrentung betrifft). Für die Einkommens-

zufriedenheit schließlich könnte wiederum größere „Realitätsnähe“ zur Erklärung des geringen Kohorteneffekts vermutet werden, für den positiven Alterseffekt bieten sich m.E. allerdings keine einfachen Erklärungen auf Anhieb an.

3.3 Analysen zur Zufriedenheitsstabilität im höheren Lebensalter

Den Anlaß zu den Analysen zur Stabilität von Zufriedenheit, die im folgenden dargestellt werden, geben diejenigen der im Kapitel 1.2 begründeten Hypothesen, welche jeweils in Vermutungen über die statistischen Kennwerte, die im Kapitel 2.2.1 behandelt wurden, übersetzt werden können: Nämlich die Gewinner-und-Verlierer-Hypothese, sowie die Hypothesen allgemeiner und erhöhter Stabilität der Zufriedenheit. Es sei hier noch einmal ausdrücklich auf die in Kapitel 1.3 aufgeführten Überlegungen zur logischen Beziehungsstruktur dieser Hypothesen untereinander erinnert: Sie sind nicht strikt logisch miteinander verknüpft in dem Sinne, daß es sich dabei um wechselseitig ausschließliche Alternativerklärungen des Phänomens einer häufig vorgefundenen paradoxen Mittelwertsstabilität der Lebenszufriedenheit im höheren Erwachsenenalter handelt.

Die Gewinner-und-Verlierer-Hypothese eines primär positiven Alterseffekts bei einem substantiellen Anteil von Zufriedenheitsverlierern im höheren Lebensalter impliziert keine allzu hohe Merkmalsstabilität im höheren Erwachsenenalter und stattdessen wachsende Varianzen im längsschnittlichen Verlauf, hervorgerufen durch viele Zufriedenheitsveränderungen in beide Richtungen, unabhängig vom Ausgangsniveau der Zufriedenheit. Dieses würde bedeuten, daß bei den Höheraltrigen keine ausgeprägte monotone und somit keine auch nur näherungsweise parallele (und natürlich erst recht keine strikte) Stabilität beobachtet werden kann. Vielmehr wurde aus diesem Effekt eine alterskorrelierte Abnahme der Stabilität von Zufriedenheit in der Phase des höheren Lebensalters vorhergesagt.

Die Hypothesen allgemeiner (lebenslanger) und im Alter erhöhter Stabilität von Zufriedenheit beruhen auf der Annahme verschiedener stabilisierender Effekte auf die Zufriedenheiten älterer Menschen. Die Frage ist hier, wie die darin postulierte hohe Stabilität hinsichtlich der im Kapitel 2.2.1 besprochenen Kennwerte bzw. Stabilitätstypen zu präzisieren ist. Tatsächlich erklären diese Hypothesen eine quer- und (möglicherweise auch längsschnittliche) Invarianz von Altersgruppenmittelwerten bei den Höheraltrigen mit einer weitestgehenden intraindividuellen Invarianz von deren Zufriedenheit – d.h., es ist im Grunde eine strikte Stabilität gemeint. Jedoch versteht es sich aus den im Kapitel 1 angestellten Überlegungen heraus von selbst, daß eine solche strikte Stabilität nicht „absolut“, sondern nur näherungsweise zu erwarten ist.

So wurde die Adaptivität von Zufriedenheitsurteilen als bedeutsamer stabilisierender Effekt benannt, welcher paradoxerweise eine gewisse intraindividuelle Variabilität impliziert. Denn adaptiv reguliert werden können ja nur Veränderungen in der Zufriedenheit,

welche infolge dieser Adaptation dann nur kurzfristig Bestand haben, so daß bei Beobachtung über einen genügend großen Zeitraum hinweg nur wenig absolute Veränderung wahrgenommen wird. Dieses würde aber bedeuten, daß zu jedem Beobachtungszeitpunkt ein kleiner Anteil solcher Personen erfaßt wird, die sich aktuell sozusagen in der Adaptationsphase mit vorübergehend entsprechend herauf- oder herabgesetzter Zufriedenheit befinden – wodurch eine gewisse Minderung von Stabilitätskoeffizienten bewirkt würde, welche ansonsten allerdings sehr hoch sein müßten, da ja infolge der Adaptivität zu jedem Zeitpunkt ein weitaus größerer Anteil der Personen sich außerhalb einer solchen Adaptationsphase auf ihrem normalen persönlichen Zufriedenheitsniveau befinden.

Aus den beiden „Stabilitätshypothesen“ wäre also bei Höheraltrigen außer einem hohen Stabilitätskoeffizienten eine gewisse, relativ geringe Varianz der in einer Beobachtungsphase stattfindenden intraindividuellen Veränderungen – bzw. eine näherungsweise strikte Stabilität – zu erwarten. Diese Erwartung enthält einen Interpretationsspielraum in der Frage, wie hoch ein „hoher“ Stabilitätskoeffizient und wie gering eine „relativ geringe“ Veränderungsvarianz ist. Für einen gegebenen Stabilitätskoeffizientenwert kann diese Frage leichter beantwortet werden, da es sich dabei ja um einen Korrelationswert handelt, dessen alltägliche und vielfältige Verwendung in allen empirischen Forschungszusammenhängen gewissermaßen einen Referenzrahmen für solche Einschätzungen liefert – beispielsweise könnten hier die von Cohen (1992) vorgeschlagenen Daumenregeln zur Einstufung eines Korrelationswertes als „niedriger“, „mittlerer“ oder „hoher“ Effekt benutzt werden. Zur Einschätzung einer vorhandenen Veränderungsvarianz kann dagegen nicht auf gewohnheitsmäßige Regeln zurückgegriffen werden, es geht hier v.a. darum, diese Varianz zu einem stabilen, zeitlich invarianten Varianzanteil in Beziehung zu setzen. Ein Vorschlag dazu wurde im Kapitel 2.2.1 vorgestellt und begründet: Der Quotient Q_{Δ} (vgl. Gleichung 2.20) kann so, wie dort diskutiert, zur Beurteilung einer bloß näherungsweise parallelen Stabilität oder, was im Grunde dasselbe meint, der relativen Größe der Veränderungsvarianz verwendet werden.

Beide Stabilitätshypothesen behaupten stabilisierende Effekte auf Zufriedenheitsurteile, jedoch im einen Fall solche, die lebenslang (zumindest über die gesamte Spanne des Erwachsenenalters hinweg) wirken, und im anderen solche, die im Lebenslauf sich entwickeln und im höheren Alter immer stärkeren Einfluß entfalten. Es besteht also der Unterschied hier in der Annahme lebenslang gleichbleibend hoher versus im höheren Alter erhöhter Stabilität. Zudem impliziert ja, wie oben bereits gesagt, die Gewinner-und-Verlierer-Hypothese im Alter sinkende Stabilität. Es muß deshalb bei den folgenden Stabilitätsanalysen nicht nur Ausmaß der Merkmalsstabilität im höheren Alter analysiert werden, sondern auch die Entwicklung der Stabilität in dieser Lebensphase. Es soll außer der intraindividuellen Veränderung sozusagen auch die Veränderung dieser Veränderung untersucht werden. Prinzipiell ermöglicht das SOEP mit den 16 jährlichen Wiederholungen verschiedener Zufriedenheitsratings solchermaßen doppelt längsschnittliche Analysen: Es kann damit eine Zeitreihe von 15 Stabilitätskoeffizienten oder sonstiger stabilitätsrelevanter Kennwerte (Veränderungs-

rianzen, Q_{Δ} usw.) auf der Basis sukzessiver Erhebungswellen berechnet und anhand dieser Zeitreihe deren Veränderung untersucht werden.

Für die Analysen wurde aus den in Frage kommenden längsschnittlichen Strukturgleichungsmodellen, die im Kapitel 2.2.2 diskutiert wurden, das QMSM ausgewählt. Gründe für diese Auswahl wurden im Rahmen der Darstellung dieser Modelle bereits angegeben, sie lassen sich knapp so zusammenfassen: Das QMSM erlaubt eine hinsichtlich Stabilitätstypus und Veränderungsverlauf „unvoreingenommene“ Analyse und zudem die unkomplizierteste Schätzung der Stabilitätskoeffizienten, die ja die zentralen Analyseparameter der Untersuchung sind. Auf der Grundlage der Modellierung der vorliegenden Zeitreihen von Zufriedenheitsmessungen mit dem QMSM werden die Gewinner-und-Verlierer-Hypothese, die Hypothese allgemeiner Stabilität und die Hypothese erhöhter Stabilität auf die Weise untersucht, daß die daraus sich ergebenden Annahmen in Parameterrestriktionen des Modells übersetzt und solchermaßen unterschiedlich restringierte Modelle miteinander verglichen werden. D.h. es werden zur Überprüfung der Hypothesen, Vergleiche von „nested models“ durchgeführt (vgl. die diesbezügliche Besprechung im Kapitel 2.3.2).

So impliziert die Gewinner-und-Verlierer-Hypothese Veränderung – nämlich längsschnittliches Wachstum – der True-Score-Varianzen von Zufriedenheit und einen Rückgang der Stabilitätskoeffizienten bei Altersgruppen Höheraltriger. Wenn diese Effekte bedeutsam wären, dann sollte ein Modell, bei dem für die Höheraltrigen über die sukzessiven jährlichen Erhebungen hinweg die Invarianz von True-Score-Varianzen und Stabilitätskoeffizienten festgelegt ist, eine deutlich schlechtere Anpassung an die vorhandenen Daten resultieren, als bei einem diesbezüglich unrestringierten Modell, bei dem dann natürlich auch die entsprechenden Parameterschätzungen ein Wachstum der Varianzen und eine Schrumpfung der Stabilität zeigen sollten. Würde dagegen mit einem diesbezüglich restringierten Modell eine nahezu gleich hohe Modellanpassung erzielt, so bedeutet dieses, daß die durch die Gewinner-und-Verlierer-Hypothese postulierten Effekte zumindest nicht dergestalt ausschlaggebend gewirkt haben, daß durch ein Veränderungsmodell ohne diese Effekte die vorhandenen Daten nicht mehr erklärbar sind.

In ähnlicher Weise sind auch die Hypothesen allgemein (lebenslang) hoher bzw. im Alter erhöhter Stabilität überprüfbar: Insbesondere führt letztere zu der Annahme, daß in höheren Altersgruppen die sukzessiven Stabilitätskoeffizienten ansteigen. Dagegen enthält die Hypothese lebenslang hoher Merkmalsstabilität zwar keinen expliziten Widerspruch zur Wirkung zusätzlich stabilisierender Effekte im höheren Alter, aber immerhin könnte man daraus die Schlußfolgerung ableiten, daß damit auch im höheren Lebensalter sozusagen keine allzu großen „Steigerungspotentiale“ mehr vorhanden sind. Somit postulieren beide Hypothesen hohe Stabilitätskoeffizienten Höheraltriger, wobei allerdings dann, wenn in diesen Altersgruppen Effekte bedeutsam wären, welche eine Zunahme der Stabilität bewirken, dieses dazu führen müßte, daß ein Modell mit der Restriktion der Invarianz der sukzessiven Stabilitätskoeffizienten eine schlechtere Anpassung an die Daten erzielt, als ein diesbezüglich unrestringiertes. Wäre dieses nicht der Fall und wäre ein durch ein solchermaßen

restringiertes Modell geschätzter invarianter Stabilitätskoeffizient bei den Höheraltrigen als „sehr hoch“ einzuschätzen, dann würde dies eine allgemeine Stabilität belegen, die im untersuchten Zeitverlauf allenfalls geringfügig angestiegen wäre.

Es wurden also verschiedene Versionen des QMSM für die SOEP-Zeitreihen von jeweils 16 sukzessiven (allgemeinen oder bereichsspezifischen) Zufriedenheitsratings mit unterschiedlichen Invarianzrestriktionen für True-Score-Varianzen und Stabilitätskoeffizienten berechnet und deren Ergebnisse verglichen. Ein spezielles designtechnisches Problem besteht in der Frage, wie dabei dem thematischen Fokus auf das höhere Lebensalter Rechnung getragen werden kann. Denn gerade die Untersuchung von „Veränderung der Veränderung“ in höheren Altersgruppen bringt eine gewisse Problematik für die Definition letzterer mit sich: Es muß entschieden werden, aufgrund welchen Lebensalters zu welchem Meßzeitpunkt eine Zuordnung zu einer Altersgruppe Höheraltriger erfolgt. Mit anderen Worten: Es muß entschieden werden, welche Geburtskohorten die Höheraltrigen bilden sollen.

Wenn beispielsweise nur eine Globalgruppe älterer Menschen untersucht werden sollte, mit einem Lebensalter von mindestens 60 Jahren als Definitionskriterium der Höheraltrigkeit, dann wäre zunächst noch zu entscheiden, die zu welchem Meßzeitpunkt mindestens 60 Jahre alten SOEP-Befragungspersonen in diese Substichprobe Höheraltriger aufgenommen werden. Würde hierfür ein Referenzzeitpunkt festgelegt, z.B. die Erstbefragung 1984, so daß die Gruppe der zum Panelbeginn mindestens 60 Jahre alten Personen über die folgenden 16 Jahre längsschnittlich analysiert wird, dann könnte ein Problem darin liegen, daß damit bei der Erfassung von „Veränderung der Veränderung“ jeweils (von Jahr zu Jahr) über ein sehr breites Altersspektrum hinweg aggregiert wird. Die hypothetischen längsschnittlichen Effekte, die im höheren Lebensalter z.B. zu einer allmählichen Zunahme der Varianz oder der Stabilitätskoeffizienten führen könnten, hätten dann bei den Mitgliedern dieser Substichprobe bereits zum ersten Meßzeitpunkt (und natürlich auch zu allen folgenden) in sehr unterschiedlicher Dauer gewirkt. Dieses wäre dann kein Problem, wenn diese Effekte sich über die gesamte höheraltrige Lebensphase fortgesetzt gleichermaßen auswirken, sich also gewissermaßen von Lebensjahr zu Lebensjahr weiter aufsummieren, so daß ein Anstieg von Varianzen oder Stabilitäten über diese gesamte Phase bis hin zum Tode linear oder zumindest monoton stattfindet: Dann wären diese Effekte im Prinzip bei allen Mitgliedern einer solchermaßen breit – über ein großes Spektrum von Geburtsjahrgängen hinweg – definierten Stichprobe zu jedem Erhebungszeitpunkt wirksam, so daß sie in Kennwerten, die über die Stichprobe aggregieren, sozusagen unverfälscht erfaßbar wären. Verhielte es sich jedoch bei diesen Effekten beispielsweise so, daß sie nur während einer Phase früher Höheraltrigkeit wirksam wären, um sich dann allmählich abzuschwächen, dann würde das bedeuten, daß nur bei den „jüngeren Alten“ die gemäß obiger Hypothesen möglichen Veränderungen von Varianzen und Stabilitäten stattfinden, bei den „älteren Alten“ dagegen nicht mehr – in der Gesamtstichprobe aller bei der Ersterhebung Älterer ab 60 würden dann während der ersten Erhebungsjahre Fälle mit viel bzw. wachsender und Fälle mit wenig bzw. wieder sinkender „Veränderung der Veränderung“ aggregiert, so daß

möglicherweise das wahre Ausmaß dieser Effekte in der frühen Höheraltrigkeit nicht erfaßt würde.

Diesem Problem könnte natürlich dadurch begegnet werden, daß Altersgruppen Höheraltriger durch engere – d.h. auf kürzere Geburtsjahrgangsspektren begrenzte – Geburtskohorten definiert werden. Wenn alle 16 *jährlichen* Wiederholungsmessungen genutzt werden sollen, dann wären idealerweise jeweils Ein-Jahres-Geburtskohorten längsschnittlich zu analysieren, so daß innerhalb einer solchen Untersuchungsgruppe bei der Berechnung der „Veränderung der Veränderung“ nur über Individuen aggregiert wird, die zu jedem Zeitpunkt dasselbe Alter haben. Ein ganz banales technisches Problem bestünde hier allerdings darin, daß im SOEP die Stichprobengrößen dieser Ein-Jahres-Altersgruppen teilweise sehr gering würden, was für die Berechnung von Strukturgleichungsmodellen insofern von Belang ist, als dieses Verfahren möglichst große Stichproben voraussetzt und die mögliche Fehlerhaftigkeit von Parameterschätzungen bei sehr kleinen Stichproben nicht übersehen werden darf (vgl. dazu z.B. Raykov & Widaman, 1995; Hu & Bentler, 1995). Dieses Problem wird noch dadurch verschärft, daß es natürlich gerade bei den höheraltrigen Menschen im Panelverlauf zu einer hohen Zahl von Ausfällen infolge von Mortalität kommt.

Um den somit angesprochenen Problemen der Untersuchungsgruppendefinition Rechnung zu tragen, wurden Stabilitätsanalysen mit dem QMSM in zwei Teilen durchgeführt:

- ▶ Es wurden Berechnungen für die Daten der breit definierten Substichprobe aller SOEP-Befragungspersonen, die bei Panelbeginn 1984 mindestens 60 Jahre alt waren, durchgeführt. Zudem werden die Ergebnisse dieser Berechnungen kontrastiert mit denen für die Gesamtstichprobe aller SOEP-Befragungspersonen. Auf diese Weise kann gewissermaßen ein allgemeines Bild der Merkmalsstabilität von Zufriedenheit im höheren Lebensalter im untersuchten Zeitraum von 1984 bis 1999 gewonnen und können deutliche systematische Unterschiede in der Stabilität „der Alten“ von der durchschnittlichen Merkmalsstabilität der Gesamtbevölkerung entdeckt werden. Diese Analysen sind im Unterkapitel 3.3.1 beschrieben.
- ▶ Dem oben formulierten Bedürfnis nach eng gefaßten Altersgruppen mit hinreichenden Stichprobengrößen wurde dadurch Rechnung getragen, daß Altersgruppen auf der Basis von Vier-Jahres-Geburtskohorten gebildet und längsschnittlich deren „Veränderungen der Veränderung“ in Vier-Jahres-Intervallen analysiert wurden – diese Vorgehensweise wird im folgenden noch genauer erläutert werden. Diese Analysen sind im Unterkapitel 3.3.2 beschrieben.

Zur Berechnung der Strukturgleichungsmodelle wurde das Programm AMOS 4 benutzt (Arbuckle & Wothke, 1999) und der darin implementierte und vorne in Kapitel 2.3.1.3 dargestellte FIML-Algorithmus zur Schätzung der Modellparameter angewandt (Arbuckle 1996). Die FIML-Schätzung bei fehlenden Werten erfordert die Spezifikation einer Kovarianz- und einer Mittelwertsstruktur für die Modellvariablen. Letztere wurde in allen Modellen auf die im Kapitel 2.2.2.1 diskutierte Weise festgelegt, nämlich durch die Aufnahme von Intercepts in die Meßmodellgleichungen für die beobachteten Variablen X_t , d.h. das QMSM ist

durch die in Kapitel 2.2.2.1 aufgeführten Gleichungen 2.30 und 2.24 definiert.

Die Analysen mit den verschiedenen Versionen des QMSM wurden wiederum jeweils für alle in den 16 Erhebungen vollständig erfragten Zufriedenheiten durchgeführt, d.h. es wurden die Zeitreihen der Meßwerte allgemeiner Lebenszufriedenheit, sowie der Zufriedenheiten mit der Gesundheit, mit dem Einkommen, mit der Wohnung, sowie mit der Freizeit analysiert.

3.3.1 Stabilität von Zufriedenheit in der Gesamtgruppe und bei älteren Befragungspersonen

Wie oben dargelegt wurde in einem ersten Auswertungsschritt eine allgemeine Analyse der Merkmalsstabilität von Zufriedenheit im höheren Lebensalter anhand einer breit definierten Höheraltrigengruppe unternommen: Diese wurden aus den Befragten der westdeutschen SOEP-Teilstichprobe mit Geburtsjahr 1924 oder früher gebildet, es handelt sich also um die bei der Erstbefragung im Jahr 1984 mindestens 60 Jahre alten Personen (Bezeichnung: Alt84), deren Zufriedenheiten somit ab 1984 über 16 Jahre hinweg mittels eines Modells vom Typ QMSM längsschnittlich untersucht wurden. Den Ergebnissen dieser Gruppe werden diejenigen der Gesamtgruppe aller westdeutschen SOEP-Befragten gegenübergestellt.

Der Einfachheit halber werden diese Analysen im folgenden auch als *allgemeine Stabilitätsanalysen* bezeichnet.

3.3.1.1 Deskription der Analysestichproben für die allgemeinen Stabilitätsanalysen

Zunächst die Ergebnisse einiger vorbereitenden deskriptivstatistischen Auswertungen zu den beiden Untersuchungsgruppen der allgemeinen Stabilitätsanalysen:

Tabelle 3.6 informiert über die tatsächlichen Fallzahlen dieser beiden Untersuchungsgruppen, d.h. die Anzahlen der für die jeweiligen Erhebungswellen tatsächlich vorhandenen Werte der allgemeine Lebenszufriedenheit (ALZ), Einkommenszufriedenheit (EIZ), Gesundheitszufriedenheit (GEZ), Wohnzufriedenheit (WOZ) und Freizeitzufriedenheit (FRZ). Diese Fallzahlen gewährleisten in jedem Fall die Durchführbarkeit der FIML-Schätzung der Strukturgleichungsmodelle, ohne daß eine durch zu kleine Stichproben bedingte Instabilität der Parameterschätzungen befürchtet werden muß (vgl. Raykov & Widaman, 1995). Wie nicht anders zu erwarten, ist die Gruppe Alt84 in deutlich stärkerem Maße von dem unvermeidlichen Stichprobenausfall im Panelverlauf betroffen, als die westdeutsche Stichprobe insgesamt: Während letztere über die 16 Jahre hinweg nur ca. ein Drittel ihrer Mitglieder verlor, reduzierte sich die Höheraltrigengruppe auf nahezu ein Fünftel ihrer Ursprungsgröße. Oder, anders ausgedrückt: Von den ungefähr 3000 Ausfällen, die in der Gesamtgruppe zu verzeichnen sind, entfiel ca. die Hälfte auf Personen, die vor 1925 geboren wurden.³⁰ Natürlich handelt es sich dabei in erster Linie um Ausfälle infolge der Mortalität der

³⁰Bei dieser Aussage ist allerdings davon abzusehen, daß durch das Weiterverfolgungskonzept des SOEP

TABELLE 3.6:
Stichprobenumfänge bei den allgemeinen Stabilitätsanalysen

Jahr	ALZ		EIZ		GEZ		WOZ		FRZ	
	Alt84	Alle	Alt84	Alle	Alt84	Alle	Alt84	Alle	Alt84	Alle
1984	1954	9039	1955	8911	1963	9060	1954	9011	1949	9020
1985	1652	8343	1649	8200	1654	8333	1652	8286	1643	8307
1986	1449	7980	1442	7904	1446	7990	1446	7965	1435	7970
1987	1356	7842	1355	7785	1358	7860	1352	7808	1347	7840
1988	1208	7450	1211	7390	1211	7466	1210	7444	1202	7448
1989	1095	7174	1093	7118	1096	7187	1095	7163	1088	7164
1990	1033	7005	1035	6972	1038	7028	1036	7009	1035	7020
1991	945	6898	957	6905	958	6961	960	6956	949	6943
1992	873	6780	876	6763	874	6804	872	6790	862	6782
1993	809	6718	794	6566	810	6735	809	6704	795	6702
1994	737	6595	731	6460	746	6625	745	6595	728	6593
1995	654	6522	652	6409	659	6543	659	6506	659	6540
1996	579	6442	574	6306	579	6449	578	6408	570	6428
1997	540	6366	530	6223	541	6363	542	6327	534	6339
1998	485	6169	479	6016	484	6161	486	6131	479	6149
1999	424	6033	417	5901	424	6022	423	5993	420	5999

älteren Teilnehmer: Tatsächlich entfallen von den 1145 Todesfällen, die der westdeutsche Datensatz im untersuchten Zeitraum verzeichnet, 798 auf die Gruppe Alt84.

Mittelwerte und Standardabweichungen der beobachteten Zufriedenheitswerte in den beiden Untersuchungsgruppen zeigen die Tabellen 3.7 und 3.8. Hier fällt zunächst bei den ALZ-Mittelwerten der Höheraltrigen ein auf den ersten Blick durchaus nicht ganz geringfügiger Rückgang über die untersuchte Zeitspanne hinweg auf, welcher der in der wissenschaftlichen Diskussion des Zufriedenheitsparadoxes implizierten Grundannahme, daß kein alterskorrelierter Abfall von durchschnittlicher Zufriedenheit im höheren Alter stattfindet, widerspricht. Allerdings ist hier folgendes zu bedenken: Zum ersten zeigt Tabelle 3.7 eine längsschnittliche Folge von Mittelwerten und es wurde in den theoretischen Überlegungen im Kapitel 1.2 ja bereits großer Nachdruck darauf gelegt, daß eine Invarianz in querschnittlichen Mittelwertvergleichen nicht notwendig auch eine solche im Längsschnitt bedeutet (die Kohorteneffekthypothese zeigt einen möglichen Mechanismus längsschnittlichen Mittelwertsabfalls bei querschnittlicher Invarianz auf). Zweitens aber ist hier die sehr breite Definition der Höheraltrigengruppe zu bedenken: In den frühen Meßzeitpunkten enthält die Gruppe Alt84 „jüngere“ und „ältere Alte“ – wenn man nun annimmt, daß eine weitgehende durchschnittliche Invarianz von Lebenszufriedenheit nur in der frühen Phase der Höheraltrigkeit (z.B. bis ca. 70 Jahre, vgl. die deskriptiven Ergebnisse in Kapitel 3.2.1)

(vgl. Kapitel 3.1) neue Befragte hinzukommen können.

TABELLE 3.7:
Allgemeine Stabilitätsanalysen: Stichprobenmittelwerte der Meßwerte

Jahr	ALZ		EIZ		GEZ		WOZ		FRZ	
	Alt84	Alle	Alt84	Alle	Alt84	Alle	Alt84	Alle	Alt84	Alle
1984	7.42	7.45	6.85	6.50	5.46	6.91	8.42	7.95	8.20	7.43
1985	7.30	7.26	6.94	6.52	5.60	6.86	8.29	7.88	8.10	7.26
1986	7.45	7.33	7.05	6.60	5.65	6.84	8.31	7.84	8.14	7.13
1987	7.23	7.13	7.14	6.61	5.63	6.77	8.37	7.90	8.14	7.19
1988	7.03	7.04	7.03	6.63	5.45	6.70	8.23	7.80	8.08	7.03
1989	7.06	7.10	7.12	6.67	5.40	6.65	8.31	7.78	8.12	6.96
1990	7.13	7.31	7.23	6.74	5.41	6.68	8.25	7.73	7.89	6.98
1991	7.32	7.40	7.47	6.97	5.38	6.66	8.50	7.79	8.39	6.88
1992	7.10	7.30	7.08	6.82	5.33	6.79	8.41	7.73	8.41	6.98
1993	7.12	7.23	7.12	6.70	5.30	6.64	8.44	7.80	8.13	6.90
1994	6.98	7.12	7.15	6.54	5.16	6.57	8.57	7.79	8.20	6.88
1995	6.77	7.07	7.29	6.58	5.14	6.61	8.32	7.75	7.89	6.92
1996	6.97	7.12	7.21	6.64	5.19	6.59	8.32	7.81	8.36	7.07
1997	6.83	6.98	7.03	6.48	5.26	6.58	8.25	7.75	8.29	6.99
1998	6.80	7.05	7.21	6.50	5.15	6.61	8.10	7.73	8.17	7.01
1999	6.75	7.08	7.17	6.55	4.93	6.56	8.11	7.72	8.38	6.99

gegeben ist und danach ein Abfall stattfindet, so wäre hier vom ersten Erhebungsjahr an für die Gruppe Alt84 ein Rückgang zu erwarten, da ja bereits zu Beginn auch „ältere Alte“ in dieser Gruppe sind. Drittens aber relativiert der Blick auf Tabelle 3.8 diesen Abfall der ALZ-Mittelwerte: Gemessen an den Streuungen der Stichproben scheint er eher gering. So beträgt die Differenz zwischen dem höchsten (im Jahr 1986) und dem niedrigsten (im Jahr 1999) ALZ-Mittelwert nur 0.7, während die Standardabweichungen im Bereich von 2.10 bis 2.28 schwanken, d.h. der tatsächliche Rückgang der Mittelwerte beträgt ungefähr nur ein Drittel der Standardabweichung des Merkmals, was ihn weit weniger dramatisch erscheinen läßt, als der absolute Wert. Zu erwähnen wäre hier auch, daß natürlich, angesichts hoher Fallzahlen, die jeweils sukzessiven Mittelwertsdifferenzen bei statistischer Testung signifikant sind (mittels konventioneller t -Tests für abhängige Messungen und $\alpha \leq 0.05$), was aber im Grunde nichts zur inhaltlichen Einschätzung der Bedeutung der jeweiligen Effektstärken für die Diskussion „paradoxe“ Mittelwertsinvarianz beiträgt, sondern allenfalls bestätigt, daß ein gewisser Rückgang nicht bloß „zufällig“ beobachtet wurde, wie gering er auch sein mag.

Bei den anderen Zufriedenheitsmittelwerten auch der Gruppe Alt84 ist keine ähnlich deutliche Tendenz des Abfalls über die Zeit erkennbar. In sehr schwachem Maße zeigt sie sich allenfalls bei der Gesundheits- und Wohnzufriedenheit, während bei der Einkommenszufriedenheit der Älteren eher ein Anstieg sichtbar ist.

Betrachtet man die Standardabweichungen in Tabelle 3.8, so wird bei keiner der Zufriedenheiten eine Zunahme der Zufriedenheitsvarianz, wie sie gemäß der Gewinner-und Verlierer-Hypothese zu erwarten wäre, sichtbar. Tendenziell sind v.a. bis ungefähr 1990 eher Rückgänge zu sehen, und zwar gleichermaßen bei den Älteren, wie in der Gesamtgruppe. Theoretisch könnte allerdings eine Zunahme der Varianz der „wahren“ Zufriedenheitswerte mit einer Abnahme der Meßfehlervarianz einhergegangen sein, so daß in der Addition dieser beiden Varianzquellen zu den jeweiligen Meßzeitpunkten eine weitestgehende Invarianz der Meßwertvarianzen resultiert – diese Möglichkeit bleibt durch das Strukturgleichungsmodell zu überprüfen.

Die Abnahme der Zufriedenheitsstreuungen in der westdeutschen Gesamtstichprobe über die ersten SOEP-Wellen, ebenso wie das leichte Absinken der Mittelwerte allgemeiner Lebenszufriedenheit in den ersten Jahren, wurde bereits von Landua (1993) berichtet und analysiert – mit der Hauptschlußfolgerung, daß dieses v.a. eine Art von Lerneffekt bezüglich der Nutzung der Zufriedenheitsskala hin zu einem mehr „rationalen Antwortverhalten“ widerspiegeln und deshalb die meßtheoretische Qualität der Zufriedenheitsratings durch diese anfänglichen Wiederholungsmessungen verbessert würde. Letzterer Aspekt, d.h. die Veränderung der Reliabilität der SOEP-Zufriedenheitsmessungen, wurde explizit von Schräpler (1995) analysiert, der dazu bereits das QMSM zur Modellierung der Reliabilitäten und Entflechtung der Konfundierung von Reliabilität und Stabilität in den manifesten Meßwertautokorrelationen nutzte: Über die jeweils ersten sechs Wellen hinweg fand er einen Anstieg der Reliabilitäten bei fast allen Zufriedenheitsmessungen (außer Gesundheitszufriedenheit) der westdeutschen Teilstichprobe.

Tabelle 3.9 zeigt die Meßwertkorrelationen zwischen den aufeinanderfolgenden Zufriedenheitsmessungen, d.h. die Autokorrelationen erster Ordnung. Wie bereits dargestellt (Kap. 2.2.2 und Anhang A.1), handelt es sich dabei gewissermaßen um die noch nicht meßfehlerbereinigten Stabilitätskoeffizienten, die ein Ausmaß von Stabilität anzeigen, von dem angenommen werden muß, daß es bei nicht perfekter Reliabilität unter dem der „wahren“ Merkmalsstabilität liegt, so daß es eine Art von theoretisch möglicher Untergrenze für letztere darstellt (unter der Voraussetzung unkorrelierter Meßfehler). Zu erkennen ist hier wiederum bei allen Zufriedenheiten ein leichter tendenzieller Anstieg der Korrelationen während der ersten Erhebungen auf ein Niveau, welches dann annähernd beibehalten wird. Auch dieses könnte mit der von Landua (1993) und Schräpler (1995) geäußerten Vermutung eines anfänglichen Lerneffekts erklärt werden: Eine allmähliche Verbesserung der Reliabilität, d.h. eine Verminderung des Varianzanteils untereinander (zwischen den Erhebungswellen) unkorrelierter Zufallseinflüsse auf die Meßwerte, würde zu höheren Meßwertkorrelationen führen. Möglicherweise ist ein Anstieg der Meßwertkorrelationen aber auch wachsenden True-Score-Korrelationen, d.h. längsschnittlich ansteigender „wahrer“ Stabilität geschuldet. Interessant hinsichtlich der Fragestellungen dieser Untersuchung ist hierbei noch die Beobachtung, daß bei der allgemeinen Lebenszufriedenheit und der Einkommenszufriedenheit die Korrelationen der Gruppe Alt84 sichtbar über denen der Gesamtgruppe liegen,

TABELLE 3.8:
Allgemeine Stabilitätsanalysen: Standardabweichungen der Meßwerte

Jahr	ALZ		EIZ		GEZ		WOZ		FRZ	
	Alt84	Alle	Alt84	Alle	Alt84	Alle	Alt84	Alle	Alt84	Alle
1984	2.28	2.09	2.54	2.58	2.82	2.65	1.92	2.26	2.16	2.47
1985	2.24	2.01	2.38	2.44	2.63	2.44	1.92	2.15	2.12	2.37
1986	2.13	1.89	2.28	2.32	2.61	2.41	1.87	2.09	2.11	2.37
1987	2.15	1.93	2.24	2.27	2.64	2.37	1.88	2.08	2.11	2.34
1988	2.23	1.96	2.28	2.25	2.64	2.37	1.85	2.03	2.04	2.33
1989	2.15	1.94	2.20	2.20	2.62	2.41	1.77	2.03	1.95	2.35
1990	2.13	1.78	2.17	2.14	2.55	2.30	1.81	2.05	1.94	2.05
1991	2.05	1.70	2.05	2.06	2.57	2.31	1.57	1.99	1.87	2.41
1992	2.06	1.70	2.10	2.05	2.58	2.32	1.63	2.01	1.79	2.29
1993	2.13	1.78	2.22	2.17	2.58	2.30	1.67	1.98	2.10	2.31
1994	2.10	1.78	2.19	2.23	2.56	2.31	1.61	2.02	1.99	2.34
1995	2.22	1.80	2.15	2.19	2.62	2.26	1.72	1.98	1.98	2.05
1996	2.12	1.75	2.22	2.18	2.55	2.26	1.70	1.96	1.86	2.25
1997	2.15	1.79	2.17	2.16	2.53	2.21	1.75	1.92	1.99	2.27
1998	2.13	1.77	2.13	2.18	2.54	2.20	1.90	1.94	2.16	2.23
1999	2.23	1.77	2.22	2.14	2.57	2.24	1.98	1.94	2.11	2.23

was entweder auf höhere Reliabilitäten, oder aber auf höhere True-Score-Stabilitäten der Älteren hinweist.

Neben diesen allgemein beschreibenden vorbereitenden Analysen wurden auch solche durchgeführt, die eine eher technische Bedeutung für die QMSM-Berechnungen haben: Tabelle 3.10 zeigt die Durchschnittswerte jeweils aller Korrelationen n -ter Ordnung, d.h. die durchschnittliche Korrelation zwischen zwei aufeinanderfolgenden Messungen (lag1), die durchschnittliche Korrelation zwischen jeweils zwei Messungen im Abstand von zwei Jahren (lag2), von drei Jahren (lag3) usw. bis hin zur Korrelation zwischen erster und sechzehnter Messung (lag15). In diesen gemittelten Korrelationen n -ter Ordnung zeigt sich in allen Fällen die Simplex-Struktur der jeweiligen Korrelationsmatrix, welche in Modellen vom Typ des QMSM vorausgesetzt wird, nämlich eine stetige Abnahme mit größer werdendem Zeitabstand, welche zumindest bis zum lag13 immer eindeutig vorhanden ist. Die gelegentliche leichte Zunahme von lag13 zu lag14 bzw. lag14 zu lag15 sollte hier nicht allzu schwer ins Gewicht fallen, denn mit steigendem „time-lag“ sind es zusehends weniger Korrelationswerte, die hier in die Mittelwertbildung miteingehen, so daß letztere sozusagen einen mit dem Zeitabstand steigenden Standardfehler enthalten: Der lag14-Mittelwert wird lediglich noch aus zwei Korrelationswerten (Messung 1984 mit Messung 1998 und Messung 1985 mit Messung 1999) errechnet, während es sich bei lag15 nur noch um eine einzige Korrelation handelt, die hier als ihr eigener Mittelwert dient und damit natürlich

TABELLE 3.9:
Allgemeine Stabilitätsanalysen: Autokorrelationen erster Ordnung der Meßwerte

Spanne	ALZ		EIZ		GEZ		WOZ		FRZ	
	Alt84	Alle	Alt84	Alle	Alt84	Alle	Alt84	Alle	Alt84	Alle
84-85	0.49	0.46	0.55	0.52	0.52	0.58	0.48	0.49	0.38	0.41
85-86	0.53	0.48	0.60	0.54	0.57	0.59	0.48	0.49	0.48	0.46
86-87	0.49	0.47	0.65	0.57	0.58	0.58	0.52	0.52	0.54	0.52
87-88	0.59	0.51	0.61	0.56	0.60	0.59	0.52	0.50	0.52	0.52
88-89	0.59	0.53	0.59	0.57	0.58	0.60	0.57	0.54	0.49	0.54
89-90	0.54	0.54	0.65	0.57	0.60	0.60	0.58	0.56	0.45	0.53
90-91	0.60	0.55	0.66	0.58	0.59	0.61	0.52	0.57	0.47	0.55
91-92	0.57	0.55	0.65	0.61	0.63	0.63	0.55	0.57	0.43	0.59
92-93	0.60	0.56	0.66	0.59	0.62	0.62	0.50	0.59	0.46	0.56
93-94	0.59	0.57	0.61	0.60	0.64	0.61	0.53	0.59	0.56	0.56
94-95	0.53	0.58	0.65	0.60	0.66	0.63	0.47	0.56	0.46	0.52
95-96	0.59	0.59	0.70	0.62	0.66	0.64	0.53	0.56	0.49	0.56
96-97	0.65	0.60	0.67	0.64	0.64	0.64	0.59	0.58	0.61	0.59
97-98	0.65	0.58	0.57	0.61	0.62	0.60	0.61	0.55	0.53	0.57
98-99	0.62	0.58	0.55	0.59	0.61	0.61	0.50	0.54	0.48	0.56

ein größeres Maß an „Zufälligkeit“ beinhaltet, als z.B. der lag1-Mittelwert, der aus immerhin 15 Korrelationen gebildet wurde. Zudem bezieht sich die *Quasi*-Simplex-Struktur ja auf die Korrelationsmatrix der True-Scores, so daß die Simplex-Struktur für die beobachteten Werte nur näherungsweise gefordert ist.

Nach der Diskussion in Kapitel 2.3.1 bleibt als weiterer eher technischer Aspekt zur Durchführung der FIML-Schätzung von Strukturgleichungsmodellen die Beurteilung der Normalverteilungsannahme für die Daten bzw. zumindest die Suche nach Hinweisen auf grobe Abweichungen von derselben. Wie dort schon dargelegt (vgl. Kapitel 2.3.1.4), wird hier lediglich anhand der univariaten Verteilungskennwerte nach solchen Hinweisen gesucht: Für die fünf Zufriedenheitstypen wurden jeweils Schiefe, Kurtosis sowie ein statistischer Test der Nullhypothese univariater Normalverteilung für jedes Erhebungsjahr und beide Untersuchungsgruppen berechnet. Die statistischen Tests wurden mit der SAS-Prozedur UNIVARIATE (SAS Institute Inc., 1990, 617-635) durchgeführt, welche durch Voreinstellung bei Stichprobengrößen größer 2000 den Kolmogoroff-Smirnoff-Test (allgemeine Beschreibung siehe z.B. bei Bosch, 1994, 117f; Hays, 1994, 853-856) und bei kleineren Stichproben der Shapro-Wilk-Test (Shapiro & Wilk, 1965) anwendet.

Die Ergebnisse sind in den Tabellen B.9 bis B.13 im Anhang B wiedergegeben. Bei allen Normalverteilungstests zeigt sich eine signifikante Teststatistik, so daß jeweils die Nullhypothese univariater Normalverteilung zurückzuweisen wäre. Jedoch erscheint hier eher das Ausmaß der Abweichung von der Verteilungsnormalität relevant insofern, als bei

TABELLE 3.10:
Allgemeine Stabilitätsanalysen: Durchschnittliche Autokorrelationen n -ter Ordnung (lag n)
der Meßwerte

lag	ALZ		EIZ		GEZ		WOZ		FRZ	
	Alt84	Alle	Alt84	Alle	Alt84	Alle	Alt84	Alle	Alt84	Alle
1	0.58	0.54	0.63	0.58	0.61	0.61	0.53	0.55	0.49	0.54
2	0.52	0.49	0.60	0.53	0.56	0.57	0.49	0.47	0.42	0.49
3	0.49	0.46	0.56	0.49	0.52	0.54	0.46	0.42	0.38	0.45
4	0.47	0.43	0.53	0.46	0.50	0.52	0.43	0.39	0.35	0.42
5	0.45	0.41	0.51	0.44	0.48	0.50	0.39	0.36	0.33	0.40
6	0.42	0.39	0.49	0.42	0.46	0.48	0.39	0.35	0.30	0.38
7	0.42	0.37	0.47	0.39	0.45	0.46	0.37	0.34	0.27	0.36
8	0.39	0.36	0.45	0.38	0.43	0.45	0.38	0.32	0.25	0.34
9	0.38	0.34	0.44	0.37	0.42	0.43	0.34	0.31	0.24	0.32
10	0.36	0.32	0.43	0.35	0.41	0.42	0.31	0.28	0.22	0.30
11	0.36	0.30	0.41	0.33	0.40	0.40	0.32	0.28	0.22	0.29
12	0.36	0.29	0.38	0.32	0.39	0.38	0.29	0.26	0.18	0.27
13	0.36	0.28	0.36	0.31	0.35	0.36	0.26	0.23	0.14	0.25
14	0.32	0.27	0.31	0.30	0.36	0.36	0.27	0.24	0.13	0.22
15	0.41	0.27	0.34	0.32	0.37	0.37	0.23	0.25	0.09	0.20

mäßigen Werten in Schiefe und Exzeß noch von einer Nichtnormalität der Daten ausgegangen werden kann, dergegenüber die FIML-Parameterschätzungen robust sind (vgl. Kapitel 2.3.1.4). Zudem sind hier die sehr niedrigen Irrtumswahrscheinlichkeiten bei Zurückweisung der Nullhypothese v.a. auch den sehr hohen Fallzahlen geschuldet – ähnlich wie beim χ^2 -Model-Fit-Test bei Strukturgleichungsmodellen ist die daraus resultierende hohe Teststärke unter dem Gesichtspunkt praktischer Bedeutsamkeit eher von Nachteil.³¹ Dieses gilt besonders für den Shapiro-Wilk-Test, dessen Vorteile eigentlich in seiner Teststärke

³¹Auch rät hier Bollen (1989, 423) wegen der implizierten multiplen Testung zur Anwendung der Bonferroni-Korrektur. Dem möchte ich allerdings nicht folgen: Da bei allen diesen Tests die Nullhypothese die eigentlich kritisch zu testende ist und letztlich die vielen univariaten Tests zur behelfswisen Überprüfung einer multivariaten Normalverteilung dienen, erscheint die Korrektur nicht sinnvoll. Denn diese wäre dann angebracht, wenn die Bestätigung der „eigentlichen“ multivariaten Hypothese aus der simultanen Gültigkeit *aller* univariaten Alternativhypothesen erfolgt, so daß die Wahrscheinlichkeit dafür, bei 16 Tests mindestens einmal irrtümlich für die H_1 entschieden zu haben, kontrolliert werden muß. Im vorliegenden Fall aber ist aus der Hypothese multivariater Normalverteilung abzuleiten, daß alle univariaten Verteilungen normal sind, d.h. die „eigentliche“ multivariate Hypothese ist bereits dann zurückzuweisen, wenn nur in einem Falle die univariate Alternativhypothese richtig ist (unabhängig von den anderen Tests), so daß, wenn überhaupt, die jeweilige α -Irrtumswahrscheinlichkeit des einzelnen Tests kontrolliert werden muß – vgl. dazu Rule (1976). Immerhin aber betrüge bei Bonferroni-Korrektur im Sinne Bollens und „multiplem“ Signifikanzniveau von $\alpha = 0.05$ das Signifikanzlevel der einzelnen Tests $\alpha_B = 0.05/16 = 0.003$, was zumindest bei allen Kolmogoroff-Smirnoff-Tests für die Verteilungen in der Gesamtgruppe nicht erreicht würde.

bei viel geringeren Stichprobengrößen, als diejenigen, welche hier zur Anwendung kamen, liegt. Die sehr niedrigen (bei Beschränkung auf drei Nachkommastellen nicht mehr darstellbaren) Irrtumswahrscheinlichkeiten in den Tabellen B.9 bis B.13 zeigen dieses sehr deutlich, v.a. auch angesichts der dazugehörigen Teststatistiken W , welche fast immer über 0.9 liegen, was auf eine eher mäßige Effektstärke schließen läßt.

Die Betrachtung der tatsächlichen Abweichungen von der Verteilungsnormalität, d.h. der Schiefe- und Kurtosiswerte, zeigt dagegen keinen einzigen Fall, der in dieser Hinsicht als besonders dramatisch erschiene. Dieses kann selbst dann noch gelten, wenn man hier strengere Kriterien anlegt, als die wie bereits erwähnt von West et al. (1995) vorgeschlagenen (Betrag des Schiefewert größer 2 und des Kurtosiswerte größer 7). Er scheint deshalb zu rechtfertigen, die in den Daten erkennbaren Abweichungen von der Voraussetzung multivariater Normalverteilung hinzunehmen und auf die Güte der FIML-Schätzungen zu vertrauen.

3.3.1.2 Analysestrategie der allgemeinen Stabilitätsanalysen

Es wurden für jede der Zeitreihen von jeweils 16 Messungen allgemeiner Lebenszufriedenheit, Einkommenszufriedenheit, Gesundheitszufriedenheit, Wohnzufriedenheit und Freizeitzufriedenheit unterschiedlich restringierte Versionen des QMSM berechnet. Eine graphische Darstellung des Analysemodells wurde bereits durch Abbildung 2.3B (Kapitel 2.2.2.1) gegeben, im vorliegenden Fall bezeichnet X_t also die jeweilige t -te Zufriedenheitsmessung der Zeitreihe ($t = 1, \dots, 16$) und T_t den dazugehörigen True-Score der Zufriedenheit. Zur Restriktion homogener True-Score-Varianzen und Stabilitätskoeffizienten wurde die in Abbildung 2.4 wiedergegebene Modellierung mittels Phantomvariablen angewandt.

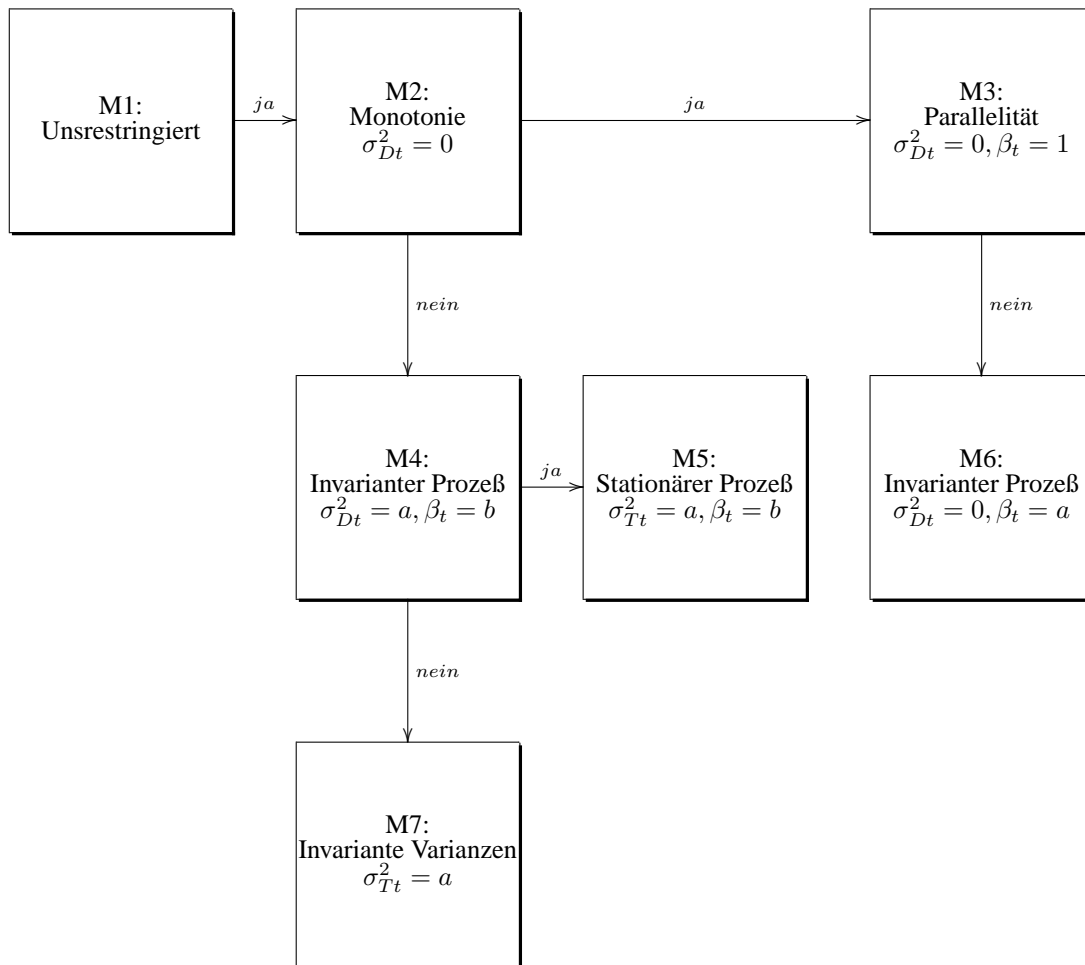
Wie eingangs des Kapitels 3.3 bereits dargelegt, besteht die Analysestrategie bei den hier angestrebten allgemeinen Stabilitätsanalysen mit dem QMSM in der schrittweisen Berechnung unterschiedlich restringierter Modellversionen. Letztendlich geht es hier darum, Annahmen über längsschnittliche Invarianzen von True-Score-Varianzen und Stabilitätskoeffizienten zu überprüfen, welche modelltechnisch als Restriktionen der Kovarianzstruktur des QMSM spezifiziert werden können. Eine Restriktion hält der Überprüfung dann stand, wenn sie keine bedeutsame Verschlechterung der Modellanpassung bewirkt, d.h. wenn das unter dieser Restriktion berechnete Modell keinen schlechteren Fit aufweist, als das ohne diese Restriktion berechnete – das Kriterium dafür, was als schlechterer Fit gelten soll, wurde in Kapitel 2.3.2 festgelegt. Dieses führt zu einem schrittweisen Vorgehen der Spezifizierung und sukzessiven Berechnung immer stärker restringierter Modelle: Es wurden sieben verschiedene Modellversionen spezifiziert und nach der logisch sich ergebenden Abfolge überprüft, die in Abbildung 3.18 dargestellt ist. Es handelt sich dabei um die folgenden Versionen des QMSM:

- **M1:** Das bezüglich der True-Score-Varianzen und Stabilitätskoeffizienten unrestrin-

gierte QMSM. Gemäß der Darlegungen in Kapitel 2.2.2.1 sind bei der Berechnung dieses Modells die Stabilitätskoeffizienten und True-Score-Varianzen an den „Rändern“ der Zeitreihe (d.h. jeweils die ersten und letzten Koeffizienten) nicht identifiziert, so daß hier nur die „inneren“ Stabilitätskoeffizienten ρ_2 bis ρ_{14} und True-Score-Varianzen $\sigma_{T_2}^2$ bis $\sigma_{T_{15}}^2$ geschätzt werden können.

- ▶ **M2:** Restriktion aller True-Score-Residualvarianzen $\sigma_{Dt}^2 = 0$ – dieses bedeutet die Restriktion auf ein Modell *absolut* monotoner Stabilität, da die True-Scores zu allen Zeitpunkten $t > 1$ keinerlei modellexogene Varianzquelle mehr besitzen, sondern ganz und gar durch den jeweils vorherigen True-Score bzw. damit letztendlich alle durch den True-Score zum ersten Zeitpunkt bestimmt sind, so daß, bei möglicher Veränderung der Varianzen (vgl. auch die Gleichungen 2.26a und 2.26b, Kapitel 2.2.2.1), die Autokorrelationen gleich Eins sind.
- ▶ **M3:** Restriktion aller True-Score-Residualvarianzen $\sigma_{Dt}^2 = 0$ und der unstandardisierten Autoregressionskoeffizienten $\beta_t = 1$ – diese letztere Festlegung zusätzlich zu der bereits in M2 gesetzten Monotonierestriktion bewirkt ein Modell *absolut* paralleler Stabilität, da damit nicht nur alle Stabilitätskoeffizienten auf den Wert 1 fixiert sind, sondern auch die Gleichheit sämtlicher True-Score-Varianzen festgelegt ist (vgl. wiederum Gleichung 2.26b).
- ▶ **M4:** Restriktion invarianter True-Score-Residualvarianzen $\sigma_{Dt}^2 = a$ und unstandardisierter Autoregressionskoeffizienten $\beta_t = b$. Hierdurch ist also keine Bedingung perfekter Merkmalsstabilität spezifiziert, wie in den Modellversionen M2 und M3. Stattdessen bedeuten diese Restriktionen eine Invarianz des Veränderungsprozesses, d.h. der Regressionsgleichung, welche die Abhängigkeit der True-Scores T_t von den vorherigen True-Scores T_{t-1} beschreibt. Dieses bedeutet aber *nicht* die Invarianz der True-Score-Varianzen oder Stabilitätskoeffizienten, sondern lediglich, daß das absolute Ausmaß der modellexogenen Veränderungsvarianz ebenso unabhängig vom Meßzeitpunkt ist, wie auch die absolute, unstandardisierte Abhängigkeit der True-Score-Werte über ein Ein-Jahres-Befragungsintervall hinweg.
- ▶ **M5:** Restriktion invarianter True-Score-Varianzen $\sigma_{T_t}^2 = a$ und unstandardisierter Autoregressionskoeffizienten $\beta_t = b$. Dieses bedeutet sozusagen ein maximal restriktives Veränderungsmodell bei nicht perfekter Stabilität: Der Veränderungsprozeß ist wie bei M4 zu jedem Zeitpunkt derselbe und er verändert auch nicht die Varianz des Merkmals. Dieses hat, wie bereits in Kapitel 2.2.2.1 dargestellt, auch die Invarianz der Stabilitätskoeffizienten ρ_t zur Konsequenz. Damit wäre also ein Veränderungsprozeß beschrieben, der völlig unabhängig vom Zeitpunkt ist und durch einen einzigen Stabilitätskoeffizienten und eine einzige True-Score-Varianz charakterisiert werden kann. Der solchermaßen modellierte Prozeß erfüllt das Merkmal der sogenannten Stationarität zweiter Ordnung (vgl. Verbeke & Molenberghs, 2000, 142; Hershberger et al., 1996, 163).
- ▶ **M6:** Restriktion aller True-Score-Residualvarianzen $\sigma_{Dt}^2 = 0$ und der Invarianz der unstandardisierten Autoregressionskoeffizienten $\beta_t = b$. Letzteres legt die längsschnittli-

ABBILDUNG 3.18:
Analysestrategie der allgemeinen Stabilitätsanalysen



che Invarianz der Veränderungsprozesse fest, welche infolge der ersten Restriktion (vgl. M2) durch perfekte monotone Stabilität gekennzeichnet sind. Es ist insofern als mögliche Ergänzung zu M2 zu verstehen.

- **M7:** Restriktion invarianter True-Score-Varianzen $\sigma_{Tt}^2 = a$ – dieses bedeutet lediglich, daß bei aller möglicherweise stattfindenden Veränderung die Varianz des Merkmals über die Zeit gleich bleibt, ansonsten sind längsschnittliche „Veränderungen der Veränderung“ (z.B. der Stabilitätskoeffizienten) möglich. M7 bezeichnet also die nach M1 am wenigsten restringierte Modellversion.

Abbildung 3.18 stellt einen Entscheidungsablauf zur Überprüfung dieser Modellversionen dar: Zunächst einmal sollte überhaupt für die unrestringierte Version M1 ein guter Model-Fit erreicht werden, andernfalls ist die Brauchbarkeit des QMSM zur Modellierung

der längsschnittlichen Entwicklung der Zufriedenheitswerte grundsätzlich in Frage gestellt. Bei guter Modellanpassung durch M1 kann diese dann zunächst mit der von M2, dem Modell perfekt monoton stabiler Veränderung, verglichen werden: Ist letztere nicht in relevantem Maße schlechter, als erstere, so scheint also die Zeitreihe der Zufriedenheitsmessungen durch ein Modell monoton stabiler True-Score-Veränderungen ebenso gut erklärbar, wie durch ein nicht weiter festgelegtes autoregressives True-Score-Modell, und es wäre dann zu überprüfen, ob sogar M3, d.h. ein Modell parallel stabiler True-Score-Veränderung, „paßt“: M3 sollte dann im Vergleich zu M2 (und M1!) nicht zu einer deutlichen Verschlechterung der Modellanpassung führen. Ist M3 aufgrund einer solchen Verschlechterung abzulehnen, so kann noch M6, die restriktivere Variante eines Modells monotoner Stabilität, überprüft werden. Führt dagegen bereits M2 zu einem im Vergleich zu M1 deutlichen Abfall der Modellanpassung, so erübrigen sich alle Modelle perfekter Merkmalsstabilität und es erscheint in diesem Falle sinnvoll, die Überprüfung der Modelle nicht perfekter Stabilität zunächst einmal mit M4 zu beginnen, d.h. bei der grundsätzlichen Frage, ob die Zeitreihe der Zufriedenheitsmessungen gut beschrieben werden kann durch ein Modell eines autoregressiven Veränderungsprozesses, welcher sich selbst über die Zeit hinweg nicht verändert. Wenn M4 angenommen werden kann, dann wäre die restriktivere Variante M5 zu überprüfen, d.h. die Möglichkeit invarianter Varianzen bei einem invarianten Veränderungsprozeß. Wird M4 abgelehnt, so bleibt noch M7, eine Modellvariante, die keine expliziten Restriktionen bezüglich der Stabilität des Merkmals enthält, sondern nur die Invarianz der True-Score-Varianzen im Zeitverlauf festlegt.

Zusätzlich zu diesen Modellvergleichen, welche Restriktionen der Kovarianzstruktur des Modells beinhalten, wurde noch ein weiterer, davon unabhängiger Vergleich des unrestringierten Modells M1 mit einer lediglich in der Mittelwertsstruktur restringierten Modellversion durchgeführt: Bei letzterem Modell, das mit M1b bezeichnet wird, wurde die Gleichheit der Intercepts $\mu_t = \mu$ (vgl. Gleichung 2.30, Kapitel 2.2.2.1) festgelegt, d.h. es handelt sich um ein Modell invarianter Mittelwerte über die untersuchte Zeitreihe. Wie bereits dargelegt (vgl. vorne S.106f), sind bei dieser Modellierung des QMSM Restriktionen der Mittelwerts- und der Kovarianzstruktur jeweils unabhängig voneinander zu überprüfen, die Modellversionen M2 bis M7 sind in der Mittelwertsstruktur unrestringiert und der Vergleich M1 mit M1b dient der zusätzlichen Überprüfung der Gleichheitsrestriktion für die Mittelwerte, welche, wie in Kapitel 2.2.1 dargelegt, die Stabilitätsanalyse vervollständigt, indem damit die Frage beantwortet wird, ob eine möglicherweise ausgeprägte (näherungsweise) parallele Stabilität auch als (näherungsweise) strikte bezeichnet werden kann: Annahme von M1b würde bedeuten, daß die Zeitreihe der Zufriedenheitsdaten durch ein „mittelwertsstabiles“ Modell genauso gut beschrieben werden kann, wie durch ein Modell freier variabler Mittelwerte.

3.3.1.3 Ergebnisse der allgemeinen Stabilitätsanalysen

Die Tabellen 3.11 bis 3.15 zeigen die Model-Fit-Kennwerte der Modelle M1 bis M7 bei Analyse der allgemeinen Lebenszufriedenheit, sowie der Zufriedenheiten mit Einkommen, Gesundheit, Wohnen, Freizeit für die beiden Untersuchungsgruppen. Es ist bereits auf den ersten Blick zu erkennen, daß für alle Zufriedenheiten bei der unrestringierten Modellversion M1 eine Anpassungsgüte erreicht wurde, die nach den in Kapitel 2.3.2 festgelegten Kriterien als gut bezeichnet werden kann: Die inkrementellen Fit-Indizes liegen sämtlichst über 0.95, die RMSEA-Werte alle deutlich unter 0.05. Dagegen waren sämtliche χ^2 -Model-Fit-Tests – nicht nur bei den Modellversionen M1, sondern bei allen berechneten Modellen – signifikant mit $\alpha \leq 0.001$, was m.E. eher auf eindrucksvolle Weise die im obengenannten Kapitel diskutierte Problematik dieser Tests bei großen Stichprobenumfängen illustriert, als einen praktisch relevanten Model-Misfit, zumal ja, wie bereits gezeigt, in den Daten moderate Abweichungen von der zugrundegelegten Normalverteilungsannahme vorliegen, welche zu einem negativen Bias der α -Irrtumswahrscheinlichkeiten führen können. Es kann also zunächst einmal festgestellt werden, daß das unrestringierte QMSM für alle untersuchten Zufriedenheitsdaten „paßt“, ein Markov-Modell latenter True-Scores kann als adäquate Modellierung der Zeitreihen von Zufriedenheitsmessungen betrachtet werden.

Bei Vergleich der Modellversionen ist wiederum in allen Tabellen und jeweils für beide Untersuchungsgruppen ein recht eindeutiges Muster zu erkennen: Während die Fit-Indizes der Modelle M4, M5 und M7 geringfügige bis gar keine Unterschiede untereinander und zu denen von M1 aufweisen, „fitten“ die Modelle M2, M3 und M6 in allen Indizes erkennbar schlechter, als die der ersten Gruppe und insbesondere M1.

Bei den Modellen für die Messungen der allgemeinen Lebenszufriedenheit (Tabelle 3.11) zeigt M2 im Vergleich zu M1 einen Abfall der Werte in den inkrementellen Fit-Indizes TLI, NFI, IFI und CFI, welcher auf den ersten Blick nicht dramatisch erscheinen mag, da er sozusagen nur ab der zweiten Nachkommastelle stattfindet. Daß sich darin aber ein durchaus substantieller Anpassungsverlust widerspiegelt, zeigt der Vergleich der RMSEA-Werte, die in beiden Untersuchungsgruppen für M1 jeweils deutlich unter dem Cutoff-Wert von 0.05 und bei M2 darüber liegen: Nach den Daumenregeln von Browne & Cudeck (1993) würde M2 in beiden Fällen allenfalls noch ein „akzeptabler“ Fit attestiert. Gemäß der in Kapitel 2.3.2 festgelegten praktischen Regeln zum Modellvergleich wäre M2 also eindeutig zurückzuweisen, da in allen Fit-Indizes eine Differenz gegenüber M1 größer 0.01 auftritt und sich auch die RMSEA-Konfidenzintervalle der beiden Modellversion nicht überschneiden. Nach dem in Abbildung 3.18 dargestellten Entscheidungsablauf erübrigt sich damit auch der Vergleich von M1 mit den noch stärker restringierten Modellen M3 und M6, der auch tatsächlich zu demselben Ergebnis führen würde, wie der mit M2.

Nach Zurückweisung von M2 wäre nun gemäß Abbildung 3.18 M1 mit M4 zu vergleichen: Für beide Untersuchungsgruppen zeigt Tabelle 3.11 allenfalls winzige Verluste in den Fit-Index-Werten. Bei der in beiden Gruppen unverändert hohen TLI- und in der

TABELLE 3.11:
Allgemeine Stabilitätsanalysen: Anpassungsgüte der QMSM für Lebenszufriedenheit

	M1	M2	M3	M4	M5	M6	M7	M1b
<i>df</i>	91	104	119	117	118	118	104	106
Alle								
χ^2	343.9	5800.8	6095.6	530.7	533.2	5974.0	464.5	1376.6
<i>TLI</i>	0.999	0.979	0.981	0.999	0.999	0.981	0.999	0.995
<i>NFI</i>	0.999	0.984	0.983	0.999	0.999	0.983	0.999	0.996
<i>IFI</i>	0.999	0.984	0.983	0.999	0.999	0.984	0.999	0.996
<i>CFI</i>	0.999	0.984	0.983	0.999	0.999	0.984	0.999	0.996
<i>RMSEA</i>	0.014	0.063	0.060	0.016	0.016	0.060	0.016	0.029
<i>RMSEA_U</i>	0.013	0.061	0.059	0.015	0.014	0.058	0.014	0.028
<i>RMSEA_O</i>	0.016	0.064	0.061	0.017	0.017	0.061	0.017	0.031
Alt84								
χ^2	173.3	850.2	919.7	234.8	254.2	877.5	201.0	459.5
<i>TLI</i>	0.997	0.979	0.980	0.997	0.997	0.981	0.997	0.990
<i>NFI</i>	0.996	0.982	0.980	0.995	0.994	0.981	0.996	0.990
<i>IFI</i>	0.998	0.984	0.983	0.997	0.997	0.983	0.998	0.992
<i>CFI</i>	0.998	0.984	0.983	0.997	0.997	0.983	0.998	0.992
<i>RMSEA</i>	0.021	0.059	0.057	0.022	0.024	0.056	0.021	0.040
<i>RMSEA_U</i>	0.016	0.055	0.053	0.018	0.020	0.052	0.017	0.036
<i>RMSEA_O</i>	0.025	0.062	0.060	0.026	0.027	0.059	0.025	0.044

df: Freiheitsgrade des Modells.

RMSEA_U/RMSEA_O: Untere/obere Grenze des *RMSEA*-90%-Konfidenzintervalls

Alle χ^2 -Werte sind signifikant mit $\alpha \leq 0.001$.

Gesamtgruppe auch gleichbleibenden NFI-, IFI- und CFI-Werten, sinken letztere in der Gruppe Alt84 jeweils nur um 0.001. Auch der Anstieg des RMSEA-Wertes um 0.002 (Gesamtgruppe) und 0.001 (Alt84) scheint jeweils kaum der Rede wert, die Konfidenzintervalle überlappen sich jeweils. Somit wäre also M4 nicht zurückzuweisen und im nächsten Entscheidungsschritt M5 im Vergleich zu M1 zu bewerten: In der Gesamtgruppe unterscheiden sich M4 und M5 in keinem einzigen der Fit-Index-Werte, so daß hier M5 genauso wenig zurückzuweisen ist, wie M4. In der Gruppe Alt84 zeigt der Vergleich von M1 und M5 einen NFI-Verlust von 0.001, IFI- und CFI-Verluste von 0.001, sowie einen RMSEA-Anstieg um 0.003 – allesamt geringfügige Differenzen, die dem festgelegten Kriterien für die Zurückweisung der restriktiveren Modellversion nicht genügen, zumal sich wiederum die RMSEA-Konfidenzintervalle deutlich überschneiden. Da somit M5 angenommen werden kann, erübrigt sich im Grunde der Vergleich von M1 mit M7, welches natürlich, da es sich um eine weniger restringierte Variante handelt als M5, ebenfalls im Vergleich mit M1 nicht zurückgewiesen werden würde.

TABELLE 3.12:
Allgemeine Stabilitätsanalysen: Anpassungsgüte der QMSM für Einkommenszufriedenheit

	M1	M2	M3	M4	M5	M6	M7	M1b
<i>df</i>	91	104	119	117	118	118	104	106
Alle								
χ^2	333.4	7817.7	7960.4	538.1	553.2	7929.9	461.0	866.5
<i>TLI</i>	0.999	0.967	0.971	0.998	0.998	0.971	0.998	0.997
<i>NFI</i>	0.999	0.975	0.974	0.998	0.998	0.974	0.999	0.997
<i>IFI</i>	0.999	0.975	0.974	0.999	0.999	0.975	0.999	0.998
<i>CFI</i>	0.999	0.975	0.974	0.999	0.999	0.975	0.999	0.998
<i>RMSEA</i>	0.014	0.073	0.069	0.016	0.016	0.069	0.016	0.023
<i>RMSEA_U</i>	0.012	0.071	0.067	0.015	0.015	0.067	0.014	0.021
<i>RMSEA_O</i>	0.015	0.074	0.070	0.017	0.018	0.070	0.017	0.024
Alt84								
χ^2	145.8	1026.1	1050.1	173.7	175.1	1044.6	166.5	251.0
<i>TLI</i>	0.998	0.974	0.977	0.999	0.999	0.977	0.998	0.996
<i>NFI</i>	0.997	0.978	0.977	0.996	0.996	0.977	0.996	0.995
<i>IFI</i>	0.999	0.980	0.980	0.999	0.999	0.980	0.999	0.997
<i>CFI</i>	0.999	0.980	0.980	0.999	0.999	0.980	0.999	0.997
<i>RMSEA</i>	0.017	0.065	0.061	0.015	0.015	0.061	0.017	0.026
<i>RMSEA_U</i>	0.012	0.062	0.058	0.010	0.010	0.058	0.012	0.022
<i>RMSEA_O</i>	0.022	0.069	0.065	0.020	0.020	0.065	0.022	0.030

df: Freiheitsgrade des Modells.

RMSEA_U/RMSEA_O: Untere/obere Grenze des *RMSEA*-90%-Konfidenzintervalls

Alle χ^2 -Werte sind signifikant mit $\alpha \leq 0.001$.

Die für die Modelle der allgemeinen Lebenszufriedenheit somit gefundene Entscheidung für M5 ergibt sich in derselben Weise auch bei den jeweiligen Modellvergleichen für die Bereichszufriedenheiten mit dem Einkommen, der Gesundheit und dem Wohnen, wie anhand der Tabellen 3.12 bis 3.14 leicht nachzuvollziehen ist: Stets würde die in Abbildung 3.18 dargestellte Entscheidungsstrategie und die in Kapitel 2.3.2 dafür festgelegten Regeln ganz eindeutig zur Auswahl der Modellversion M5 führen. Allenfalls bei den Einkommenszufriedenheitsmodellen für die Gesamtgruppe (Tabelle 3.12) zeigt sich für eines der Kriterien zur Auswahl des sparsameren Modells eine „Beinahe-Verfehlung“, nämlich in der Überschneidung der *RMSEA*-Konfidenzintervalle: Hier entspricht der obere Wert für M1 dem unteren für M5, was aber, bei einer nur geringen *RMSEA*-Differenz von 0.002, noch als Erfüllung des Kriteriums gewertet werden könnte.

Beim Vergleich der Freizeitzufriedenheitsmodelle (Tabelle 3.15) zeigt sich im Prinzip dasselbe Muster, wie bei den anderen Zufriedenheiten: M2, M3 und M6 „fitten“ deutlich

TABELLE 3.13:

Allgemeine Stabilitätsanalysen: Anpassungsgüte der QMSM für Gesundheitszufriedenheit

	M1	M2	M3	M4	M5	M6	M7	M1b
<i>df</i>	91	104	119	117	118	118	104	106
Alle								
χ^2	276.6	5652.0	5819.6	437.3	442.3	5789.7	321.5	1091.9
<i>TLI</i>	0.999	0.977	0.979	0.999	0.999	0.979	0.999	0.996
<i>NFI</i>	0.999	0.982	0.981	0.999	0.999	0.981	0.999	0.997
<i>IFI</i>	0.999	0.982	0.982	0.999	0.999	0.982	0.999	0.997
<i>CFI</i>	0.999	0.982	0.982	0.999	0.999	0.982	0.999	0.997
<i>RMSEA</i>	0.012	0.062	0.058	0.014	0.014	0.059	0.012	0.026
<i>RMSEA_U</i>	0.010	0.060	0.057	0.013	0.013	0.057	0.011	0.024
<i>RMSEA_O</i>	0.014	0.063	0.060	0.015	0.015	0.060	0.014	0.027
Alt84								
χ^2	153.9	841.4	904.8	191.3	204.0	862.3	170.6	380.0
<i>TLI</i>	0.997	0.972	0.974	0.997	0.997	0.975	0.997	0.990
<i>NFI</i>	0.996	0.975	0.974	0.994	0.994	0.975	0.995	0.989
<i>IFI</i>	0.998	0.978	0.977	0.998	0.997	0.978	0.998	0.992
<i>CFI</i>	0.998	0.978	0.977	0.998	0.997	0.978	0.998	0.992
<i>RMSEA</i>	0.018	0.058	0.056	0.017	0.019	0.055	0.018	0.035
<i>RMSEA_U</i>	0.013	0.055	0.053	0.013	0.014	0.052	0.013	0.031
<i>RMSEA_O</i>	0.023	0.062	0.060	0.022	0.023	0.058	0.022	0.039

df: Freiheitsgrade des Modells.*RMSEA_U*/*RMSEA_O*: Untere/obere Grenze des *RMSEA*-90%-KonfidenzintervallsAlle χ^2 -Werte sind signifikant mit $\alpha \leq 0.001$.

schlechter, als die anderen Modellversionen, und die in den Fit-Indizes erkennbare Verschlechterung der Anpassungsgüte von M4, M5 und M7 gegenüber M1 scheint nur gering. Jedoch tritt letztere deutlicher zutage, als bei den vorherigen Vergleichen. Dieses wird schon an den χ^2 -Werten selbst deutlich: Bei diesen, die ja die Diskrepanz der modellimplizierten an die empirische Kovarianzmatrix (und hier auch zusätzlich des modellimplizierten an den empirischen Mittelwertsvektor) absolut und ohne Adjustierung an die Sparsamkeit des Modells ausdrücken, ist auf jeden Fall ein Anstieg bei zusätzlichen Restriktionen zu erwarten, und ein solcher zeigt sich auch jeweils in den Tabellen 3.11 bis 3.14 bei Vergleich der M1- χ^2 -Werte mit allen anderen. Jedoch ist dieser Anstieg in der Tabelle 3.15 doch deutlich größer, als in den vorherigen Fällen. Bei den Werten der Fit-Indizes, die sozusagen eine Belohnung für die Sparsamkeit von Modellen beinhalten, zeigt sich dieser Unterschied weniger deutlich – nämlich bei den inkrementellen Indizes im Vergleich von M1 mit M4, M5 und M7 wiederum beinahe gar nicht. Doch immerhin ist der Anstieg der *RMSEA*-Werte zumindest „tendenziell“ etwas größer, als er zuvor bei den Modellen der anderen

TABELLE 3.14:
Allgemeine Stabilitätsanalysen: Anpassungsgüte der QMSM für Wohnzufriedenheit

	M1	M2	M3	M4	M5	M6	M7	M1b
<i>df</i>	91	104	119	117	118	118	104	106
Alle								
χ^2	448.5	7938.0	8292.7	569.5	571.9	8286.8	480.4	573.6
<i>TLI</i>	0.998	0.971	0.974	0.999	0.999	0.973	0.999	0.998
<i>NFI</i>	0.999	0.978	0.977	0.998	0.998	0.977	0.999	0.998
<i>IFI</i>	0.999	0.978	0.977	0.999	0.999	0.977	0.999	0.999
<i>CFI</i>	0.999	0.978	0.977	0.999	0.999	0.977	0.999	0.999
<i>RMSEA</i>	0.017	0.073	0.070	0.017	0.017	0.070	0.016	0.018
<i>RMSEA_U</i>	0.015	0.072	0.069	0.015	0.015	0.069	0.015	0.016
<i>RMSEA_O</i>	0.018	0.075	0.071	0.018	0.018	0.072	0.018	0.019
Alt84								
χ^2	172.7	786.8	847.8	254.4	254.4	847.1	217.8	263.7
<i>TLI</i>	0.998	0.984	0.985	0.997	0.997	0.985	0.997	0.996
<i>NFI</i>	0.997	0.986	0.985	0.995	0.995	0.985	0.996	0.995
<i>IFI</i>	0.999	0.988	0.987	0.998	0.998	0.987	0.998	0.997
<i>CFI</i>	0.999	0.988	0.987	0.998	0.998	0.987	0.998	0.997
<i>RMSEA</i>	0.021	0.056	0.054	0.024	0.024	0.054	0.023	0.027
<i>RMSEA_U</i>	0.016	0.052	0.051	0.020	0.020	0.051	0.019	0.023
<i>RMSEA_O</i>	0.025	0.060	0.058	0.028	0.027	0.058	0.027	0.031

df: Freiheitsgrade des Modells.

RMSEA_U/RMSEA_O: Untere/obere Grenze des *RMSEA*-90%-Konfidenzintervalls

Alle χ^2 -Werte sind signifikant mit $\alpha \leq 0.001$.

Zufriedenheiten war: Er beträgt hier immerhin 0.006 (Vergleich M1-M4, M1-M5, M1-M7) für die Gesamtgruppe und 0.004 (Vergleich M1-M4, M1-M5) bzw. 0.003 (Vergleich M1-M7) für die Gruppe Alt84. Dieses erfüllt zwar immer noch das festgelegte Kriterium, nur einen Anstieg ≥ 0.01 als relevante Fit-Verschlechterung zu betrachten, zudem ist aber in der Gesamtgruppe das Kriterium sich überschneidender *RMSEA*-Konfidenzintervalle bei keinem der Vergleiche mit M1 erfüllt. Somit könnte hier aus den Ergebnissen der Modellrechnungen für die Freizeitzufriedenheitswerte zwar ebenfalls eine Entscheidung für die Modellvariante M5 abgeleitet werden, jedoch versehen mit einem gewissen Vorbehalt für die Gesamtgruppe: M5 scheint noch annehmbar, wenngleich dieses Modell hier sozusagen die tatsächlichen Verhältnisse etwas mehr vereinfacht, als bei den anderen Zufriedenheiten.

Für die Gesamtgruppe aller westdeutschen wie auch die Subgruppe aller zum Panelbeginn mindestens 60-jährigen SOEP-Befragungspersonen kann also M5 als dasjenige sparsamste QMSM angenommen werden, durch das die Zeitreihendaten der verschiedenen Zu-

TABELLE 3.15:
Allgemeine Stabilitätsanalysen: Anpassungsgüte der QMSM für Freizeitzufriedenheit

	M1	M2	M3	M4	M5	M6	M7	M1b
<i>df</i>	91	104	119	117	118	118	104	106
Alle								
χ^2	274.3	6590.9	7152.5	633.1	658.9	7054.0	575.5	721.5
<i>TLI</i>	0.999	0.973	0.974	0.998	0.998	0.974	0.998	0.997
<i>NFI</i>	0.999	0.979	0.977	0.998	0.998	0.977	0.998	0.998
<i>IFI</i>	0.999	0.979	0.977	0.998	0.998	0.978	0.998	0.998
<i>CFI</i>	0.999	0.979	0.977	0.998	0.998	0.978	0.998	0.998
<i>RMSEA</i>	0.012	0.067	0.065	0.018	0.018	0.065	0.018	0.020
<i>RMSEA_U</i>	0.010	0.065	0.064	0.016	0.017	0.064	0.017	0.019
<i>RMSEA_O</i>	0.014	0.068	0.066	0.019	0.019	0.066	0.019	0.022
Alt84								
χ^2	166.0	957.8	1030.8	261.2	264.4	1028.2	218.9	278.0
<i>TLI</i>	0.998	0.977	0.979	0.997	0.997	0.979	0.997	0.995
<i>NFI</i>	0.997	0.981	0.979	0.995	0.995	0.979	0.996	0.994
<i>IFI</i>	0.998	0.983	0.981	0.997	0.997	0.982	0.998	0.996
<i>CFI</i>	0.998	0.983	0.981	0.997	0.997	0.982	0.998	0.996
<i>RMSEA</i>	0.020	0.063	0.061	0.024	0.024	0.061	0.023	0.029
<i>RMSEA_U</i>	0.015	0.059	0.057	0.020	0.020	0.057	0.019	0.025
<i>RMSEA_O</i>	0.025	0.066	0.064	0.028	0.028	0.064	0.027	0.033

df: Freiheitsgrade des Modells.

RMSEA_U/RMSEA_O: Untere/obere Grenze des *RMSEA*-90%-Konfidenzintervalls

Alle χ^2 -Werte sind signifikant mit $\alpha \leq 0.001$.

friedenheitsmessungen noch zufriedenstellend beschrieben werden. Somit erwiesen sich Veränderungsmodelle mit perfekter Merkmalsstabilität (nämlich M2, M3 und M6) generell als nicht angemessen, die Stabilität von Zufriedenheiten erscheint auf jeden Fall als eine bestenfalls näherungsweise. Jedoch kann die Veränderung der Zufriedenheiten gut charakterisiert werden durch das sparsamste Prozeßmodell der Stationarität (zweiter Ordnung): Weder finden sich bei diesen breit gefaßten Untersuchungsgruppen Hinweise auf deutliche Veränderungen in den True-Score-Varianzen der Zufriedenheiten, noch auf deutliche Veränderungen der Stabilitätskoeffizienten während des untersuchten Zeitraums. Um es also noch einmal mit Blick auf die Untersuchungshypothesen deutlich zu formulieren: Die Annahme von M5 für die „Altengruppe“ (Alt84) impliziert sowohl *keine* Zunahme der Varianz, als auch *keine* Zunahme der Stabilität bei diesen älteren SOEP-Teilnehmern über den untersuchten Befragungszeitraum hinweg!

Das stationäre Prozeßmodell M5 enthält nur noch sehr wenige Parameter der längs-

TABELLE 3.16:
Allgemeine Stabilitätsanalysen: Geschätzte True-Score-Varianzen

		ALZ		EIZ		GEZ		WOZ		FRZ	
		Alt84	Alle	Alt84	Alle	Alt84	Alle	Alt84	Alle	Alt84	Alle
M1	1985	3.07	2.25	3.66	3.69	4.01	3.74	1.94	2.68	2.37	2.64
	1986	2.56	1.94	3.39	3.24	4.51	3.77	1.80	2.48	2.82	3.06
	1987	2.79	1.97	3.50	3.24	4.54	3.52	2.18	2.63	2.85	3.08
	1988	3.34	2.15	3.14	3.11	4.81	3.67	1.95	2.43	2.44	3.02
	1989	3.18	2.21	3.31	3.23	4.52	3.85	1.93	2.54	2.25	3.26
	1990	3.30	1.92	3.03	2.79	4.37	3.45	1.91	2.66	1.94	2.30
	1991	2.93	1.68	3.02	2.65	4.66	3.54	1.37	2.42	1.84	3.43
	1992	2.85	1.72	3.10	2.69	5.11	3.85	1.59	2.59	1.51	3.09
	1993	3.55	2.02	3.39	3.09	4.80	3.59	1.58	2.54	2.64	3.11
	1994	3.25	1.94	3.27	3.12	5.23	3.66	1.45	2.50	2.53	3.11
	1995	3.14	2.04	3.37	3.10	5.23	3.67	1.84	2.40	2.18	2.31
	1996	3.52	1.97	3.68	3.25	5.14	3.72	1.88	2.35	2.22	3.11
	1997	3.90	2.15	3.33	3.22	5.62	3.62	2.20	2.37	2.64	3.26
	1998	3.52	2.01	2.60	2.95	4.77	3.29	2.36	2.28	2.97	2.84
M5		3.18	1.98	3.32	3.13	4.54	3.69	1.81	2.47	2.27	2.91

schnittlichen Veränderung bzw. Stabilität der Zufriedenheiten: Jeweils nur einen Stabilitätskoeffizienten ρ , eine True-Score-Varianz σ_T^2 und somit auch nur einen Quotienten Q_Δ , der nach den im Kapitel 2.2.1 vorgestellten Überlegungen als Indikator näherungsweise Parallelität zur weiteren Charakterisierung der Merkmalsstabilität hilfreich sein kann (vgl. dort Gleichung 2.20). Zwar wird die Interpretation der Ergebnisse der Parameterschätzungen durch diese erfreuliche Sparsamkeit sehr erleichtert, jedoch sollen hier auch die Ergebnisse für M1 zumindest berichtet werden, um auf diese Weise sicherzustellen, daß durch die Entscheidung für M5 allein aufgrund letztendlich subjektiv festgelegter Regeln zum Vergleich bestimmter Fit-Indizes (und Ignorierung der $\Delta\chi^2$ -Tests) nicht etwa doch gewisse längsschnittliche Trends in diesen Parametern unentdeckt bleiben, die von Belang sein könnten. Tabelle 3.16 zeigt darum geschätzte True-Score-Varianzen der Modellversionen M1 und M5, Tabelle 3.17 zeigt entsprechend die aus M1 und M5 geschätzten Stabilitätskennwerte, d.h. Stabilitätskoeffizienten und Q_Δ -Werte. Zu beachten ist, daß bei den M1-Ergebnissen die jeweiligen „äußeren“ Kennwerte für die Zeitreihe, also die True-Score-Varianzen für 1984 und 1999, sowie die Stabilitätskoeffizienten und Q_Δ -Werte für das erste und letzte Befragungsintervall fehlen: Diese sind, wie in Kapitel 2.2.2.1 und genauer in Anhang A.2 dargelegt wurde, in der unrestringierten Modellversion nicht identifiziert.

Die genaue Betrachtung der durch M1 berechneten True-Score-Varianzen der allgemeinen Lebenszufriedenheit in Tabelle 3.16 zeigt nun bei Alt84 und in der Gesamtgruppe durchaus gewisse Schwankungen über die untersuchte Zeitspanne hinweg, jedoch kaum

einen erkennbaren systematischen Trend: Allenfalls ansatzweise könnte eine Zunahme der True-Score-Varianzen der Älteren (Alt84) zum Ende hin aufgrund der letzten drei relativ hohen Werte für die Jahre 1996, 1997 und 1998 wahrgenommen werden. Interessanterweise war ja bei den Streuungen der beobachteten Werte eher eine Abnahme mit der Zeit beobachtet worden: Bei Zugrundlegung des im QMSM implizierten einfachen Meßmodells resultiert eine Abnahme der Meßwertvarianzen bei Zunahme der True-Score-Varianzen aus einer Abnahme der Meßfehlervarianzen bzw. wachsender Reliabilität der Messungen. Entsprechende Hinweise auf eine zumindest anfänglich während der ersten Befragungswellen stattfindende Zunahme der Reliabilitäten der Zufriedenheitsmessungen im SOEP wurden, wie schon erwähnt, von Landua (1993) und Schräpler (1995) gefunden und scheinen sich auch hier, wie noch gezeigt wird, zu bestätigen. Hier bleibt vorerst festzuhalten, daß ein schwach erkennbarer Anstieg der True-Score-Varianzen der Lebenszufriedenheit bei den Älteren kaum diskussionswürdig scheint, wie auch insgesamt hinsichtlich der Varianzschwankungen darauf hinzuweisen ist, daß diese natürlich deutlich geringer erschienen, wenn Standardabweichungen berichtet worden wären (wenn also die Quadrierung der Zufriedenheitsskalierung, die in den Varianzen „steckt“, rückgängig gemacht worden wäre). Es scheint darum nachvollziehbar, daß die Restriktion invarianter True-Score-Varianzen hier zu Modellversionen führt, deren Model-Fit kaum schlechter zu beurteilen ist. Wenn überhaupt, dann ist ein ähnlicher schwacher Trend zu einer Zunahme der Varianzen bei der Gesundheitszufriedenheit in der Gruppe Alt84 ansatzweise erkennbar, ansonsten sind in den sonstigen Spalten der Tabelle 3.16 keine erwähnenswerten Tendenzen einer Zu- oder Abnahme über die Zeit hinweg wahrzunehmen.

Interessanter, weil sehr deutlich zutage tretend, sind die jeweiligen Unterschiede der Varianzen von Alt84 im Vergleich zur Gesamtgruppe, die sowohl in den M1-, als in den M5-Ergebnissen festzustellen sind: Bei der allgemeinen Lebenszufriedenheit beträgt der Alt84-Varianzwert mehr als das 1.5-fache des Varianzwerts der Gesamtgruppe. Auch bei den anderen Zufriedenheiten zeigen sich Varianzunterschiede zwischen den beiden Gruppen: Die Substichprobe der Älteren variiert wiederum stärker in den True-Scores der Zufriedenheit mit der Gesundheit – der Bereich, der in beiden Gruppen die jeweils größte Varianz zeigt. Dagegen variieren die Älteren bei der Wohnzufriedenheit und der Freizeitzufriedenheit unterdurchschnittlich. Interessanterweise ist die Wohnzufriedenheit der Bereich mit der geringsten Varianz bei den Älteren, während dieses in der Gesamtgruppe die allgemeine Lebenszufriedenheit ist.

Bei den durch M1 geschätzten Stabilitätskoeffizienten, die in Tabelle 3.17 wiedergegeben sind, sind keine Trends einer monotonen Veränderung über die Zeitreihe hinweg erkennbar, es zeigen sich nur geringe Schwankungen dieser Werte, von einzelnen „Ausreißern“ abgesehen. Als letztere können v.a. drei Einsen gelten, die in der Tabelle auftauchen und für diese jeweiligen Erhebungsintervalle bis auf geringe Rundungsungenauigkeiten nahezu perfekte Stabilität der True-Scores anzeigen würden – m.E. sollten diese drei Schätzungen mit einer gewissen Vorsicht betrachtet werden bzw. verdeutlichen diese die

TABELLE 3.17:
Allgemeine Stabilitätsanalysen: Geschätzte Stabilitätskoeffizienten und Quotienten Q_{Δ}

		ALZ		EIZ		GEZ		WOZ		FRZ	
		Alt84	Alle	Alt84	Alle	Alt84	Alle	Alt84	Alle	Alt84	Alle
M1	$\hat{\rho}_2$	0.93	0.88	0.92	0.90	0.94	0.95	0.92	0.85	0.88	0.91
	$\hat{\rho}_3$	0.88	0.89	0.96	0.93	0.92	0.93	0.93	0.88	0.86	0.92
	$\hat{\rho}_4$	1.00	0.94	0.96	0.91	0.95	0.95	0.89	0.85	0.91	0.91
	$\hat{\rho}_5$	0.97	0.93	0.94	0.90	0.95	0.95	0.96	0.91	0.89	0.90
	$\hat{\rho}_6$	0.91	0.92	0.95	0.91	0.98	0.96	0.97	0.90	0.90	0.91
	$\hat{\rho}_7$	0.95	0.93	0.96	0.94	0.93	0.96	0.91	0.90	0.95	0.92
	$\hat{\rho}_8$	0.96	0.95	0.94	0.95	0.96	0.98	0.98	0.90	0.91	0.94
	$\hat{\rho}_9$	0.96	0.94	0.95	0.91	0.95	0.95	0.93	0.91	0.87	0.93
	$\hat{\rho}_{10}$	0.94	0.93	0.93	0.93	0.95	0.95	1.00	0.94	1.00	0.93
	$\hat{\rho}_{11}$	0.94	0.95	0.95	0.94	0.99	0.97	0.89	0.89	0.85	0.93
	$\hat{\rho}_{12}$	0.99	0.94	0.97	0.94	0.99	0.97	0.91	0.90	0.92	0.94
	$\hat{\rho}_{13}$	0.94	0.95	0.95	0.94	0.93	0.96	0.93	0.91	0.99	0.93
	$\hat{\rho}_{14}$	0.97	0.93	0.97	0.94	0.94	0.94	0.98	0.89	0.89	0.94
	M5	$\hat{\rho}$	0.94	0.93	0.95	0.93	0.95	0.96	0.94	0.90	0.90
M1	$Q_{\Delta 2}$	0.14	0.24	0.15	0.20	0.14	0.10	0.15	0.29	0.27	0.21
	$Q_{\Delta 3}$	0.26	0.21	0.08	0.15	0.15	0.13	0.17	0.24	0.27	0.16
	$Q_{\Delta 4}$	0.02	0.12	0.08	0.17	0.10	0.10	0.20	0.28	0.18	0.18
	$Q_{\Delta 5}$	0.06	0.15	0.13	0.21	0.10	0.11	0.07	0.19	0.21	0.20
	$Q_{\Delta 6}$	0.17	0.15	0.09	0.18	0.05	0.08	0.06	0.21	0.19	0.18
	$Q_{\Delta 7}$	0.11	0.13	0.08	0.12	0.15	0.08	0.17	0.19	0.10	0.25
	$Q_{\Delta 8}$	0.08	0.11	0.13	0.09	0.09	0.05	0.04	0.20	0.17	0.11
	$Q_{\Delta 9}$	0.10	0.14	0.10	0.19	0.09	0.10	0.14	0.18	0.46	0.14
	$Q_{\Delta 10}$	0.12	0.13	0.14	0.14	0.11	0.09	0.04	0.13	0.01	0.13
	$Q_{\Delta 11}$	0.12	0.10	0.10	0.11	0.02	0.07	0.25	0.22	0.27	0.14
	$Q_{\Delta 12}$	0.04	0.12	0.07	0.12	0.02	0.06	0.16	0.21	0.16	0.17
	$Q_{\Delta 13}$	0.14	0.12	0.09	0.12	0.15	0.07	0.16	0.17	0.02	0.15
	$Q_{\Delta 14}$	0.05	0.14	0.07	0.11	0.12	0.11	0.04	0.22	0.24	0.12
	M5	Q_{Δ}	0.13	0.13	0.10	0.14	0.10	0.08	0.12	0.21	0.19

Notwendigkeit von Restriktionen der Stabilitätskoeffizienten, da hier allem Anschein nach das unrestringierte Modell Lösungen erlaubt, die zwar eine gute Likelihood aufweisen, jedoch schlicht unrealistisch sind. Die Entscheidung für M5 ist also insgesamt hier nicht nur anhand der sonstigen Werte nachvollziehbar, sondern vermeidet dazu noch dieses Problem.

Insgesamt sind die Werte der Stabilitätskoeffizienten sehr hoch: So liegen sie z.B. alle deutlich über dem von Cohen (1992) vorgeschlagenen Cutoff-Wert zur Effektstärkewertung eines Korrelationswerts als „hoch“ und bedeutet der Wert 0.94, der nach M5 die Stabilität der True-Scores allgemeiner Lebenszufriedenheit kennzeichnet, daß die True-

Score-Verteilungen zweier aufeinanderfolgender Meßzeitpunkte jeweils 88% „gemeinsame Varianz“ besitzen. Für die „wahren“ Werte der Zufriedenheiten wären demnach also insgesamt sehr hohe Stabilitäten charakteristisch, welche deutlich über den durch die Meßfehler „verwässerten“ Stabilitäten der Meßwerte liegen (vgl. Tabelle 3.9 bzw. in Tabelle 3.10 die lag-1-Korrelationen).

Die Unterschiede zwischen den Alt84-Stabilitäten und denen der Gesamtgruppe sind gering und scheinen vernachlässigbar. Auch die höchste Differenz bei den Koeffizienten der Wohnzufriedenheitsstabilität erscheint nicht mehr allzu bedeutend, wenn man sie in Varianzaufklärungsanteile umrechnet: Dem $R^2 = 0.88$ der Älteren steht ein $R^2 = 0.81$ der Gesamtgruppe gegenüber. Es ist also nicht erkennbar, daß die Stabilität der Subgruppe älterer Befragungspersonen über der aller Erwachsener liegt.

Genauso, wie bei den Stabilitätskoeffizienten, sind auch bei den aus den M1-Ergebnissen berechneten Q_{Δ} -Werten keine systematischen Trends über den Zeitverlauf hinweg, sondern lediglich eine gewisse Schwankungsbreite in allen Spalten der Tabelle 3.17 zu erkennen. Es handelt sich, daran sei nochmals erinnert, um den Quotienten $\sigma_{\Delta}^2/\sigma_t^2$ in einem Erhebungsintervall t bis $t + 1$, d.h. um das Verhältnis der Veränderungsvarianz zur „Ausgangsvarianz“ der True-Scores in diesem Intervall – trotz der Schwankungen der Werte zeigen sich die Veränderungsvarianzen in dieser Relation stets sehr gering. Angesichts der sehr hohen Stabilitätswerte kann deshalb hier eine näherungsweise Parallelität der Verläufe konstatiert werden.

Insgesamt also zeigen die Analysen des QMSM bei der Gesamtgruppe aller Befragter und genauso bei der der 1924 oder früher Geborenen eine sehr hohe und näherungsweise parallele Stabilität der True-Scores in allen untersuchten Zufriedenheiten, welche selbst in ihrem zeitlichen Verlauf über die 16-jährige Untersuchungsperiode hinweg durch Stationarität gekennzeichnet ist, d.h. sich kaum verändert. Belege für eine erhöhte Stabilität v.a. der Lebenszufriedenheit der Älteren wurden hier nicht gefunden.

Bei hoher näherungsweise paralleler True-Score-Stabilität in allen untersuchten Zufriedenheitsbereichen bleibt nun noch der Aspekt der Mittelwertsverschiebungen zu beachten: Der Vergleich der Modellversionen M1 und M1b – bei der die Invarianz aller 16 Mittelwerte festgelegt ist – dient der Beurteilung der Frage, ob bei den Zufriedenheiten möglicherweise sogar eine annähernd strikte Merkmalsstabilität vorliegt. In der jeweils letzten Spalte der Tabellen 3.11 bis 3.15 sind die Model-Fit-Kennwerte für die Modellversion M1b abgebildet. Nach den vorgegebenen Vergleichskriterien ergibt sich hier nicht über alle Zufriedenheiten hinweg dasselbe eindeutige Ergebnis: Prinzipiell schlägt sich die Restriktion invarianter Mittelwerte kaum in den Werten der inkrementellen Fit-Indizes TLI, NFI, IFI und CFI nieder, so daß hier praktisch nie ein nennenswerter Abfall bei M1b im Vergleich zu M1 resultiert. Sensitiv gegenüber der Restriktion zeigt sich dagegen der RMSEA: Bei der allgemeinen Lebenszufriedenheit (Tabelle 3.11) sowie der Zufriedenheit mit der Gesundheit (Tabelle 3.13) ist jeweils in beiden Untersuchungsgruppen der Anstieg des RMSEA-Wertes > 0.01 zu erkennen und überlappen sich die Konfidenzintervalle nicht, so daß also stets das

TABELLE 3.18:
Allgemeine Stabilitätsanalysen (M1): Geschätzte Mittelwerte[†]

	ALZ		EIZ		GEZ		WOZ		FRZ	
	Alt84	Alle	Alt84	Alle	Alt84	Alle	Alt84	Alle	Alt84	Alle
1984	7.40	7.50	6.85	6.48	5.44	7.14	8.42	7.93	8.19	7.45
1985	7.26	7.30	6.93	6.49	5.55	7.06	8.30	7.86	8.09	7.27
1986	7.36	7.35	6.99	6.56	5.53	6.98	8.29	7.81	8.08	7.15
1987	7.13	7.14	7.08	6.56	5.47	6.89	8.32	7.85	8.08	7.20
1988	6.85	7.03	6.97	6.57	5.20	6.78	8.19	7.75	7.99	7.02
1989	6.83	7.08	7.05	6.61	5.10	6.71	8.25	7.73	8.00	6.96
1990	6.88	7.28	7.16	6.69	5.07	6.71	8.19	7.68	7.80	6.99
1991	7.05	7.36	7.38	6.93	5.01	6.67	8.44	7.75	8.27	6.89
1992	6.82	7.27	7.00	6.78	4.87	6.77	8.34	7.69	8.29	6.99
1993	6.76	7.18	7.03	6.65	4.82	6.59	8.36	7.76	7.98	6.91
1994	6.58	7.06	7.02	6.48	4.55	6.48	8.50	7.76	8.01	6.88
1995	6.39	7.00	7.17	6.53	4.48	6.49	8.26	7.74	7.77	6.95
1996	6.47	7.04	7.06	6.58	4.45	6.44	8.23	7.79	8.16	7.09
1997	6.31	6.87	6.87	6.42	4.45	6.39	8.12	7.72	8.06	7.00
1998	6.25	6.94	7.06	6.45	4.38	6.41	7.98	7.70	7.92	7.03
1999	6.12	6.97	6.96	6.50	4.08	6.33	7.97	7.70	8.12	7.00

[†] Geschätzte Intercepts der Meßmodellgleichungen für die beobachteten Variablen X_{ti}

restringierte Modell M1b zurückzuweisen wäre. Bei der Einkommenszufriedenheit (Tabelle 3.12) gilt selbiges für die Gesamtgruppe, während bei der Gruppe Alt84 die Kriterien für die Annahme des restringierteren Modells knapp erfüllt wären. Letzteres wiederum gilt auch für die Freizeitzufriedenheit (Tabelle 3.15) in der Gruppe Alt84, während hier bei der Gesamtgruppe die Entscheidung sozusagen am schwersten fällt, da einerseits der RMSEA-Anstieg ≤ 0.01 ist, andererseits sich aber die Konfidenzintervalle nicht überlappen. Bei der Wohnzufriedenheit (Tabelle 3.14) dagegen erfüllt M1b sowohl bei Alt84, als auch in der Gesamtgruppe die Annahmekriterien. Insgesamt allerdings zeigt das restringierte Modell M1b in jedem Falle noch Fit-Kennwerte, die, von der vergleichenden Betrachtung abgesehen, jeweils noch eine gute Anpassungsgüte anzeigen, so daß die Veränderungen der (geschätzten) Mittelwerte in jedem Fall nicht so ausgeprägt sind, daß nicht auch Modelle invarianter zentraler Tendenzen noch gut „auf die Daten passen“ würden. Offenbar aber ist die längsschnittliche Varianz der Mittelwerte am ausgeprägtesten bei der allgemeinen Lebenszufriedenheit und der Zufriedenheiten mit der Gesundheit, sowie auch in der Gesamtgruppe aller Befragter bei der Einkommenszufriedenheit.

Die dazugehörigen Mittelwertsschätzungen des unrestringierten Modells M1 (die auch bei den Versionen M2 bis M7 identisch sind) zeigt Tabelle 3.18. Sie weichen in der Gesamtgruppe meist nur geringfügig, bei der Gruppe Alt84 zum Teil jedoch schon etwas deutlicher

von den in Tabelle 3.7 angegebenen Stichprobenmittelwerten ab. Diese Abweichungen ergeben sich infolge der fehlenden Werte in der Datenmatrix und der Anwendung des FIML-Algorithmus, bei der die Schätzung der Mittelwerte nicht auf einer „available-case analysis“ (paarweiser Ausschluß) beruht, wie bei den Werten in Tabelle 3.7, – es handelt sich jeweils um Schätzungen des Populationsmittelwerts unter Einbeziehung der Fälle mit fehlenden Werten.³² Folglich kann es nicht überraschen, daß die Abweichungen in den späteren Erhebungswellen und in der Gruppe Alt84 größer erscheinen, da hier ja hier jeweils größere Anteile von Dropouts vorhanden sind.

Es zeigt sich, daß dort, wo die Restriktion invarianter Mittelwerte zu besonders deutlichen Fit-Verlusten geführt hat, nämlich bei der allgemeinen Lebenszufriedenheit und bei der Gesundheitszufriedenheit, die Veränderung der Mittelwerte jeweils in beiden Gruppen durchaus als Abnahme über die Zeit beschreiben werden kann. Lediglich bei den Durchschnittswerten der Einkommenszufriedenheit in der Gesamtgruppe scheint aus der Veränderung kein klarer Trend herauslesbar, allerdings erscheinen diese Veränderungen auch sehr gering, so daß es beinahe etwas überrascht, daß hier M1b zurückgewiesen wurde. Das Fazit dieser Betrachtungen kann wie folgt lauten: Es zeigt sich in den Mittelwerten der Einkommens-, Wohn- und Freizeit Zufriedenheit jeweils in beiden Untersuchungsgruppen eine mehr oder weniger ausgeprägte unsystematische Invarianz von insgesamt geringem Ausmaß, während die Durchschnittswerte der allgemeinen Lebenszufriedenheit und der Gesundheitszufriedenheit durch einen deutlicheren Trend der Abnahme über die Untersuchungszeitspanne charakterisierbar sind, welcher jeweils in der Gruppe der Älteren deutlicher zutage tritt, als in der Gesamtgruppe aller Befragter.

Die Darstellung der Ergebnisse der Modellrechnungen für die allgemeinen Stabilitätsanalysen wird vervollständigt durch die geschätzten Reliabilitäten – nämlich der „inneren Reliabilitäten“ des unrestringierten Modells M1, welche in Tabelle 3.19 angegeben sind. Sie liegen in den ersten Untersuchungswellen ungefähr auf demselben Niveau, wie die von Schräpler (1995) anhand derselben Daten und eines QMSM – allerdings unter fallweisem Ausschluß fehlender Werte – geschätzten. Eine sehr schwache Tendenz zur Zunahme der Reliabilitäten im Panelverlauf scheint in allen Spalten der Tabelle erkennbar, auch liegen bei der Lebens-, Einkommens- und Gesundheitszufriedenheit die Reliabilitäten der Gruppe Alt84 in den meisten Jahren etwas über denen der Gesamtgruppe aller Befragter, während es sich bei der Wohnzufriedenheit eher umgekehrt zu verhalten und bei der Freizeit Zufriedenheit keine solche Tendenz erkennbar scheint. Insgesamt sind diese Unterschiede jedoch jeweils eher gering und befinden sich die Reliabilitäten durchweg auf einem Niveau, welches man vielleicht mit „hoch, aber nicht optimal“ bezeichnen könnte – nämlich sozusagen nicht optimal für für Anwendungen der Zufriedenheitsratings, die keine Modellierung der latenten True-Scores beinhalten, und hoch genug für eine solche Modellierungen in Struk-

³²Im vorliegenden Fall sind die Fälle mit fehlenden Werten oft die im Panelverlauf verstorbenen Befragungspersonen, deren Werte nicht im strengen Wortsinne „fehlen“, sondern tatsächlich nicht existieren. Der Sinn des FIML-Verfahren für diese Art der Dropouts wurde bereits in Kapitel 2.3.1 erörtert.

TABELLE 3.19:
Allgemeine Stabilitätsanalysen (M1): Geschätzte Reliabilitäten[†]

	ALZ		EIZ		GEZ		WOZ		FRZ	
	Alt84	Alle	Alt84	Alle	Alt84	Alle	Alt84	Alle	Alt84	Alle
1985	0.60	0.56	0.65	0.62	0.58	0.62	0.53	0.58	0.52	0.47
1986	0.55	0.54	0.65	0.60	0.65	0.64	0.52	0.57	0.62	0.55
1987	0.59	0.52	0.69	0.62	0.63	0.62	0.61	0.60	0.63	0.57
1988	0.62	0.55	0.60	0.61	0.66	0.64	0.57	0.58	0.56	0.56
1989	0.63	0.57	0.68	0.66	0.62	0.64	0.61	0.61	0.57	0.60
1990	0.65	0.60	0.64	0.60	0.63	0.63	0.58	0.63	0.50	0.55
1991	0.63	0.57	0.71	0.62	0.66	0.64	0.54	0.61	0.51	0.60
1992	0.61	0.58	0.70	0.64	0.71	0.68	0.59	0.64	0.46	0.60
1993	0.71	0.62	0.68	0.65	0.67	0.65	0.55	0.64	0.58	0.59
1994	0.65	0.60	0.67	0.62	0.71	0.65	0.51	0.61	0.60	0.57
1995	0.59	0.62	0.72	0.64	0.69	0.68	0.58	0.61	0.54	0.55
1996	0.69	0.62	0.74	0.68	0.70	0.68	0.61	0.61	0.60	0.62
1997	0.75	0.65	0.68	0.68	0.76	0.68	0.65	0.64	0.62	0.63
1998	0.67	0.62	0.56	0.62	0.65	0.63	0.60	0.60	0.59	0.57

[†] R^2 -Werte der Meßmodellgleichungen für die beobachteten Variablen X_{ti}

turgleichungsmodellen. Die Reliabilitäten der Zufriedenheitsratings sind nicht Gegenstand der theoretischen Fragestellung, die die vorliegende Untersuchung veranlaßt haben und sollen deshalb im folgenden auch nicht weiter besprochen werden, ihre Auflistung in Tabelle 3.19 dient hier nur der Information des Lesers.

3.3.1.4 Zusammenfassung und Diskussion der allgemeinen Stabilitätsanalysen

Das m.E. herausragende Ergebnis der allgemeinen Stabilitätsanalysen ist die durchweg sehr hohe Merkmalsstabilität der Zufriedenheiten, die dort sowohl für die Gesamtgruppe aller Befragter, als auch für die Subgruppe derjenigen Untersuchungsteilnehmer, die zum Panelbeginn 1984 mindestens 60 Jahre alt waren, gefunden wurde: Alle geschätzten Stabilitätskoeffizienten, d.h. Korrelationen der meßfehlerbereinigten True-Scores der Zufriedenheiten, liegen sehr nahe am theoretisch möglichen Maximalwert, zudem zeigen die berechneten Quotienten Q_{Δ} , daß die Varianzen der True-Score-Veränderungen in Relation zur „Ausgangsvarianz“ des jeweiligen Erhebungsintervalls sehr niedrig sind, so daß insgesamt annähernd parallele Stabilität der Zufriedenheiten konstatiert werden konnte. Kurz gesagt: Zufriedenheitsbewertungen sind sehr stabil. Sie sind es offenbar über die gesamte Spanne des Erwachsenenalters hinweg und nicht nur im höheren Lebensalter.

Einschränkend sind hier lediglich die längsschnittlichen Veränderungen der Durchschnittswerte zu berücksichtigen, die bei der allgemeinen Lebenszufriedenheit und Gesund-

heitszufriedenheit, sowie auch in der Gesamtgruppe aller Befragter bei der Einkommenszufriedenheit immerhin ausgeprägt genug waren, um zur Zurückweisung eines Modells invarianter Mittelwerte zu führen (mit dem allerdings jeweils noch eine gute Modellanpassungsgüte erreicht wurde): In diesen Fällen könnte die Stabilität dann konsequenterweise nicht als annähernd strikte bezeichnet werden. Bei hoher und annähernd paralleler Stabilität bedeutet dieses gewissermaßen allgemeingültige, interindividuell kaum variierende Veränderungseffekte, die zwischen einzelnen Erhebungszeitpunkten bei allen (bzw. den meisten) Personen der untersuchten Population wirksam waren – eben Mittelwertverschiebungen bei weitestgehender Invarianz der individuellen Abweichungen vom jeweiligen Mittelwert.

Sieht man davon ab, so reicht auch das Ergebnis einer näherungsweise bloß parallelen Merkmalsstabilität der Zufriedenheiten, insbesondere der allgemeinen Lebenszufriedenheit, aus, um daraus einen bedeutsamen empirischen Beleg für die Relevanz der Hypothese allgemeiner lebenslanger Stabilität von Zufriedenheit abzuleiten: Personen bewahren weitestgehend ihre „relative Position“ über die Zeit hinweg. Dieses Ergebnis ist keinesfalls trivial: Wie im Theorieteil dieser Arbeit dargestellt, wurde die Stabilität der Lebenszufriedenheit, die in der Theorie des SWB v.a. durch trait-theoretische Konzepte akzentuiert wurde, in der diesbezüglichen Fachliteratur durchaus kontrovers diskutiert und variieren auf empirischen Ergebnissen begründete Einschätzungen des Ausmaßes dieser Stabilität. Nach den hier geschätzten Stabilitätskoeffizienten erscheint jedenfalls die Merkmalsstabilität erheblich höher, als ein vorläufiger Gesamteindruck, der aus der vorne dargestellten diesbezüglichen Diskussion gewonnen werden könnte, vermuten läßt. Tatsächlich sind empirische Ergebnisse, auf denen Aussagen über die Stabilität von Lebenszufriedenheit begründet wurden, meist aus statistischen Analysen abgeleitet, die keine Modellierung latenter True-Scores beinhalten und deshalb die Reliabilitätsproblematik außer Acht lassen. Es wurde bereits darauf hingewiesen und im Anhang A.1 formal gezeigt, daß längsschnittliche serielle Korrelationen von Meßwerten in der Regel die Zusammenhänge der „wahren“ Werte und somit die tatsächliche Stabilität des untersuchten Merkmals unterschätzen. Die berichteten Ergebnisse belegen dieses nun auf sehr eindrucksvolle Weise: Die geschätzten Ein-Jahres-Korrelationen der True-Scores liegen erheblich über den entsprechenden Meßwertkorrelationen in der Stichprobe (Tabelle 3.9) und zeichnen sozusagen ein Bild der Stabilität, das sich von dem unterscheidet, welches letztere vermitteln würden. Dieser Umstand ist als ein weiteres, für die Wohlbefindentheorie insgesamt relevantes Ergebnis der allgemeinen Stabilitätsanalysen hervorzuheben: Obwohl die Stabilität der Zufriedenheitsmeßwerte in einem mittleren Bereich (des möglichen Ausmaßes) liegt, ist die Stabilität der True-Scores sehr hoch. Zufriedenheiten wären demnach also in *jedem* – und folglich auch im höheren – Lebensalter sehr stabil.

Mit Blick auf die Hypothese altersspezifisch erhöhter Stabilität ist festzuhalten, daß sich bei den Analysen kein deutlicher Anstieg der geschätzten Stabilitätskoeffizienten über die Zeitreihe hinweg bei der Gruppe der Höheraltrigen zeigte bzw. auch für diese Grup-

pe ein stationäres Prozeßmodell für alle untersuchten Zufriedenheiten ausgewählt werden konnte. Somit sind sozusagen die Zufriedenheiten dieser Alten während ihres weiteren Alters nicht stabiler geworden, als sie es von Anfang an sowieso schon waren. Wie schon dargelegt, verhindert hier allerdings die hinsichtlich der einbezogenen Spanne von Geburtsjahrgängen sehr breite Definition der Höheraltrigen Gruppe eine eindeutige Bewertung dieses Ergebnisses dergestalt, daß daraus auf das Nichtvorhandensein irgendeines Alterseffekts auf die Stabilität geschlossen werden könnte. Denn es ist immerhin möglich, daß ein solcher Effekt nicht über die gesamte Spanne der Höheraltrigkeit gleichbleibend wirksam ist, sondern z.B. eine Zunahme der Stabilität während einer ersten Phase des höheren Lebensalters stattfindet und danach verebbt: Dann wäre bei der untersuchte Höheraltrigen Gruppe eine allenfalls schwache Abbildung dieser Zunahme zu erwarten, welche zudem gegen Ende der Untersuchungsspanne abflaut, da anfangs diese Gruppe noch Alte aus dem gesamten Spektrum der Höheraltrigkeit (≥ 60) enthält, sich im Verlauf aber zu einer Gruppe der „ältesten Alten“ wandelt (1999 ist es die Gruppe der mindestens 76-Jährigen). Viel eher müßten sich allerdings im höheren Alter erhöhte Stabilitäten darin zeigen, daß auch in der breit definierten Höheraltrigen Gruppe die Stabilitäten insgesamt über denen der Gesamtgruppe aller Befragter liegen, da ja zur Berechnung letzterer auch die dann „instabilen“ jüngeren Personen miteinbezogen werden. Solche Unterschiede zeigten sich jedoch allenfalls in einem vernachlässigbaren Ausmaß (vgl. Tabelle 3.17), so daß hier zu resümieren bleibt, daß aus den allgemeinen Stabilitätsanalysen sich keinerlei Hinweise auf eine im höheren Lebensalter wachsende intraindividuelle Stabilität von Zufriedenheitsurteilen ergeben. Die folgenden altersspezifischen Stabilitätsanalysen sollten diesbezüglich allerdings aussagekräftiger sein.

Die Betrachtung der geschätzten True-Score-Varianzen ist v.a. zur Bewertung der Gewinner-und-Verlierer-Hypothese relevant. Zum einen wuchsen die Varianzen der Zufriedenheiten der Höheraltrigen während der untersuchten Zeitspanne *nicht* an. Andererseits zeigten sich deutliche Varianzunterschiede bei den Höheraltrigen im Vergleich zur Gesamtgruppe: In der Gruppe der Älteren variiert demnach die „wahre“ Lebenszufriedenheit zu jedem Zeitpunkt stärker, als insgesamt bei allen erwachsenen Personen (vgl. Tabelle 3.16). Während also im längsschnittlichen Verlauf kein Anstieg der Lebenszufriedenheitsvarianz älterer Menschen beobachtet wurde, erscheint hier – beim sozusagen querschnittlichen Vergleich der Werte aus einer sehr großen Zeitspanne von 16 Jahren – durchaus ein Effekt des höheren Lebensalters erkennbar, der der Gewinner-und-Verlierer-Hypothese entspräche. Zudem gilt auch hier wiederum, daß dann, wenn diese Hypothese die Lebenszufriedenheit im höheren Alter zutreffend beschriebe, bei diesen allgemeinen Stabilitätsanalysen ein allzu deutlicher längsschnittlicher Anstieg der Varianzen in der Gruppe Alt84 nicht zwingend zu erwarten wäre: Wieder würde, wenn z.B. tatsächlich beim Übergang in die Hochaltrigkeit allmählich die Varianz der Zufriedenheiten ansteige, dieser Anstieg später aber wieder abflaute, diese Varianzveränderung hier wohl weitestgehend dadurch „aufgezehrt“, daß ja bereits von Anfang an Personen aus dem gesamten Altersspektrum der Höheraltrigkeit die-

se Stichprobe bilden. In diesem Fall sollte aber die Gruppe Alt84, die dann ja zum größten Teil den Altersbereich während des und nach dem Anstieg der Varianzen auf ein hohes Niveau abdeckt, insgesamt höhere Varianzen aufweisen, als die Gesamtgruppe, die von Personen in den Lebensaltersbereichen vor der Wirksamkeit dieses Effekts dominiert wird. Hinsichtlich der Frage eines längsschnittlichen Effekts wachsender Varianzen im höheren Lebensalter sollten die nachfolgenden altersspezifischen Stabilitätsanalysen, mit ihren enger gefaßten Altersgruppen, aufschlußreicher sein. Vorerst ist hier, aufgrund der in Tabelle 3.16 berichteten Werte, der – wie gesagt eigentlich querschnittliche – Befund einer gegenüber der Gesamtgruppe aller Erwachsener erhöhten True-Score-Varianz der Lebenszufriedenheit Höheraltriger festzuhalten, der in Richtung der Gewinner-und-Verlierer-Hypothese weist.

Ein ähnlich deutlicher Gruppenunterschied der True-Score-Varianzen, wie bei der Lebenszufriedenheit, zeigt sich bei der Zufriedenheit mit der Gesundheit. Daß dem so ist und daß überhaupt dieser Bereich der bei beiden Gruppen am stärksten variierende ist, scheint ohne weiteres plausibel erklärbar: Man könnte annehmen, daß Veränderungen im Bereich Gesundheit – d.h. gesundheitliche Probleme – für die Betroffenen unmittelbar physisch erlebte Verluste bedeuten, welche auch durch kognitive Adaptationen kaum abgemildert werden können, so daß sie sich auch entsprechend unmittelbarer in diesbezüglichen Bewertungen niederschlagen. Dieses würde zunächst die im Vergleich zu den anderen Bereichen größere Varianz erklären und dann auch die größere Varianz bei den Älteren im Vergleich zur Gesamtgruppe, denn die Älteren sind von gesundheitlichen Problemen in stärkerem Maße betroffen. Es ist hier daran zu erinnern, daß ja gerade diese Häufung gesundheitlicher Probleme im höheren Lebensalter ein besonders wichtiger Grund für die Erwartung war, welche durch die Befunde des Zufriedenheitsparadoxes nicht erfüllt wurde, und daß sie genauso bei der Begründung der Gewinner-und-Verlierer-Hypothese sozusagen die Hauptrolle spielte, d.h. daß es v.a. die von jedem Menschen im höheren Lebensalter erwarteten körperlichen Einschränkungen sein sollten, die möglicherweise nicht nur dann, wenn sie sich erfüllen, zu Zufriedenheitsverlusten führen, sondern auch dann, wenn sie sich nicht erfüllen, zu Zufriedenheitsgewinnen. Deshalb wäre bei Gültigkeit der Gewinner-und-Verlierer-Hypothese gerade auch in dieser Bereichszufriedenheit ein entsprechender ähnlicher Effekt zu erwarten, wie für die allgemeine Lebenszufriedenheit. Somit könnte hier der Varianzunterschied zwischen der Gruppe der Älteren und der Gesamtgruppe als erneuter Hinweis auf einen solchen Effekt gewertet werden.

Allerdings erschöpft sich diese Gewinner-und-Verlierer-Hypothese nicht in der Annahme wachsender Varianzen in der (zumindest anfänglichen) Phase des höheren Lebensalters, sondern impliziert auch eine gewisse Instabilität der Zufriedenheiten: Die Hypothese nimmt ja an, daß Veränderung stattfindet, durch welche die Varianzen wachsen, sie ist somit zwar nicht prinzipiell logisch gänzlich unvereinbar mit einer hohen monotonen Stabilität des Merkmals, jedoch mit einer hohen Parallelität der Merkmalsverläufe, wie sie hier näherungsweise aufgrund der niedrigen Q_{Δ} -Werte konstatiert wurde. Damit ist eine gewisse Widersprüchlichkeit benannt, die in den berichteten Ergebnissen enthalten ist: Ei-

nerseits führt die Annahme einer über das Erwachsenenalter hinweg gleichermaßen sehr hohen und annähernd parallelen Stabilität zwingend zu der Annahme allenfalls geringfügig sich verändernder Varianzen (die ja auch der Entscheidung für das stationäre Modell M5 mit invarianten Varianzen entspricht), andererseits aber sind die Varianzen der Lebens- und Gesundheitszufriedenheit in der Subgruppe der Höheraltrigen deutlich erhöht. Anders ausgedrückt: Wie sollten mit den für die Gesamtgruppe errechneten Stabilitätskoeffizienten und Q_{Δ} -Werten die jüngeren in dieser Gesamtgruppe jemals die Varianzen der Subgruppe der Älteren erreichen, wenn sie selbst alt geworden sind? Den wenn schon die Varianzen der Gesamtgruppe, die ja die Älteren einschließt, unter denen der höheraltrigen Subgruppe liegen, dann müssen zumindest bei Teilen der jüngeren Personen diese Varianzen noch viel niedriger sein.

Es liegt auf der Hand, daß das so formulierte Problem nur solange eines ist, wie jeglicher „zeitbezogene“ Effekt als monotoner Alterseffekt über die gesamte Lebensspanne hinweg konzeptualisiert wird. Andernfalls könnte hier beispielsweise ein Anzeichen für einen Kohorteneffekt gesehen werden: Möglicherweise sind die True-Score-Varianzen der Lebens- und Gesundheitszufriedenheit in der Gruppe der 1924 oder früher Geborenen nicht irgendwann einmal auf ihr vergleichsweise hohes Niveau angewachsen, sondern befanden sich schon immer und unverändert auf demselben, während auch jüngere Subgruppen mit niedriger Varianz später einmal, wenn sie gealtert sind, diese niedrige Varianz beibehalten werden. Möglicherweise aber findet bei diesen Varianzen auch ein Anstieg regelmäßig während einer früheren Lebensphase statt: Dieses würde bedeuten, daß in dieser Phase auch die Veränderungsvarianzen bzw. die Koeffizienten Q_{Δ} erhöht und die Verläufe weniger parallel sind. Mit dieser Überlegung wird aber m.E. der thematische Rahmen der vorliegenden Untersuchung, der ja auf die Zufriedenheiten des höheren Lebensalters eingegrenzt ist, verlassen, so wie auch eine Kohortenhypothese bezüglich der Zufriedenheitsvarianzen weder theoretisch begründet wurde, noch m.E. zur Erklärung des Zufriedenheitsparadoxes beiträgt – zur weiteren Abklärung dieser beiden Möglichkeiten wären zusätzliche empirische Analysen erforderlich, die hier aber nicht unternommen werden sollen, da es sich gewissermaßen um eine Abschweifung vom Thema handeln würde. Es bleibt jedoch zu diesem Punkt festzuhalten, daß die hohe annähernd parallele Stabilität – und insbesondere die niedrigen Q_{Δ} -Werte – der True-Scores auch der Lebens- und Gesundheitszufriedenheit in der Älterengruppe der Gewinner-und-Verlierer-Hypothese widersprechen, trotz der mit dieser Hypothese konformen erhöhten Varianzen bei den Älteren, welche somit möglicherweise nicht durch diese Hypothese zu erklären sind.

Ein weiterer interessanter und diskussionswürdiger Befund sind die in Tabelle 3.18 aufgezeigten längsschnittlichen Veränderungen der geschätzten Mittelwerte. Wie schon gesagt ist dort, wo die Restriktion unveränderter Mittelwerte zu den deutlichsten Verschlechterungen der Modellanpassungsgüte geführt hat, nämlich bei beiden Untersuchungsgruppen jeweils bei der Lebens- und Gesundheitszufriedenheit, ein Absinken dieser Werte über den untersuchten Zeitraum hinweg festzustellen. Dieses Ergebnis paßt zu den Befunden

aus den deskriptiven Analysen (vgl. Kapitel 3.2.1), die ja bei genau diesen beiden Zufriedenheiten einen längsschnittlichen Abfall zeigen – allerdings zumindest im Falle der Lebenszufriedenheit nur im höheren Lebensalter. Es könnten jedoch durchaus die in der Gesamtgruppe enthaltenen Älteren sein, die auch für diese Untersuchungsgruppe ein längsschnittliches Absinken der Durchschnittswerte bewirken, wozu es gut passen würde, daß bei den Älteren die längsschnittliche Negativtendenz deutlicher ausgeprägt ist (ähnliches wurde ja auch schon bei den deskriptiven Werten der Tabelle 3.7 beobachtet). Auch die bei den Analysen zur Kohorteneffekthypothese errechneten Alterseffekte ließen bereits den hier zutage tretenden Mittelwertsabfall erwarten (vgl. Kapitel 3.2.2). Lägen nicht bereits diese Ergebnisse aus den vorangegangenen Analyseschritten vor, so würde hier wiederum das breite Geburtsjahrgangsspektrum der Untersuchungsgruppe der Älteren eine eindeutige diesbezügliche Interpretation verbieten: Auch die in Tabelle 3.18 aufgeführten Werte zeigen *nicht* einen kontinuierlichen Abstieg durchschnittlicher Lebenszufriedenheit im höheren Erwachsenenalter. Denn es könnten die in dieser Gruppe von Jahr zu Jahr in wachsendem Anteil enthaltenen „älteren“ und „ältesten Alten“ sein, deren Zufriedenheitsverluste ein Absinken der Durchschnittswerte verursachen, während ein solches Absinken bei den jüngeren Höheraltrigen nicht stattfindet. Die vorherigen Analysen haben aber gezeigt, daß auch schon während der frühen Höheraltrigkeit ein derartiger Negativtrend besteht. Auf die prinzipielle Möglichkeit, daß es sich bei diesem Negativtrend eher um einen Perioden-, als um einen Alterseffekt handeln könnte, wurde dort auch bereits ausgiebigst hingewiesen.

Bei der einleitenden Deskription der beiden Untersuchungsgruppen wurden bereits die Mittelwertsdifferenzen der Tabelle 3.7 in Relation zur Streuung der Meßwerte betrachtet und für praktisch sehr gering befunden. Vergleicht man nun die Werte der Tabelle 3.18 mit den geschätzten Streuungen der True-Scores, so erscheinen sie in anderem Licht: Beispielsweise beträgt die aus M5 geschätzte Alt84-Standardabweichung der True-Scores allgemeiner Lebenszufriedenheit $\sqrt{3.18} = 1.78$ (vgl. Tabelle 3.16) und die maximale Mittelwertsdifferenz 1.28 (vgl. Tabelle 3.18) – sie erscheint also nicht mehr ganz und gar gering im Verhältnis zu dieser Streuung. Dieses ist auch zu erwarten, da ja die „wahre“ Varianz infolge der Unkorreliertheit von Meßfehlern und True-Scores zwangsläufig unter der Varianz der Meßwerte liegen muß. Jedenfalls sind die in Tabelle 3.18 sichtbaren Mittelwertsdifferenzen nicht mehr einfach dadurch hinweg zu diskutieren, daß sie angesichts der True-Score-Streuungen als geringfügig eingeschätzt werden.

Andererseits ist bei der Bewertung der Mittelwertsveränderungen auch noch einmal zu berücksichtigen, daß die Restriktion invarianter Mittelwerte in allen Fällen zu Model-Fit-Werten geführt hat, die für sich genommen noch durchaus als gut zu bewerten wären. In gewisser Weise bewegten sich hier die Vergleiche des unrestringierten mit dem restringierten Modell stets in einem Grenzbereich, in dem die festgelegten Kriterien für Annahme oder Zurückweisung des restringierten Modells jeweils nur knapp über- oder unterschritten werden. Dieser Umstand, und auch die schlichte Betrachtung der Mittelwerte bzw. Mittelwertsdifferenzen in Tabelle 3.18 läßt zumindest die Veränderungen der Durchschnittswerte

in der Gesamtgruppe als durchaus noch moderate erscheinen, auch wenn man nicht soweit gehen möchte, in diesen Fällen von annähernd strikter Zufriedenheits- und Gesundheitsstabilität zu sprechen.

Es bleiben noch einige interessante Befunde zu erwähnen, welche allerdings für die Bewertung der Merkmalsstabilität im Allgemeinen und der hier zugrundegelegten Hypothesen im Besonderen weniger relevant sind: Ein Gruppenunterschied der True-Score-Varianzen zeigt sich auch bei der Wohn- und der Freizeit Zufriedenheit, jedoch hier in umgekehrter Richtung, wie bei Lebens- und Gesundheitszufriedenheit. Die Älteren variieren insgesamt am geringsten bei der Wohnzufriedenheit, relativ gering bei der Freizeit Zufriedenheit, und in beiden Bereichen deutlich weniger, als die Gesamtgruppe. Auch sind dies die beiden Bereiche, bei denen die Älteren die höchsten Mittelwerte aufweisen. Die Zufriedenheit mit der Freizeit und ganz besonders die Wohnzufriedenheit scheinen hier also bei den Personen der Höheraltrigen Gruppe interindividuell konstant hoch ausgeprägt. Zur Erklärung diese Befunde könnte man beispielsweise annehmen, daß es sich bei Wohnen und Freizeit um solche Lebensbereiche handelt, in denen bei der Mehrzahl der erwachsenen Menschen bis zum Übergang ins höhere Lebensalter eine jeweils individuell optimierte Anpassung (oder auch schlicht Gewöhnung) stattgefunden hat, so daß die Zustände in diesen Bereichen von den meisten Älteren positiv bewertet werden können. Solche Erklärungen aber fallen in den Themenbereich der auf diese Bereiche spezialisierten ökogerontologischen Forschung und sollen hier nicht weiter vertieft werden.

Auch fällt auf, daß bei den allgemeinen Stabilitätsanalysen dort, wo die Ergebnisse für die einzelnen Zufriedenheitsbereiche unterschiedlich ausfielen (Gruppenunterschiede von Varianzen, Gruppenunterschiede und längsschnittliche Veränderung von Mittelwerten), jeweils starke Ähnlichkeiten der Lebens- und Gesundheitszufriedenheit bestehen: Dieses deutet auf einen gewissen Zusammenhang zwischen diesen beiden Zufriedenheiten hin, welcher sich übrigens auch bei der Betrachtung der Korrelationen der fünf Zufriedenheitsmessungen in den einzelnen Erhebungsjahren, die im Verlauf der Arbeiten an dieser Untersuchung zu explorativen Zwecken berechnet wurden (aber hier nicht ausführlich dargestellt werden) zeigte.³³ Eine in allen Erhebungsjahren hohe Korrelation von Lebens- und Gesundheitszufriedenheitsmeßwerten läßt vermuten, daß die Gesundheitszufriedenheit in besonders starkem Maße zur Lebenszufriedenheit beiträgt oder, dem konstruktionistischen Ansatz in der Top-down/Bottom-Up-Kontroverse (vgl. Kapitel 1.1.4.7) entsprechend, daß

³³Zur Veranschaulichung hier nur die Mittelwerte dieser jeweils 16 Korrelationen in der westdeutschen Stichprobe:

	<i>EIZ</i>	<i>GEZ</i>	<i>WOZ</i>	<i>FRZ</i>
ALZ	0.43	0.47	0.35	0.34
EIZ		0.27	0.42	0.28
GEZ			0.20	0.19
WOZ				0.35

Deutlich erkennbar ist der hohe durchschnittliche Zusammenhang von Lebens- und Gesundheitszufriedenheitsmeßwerten.

diese beiden Zufriedenheiten durch ähnliche Inferenzen generiert werden. Dieses würde erklären, daß für die beiden Bereiche sowohl stark ähnliche Gruppenunterschiede bei den True-Score-Varianzen und Mittelwerten, als auch ähnliche Mittelwertsabfälle über die untersuchte Zeitreihe hinweg gefunden wurden. Natürlich müßten zur genaueren Analyse solcher Zusammenhänge wiederum die meßfehlerbereinigten True-Score-Korrelationen modelliert werden.

Damit sind die m.E. untersuchungsrelevanten und darüber hinaus noch möglicherweise interessanten Befunde aus den QMSM-Berechnungen mit den Untersuchungsgruppen aller und aller 1984 mindestens 60-Jährigen westdeutschen SOEP-Befragten besprochen. Die weitere theoretische Bewertung dieser Ergebnisse soll unter Einbeziehung der Befunde aus den anderen Analyseschritten zum Abschluß dieser Arbeit erfolgen.

3.3.2 Entwicklung der Stabilität von Zufriedenheit in ausgewählten Geburtsjahrgangsgruppen älterer Befragungspersonen

Wie eingangs des Kapitels 3.3 dargelegt, soll mit den nun folgenden Stabilitätsanalysen der Problematik von über ein breites Geburtsjahrgangsspektrum definierten Untersuchungsgruppen begegnet werden, welche, wie oben schon mehrfach gezeigt, die Interpretation v.a. der bei den QMSM-Berechnungen der allgemeinen Stabilitätsanalysen gefundenen Indikatoren von „Veränderung der Veränderung“ im höheren Lebensalter stark einschränkt. Diese Problematik besteht darin, daß mögliche Veränderungen der Mittelwerte, True-Score-Varianzen und/oder Stabilitäten, die im höheren Alter stattfinden könnten, dann in einer wie auch immer breit definierten Gruppe Höheraltriger nicht oder nur abgeschwächt gefunden würden, wenn sie über diesen Altersabschnitt hinweg nicht als annähernd lineare Alterseffekte wirksam wären. Wenn also z.B. Varianzen oder Stabilitäten in der früheren Phase der Höheraltrigkeit wachsen, später dann aber sich nicht mehr verändern, so würde in einer breit definierten Gruppe von Älteren die Veränderung bei den jüngeren und die Invarianz bei den älteren Mitgliedern zu einer allenfalls mäßigen Veränderung in dieser Gruppe zusammengefaßt, mit zudem noch absteigender Tendenz im längsschnittlichen Verlauf aufgrund des stetigen Anstiegs des Mindestalters und Anteils der „alten Alten“ in der Höheraltrigen Gruppe.

Es folgen darum Untersuchungen an enger definierten Geburtsjahrgangsgruppen Älterer, sie werden im folgenden der Einfachheit halber als *altersspezifische Stabilitätsanalysen* bezeichnet.

3.3.2.1 Stichproben, Design und Strategie der altersspezifischen Stabilitätsanalysen

Es liegt auf der Hand, daß die oben genannten Probleme eines breiten Altersspektrums in den untersuchten Höheraltrigen Gruppen dann nicht bestünden, wenn jeweils die Zufriedenheiten von Ein-Jahres-Alters- bzw. Geburtsjahrgangsgruppen untersucht würden,

wenn also für Gruppen jeweils gleichaltriger Älterer die Kennwerte des QMSM über die Zeitreihe der 16 jährlichen SOEP-Zufriedenheitsmessungen geschätzt würden. Es wurde aber ebenfalls bereits dargelegt, daß dieses nicht realisierbar ist, weil die so definierten Untersuchungsgruppen schlicht zu geringe Stichprobenumfänge aufweisen würden, um eine FIML-Schätzungen von verlässlichem Wert zu ermöglichen (vgl. dazu die Fallzahlen in den Zellen der Tabelle B.8 im Anhang B). Darum wurde für die folgenden Analysen sozusagen ein Mittelweg beschritten, der als Kompromiß zwischen den Erfordernissen ausreichender Stichprobenumfänge und enger definierter Geburtsjahrganggruppen geeignet erscheint, ein genaueres Bild von den Veränderungen der True-Score-Verteilungen der Zufriedenheiten, die möglicherweise innerhalb der Phase des höheren Lebensalters sich vollziehen, zu gewinnen.

Dieser Kompromiß ist aus einer einfachen Überlegung erklärbar: Auch dann, wenn Ein-Jahres-Alters- bzw. Geburtsjahrganggruppen untersucht würden, wäre ja dabei das Lebensalter bzw. der Geburtszeitpunkt aller Mitglieder einer Gruppe genaugenommen nicht identisch, sondern die Gruppen würden wiederum jeweils nur das durch ein Kalenderjahr definierte Intervall von Lebensaltern bzw. Geburtszeitpunkten umfassen. Das Jahr ist hier sozusagen die Maßeinheit der Skalierung der eigentlich kontinuierlichen Variablen Geburtszeitpunkt oder Lebensalter. Genauso bedeuten die jährlichen Wiederholungsmessungen des SOEP eine Skalierung der Variablen Untersuchungszeitpunkt in Jahreseinheiten, genaugenommen fanden die Befragungen während irgendeines spezifischeren Zeitintervalls innerhalb des jeweiligen Erhebungsjahres statt.³⁴ Bei der Festlegung von Ein-Jahres-Geburtsjahrganggruppen würden also genaugenommen nicht wirklich Gleichaltrige untersucht, sie beinhaltet hier aber die genauestmögliche Unterscheidung von Gruppen insofern, als in den SOEP-Daten keine genaueren Angaben über Geburtszeitpunkte vorliegen. Auch erscheint sie im vorliegenden Fall auch deshalb besonders geeignet, weil sie mit der im SOEP-Design einjähriger Befragungsintervalle implizierten Skalierung der Untersuchungszeitpunkte korrespondiert, so daß es bei Vergleichen zwischen aufeinanderfolgenden Meßzeitpunkten t und $t + 1$ nicht mehr zu Überlappungen des Altersspektrums einer Untersuchungsgruppe zu t und $t + 1$ kommen kann und dabei jeweils dieselbe Altersdifferenz involviert ist, wie bei querschnittlichen Vergleichen „benachbarter“ Altersgruppen. Wenn nun aber die Stichprobenumfänge der Ein-Jahres-Geburtsjahrganggruppen unzureichend sind, so bietet es sich aufgrund dieser Überlegungen an, die Einheit der Zeiteinteilung sozusagen einfach zu vergrößern: Beispielsweise könnte die Einheit einfach verdoppelt werden, so daß also Zwei-Jahres-Geburtsjahrganggruppen gebildet und korrespondierend dazu die Veränderung in Zwei-Jahres-Intervallen betrachtet würde. Die Frage dabei wäre dann, ob dadurch ausreichende Stichprobenumfänge für die Untersuchungsgruppen erreicht würden und ob die grobe Zeiteinteilung noch als ausreichend zur Erfassung der zu untersuchenden Veränderungen bewertet werden könnte.

Dem Design den folgenden altersspezifischen Stabilitätsanalysen wurde nun dement-

³⁴Genauere Informationen zum Zeitpunkt sind allerdings in den SOEP-Datensätzen enthalten.

sprechend eine Zeiteinteilung in Vier-Jahres-Abschnitte zugrundegelegt. D.h. es wurden Vier-Jahres-Geburtsjahrganggruppen gebildet und die Veränderung der Zufriedenheiten in Vier-Jahres-Intervallen modelliert. Der ursprüngliche Grund für ausgerechnet diese Einteilung ist banal: Der untersuchte Gesamtzeitraum von 16 Jahren ist in vier gleichlange Abschnitte unterteilbar und die Definition von Vier-Jahres-Geburtsjahrganggruppen ergab befriedigende Stichprobengrößen für die SEM-Auswertungen. Bezüglich der zu untersuchenden Veränderung wurde eine Modellierung gewählt, die die grundlegenden Stabilitätskennwerte wie schon bei den allgemeinen Stabilitätsanalysen auf der Basis der einjährigen Erhebungsintervalle berechnet, jedoch die Veränderung dieser Kennwerte selbst auf der Basis der Vier-Jahres-Intervalle – dieses wird im folgenden noch anschaulicher erläutert werden. Dieses Zeitraster erscheint einerseits „fein“ genug zur Untersuchung der stattfindenden Zufriedenheitsveränderungen (Ein-Jahres-Intervalle) und andererseits auch nicht zu grob zur Analyse der „Veränderung der Veränderung“, für welche aufgrund der Ergebnisse der allgemeinen Stabilitätsanalysen zu vermuten ist, daß sie auf jeden Fall nicht in einer Geschwindigkeit sich vollzieht, die die Erfassung in kürzeren als den Vier-Jahres-Intervallen notwendig macht.

Untersucht wurden fünf solche Gruppen älterer SOEP-Befragungspersonen, die m.E. das interessierende Altersspektrum, für das mögliche „Veränderung der Veränderung“ untersucht werden soll, abdecken – nämlich mit Bezug auf das Alter zum Panelstart 1984 die folgenden Gruppen: 52- bis 55-Jährige (Gruppe 1: Geburtsjahre 1929 bis 1932), 56- bis 59-Jährige (Gruppe 2: Geburtsjahre 1925 bis 1928), 60- bis 63-Jährige (Gruppe 3: Geburtsjahre 1921 bis 1924), 64- bis 67-Jährige (Gruppe 4: Geburtsjahre 1917 bis 1920) und 68- bis 71-Jährige (Gruppe 5: Geburtsjahre 1913 bis 1916).

Es wurden nicht, wie für die beiden überlappenden Untersuchungsgruppen der allgemeinen Stabilitätsanalysen, getrennte SEM-Berechnungen für jede dieser Untersuchungsgruppen durchgeführt, sondern es wurden sogenannte Multisample-Modelle berechnet. Dieses Verfahren dient der simultanen subgruppenweisen SEM-Parameterschätzung und ist einfach durch eine Erweiterung der empirischen Varianz-Kovarianz-Matrix und des empirischen Mittelwertsvektors, für die die Modellanpassung erfolgt, zu charakterisieren: Es handelt sich nicht mehr um eine einzige „globale“, über alle Mitglieder der Analytestichprobe hinweg berechnete bzw. für die Gesamtpopulation gültige empirische Varianz-Kovarianz-Matrix, sondern um mehrere für Subgruppen der Population bestimmte gruppenspezifische Varianz-Kovarianz-Matrizen und Mittelwertsvektoren und entsprechend gruppenspezifische Parametervektoren, die geschätzt werden. Die Simultanität der Parameterschätzung besteht darin, daß nicht schlicht für jede Subgruppe eine eigene, von den anderen abgetrennte Modellanpassung durchgeführt, sondern stattdessen für das Gesamtmodell über alle Gruppen hinweg eine Anpassungsfunktion als gewichtete Kombination der gruppenspezifischen Anpassungsfunktionen gebildet und iterativ optimiert wird. Da das Verfahren zum Standard der SEM-Methoden gehört, wird auf dessen ausführlichere Darstellung verzichtet und kann stattdessen auf solche z.B. bei Bollen (1989, 355ff) oder Bentler (1995, 149ff)

verwiesen werden. Der für die folgenden Analysen entscheidende Vorteil, den die Vorgehensweise z.B. gegenüber der nicht simultanen Schätzung der Modelle für verschiedene Gruppen bietet, besteht in der Möglichkeit von Parameterrestriktionen zwischen den Gruppen: Beispielsweise können Modellversionen spezifiziert werden, bei denen die Gleichheit bestimmter Pfadkoeffizienten in allen (oder nur bestimmten) Subgruppen festgelegt ist, und gegen ein diesbezüglich unrestringiertes Modell, in dem diese Koeffizienten zwischen den Gruppen frei variieren, getestet werden.

Unterschiedlich restringierte Versionen des QMSM wurden als solchermaßen für die obengenannten fünf Untersuchungsgruppen spezifizierten Multisample-Modelle für die Zeitreihe aller 16 Zufriedenheitsmessungen berechnet, jedoch wurde diese Zeitreihe in Vier-Jahres-Intervalle dadurch unterteilt, daß die gesamte autoregressive Kette gewissermaßen in vier Subketten unterteilt wurde, innerhalb derer jeweils Stationarität zweiter Ordnung – also invariante True-Score-Varianzen und Stabilitätskoeffizienten – festgelegt wurden. Zur formalen Darstellung dieser Vorgehensweise wird zunächst einmal die Modellgleichung 2.24 (Kapitel 2.2.2.1) um einen Index j ergänzt, der die Gruppenzugehörigkeit anzeigt (also $j = 1, 2, 3, 4, 5$):

$$T_{tji} = \beta_{(t-1)j} T_{(t-1)ji} + D_{tji} \quad (t = 2, \dots, k) \quad (3.1)$$

T_{tji} ist hier also der True-Score der Observation i aus der Gruppe j zum Zeitpunkt t und dementsprechend wären die hier besonders interessierenden indirekten Modellparameter $\sigma_{T_{tj}}^2$, d.h. die True-Score-Varianzen zum Zeitpunkt t in der Gruppe j , und ρ_{tj} , d.h. die Stabilitätskoeffizienten der Gruppe j im Zeitintervall t .

Die Zeitreihe über 16 Meßzeitpunkte wird in vier Teile unterteilt, nämlich die jeweils vier Meßzeitpunkte von 1984 bis 1987, 1988 bis 1991, 1992 bis 1995 und 1996 bis 1999. In jedem dieser Intervalle wird Invarianz der True-Score-Varianzen und Stabilitätskoeffizienten festgelegt, es ist jedoch eine Veränderung dieser Parameter zwischen den Intervallen möglich. Die Modellrestriktionen zur Erzeugung dieser „Binnenstationarität“ lauten für die True-Score-Varianzen:

$$\sigma_{T1j}^2 = \sigma_{T2j}^2 = \sigma_{T3j}^2 = \sigma_{T4j}^2 = \sigma_{T1j}^{*2} \quad (3.2a)$$

$$\sigma_{T5j}^2 = \sigma_{T6j}^2 = \sigma_{T7j}^2 = \sigma_{T8j}^2 = \sigma_{T2j}^{*2} \quad (3.2b)$$

$$\sigma_{T9j}^2 = \sigma_{T10j}^2 = \sigma_{T11j}^2 = \sigma_{T12j}^2 = \sigma_{T3j}^{*2} \quad (3.2c)$$

$$\sigma_{T13j}^2 = \sigma_{T14j}^2 = \sigma_{T15j}^2 = \sigma_{T16j}^2 = \sigma_{T4j}^{*2} \quad (3.2d)$$

Sowie für die Autoregressionskoeffizienten:

$$\beta_{1j} = \beta_{2j} = \beta_{3j} = \beta_{1j}^* \quad (3.3a)$$

$$\beta_{5j} = \beta_{6j} = \beta_{7j} = \beta_{2j}^* \quad (3.3b)$$

$$\beta_{9j} = \beta_{10j} = \beta_{11j} = \beta_{3j}^* \quad (3.3c)$$

$$\beta_{13j} = \beta_{14j} = \beta_{15j} = \beta_{4j}^* \quad (3.3d)$$

Zusammen mit den Restriktionen 3.2a bis 3.2d bedingen die Restriktionen 3.3a bis 3.3d auch die jeweilige Gleichheit der entsprechenden standardisierten β -Werte, also der entsprechenden Stabilitätskoeffizienten ρ_{tj} , wie bereits in Kapitel 2.2.2.1 dargelegt wurde. Unrestringiert bleiben dagegen die Koeffizienten β_{4j} , β_{8j} und β_{12j} – sie werden schlicht nicht weiter berücksichtigt. Diese Modellierung bewirkt, daß *zwischen* den Vier-Jahres-Intervallen die True-Score-Varianzen und Stabilitätskoeffizienten sich verändern können und somit die Veränderung dieser *innerhalb* der Vier-Jahres-Intervalle invarianten Kennwerte analysierbar wird: Es können die durch diese Gleichheitsrestriktionen gewissermaßen „gemittelten“ Ein-Jahres-Stabilitäten (hier benannt als β_{1j}^* , β_{2j}^* , β_{3j}^* , β_{4j}^*) und Varianzen (σ_{T1j}^{*2} , σ_{T2j}^{*2} , σ_{T3j}^{*2} , σ_{T4j}^{*2}) der True-Scores der vier aufeinanderfolgenden Vier-Jahres-Intervalle miteinander verglichen werden. Zudem werden die Parameter im Multisample-QMSM für jeweils jede der fünf Gruppen berechnet, so daß also auch Vergleiche der über Vier-Jahres-Intervalle aggregierten Varianzen und Stabilitäten verschiedener Vier-Jahres-Geburtskohorten oder Vier-Jahres-Altersgruppen möglich sind. Tabelle 3.20 zeigt das somit spezifizierte Design für die altersspezifischen Stabilitätsanalysen.

Mit den hier zu untersuchenden Veränderungen von True-Score-Varianzen und Stabilitäten über das Lebensalter sind Alterseffekte gemeint. Dazu sei zunächst noch einmal an die Überlegungen des Kapitels 2.1 erinnert: Aufgrund der Abhängigkeit der Variablen Lebensalter, Geburtsjahr und Meßzeitpunkt ist eine strikte Isolierung von Effekten dieser drei Variablen nicht möglich und es sind prinzipiell, erst recht unter Einbeziehung möglicher Interaktionen, eine Vielzahl möglicher Überlagerungsmuster möglich, die es praktisch unmöglich machen, aus irgendeinem empirischen Befund logisch genau auf das Ausmaß dieser Effekte zurückzuschließen. Deshalb steht auch hier, wie genaugenommen bereits im vorherigen Kapitel, die Untersuchung vor dem prinzipiellen Problem der Trennung dieser Alterseffekte von den anderen zeitvariablen Effekten. Bei längsschnittlichen Vergleichen innerhalb von Geburtskohortengruppen, d.h. bei horizontalen Vergleichen entlang der Zeilen in Tabelle 3.20, ist der Kohorteneffekt ausgeschlossen, jedoch nicht der Periodeneffekt: Eine vorgefundene längsschnittliche Veränderung kann dem Periodeneffekt geschuldet oder aber ein Alterseffekt von einem Periodeneffekt solchermaßen überlagert sein, daß z.B. gar keine längsschnittliche Veränderung mehr erkennbar wird, oder gar eine zum eigentlichen Alterseffekt gegenläufige. Ein vom Periodeneffekt unbefleckter Alterseffekt wäre allenfalls bei vertikalen Vergleichen entlang der Spalten in Tabelle 3.20 zu beobachten, da innerhalb der Spalten der Meßzeitpunkt konstant gehalten ist – allerdings ist dann wieder die mögliche Überlagerung von Alter- und Kohorteneffekten zu berücksichtigen. Auch die drit-

TABELLE 3.20:
Design der altersspezifischen Stabilitätsanalysen

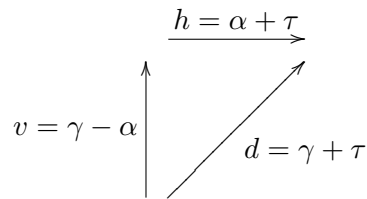
	1984-1987	1988-1991	1992-1995	1996-1999
Gruppe 1: 52-55	$\beta_{t1} = \beta_{11}^*$ $\sigma_{Tt1}^2 = \sigma_{T11}^{*2}$	$\beta_{t1} = \beta_{21}^*$ $\sigma_{Tt1}^2 = \sigma_{T21}^{*2}$	$\beta_{t1} = \beta_{31}^*$ $\sigma_{Tt1}^2 = \sigma_{T31}^{*2}$	$\beta_{t1} = \beta_{41}^*$ $\sigma_{Tt1}^2 = \sigma_{T41}^{*2}$
Gruppe 2: 56-59	$\beta_{t2} = \beta_{12}^*$ $\sigma_{Tt2}^2 = \sigma_{T12}^{*2}$	$\beta_{t2} = \beta_{22}^*$ $\sigma_{Tt2}^2 = \sigma_{T22}^{*2}$	$\beta_{t2} = \beta_{32}^*$ $\sigma_{Tt2}^2 = \sigma_{T32}^{*2}$	$\beta_{t2} = \beta_{42}^*$ $\sigma_{Tt2}^2 = \sigma_{T42}^{*2}$
Gruppe 3: 60-63	$\beta_{t3} = \beta_{13}^*$ $\sigma_{Tt3}^2 = \sigma_{T13}^{*2}$	$\beta_{t3} = \beta_{23}^*$ $\sigma_{Tt3}^2 = \sigma_{T23}^{*2}$	$\beta_{t3} = \beta_{33}^*$ $\sigma_{Tt3}^2 = \sigma_{T33}^{*2}$	$\beta_{t3} = \beta_{43}^*$ $\sigma_{Tt3}^2 = \sigma_{T43}^{*2}$
Gruppe 4: 64-67	$\beta_{t4} = \beta_{14}^*$ $\sigma_{Tt4}^2 = \sigma_{T14}^{*2}$	$\beta_{t4} = \beta_{24}^*$ $\sigma_{Tt4}^2 = \sigma_{T24}^{*2}$	$\beta_{t4} = \beta_{34}^*$ $\sigma_{Tt4}^2 = \sigma_{T34}^{*2}$	$\beta_{t4} = \beta_{44}^*$ $\sigma_{Tt4}^2 = \sigma_{T44}^{*2}$
Gruppe 5: 68-71	$\beta_{t5} = \beta_{15}^*$ $\sigma_{Tt5}^2 = \sigma_{T15}^{*2}$	$\beta_{t5} = \beta_{25}^*$ $\sigma_{Tt5}^2 = \sigma_{T25}^{*2}$	$\beta_{t5} = \beta_{35}^*$ $\sigma_{Tt5}^2 = \sigma_{T35}^{*2}$	$\beta_{t5} = \beta_{45}^*$ $\sigma_{Tt5}^2 = \sigma_{T45}^{*2}$

te mögliche Richtung des Vergleichs, nämlich die entlang der gegendiagonalen Linien (von links unten nach rechts oben) der Tabelle 3.20, bringt keinen isolierten Effekt zutage, hier können Kohorten- und Periodeneffekt überlagert sein.

Man sieht sich hier also wiederum mit dem „muddle“ konfrontiert, das nach dem Diktum von Donaldson & Horn (1992) durch die Überlagerung von Alters-, Kohorten- und Periodeneffekten entsteht. Allerdings geht es an dieser Stelle der Untersuchung auch nicht um genauere Abschätzungen dieser drei Effekte, sondern lediglich um einen von den dreien: Bezüglich der Veränderung von True-Score-Varianzen und/oder Stabilitäten wurde keine „inhaltliche“ Hypothese eines Kohorten- oder Periodeneffekts formuliert und sollten solche Effekte denn vorhanden sein, so interessieren sie hier allenfalls im Ausblick auf weitere Theorieentwicklungen zum Thema der Zufriedenheit im höheren Lebensalter. Sie sind nur insofern relevant, weil ihre prinzipielle Möglichkeit eine Störhypothese für alle Entdeckungen eines Alterseffekts darstellt, gegen die möglicherweise gefundene längsschnittliche Effekte abzusichern wären.

Dieses eingeschränkte Interesse erleichtert die Problematik der Isolation dieser Effekte im Design der Tabelle 3.20: Man kann das Problem nun sozusagen ausschließlich unter dem Gesichtspunkt der möglichen Interpretation von gefundenen Effekten (z.B. „zeilenweise“ ansteigende Stabilitätskoeffizienten β_{tj}^* oder True-Score-Varianzen σ_{Ttj}^{*2}) als *Alterseffekte* betrachten. Zur weiteren Vereinfachung sei von allen möglichen Interaktionseffekten von Alter, Kohorte und Periode abgesehen – irgendwelche Vorannahmen müssen, wie bereits in Kapitel 2.1 dargelegt wurde, nun einmal gemacht werden. Man betrachte nun nur ein beliebiges Teilquadrat aus dem in Tabelle 3.20 abgebildeten Design, also einen Ausschnitt

aus zwei aufeinanderfolgenden Vier-Jahres-Meßzeitpunkt- und Kohortenintervallen: Wäre in diesem Quadrat ein horizontaler Effekt h zu beobachten, nämlich die Veränderung in irgendeinem Kennwert (z.B. dem Stabilitätskoeffizienten $\beta_{t,j}^*$) zwischen der rechten und der linken Spalte (wenn man Interaktionen ausschließt, so ist diese Veränderung in beiden Zeilen des Quadrats dieselbe), so sind in h die beiden möglichen Effekte des Alters α und des Meßzeitpunkts τ enthalten: $h = \alpha + \tau$ (horizontale Fortschaltung von Alter und Meßzeitpunkt um eine Einheit). Entsprechend sei v ein vertikaler Effekt der Veränderung zwischen der unteren und der oberen Spalte und γ der Kohorteneffekt des Geburtszeitpunktes, so daß $v = \gamma - \alpha$ (vertikale Fortschaltung der Geburtszeit und Rückschaltung des Alters um eine Einheit von unten nach oben). Bezeichnet man den Effekt auf der Gegendiagonalen zwischen der rechten unteren und linken oberen Zelle als d , dann gilt $d = \gamma + \tau$ (gegen diagonale Fortschaltung von Geburts- und Meßzeitpunkt um eine Einheit). Zur graphischen Veranschaulichung:



Fatal für jegliche Analyse der Effekte wäre eine Extremsituation völliger wechselseitiger Neutralisierung von Alters-, Kohorten- und Periodeneffekten: Wenn $\alpha = -\tau$ und $\alpha = \gamma$, dann ist $h = v = d = 0$, d.h. es gäbe keine Unterschiede zwischen den Zellen und es wäre keinerlei Alters-, Kohorten- oder Periodeneffekt zu entdecken (bzw. wäre diese Entdeckung nur dann möglich, wenn der beobachtete Zeitausschnitt über eine Veränderung des Periodeneffekts hinaus verlängert wird). Diese Situation muß schlicht von vornherein ausgeschlossen werden und es gibt m.E. auch gute „inhaltliche“ Gründe für diesen Ausschluß, da es sich bei dieser Situation doch eher um eine *sehr* abstrakte Spitzfindigkeit handelt, als um eine Möglichkeit, für deren theoretische Annahme sich plausible Gründe finden ließen.

Schließt man also diese Möglichkeit völliger Neutralisierung der Effekte aus, so könnte ein nicht vorhandener horizontaler Effekt $h = 0$ zweierlei bedeuten: Kein Alters- und kein Periodeneffekt, d.h. $\alpha = \tau = 0$, oder sich wechselseitig neutralisierende Alters- und Periodeneffekte, d.h. $\alpha = -\tau$. Im letzteren Falle aber könnten, wenn die Situation völliger wechselseitiger Neutralisierung von α , γ und τ ausgeschlossen ist, wenn also nicht auch noch gilt $\alpha = \gamma$ (so daß dann auch $\gamma = -\tau$), der vertikale und der gegen diagonale Effekt nicht Null sein und es würde gelten: $v = \gamma - \alpha = \gamma + \tau = d$. Wenn aber $\alpha = \tau = 0$, dann folgt $d = v = \gamma$, d.h. der vertikale und gegen diagonale Effekt könnten in diesem Falle genau dann Null werden, wenn zusätzlich auch $\gamma = 0$. Es wurde somit immerhin gezeigt: Horizontaler, vertikaler und gegen diagonaler Effekt können (bei Ausschluß vollständiger wechselseitiger Neutralisierung) nur dann gleichzeitig nicht vorhanden sein, wenn gilt $\alpha = \gamma = \tau = 0$. Wenn jedoch nur der horizontale Effekt $h = 0$ ist, jedoch

$v \neq 0$ und $d \neq 0$, dann folgt aus dem soeben gezeigten, daß der Kohorteneffekt $\gamma \neq 0$ sein *muß*. Ebenso läßt sich leicht zeigen, daß wenn $v = 0$ und $h \neq 0$, $d \neq 0$, dann $\tau \neq 0$, und wenn $d = 0$ und $h \neq 0$, $v \neq 0$, dann $\alpha \neq 0$. Mit anderen Worten: Es können horizontaler, vertikaler und gegendiagonaler Designeffekt nur entweder alle drei zusammen Null sein, oder jeweils nur einer von den dreien – in letzterem Fall ist dann immer derjenige der zeitvariablen Effekte des Alters, der Kohorte oder der Periode, der in dem nicht vorhandenen Designeffekt *nicht* enthalten ist, *nicht* Null. Wenn also z.B. kein Gegendiagonaleffekt voranden wäre ($d = 0$), jedoch ein horizontaler ($h \neq 0$) und vertikaler ($v \neq 0$), so folgt daraus die Existenz des Alterseffekts ($\alpha \neq 0$).

Aus diesen Überlegungen heraus wurde nun eine Vorgehensweise zur Prüfung der Existenz horizontaler, vertikaler oder gegendiagonaler Effekte auf die True-Score-Varianzen und Stabilitätskoeffizienten des in Tabelle 3.20 abgebildeten Designs entwickelt, die sich wiederum des Vergleichs restringierter mit weniger restringierten Modellversionen des für dieses Design spezifizierten Multisample-QMSM bedient. Die restringierteren Modellversionen sind solche, in denen horizontale, vertikale oder gegendiagonale Effekte durch die im folgenden gezeigten Gleichheitsrestriktion für die horizontalen, vertikalen und gegendiagonalen „Parameterketten“ ausgeschlossen werden. Deren Model-Fit-Kennwerte werden mit denen eines unrestringierten Modells (d.h. das lediglich durch die Restriktionen 3.2a bis 3.3d beschriebene QMSM) verglichen und die im Kapitel 2.3.2 aufgestellten und bereits bei den allgemeinen Stabilitätsanalysen angewandten Entscheidungskriterien befolgt. Die verschiedenen somit untersuchten Modellversionen sind:

- ▶ **M1:** Unrestringiertes Modell, d.h. das durch die Restriktionen 3.2a bis 3.3d spezifizierte QMSM.
- ▶ **M2:** Völlige horizontale, vertikale und gegendiagonale Invarianz. D.h. es wird die Gleichheit aller True-Score-Varianzen und aller unstandardisierten Autoregressionskoeffizienten in allen Zellen des durch Tabelle 3.20 dargestellten Designs festgelegt:

$$\sigma_{T't'j}^{*2} = \text{konst.} \quad \text{und} \quad \beta_{t'j}^* = \text{konst.} \quad (t' = 1 \dots 4)$$

Dieses Modell bedeutet also: Keine horizontalen, vertikalen oder gegendiagonalen Veränderungen von True-Score-Varianzen und Stabilitätskoeffizienten – und somit, gemäß der oben dargestellten Überlegungen auch *keine Alters-, Kohorten- und Periodeneffekte* auf diese Kennwerte.

- ▶ **M3:** Gegendiagonale Invarianz, d.h. jeweils Gleichheit aller Vier-Jahres-True-Score-Varianzen und Autoregressionskoeffizienten innerhalb aller diagonal von links unten nach rechts oben benachbarten Zellen des Designs. Die entsprechenden Restriktionen lauten:

$$\begin{aligned} \sigma_{T'12}^{*2} &= \sigma_{T'21}^{*2} \quad \text{und} \quad \beta_{12}^* = \beta_{21}^* \\ \sigma_{T'13}^{*2} &= \sigma_{T'22}^{*2} = \sigma_{T'31}^{*2} \quad \text{und} \quad \beta_{13}^* = \beta_{22}^* = \beta_{31}^* \\ \sigma_{T'14}^{*2} &= \sigma_{T'23}^{*2} = \sigma_{T'32}^{*2} = \sigma_{T'41}^{*2} \quad \text{und} \quad \beta_{14}^* = \beta_{23}^* = \beta_{32}^* = \beta_{41}^* \end{aligned}$$

$$\begin{aligned}\sigma_{T15}^{*2} &= \sigma_{T24}^{*2} = \sigma_{T33}^{*2} = \sigma_{T42}^{*2} & \text{und} & \beta_{15}^* = \beta_{24}^* = \beta_{33}^* = \beta_{42}^* \\ \sigma_{T25}^{*2} &= \sigma_{T34}^{*2} = \sigma_{T43}^{*2} & \text{und} & \beta_{25}^* = \beta_{34}^* = \beta_{43}^* \\ \sigma_{T35}^{*2} &= \sigma_{T44}^{*2} & \text{und} & \beta_{35}^* = \beta_{44}^*\end{aligned}$$

Dieses bedeutet *nicht*, daß keine Kohorten- oder Periodeneffekte vorhanden wären. Diese könnten sich auch wechselseitig neutralisiert haben. Es bedeutet jedoch – vgl. die obigen Überlegungen –, daß dann, wenn M2 (und M4 und M5) nicht gilt, ein von Null verschiedener *Alterseffekt* vorhanden ist.

- **M4:** Horizontale Invarianz, d.h. jeweils Gleichheit aller Vier-Jahres-True-Score-Varianzen und Autoregressionskoeffizienten innerhalb jeder Zeile des Designs. Die entsprechenden Restriktionen lauten:

$$\begin{aligned}\sigma_{T11}^{*2} &= \sigma_{T21}^{*2} = \sigma_{T31}^{*2} = \sigma_{T41}^{*2} & \text{und} & \beta_{11}^* = \beta_{21}^* = \beta_{31}^* = \beta_{41}^* \\ \sigma_{T12}^{*2} &= \sigma_{T22}^{*2} = \sigma_{T32}^{*2} = \sigma_{T42}^{*2} & \text{und} & \beta_{12}^* = \beta_{22}^* = \beta_{32}^* = \beta_{42}^* \\ \sigma_{T13}^{*2} &= \sigma_{T23}^{*2} = \sigma_{T33}^{*2} = \sigma_{T43}^{*2} & \text{und} & \beta_{13}^* = \beta_{23}^* = \beta_{33}^* = \beta_{43}^* \\ \sigma_{T14}^{*2} &= \sigma_{T24}^{*2} = \sigma_{T34}^{*2} = \sigma_{T44}^{*2} & \text{und} & \beta_{14}^* = \beta_{24}^* = \beta_{34}^* = \beta_{44}^* \\ \sigma_{T15}^{*2} &= \sigma_{T25}^{*2} = \sigma_{T35}^{*2} = \sigma_{T45}^{*2} & \text{und} & \beta_{15}^* = \beta_{25}^* = \beta_{35}^* = \beta_{45}^*\end{aligned}$$

Bei Ungültigkeit von M2 bedeutet dieses Modell, daß ein von Null verschiedener *Kohorteneffekt* vorhanden ist.

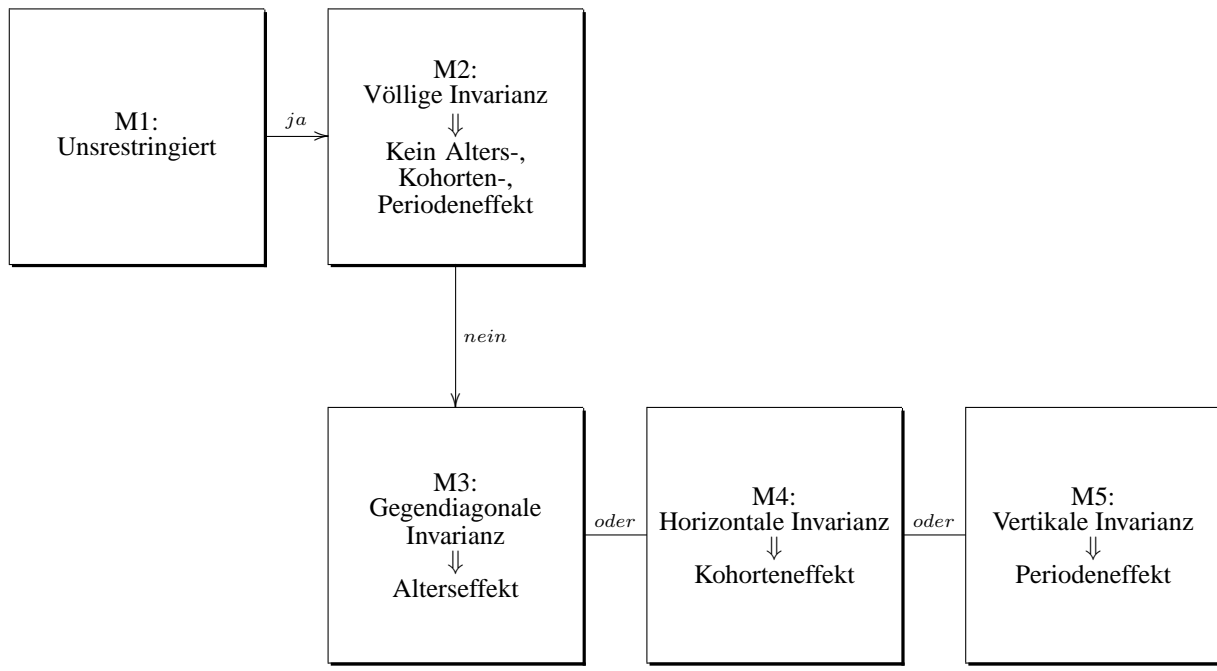
- **M5:** Vertikale Invarianz, d.h. jeweils Gleichheit aller Vier-Jahres-True-Score-Varianzen und Autoregressionskoeffizienten innerhalb jeder Spalte des Designs. Die entsprechenden Restriktionen lauten:

$$\begin{aligned}\sigma_{T11}^{*2} &= \sigma_{T12}^{*2} = \sigma_{T13}^{*2} = \sigma_{T14}^{*2} = \sigma_{T15}^{*2} & \text{und} & \beta_{11}^* = \beta_{12}^* = \beta_{13}^* = \beta_{14}^* = \beta_{15}^* \\ \sigma_{T21}^{*2} &= \sigma_{T22}^{*2} = \sigma_{T23}^{*2} = \sigma_{T24}^{*2} = \sigma_{T25}^{*2} & \text{und} & \beta_{21}^* = \beta_{22}^* = \beta_{23}^* = \beta_{24}^* = \beta_{25}^* \\ \sigma_{T31}^{*2} &= \sigma_{T32}^{*2} = \sigma_{T33}^{*2} = \sigma_{T34}^{*2} = \sigma_{T35}^{*2} & \text{und} & \beta_{31}^* = \beta_{32}^* = \beta_{33}^* = \beta_{34}^* = \beta_{35}^* \\ \sigma_{T41}^{*2} &= \sigma_{T42}^{*2} = \sigma_{T43}^{*2} = \sigma_{T44}^{*2} = \sigma_{T45}^{*2} & \text{und} & \beta_{41}^* = \beta_{42}^* = \beta_{43}^* = \beta_{44}^* = \beta_{45}^*\end{aligned}$$

Bei Ungültigkeit von M2 bedeutet dieses Modell, daß ein von Null verschiedener *Periodeneffekt* vorhanden ist.

Der Test bzw. Vergleich dieser Modellversionen kann nun dem einfachen Schema folgen, das in Abbildung 3.19 dargestellt ist: Es wird Modell M2 mit dem unrestringierten Modell M1 verglichen: Wenn die Anpassungsgüte von M2 nicht wesentlich schlechter ist, als die von M1, dann kann M2 angenommen werden und damit auch implizit die Hypothese eines wie auch immer (nicht notwendigerweise linear) wirksamen Alterseffekts – sowie auch von Kohorten- und Periodeneffekten – verworfen werden. Wenn M2 nicht angenommen wird, dann bliebe noch, die drei jeweils weniger restriktiven Modellversionen M3, M4

ABBILDUNG 3.19:
Analysestrategie der altersspezifischen Stabilitätsanalysen



und M5 mit M1 zu vergleichen: Genaugenommen ist für die eigentlich zu überprüfende Frage eines Alterseffekts auf True-Score-Varianzen und Stabilitätskoeffizienten lediglich M3 von Interesse: Gemäß der oben dargelegten Zusammenhänge würde dann, wenn nicht gar kein zeitvariabler Effekt vorhanden ist (M2), das Fehlen eines gegendiagonalen Effekts das Vorhandensein eines Alterseffekts anzeigen – im Falle der Zurückweisung von M2 wäre also die Annahme von M3 ein starker Beleg für das Vorhandensein eines Alterseffekts. Leider kann aber nicht umgekehrt von der Ablehnung von M3 auf das Nichtvorhandensein eines Alterseffekts geschlossen werden. Aber immerhin schließt M3 die hier im Sinne der untersuchungsrelevanten Hypothesen eigentlich realistischerweise vermutete Alternative zum Nulleffektmodell M2 ein, nämlich die eines Alterseffekts *ohne* Kohorten- und Periodeneffekte, und die Annahme von M3 bedeutet sozusagen die im Untersuchungsdesign strengstmögliche Bestätigung eines vorhandenen Alterseffekts. Die Modelle M4 und M5 werden dazu nur der Vollständigkeit halber mitgetestet, um des möglichen Gewinns für die weitere Theorieentwicklung willen, welcher bei Annahme einer dieser Modellversionen aus der dann starken Evidenz für einen Kohorten- oder Periodeneffekt gezogen werden könnte.

Eine weitere hinterfragbare Eigenart der beschriebenen Vorgehensweise besteht in dem Umstand, daß hier die jeweiligen Gleichheitsrestriktionen der Modellvarianten für

True-Score-Varianzen und Stabilitätskoeffizienten gleichermaßen festgesetzt und damit die möglichen Alters-, Kohorten- und Periodeneffekte auf diese beiden Parameter simultan und nicht getrennt voneinander überprüft werden. Genaugenommen müßte die Untersuchung der alterskorrelierten Veränderung von True-Score-Varianzen und Stabilitätskoeffizienten in getrennten Schritten erfolgen, in dem einmal die Modelle M1 bis M5 jeweils nur durch die genannten Gleichheitsrestriktionen für die True-Score-Varianzen und dann nur durch diejenigen für Stabilitätskoeffizienten spezifiziert und getestet würden. Der etwas „bequemere“ Weg, Alterseffekte auf beide Parameter simultan zu untersuchen, wurde hier zum einen zur Vereinfachung und Verkürzung dieses Analyseschritts gewählt, zum anderen aber v.a. deswegen, weil die getrennte Untersuchung der Effekte auf Varianzen und Stabilitäten mit enormen praktischen Schwierigkeiten verbunden gewesen wäre: Wie in Kapitel 2.2.2.1 dargestellt wurde, erfordern Gleichheitsrestriktionen für True-Score-Varianzen und Stabilitätskoeffizienten des QMSM komplexe nichtlineare Restriktionen, die nicht bei allen derzeit marktgängigen Versionen der SEM-Softwarepakete spezifizierbar sind. Ein Lösungsweg, welcher auch hier beschränkt wurde, bietet die von Rudinger & Rietz (1993; 1998) gezeigte Methode zur Restriktion invarianter Varianzen mittels von Phantomvariablen – sind auf diese Weise die sukzessiven Varianzen gleichgesetzt, so bewirkt dann die einfache Gleichheitsrestriktion für die sukzessiven unstandardisierten Autoregressionskoeffizienten auch die Invarianz der Stabilitätskoeffizienten. Für letztere ist damit aber diese Lösung an die vorherige Restriktion gleicher True-Score-Varianzen geknüpft. Deshalb würde die Festlegung horizontaler, vertikaler und gegendagonaler Invarianz der Stabilitätskoeffizienten, wenn sie ohne entsprechende Restriktionen der True-Score-Varianzen spezifiziert werden sollte, die Festlegung komplexer nichtlinearer Restriktionen erfordern. Unter den SEM-Programmversionen, die zum Zeitpunkt dieser Analysen zur Verfügung standen, erlaubte nur die SAS Prozedur CALIS die Festlegung nichtlinearer Parameterrestriktionen – allerdings ermöglicht die Prozedur keine FIML-Schätzung unter Einbeziehung der Fälle mit teilweise fehlenden Werten. Eine getrennte Überprüfung der Alterseffekte bei True-Score-Varianzen und Stabilitätskoeffizienten hätte letztlich einen Aufschub der Analysen bis zur Verfügbarkeit (und Anschaffung) neuester Softwareangebote, die sowohl FIML-Schätzung, als auch Spezifikation komplexer nichtlinearer Restriktionen erlauben, erfordert. Angesichts dieser Schwierigkeit erscheint die Vorgehensweise der simultanen Restriktion von horizontalen, vertikalen und gegendagonalen Unterschieden der True-Score-Varianzen und Stabilitätskoeffizienten begründbar. Würde diese Restriktionen zurückgewiesen, dann wäre gegebenenfalls anhand der Schätzungen des unrestringierten Modells zu einzuschätzen, ob diese Zurückweisung eher auf Unterschiede in den True-Score-Varianzen oder durch solche in den Stabilitätskoeffizienten hervorgerufen wurde.

Es werden im folgenden die Ergebnisse der Modellvergleiche und sowie Detailergebnisse für das daraus ausgewählte Modell berichtet. Auf ausführliche Deskription der Daten der einzelnen Untersuchungsgruppen, wie sie noch für die allgemeinen Stabilitätsanalysen durchgeführt worden war, wird hier verzichtet, da die daraus sich ergebende Masse sehr

TABELLE 3.21:

Altersspezifische Stabilitätsanalysen: Anpassungsgüte der QMSM für allgemeine Lebenszufriedenheit

	M1	M2	M3	M4	M5
<i>df</i>	609	647	633	639	641
χ^2	1038.4	1223.3	1110.1	1108.5	1196.7
<i>TLI</i>	0.993	0.992	0.993	0.993	0.992
<i>NFI</i>	0.986	0.983	0.985	0.985	0.983
<i>IFI</i>	0.994	0.992	0.993	0.993	0.992
<i>CFI</i>	0.994	0.992	0.993	0.993	0.992
<i>RMSEA</i>	0.017	0.019	0.018	0.018	0.019
<i>RMSEA_U</i>	0.015	0.018	0.016	0.016	0.017
<i>RMSEA_O</i>	0.019	0.021	0.020	0.019	0.021

df: Freiheitsgrade des Modells.

RMSEA_U/RMSEA_O: Untere/obere Grenze des *RMSEA*-90%-Konfidenzintervalls.

Alle χ^2 -Werte sind signifikant mit $\alpha \leq 0.001$.

spezifischer Information m.E. kaum wesentlich zur Klärung der über die allgemeinen Stabilitätsanalysen hinaus bestehenden Fragen beitragen würde. Eine Überprüfung der univariaten Normalverteilungen in den einzelnen Subgruppen ergab ähnliche Resultate, wie für die größeren Gruppen der allgemeinen Stabilitätsanalysen berichtet, so daß die Anwendbarkeit des FIML-Schätzalgorithmus hier genauso gerechtfertigt erscheint, wie dort. Die für diese Analysen zur Verfügung stehenden Stichprobengrößen sind leicht aus der Tabelle B.8 im Anhang B herauszulesen und werden deshalb ebenfalls nicht mehr eigens tabelliert.

3.3.2.2 Ergebnisse der altersspezifischen Stabilitätsanalysen

Die Model-Fit-Kennwerte der Modellversionen M1 bis M5 für die Meßwertreihe der Beurteilungen allgemeiner Lebenszufriedenheit zeigt Tabelle 3.21. Unter Zugrundelegung der in Kapitel 2.3.2 aufgestellten Regeln zum Vergleich der „nested models“ und des durch Abbildung 3.19 dargestellten Entscheidungsablaufs ergibt sich die eindeutige Entscheidung für die Modellvariante M2, d.h. für das Modell ohne alle Alters-, Kohorten- oder Periodeneffekte: Relevante Unterschiede in den Indikatoren der Modellanpassungsgüte sind zwischen allen Modellvarianten praktisch nicht festzustellen, auch wenn wiederum die $\Delta\chi^2$ -Tests signifikante Differenzen zwischen den Modellvarianten M1 und jeweils M3, M4, M5, sowie zwischen jeweils M3, M4, M5 und M2 anzeigen – welche aber, wie in Kapitel 2.3.2 begründet, nicht berücksichtigt werden sollen.

Die Ergebnisse der Modellberechnungen für die Einkommenszufriedenheit (Tabelle 3.22), Gesundheitszufriedenheit (Tabelle 3.23), Wohnzufriedenheit (Tabelle 3.24) und Frei-

TABELLE 3.22:

Altersspezifische Stabilitätsanalysen: Anpassungsgüte der QMSM für Einkommenszufriedenheit

	M1	M2	M3	M4	M5
<i>df</i>	609	647	633	639	641
χ^2	1041.9	1184.6	1083.2	1167.1	1149.7
<i>TLI</i>	0.993	0.992	0.993	0.992	0.992
<i>NFI</i>	0.985	0.983	0.984	0.983	0.983
<i>IFI</i>	0.994	0.992	0.993	0.992	0.993
<i>CFI</i>	0.994	0.992	0.993	0.992	0.993
<i>RMSEA</i>	0.017	0.019	0.017	0.019	0.018
<i>RMSEA_U</i>	0.016	0.017	0.016	0.017	0.017
<i>RMSEA_O</i>	0.019	0.020	0.019	0.020	0.020

df: Freiheitsgrade des Modells.

RMSEA_U/RMSEA_O: Untere/obere Grenze des *RMSEA*-90%-Konfidenzintervalls.

Alle χ^2 -Werte sind signifikant mit $\alpha \leq 0.001$.

zeitzufriedenheit (Tabelle 3.25) decken sich gänzlich mit denen für die Lebenszufriedenheitsmodelle: In keinem Fall zieht irgendeine der Restriktionen eine Verschlechterung der Anpassungsgütekennwerte nach sich, stets erscheinen die Unterschiede in den aufgeführten Kennwerten marginal. Es wäre somit in allen Fällen das Modell M2 auszuwählen und das bedeutet: Es wurde bei keiner der untersuchten Zufriedenheiten ein Hinweis auf bedeutsame Alters-, Kohorten- oder Periodeneffekte gefunden. Oder anders ausgedrückt: Die Annahme von wie auch immer gerichteten Alterseffekten der True-Score-Varianzen oder Stabilitäten ist nicht notwendig zur Beschreibung der Datenstruktur der bei diesen Analysen untersuchten Personen aus dem Höheraltrigkeitsspektrum.

Die Parameterschätzungen für die Modellversion M2 – die geschätzten Vier-Jahres-Stabilitätskoeffizienten $\hat{\rho}^*$ und die geschätzten Vier-Jahres-True-Score-Varianzen $\hat{\sigma}_T^{*2}$ –, sowie die daraus für die Vier-Jahres-Intervalle resultierenden Q_Δ -Werte, sind in Tabelle 3.26 angegeben. In den unteren Zeilen der Tabelle sind dort um der vollständigen Ergebnisdokumentation willen auch die Schätzungen $\hat{\rho}_t$ für die jeweiligen Stabilitätskoeffizienten zwischen den Vier-Jahres-Intervallen, welche ja auch bei M2 unrestringiert blieben, angegeben. Es gibt m.E. keine Auffälligkeiten in diesen Ergebnissen, die eine Korrektur des bei den allgemeinen Stabilitätsanalysen für die Gruppe Alt84 gefundenen Bildes nötig scheinen lassen (vgl. dort Tabelle 3.17): Insgesamt zeigen sich wieder sehr hohe Stabilitätskoeffizienten und niedrige Quotienten Q_Δ , welche zusammengenommen das Bild einer hohen näherungsweise Stabilität aller Zufriedenheiten ergeben. Die geschätzten Werte der Stabilitätskoeffizienten sind bei Testung mit dem SEM-üblichen approximativen *t*-Test allesamt statistisch signifikant mit $p \leq 0.001$, was angesichts ihrer Höhe wohl kaum überrascht. Die

TABELLE 3.23:
 Altersspezifische Stabilitätsanalysen: Anpassungsgüte der QMSM für Gesundheitszufriedenheit

	M1	M2	M3	M4	M5
<i>df</i>	609	647	633	639	641
χ^2	952.4	1060.1	990.8	1006.3	1050.4
<i>TLI</i>	0.993	0.992	0.993	0.993	0.992
<i>NFI</i>	0.983	0.981	0.982	0.982	0.981
<i>IFI</i>	0.994	0.992	0.994	0.993	0.993
<i>CFI</i>	0.994	0.992	0.993	0.993	0.993
<i>RMSEA</i>	0.015	0.016	0.015	0.016	0.016
<i>RMSEA_U</i>	0.014	0.015	0.014	0.014	0.015
<i>RMSEA_O</i>	0.017	0.018	0.017	0.017	0.018

df: Freiheitsgrade des Modells.

RMSEA_U/RMSEA_O: Untere/obere Grenze des *RMSEA*-90%-Konfidenzintervalls.

Alle χ^2 -Werte sind signifikant mit $\alpha \leq 0.001$.

TABELLE 3.24:
 Altersspezifische Stabilitätsanalysen: Anpassungsgüte der QMSM für Wohnzufriedenheit

	M1	M2	M3	M4	M5
<i>df</i>	609	647	633	639	641
χ^2	1120.3	1305.4	1198.3	1252.9	1268.5
<i>TLI</i>	0.993	0.992	0.993	0.992	0.992
<i>NFI</i>	0.986	0.984	0.986	0.985	0.985
<i>IFI</i>	0.994	0.992	0.993	0.993	0.992
<i>CFI</i>	0.994	0.992	0.993	0.993	0.992
<i>RMSEA</i>	0.019	0.021	0.019	0.020	0.020
<i>RMSEA_U</i>	0.017	0.019	0.018	0.018	0.019
<i>RMSEA_O</i>	0.021	0.022	0.021	0.022	0.022

df: Freiheitsgrade des Modells.

RMSEA_U/RMSEA_O: Untere/obere Grenze des *RMSEA*-90%-Konfidenzintervalls.

Alle χ^2 -Werte sind signifikant mit $\alpha \leq 0.001$.

TABELLE 3.25:
 Altersspezifische Stabilitätsanalysen: Anpassungsgüte der QMSM für Freizeitzufriedenheit

	M1	M2	M3	M4	M5
<i>df</i>	609	647	633	639	641
χ^2	1068.1	1202.1	1101.5	1163.2	1192.8
<i>TLI</i>	0.993	0.992	0.993	0.992	0.992
<i>NFI</i>	0.985	0.983	0.984	0.984	0.983
<i>IFI</i>	0.993	0.992	0.993	0.993	0.992
<i>CFI</i>	0.993	0.992	0.993	0.993	0.992
<i>RMSEA</i>	0.018	0.019	0.018	0.019	0.019
<i>RMSEA_U</i>	0.016	0.017	0.016	0.017	0.017
<i>RMSEA_O</i>	0.020	0.021	0.019	0.020	0.021

df: Freiheitsgrade des Modells.

RMSEA_U/RMSEA_O: Untere/obere Grenze des *RMSEA*-90%-Konfidenzintervalls.

Alle χ^2 -Werte sind signifikant mit $\alpha \leq 0.001$.

Q_{Δ} -Werte liegen hier sogar noch etwas unter und die Stabilitätskoeffizienten etwas über denen für die Gruppe Alt84 bei den allgemeinen Stabilitätsanalysen. Angesichts ihres sehr geringen Ausmaßes erscheinen mir diese Unterschiede jedoch weder interpretationswürdig, noch interpretierbar.

Bei den frei schätzbaren „Übergangsstabilitäten“ zwischen den Vier-Jahres-Intervallen zeigen sich einige Auffälligkeiten, nämlich zum einen einige Maximalwerte $\hat{\rho}_t = 1$ und einige vergleichsweise niedrigen Koeffizienten $\hat{\rho}_t < 0.9$. Erstere sind natürlich durch Rundung auf zwei Nachkommastellen zustande gekommen. Auch wenn man in diesen Fällen daran zweifeln würde, daß dieser Wert realistisch ist und in der untersuchten Population in diesem Zeitintervall tatsächlich eine absolute Stabilität von Zufriedenheiten bestand, so stellen diese Werte m.E. die Gültigkeit der Parameterschätzungen nicht grundsätzlich in Frage – auch sie sind mit einem Standardfehler behaftete Schätzungen eines Populationswertes (allerdings kann es bei der SEM-Parameterschätzung auch vorkommen, daß ein standardisierter Koeffizientenwert außerhalb des theoretisch möglichen Wertebereichs von -1 bis 1 liegt, in diesem Fall ist die Schätzung dann nicht mehr ohne weiteres verwendbar und im Grunde eine Revision der gesamten Modellspezifikation angebracht). Den sehr hohen Koeffizienten stehen auch einige eher niedrige gegenüber: Möglicherweise ist beides dadurch erklärbar, daß durch die Erzwingung einer einheitlichen Kovarianz der True-Scores innerhalb der Vier-Jahres-Intervalle eben Kovarianzen zwischen True-Scores an den Übergängen der Intervalle bisweilen gewissermaßen künstlich aufgebläht oder verringert werden. Man kann diese Überlegung *sehr* vereinfachend und ohne entsprechende formale Ableitungen so veranschaulichen: Wenn beispielsweise bei völlig freier Modellierung (wie bei M1 in den allgemeinen Stabilitätsanalysen, d.h. ohne Vier-Jahres-Intervalle usw.) die True-Scores

TABELLE 3.26:

Altersspezifische Stabilitätsanalysen (M2): True-Score-Varianz, Stabilitätskoeffizienten und Q_{Δ}

		ALZ	EIZ	GEZ	WOZ	FRZ
	$\hat{\rho}^*$	0.95	0.96	0.96	0.96	0.91
	$\hat{\sigma}_T^{*2}$	1.61	1.78	1.94	1.35	1.59
	Q_{Δ}^*	0.11	0.08	0.09	0.07	0.17
Gruppe 1	$\hat{\rho}_4$	0.96	0.88	0.98	0.82	0.87
	$\hat{\rho}_8$	0.96	0.91	1.00	0.82	0.83
	$\hat{\rho}_{12}$	0.99	0.95	0.99	1.00	0.95
Gruppe 2	$\hat{\rho}_4$	0.94	0.92	0.94	0.83	0.86
	$\hat{\rho}_8$	1.00	0.96	1.00	0.86	0.93
	$\hat{\rho}_{12}$	1.00	1.00	0.99	1.00	1.00
Gruppe 3	$\hat{\rho}_4$	0.97	0.96	0.93	0.85	0.91
	$\hat{\rho}_8$	0.97	0.99	0.97	0.96	0.98
	$\hat{\rho}_{12}$	0.94	0.96	0.96	0.88	0.98
Gruppe 4	$\hat{\rho}_4$	0.97	0.94	0.98	0.85	0.91
	$\hat{\rho}_8$	0.92	0.94	0.95	0.97	0.88
	$\hat{\rho}_{12}$	0.91	0.93	1.00	0.88	0.85
Gruppe 5	$\hat{\rho}_4$	0.93	0.98	0.96	0.89	1.00
	$\hat{\rho}_8$	0.86	0.93	1.00	0.89	0.82
	$\hat{\rho}_{12}$	0.88	0.96	0.92	0.83	0.78

1987 mit denen 1986 und mit denen 1988 gleichermaßen hoch kovariieren würden, so könnten durch die Restriktion gleicher Kovarianzen innerhalb des Intervalls von 1984 bis 1987 z.B. True-Scores für 1987 erzwungen werden, welche höher (oder niedriger) mit denen von 1986 kovariieren, als es ohne Restriktion der Fall wäre, und die dann eventuell niedriger (oder höher) mit den True-Scores von 1988 kovariieren. Die „Übergangsstabilitäten“ sind also eventuell durch die Restriktion von Vier-Jahres-Intervallen mit invarianten Stabilitäten verzerrt und sollten auch darum nicht weiter interpretiert werden.

Wie schon bei den allgemeinen Stabilitätsanalysen könnte auch hier noch ein Blick auf die Parameterschätzungen aus der unrestringierten Modellversion M1 geworfen werden, denn immerhin wurde ja bei der Formulierung der Kriterien zur Modellauswahl nicht auf einen vorhandenen Common-Sense der SEM-Anwendergemeinde zurückgegriffen, sondern diese Kriterien in Kapitel 2.3.2 relativ frei begründet und festgelegt, so daß zur Sicherheit zumindest darauf geachtet werden sollte, ob bei den Ergebnissen des unrestringierten Modells sich auffällige Abweichungen zeigen, durch die diejenigen des ausgewählten Modells in Frage gestellt würden. Die tabellarische Auflistung der M1-Schätzungen der Vier-Jahres Stabilitäten und True-Score-Varianzen zeigt Tabelle B.14 im Anhang B. Die

besagten Auffälligkeiten scheinen dort kaum erkennbar.

Jedoch wurde in der Diskussion der Ergebnisse der allgemeinen Stabilitätsanalysen „jenseits“ der dort getroffenen Entscheidung für das stationäre – d.h. in den längsschnittlichen Varianzen und Kovarianzen invariante – Modell noch gewisse vage Anzeichen für ein gemäß der Gewinner-und-Verlierer-Hypothese zu erwartendes Wachstum der True-Score-Varianzen über das höhere Alter hinweg gefunden. Um diesem speziellen Aspekt etwas genauer nachzuspüren, werden in Abbildung 3.20 die aus der unrestringierten Modellversion M1 für die allgemeine Lebenszufriedenheit geschätzten True-Score-Varianzen in dem durch Abbildung 3.19 gezeigten 5×4 -Felder-Schema aufgelistet, zudem werden über die Gegendiagonalen hinweg die Mittelwerte dieser Varianzschätzungen berechnet und gezeigt (Fettdruck). Es handelt sich um die über die Kohorten und Meßzeitpunkte hinweg gemittelten durchschnittlichen Varianzen pro durchlaufener Altersspanne, da die Gegendiagonalen jeweils „altersgleiche“ Zellen enthalten, d.h. in den Zellen innerhalb einer Gegendiagonale die Vier-Jahres-Geburtsjahrgangsgruppen jeweils dieselbe Lebensaltersspanne durchlaufen. Die Veränderung dieser Durchschnittsvarianzen könnten somit als Ausdruck eines tatsächlichen Alterseffekts angesehen werden: Es zeigt sich tatsächlich ein tendenzieller Anstieg von oben nach unten, d.h. mit wachsendem Alter. Wie vorne bereits dargelegt wurde, läßt die durch die Gewinner-und-Verlierer-Hypothese beschriebene Dynamik intraindividuelle Veränderung im höheren Lebensalter aber nicht nur wachsende Varianzen, sondern auch sinkende Parallelität der intraindividuellen Merkmalsverläufe erwarten. Deshalb wurden in Abbildung 3.20 auch die aus M1 resultierenden Quotienten Q_{Δ} in derselben Weise aufgelistet, wie die True-Score-Varianzen: Auch hier erscheint eine allerdings nicht ganz eindeutige, ungefähre Wachstumstendenz bei den lebensaltersspezifischen Durchschnittswerten (Fettdruck), die tatsächlich über das Alter abnehmende Parallelität anzeigen würde. Die entsprechenden durchschnittlichen gegendiagonalen Mittelwerte der in Tabelle B.14 im Anhang B aufgelisteten Vier-Jahres-Stabilitäteskoeffizienten wären übrigens (von oben nach unten): 0.95, 0.93, 0.95, 0.93, 0.89 und 0.91. Hier sind die geringsten Werte am Ende der Altersspanne zu finden, es könnte somit auch eine Tendenz zu im sehr hohen Alter sinkender Stabilität herausgelesen werden – welche sich allerdings nur sehr schwach zeigt.

Insgesamt wird über die Interpretierbarkeit der in Abbildung 3.20 gezeigten Tendenzen noch in der folgenden Diskussion nachzudenken sein. Da es sich dabei in jedem Fall gewissermaßen nur um ein Anhängsel zu den Modellvergleichen handelt, der auf Alterseffekte unterhalb eines Fit-relevanten Ausmaßes zielt, welche eventuell noch zur Bewertung der Relevanz von Gewinner-und-Verlierer-Hypothese und der Hypothese erhöhter Stabilität beitragen könnten, erscheint hier die Betrachtung der Lebenszufriedenheit ausreichend. Es wird also auf die entsprechenden Darstellungen für die untersuchten Bereichszufriedenheiten verzichtet – sie können bei Bedarf aus den Werten der im Anhang B aufgeführten Tabelle B.14 erstellt werden.

Auf die Auflistung der bei den altersspezifischen Stabilitätsanalysen resultierenden Mittelwertsschätzungen (subgruppenspezifische Intercepts der Meßmodellgleichungen) und

ABBILDUNG 3.20:

Altersspezifische Stabilitätsanalysen (M1), allgemeine Lebenszufriedenheit: Geschätzte True-Score-Varianzen und Quotienten Q_{Δ}

True-Score-Varianzen in M1:					Quotienten Q_{Δ} in M1:						
				1.51					0.09		
				1.51					0.14		
				1.62					0.09		
Gruppe 1:	1.63	1.53	1.52	1.51	1.66	Gruppe 1:	0.14	0.06	0.06	0.06	0.14
Gruppe 2:	1.50	1.53	1.55	1.66	1.86	Gruppe 2:	0.12	0.10	0.06	0.06	0.22
Gruppe 3:	1.48	1.57	1.65	1.78	1.81	Gruppe 3:	0.24	0.10	0.10	0.20	0.17
Gruppe 4:	1.85	1.79	1.96	2.04		Gruppe 4:	0.14	0.08	0.16	0.06	
Gruppe 5:	1.68	1.85	1.58	1.82		Gruppe 5:	0.32	0.30	0.28	0.18	

Fettdruck: Mittelwert der Gegendiagonale, d.h. pro Lebensaltersintervall.

Reliabilitäten (subgruppenspezifische R^2 -Werte der Meßmodellgleichungen) wird ebenfalls verzichtet, da diese Werte zur Untersuchung der theoretischen Fragestellungen, um derentwillen diese Analysen durchgeführt wurden, irrelevant sind und auch die mit der großen Zahl solcher gruppenspezifischer Werte verbundenen Überschüsse an Information kaum tiefgreifende Einsichten für explorative Erkundungen versprechen. Es wurden in den vorangegangenen Abschnitten dieser Untersuchung bereits vielfältige mittelwertsbezogene Analysen durchgeführt, die Zellmittelwerte des den altersspezifischen Stabilitätsanalysen zugrundegelegten Designs versprechen hier keine zusätzlichen Aufschlüsse. Auch scheint es ausreichend, die Reliabilitätsfrage anhand der breiteren Stichproben der allgemeinen Stabilitätsanalysen zu behandeln und nicht differentielle Reliabilitäten in enger gefaßten Geburtskohortengruppen zu untersuchen – jedenfalls ist z.B. die Frage von Alters- oder Kohorteneffekten auf die Reliabilität der Zufriedenheitsratings nicht das Thema dieser Untersuchung.

3.3.2.3 Zusammenfassung und Diskussion der altersspezifischen Stabilitätsanalysen

Die altersspezifischen Stabilitätsanalysen wurden durchgeführt, um die Ungenauigkeiten, welche bei den allgemeinen Stabilitätsanalysen aus der breiten Definition der Untersuchungsgruppe für die Erfassung längsschnittlicher Veränderung *in den Kennwerten der Stabilitätsanalysen* resultierten, abzumildern: Letztlich sollte untersucht werden, ob und in welchem Ausmaße im Bereich der Höheraltrigkeit alterskorrelierte Entwicklungen von True-Score-Varianzen und Stabilitätskoeffizienten vorliegen, die durch die Aggregation

eines breiten Geburtsjahrgangsspektrums in einer Höheraltrigengruppe eventuell dann verdeckt würden, wenn sie nicht über das gesamte höhere Lebensalter hinweg linear verliefen. Es wurde deshalb ein Modell spezifiziert, das die prinzipielle Unterscheidung der längsschnittlichen Entwicklung in enger definierten Geburtsjahrganggruppen und Überprüfung der innerhalb derselben – möglicherweise auch nicht strikt linear – stattfindenden Parameterveränderungen erlaubt.

Das herausragende Ergebnis dieser zusätzlichen Stabilitätsanalysen besteht darin, daß dabei diejenigen Alterseffekte, um deren willen der Aufwand betrieben wurde, insgesamt nicht in nennenswerter Stärke vorgefunden wurden. Für alle untersuchten Zufriedenheiten führte die restriktive Modellvariante M2 gegenüber der am wenigsten restriktiven kaum zu einer (in den aufgeführten Model-Fit-Kennwerten) erkennbaren Verschlechterung der Modellanpassungsgüte. Also „paßt“ dieses Modell, das durch invariante True-Score-Varianzen über den gesamten Meßzeitraum hinweg und invariante Stabilitätskoeffizienten innerhalb vier aufeinanderfolgender Vier-Jahres-Intervalle in allen einbezogenen Geburtsjahrganggruppen gekennzeichnet ist, genauso gut zu den Daten, wie ein Modell, das keine diesbezüglichen Beschränkungen enthält und zeitvariable Effekte des Alters, Geburtsjahrgangs oder Meßzeitpunkts zuläßt. Aus der Hinzufügung zu den Ergebnissen der allgemeinen Stabilitätsanalysen kann daraus die klare Schlußfolgerung abgeleitet werden: Die Daten zeigen weder deutliche Hinweise auf eine im höheren Lebensalter wachsende Stabilität, noch auf wachsende True-Score-Varianzen der untersuchten Zufriedenheiten.

Somit steht am Ende eines aufwendigen – und hier auch entsprechend aufwendig beschriebenen – Datenanalyseschrittes ein sehr kurzes Fazit und m.E. reflektiert diese Knappheit nicht einen Mangel an berichtenswerten und diskussionswürdigen Resultaten, sondern die erfreuliche Eindeutigkeit der Ergebnisse, die oben dargestellt wurden: Weder fiel in irgendeinem Falle die Entscheidung nach den spezifizierten Kriterien nur sehr knapp zugunsten der Modellversion M2, noch zeigten sich für die verschiedenen Zufriedenheiten diesbezügliche Unterschiede.

Die Annahme der Modellversion M2 ist von hoher Relevanz für die Bewertung derjenigen beiden Hypothesen, durch die Alterseffekte auf die Merkmale intraindividuelle Veränderung der Zufriedenheit vorhergesagt wurden. Das durch die Gewinner-und-Verlierer-Hypothese postulierte Wachstum der Merkmalsvarianz über das höhere Alter – bzw. zumindest über die Phase des Übergangs in die und Durchlaufens der frühen Höheraltrigkeit – hinweg und auch die wachsende „Disparallelität“ der intraindividuellen Veränderungen infolge der Diversifizierung von Zufriedenheitsgewinnern und Zufriedenheitsverlierern ist durch M2 ausgeschlossen, ebenso eine über diese Lebensphase hinweg allmählich wachsende Zufriedenheitsstabilität infolge z.B. erhöhter adaptiver Kompetenz oder größerer Retrospektivität der Urteilsperspektive. Es erscheint schlicht so, daß die Zufriedenheiten während des gesamten Erwachsenenalters sehr stabil sind und es unverändert bis ins höchste Alter hinein auch bleiben, und es scheint mit der Höheraltrigkeit keine altersspezifische Entwicklungsdynamik verbunden zu sein.

Nun bedeutet allerdings die Annahme des höher restringierten SEM nicht zwangsläufig, daß die dort ausgeschlossenen Effekte nicht auf einem niedrigeren Niveau vorhanden wären, es bedeutet lediglich, daß sie nicht so stark ausgeprägt sind, daß man zur „Erklärung der Daten“ nicht auch ganz gut ohne sie auskäme. Es erscheint darum durchaus sinnvoll, noch einmal sozusagen mit geschärftem Blick auf die Parameterschätzungen des unrestringierten Modells M1 zu schauen, um dort auch nach schwachen Anzeichen solcher Effekte zu suchen. Dieses wurde hier nur für das unrestringierte Modell der in erster Linie relevanten allgemeinen Lebenszufriedenheit getan und diese Detailbetrachtung wurde durch die Abbildung 3.20 zusammengefaßt und dargestellt: Es hat den Anschein, als bestünden zumindest schwache Effekte, die in die durch die Gewinner-und-Verlierer-Hypothese vorgegebene Richtung verweisen – nämlich eben doch eine tendenzielle Zunahme von True-Score-Varianzen über das Lebensalter, welche darin sich zeigt, daß die gemittelten Varianzen der untersuchten Vier-Jahres-Geburtsjahrgangskohorten während des Durchlaufens des jeweils gleichen Altersspektrums mit der Fortschaltung des Alters leicht ansteigen. Ebenso war auch ein, allerdings weniger eindeutiger, schwacher Trend zum Abfall der Parallelität der intraindividuellen Merkmalsveränderung zu erkennen, nämlich im Anstieg der Durchschnittswerte der für die untersuchten Vier-Lebensjahres-Alterspektren spezifischen Quotienten Q_{Δ} .

Insgesamt aber steht die Bewertung dieser in Abbildung 3.20 gezeigten Durchschnittswerte auf recht unsicherem Grund, da es ihnen sozusagen an inferenzstatistischem Gehalt mangelt. Es sind keine auch nur näherungsweise Verteilungseigenschaften dieser Mittelwerte bekannt, so daß beispielsweise kaum abzuschätzen ist, wie sehr der mit dem sehr hohen Lebensalter einhergehende Stichprobenschwund die Interpretierbarkeit der Durchschnittswerte für die späten Lebensalterbereiche mindert. Hier könnten z.B. Signifikanztests zum Vergleich der sukzessive aufeinanderfolgenden Varianzmittelwerte hilfreich sein. Eine indirekte und sozusagen teilweise Signifikanztestung für die Varianzen hat allerdings stattgefunden, nämlich durch den $\Delta\chi^2$ -Test beim Vergleich der Modellversionen M1 und M2, welcher wie erwähnt die statistische Signifikanz ($p \leq 0.001$) des Diskrepanzzuwachses anzeigt, der u.a. auch aus der Restriktion gleicher Varianzen in allen Zellen des Designs resultiert. Es wurde aber auch bereits dargelegt, daß diese Signifikanz im Grunde nur von geringer Relevanz ist, was sich ja auch daran zeigt, daß das Modell ohne Unterschiede in den True-Score-Varianzen praktisch dieselbe Anpassungsgüte erzielt, wie das unrestringierte. Die tatsächliche Bedeutsamkeit der in Abbildung 3.20 aufgezeigten Tendenzen hinsichtlich der Frage nach der Bestätigung der Gewinner-und-Verlierer-Hypothese bleibt also unklar und es kann lediglich festgehalten werden, daß darin schwache Hinweise auf die durch diese Hypothese prognostizierten Effekte sichtbar werden.

Insgesamt bestätigen also die Ergebnisse der altersspezifischen Stabilitätsanalysen das Bild, das aus den allgemeinen Stabilitätsanalysen gewonnen wurde und liegt die nicht gering zu schätzende Bedeutung dieser Ergebnisse gewissermaßen darin, daß sie diesem Bild nichts weiter hinzufügen: Lebenszufriedenheit, und Zufriedenheiten überhaupt, erscheinen

insgesamt hoch stabil, die im Kapitel 1.2.3.3 begründete allgemeine Stabilitätshypothese erfährt durch die hier dargestellten Datenanalysen deutlichste Bestätigung. Keine Bestätigung erfahren die Hypothese erhöhter Stabilität und die Gewinner-und-Verlierer-Hypothese – allenfalls schwache Hinweise sprechen dafür, daß in sehr schwacher Ausprägung eine Entwicklungsdynamik wirksam sein könnte, wie sie durch letztere beschrieben wurde.

4 Zusammenfassung und abschließende Bewertung

Der „gerontologische Blick“ auf das SWB ist in starkem Maße durch dasjenige empirische Phänomen beeinflusst, welches als Wohlbefindens- oder Zufriedenheitsparadox des höheren Lebensalters bezeichnet wurde: In zumeist querschnittlichen Untersuchungen wurde kaum ein mit dem Alter einhergehender Abfall – und bisweilen sogar eher ein Anstieg – von Meßwerten der Lebenszufriedenheit gefunden, was angesichts von in dieser Lebensaltersphase typischen Verschlechterungen der objektiven Lebenssituation der Personen, z.B. infolge körperlicher oder sozialer Einbußen, nicht erwartet wurde. Die Ansätze zur Erklärung dieses sogenannten Paradoxes, so könnte man etwas überspitzt formulieren, scheinen oft darunter zu leiden, daß eine einheitlich zugrundegelegte theoretische Konzeption des Konstruktes SWB – daß *die* Theorie des SWB – fehlt und beispielsweise die Begriffe SWB und Lebenszufriedenheit oft synonym gebraucht werden oder nicht explizit zwischen affektiven und kognitiven Komponenten dessen, was alles durch den Begriff SWB beschrieben werden kann, unterschieden wird.

Es wurde darum im Kapitel 1.1 zunächst versucht, einen aktuellen „Stand der Dinge“ in der theoretischen Konzeptualisierung des SWB zu erschließen bzw. jene „Theorielinien“ zumindest kurz zu umreißen, welche dann für die gerontologische Diskussion des Zufriedenheitsparadoxes von besonderer Relevanz sein könnten. Im Kapitel 1.2 wurden dann mögliche Entwicklungen des SWB bzw. dessen kognitiver Komponente Zufriedenheit beim Übergang in die und Durchlaufen der Phase des höheren Lebensalters erörtert, welche auch zur Erklärung der paradoxen Stabilität von Altersgruppenmittelwerten im querschnittlichen Vergleich beitragen können. Dabei zeigte die Betrachtung diesbezüglicher wissenschaftlicher Veröffentlichungen eine gewisse Tendenz, diese Mittelwertstabilität mit tatsächlicher Stabilität des Merkmals Zufriedenheit gleichzusetzen: Es scheint so, als sei diese Vermutung hoher Merkmalsstabilität der Lebenszufriedenheit im höheren Lebensalter sozusagen derjenige Eindruck aus den Befunden des Zufriedenheitsparadoxes, der sich dem besagten gerontologischen Blick hauptsächlich eingeprägt hat. Jedoch zeigt die Invarianz von Zufriedenheitsmittelwerten in querschnittlichen Vergleichen nicht per se die intraindividuelle Invarianz von Zufriedenheit im längsschnittlichen Verlauf an und kann deshalb aufgrund der Befunde ersterer nicht so ohne Weiteres die hohe Merkmalsstabilität von Zufriedenheit im höheren Lebensalter als empirisch belegte Tatsache angenommen werden, wie dies bei manchen diesbezüglichen gerontologischen Veröffentlichungen der Fall scheint (vgl. z.B. Staudinger, 2000). Vielmehr erscheint diese intraindividuelle Stabilität fraglich und kann nur aus längsschnittlichen Daten empirisch belegt werden. Bei den theoretischen Erörterungen im Kapitel 1.2 wurden somit auch Vermutungen bezüglich möglicher Entwicklungsdynamiken der Zufriedenheit im höheren Alter gefunden, die nicht die intraindividuelle Stabilität des Merkmals beinhalteten und doch das Zufriedenheitsparadox erklären könnten. Es wurden

so schließlich vier Hypothesen formuliert, die als Kohorteneffekthypothese, Gewinner-und-Verlierer-Hypothese, Hypothese allgemeiner Stabilität und Hypothese erhöhter Stabilität benannt und der hier vorliegenden Untersuchung zugrundegelegt wurden.

Die Kohorteneffekthypothese könnte aus der Möglichkeit abgeleitet werden, daß zwar im höheren Lebensalter typischerweise alterskorrelierte *intraindividuelle* Zufriedenheitsverluste stattfinden, jedoch auch ein negativer Effekt des Geburtsjahres besteht, gewissermaßen infolge eines mit der in problematischen historischen und sozialen Umständen durchlebten Zeit sinkenden Anspruchsniveaus, welches positive Zufriedenheitsurteile erleichtert. In querschnittlichen Vergleichen könnte dieser negative Kohorteneffekt den negativen Alterseffekt solchermaßen neutralisieren, daß keine Mittelwertsdifferenzen zwischen fortlaufenden Altersgruppen sichtbar werden.

Aus der Möglichkeit, daß das Erreichen der Hochaltrigkeit als solches einen positiven Effekt auf Zufriedenheit bewirken könnte, wurde die Gewinner-und-Verlierer-Hypothese abgeleitet: Negativ gefärbte Stereotype bezüglich des höheren Alters könnten sozusagen dann, wenn erwartete Verschlechterungen ausbleiben, zu einem positivem „Überraschungseffekt“ führen und Zufriedenheit Älterer begünstigen, während auf der anderen Seite bei denen, die drastische Verluste erleben, auch die Lebenszufriedenheit abnimmt. Die Zufriedenheiten Höheraltriger wäre demnach durch eine Diversifizierung intraindividuellere Veränderung – mit vielen Zufriedenheitsgewinnern und vielen Zufriedenheitsverlierern – gekennzeichnet, mit über das Alter hinweg wachsender Varianz des Merkmals und möglicherweise infolge wechselseitiger Ausmittelung von Gewinnern und Verlierern weitgehend invarianten durchschnittlichen Zufriedenheiten.

Die Vermutung hoher intraindividuellere Stabilität des SWB ist aus den Hauptlinien der theoretischen Behandlung des SWB – v.a. aus den trait-theoretischen Konzeptionen und aus den Theorien der Adaptivität des SWB – ableitbar: Wenn Zufriedenheit gewissermaßen lebenslang intraindividuell „trait-artig“ stabil und nur in Maßen durch den Wechsel äußerer Lebensumstände veränderbar ist, dann würde dieses auch erklären, warum sie im höheren Lebensalter nicht infolge der dort gehäuften objektiven Verschlechterungen deutlich sinkt. Es ergibt sich daraus eine Hypothese allgemeiner Stabilität von Zufriedenheitsurteilen, welche das besagte Paradox erklären könnte – jedoch nicht mit diesem gleichgesetzt werden darf, da letzteres eben nur die Invarianz von Zufriedenheitsmittelwerten in querschnittlichen Vergleichen und *nicht* die intraindividuelle Invarianz von Zufriedenheit im längsschnittlichen Verlauf bezeichnet.

Möglicherweise aber wächst auch die intraindividuelle Stabilität von Zufriedenheitsurteilen erst im höheren Lebensalter, so daß gerade in dieser Phase Zufriedenheitsurteile zusehends weniger von den objektiven Lebensumständen abhängen. Theoretische Gründe für eine derartige Annahme wurden in einer möglicherweise im Alter steigenden Retrospektivität bei der Bildung von globalen Lebenszufriedenheitsurteilen, oder in einer durch gewachsene Lebenserfahrung verstärkten Kompetenz zur adaptiven Regulierung kritischer Lebensereignisse gefunden.

Bei der Erörterung dieser Möglichkeiten zeigte sich v.a. die Notwendigkeit längsschnittlicher Untersuchungen: Alle Hypothesen benennen längsschnittliche Effekte von Veränderung über das Lebensalter hinweg und bei drei dieser Hypothesen (Gewinner- und Verlierer, allgemeine und erhöhte Stabilität) ist nicht nur die Veränderung aggregierter Verteilungskennwerte (Mittelwerte der Zufriedenheit) angesprochen, sondern „innerhalb der Person“ stattfindende Veränderung. Die Untersuchung letzterer erfordert somit eine anspruchsvolle statistische Längsschnittmethodologie, die über den weitverbreiteten Vergleich von zentralen Tendenzen in Meßwiederholungsdesigns hinausreicht und die Modellierung intraindividuelle Veränderung beinhaltet. In dieser Hinsicht könnte die bei dieser Untersuchung aus den inhaltlichen Fragestellungen resultierende methodische Problematik als gewissermaßen repräsentativ für viele andere psychogerontologische Forschungsgebiete angesehen werden: Es liegt in der Natur dieser Disziplin, daß letztendlich immer wieder die Frage nach Veränderungen – und zwar insbesondere intraindividuellen Veränderungen –, die über das (höhere) Lebensalter hinweg sich vollziehen, gestellt werden muß. Neuere Entwicklungen statistischer Längsschnittmethodologie, wie sie v.a. aus der Modellierung längsschnittlicher Strukturgleichungsmodelle resultieren, sind für die psychogerontologische Forschung darum von grundsätzlicher Bedeutung, und so war es ein besonderes Anliegen dieser Studie, einen möglichst aktuellen „state of the art“ der zur Analyse intraindividuelle Veränderung bzw. Stabilität verfügbaren statistischen Modellierungen zu dokumentieren und anzuwenden.

Der Datensatz, auf den zurückgegriffen werden konnte, ist das deutsche Sozio-oekonomische Panel, welches in einer zum Panelstart bevölkerungsrepräsentativen westdeutschen Haushaltsstichprobe mit sehr hohen Fallzahlen seit 1984 mit jährlichen Wiederholungsmessungen durchgeführt wird und vollständige Zeitreihen von Ein-Item-Ratings der aktuellen Lebenszufriedenheit, sowie verschiedener Bereichszufriedenheiten (mit Einkommen, Gesundheit, Wohnen, Freizeit) enthält.

Die empirischen Datenanalysen wurden in zwei Abschnitten durchgeführt: In Kapitel 3.2 wurden Analysen von Zufriedenheitsmittelwerten dargestellt. Zunächst wurden die durchschnittlichen Zufriedenheiten von Lebensalters- und/oder Geburtsjahrganggruppen in den einzelnen SOEP-Befragungen rein deskriptiv untersucht mit dem Ziel, den Befund des Zufriedenheitsparadoxes in den hier analysierten Daten zu replizieren bzw. ganz allgemein ein differenziertes Bild des typischen Verlaufs durchschnittlicher Zufriedenheiten über die höhere Lebensaltersspanne hinweg zu gewinnen und bereits auf diesem bloß deskriptiven Analyselevel möglicherweise aufscheinende Unterschiede zwischen quer- und längsschnittlichen Mittelwertverläufen zu erkennen (Kapitel 3.2.1). Zum zweiten wurden die Untersuchungen zur Kohorteneffekthypothese im Rahmen dieser mittelwertsbezogenen Analysen durchgeführt, da ja die in dieser Hypothese thematischen Alters- und Kohorteneffekte die Entwicklung durchschnittlicher Zufriedenheiten über das Lebensalter bzw. die Geburtsjahrgänge hinweg betreffen (Kapitel 3.2.2). Im weiteren wurden dann die im Kapitel 3.3 dargestellten Stabilitätsanalysen durchgeführt: Diese sind Analysen auf der Basis der

Modellierung intraindividuelle Veränderung von Merkmalsausprägungen, durch die die Stabilitäten der untersuchten Zufriedenheiten allgemein, d.h. ohne Vorabbeschränkung auf ein spezielles Prozeßmodell, analysiert und im Speziellen die aus Gewinner-und-Verlierer-Hypothese, sowie den Hypothesen allgemeiner und erhöhter Stabilität ableitbaren Veränderungscharakteristika überprüft werden sollten. Diese Stabilitätsanalysen erfolgten wiederum in zwei Schritten, nämlich zuerst als Analyse der längsschnittlichen Merkmalsverläufe in über breite Geburtsjahrgangsspektren hinweg definierten Untersuchungsgruppen (Kapitel 3.3.1) und ergänzend, zur Aufklärung möglicher Alterseffekte der Parameter, die die intraindividuelle Veränderung beschreiben, in über engere Geburtsjahresintervalle definierten Subgruppen (Kapitel 3.3.2).

Eine diese Analysen vorbereitende – und darüber hinaus dem obengenannten Untersuchungsziel der Dokumentation eines aktuellen „Standes der Dinge“ in den diesbezüglichen Modellierungsmöglichkeiten verpflichtete – Darstellung und Erörterung der statistischen Methodologie wurde in Kapitel 2 unternommen.

4.1 Zusammenfassung der Ergebnisse

Die Ergebnisse der Analysen wurden bereits an Ort und Stelle in den jeweils abschließenden Unterkapiteln 3.2.1.3, 3.2.2.2, 3.3.1.4 und 3.3.2.3 diskutiert, so daß im folgenden nur noch die Kernpunkte dieser Diskussionen aufzulisten bleiben, ehe diese Untersuchung mit abschließenden und ausblickenden Bemerkungen abgeschlossen werden kann.

4.1.1 Deskriptiven Analysen von Zufriedenheitsmittelwerten

Die vielfältigen empirischen Befunde des Zufriedenheitsparadoxes, die eigentlich die hier vorgelegte Untersuchung initiiert haben, konnten für die vorliegenden Meßwerte allgemeiner Lebenszufriedenheit repliziert werden: Bei der querschnittlichen Betrachtung des Zusammenhangs von Lebensalter und Lebenszufriedenheit in den einzelnen SOEP-Erhebungsjahren zeigen die durchschnittlichen Lebenszufriedenheitswerte pro Lebensalter keinen Abfall bis weit ins höhere Alter hinein: In allen untersuchten querschnittlichen Verläufen (d.h. in den 16 SOEP-Erhebungswellen) fand erst ab ungefähr der Mitte des achten Lebensjahrzehnts ein moderates Absinken der Durchschnittswerte mit fortschreitendem Alter statt. „Paradoxe“ Mittelwertstabilität ist also nicht uneingeschränkt über die gesamte hochaltrige Lebensphase hinweg festzustellen, sondern es kann hier eine mittelwertstabile Phase früher Höheraltrigkeit von einer durch kontinuierliche leichte durchschnittliche Zufriedenheitsverluste gekennzeichneten zweiten Phase später Höheraltrigkeit unterschieden werden.

Dieses aus der querschnittlichen Perspektive gewonnene Bild verändert sich bei Betrachtung der längsschnittlichen Entwicklung der allgemeinen Lebenszufriedenheit: Wenn jeweils die sukzessiven Mittelwerte von Ein-Jahres-Geburtsjahrganggruppen über die 16

untersuchten Erhebungsjahre hinweg verfolgt werden, dann zeigt sich während des gesamten untersuchten Altersspektrums des Übergangs in die und Durchlaufens der Höheraltrigkeit (hier: ab dem fünfundvierzigsten bis zum neunzigsten Lebensjahr) ein fortgesetzter Abfall und es sind über diese gesamte Altersspanne hinweg die intraindividuellen Veränderungen der Lebenszufriedenheit von Lebensjahr zu Lebensjahr im Durchschnitt negativ.

Die entsprechende Untersuchung der Durchschnittswerte spezifischer Bereichszufriedenheiten mit dem Einkommen, der Gesundheit, dem Wohnen und der Freizeit ergab teilweise stark von den für die allgemeine Lebenszufriedenheit gefundenen Charakteristika abweichende Befunde. Das höhere Erwachsenenalter erschien hier als hinsichtlich der Einkommens-, Wohn- und Freizeit Zufriedenheit durchaus günstige Lebensphase, bei einer erwartbar negativen Entwicklung der Zufriedenheit mit der Gesundheit: Es wurde ein Anstieg der Einkommenszufriedenheit während der Phase früher Höheraltrigkeit und ein sehr deutlicher positiver „Wachstumsschub“ der Freizeit Zufriedenheit beim Übergang in die Höheraltrigkeit vorgefunden, besonders letztere wurde von Menschen, die das sechzigste Lebensjahr bereits vollendet haben, insgesamt deutlich besser bewertet, als von jüngeren Personen. Die Wohnzufriedenheit befand sich in allen Altersphasen auf einem über das Alter hinweg invariant sehr hohen, stabilen Niveau. Lediglich ein deutlicher Abfall der Gesundheitszufriedenheit erschien hier als negative Begleiterscheinung des Alterns im höheren Lebensalter – dieser Abfall wurde im übrigen schon über recht frühe Lebensaltersjahre vor dem Altersbereich, den man mit dem Begriff der Höheraltrigkeit belegen könnte, beobachtet und schien gerade während des Übergangs in die Höheraltrigkeit durch eine kurze Phase der Mittelwertstabilität unterbrochen. Wie schon in Unterkapitel 3.2.1.3 gesagt wurde, bieten bestimmte alterstypische Entwicklungen der objektiven Lebensumstände Höheralt-riger naheliegende Erklärungen für einige dieser Befunde (z.B. wachsende Morbidität im höheren Alter zur Erklärung sinkender Gesundheitszufriedenheit, Austritt aus dem Arbeitsleben zur Erklärung ansteigender Freizeit Zufriedenheit), während manches Ergebnis auch überraschen mag – jedenfalls gehören solche Erklärungen nicht zum Thema der vorgestellten Untersuchung und sind die in Kapitel 3.2.1 berichteten speziellen Befunde zu den Bereichszufriedenheiten eher von allgemeingerontologischem und jenseits der hier untersuchten Fragestellungen bestehendem Interesse. Insgesamt zeigt sich, auch dieses wurde bereits festgehalten, in den Bereichszufriedenheiten das höhere Erwachsenenalter als Lebensphase, die nicht durch einfache Negativstereotype zu beschreiben ist, vielmehr erscheint die Höheraltrigkeit durch erwartbare Verluste *und* möglicherweise unverhoffte Zuwächse in der subjektiven Einschätzung verschiedener Lebensumstände gekennzeichnet.

4.1.2 Analysen zur Kohorteneffekthypothese

Zur Untersuchung der Hypothese einer Überlagerung und querschnittlichen Neutralisierung negativer Effekte des Lebensalters und des Geburtsjahres wurden Multilevel-Regressionsmodelle mit linearen Effekten von Geburtsjahr und Lebensalter auf die Zufriedenheitswer-

te der SOEP-Befragungspersonen berechnet. Die Ergebnisse dieser Berechnungen zeigten für die Werte der allgemeinen Lebenszufriedenheit in der untersuchten Gruppe der Geburtsjahrgänge 1924 bis 1939, d.h. der bei Panelstart 45- bis 60-Jährigen, die also im Zeitraum der SOEP-Erhebungsjahre 1984 bis 1999 ungefähr die Lebensaltersphase des Übergangs in die und Durchlaufens der „frühen Höheraltrigkeit“ repräsentieren, ein Ergebnis, das überraschend genau der Kohorteneffekthypothese entsprach, nämlich (als Regressionskoeffizienten) geschätzte Alters- und Kohorteneffekte, die beide negativ gerichtet, annähernd betragsgleich und zudem statistisch signifikant waren und aus denen somit eine lineare Entwicklung der Durchschnittswerte von $\text{Alter} \times \text{Geburtsjahres-Mittelwerten}$ über die Meßzeitdimension vorherzusagen wäre, die dem durch die Kohorteneffekthypothese postulierten und in Abbildung 1.1 (Unterkapitel 1.2.3.1) gezeigten Verlaufsschema entspricht.

Bei der Bewertung dieses Befunds ist jedoch die grundsätzliche methodische Problematik der Untersuchung von Alters- und Kohorteneffekten zu berücksichtigen: Die prinzipielle Möglichkeit eines dritten Typus von zeitvariablen Effekten, nämlich von Periodeneffekten (der Meßzeitpunkte) macht infolge der vollständigen Abhängigkeit der dann als Prädiktoren fungierenden Zeitdimensionen (Meßzeitpunkt, Alter, Geburtszeitpunkt) eine Isolation der drei Effektypen unmöglich und erzeugt unvermeidliche Unsicherheiten bezüglich der Interpretation jeglichen vorgefundenen Effekts der Variation in einer der drei Zeitdimensionen. Um überhaupt Alters- und Kohorteneffekte untersuchen zu können, sind letztlich unbewiesene Vorabannahmen über einen oder alle dieser Effekte vorauszusetzen. Bei dem berechneten Multilevel-Regressionsmodell bestand diese Vorabannahme in der Linearität und im Ausschluß von Interaktionen der Effekte. Unter dem Vorbehalt der Gültigkeit dieser Linearitätsannahmen sind die errechneten Alters- und Kohorteneffektkoeffizienten mit der Ungenauigkeit versehen, daß ein möglicherweise vorhandener Periodeneffekt in ihrem jeweiligen Betrag mit aufsummiert wäre, so daß sie nur unter der Annahme geringfügiger Periodeneffekte als unverzerrter Ausdruck von Alters- und Kohorteneffekten zu bewerten sind.

Die Modellierung linearer Alters- und Kohorteneffekte wurde damit gerechtfertigt, daß hier eine beschränkte Spanne von Geburtsjahrgängen und Lebensaltern in die Analyse einbezogen wurde: Innerhalb solchermaßen kleinerer Abschnitte des Geburtsjahres- und Lebensaltersspektrums sollten nichtlineare Veränderungen im Trend des Kohorten- bzw. Alterseffekts nicht so gravierend sein, daß eine näherungsweise lineare Anpassung an die Trendverläufe nicht mehr möglich wäre. Geringfügige Nichtlinearitäten der Effekte aber wären hinsichtlich der inhaltlichen Fragestellung nicht relevant, es interessierten im Grunde nur die groben linearen Trends über die Zeitdimensionen.

Allerdings ist logisch nicht auszuschließen, daß die Ergebnisse eher einen vorhandenen Periodeneffekt, denn Alters- und Kohorteneffekte zeigen. Dagegen spricht allerdings m.E. schlicht auch die mangelnde theoretische Begründung eines solchen Effekts, der über den gesamten Untersuchungszeitraum von 16 Jahren gewirkt haben sollte – wie bereits gesagt wären unter der Voraussetzung eines negativen Alters- und Kohorteneffekts einerseits oder

unter der Voraussetzung eines negativen Periodeneffekts andererseits die gezeigten Befunde gleichermaßen wahrscheinlich, umgekehrt erscheint aber unter der Voraussetzung der Befunde das Vorhandensein von negativem Alters- und Kohorteneffekt zumindest subjektiv wahrscheinlicher, als das eines negativen Periodeneffekts.

Abgesehen davon wurde in den graphischen Abbildungen der längsschnittlichen Verläufe der durchschnittlichen Lebenszufriedenheiten verschiedener Geburtskohortengruppen ein gewisser „Gipfel“ um das Jahr 1991 herum gefunden, welcher durchaus als Hinweis auf einen kurzfristigen – und damit nicht über die gesamte Erhebungszeitspanne hinweg linearen – positiven Periodeneffekt bewertet werden könnte und für den auch mit dem zu dieser Zeit in Deutschland erlebten historischen Umbruch der deutschen Wiedervereinigung eine vordergründig plausible Erklärung gefunden werden könnte. Insgesamt erschien dieser kurze Ausschlag in den Verlaufskurven allerdings zu geringfügig, um ernsthaftere Bedenken bezüglich der Gültigkeit des linearen Multilevel-Regressionsmodells zu begründen. Zudem ergab die Modellberechnung für andere, sozusagen in der Geburtsjahresdimension rückwärts verschobene Personenengruppen bei der Gruppe der 1914 bis 1929 Geborenen einen nur noch sehr niedrigen Kohorteneffektwert: Ein linearer Periodeneffekt „im“ Betrag des Kohorteneffekts aber müßte auch in unterschiedlichen Geburtsjahresgruppen stets derselbe sein, so daß dieser niedrigere Kohorteneffektwert einen gewissen Rückschluß auf einen entsprechend niedrigen Periodeneffektwert zuläßt – dieses ist hier, in der abschließenden Zusammenfassung, kaum hinreichend genau darzulegen, es sei hier auf die diesbezüglichen Überlegungen im Unterkapitel 3.2.2.2 verwiesen.

Eingedenk gewisser methodischer Unsicherheiten, die v.a. aus der prinzipiellen und m.E. eigentlich unrealistischen Möglichkeit eines Periodeneffekts resultieren, können die Ergebnisse der Berechnung des Multilevel-Regressionsmodells für die allgemeine Lebenszufriedenheit als Bestätigung der Kohorteneffekthypothese gewertet werden: Allem Anschein nach gibt es auch während der Phase des Übergangs in die und des Durchlaufens der frühen Höheraltrigkeit einen negativen Alterseffekt, welcher allerdings bei querschnittlichen Untersuchungen nicht erscheinen kann, weil er durch einen entsprechend negativen Kohorteneffekt des Geburtszeitpunkts überlagert wurde. Die Bedeutung der Befunde „paradoxe“ Mittelwertstabilität aus Querschnittstudien, die meist ungefähr während des Zeitraums der ersten SOEP-Befragungen durchgeführt wurden, ist damit zumindest in Frage gestellt: Die Zufriedenheitsmittelwerte erscheinen gar nicht über das Alter hinweg stabil!

Ob allerdings der vorgefundene negative Alterseffekt groß genug ist, um letztendlich die Diskussion um die Befunde querschnittlicher Stabilität insgesamt einfach damit abzutun, daß es sich bei dem sogenannten Paradox letztlich nur um ein Artefakt handle und alle Erwartungen eines alterskorrelierten Verlustes von Lebenszufriedenheit durch den gefundenen Effekt erfüllt würden, ist hier nicht weiter zu beurteilen. Denn es ist damit der Aspekt praktischer Bedeutsamkeit der Alterseffektgröße angesprochen, zu deren Beurteilung hier m.E. die rationalen Kriterien fehlen. Festzuhalten bleibt, daß ein, wenn auch nicht allzu groß scheinender, negativer Alterseffekt, sowie ein negativer Kohorteneffekt und Belege

für die Gültigkeit der Kohorteneffekthypothese gefunden wurde.

Für die Bereichszufriedenheiten wurde kein derartiges Ergebnis gefunden, was bereits aufgrund der deskriptiven Analysen, bei denen sich kein der Kohorteneffekthypothese entsprechendes Verlaufsmuster zeigte, nicht mehr überraschen konnte. Dieses schmälert m.E. nicht die Bedeutung der Ergebnisse für die allgemeine Lebenszufriedenheit, da es bei der Hypothese ja letztlich um diese geht.

4.1.3 Analyse der intraindividuellen Merkmalsstabilität

Für die Durchführung der Analysen zur Merkmalsstabilität der untersuchten Zufriedenheiten waren einige methodische Probleme solcher Analysen zu klären und entsprechende Entscheidungen bezüglich der statistischen Vorgehensweise zu treffen – dieses wurde in Kapitel 2 ausführlich dargestellt und wird im folgenden Unterkapitel noch einmal zusammengefaßt.

Die Stabilitätsanalysen wurden dann in zwei Schritten durchgeführt: Bei den allgemeinen Stabilitätsanalysen wurde das spezifiziertere längsschnittliche Strukturgleichungsmodell für die Gesamtgruppe aller erwachsener westdeutscher und für die Gruppe aller beim Panelstart 1985 mindestens 60 Jahre alten SOEP-Befragungspersonen berechnet. Hierbei wurden also die Stabilitätskoeffizienten der intraindividuellen Veränderungen zwischen den und die True-Score-Varianzen zu den einzelnen Meßzeitpunkten über die Gesamtgruppe bzw. eine Gesamtgruppe Älterer hinweg berechnet, ohne Differenzierung spezifischerer Geburtsjahrgangsgruppen. Es wurden sieben Modellversionen formuliert, die u.a. die Restriktion auf ein Veränderungsmodell absoluter monotoner und absolut paralleler Stabilität, sowie auf ein Modell unvollständiger Stabilität und Stationarität zweiter Ordnung beinhalteten. Eine zusätzlich berechnete Modellversion enthielt eine Restriktion der Mittelwertsstruktur des Modells, nämlich invarianter Zufriedenheitsmittelwerte zu allen Meßzeitpunkten.

Die allgemeinen Stabilitätsanalysen zeigten m.E. zwei herausragende Ergebnisse, nämlich die Entscheidung für das stationäre Veränderungsmodell und die sehr hohe Merkmalsstabilität der Zufriedenheiten, die dort sowohl für die Gesamtgruppe aller Befragter, als auch für die Subgruppe der zum Panelbeginn 1984 Älteren in allen berechneten Modellversionen gefunden wurde.

Zwar mußten diejenigen Modellversionen, durch deren Restriktionen vollständige Stabilität im Sinne von Stabilitätskoeffizienten gleich Eins spezifiziert wird, gemäß der festgesetzten Entscheidungskriterien zurückgewiesen werden, doch bewegten sich die Koeffizienten des unrestringierten – mit über die Zeitreihe hinweg variablen, frei schätzbaren Koeffizienten – und auch des letztlich angenommenen stationären Modells durchweg sehr nahe an diesem theoretisch möglichen Maximalwert. Zufriedenheitsbewertungen sind offenbar ganz allgemein und über die gesamte Spanne des Erwachsenenalters hinweg hochstabil! Zudem zeigten die Quotienten Q_{Δ} an, daß die Varianzen der True-Score-Veränderungen in Relation zur „Ausgangsvarianz“ des jeweiligen Erhebungsintervalls sehr niedrig sind, so

daß insgesamt annähernd parallele Stabilität der Zufriedenheiten konstatiert werden konnte.

Zur genaueren Bewertung des Typus der Stabilität sind auch die Mittelwertsveränderungen relevant: Eine annähernd parallele Stabilität würde sozusagen dadurch zur annähernd strikten, daß auch noch die Mittelwerte weitestgehend invariant wären. Es wurde die unrestringierte Modellversion auch mit einer hinsichtlich der Kovarianzstruktur unrestringierten, jedoch in der Mittelwertsstruktur zur Invarianz der Zufriedenheitsmittelwerte über die Zeit restringierten Modellversion verglichen. Bei der allgemeinen Lebenszufriedenheit und Gesundheitszufriedenheit, sowie auch in der Gesamtgruppe aller Befragter bei der Einkommenszufriedenheit waren allerdings die durchschnittlichen Veränderungen immerhin ausgeprägt genug, um zur Zurückweisung des Modells invarianter Mittelwerte zu führen. Zwar wurde auch mit letzterem noch in allen diesen Fällen eine gute Modellanpassungsgüte erreicht, jedoch kann hier die Stabilität des Merkmals konsequenterweise nicht als annähernd strikte bezeichnet werden. Bei der dennoch hohen und annähernd parallelen Stabilität bedeutet diese Zurückweisung ausgerechnet der Invarianz der Mittelwerte: Verschiebungen des Mittelwerts bei großer Invarianz der individuellen Abweichungen vom jeweiligen Mittelwert.

Auch die näherungsweise „bloß“ parallele Merkmalsstabilität der Lebenszufriedenheit nicht nur in der Untersuchungsgruppe der 1984 Älteren, sondern auch in der Gesamtgruppe aller westdeutschen SOEP-Befragten, bedeutet einen empirischen Beleg für die Relevanz der Hypothese allgemeiner, lebenslanger Stabilität von Zufriedenheit: Personen bewahren die „relative Position“ ihrer Lebenszufriedenheit weitgehend bei. Angesichts eines in der diesbezüglichen Theoriediskussion – um Trait-Konzepte des SWB und Top-down-Einflüsse intraindividuell stabiler Persönlichkeitsmerkmale sowie des Ausmaßes, in dem allgemein adaptativ relativierende Beurteilungsstrategien zu solcher Stabilität beitragen könnten – noch nicht vorhandenen Konsens über die Stabilität des Konstruktes und der auf empirischen Ergebnissen begründeten variierenden Einschätzungen des Ausmaßes dieser Stabilität (vgl. Kapitel 1.1.4.4) erscheinen die hier gefundenen sehr hohen Stabilitäten als der vielleicht bedeutsamste Befund, der mitgeteilt werden kann. Diese sehr hohen Stabilitätskoeffizienten resultieren aus der Modellierung latenter True-Scores und sind insofern meßfehlerbereinigt – würden die Reliabilitäten der Messungen außer Acht gelassen und lediglich die Ein-Jahres-Korrelationen der Meßwerte berechnet, so würden diese den Eindruck deutlich niedrigerer Stabilitäten vermitteln, wie er aus den Ergebnissen vieler empirischer Untersuchungen, in denen die Stabilität von Lebenszufriedenheit auf solche Weise untersucht wurde, entstehen kann.

Die letztlich für alle Zufriedenheiten ausgewählte Modellvariante ist die eines durch Stationarität zweiter Ordnung gekennzeichneten Veränderungsmodells, d.h. es sind hier über die Zeitreihe hinweg invariante True-Score-Varianzen und Stabilitätskoeffizienten festgelegt, die längsschnittliche Veränderung ist dabei also durch lediglich einen einzigen Ein-Jahres-Stabilitätskoeffizienten, der den Zusammenhang sukzessiver True-Score-Werte für

alle 15 Intervalle beschreibt, und durch eine zu allen Meßzeitpunkten gleiche Varianz der True-Scores charakterisiert. Die Bedeutung, die die Entscheidung für dieses Modell auch für die Gruppe der 1984 mindestens 60-Jährigen hinsichtlich der Hypothese altersspezifisch erhöhter Stabilität und der Gewinner-und-Verlierer-Hypothese hat, liegt auf der Hand: Es wurden damit weder Hinweise auf im höheren Alter längsschnittlich ansteigende True-Score-Varianzen, noch für steigende Stabilität gefunden – und somit also keine Belege für diejenigen Effekte, die mit diesen beiden Hypothesen postuliert werden.

Jedoch behindert die hinsichtlich der einbezogenen Spanne von Geburtsjahrgängen sehr breite Definition der Höheraltrigengruppe allzu eindeutige Schlüsse auf das Nichtvorhandensein irgendeines Alterseffekts auf Stabilität oder True-Score-Varianzen. Denn es wäre ja denkbar – und aus den Überlegungen heraus, mit denen diese beiden Hypothesen begründet sind (vgl. Kapitel 1.2.3.1 und 1.2.3.4), m.E. sogar realistisch anzunehmen –, daß derartige Effekte nicht gleichermaßen über die gesamte Spanne des hohen Lebensalters hinweg, sondern v.a. beim Übergang in die und Durchlaufen der frühen Höheraltrigkeit wirksam sind und sich später dann abschwächen. In diesem Fall aber wäre zu erwarten, daß sie sich in der breit definierten Höheraltrigengruppe nur schwach abbilden, wegen der in die Analyse miteinbezogenen „alten“ Alten, deren Anteil an dieser Gruppe zudem mit fortschreitendem Meßzeitpunkt wächst. Starke Alterseffekte auf Stabilitätskoeffizienten oder True-Score-Varianzen während der frühen Höheraltrigkeit müßten dann allerdings dennoch soweit „durchschlagen“, daß es auch bei der breit definierten Älterengruppe zu einer Entscheidung gegen das stationäre Veränderungsmodell kommt.

Allerdings könnten sich dann, wenn es während der früheren Höheraltrigkeit zu einem Anstieg der True-Score-Varianzen oder der Stabilitätskoeffizienten kommt und die dann gestiegenen Kennwerte im weiteren Altersverlauf der späten Höheraltrigkeit ohne weitere Veränderung auf hohem Niveau gehalten werden, solche Alterseffekte auch darin zeigen, daß die entsprechenden Kennwerte in der Gruppe der 1984 mindestens 60-Jährigen über denen der Gesamtgruppe liegen, da sich letztere zu einem großen Teil aus Personen aus jenen jüngeren Lebensaltersbereichen zusammensetzt, in denen dann ja diese Alterseffekte noch nicht wirksam wären. Solchermaßen gegenüber der Gesamtgruppe deutlich erhöhte Stabilitätskoeffizienten der Lebenszufriedenheit (und auch der anderen Zufriedenheiten) wurden für die Gruppe der Älteren nicht errechnet, es fand sich auch hier kein Beleg für eine im höheren Lebensalter erhöhte Merkmalsstabilität. Jedoch zeigen die Ergebnisse eine deutlich höhere True-Score-Varianz der Lebenszufriedenheit bei den Höheraltrigen im Vergleich zur Gesamtgruppe. Während also im längsschnittlichen Verlauf kein Anstieg der Lebenszufriedenheitsvarianz älterer Menschen erkennbar wird, erscheint beim Vergleich von Älteren und Gesamtgruppe ein der Gewinner-und-Verlierer-Hypothese entsprechender Effekt des höheren Lebensalters erkennbar. Rein „technisch“ betrachtet könnte dieses aber auch wiederum als Ausdruck eines Kohorteneffekts interpretiert werden, da es sich ja genau genommen dabei um einen querschnittlichen Vergleich handelt.

Ein ähnlicher deutlicher Gruppenunterschied der True-Score-Varianzen in der Höheral-

trigengruppe und der Gesamtgruppe zeigte sich ansonsten nur noch bei der Gesundheitszufriedenheit und gerade auch darin könnte ein Hinweis auf die Bedeutung der Gewinner-und-Verlierer-Hypothese gesehen werden, wenn man die besonders bezüglich gesundheitlicher Probleme und körperlichen Einschränkungen allgemein bestehenden Erwartungen an die Höheraltrigkeit berücksichtigt, die möglicherweise nicht nur im Falle ihrer Erfüllung zu Zufriedenheitsverlusten führen, sondern auch im Falle ihrer Nichterfüllung zu Zufriedenheitsgewinnen: Speziell der gesundheitliche Bereich könnte als derjenige angesehen werden, der die bei Jüngeren hinsichtlich des eigenen Alters bestehenden Erwartungen in besonderer Weise prägt und deshalb wären dann besonders die aus der Gewinner-und-Verlierer-Hypothese resultierenden Effekte für diese Bereichszufriedenheit ähnlich zu erwarten, wie für die allgemeine Lebenszufriedenheit.

Weitere erwähnenswerte – wenn auch nicht für die Bewertung der Merkmalsstabilität im Allgemeinen und der hier zugrundegelegten Hypothesen im besonderen relevante – Befunde aus den allgemeinen Stabilitätsanalysen sind: Zum einen ein erkennbarer Unterschied der True-Score-Varianzen der Lebens- und Gesundheitszufriedenheit bei den Älteren und bei der Gesamtgruppe aller Befragter – in beiden Bereichen variierten die Älteren deutlich geringer. Zum zweiten fiel eine gewisse Ähnlichkeit von Lebens- und Gesundheitszufriedenheit bei den Befunden der allgemeinen Stabilitätsanalysen auf (hinsichtlich der Gruppenunterschiede von Varianzen und der Gruppenunterschiede und längsschnittlichen Veränderung von Mittelwerten), in der sich ein Zusammenhang zwischen diesen beiden Zufriedenheiten andeutet, welcher auch durch die einfachen bivariaten Meßwertkorrelationen sich zeigt.

Zusätzlich zu den als allgemeine Stabilitätsanalysen bezeichneten Auswertungen sollte durch die als altersspezifische Stabilitätsanalysen bezeichneten Berechnungen schließlich noch das Ausmaß längsschnittlicher Veränderung *in den Kennwerten der Stabilitätsanalysen* genauer geklärt werden, als dies vermittels der allgemeinen Stabilitätsanalysen möglich war, d.h es wurden hiermit die im Bereich der Höheraltrigkeit möglicherweise vorhandenen Alterseffekte auf die Entwicklungen von True-Score-Varianzen und Stabilitätskoeffizienten untersucht. Es sollte also die „Veränderung der Veränderung“ sozusagen intensiver modelliert werden, als dieses bereits bei den allgemeinen Stabilitätsanalysen durch die Überprüfung der stationären Modellvariante der Fall war.

Anstatt einer über viele Geburtsjahrgänge hinweg breit definierten Höheraltrigengruppe wurden dazu mehrere enger definierte Vier-Jahres-Geburtsjahrgangsgruppen – als Kompromiß aus der Notwendigkeit größerer Fallzahlen in den Subgruppen und zumindest nahe an der genauestmöglichen Unterscheidung von Ein-Jahres-Geburtsjahrgangsgruppen liegenden Subgruppendefinitionen – längsschnittlich analysiert, wozu das bereits bei den allgemeinen Stabilitätsanalysen zugrundegelegte Strukturgleichungsmodell zum Multisample-Modell modifiziert wurde, durch welches eine simultane subgruppenspezifische Parameterschätzung und auch Gleichheitsrestriktionen zwischen den Gruppen ermöglicht war. Zur Untersuchung der Alterseffekte auf die relevanten Parameter des Modells wurden dann

in verschiedenen Modellversionen Parameterrestriktionen innerhalb der und zwischen den Gruppen solchermaßen spezifiziert, daß dabei jeweils eine der möglichen Überlagerungen von zwei der drei möglichen zeitvariablen Effekte (Alter, Kohorte, Periode) auf Null gesetzt wurde – die Logik und genaue Form des hier zugrundegelegten Analysedesigns kann nicht in wenigen Sätzen hinreichend dargelegt werden, es ist hier auf die ausführliche Darstellung im vorherigen Text zu verweisen (Kapitel 3.3.2). Entscheidend ist, daß auf diese Weise eine indirekte Überprüfung dieser Effekte ermöglicht wurde, welche allerdings mit den bereits erwähnten prinzipiellen Unsicherheiten behaftet ist, die jedem Versuch einer strengen Isolation der drei zeitvariablen Effekte letztendlich scheitern lassen. Immerhin war bei dieser Vorgehensweise keine Linearitätsannahme impliziert, was hier darum wichtig erschien, da ja das Hauptmotiv für diese Analysen aus der Vermutung von über die Höheraltrigkeitsspanne hinweg nichtlinearen Veränderungen in True-Score-Varianzen oder Stabilitätskoeffizienten (nämlich früher Anstieg und spätere Stabilisierung) abgeleitet wurde – wenn die Linearität solcher Effekte garantiert wäre, dann wären im Grunde die allgemeinen Stabilitätsanalysen ausreichend gewesen.

Das wesentliche Ergebnis dieser zusätzlichen Stabilitätsanalysen bestand darin, daß keine bedeutsamen Alterseffekte auf True-Score-Varianzen oder Stabilitätskoeffizienten gefunden wurden: Im Vergleich der verschieden restringierten Modellversionen führte jene, bei der die Gleichheit von True-Score-Varianzen und Stabilitätskoeffizienten über alle Gruppen und alle Meßzeitintervalle hinweg festgelegt war, gegenüber der diesbezüglich unrestringierten Variante bei keiner der untersuchten Zufriedenheiten zu einer nennenswerten Verschlechterung der Modellanpassungsgüte. Dieses somit angenommene Modell aber bedeutet schlicht: Keine zeitvariablen Effekte! Fügt man diesem Ergebnis noch diejenigen der allgemeinen Stabilitätsanalysen hinzu, so ergibt sich der m.E. eindeutige Befund, daß in den untersuchten Daten die durch Gewinner-und-Verlierer-Hypothese und durch die Hypothese erhöhter Merkmalsstabilität postulierten Alterseffekte zumindest nicht in größerem Maße vorhanden waren. Das angenommene restriktive Modell schließt das durch die Gewinner-und-Verlierer-Hypothese postulierte Wachstum der Merkmalsvarianz über das höhere Alter hinweg und auch eine wachsende „Disparallelität“ der intraindividuellen Veränderungen infolge der Diversifizierung von Zufriedenheitsgewinnern und Zufriedenheitsverlierern aus. Ebenso schließt es eine über diese Lebensphase hinweg allmählich wachsende Zufriedenheitsstabilität infolge z.B. erhöhter adaptiver Kompetenz oder größerer Retrospektivität der Urteilsperspektive aus, wie sie durch die Hypothese erhöhter Stabilität vermutet wurde. Allem Anschein nach sind die Zufriedenheiten während des gesamten Erwachsenenalters sehr stabil und belieben es unverändert bis ins höchste Alter hinein, das höhere Lebensalter scheint keine diesbezüglichen altersspezifische Entwicklungsdynamiken mit sich zu bringen.

Gewissermaßen jenseits der angenommenen restriktiven Modellversion ohne jeglichen zeitvariablen Effekt auf True-Score-Varianzen und Stabilitätskoeffizienten wurden allerdings in den Parameterschätzungen der unrestringierten Variante gewisse Hinweise auf

zumindest schwache Effekte, die der Gewinner-und-Verlierer-Hypothese entsprächen, gefunden: Dort zeigte sich eine tendenzielle Zunahme von True-Score-Varianzen über das Lebensalter und ein sozusagen zu erahnender schwacher Trend eines Anstieg der Quotienten Q_{Δ} über das höhere Alter hinweg (was dann erhöhte Disparallelität der intraindividuellen Merkmalsverläufe bedeuten würde). Die Interpretierbarkeit dieses Befundes wurde aber bereits in Frage gestellt (vgl. in Unterkapitel 3.3.2.3), es bleibt diesbezüglich lediglich festzuhalten, daß gewisse vage Hinweise erkennbar wurden, daß die aus der Gewinner-und-Verlierer-Hypothese postulierten Effekte eine gewisse und wohl eher schwache Rolle bei der im höheren Erwachsenenalter typischen Entwicklung der Zufriedenheit spielen könnten. Dagegen fand sich für die Hypothese erhöhter Merkmalsstabilität im höheren Erwachsenenalter keinerlei Bestätigung durch die hier vorgestellten Ergebnisse.

4.2 *Statistische Methoden der längsschnittlichen Untersuchungen*

Die Dokumentation und Anwendung neuerer Entwicklungen statistischer Längsschnittmethoden – insbesondere zur Modellierung von intraindividuelle Veränderung – wurde als explizites Ziel dieser Untersuchung zugrundegelegt, welches sozusagen zusätzlich zu den inhaltlich-theoretischen Untersuchungszielen einen gewissen eigenständigen Wert für die psychogerontologische Erforschung von *Altern*sprozessen besitzt, welche mit dem methodischen Problem der Erfassung und Untersuchung intraindividuelle Veränderung in besonderem Maße konfrontiert ist. Es wurde diesem Ziel v.a. durch eine ausführlichere – als für die Dokumentation „nur inhaltlicher“ Untersuchungen üblich – Darstellung und Diskussion der statistischen Methoden im Kapitel 2 entsprochen.

Zusammenfassend ist zunächst auf die Modellierung von Alters- und Kohorteneffekt durch das in Kapitel 2.1 formulierte Multilevel-Regressionsmodell hinzuweisen. Die Vor- und Nachteile desselben wurden dort bereits ausgiebig diskutiert, es bleibt hier nur als Fazit, daß damit eine prinzipiell auch in anderen Forschungszusammenhängen anwendbare Möglichkeit zur Untersuchung zeitvariabler Effekte vorgeschlagen wurde, welche, nebenbei bemerkt, auch den „state of the art“ des Umgangs mit fehlenden Werten gewährleistet.

Besonders ausführlich wurde die Methodik der längsschnittlichen Analyse von Merkmalsstabilität erörtert. Die wesentlichen Gesichtspunkte dieser Erörterungen lassen sich wie folgt zusammenfassen:

- Es wurden die möglichen statistischen Kennwerte, sowie die formalen Beziehungen zwischen diesen Kennwerten, die bei der Analyse intraindividuelle Veränderung und Stabilität relevant sind, im Kapitel 2.2.1 besprochen. Als wesentlicher Kennwert intraindividuelle Stabilität (der Meßwerte und/oder der True-Scores) wurde die Korrelation der Werte im Meßwiederholungsdesign zwischen verschiedenen Meßzeitpunkten benannt und es wurde die von Tisak & Meredith (1990) vorgeschlagene Unterscheidung

von drei Stabilitätstypen mittels der besprochenen Kennwerte vorgenommen. Zudem wurde in diesem Zusammenhang der Quotient Q_{Δ} zur Relativierung der Varianz stattfindender Veränderung an der unveränderten Varianz der Werte vorgeschlagen.

- ▶ Es wurde die Relevanz einer expliziten Berücksichtigung der Reliabilität der untersuchten Meßwerte bei Analysen längsschnittlicher Veränderung bzw. Stabilität dargelegt (Kapitel 2.2.2 und Anhang A.1): Längsschnittliche Autokorrelationen von Meßwerten sind durch die in ihnen enthaltenen Meßfehleranteile reduziert, zur Schätzung der Stabilität der „wahren“ Werte untersuchter Konstrukte sind deshalb Schätzungen meßfehlerbereinigter Korrelationen vonnöten. Solche Schätzungen können mittels längsschnittlicher Strukturgleichungsmodelle vorgenommen werden, da diese statistische Methodologie eine explizite Modellierung von beobachteten Meßwerten und latenten True-Scores bzw. der Meßmodelle, durch die diese beiden verbunden sind, ermöglicht.
- ▶ Es wurden verschiedene Typen längsschnittlicher Strukturgleichungsmodelle dargestellt und diskutiert. Angesichts der großen Flexibilität der mit dieser Methode gegebenen Modellierungsmöglichkeiten mußte dieses allerdings auf den für die vorliegende Untersuchung relevanten Fall der Modellierung von Ein-Item-Zeitreihen, d.h. auf Meßwiederholungen jeweils nur eines Meßwerts für das zu untersuchende latente Konstrukt, eingeschränkt werden. Ausführlicher dargestellt wurde das bereits von Jöreskog (1970) vorgestellte Quasi-Markov-Simplex-Modell (Unterkapitel 2.2.2.1), sowie ein Modell latenter Differenzkomponenten (Unterkapitel 2.2.2.2), ein Modell latenter Differenzbeträge (Unterkapitel 2.2.2.3) und die einfachste lineare Version des sogenannten Latent-Growth-Curve-Models (Unterkapitel 2.2.2.4). Letzteres ist unter den längsschnittlichen Strukturgleichungsmodellen dasjenige, welches in jüngster Vergangenheit noch am ehesten bekannt geworden ist und es wurde auch dargelegt, daß dieser Modelltypus nicht immer die optimale Wahl längsschnittlicher Modellierung darstellt: In Fällen wie dem bei dieser Untersuchung vorliegenden, wo keine vorab festzulegende Entscheidung für ein spezifisches (z.B. über die gesamte Zeitreihe lineares) Verlaufsmodell der intraindividuellen Veränderung möglich oder erwünscht ist, sind solche Modelle zu restriktiv und eventuell die vorgenannten Modelltypen vorzuziehen.
- ▶ Es wurde die Problematik des Umgangs mit fehlenden Werten ausführlich erörtert (Unterkapitel 2.3.1), welche gerade bei längsschnittlichen Untersuchungen mit dem dort im Grunde unvermeidlichen Dropout von Untersuchungsteilnehmern besonders dringlich ist. Der „Stand der Dinge“ in der diesbezüglichen Methodendiskussion beruht auf der Unterscheidung verschiedener stochastischer „Mechanismen“ des Datenausfalls, welche im Allgemeinen mit den Begriffen MCAR („missing completely at random“), MAR („missing at random“) und MNAR („missing not at random“) benannt werden. Als Ergebnis dieser Erörterungen ist festzuhalten, daß herkömmliche und oft gewohnheitsmäßig angewandte Verfahren zum Umgang mit fehlenden Werten, wie der fallweise Ausschluß derselben, nur dann zu unverzerrten Ergebnissen führen, wenn die Gültigkeit der sehr restriktiven Bedingung MCAR garantiert ist, ansonsten jedoch allenfalls bei einer vernachlässig-

baren Zahl fehlender Werte akzeptabel sind. Dagegen sind Maximum-Likelihood-Parameterschätzungen, die alle vorhandenen Werte miteinbeziehen, den stochastischen Prozeß des Datenausfalls aber nicht weiter berücksichtigen, auch dann anwendbar, wenn die weniger restriktive Bedingung MAR vorliegt. Solche Verfahren sind darum vorzuziehen und stellen in gewisser Weise bis auf Weiteres (d.h. bis zur Entwicklung überzeugender und leicht anwendbarer Algorithmen zum Umgang mit MNAR-Datenausfällen) den „state of the art“ zum Umgang mit fehlenden Werten dar. Der in dieser Untersuchung angewandte FIML-Algorithmus ist ein solches Verfahren zur Schätzung von Strukturgleichungsmodellen.

Ergänzend zu diesen wesentlichen Gesichtspunkten der hier angewandten längsschnittlichen Modellierung intraindividuelle Veränderung sind noch zwei weitere methodische Aspekte der im vorherigen vorgestellten Analysen festzuhalten, die m.E. nicht zum selbstverständlichen Standard solcher Analysen gehören. Nämlich zum einen bei den Auswertungen des Kapitels 3.2.1 der weitestgehende Verzicht auf inferenzstatistische Analysen und insbesondere auf Signifikanztestungen und die stattdessen gewählte Beschränkung auf die Deskription graphisch abgebildeter Merkmalsverläufe: Diese möglicherweise etwas anspruchslos erscheinende Vorgehensweise war m.E. besser geeignet, einen Eindruck der im vorhandenen Datensatz tatsächlich über die drei verschiedenen Zeitdimensionen Lebensalter, Geburtszeitpunkt und Meßzeitpunkt hinweg stattfindenden Zufriedenheitsentwicklungen zu gewinnen, als es statistische Modellierungen ermöglichen würden, die bereits gewisse Vorannahmen über die fraglichen zeitvariablen Effekte (z.B. über deren Linearität) enthalten. Zwar könnten für diese Analysen auch einfachste inferenzstatistische Verfahren erwogen werden, die bezüglich der interessierenden Effekte relativ voraussetzungsarm wären, nämlich v.a. einfache Signifikanztests der jeweils über die Zeitachsen hinweg sukzessiven Mittelwerte der aus den Kombinationen von Lebensalter \times Geburtsjahr \times Erhebungsjahr sich ergebenden Zellen, jedoch würde angesichts der Vielzahl solcher Vergleiche dann das Problem der multiplen Testung sehr bedeutsam, ganz abgesehen davon, daß der Gewinn dieser Vorgehensweise für das letztendlich explorative Untersuchungsziel bei diesem Analyseschritt kaum erkennbar scheint. Die einfachen graphischen Verlaufsabbildung genügten m.E. durchaus diesem Ziel: Sie zeigten die diskutierten Ergebnisse und machten auch geringe Veränderungen über die Zeitachsen hinweg sichtbar. Die *an dieser Stelle* erfolgte Abkehr von der Gewohnheit, jeglichen empirischen Vergleich von Mittelwerten mit einer entsprechenden Angabe zu dessen statistischer Signifikanz zu garnieren, soll also hier – auch unter Verweis auf die andernorts bereits ausgiebig diskutierte Problematik dieser Gewohnheit (vgl. z.B. Frick, 1996; Sedlmeier, 1996) – noch einmal mit Nachdruck vertreten werden.

4.3 *Abschließende Bewertung und Ausblick*

Die vorgelegte Untersuchung hat gezeigt, daß der als Zufriedenheitsparadox des höheren Alters bezeichnete Befund – bis weit ins hohe Lebensalter hinein stabiler Zufriedenheitsmittelwerte bei querschnittlichen Altersgruppenvergleichen – nicht mit dem Fehlen alterskorrelierter Zufriedenheitsverluste in diesem Lebensabschnitt gleichzusetzen ist: Obwohl er auch anhand der querschnittlichen Betrachtung der Zufriedenheitsmeßwerte in den einzelnen Erhebungsjahren der untersuchten SOEP-Daten repliziert werden konnte, ergaben die längsschnittlichen Analysen bei der allgemeinen Lebenszufriedenheit (sowie bei der Zufriedenheit mit der Gesundheit) deutliche Hinweise auf einen über die Phase der Höheraltrigkeit hinweg laufenden Negativtrend. Insbesondere für die Vermutung, daß bei den während der letzten beiden Dekaden des zwanzigsten Jahrhunderts älteren Personen ein solcher negativer Alterseffekt durch einen ungefähr betragsgleichen negativen Kohorteneffekt des Geburtsjahres solchermaßen überlagert sein könnte, daß beide sich im Querschnitt neutralisieren, konnten empirische Belege gefunden werden. Einschränkend ist allerdings hinzuzufügen, daß aus den durchgeführten Analysen der Zeitraum von Geburtsjahrgängen, über den hinweg dieser Kohorteneffekt in der genannten Weise wirkt, nicht eindeutig eingegrenzt ist.

Das absolute Ausmaß des gefundenen negativen Alterseffekts auf die Lebenszufriedenheit könnte jedoch möglicherweise als eher geringfügig bewertet werden – dieses ist in Ermangelung rationaler Kriterien für eine solche Bewertung nicht ohne Weiteres zu entscheiden. Jedenfalls *erscheint* der tatsächliche Abfall der Lebenszufriedenheitswerte über das höhere Lebensalter hinweg, der hier beobachtet werden konnte, nicht allzu groß, ebenso wie die sukzessiven Differenzen der aus dem Strukturgleichungsmodell für die Gruppe der 1984 Älteren geschätzten Mittelwerte der einzelnen SOEP-Erhebungsjahre. Aber wie auch immer man das Ausmaß dieses Negativtrends beurteilt: Er ist jedenfalls vorhanden und immerhin groß genug, um bei der statistischen Modellierung durch die Multilevel-Regressionsanalyse sich als statistisch signifikanter Effekt zu zeigen und den Versuch, ihn bei der Modellierung eines längsschnittlichen Strukturgleichungsmodells einzusparen, mit einer spür- und nicht mehr ganz vernachlässigbaren Verschlechterung der Modellanpassungsgüte zu bestrafen. Entgegen dem Bild, das gelegentlich bei der Diskussion des Zufriedenheitsparadoxes gezeichnet wird, kennzeichnet also ein durchschnittlicher Verlust an Lebenszufriedenheit das Altern in den höheren Lebensaltersphasen.

Von diesem möglicherweise nur als geringfügig zu bewertenden negativen Alterseffekt auf die Lebenszufriedenheit (und Gesundheitszufriedenheit) über das höhere Lebensalter abgesehen, erwies sich diese und ebenso auch die untersuchten Bereichszufriedenheiten als intraindividuell sehr hoch stabil. Die Zufriedenheiten könnten kaum stabiler sein, als durch die für die meßfehlerbereinigten „wahren“ Werte geschätzten Stabilitätskoeffizienten angezeigt – in der Regel lagen diese nahe beim möglichen Maximalwert. Zudem zeigten sich relativ geringe Varianzen der Merkmalsveränderung (niedrige Q_{Δ} -Werte), so daß der

Typus dieser Stabilität als annähernd parallel charakterisiert werden konnte.

Dieses Ergebnis scheint über das spezifisch gerontologische Interesse hinaus für die Zufriedenheitsforschung überhaupt relevant, da es so nicht nur für höheraltrige Personengruppen, sondern auch für die Gesamtgruppe aller westdeutscher SOEP-Befragungspersonen gefunden wurde. Es widerspricht dem Eindruck moderater Stabilität von Zufriedenheit, wie er aus der Analyse von Meßwertkorrelationen, die gegenüber den True-Score-Korrelationen reduziert sind, entstehen kann. Jedenfalls schienen die untersuchten Zufriedenheiten allesamt weitestgehend von stabilisierenden Einflüssen geprägt – welche aber durch die vorgelegte Untersuchung als solche nicht erfaßt wurden: Sind es nun intraindividuell stabile Persönlichkeitseigenschaften wie Neurotizismus und Extra-/Introversion, die mittelbar – über die affektive Komponente – oder wie auch immer unmittelbar stabilisierend auf die Zufriedenheiten gewirkt haben, oder waren die Ein-Jahres-Befragungsintervalle nicht eng genug, um kurzfristige intraindividuelle Ausschläge ausreichend zu erfassen, so daß die gefundene Stabilität v.a. der längerfristigen adaptiven Regulierung von Zufriedenheitsurteilen geschuldet ist? Alle diese theoretischen Möglichkeiten wurden vorne bereits diskutiert (Kapitel 1.1) und zudem muß auch noch eine weitere Möglichkeit in Betracht gezogen werden, nämlich daß die Stabilität der Zufriedenheiten schlicht eine weitestgehende Stabilität äußerer Lebensbedingungen reflektiert, daß also Personen nicht infolge komplexer psychologischer Wirkgefüge nicht in ihren Urteilen schwanken, sondern schlicht darum, weil kein Anlaß zu solcher Schwankung besteht.

Es steht aber immerhin außer Frage, daß bestimmte äußere gesellschaftliche Rahmenbedingungen – man denke z.B. an die Folgen der fortschreitenden Technisierung unserer Lebensumwelten mit den daraus resultierenden Veränderungen des Arbeitslebens, der kulturellen Angebote und Freizeitmöglichkeiten – sich im untersuchten Zeitraum für alle untersuchten Personen gleichermaßen verändert haben und daß andere Lebensumstände (wie z.B. das Krankheitsgeschehen) im Verlauf eines solchen Zeitraumes sich bei vielen Personen verändern. Es müßte dann zumindest für diese sich verändernden Umstände geschlossen werden, daß sie nicht allzu wohlbefindensrelevant sind. Allerdings zeigte sich gerade für den Bereich der Gesundheit eine deutlicher alterskorrelierte Abnahme der Zufriedenheiten im höheren Lebensalter, in dem v.a. vermehrt gesundheitliche Einbußen stattfinden (vgl. Kapitel 1.2.1), bei dennoch hohen Stabilitätskoeffizienten und relativ geringen Veränderungsvarianzen. Dieses erscheint mit Blick auf theoretische Konzepte zur Zufriedenheit besonders relevant: Allgemein entwickelt sie sich mit dem Alter nach unten und reagiert damit offenbar auf alterstypische äußere Veränderungen, jedoch behalten Personen dabei weitgehende ihre *relative* Zufriedenheit im Vergleich zu ihren jeweiligen Altersgenossen. Ein Hochzufriedener bleibt demnach sozusagen stets relativ hochzufrieden, auch wenn beispielsweise gesundheitliche Probleme seine Zufriedenheit mindern, – dieses würde m.E. sehr für die trait-theoretischen Sichtweisen sprechen: Es ist nicht so, daß alle Veränderung durch Neuadjustierungen internaler Soll-Zustände gänzlich adaptiv reguliert werden kann, aber es scheint so, daß über die Lebensspanne hinweg bei den Personen eine gewisse dis-

positionelle Tendenz zu mehr oder weniger Zufriedenheit besteht.

Für ein Anwachsen der Stabilität von Zufriedenheit im höheren Alter, wie es im Kapitel 1.2.3.4 v.a. mit einer möglicherweise stärker retrospektiven Fokussierung bei der Generierung von Lebenszufriedenheitsurteilen und mit im Lebenslauf gewachsener adaptiver Kompetenz hypothetisch begründet wurde, konnten keine Hinweise gefunden werden. Die Hypothese muß im Fazit dieser Untersuchung schlicht als nicht bestätigt gewertet werden, die Alten waren hier nicht stabiler in ihren Zufriedenheitsurteilen, als jüngere Befragungspersonen.

Ähnliches ist, mit leichten Einschränkungen, bezüglich der als Gewinner-und-Verlierer-Hypothese bezeichneten Vermutung zu sagen: Die Zufriedenheiten waren nicht durch eine mit höherem Alter fortschreitende Differenzierung der intraindividuellen Verläufe gekennzeichnet, es fand sich kein deutlicher Hinweis darauf, daß viele Fälle mit größeren Zufriedenheitsverlusten einhergehen mit vielen Fällen, bei denen die Zufriedenheit im Alter wächst. Die besagten Einschränkungen beziehen sich auf die oben bereits diskutierten geringfügigeren Hinweise darauf, daß eine solche Entwicklungsdynamik in schwacher Ausprägung vielleicht doch vorhanden sein könnte. Die Gewinner-und-Verlierer-Hypothese sollte deshalb aus den Ergebnissen der Untersuchung heraus sozusagen nicht hundertprozentig zurückgewiesen, sondern diese Zurückweisung mit Fragezeichen versehen und die Hypothese an zukünftige spezifischere Forschung verwiesen werden.

Bezüglich aller somit aufgeführten hauptsächlichen Resultate der vorgestellten Untersuchung könnten natürlich auch gewisse prinzipielle Vorbehalte geäußert werden. Diese bestünden m.E. hinsichtlich der Validität der verwendeten Zufriedenheitsratings, der Genauigkeit der Ein-Jahres-Zeitintervalle und der Einbindung der Zufriedenheiten in komplexere Abhängigkeitsgefüge.

Es könnte also zunächst gefragt werden, ob es sich bei dem, was vorne als Zufriedenheit bezeichnet wurde und auch bei den True-Scores der Strukturgleichungsmodelle tatsächlich um Zufriedenheiten handelt. Diese Frage war hier nicht zu klären, alle Aussagen in der Diskussion der Ergebnisse erfolgten gewissermaßen im Vertrauen darauf, daß die Befragungspersonen auf die Frage nach Zufriedenheit auch ihre Zufriedenheit und nicht etwas ganz anderes beurteilen. Allerdings berichtet Veenhoven (1996) in seinem Review zur Zufriedenheitsforschung auch von empirischen Studien zur Validität dieser Items, in denen selbige durchaus positiv bewertet wurde.

Die Frage nach der Genauigkeit der Zeitintervalle der untersuchten Zeitreihe ist mit Blick auf die theoretische Fragestellung der Untersuchung zu stellen: Welche Veränderungsdauer interessiert überhaupt? Möglicherweise ergäbe die Untersuchung von Zufriedenheitsveränderungen in Wochen- oder Tagesabständen einen ganz anderen Eindruck von der Stabilität des Merkmals. Es wurde aber vorne (vgl. Kapitel 1.2.3.3) das inhaltliche Interesse dieser Untersuchung auf die längerfristige Stabilität des Merkmals gelenkt: Daß kurzfristige Ausschläge des Wohlbefindens reaktiv infolge von kurzfristigen Veränderungen relevanter Lebensumstände stattfinden, ist anzunehmen – die Annahme adaptiver Regula-

tion bezieht sich ja gerade auf die längerfristigen „Rückveränderungen“ der solchermaßen kurzfristigen Zufriedenheitsneubewertungen. Die in dieser Untersuchung hinterfragte intraindividuelle Merkmalsstabilität ist die auf lange Sicht, und diesbezüglich erscheinen die Jahresintervalle als durchaus vernünftige Zeiteinteilung der Meßwiederholungen. Ganz abgesehen davon konnte hier ja über diese Zeiteinteilung nicht vorweg entschieden werden, da es sich um die Sekundäranalyse eines vorhandenen Datensatzes handelt.

Die sekundäranalytische Natur der Untersuchung ist auch hinsichtlich des letzten der obengenannten einschränkenden Gesichtspunkte zu bedenken: Es wäre wünschenswert und lohnend, die Untersuchung der Stabilität von Zufriedenheit gewissermaßen zu Verzahnen mit denjenigen Variablen, deren Relevanz für die intraindividuelle Zufriedenheitsentwicklung anzunehmen ist. Es wurden bereits solche Variablen genannt: Persönlichkeitskonstrukte wie Extra-/Introversion und Neurotizismus einerseits und die Veränderung tatsächlicher, objektiver relevanter Lebensumstände der Personen andererseits sind natürlich wichtige „Kandidaten“, von denen die Stabilität der Zufriedenheitsurteile abhängen könnte. Jedoch sind zum einen im SOEP kaum psychologische Variablen erhoben – die genannten Trait-Ausprägungen sind dort natürlich nicht gemessen – und wäre zum anderen, bezüglich relevanter objektiver Lebensbedingungen, auch diese Relevanz vorab festzustellen. Denn Untersuchungen zum Zusammenhang von Wohlbefinden mit spezifischen psychologischen, sozialen und gesundheitlichen Variablen gibt es in großer Zahl, jedoch resultiert aus ihnen bislang nicht *der* Zufriedenheitsprädiktor schlechthin (vgl. Diener et al., 1999; Veenhoven, 1996).

Die Untersuchung der Abhängigkeit intraindividuelle Zufriedenheitsstabilität von spezifischen Variablen liefe also eher darauf hinaus, mehrere Einflußgrößen zu benennen, die zusammengenommen möglicherweise in relevantem Ausmaß varianzaufklärend wären und es wäre dann wieder zu fragen, ob im verwendeten Datensatz diese Informationen in hinreichender Spezifität und Genauigkeit vorhanden sind. Beispielsweise kann ein solch umfassender Survey wie das SEOP natürlich kaum allzu detaillierte, „harte“ Gesundheitsdaten (Arztdiagnosen, Anamnesedaten) enthalten, gerade diese scheinen aber im Zusammenhang mit der Entwicklung der Zufriedenheit älterer Menschen besonders interessant. Kurzum: Die Untersuchung der Abhängigkeiten der Zufriedenheitsstabilität von psychischen, sozialen oder gesundheitlichen Einflußgrößen ist selbstverständlich von großer theoretischer Relevanz für die Zufriedenheitsforschung – damit ist aber m.E. kein wirklicher Einwand gegen die hier dargestellte Untersuchung genannt, sondern sind schlicht unterschiedliche Fragestellungen angesprochen, die jeweils für sich genommen berechtigt sind. Die hier vorgelegte Untersuchung zielte auf die statistische Analyse vorhandener Merkmalsstabilität als solcher (und sonstiger entwicklungs-dynamischer Effekte), wofür das SOEP, mit seinen großen Fallzahlen und über einen langen Zeitraum hinweg fortgeführten Meßwiederholungen, eine lohnende Datenbasis bot, die es für die Untersuchung spezifischer Zusammenhänge infolge der naturgemäß eingeschränkten Spezifität der erhobenen Informationen nicht geboten hätte.

Abschließend ist ein weiterer möglicher Effekt zu erwähnen, der in dieser Untersuchung nicht berücksichtigt wurde: Zur Erklärung des Zufriedenheitsparadoxes könnte auch eine spezifische Selektivität der in sozialwissenschaftlichen Studien untersuchten Stichproben Älterer beitragen. Denn möglicherweise bleiben v.a. diejenigen lange am Leben und/oder sind für sozialwissenschaftliche Datenerhebungen zugänglich, deren Lebensumstände durch geringe Einbußen und Verschlechterungen geprägt sind. Dieses würde dann zu einer positiven Selektion von Probanden mit solchermaßen „günstigen“ Bedingungen und deswegen hoher und stabiler Zufriedenheit führen (vgl. dazu Baltes et al., 1988). Jedoch ist Vorsicht angebracht, wenn deshalb ein positiver Bias von Zufriedenheitsmittelwerten in den älteren Stichproben angenommen würde: Eigentlich muß zwischen selektiver Mortalität im Wortsinne und sonstigem Stichprobenausfall, der mit reduzierter Zufriedenheit einhergeht, unterschieden werden. Denn die Verstorbenen haben auch keine schlechte Zufriedenheit mehr, sie gehören überhaupt nicht mehr zur Grundgesamtheit der mehr oder weniger Zufriedenen, und die für eine Stichprobe gefundenen Ergebnisse können nicht dadurch verzerrt sein, daß ihr keine Verstorbenen angehören. Anders dagegen diejenigen, die z.B. aufgrund problematischer Lebensumstände nicht in die Stichprobe aufgenommen werden konnten: Diese Mitglieder der Grundgesamtheit aller mehr oder weniger Zufriedenen sind in der Stichprobe dann nicht adäquat repräsentiert, so daß eine daraus resultierende Verzerrung der Ergebnisse denkbar ist. Eine solchermaßen sich ergebende Selektivitätshypothese könnte also auch (quer- und längsschnittliche) Stabilität von Zufriedenheitsmittelwerten über das höhere Lebensalter hinweg mit erklären: Es finden im höheren Alter viele Zufriedenheitsverluste statt, welche jedoch in den Stichproben nicht auftauchen. Diese Hypothese wurde aber hier nicht untersucht, weil ein solcher Effekt mit dem verwendeten Datenmaterial m.E. kaum einigermaßen eindeutig nachweisbar wäre. Denn die Untersuchung von Selektivitätseffekten steht vor dem prinzipiellen Problem, daß dadurch Aussagen über die Merkmale von Daten angestrebt werden, die gar nicht erhoben wurden: „Unfortunately ...selectivity analyses are confronted with a fundamental paradox: In order to optimally document the degree and nature of selectivity, precisely the information that is missing must be known“ (Little et al., 2000, 190). Hiermit ist also noch ein Paradox benannt und für die weitere gerontologische Wohlbefindensforschung könnte der Versuch lohnen, es aufzulösen und wie auch immer empirische Erkenntnisse darüber zu gewinnen, wie selektiv untersuchte Stichproben hinsichtlich Indikatoren des SWB sind.

Damit ist die vorgelegte Untersuchung an ihrem Ende angelangt. Wer auch immer bis hierher gelesen hat, wird selbiges hoffentlich nicht herbeigesehnt haben.

Anhang A *Ergänzungen zum Kapitel 2*

A.1 *Zusammenhang zwischen Meßwertkorrelationen und Merkmalsstabilität*

Im folgenden soll genauer dargestellt werden, wie Korrelationen der Meßwerte bei Meßwiederholungen von der Reliabilität der Messungen und der Stabilität des gemessenen Merkmals abhängen. Dazu werden zwei Messungen eines Merkmals unter Zugrundelegung des Modells der klassischen Testtheorie (siehe z.B. bei Stumpf, 1996; Steyer & Eid, 1993) betrachtet: Durch Gleichung 2.22 in Kapitel 2.2.2 wurde ein Meßwert X_{ti} eines Merkmalsträgers i zum Meßzeitpunkt t als Summe von True-Scores und Meßfehler spezifiziert. Zur Betrachtung des Zusammenhangs zwischen zwei Meßzeitpunkten t_1 und t_2 (mit $t_1 < t_2$) kann Gleichung 2.22 noch einmal wie folgt ausgedrückt werden (der Einfachheit halber wird im folgenden der Index i weggelassen):

$$\begin{aligned} X_1 &= T_1 + E_1 \\ X_2 &= T_2 + E_2 \end{aligned} \tag{A.1}$$

Nach den Axiomen der klassischen Testtheorie gilt dann:³⁵:

$$\begin{aligned} \mu_{E1} &= \mu_{E2} = 0 \\ \sigma_{E1E2} &= \sigma_{T1E1} = \sigma_{T2E2} = \sigma_{T1E2} = \sigma_{T2E1} = 0 \end{aligned} \tag{A.2}$$

A.2 spezifiziert also die Unabhängigkeit von True-Scores und Meßfehlern (sowohl desselben, als auch unterschiedlicher Zeitpunkte), die Unabhängigkeit der Meßfehler E_1 und E_2 , sowie Meßfehlermittelwerte gleich Null. Letztere Festlegung bedingt, daß die (Populations-)Mittelwerte der Meßwerte X_1 und X_2 gleich den jeweiligen True-Score-Mittelwerten μ_{T1} und μ_{T2} sind und man kann nun auch diese beiden Mittelwerte gleich Null setzen, wenn man annimmt, bei den X - und T -Werten handele es sich um Abweichungen vom jeweiligen (Populations-)Mittelwert bzw., mit anderen Worten, die X - und T -Skalen seien jeweils auf den Mittelwert Null verschoben. Dadurch vereinfachen sich die folgenden formalen Ausdrücke, ohne daß die Gültigkeit der Ableitungen eingeschränkt würde.

Unter Zugrundelegung des durch A.1 spezifizierten Modells können die Reliabilitäten der beiden Messungen X_1 und X_2 als Anteil der jeweiligen True-Score- an den Meßwertvarianzen ausgedrückt werden:

³⁵Es gilt die in Kapitel 2.2.1 und 2.2.2 befolgte Terminologie, d.h. $\mu_{...}$ bezeichnet Mittelwerte, $\sigma_{...}^2$ bezeichnet Varianzen und $\sigma_{...}$ Kovarianzen. Insbesondere ist σ_j^2 die Varianz der Meßwerte X_j , während $\sigma_{E_j}^2$ die Varianz des Meßfehlers E_j und $\sigma_{T_j}^2$ die Varianz der True-Scores T_j bezeichnet.

$$\text{Rel}_1 = \frac{\sigma_{T1}^2}{\sigma_1^2} = \frac{\sigma_{T1}^2}{\sigma_{T1}^2 + \sigma_{E1}^2} \quad \text{und} \quad \text{Rel}_2 = \frac{\sigma_{T2}^2}{\sigma_2^2} = \frac{\sigma_{T2}^2}{\sigma_{T2}^2 + \sigma_{E2}^2} \quad (\text{A.3})$$

Wegen der Unabhängigkeit der Meßfehler beruht natürlich die Kovarianz der Meßwerte gänzlich auf der Kovarianz der True-Scores, wie die folgende Ableitung der Meßwertkovarianz zeigt (die Ausdrücke $E[\dots]$ bezeichnen hier Erwartungswerte):

$$\begin{aligned} \sigma_{12} &= E[X_1 X_2] \\ &= E[(T_1 + E_1)(T_2 + E_2)] \\ &= E[T_1 T_2 + T_1 E_2 + T_2 E_1 + E_1 E_2] \\ &= \sigma_{T_1 T_2} + \sigma_{T_1 E_2} + \sigma_{T_2 E_1} + \sigma_{E_1 E_2} \\ &= \sigma_{T_1 T_2} \end{aligned} \quad (\text{A.4})$$

Die Korrelation der Meßwerte ist nun allerdings nicht mit der Korrelation der wahren Werte T_1 und T_2 identisch (die Ausdrücke σ_{\dots}^2 bezeichnen Varianzen):³⁶

$$\begin{aligned} \rho_{12} &= \frac{\sigma_{12}}{\sigma_1 \sigma_2} \\ &= \frac{\rho_{T_1 T_2} \sigma_{T_1} \sigma_{T_2}}{\sigma_1 \sigma_2} \\ &= \rho_{T_1 T_2} \sqrt{\text{Rel}_1 \text{Rel}_2} \end{aligned} \quad (\text{A.5})$$

Die Ableitung A.5 zeigt: Die Korrelation der Meßwerte X_1 und X_2 ist gleich dem Produkt der Korrelation der wahren Werte T_1 und T_2 mit dem geometrischen Mittel der beiden Reliabilitäten. Nur wenn beide Reliabilitäten gleich 1 sind, d.h. wenn fehlerfrei gemessen wurde, ist die Meßwert- gleich der True-Score-Korrelation, ansonsten, bei $\text{Rel}_1 < 1$ und/oder $\text{Rel}_2 < 1$, liegt die Meßwertkorrelation *unter* der True-Score-Korrelation.

Dieses gilt unter den oben in A.2 spezifizierten Annahmen der klassischen Testtheorie – insbesondere der Unkorreliertheit der Meßfehler untereinander und mit den wahren Werten. Wenn diese Annahmen gelockert werden, dann könnte die Meßwertkorrelation theoretisch auch über der True-Score-Korrelation liegen. V.a. könnte wohl die Annahme der

³⁶Beim Umformen werden folgende Beziehungen benutzt:

Korrelation der True-Scores:

$$\rho_{T_1 T_2} = \frac{\sigma_{T_1 T_2}}{\sigma_{T_1} \sigma_{T_2}}$$

Somit gilt für die Kovarianz der True-Scores:

$$\sigma_{T_1 T_2} = \rho_{T_1 T_2} \sigma_{T_1} \sigma_{T_2}$$

Unabhängigkeit der Meßfehler bei Wiederholungsmessungen in Zweifel gezogen werden, wenn Grund zu der Annahme besteht, daß reliabilitätsmindernde Fehlerquellen zeitlich invariant bestehen bleiben. In diesem Fall würde also in A.2 die Voraussetzung $\sigma_{E_1E_2} = 0$ nicht mehr gelten. Die Konsequenz für den Zusammenhang von True-Score- und Meßwertkorrelation ist leicht einsehbar. Zunächst würde Gleichung A.4 wie folgt verändert:

$$\sigma_{12} = \sigma_{T_1T_2} + \sigma_{E_1E_2} \quad (\text{A.6})$$

Und somit ergibt sich – anstelle von A.5 – für die Korrelation von X_1 und X_2 :³⁷

$$\begin{aligned} \rho_{12} &= \frac{\sigma_{12}}{\sigma_1\sigma_2} \\ &= \frac{\sigma_{T_1T_2} + \sigma_{E_1E_2}}{\sigma_{X_1}\sigma_{X_2}} \\ &= \frac{\rho_{T_1T_2}\sigma_{T_1}\sigma_{T_2} + \rho_{E_1E_2}\sigma_{E_1}\sigma_{E_2}}{\sigma_1\sigma_2} \\ &= \rho_{T_1T_2}\sqrt{\text{Rel}_1\text{Rel}_2} + \rho_{E_1E_2}\sqrt{(1 - \text{Rel}_1)(1 - \text{Rel}_2)} \end{aligned} \quad (\text{A.7})$$

Der Zusammenhang von Meßwert- und True-Score-Korrelation, der in A.7 zum Ausdruck kommt, ist nicht mehr so unmittelbar anschaulich, wie noch in A.6. Immerhin wird ersichtlich, daß ρ_{12} auch über $\rho_{T_1T_2}$ liegen kann. Wenn man z.B. der Einfachheit halber annimmt, die Reliabilität bliebe bei den Wiederholungsmessungen konstant, d.h. $\text{Rel}_1 = \text{Rel}_2$, dann wird die Meßwertkorrelation kleiner/gleich/größer der True-Score-Korrelation je nachdem, ob letztere größer/gleich/kleiner der Korrelation der Meßfehler ist:

- Aus $\rho_{T_1T_2} = \rho_{E_1E_2}$ folgt:

$$\rho_{12} = \rho_{T_1T_2}\text{Rel} + \rho_{T_1T_2}(1 - \text{Rel}) = \rho_{T_1T_2}$$

- Der Fall $\rho_{T_1T_2} > \rho_{E_1E_2}$ – oder, anders formuliert: $\rho_{E_1E_2} = \rho_{T_1T_2} - a$ mit $a > 0$ – ergibt:

$$\rho_{12} = \rho_{T_1T_2}\text{Rel} + (\rho_{T_1T_2} - a)(1 - \text{Rel}) = \rho_{T_1T_2} + (\text{Rel} - 1)a$$

Und somit, da $\text{Rel} < 1$:

$$\rho_{12} < \rho_{T_1T_2}$$

- Der umgekehrte Fall $\rho_{T_1T_2} < \rho_{E_1E_2}$ – oder, anders formuliert: $\rho_{E_1E_2} = \rho_{T_1T_2} + a$ mit $a > 0$ – ergibt entsprechend:

³⁷Siehe wiederum Fußnote 36. Außerdem gilt:

$$1 - \text{Rel}_1 = \frac{\sigma_{E_1}^2}{\sigma_1^2} \quad \text{und} \quad 1 - \text{Rel}_2 = \frac{\sigma_{E_2}^2}{\sigma_2^2}$$

$$\rho_{12} = \rho_{T_1 T_2} \text{Rel} + (\rho_{T_1 T_2} + a)(1 - \text{Rel}) = \rho_{T_1 T_2} + (1 - \text{Rel})a$$

Und somit:

$$\rho_{12} > \rho_{T_1 T_2}$$

In einem Meßmodell mit invarianter Reliabilität und korrelierten Meßfehlern läge die Meßwertkorrelation also dann über der True-Score-Korrelation, wenn letztere kleiner ist, als die Korrelation der Meßfehler. Würde das Modell um variable Reliabilitäten erweitert, so würden natürlich die entsprechenden Zusammenhänge komplizierter.

An dieser Stelle aber soll es genügen, gezeigt zu haben, daß das Modell der klassischen Testtheorie eine generelle Unterschätzung von Stabilität durch Meßwertkorrelationen impliziert, wenn die Reliabilität der Messungen nicht berücksichtigt wird. Dieses gilt tendenziell auch dann, wenn bei hoher Merkmalsstabilität (der „wahren“ Werte) geringe Meßwertkorrelationen vorliegen.

A.2 Identifikation des QMSM

Das in Kapitel 2.2.2.1 vorgestellte autoregressive Modell erster Ordnung (sogenanntes Quasi-Markov-Simplex-Modell, vgl. Abbildung 2.3.B) wurde durch die Modellgleichungen 2.22 und 2.24 spezifiziert. Es wurde bereits auf Veröffentlichungen verwiesen (Jöreskog, 1970; Werts et al., 1971), die gezeigt hatten, daß in dem Modell lediglich die „inneren“ Reliabilitäten und Stabilitäten, d.h. also in einem Modell mit k Meßzeitpunkten die Reliabilitäten der Messungen zu $t = 2$ bis $t = k - 1$, sowie die Autokorrelationen erster Ordnung zwischen den True-Scores $T_2 \dots T_{k-1}$, identifiziert sind: „Generalizing to multiple wave panel studies, we may state that, when the assumptions of the Wiley and Wiley structural model [*gemeint ist das oben dargestellte Quasi-Markov-Simplex-Modell, O.S.*] are given, error variances, true score variances, and standardized regression coefficients are identified for all but the first and last measures“ (Werts et al., 1971, 111). Da das QMSM das wichtigste Modell für die vorne dargestellten Stabilitätsanalysen war, soll diese Modellidentifikation im folgenden noch einmal explizit gezeigt werden.

Die Identifizierbarkeit der einzelnen Modellparameter wird dadurch festgestellt, daß das Gleichungssystem der durch die sogenannte implizite Kovarianzmatrix (implied covariance matrix) des Modells spezifizierten Zusammenhänge zwischen Varianzen und Kovarianzen der beobachteten Variablen und den zu schätzenden Modellparametern nach letzteren hin aufgelöst wird. Es gilt: „If an unknown parameter in θ can be written as a function of one or more elements of Σ , that parameter is identified“ (Bollen, 1989, 89, mit θ ist hier der Vektor der Modellparameter und mit Σ die Varianz-Kovarianz-Matrix der beobachteten Variablen bezeichnet). Unter Anwendung allgemein bekannter Regeln für die Dekomposition von Varianzen und Kovarianzen (z.B. Hays, 1994) lassen sich für das QMSM diese

entsprechenden Gleichungen wie folgt zusammenfassen (zur Veranschaulichung vgl. wiederum Abbildung 2.3.B, Kapitel 2.2.2.1):

Für die Varianzen σ_t^2 der beobachteten Variablen X_{ti} zum Meßzeitpunkt t gilt (vgl. auch Gleichung 2.26b, S. 100):

$$\sigma_t^2 = \left[\prod_{i=1}^{t-1} \beta_i^2 \right] \sigma_{T1}^2 + \sum_{j=2}^{t-1} \left[\prod_{k=j}^{t-1} \beta_k^2 \right] \sigma_{Dj}^2 + \sigma_{Dt}^2 + \sigma_{Et}^2 \quad (\text{A.8})$$

(für: $\beta_0 = 1, \sigma_{D1}^2 = 0$ und $\sum_2^{t-1} = 0$ wenn $t - 1 < 2$)

Für die Kovarianzen $\sigma_{tt'}$ der beobachteten Variablen X_{ti} zu den Meßzeitpunkten t und t' , mit $t < t'$, gilt:

$$\sigma_{tt'} = \left[\prod_{i=0}^t \beta_i \prod_{j=1}^{t'-1} \beta_j \right] \sigma_{T1}^2 + \sum_{l=1}^t \left[\prod_{i=l}^{t-1} \beta_i \prod_{j=l}^{t'-1} \beta_j \right] \sigma_{Dl}^2 \quad (\text{A.9})$$

(für: $\beta_0 = 1, \sigma_{D1}^2 = 0$)

Die beiden Formeln A.8 und A.9 sind wohl ohne weiteres wenig anschaulich und offenen die möglichen Auflösungen für die unbekannt Parameter $\beta_t, \sigma_{Dt}^2, \sigma_{T1}^2, \sigma_{Et}^2$ nicht auf den ersten Blick. Zur Veranschaulichung zeige ich die Auflösungen am Beispiel eines QMSM mit $k = 4$, d.h. mit vier Meßzeitpunkten:

Aus Gleichung A.8 erhält man die folgenden Gleichungen für die Varianzen der vier beobachteten Variablen:

$$\sigma_1^2 = \sigma_{T1}^2 + \sigma_{E1}^2 \quad (\text{A.8a})$$

$$\sigma_2^2 = \beta_1^2 \sigma_{T1}^2 + \sigma_{D2}^2 + \sigma_{E2}^2 \quad (\text{A.8b})$$

$$\sigma_3^2 = \beta_1^2 \beta_2^2 \sigma_{T1}^2 + \beta_2^2 \sigma_{D2}^2 + \sigma_{D3}^2 + \sigma_{E3}^2 \quad (\text{A.8c})$$

$$\sigma_4^2 = \beta_1^2 \beta_2^2 \beta_3^2 \sigma_{T1}^2 + \beta_2^2 \beta_3^2 \sigma_{D2}^2 + \beta_3^2 \sigma_{D3}^2 + \sigma_{D4}^2 + \sigma_{E4}^2 \quad (\text{A.8d})$$

Für die Kovarianzen der beobachteten Variablen ergibt sich aus Gleichung A.9:

$$\sigma_{12} = \beta_1 \sigma_{T1}^2 \quad (\text{A.9a})$$

$$\sigma_{13} = \beta_1 \beta_2 \sigma_{T1}^2 \quad (\text{A.9b})$$

$$\sigma_{14} = \beta_1 \beta_2 \beta_3 \sigma_{T1}^2 \quad (\text{A.9c})$$

$$\sigma_{23} = \beta_1^2 \beta_2 \sigma_{T1}^2 + \beta_2 \sigma_{D2}^2 \quad (\text{A.9d})$$

$$\sigma_{24} = \beta_1^2 \beta_2 \beta_3 \sigma_{T1}^2 + \beta_2 \beta_3 \sigma_{D2}^2 \quad (\text{A.9e})$$

$$\sigma_{34} = \beta_1^2 \beta_2^2 \beta_3 \sigma_{T1}^2 + \beta_2^2 \beta_3 \sigma_{D2}^2 + \beta_3 \sigma_{D3}^2 \quad (\text{A.9f})$$

Auflösungen für unbekannte Modellparameter ergeben sich wie folgt:

- Identifikation von β_2 und β_3 :

A.9a in A.9b ergibt:

$$\sigma_{13} = \sigma_{12}\beta_2 \implies \beta_2 = \sigma_{13}/\sigma_{12}$$

A.9b in A.9c ergibt:

$$\sigma_{14} = \sigma_{13}\beta_3 \implies \beta_3 = \sigma_{14}/\sigma_{13}$$

- Identifikation von σ_{D3}^2 :

A.9e in A.9f ergibt:

$$\sigma_{34} = \beta_2\sigma_{24} + \beta_3\sigma_{D3}^2 \implies \sigma_{D3}^2 = (\sigma_{34} - \beta_2\sigma_{24})/\beta_3$$

Und mit den Auflösungen für β_2 und β_3 erhält man:

$$\sigma_{D3}^2 = (\sigma_{34} - \sigma_{13}\sigma_{24}/\sigma_{12})/(\sigma_{13}/\sigma_{14})$$

Ein Wert für σ_{D3}^2 ergibt sich auch durch Einsetzen von A.9d in A.9f (σ_{D3}^2 ist überidentifiziert).

- Identifikation von σ_{E2}^2 und σ_{E3}^2 :

A.9d kann umgeschrieben werden:

$$\sigma_{23}/\beta_2 = \beta_1^2\sigma_{T1}^2 + \sigma_{D2}^2$$

Diese Umformung eingesetzt in A.8b ergibt:

$$\sigma_2^2 = \sigma_{23}/\beta_2 + \sigma_{E2}^2 \implies \sigma_{E2}^2 = \sigma_2^2 - \sigma_{23}\sigma_{12}/\sigma_{13}$$

Für σ_{E3}^2 kann nun ein Wert dadurch gefunden werden, daß wiederum die obige Umformung von A.9d und die bereits gefundene Lösung für σ_{D3}^2 in A.8c eingesetzt werden, oder dadurch, daß A.9f analog zur Bestimmung von σ_{E2}^2 umgeformt und in A.8c eingesetzt wird – letztere Auflösung ergibt:

$$\sigma_3^2 = \sigma_{34}/\beta_3 + \sigma_{E3}^2 \implies \sigma_{E3}^2 = \sigma_3^2 - \sigma_{34}\sigma_{13}/\sigma_{14}$$

Wie man nun erkennen kann, sind für σ_{T1}^2 , β_1 , σ_{D2}^2 , σ_{D4}^2 , σ_{E1}^2 und σ_{E4}^2 noch keine Lösungen gefunden und können auch ohne zusätzliche Restriktionen keine gefunden werden: σ_{E1}^2 kommt nur in Gleichung A.8a gemeinsam mit σ_{T1}^2 vor und könnte durch diese Gleichung nur bestimmt werden, wenn σ_{T1}^2 aus den anderen Gleichungen bestimmt würde. In denen aber taucht σ_{T1}^2 stets nur entweder im Produkt $\beta_1\sigma_{T1}^2$ (in A.9a bis A.9c) oder in der Kombination $\beta_1^2\sigma_{T1}^2 + \sigma_{D2}^2$ (in A.8b bis A.8d und A.9d bis A.9f) auf und man kann sich anhand dieser Gleichungen leicht davon überzeugen, daß diese beiden Ausdrücke dann jeweils nur als ganze berechenbar sind und keiner der drei in ihnen enthaltenen Parameter

einzel identifiziert ist. Auch σ_{D4}^2 und σ_{E4}^2 kommen wiederum nur gemeinsam in der Gleichung A.8d vor und können aus dieser nicht getrennt berechnet werden.

Wichtig für Analysen der Merkmalsstabilität ist nun jedoch noch die Identifikation „indirekter“ Modellparameter, nämlich der Stabilitätskoeffizienten ρ_t ($t = 1, \dots, k-1$), sowie der True-Score-Varianzen σ_{Tt}^2 ($t = 2, \dots, k$) und Reliabilitätskoeffizienten Rel_t ($t = 1, \dots, k$). Erstere sind die Autokorrelationen erster Ordnung, d.h. die standardisierten β -Koeffizienten (vgl. Gleichung 2.25), und letztere die quadrierten Korrelationen der True-Score-Werte T_{ti} mit den Meßwerten X_{ti} bzw. der jeweilige Anteil der True-Score-Varianz σ_{Tt}^2 an der Meßwertvarianz σ_t^2 (vgl. Gleichung A.3).

Da weder σ_{T1}^2 , noch σ_{E1}^2 identifiziert sind, kann auch für den ersten Meßzeitpunkt nicht der Anteil der True-Score- an der Meßwertvarianz ohne weitere Information bestimmt werden, Rel_1 ist also nicht identifiziert. Identifiziert ist dann allerdings die zweite True-Score-Varianz σ_{T1}^2 : Es handelt sich eben um jenen Ausdruck $\beta_1^2 \sigma_{T1}^2 + \sigma_{D2}^2$, der, wie oben dargelegt, z.B. aus Gleichung A.9d (und der Lösung für β_2) berechnet werden kann. Wenn σ_{T2}^2 bestimmt ist, dann kann, da auch β_2 und σ_{D3}^2 identifiziert sind, σ_{T3}^2 berechnet werden. Da also die True-Score-Varianzen der beiden „inneren“ Meßzeitpunkte (ebenso wie die dazugehörigen Fehlervarianzen) identifiziert sind, können für diese beiden Messungen auch die Reliabilitäten Rel_2 und Rel_3 berechnet werden. Die True-Score-Varianz σ_{T4}^2 ist dagegen nicht mehr berechenbar, da σ_{D4}^2 nicht identifiziert ist, folglich bleibt also auch Rel_4 unidentifiziert.

Die Berechnung der Stabilitätskoeffizienten, d.h. standardisierter Autoregressionskoeffizienten, setzt nicht nur die Identifikation der unstandardisierten β -Koeffizienten voraus, sondern auch der beiden True-Score-Varianzen der jeweiligen Kriteriums- und Prädiktorwerte (vgl. Gleichung 2.25). Deshalb sind ρ_1 und ρ_3 , d.h. die beiden „äußeren“ Korrelationen der Markov-Kette, nicht identifiziert (da eben die äußeren True-Score-Varianzen nicht identifiziert sind) und ist lediglich der „innere“ Stabilitätskoeffizient ρ_2 berechenbar.

Diese Grundregel, daß lediglich die „äußeren“ Reliabilitäts- und Stabilitätskoeffizienten nicht identifiziert und die „inneren“ berechenbar sind, gilt auch für längere Markov-Ketten: Weitere Messungen erhöhen lediglich die Anzahl der Elemente in der impliziten Kovarianzmatrix des Modells bzw. der durch A.8 und A.9 spezifizierten Gleichungen, diese weiteren Gleichungen können leicht nach dem durch das in A.8a bis A.9f erkennbaren Schema fortgeschrieben werden, die Auflösung der darin enthaltenen Parameter kann dann jeweils ebenfalls nach dem Schema erfolgen, wie es oben für $k = 4$ dargestellt wurde.

Zudem zeigt die dargestellte Auflösung der Gleichungen A.8a bis A.9f, daß alle Parameter für $t = 1$ dann berechnet werden können, wenn entweder σ_{T1}^2 oder σ_{E1}^2 durch weitere Restriktionen identifizierbar gemacht werden: Liegt z.B. für σ_{E1}^2 eine Lösung vor, so kann über A.8a auch σ_{T1}^2 berechnet werden und folglich auch β_1 (z.B. über A.9a) und σ_{D2}^2 . Genauso würden die „letzten“ Parameter σ_{Dk}^2 und σ_{Ek}^2 dann berechenbar, wenn einer der beiden oder aber die letzte True-Score-Varianz σ_{Tk}^2 über Restriktionen identifiziert würde.

A.3 Identifikation des LDCM

Zur Identifikation des durch Abbildung 2.5 bzw. Gleichung 2.31 (Kapitel 2.2.2.2) beschriebenen LDCM sind die durch die modellimplizierte Kovarianzmatrix spezifizierten Zusammenhänge zwischen Varianzen und Kovarianzen der beobachteten Variablen und den zu schätzenden Modellparametern wiederum so aufzulösen, daß letztere vollständig durch erstere ausgedrückt werden können. Die entsprechenden Gleichungen lassen sich einfach wie folgt zusammenfassen:

Für die Varianzen σ_t^2 der beobachteten Variablen X_{ti} zum Meßzeitpunkt t gilt ($\sigma_{\Delta t}^2$ bezeichnet die Varianz der Differenzkomponente Δ_{ti}):

$$\sigma_t^2 = \sum_{j=1}^t \sigma_{\Delta j}^2 + \sigma_{Et}^2 \quad (\text{A.10})$$

Für die Kovarianzen $\sigma_{tt'}$ der beobachteten Variablen X_{ti} zu den Meßzeitpunkten t und t' , mit $t < t'$, gilt:

$$\sigma_{tt'} = \sum_{j=1}^t \sigma_{\Delta j}^2 \quad \text{für alle } t < t' \quad (\text{A.11})$$

Bei insgesamt k Meßzeitpunkten tauchen nun die „letzten“ beiden Modellparameter, $\sigma_{\Delta k}^2$ und σ_{Ek}^2 , *beide zusammen* nur in einer einzigen Gleichung auf, nämlich in der Gleichung für die letzte beobachtete Varianz:

$$\sigma_k^2 = \sigma_{\Delta 1}^2 + \dots + \sigma_{\Delta k}^2 + \sigma_{Ek}^2$$

Ansonsten sind diese beiden Parameter in keiner Gleichung vorhanden, so daß keiner von beiden gänzlich als Funktion der beobachtbaren Varianzen/Kovarianzen ausgedrückt werden kann (da ja auch nach allen möglichen Umformungen des Gleichungssystems stets diese beiden Unbekannten in dieser Gleichung verbleiben). Alle anderen unbekannt Parameter können als Funktion der beobachtbaren Varianzen/Kovarianzen ausgedrückt werden, wie man durch entsprechende Umformung des Gleichungssystems leicht feststellen kann: Für $t = 1$ ergibt A.11 stets $\sigma_{1t'} = \sigma_{\Delta 1}^2$, unabhängig von t' , d.h. gemäß dem LDCM wären alle Populationskovarianzen der Meßwerte des ersten mit denen irgendeines anderen Meßzeitpunkts stets gleich der Varianz der ersten Differenzkomponente (welche umgekehrt dadurch überidentifiziert ist). Entsprechend gilt für $t = 2$ stets $\sigma_{2t'} = \sigma_{\Delta 1}^2 + \sigma_{\Delta 2}^2$, was dann nach Auflösung von $\sigma_{\Delta 1}^2$ ebenfalls aufzulösen wäre, – usw. für $\sigma_{3t'}, \dots, \sigma_{(k-1)t'}$. Die somit erhaltenen Lösungen für $\sigma_{\Delta 1}^2$ bis $\sigma_{\Delta (k-1)}^2$ können dann in die Gleichungen A.10 eingesetzt werden und führen dort zu Lösungen für σ_{E1}^2 bis $\sigma_{E(t-1)}^2$.

Um eine vollständige Schätzbarkeit des Modells zu erreichen, muß also einer der beiden unidentifizierten Parameter $\sigma_{\Delta k}^2$ oder σ_{Ek}^2 durch eine Restriktion identifizierbar gemacht

werden, d.h. der Wert, der aus den Gleichungen A.10 und A.11 nicht berechenbar ist, muß „irgendwie sonst“ festgelegt werden. Dieses leistet z.B. die im Kapitel 2.2.2.2 durch Gleichung 2.33 erwähnte Restriktion gleicher Meßfehlervarianzen, durch die $\sigma_{E_k}^2$ aus den Gleichungen A.10 für die Zeitpunkte $1, \dots, k - 1$ berechnet werden kann.

Anhang B Weitere Ergebnistabellen

Im folgenden werden Ergebnistabellen aufgelistet, die die Ergebnisdarstellungen in den vorherigen Kapiteln ergänzen und vervollständigen sollen. Dabei werden folgende Abkürzungen benutzt:

ALZ = allgemeine Lebenszufriedenheit,

EIZ = Einkommenszufriedenheit,

GEZ = Gesundheitszufriedenheit,

WOZ = Wohnzufriedenheit,

FRZ = Freizeitzufriedenheit.

TABELLE B.1:

Ergänzung Kapitel 3.1: Stichprobengrößen des SOEP nach Substichprobe und Befragungsjahr

Jahr	Westdeutsche		Ausländer		Ostdeutsche		Immigranten		Total	
	HH [†]	Pers. [‡]	HH [†]	Pers. [‡]	HH [†]	Pers. [‡]	HH [†]	Pers. [‡]	HH [†]	Pers. [‡]
1984	4528	9076	1393	3169					5921	12245
1985	4141	8372	1181	2718					5322	11090
1986	3962	8009	1128	2637					5090	10646
1987	3910	7868	1116	2648					5026	10516
1988	3743	7481	1071	2542					4814	10023
1989	3647	7201	1043	2509					4690	9710
1990	3612	7036	1028	2483	2179	4453			6819	13972
1991	3613	6974	1056	2493	2030	4202			6699	13669
1992	3585	6821	1060	2484	2020	4092			6665	13397
1993	3603	6747	1064	2459	1970	3973			6637	13179
1994	3577	6637	1023	2364	1959	3945	236	471	6795	13417
1995	3526	6567	982	2231	1938	3892	522	1078	6968	13768
1996	3485	6454	960	2152	1951	3882	498	1023	6894	13511
1997	3458	6378	931	2089	1942	3844	479	972	6810	13283
1998	3387	6184	898	1961	1886	3730	438	885	7679*	14692*
1999	3325	6045	858	1864	1894	3709	425	838	7392*	13696*

[†] Anzahl befragter Haushalte; [‡] Anzahl befragter Personen.

* Summe inklusive Auffrischungsstichprobe.

Quelle: Infratest Sozialforschung (2001).

TABELLE B.2:

Ergänzung Kapitel 3.1: Mittelwerte und Standardabweichungen der untersuchten Zufriedenheitsmessungen im SOEP (Westdeutsche)

Jahr	ALZ		EIZ		GEZ		WOZ		FRZ	
	<i>m</i>	<i>s</i>	<i>m</i>	<i>s</i>	<i>m</i>	<i>s</i>	<i>m</i>	<i>s</i>	<i>m</i>	<i>s</i>
1984	7.44	2.18	6.68	2.55	5.94	2.76	8.31	2.00	7.67	2.46
1985	7.27	2.13	6.64	2.42	5.99	2.53	8.20	1.94	7.54	2.36
1986	7.33	2.03	6.76	2.29	5.94	2.52	8.20	1.91	7.50	2.39
1987	7.17	2.01	6.79	2.27	5.97	2.47	8.31	1.84	7.64	2.30
1988	7.05	2.08	6.78	2.27	5.83	2.53	8.17	1.83	7.48	2.31
1989	7.06	2.06	6.80	2.22	5.73	2.51	8.19	1.79	7.52	2.25
1990	7.20	1.96	6.88	2.19	5.77	2.39	8.17	1.80	7.47	2.07
1991	7.40	1.82	7.17	2.04	5.81	2.39	8.35	1.63	7.66	2.25
1992	7.23	1.87	6.90	2.10	5.77	2.47	8.30	1.67	7.76	2.19
1993	7.22	1.94	6.87	2.22	5.69	2.39	8.31	1.69	7.57	2.25
1994	7.05	1.94	6.82	2.20	5.60	2.40	8.34	1.73	7.66	2.20
1995	6.98	1.99	6.96	2.10	5.65	2.42	8.22	1.72	7.60	2.02
1996	7.08	1.91	6.94	2.20	5.64	2.36	8.20	1.77	7.94	2.09
1997	6.93	1.91	6.72	2.15	5.62	2.33	8.14	1.76	7.94	2.08
1998	6.92	1.96	6.81	2.16	5.65	2.35	8.05	1.81	7.85	2.10
1999	6.96	1.93	6.88	2.15	5.55	2.36	8.04	1.83	7.99	2.12

TABELLE B.3:
Ergänzung Kapitel 3.2.1: ALZ – Mittelwerte der Ein-Jahres-Altersgruppen

Alter	SOEP-Befragungswelle															
	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99
45	7.6	7.1	7.5	7.3	7.0	6.6	7.7	7.4	7.3	7.1	6.9	6.8	6.9	6.8	6.8	6.8
46	7.3	7.2	7.0	7.2	6.9	6.9	6.9	7.5	7.2	6.8	7.2	6.8	7.0	6.8	6.8	6.8
47	7.6	7.2	7.2	6.8	7.0	7.2	7.2	7.3	7.4	7.2	6.9	7.1	7.1	7.0	6.8	6.8
48	7.8	7.4	7.0	7.0	6.8	7.3	7.3	7.2	7.3	7.6	7.1	6.6	7.1	6.7	7.0	7.0
49	7.4	7.1	7.3	7.0	7.0	6.7	7.4	7.3	7.3	7.1	7.5	7.1	6.8	7.1	6.9	6.9
50	7.3	7.2	7.4	6.9	6.8	6.9	7.3	7.1	7.3	7.3	6.8	7.3	7.2	6.6	7.1	7.1
51	7.4	7.4	7.0	7.2	7.1	7.0	7.3	7.4	7.2	7.3	7.0	6.9	7.3	7.0	6.7	6.7
52	7.0	7.1	6.9	6.7	6.9	6.9	7.2	7.5	7.3	7.1	7.2	6.7	6.9	7.1	7.1	7.1
53	7.5	7.0	7.1	7.0	6.7	7.0	7.0	7.3	7.3	7.3	7.2	7.1	6.8	6.6	7.2	7.2
54	7.3	7.1	7.0	7.0	7.0	6.7	7.2	7.3	7.3	7.2	7.1	6.9	7.2	6.5	6.8	6.8
55	7.7	7.0	7.1	6.7	7.0	6.5	6.7	7.1	7.2	7.3	7.1	6.8	7.0	6.7	6.9	6.9
56	7.4	7.7	7.2	7.1	7.1	6.9	7.2	6.8	7.3	6.8	7.1	6.8	6.9	6.9	7.1	7.1
57	7.2	7.3	7.7	7.1	7.1	7.2	6.7	7.7	7.0	7.4	6.7	7.0	7.0	7.0	7.0	7.0
58	7.4	7.1	7.2	7.7	6.8	7.0	7.1	7.2	7.1	6.9	7.2	6.7	7.0	7.1	7.0	7.0
59	7.4	7.2	7.1	7.2	7.6	6.7	7.1	7.2	7.2	7.2	6.8	7.1	7.1	6.9	7.1	7.1
60	7.5	7.3	7.3	7.3	7.0	7.6	6.8	6.9	7.3	6.8	7.1	7.0	7.4	7.1	7.0	7.0
61	7.4	7.4	7.6	7.4	7.1	7.2	7.8	7.3	7.4	7.4	7.1	7.0	6.9	7.2	7.0	7.0
62	7.9	7.5	7.5	7.5	7.2	7.1	7.2	7.6	7.1	7.3	7.2	7.0	6.9	6.5	7.0	7.0
63	7.4	7.5	7.7	7.5	7.3	7.3	7.4	7.6	7.6	7.2	6.9	7.2	7.1	6.6	6.7	6.7
64	7.2	7.2	7.3	7.3	7.1	7.4	7.6	7.6	7.3	7.5	6.8	6.8	7.2	7.1	7.1	7.1
65	7.2	7.0	7.4	7.6	7.2	7.2	7.6	7.9	7.5	7.5	7.4	7.1	7.0	6.9	6.8	6.8
66	7.4	7.8	7.2	7.5	7.3	7.1	7.5	7.4	7.4	7.4	6.8	7.8	7.0	6.8	6.8	6.8
67	7.1	7.0	8.2	7.0	7.5	7.3	7.0	7.5	7.2	7.5	7.2	7.4	7.6	6.9	6.9	6.9
68	7.0	7.2	7.5	7.6	7.1	7.4	7.2	7.0	7.3	7.6	7.2	7.1	7.3	7.5	6.7	6.7
69	7.4	7.4	7.7	7.4	7.1	6.8	7.5	7.7	7.2	7.7	7.0	7.2	7.0	7.1	7.2	7.2
70	7.8	6.8	7.5	7.3	7.2	7.7	6.7	7.8	7.3	6.8	7.2	7.0	7.4	6.8	6.8	6.8
71	7.4	7.6	7.1	6.9	7.1	7.2	7.7	7.2	7.4	7.4	6.9	7.3	7.3	7.0	6.7	6.7
72	7.7	7.7	7.6	7.4	6.6	6.7	7.1	7.9	6.8	7.4	7.3	6.8	7.4	7.1	7.3	7.3
73	7.4	7.1	7.7	7.2	7.2	6.6	7.3	7.1	7.7	6.8	7.1	7.0	6.9	7.2	7.1	7.1
74	7.6	7.1	7.7	7.1	7.1	6.8	7.3	7.6	6.4	7.4	6.7	7.3	7.3	6.6	7.1	7.1
75	7.2	7.8	6.8	7.3	6.6	7.0	6.7	7.4	7.3	7.1	6.9	6.9	7.0	7.0	6.8	6.8
76	7.4	7.3	8.1	7.4	6.8	7.4	7.0	6.7	7.2	7.2	6.8	6.8	6.9	7.0	7.4	7.4
77	7.7	7.3	7.3	7.0	7.0	6.8	6.6	6.4	6.7	6.8	6.7	6.4	7.0	7.1	6.8	6.8
78	6.9	7.0	7.4	6.5	6.6	7.1	7.0	6.8	6.9	7.1	6.3	6.4	6.4	6.7	6.6	6.6
79	7.4	7.2	7.5	7.6	6.4	7.2	6.7	6.6	7.2	6.4	6.4	6.8	7.0	6.1	7.0	7.0
80	7.0	7.1	7.2	7.3	7.1	6.4	7.1	7.2	6.1	7.0	6.8	6.3	6.1	7.1	5.9	5.9
81	7.0	7.4	7.2	6.6	6.9	6.8	6.8	7.2	6.8	6.6	6.2	6.5	6.3	6.1	6.4	6.4
82	6.7	7.1	6.7	7.5	6.6	7.4	7.3	6.9	7.5	6.6	6.7	6.2	6.9	6.2	5.5	5.5
83	6.2	6.9	7.2	6.9	7.0	6.9	6.6	7.2	6.2	7.1	7.4	5.7	7.1	6.2	6.8	6.8
84	6.4	6.1	7.4	6.5	6.1	7.4	6.9	6.8	7.2	6.3	7.5	6.2	6.8	7.2	7.4	7.4
85	5.8	6.9	6.9	6.4	6.8	7.4	7.2	5.9	6.7	6.7	6.4	7.2	5.8	7.0	6.9	6.9
86	.	6.4	6.8	6.7	5.9	6.3	7.1	7.9	7.0	7.3	6.4	5.4	7.1	5.1	6.6	6.6
87	7.8	-	6.9	6.1	6.6	6.5	-	6.4	6.9	6.5	6.5	5.9	6.5	7.0	6.4	6.4
88	5.8	-	-	6.5	-	-	7.1	-	6.8	-	-	6.6	-	6.4	7.1	7.1
89	-	-	-	-	-	-	-	7.1	-	7.3	-	-	-	-	-	-
90	-	-	-	-	-	-	-	-	-	-	7.0	-	-	-	-	-

TABELLE B.4:
Ergänzung Kapitel 3.2.1: EIZ – Mittelwerte der Ein-Jahres-Altersgruppen

Alter	SOEP-Befragungswelle															
	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99
45	6.6	6.6	6.9	6.9	6.7	6.8	7.1	7.3	7.1	6.7	6.5	6.5	6.5	6.6	5.9	5.9
46	6.5	6.4	6.6	6.9	6.7	6.6	6.9	7.6	7.2	6.8	6.7	6.5	6.9	6.6	6.6	6.6
47	6.4	6.4	6.6	6.7	6.8	6.8	6.7	7.3	7.4	6.9	6.4	6.6	6.7	6.6	6.2	6.2
48	6.7	6.3	6.3	6.6	6.7	7.1	6.9	6.9	6.9	7.3	6.9	6.3	6.7	6.2	6.6	6.6
49	6.5	6.5	6.4	6.3	6.7	6.5	6.9	7.3	6.9	6.9	7.0	6.8	6.2	6.7	6.4	6.4
50	6.4	6.1	6.5	6.5	6.1	6.5	6.8	7.1	7.1	6.6	7.0	6.8	6.8	6.1	6.5	6.5
51	6.2	6.2	6.1	6.5	6.7	6.2	6.7	7.1	7.0	7.0	6.7	6.7	6.8	6.6	6.4	6.4
52	6.0	6.2	6.4	6.1	6.5	6.4	6.3	7.0	6.9	7.0	6.8	6.5	6.8	6.8	6.4	6.4
53	6.7	6.0	6.5	6.3	6.2	6.3	6.7	6.8	7.0	6.8	6.8	6.8	6.4	6.7	6.4	6.4
54	6.3	6.2	6.2	6.4	6.6	6.4	6.6	6.9	6.6	6.5	6.8	6.7	6.6	6.3	6.7	6.7
55	6.5	6.2	6.7	6.1	6.4	6.5	6.4	6.8	6.9	6.4	6.7	6.5	7.1	6.7	6.4	6.4
56	6.8	6.7	6.3	6.7	6.5	6.4	6.5	6.5	6.6	6.4	6.0	6.8	6.7	6.4	6.7	6.7
57	6.3	6.7	7.1	6.6	6.6	6.3	6.2	7.2	6.6	6.5	6.1	6.4	6.4	6.4	6.5	6.5
58	6.5	6.5	6.6	6.9	6.3	6.7	6.2	6.6	6.8	6.3	6.4	6.1	6.4	6.4	6.6	6.6
59	6.6	6.3	6.6	6.5	6.9	6.3	6.7	6.4	6.4	6.5	6.3	6.3	6.3	5.8	6.4	6.4
60	6.8	6.5	6.7	6.8	6.7	6.9	6.1	6.8	6.3	6.6	6.0	6.3	6.9	6.0	6.0	6.0
61	6.4	6.5	7.0	6.8	7.1	6.6	7.2	6.7	6.7	6.4	6.3	6.4	6.5	6.7	6.0	6.0
62	6.9	6.9	6.7	6.8	6.9	6.6	7.1	7.5	6.3	6.4	6.7	6.9	6.4	6.3	6.3	6.3
63	6.9	6.9	7.1	7.0	6.8	7.0	6.7	7.2	7.3	6.5	6.4	6.4	6.9	6.4	6.8	6.8
64	6.8	7.0	6.7	7.0	6.6	7.1	6.7	7.3	6.9	7.2	6.2	6.6	6.7	6.5	6.0	6.0
65	7.1	6.8	7.1	7.0	6.9	6.9	6.9	7.3	7.2	7.0	7.0	6.8	6.6	6.1	6.5	6.5
66	6.6	7.4	6.9	7.6	6.9	6.8	7.0	7.5	6.8	6.9	6.7	7.2	6.5	6.6	6.2	6.2
67	7.0	6.7	7.6	6.8	7.6	7.1	6.8	7.4	7.0	6.8	7.0	7.1	7.1	6.3	6.3	6.3
68	6.3	6.8	6.6	7.3	6.8	7.5	7.1	7.3	7.0	6.9	6.8	7.2	7.2	7.1	6.3	6.3
69	6.8	6.7	7.2	6.7	7.1	7.0	7.8	7.7	6.8	6.7	7.0	6.8	7.1	6.8	7.0	7.0
70	6.9	7.0	6.7	7.2	6.8	7.8	7.1	7.8	7.5	6.8	7.0	7.4	7.2	6.9	6.8	6.8
71	6.8	7.2	7.0	7.0	7.4	6.7	7.5	7.2	7.6	7.6	6.8	7.4	7.3	6.7	6.6	6.6
72	7.3	7.1	7.0	7.1	6.5	7.1	7.3	8.0	7.0	7.3	7.3	6.7	7.0	7.1	6.9	6.9
73	6.9	6.9	7.5	7.0	7.3	6.4	7.5	7.0	7.3	6.7	7.2	7.5	6.9	6.8	7.2	7.2
74	7.2	7.3	6.9	7.3	6.8	7.5	6.7	7.6	6.5	7.5	7.0	7.7	7.7	7.0	7.0	7.0
75	6.7	7.2	6.8	7.6	7.1	6.9	7.1	7.2	7.1	6.7	7.1	7.3	7.5	7.1	7.3	7.3
76	6.6	6.9	7.2	7.7	7.3	7.0	6.9	7.2	6.7	6.6	6.5	7.8	7.2	7.3	7.7	7.7
77	7.1	7.1	6.5	7.3	7.4	7.2	7.0	7.3	6.9	6.5	7.7	6.4	7.3	6.2	7.3	7.3
78	6.2	6.9	6.9	6.6	7.2	7.2	7.7	7.2	6.9	7.3	6.6	8.1	6.5	6.7	7.3	7.3
79	7.2	6.5	7.3	7.5	6.8	7.2	7.2	7.5	7.1	6.9	7.4	7.6	7.2	5.5	7.4	7.4
80	6.9	6.8	6.9	7.1	6.4	7.2	7.5	7.1	6.8	6.8	6.5	7.6	7.0	7.1	6.4	6.4
81	6.5	7.0	6.4	6.5	7.0	7.0	7.1	7.8	6.9	7.0	7.1	7.2	7.4	6.6	7.1	7.1
82	6.3	5.8	7.2	7.4	6.4	7.6	7.2	7.2	7.0	6.1	6.2	7.2	6.5	6.7	6.0	6.0
83	6.3	6.4	7.0	7.3	6.8	6.3	7.2	7.3	6.9	6.4	6.2	6.3	7.4	6.2	7.2	7.2
84	7.2	6.8	6.7	6.6	7.7	6.6	6.4	7.4	6.8	6.2	7.3	6.5	6.1	6.9	6.6	6.6
85	6.4	6.1	7.3	6.2	6.8	7.4	6.4	6.7	7.1	6.9	6.7	7.3	6.8	6.6	7.1	7.1
86	-	7.3	7.9	7.7	6.7	7.5	8.1	6.2	6.9	7.1	6.5	7.4	6.4	5.8	5.9	5.9
87	4.8	-	7.6	7.1	7.5	6.9	-	8.4	6.0	6.0	7.1	5.6	7.0	7.0	6.4	6.4
88	6.3	-	-	6.9	-	-	8.5	-	7.7	-	-	7.0	-	7.4	6.7	6.7
89	-	-	-	-	-	-	-	7.7	-	7.1	-	-	-	-	-	-
90	-	-	-	-	-	-	-	-	-	-	7.8	-	-	-	-	-

TABELLE B.5:
Ergänzung Kapitel 3.2.1: GEZ – Mittelwerte der Ein-Jahres-Altersgruppen

Alter	SOEP-Befragungswelle															
	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99
45	6.9	6.9	7.1	6.6	6.8	6.4	7.1	7.1	6.7	7.1	6.4	6.6	6.3	6.3	6.2	6.2
46	6.6	6.6	6.4	6.7	6.5	6.7	6.4	6.7	7.3	6.3	7.1	6.5	6.6	6.6	6.2	6.2
47	6.8	6.6	6.6	6.5	6.5	6.3	6.6	6.8	7.3	6.7	6.3	6.4	6.3	6.5	6.2	6.2
48	7.0	6.8	6.3	6.6	6.3	6.6	6.5	6.7	6.8	7.0	6.8	6.1	6.7	6.4	6.8	6.8
49	6.4	6.7	6.4	6.6	6.4	6.2	6.4	6.4	6.9	6.3	6.8	6.6	6.1	6.7	6.5	6.5
50	6.4	6.6	6.7	6.6	6.2	6.1	6.5	6.3	6.5	6.7	6.3	6.7	6.6	6.2	6.5	6.5
51	6.2	6.3	6.3	6.8	6.5	6.2	6.3	6.4	6.6	6.3	6.6	6.1	6.6	6.5	6.2	6.2
52	6.1	6.2	6.0	6.0	6.5	6.4	6.2	6.4	6.3	6.3	6.3	6.2	6.0	6.3	6.4	6.4
53	6.3	6.1	6.2	5.9	6.0	6.2	6.2	6.4	6.3	6.1	6.3	6.1	6.1	5.9	6.1	6.1
54	6.0	6.0	6.2	6.4	5.6	6.0	6.4	6.2	6.4	6.0	6.1	6.2	6.6	5.9	6.0	6.0
55	6.0	5.9	6.1	5.8	6.1	5.5	5.9	6.1	6.1	5.9	6.2	5.9	6.3	5.9	6.2	6.2
56	6.1	6.0	5.8	6.0	6.0	5.6	5.5	6.0	6.0	5.9	5.9	5.9	5.9	6.3	6.0	6.0
57	5.7	5.8	6.2	5.8	5.8	5.8	5.9	5.7	6.1	6.0	5.6	5.9	6.0	5.7	6.1	6.1
58	6.2	5.9	5.4	6.0	5.7	5.7	5.8	5.6	5.5	5.9	6.4	5.7	5.9	6.0	6.1	6.1
59	5.6	6.0	5.5	5.8	6.3	5.5	5.9	5.7	5.8	5.6	5.5	6.3	5.9	5.6	6.0	6.0
60	5.8	6.0	6.2	6.1	5.8	6.1	5.5	5.9	5.7	5.9	5.1	5.7	6.3	5.8	5.9	5.9
61	5.9	5.5	6.1	6.3	5.7	5.9	6.4	5.8	5.7	6.1	5.8	6.0	5.9	6.3	6.0	6.0
62	6.1	6.0	6.0	6.1	6.3	5.7	6.1	6.3	5.6	6.1	6.0	5.9	5.8	5.6	5.8	5.8
63	5.6	6.4	6.1	6.2	6.2	6.0	5.5	6.0	6.4	5.5	6.0	5.7	5.7	5.0	6.0	6.0
64	5.2	5.6	6.0	6.2	5.9	5.9	6.4	5.9	5.9	6.2	5.5	5.9	5.8	5.6	5.4	5.4
65	5.8	5.7	5.9	6.0	5.9	6.1	5.9	6.6	6.2	6.2	6.0	5.7	6.1	5.8	5.9	5.9
66	5.0	5.8	5.7	5.9	5.8	5.7	5.7	6.0	6.4	5.8	6.0	6.3	5.4	5.6	5.7	5.7
67	5.8	5.6	5.8	5.5	6.3	5.9	5.7	5.6	6.0	5.9	5.5	5.9	6.0	5.8	5.2	5.2
68	5.3	5.7	5.7	5.8	5.7	5.6	5.7	5.6	6.1	5.7	5.7	5.7	5.6	6.0	5.7	5.7
69	5.6	5.9	5.9	5.5	5.9	4.8	6.2	5.4	5.7	5.8	5.6	5.8	5.8	6.0	6.2	6.2
70	5.7	5.2	5.6	6.1	5.8	5.3	4.8	5.9	5.7	5.4	5.7	5.7	5.9	5.3	5.7	5.7
71	5.7	6.4	5.3	5.7	5.5	5.1	6.0	5.1	5.8	5.7	5.2	5.6	5.5	6.0	5.7	5.7
72	5.8	5.9	6.1	5.5	5.6	5.9	5.2	6.2	5.1	6.1	5.2	5.3	5.6	5.8	6.0	6.0
73	4.7	5.9	5.9	5.8	5.3	5.2	5.6	5.1	5.8	4.8	5.3	5.2	5.4	5.6	5.7	5.7
74	5.7	5.1	5.9	5.5	5.3	4.9	5.4	5.8	4.9	5.4	5.4	5.7	5.4	5.3	5.6	5.6
75	5.2	6.0	5.3	6.0	4.9	5.0	5.1	5.9	5.2	5.1	5.6	4.6	5.3	5.4	5.5	5.5
76	4.9	5.2	5.7	5.3	5.4	5.8	5.6	4.6	5.4	5.3	5.3	6.1	4.7	5.5	5.6	5.6
77	6.3	5.1	5.2	5.7	4.8	5.7	5.2	5.1	4.8	5.1	5.0	4.4	5.8	5.0	5.2	5.2
78	5.0	5.6	5.1	4.9	5.3	4.9	5.7	5.4	4.9	4.7	4.7	5.0	4.7	5.5	4.9	4.9
79	5.0	4.9	6.0	6.1	4.7	5.5	4.6	5.1	5.4	4.9	4.7	5.7	5.0	4.5	5.0	5.0
80	5.6	5.8	4.0	6.1	4.8	4.3	5.6	4.8	4.9	5.8	4.7	4.3	5.1	5.1	4.0	4.0
81	4.9	5.1	5.6	4.7	5.5	5.4	4.4	6.2	4.1	5.2	5.2	4.9	4.5	4.4	4.9	4.9
82	4.0	4.9	5.1	5.8	4.6	5.6	5.3	4.7	5.8	4.3	5.4	4.8	4.3	4.5	3.7	3.7
83	3.8	4.9	4.8	5.2	4.6	5.6	5.2	4.8	4.0	5.5	4.1	5.4	4.9	4.7	5.2	5.2
84	4.3	3.5	5.0	4.3	4.1	5.2	4.3	5.6	4.9	3.9	6.2	4.6	5.9	5.6	4.8	4.8
85	4.1	3.7	4.0	4.0	4.5	5.9	5.6	4.9	4.2	4.5	3.8	6.0	4.6	5.7	4.5	4.5
86	-	4.7	4.3	4.0	4.9	3.7	5.1	5.2	5.1	5.2	4.8	4.0	6.0	4.7	5.9	5.9
87	4.1	-	4.8	2.8	4.5	5.0	-	4.8	5.3	5.4	5.0	4.5	4.8	5.7	4.6	4.6
88	4.5	-	-	4.3	-	-	4.9	-	4.2	-	-	4.8	-	4.6	5.7	5.7
89	-	-	-	-	-	-	-	4.7	-	4.5	-	-	-	-	-	-
90	-	-	-	-	-	-	-	-	-	-	4.2	-	-	-	-	-

TABELLE B.6:
Ergänzung Kapitel 3.2.1: WOZ – Mittelwerte der Ein-Jahres-Altersgruppen

Alter	SOEP-Befragungswelle															
	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99
45	8.2	8.0	8.2	8.1	8.2	7.4	8.2	7.9	7.6	7.8	7.7	7.9	8.0	7.7	7.4	7.4
46	8.1	8.0	7.8	8.2	7.8	7.9	7.5	8.1	8.0	7.9	7.7	8.1	8.0	7.7	7.8	7.8
47	8.1	8.0	8.0	7.9	8.1	7.8	7.8	7.8	8.0	7.9	8.0	7.7	7.9	8.0	7.7	7.7
48	8.0	7.9	8.0	7.8	7.9	8.3	7.8	8.0	7.8	8.1	7.9	7.9	7.8	7.7	7.9	7.9
49	8.1	8.1	7.8	8.0	8.1	7.8	7.9	8.1	8.1	7.7	7.9	8.0	7.9	7.9	7.8	7.8
50	8.2	8.0	8.2	7.9	7.9	7.9	7.8	8.0	8.1	7.9	7.9	8.2	7.8	7.9	7.6	7.6
51	7.8	8.2	8.0	8.1	8.0	7.8	8.2	8.1	7.9	8.2	7.8	7.8	8.0	7.8	8.1	8.1
52	8.0	7.7	8.3	8.4	8.1	7.8	7.9	8.2	8.1	8.0	7.9	7.8	8.0	8.2	7.7	7.7
53	8.4	7.9	7.5	8.3	8.0	7.9	7.8	8.1	8.1	7.9	7.8	7.8	8.2	7.6	7.8	7.8
54	8.1	8.2	8.0	8.0	8.2	8.3	7.8	8.0	8.0	8.0	8.1	8.0	8.3	8.1	7.6	7.6
55	8.3	8.2	8.4	8.3	8.1	8.1	8.0	7.9	8.0	8.3	8.2	7.8	8.2	8.0	8.3	8.3
56	8.4	8.5	8.1	8.5	8.2	7.8	8.1	8.0	7.8	8.0	7.9	7.9	8.0	7.9	7.9	7.9
57	8.3	8.4	8.6	8.4	8.3	8.1	8.1	8.3	8.1	8.0	7.9	8.0	7.9	8.0	8.1	8.1
58	8.5	7.8	8.1	8.5	8.0	8.2	8.3	8.1	8.1	8.2	8.2	8.0	8.2	7.9	8.0	8.0
59	8.3	7.9	8.1	8.2	8.2	8.0	8.1	8.1	8.1	8.2	8.1	7.9	8.1	8.0	7.7	7.7
60	8.1	8.1	8.3	8.3	8.4	8.4	8.1	8.2	8.4	8.0	8.0	7.6	7.9	7.8	8.0	8.0
61	8.3	7.9	8.1	8.3	8.0	8.3	8.4	8.2	8.2	8.1	8.2	8.2	8.1	8.1	7.8	7.8
62	8.5	8.2	8.2	8.6	8.3	8.0	8.3	8.5	8.1	8.2	8.3	8.2	8.1	8.1	7.6	7.6
63	8.4	8.3	8.2	8.3	8.1	8.4	8.2	8.6	8.7	8.1	7.9	8.0	8.3	8.0	8.1	8.1
64	8.4	8.5	8.1	8.3	8.1	8.2	8.2	8.4	8.5	8.5	8.0	8.2	8.2	8.1	8.0	8.0
65	8.2	8.2	8.3	8.6	8.4	8.1	8.3	8.7	8.3	8.3	8.6	7.7	7.8	8.0	8.2	8.2
66	8.3	8.4	8.1	8.6	8.5	8.1	7.7	8.5	8.7	8.2	8.1	8.8	8.0	8.0	7.9	7.9
67	8.1	8.2	8.8	8.4	8.5	8.6	7.9	8.2	8.4	8.7	8.3	8.3	8.5	7.8	7.9	7.9
68	8.3	8.3	8.3	8.4	8.2	8.5	8.3	8.4	8.2	8.4	8.7	8.1	8.2	8.4	7.9	7.9
69	8.3	7.9	8.5	8.2	8.0	8.1	8.6	8.7	8.1	8.1	8.3	8.4	8.1	8.2	8.4	8.4
70	8.3	8.5	7.9	8.7	8.0	8.2	8.2	8.5	8.7	8.2	8.5	8.3	8.4	8.1	7.9	7.9
71	8.3	8.2	8.5	7.9	8.4	8.2	8.1	8.6	8.2	8.8	8.3	8.4	8.5	8.2	7.7	7.7
72	8.6	8.4	8.1	8.6	8.0	8.6	8.4	8.9	8.3	8.3	8.5	8.2	8.3	8.5	8.4	8.4
73	8.4	8.0	8.5	8.5	8.5	8.7	8.4	8.3	8.8	8.4	8.1	8.5	7.9	8.2	8.4	8.4
74	8.5	8.2	8.0	8.3	8.2	8.4	8.7	8.7	7.9	8.7	8.6	8.6	8.7	8.3	7.8	7.8
75	8.5	8.3	7.9	8.1	8.1	8.5	8.3	9.0	8.6	8.3	8.8	8.3	8.3	8.3	8.3	8.3
76	8.2	8.5	8.3	8.6	8.3	8.5	8.0	8.4	8.4	8.7	8.7	8.3	8.2	8.5	8.4	8.4
77	8.6	8.4	8.3	7.9	8.3	8.3	8.0	8.4	8.3	8.1	8.6	8.0	8.0	8.4	8.5	8.5
78	8.7	8.3	8.2	8.1	7.9	8.2	8.4	8.3	8.2	8.4	8.2	8.5	8.0	8.0	8.0	8.0
79	8.7	8.4	8.5	8.0	8.1	8.1	8.2	8.5	8.7	8.2	8.7	8.7	8.7	7.1	8.1	8.1
80	8.1	8.3	8.2	8.3	7.7	7.8	8.4	8.1	8.0	8.4	8.8	8.4	8.6	8.4	6.9	6.9
81	8.7	8.5	8.4	8.3	8.2	7.9	8.3	8.4	8.3	8.3	8.8	8.0	8.3	8.3	8.0	8.0
82	7.8	8.2	8.5	8.5	7.9	8.6	8.3	8.4	8.2	8.1	8.2	8.1	8.1	8.1	7.3	7.3
83	8.1	7.7	8.1	8.3	8.6	8.3	8.0	8.2	8.3	7.8	8.8	7.8	7.9	7.5	8.2	8.2
84	8.9	8.1	8.1	8.0	8.1	8.6	8.4	8.8	8.3	8.4	8.6	8.3	8.1	8.3	8.1	8.1
85	7.4	7.6	8.6	7.7	7.7	8.8	8.9	8.5	8.6	8.2	8.9	8.3	8.4	7.6	8.3	8.3
86	-	7.7	8.7	8.6	7.7	8.5	8.4	8.4	8.4	8.5	7.7	8.3	8.1	7.8	7.9	7.9
87	7.8	-	8.1	8.1	8.6	7.5	-	8.9	9.3	9.0	8.8	6.9	8.3	8.3	8.1	8.1
88	8.0	-	-	7.9	-	-	8.3	-	8.7	-	-	8.0	-	8.5	8.1	8.1
89	-	-	-	-	-	-	-	7.4	-	8.4	-	-	-	-	-	-
90	-	-	-	-	-	-	-	-	-	-	9.5	-	-	-	-	-

TABELLE B.7:
Ergänzung Kapitel 3.2.1: FRZ – Mittelwerte der Ein-Jahres-Altersgruppen

Alter	SOEP-Befragungswelle															
	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99
45	6.9	7.0	7.0	7.0	6.5	6.5	6.3	6.2	6.7	6.8	6.2	6.4	6.7	6.5	6.2	6.2
46	7.1	6.7	6.8	7.2	7.0	6.1	6.6	6.5	6.6	6.5	6.8	6.6	6.5	6.8	6.5	6.5
47	7.0	7.0	6.8	6.8	7.2	6.4	6.9	6.3	6.8	6.5	6.5	6.8	6.6	6.8	6.8	6.8
48	7.0	7.0	6.8	6.8	7.0	6.9	6.9	6.7	6.5	6.5	6.2	6.8	7.0	6.6	6.9	6.9
49	7.0	6.9	6.9	7.3	6.7	6.6	6.9	6.6	6.5	6.1	6.7	6.8	6.7	7.0	6.6	6.6
50	7.1	6.9	6.8	7.0	6.9	6.7	6.8	6.7	6.7	6.3	6.2	6.5	6.9	7.0	6.9	6.9
51	6.5	6.7	6.7	7.1	6.7	7.0	6.9	6.4	7.0	6.6	6.5	6.6	6.8	6.7	7.0	7.0
52	6.8	6.7	6.6	6.7	6.8	6.4	7.0	6.5	6.5	6.7	6.8	6.8	6.5	6.6	6.5	6.5
53	7.6	6.8	6.5	7.2	6.5	6.7	6.7	7.0	6.8	6.3	6.7	7.1	7.1	6.3	6.2	6.2
54	7.4	7.1	6.8	6.5	7.1	6.6	6.9	6.6	6.9	6.6	6.7	7.3	7.4	6.7	6.5	6.5
55	7.3	7.4	7.6	6.9	6.4	6.9	6.6	6.5	6.8	6.9	6.8	6.4	7.6	7.0	6.7	6.7
56	7.3	7.3	7.1	7.6	7.0	6.2	6.8	6.4	6.8	6.4	6.9	6.8	7.1	7.2	7.3	7.3
57	7.2	7.0	7.5	7.3	7.0	7.0	6.4	6.9	6.9	6.8	6.6	7.1	7.1	6.7	7.5	7.5
58	7.3	7.3	6.9	7.4	6.9	7.3	7.2	6.2	6.8	6.7	6.9	7.0	7.6	7.1	7.3	7.3
59	7.9	7.5	7.3	7.4	7.5	7.2	7.3	6.9	6.6	6.9	6.5	7.3	7.5	7.3	7.0	7.0
60	7.9	7.7	7.4	7.9	7.1	7.5	7.4	7.3	7.5	6.8	6.8	7.0	7.5	7.3	7.5	7.5
61	7.7	7.9	7.7	7.8	7.7	7.5	7.6	7.6	7.6	6.9	7.2	7.5	7.5	7.7	7.3	7.3
62	8.3	7.6	7.6	8.3	8.0	7.7	7.5	7.5	7.4	7.7	7.6	7.3	7.9	7.5	7.6	7.6
63	8.3	8.2	7.6	8.2	7.9	8.0	7.8	7.8	7.6	7.3	7.4	7.7	7.5	8.0	7.6	7.6
64	7.9	8.3	8.2	8.0	7.9	7.6	8.0	8.3	7.9	7.8	7.6	7.4	8.1	7.6	7.6	7.6
65	8.2	8.0	8.1	8.0	8.0	8.0	7.9	8.4	8.3	7.9	8.0	7.3	7.8	7.6	7.7	7.7
66	8.0	8.2	8.0	8.2	8.1	8.1	7.8	8.2	8.4	8.1	7.9	8.2	7.8	8.2	8.0	8.0
67	8.7	7.5	8.3	8.2	8.3	7.8	7.8	8.2	8.3	8.3	7.9	7.6	8.1	7.5	7.6	7.6
68	8.1	8.5	7.9	8.4	7.8	7.9	7.8	8.1	8.4	7.7	8.4	7.8	7.7	8.1	7.8	7.8
69	8.2	7.8	8.3	7.5	8.1	7.9	8.4	8.3	8.1	8.0	7.9	8.1	8.0	8.1	8.0	8.0
70	8.2	7.6	8.1	8.3	7.8	8.0	7.8	8.6	8.3	7.7	8.2	8.2	8.5	8.0	7.7	7.7
71	8.5	8.4	7.8	8.4	8.1	8.1	8.3	8.1	8.3	8.4	7.6	8.2	8.5	8.1	7.5	7.5
72	8.1	8.3	8.2	8.2	7.6	8.4	7.5	8.8	8.4	8.3	7.7	7.7	8.4	8.2	8.4	8.4
73	8.3	7.9	8.4	8.2	8.0	8.4	8.3	7.8	8.8	8.2	7.8	8.0	7.8	8.4	8.0	8.0
74	8.4	8.3	8.0	8.4	8.0	7.5	8.3	8.7	7.8	8.3	8.6	8.2	8.0	8.0	7.7	7.7
75	8.0	8.3	8.0	8.0	7.9	8.3	7.6	8.0	8.7	7.4	8.5	7.8	8.5	7.9	7.7	7.7
76	8.2	7.8	8.2	8.4	8.3	8.2	7.8	8.0	8.0	8.5	7.4	8.2	8.6	8.4	8.1	8.1
77	8.3	8.2	7.8	8.1	7.9	8.4	7.7	8.3	8.1	7.3	8.3	7.8	7.4	8.5	8.2	8.2
78	8.4	8.1	8.2	7.4	8.0	8.6	7.8	8.2	8.0	7.7	7.8	8.7	7.3	7.5	8.7	8.7
79	8.3	8.1	8.0	8.2	7.5	8.2	7.6	8.1	9.0	8.0	7.9	8.1	8.9	7.2	7.8	7.8
80	8.1	8.4	7.9	8.4	8.1	8.0	8.4	8.3	8.1	8.4	8.0	7.8	8.4	8.7	7.5	7.5
81	8.3	8.0	8.3	7.8	8.4	8.4	7.4	8.1	8.4	7.9	8.2	7.1	7.9	7.7	8.3	8.3
82	7.5	8.2	8.1	8.0	7.3	8.1	8.2	8.4	7.8	6.8	6.9	8.0	7.7	7.6	6.7	6.7
83	7.1	7.2	7.8	7.7	8.4	8.3	7.4	8.4	7.8	7.2	8.6	6.9	8.5	7.2	8.8	8.8
84	7.5	7.6	8.1	7.4	8.0	8.7	7.5	8.5	8.5	8.0	7.0	8.0	7.5	8.1	8.5	8.5
85	6.8	7.1	8.3	6.8	7.9	7.9	8.3	8.4	7.7	8.5	8.4	7.8	7.5	7.8	8.1	8.1
86	-	8.1	8.9	6.8	8.1	7.9	7.2	8.3	7.0	7.0	8.1	7.6	8.8	7.9	6.9	6.9
87	6.4	-	8.3	7.6	8.3	8.3	-	8.7	8.0	6.4	7.4	7.2	8.5	8.7	7.6	7.6
88	7.3	-	-	7.5	-	-	8.2	-	8.6	-	-	7.0	-	8.8	7.9	7.9
89	-	-	-	-	-	-	-	7.8	-	7.3	-	-	-	-	-	-
90	-	-	-	-	-	-	-	-	-	-	9.1	-	-	-	-	-

TABELLE B.8:
Ergänzung Kapitel 3.2.1: ALZ – Anzahl der Werte in den Ein-Jahres-Altersgruppen

Alter	SOEP-Befragungswelle															
	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99
45	199	194	152	142	112	133	62	91	111	113	120	126	104	100	89	89
46	203	174	185	148	131	106	126	63	90	109	111	119	124	102	93	93
47	185	182	165	175	144	122	101	123	62	87	103	110	115	120	97	97
48	152	172	171	163	163	135	119	96	117	60	85	97	109	115	121	121
49	171	139	165	166	159	150	124	116	96	111	60	81	96	101	112	112
50	127	147	132	158	160	148	138	122	108	92	107	58	76	95	94	94
51	116	121	131	125	147	151	141	137	118	109	92	101	59	76	87	87
52	119	111	115	128	119	132	139	136	133	112	103	86	96	59	72	72
53	127	110	102	109	118	113	127	138	132	131	108	97	84	92	54	54
54	155	113	103	101	104	109	106	124	132	129	123	100	98	82	90	90
55	145	141	106	97	95	97	108	104	120	133	127	124	96	96	75	75
56	152	132	132	102	91	87	89	103	98	117	127	127	117	95	92	92
57	143	138	121	129	98	86	81	83	101	99	113	126	122	113	95	95
58	137	124	130	119	123	92	82	79	82	95	95	113	119	117	106	106
59	134	115	122	125	108	117	89	81	76	77	92	98	104	116	112	112
60	125	117	102	112	118	105	110	89	79	75	70	88	98	103	110	110
61	123	113	99	102	110	110	104	108	86	75	76	67	90	94	99	99
62	132	111	103	97	97	100	107	103	103	82	72	71	61	87	89	89
63	139	110	100	98	95	90	94	97	100	98	80	74	70	60	77	77
64	107	123	99	94	90	92	88	92	93	95	96	75	71	64	57	57
65	73	98	118	96	88	83	88	88	90	87	91	92	75	72	58	58
66	65	61	88	110	92	83	83	86	82	88	83	81	87	71	70	70
67	52	54	55	79	100	88	80	81	82	79	81	80	82	82	71	71
68	65	48	46	57	76	92	88	78	80	78	75	77	79	82	77	77
69	95	55	48	46	51	68	90	83	75	76	77	71	77	72	78	78
70	89	84	52	45	44	49	68	85	80	73	75	75	67	73	67	67
71	89	75	73	49	42	41	47	67	83	77	75	71	72	63	69	69
72	86	74	71	70	47	40	41	45	60	79	69	69	71	66	55	55
73	82	79	67	70	60	44	37	35	42	59	77	68	61	67	65	65
74	88	72	64	61	63	52	40	34	33	40	54	71	58	57	64	64
75	82	68	59	56	50	62	51	38	31	31	34	50	70	54	52	52
76	75	69	58	53	52	45	57	49	32	32	29	29	47	67	51	51
77	61	62	61	56	44	46	44	50	45	30	27	26	27	41	61	61
78	52	50	47	56	53	39	45	42	44	40	27	23	23	27	39	39
79	45	42	45	42	47	44	38	39	34	42	35	23	23	22	25	25
80	41	31	36	44	35	40	39	34	36	31	38	35	21	22	22	22
81	33	35	26	33	41	32	38	38	32	28	29	30	29	23	22	22
82	42	21	28	27	26	36	30	34	35	30	27	26	27	27	20	20
83	26	33	18	24	23	23	33	27	32	30	25	24	23	26	23	23
84	33	19	25	13	21	19	21	29	22	30	27	22	21	22	22	22
85	27	22	15	24	11	16	14	17	25	21	25	23	19	19	19	19
86	9	20	15	15	17	11	16	13	13	20	18	19	21	15	14	14
87	10	6	16	13	13	13	5	14	11	10	19	11	12	19	12	12
88	12	8	4	15	9	8	11	3	13	9	9	13	7	11	15	15
89	5	7	3	2	9	6	6	10	2	12	6	7	8	7	8	8
90	3	4	6	4	2	9	5	6	6	1	10	4	5	8	6	6

TABELLE B.9:
Ergänzung Kapitel 3.3.1: ALZ, univariate Normalverteilungseigenschaften

ALZ	Alle				Alt84			
	Schiefe	Kurtosis	D^1	α^2	Schiefe	Kurtosis	W^3	α^2
1984	-0.972	1.078	0.175	0.010	-0.938	0.752	0.895	0.000
1985	-0.900	0.998	0.176	0.010	-0.863	0.577	0.910	0.000
1986	-0.904	1.063	0.187	0.010	-1.020	1.121	0.900	0.000
1987	-0.858	0.878	0.174	0.010	-0.968	1.008	0.910	0.000
1988	-0.839	0.709	0.179	0.010	-0.827	0.526	0.923	0.000
1989	-0.851	0.776	0.188	0.010	-0.723	0.444	0.929	0.000
1990	-0.938	1.367	0.196	0.010	-0.766	0.448	0.926	0.000
1991	-1.014	1.706	0.206	0.010	-0.937	0.951	0.912	0.000
1992	-0.926	1.425	0.199	0.010	-0.796	0.617	0.925	0.000
1993	-0.994	1.507	0.194	0.010	-0.857	0.569	0.920	0.000
1994	-0.961	1.428	0.185	0.010	-0.816	0.714	0.927	0.000
1995	-0.921	1.225	0.185	0.010	-0.770	0.433	0.931	0.000
1996	-0.974	1.453	0.188	0.010	-0.770	0.619	0.926	0.000
1997	-0.869	0.940	0.191	0.010	-0.710	0.402	0.937	0.000
1998	-1.004	1.389	0.195	0.010	-0.768	0.533	0.928	0.000
1999	-1.030	1.360	0.198	0.010	-0.718	0.225	0.929	0.000

¹ Kolmogorov-Teststatistik D .

² Rechtsseitige Irrtumswahrscheinlichkeit α bei Annahme H_1 .

³ Shapiro-Wilk-Teststatistik W .

TABELLE B.10:
Ergänzung Kapitel 3.3.1: EIZ, univariate Normalverteilungseigenschaften

EIZ	Alle				Alt84			
	Schiefe	Kurtosis	D^1	α^2	Schiefe	Kurtosis	W^3	α^2
1984	-0.972	1.078	0.175	0.010	-0.938	0.752	0.895	0.000
1985	-0.900	0.998	0.176	0.010	-0.863	0.577	0.910	0.000
1986	-0.904	1.063	0.187	0.010	-1.020	1.121	0.900	0.000
1987	-0.858	0.878	0.174	0.010	-0.968	1.008	0.910	0.000
1988	-0.839	0.709	0.179	0.010	-0.827	0.526	0.923	0.000
1989	-0.851	0.776	0.188	0.010	-0.723	0.444	0.929	0.000
1990	-0.938	1.367	0.196	0.010	-0.766	0.448	0.926	0.000
1991	-1.014	1.706	0.206	0.010	-0.937	0.951	0.912	0.000
1992	-0.926	1.425	0.199	0.010	-0.796	0.617	0.925	0.000
1993	-0.994	1.507	0.194	0.010	-0.857	0.569	0.920	0.000
1994	-0.961	1.428	0.185	0.010	-0.816	0.714	0.927	0.000
1995	-0.921	1.225	0.185	0.010	-0.770	0.433	0.931	0.000
1996	-0.974	1.453	0.188	0.010	-0.770	0.619	0.926	0.000
1997	-0.869	0.940	0.191	0.010	-0.710	0.402	0.937	0.000
1998	-1.004	1.389	0.195	0.010	-0.768	0.533	0.928	0.000
1999	-1.030	1.360	0.198	0.010	-0.718	0.225	0.929	0.000

^{1 2 3} Siehe Anmerkungen Tabelle B.9.

TABELLE B.11:
Ergänzung Kapitel 3.3.1: GEZ, univariate Normalverteilungseigenschaften

GEZ	Alle				Alt84			
	Schiefe	Kurtosis	D^1	α^2	Schiefe	Kurtosis	W^3	α^2
1984	-0.972	1.078	0.175	0.010	-0.938	0.752	0.895	0.000
1985	-0.900	0.998	0.176	0.010	-0.863	0.577	0.910	0.000
1986	-0.904	1.063	0.187	0.010	-1.020	1.121	0.900	0.000
1987	-0.858	0.878	0.174	0.010	-0.968	1.008	0.910	0.000
1988	-0.839	0.709	0.179	0.010	-0.827	0.526	0.923	0.000
1989	-0.851	0.776	0.188	0.010	-0.723	0.444	0.929	0.000
1990	-0.938	1.367	0.196	0.010	-0.766	0.448	0.926	0.000
1991	-1.014	1.706	0.206	0.010	-0.937	0.951	0.912	0.000
1992	-0.926	1.425	0.199	0.010	-0.796	0.617	0.925	0.000
1993	-0.994	1.507	0.194	0.010	-0.857	0.569	0.920	0.000
1994	-0.961	1.428	0.185	0.010	-0.816	0.714	0.927	0.000
1995	-0.921	1.225	0.185	0.010	-0.770	0.433	0.931	0.000
1996	-0.974	1.453	0.188	0.010	-0.770	0.619	0.926	0.000
1997	-0.869	0.940	0.191	0.010	-0.710	0.402	0.937	0.000
1998	-1.004	1.389	0.195	0.010	-0.768	0.533	0.928	0.000
1999	-1.030	1.360	0.198	0.010	-0.718	0.225	0.929	0.000

^{1 2 3} Siehe Anmerkungen Tabelle B.9.

TABELLE B.12:
Ergänzung Kapitel 3.3.1: WOZ, univariate Normalverteilungseigenschaften

WOZ	Alle				Alt84			
	Schiefe	Kurtosis	D^1	α^2	Schiefe	Kurtosis	W^3	α^2
1984	-0.972	1.078	0.175	0.010	-0.938	0.752	0.895	0.000
1985	-0.900	0.998	0.176	0.010	-0.863	0.577	0.910	0.000
1986	-0.904	1.063	0.187	0.010	-1.020	1.121	0.900	0.000
1987	-0.858	0.878	0.174	0.010	-0.968	1.008	0.910	0.000
1988	-0.839	0.709	0.179	0.010	-0.827	0.526	0.923	0.000
1989	-0.851	0.776	0.188	0.010	-0.723	0.444	0.929	0.000
1990	-0.938	1.367	0.196	0.010	-0.766	0.448	0.926	0.000
1991	-1.014	1.706	0.206	0.010	-0.937	0.951	0.912	0.000
1992	-0.926	1.425	0.199	0.010	-0.796	0.617	0.925	0.000
1993	-0.994	1.507	0.194	0.010	-0.857	0.569	0.920	0.000
1994	-0.961	1.428	0.185	0.010	-0.816	0.714	0.927	0.000
1995	-0.921	1.225	0.185	0.010	-0.770	0.433	0.931	0.000
1996	-0.974	1.453	0.188	0.010	-0.770	0.619	0.926	0.000
1997	-0.869	0.940	0.191	0.010	-0.710	0.402	0.937	0.000
1998	-1.004	1.389	0.195	0.010	-0.768	0.533	0.928	0.000
1999	-1.030	1.360	0.198	0.010	-0.718	0.225	0.929	0.000

^{1 2 3} Siehe Anmerkungen Tabelle B.9.

TABELLE B.13:
Ergänzung Kapitel 3.3.1: FRZ, univariate Normalverteilungseigenschaften

FRZ	<i>Alle</i>				<i>Alt84</i>			
	Schiefe	Kurtosis	D^1	α^2	Schiefe	Kurtosis	W^3	α^2
1984	-0.972	1.078	0.175	0.010	-0.938	0.752	0.895	0.000
1985	-0.900	0.998	0.176	0.010	-0.863	0.577	0.910	0.000
1986	-0.904	1.063	0.187	0.010	-1.020	1.121	0.900	0.000
1987	-0.858	0.878	0.174	0.010	-0.968	1.008	0.910	0.000
1988	-0.839	0.709	0.179	0.010	-0.827	0.526	0.923	0.000
1989	-0.851	0.776	0.188	0.010	-0.723	0.444	0.929	0.000
1990	-0.938	1.367	0.196	0.010	-0.766	0.448	0.926	0.000
1991	-1.014	1.706	0.206	0.010	-0.937	0.951	0.912	0.000
1992	-0.926	1.425	0.199	0.010	-0.796	0.617	0.925	0.000
1993	-0.994	1.507	0.194	0.010	-0.857	0.569	0.920	0.000
1994	-0.961	1.428	0.185	0.010	-0.816	0.714	0.927	0.000
1995	-0.921	1.225	0.185	0.010	-0.770	0.433	0.931	0.000
1996	-0.974	1.453	0.188	0.010	-0.770	0.619	0.926	0.000
1997	-0.869	0.940	0.191	0.010	-0.710	0.402	0.937	0.000
1998	-1.004	1.389	0.195	0.010	-0.768	0.533	0.928	0.000
1999	-1.030	1.360	0.198	0.010	-0.718	0.225	0.929	0.000

^{1 2 3} Siehe Anmerkungen Tabelle B.9.

TABELLE B.14:
Ergänzung Kapitel 3.3.2: Modell 1, Vier-Jahres-Stabilitätskoeffizienten und True-Score-Varianzen

Gruppe	Zeit	ALZ		EIZ		GEZ		WOZ		FRZ	
		$\hat{\rho}_{t'j}^*$	$\hat{\sigma}_{t'j}^{*2}$	$\hat{\rho}_{t'j}^*$	$\hat{\sigma}_{t'j}^{*2}$	$\hat{\rho}_{t'j}^*$	$\hat{\sigma}_{t'j}^{*2}$	$\hat{\rho}_{t'j}^*$	$\hat{\sigma}_{t'j}^{*2}$	$\hat{\rho}_{t'j}^*$	$\hat{\sigma}_{t'j}^{*2}$
1	1	0.93	1.63	0.93	1.97	0.95	1.98	0.90	1.41	0.90	1.89
	2	0.97	1.53	0.92	1.78	0.97	1.92	0.97	1.33	0.89	1.82
	3	0.97	1.52	0.95	1.68	0.96	1.82	0.97	1.29	0.91	1.68
	4	0.97	1.51	1.00	1.59	0.99	1.81	0.98	1.34	0.93	1.61
2	1	0.94	1.50	0.89	1.85	0.96	1.90	0.93	1.37	0.87	1.76
	2	0.95	1.53	0.97	1.79	0.96	1.80	0.99	1.31	0.94	1.56
	3	0.97	1.55	0.98	1.72	0.97	1.92	0.97	1.32	0.92	1.49
	4	0.97	1.66	0.99	1.70	0.97	1.82	0.99	1.39	0.90	1.69
3	1	0.88	1.48	0.96	1.92	0.93	1.88	0.96	1.51	0.90	1.51
	2	0.95	1.57	0.95	1.76	0.95	1.90	0.97	1.38	0.88	1.44
	3	0.95	1.65	0.98	1.73	0.97	2.00	0.97	1.39	0.94	1.47
	4	0.90	1.78	0.97	1.79	0.94	2.09	0.94	1.57	0.94	1.54
4	1	0.93	1.85	0.99	1.83	0.93	2.19	0.87	1.39	0.87	1.47
	2	0.96	1.79	0.98	1.79	0.97	2.13	0.95	1.26	0.89	1.47
	3	0.92	1.96	0.93	1.77	0.96	2.30	0.99	1.24	0.92	1.25
	4	0.97	2.04	0.91	1.78	0.95	2.20	0.87	1.53	0.84	1.67
5	1	0.84	1.68	0.94	1.78	0.90	2.00	0.88	1.43	0.93	1.43
	2	0.85	1.85	0.94	1.77	0.93	2.04	0.97	1.17	0.93	1.45
	3	0.86	1.58	0.93	1.71	0.90	2.11	0.86	1.23	0.83	1.53
	4	0.91	1.82	0.86	1.91	0.71	2.16	0.60	1.66	0.61	1.82

Literatur

- Albarede, J.L., Garry, P.J. & Vellas, P. (Hrsg.) (1993). *Facts and research in gerontology, Vol. 7*. Paris: Serdi.
- Allison, P.D. (1987). Estimation of linear models with incomplete data. In: Clogg, C.C. (Hrsg.), *Sociological methodology*, S. 71–103. San Francisco: Jossey-Bass.
- Arbuckle, J.L. (1996). Full information estimation in the presence of incomplete data. In: Marcoulides, G.A. & Schumacker, R.E. (Hrsg.), *Advanced statistical modeling*, S. 243–277. Mahwah: Lawrence Erlbaum.
- Arbuckle, J.L. & Wothke, W. (1999). *AMOS 4.0 user's guide*. Chicago: SmallWaters.
- Arminger, G. (1979). *Faktorenanalyse*. Stuttgart: Teubner.
- Baltes, P.B. (1968). Longitudinal and cross-sectional sequences in the study of age and generation effects. *Human Development*, 11, 145–171.
- Baltes, P.B. (1973). Prototypical paradigms and questions in life-span research on development and aging. *The Gerontologist*, 13, 458–467.
- Baltes, P.B. & Baltes, M.M. (1989). Erfolgreiches Altern: Mehr Jahre und mehr Leben. In: Baltes, M.M., Kohli, M. & Sames, K. (Hrsg.), *Erfolgreiches Altern*, S. 5–10. Bern, Stuttgart, Toronto: Hans Huber.
- Baltes, P.B. & Baltes, M.M. (1990). Psychological perspectives on successful aging: The model of selective optimization with compensation. In: Baltes, P.B. & Baltes, M.M. (Hrsg.), *Successful aging: Perspectives from the behavioral sciences*, S. 1–34. Cambridge: Cambridge University Press.
- Baltes, P.B., Reese, H.W. & Nesselroade, J.R. (Hrsg.) (1988). *Life-span developmental psychology: Introduction to research methods*. Hillsdale: Lawrence Erlbaum.
- Baur, P.A. & Okun, M.A. (1983). Stability of life satisfaction in late life. *The Gerontologist*, 23, 261–265.
- Becker, P. (1991). Theoretische Grundlagen. In: Abele, A. & Becker, P. (Hrsg.), *Wohlbefinden*, S. 13–49. Weinheim, München: Juventa.
- Bell, B., Rose, C.L. & Damon, A. (1972). The Normative Aging Study: An interdisciplinary and longitudinal study of health and aging. *Aging and Human Development*, 3, 5–17.
- Bentler, P.M. (1995). *EQS structural equations program manual*. Encino: Multivariate Software, Inc.
- Bentler, P.M. & Bonnet, D.G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, 88, 588–606.

- Bijleveld, C., Mooijaart, A., van der Kamp, L. & van der Kloot, W. (1998). Structural equation models for longitudinal data. In: Bijleveld, C. & van der Kamp, L. (Hrsg.), *Longitudinal data analysis*, S. 207–268. London, Thousand Oaks, New Dehli: Sage.
- Bollen, K.A. (1989). *Structural equations with latent variables*. New York: Wiley.
- Bollen, K.A. & Long, J.S. (Hrsg.) (1993). *Testing structural equation models*. Newbury Park, London, New Dehli: Sage.
- Bosch, K. (1994). *Elementare Einführung in die angewandte Statistik* (5 Aufl.). Braunschweig: Vieweg.
- Bradburn, N.M. (1969). *The structure of psychological well-being*. Chicago: Aldine.
- Bradburn, N.M. & Caplovitz, D. (1965). *Reports on happiness: A pilot study of behavior related to mental health*. Chicago: Aldine.
- Brandstätter, H. (1987). *Emotional responses to everyday life situations: An individual difference approach*. Thesenpapier präsentiert auf dem Kolloquium „Subjektives Wohlbefinden“, Bad Homburg.
- Brandstädter, J. (1990). Entwicklung im Lebensablauf. Ansätze und Probleme der Lebensspannen-Entwicklungspsychologie. In: Mayer, K.U. (Hrsg.), *Lebensverläufe und sozialer Wandel. Sonderheft 31/1990 der „Kölner Zeitschrift für Soziologie und Sozialpsychologie“*, S. 322–350. Opladen: Westdeutscher Verlag.
- Brandstädter, J. & Greve, W. (1992). Das Selbst im Alter: Adaptive und protektive Mechanismen. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 18, 269–297.
- Brandstädter, J. & Greve, W. (1994). The aging self: Stabilizing and protective processes. *Developmental Review*, 14, 52–80.
- Brandstädter, J. & Rothermund, K. (1998). Bewältigungspotentiale im höheren Alter: adaptive und protektive Prozesse. In: Kruse, A. (Hrsg.), *Psychosoziale Gerontologie, Band 1: Grundlagen*, S. 223–237. Göttingen, Bern, Toronto, Seattle: Hogrefe.
- Braun, M. & Mohler, P.P. (1991). Die allgemeine Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS). *ZUMA-Nachrichten*, 29, 7–25.
- Brown, R.L. (1994). Efficacy of the indirect approach for estimating structural equation models with missing data: A comparison of five methods. *Structural Equation Modeling*, 1, 287–316.
- Browne, M.W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, 37, 62–83.
- Browne, M.W. & Cudeck, R. (1993). Alternative ways of assessing model fit. In: Bollen, K.A. & Long, J.S. (Hrsg.), *Testing structural equation models*, S. 136–162. Newbury Park, London, New Dehli: Sage.

- Bryk, A.S. & Raudenbush, S.W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin*, 101, 147–158.
- Bryk, A.S. & Raudenbush, S.W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park: Sage.
- Buhlman, T. (1996). Determinanten des subjektiven Wohlbefindens. In: Zapf, W. & Habich, R. (Hrsg.), *Wohlfahrtsentwicklung im vereinten Deutschland*, S. 79–96. Berlin: Edition Sigma.
- Burkhauser, R.V., Kreyenfeld, M. & Wagner, G.G. (1997). The German socio-economic panel: A representative sample of reunited Germany and its parts. *Vierteljahrshefte zur Wirtschaftsforschung*, 66, 7–16.
- Busse, E.W. (1993). Duke University longitudinal studies of aging. *Zeitschrift für Gerontologie*, 26, 123–126.
- Butt, D.S. & Beiser, M. (1987). Successful aging: A theme for international psychology. *Psychology and Aging*, 2, 87–94.
- Byrne, B.M. (2001). *Structural equation modeling with AMOS*. Mahwah: Lawrence Erlbaum.
- Campbell, A., Converse, P.E. & Rogers, W.L. (1976). *The quality of American life*. New York: Russel Sage Foundation.
- Carstensen, L.L. (1993). Motivation for social contact across the life span: A theory of socioemotional selectivity. *Nebraska Symposium on Motivation*, 40, 209–254.
- Carstensen, L.L. (1998). A life span approach to social motivation. In: Heckhausen, J. & Dweck, C. (Hrsg.), *Motivation and self-regulation across the life span*, S. 341–364. New York: Cambridge University Press.
- Carstensen, L.L., Isaacowitz, D.M. & Charles, S.T. (1999). Taking time seriously. A theory of socioemotional selectivity. *American Psychologist*, 54, 165–181.
- Chen, C. (2001). Aging and life satisfaction. *Social Indicators Research*, 54, 57–79.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155–159.
- Collins, L.M. & Sayer, A.G. (Hrsg.) (2001). *New methods for the analysis of change*. Washington: American Psychological Association.
- Costa, P.T. & McCrae, R.R. (1980). Influence of extraversion and neuroticism on subjective well-being: Happy and unhappy people. *Journal of Personality and Psychology*, 38, 668–687.
- Costa, P.T. & McCrae, R.R. (1984). Personality is a lifelong determinant of wellbeing. In: Malatesta, C.Z. & Izard, C.E. (Hrsg.), *Emotion in adult development*, S. 141–157. Beverly Hills, London, New Dehli: Sage.
- Costa, P.T. & McCrae, R.R. (1993). Psychological research in the Baltimore Longitudinal Study of Aging. *Zeitschrift für Gerontologie*, 26, 138–141.

- Costa, P.T., McCrae, R.R. & Arenberg, D. (1983). Recent longitudinal research on personality and aging. In: Schaie, K.W. (Hrsg.), *Longitudinal studies of adult psychological development*, S. 222–265. New York, London: Guilford Press.
- Costa, P.T., McCrae, R.R. & Zonderman, A.B. (1987a). Environmental and dispositional influences on well-being: Longitudinal follow-up of an American national sample. *British Journal of Psychology*, 78, 299–306.
- Costa, P.T., Zonderman, A.B., McCrae, R.R., Cornoni-Huntley, J., Locke, B.Z. & Barbano, H.E. (1987b). Longitudinal analyses of psychological well-being in a national sample: Stability of mean levels. *Journal of Gerontology*, 42, 50–55.
- Creed, P.A. & Evans, B.M. (2002). Personality, well-being, and deprivation theory. *Personality and Individual Differences*, 33, 1045–1054.
- Cumming, E. & Henry, W.E. (1961). *Growing old, the process of disengagement*. New York: Basic Books.
- Dempster, A.P., Laird, N.M. & Rubin, D.B. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, Series B*, 39, 1–38.
- DeNeve, K.M. & Cooper, H. (1998). The happy personality: A meta-analysis of 137 personality traits and subjective well-being. *Psychological Bulletin*, 124, 197–229.
- Diener, E. (1984). Subjective well-being. *Psychological Bulletin*, 95, 542–575.
- Diener, E. & Emmons, R.A. (1984). The independence of positive and negative affect. *Journal of Personality and Social Psychology*, 47, 1105–1117.
- Diener, E. & Iran-Nejad, A. (1986). The relationship in experience between various types of affect. *Journal of Personality and Social Psychology*, 50, 1031–1038.
- Diener, E., Larsen, R.J. & Emmons, R.A. (1984). Person \times situation interactions: Choice of situations and congruence response models. *Journal of Personality and Social Psychology*, 47, 580–592.
- Diener, E. & Larsen, R.L. (1993). The experience of emotional well-being. In: Lewis, M. & Haviland, J.M. (Hrsg.), *Handbook of emotions*, Kap. 28, S. 405–415: Guilford Press.
- Diener, E. & Lucas, R.E. (1999). Personality and subjective well-being. In: Kahneman, D. & Diener, E. (Hrsg.), *Well-being: The foundations of hedonic psychology*, S. 213–229. New York: Russel Sage Foundation.
- Diener, E., Oishi, S. & Lucas, R.E. (2003). Personality, culture, and subjective well-being: Emotional and cognitive evaluations of life. *Annual Review of Psychology*, 54, 403–425.
- Diener, E., Sandvik, E. & Pavot, W. (1991). Happiness is the frequency, not intensity of positive versus negative affect. In: Strack, F., Argyle, M. & Schwarz, N. (Hrsg.), *Subjective well-being*, S. 119–139. Oxford: Pergamon Press.

- Diener, E., Sandvik, E., Pavot, W. & Fujita, F. (1992). Extraversion and subjective well-being in a U.S. national probability sample. *Journal of Research in Personality*, 26, 205–215.
- Diener, E., Smith, H. & Fujita, F. (1995). The personality structure of affect. *Journal of Personality and Social Psychology*, 69, 130–141.
- Diener, E. & Suh, E. (1997). Age and subjective well-being: An international analysis. *Annual Review of Gerontology and Geriatrics*, 17, 304–324.
- Diener, E., Suh, E.M., Lucas, R.E. & Smith, H.L. (1999). Subjective well-being: Three decades of progress. *Psychological Bulletin*, 125, 276–302.
- Diggle, P.J., Heagerty, P., Liang, K.-Y. & Zeger, S.L. (2002). *Analysis of longitudinal data* (2. Aufl.). Oxford: Oxford University Press.
- Dittmann-Kohli, F. (1995). *Das persönliche Sinnsystem*. Göttingen, Bern, Toronto, Seattle: Hogrefe.
- Donaldson, G. & Horn, J.L. (1992). Age, cohort, and time developmental muddles: Easy in practice, hard in theory. *Experimental Aging Research*, 18, 213–222.
- Doyle, D. & Forehand, M.J. (1984). Life satisfaction and old age. *Research on Aging*, 6, 432–488.
- Ekerdt, D.J., Bossé, R. & Levkoff, S. (1985). An empirical test for phases of retirement: Findings from the Normative Aging Study. *Journal of Gerontology*, 40, 95–101.
- Enders, C. K. (2001a). The impact of nonnormality on full information maximum likelihood estimation for structural equation models with missing data. *Psychological Methods*, 6, 352–370.
- Enders, C. K. (2001b). A primer on maximum likelihood algorithms available for use with missing data. *Structural Equation Modeling*, 8, 128–141.
- Enders, C. K. & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling*, 8, 430–457.
- Fan, X., Thompson, B. & Wang, L. (1999). Effects of sample size, estimation methods, and model specification on structural equation modeling fit indexes. *Structural Equation Modeling*, 6, 56–83.
- Feist, G.S., Booner, T.E., Jacobs, S.F., Miles, M. & Tan, V. (1995). Integrating top-down and bottom-up structural models of subjective well-being: A longitudinal investigation. *Journal of Personality and Social Psychology*, 68, 138–150.
- Ferring, D. & Filipp, S.-H. (1997). Subjektives Wohlbefinden im Alter: Struktur- und Stabilitätsanalysen. *Psychologische Beiträge*, 39, 236–258.
- Filipp, S.-H. & Mayer, A.-K. (1999). *Bilder des Alters*. Stuttgart: Kohlhammer.
- Fingerman, K.L. & Perlmutter, M. (1995). Future time perspective and life events across adulthood. *Journal of General Psychology*, 122, 95–111.

- Fox, C.R. & Kahneman, D. (1992). Correlation, causes, and heuristics in survey of life satisfaction. *Social Indicators Research*, 27, 221–234.
- Frick, R.W. (1996). The appropriate use of null hypothesis testing. *Psychological Methods*, 1, 379–390.
- Fries, J.F. (1991). Medical perspectives upon successful aging. In: Baltes, P.B. & Baltes, M.M. (Hrsg.), *Successful aging: Perspectives from the behavioral sciences*, S. 35–49. New York: Cambridge University Press.
- Goldstein, H. (1986). Efficient statistical modelling of longitudinal data. *Annals of Human Biology*, 13, 129–141.
- Goldstein, H. (1995). *Multilevel statistical models*. London: Edward Arnold.
- Graham, J. W. & Hofer, S. M. (2000). Multiple imputation in multivariate research. In: Little, T.D., Schnabel, K.U. & Baumert, J. (Hrsg.), *Modeling longitudinal and multilevel data: Practical issues, applied approaches, and specific examples*, S. 201–218. Mahwah: Lawrence Erlbaum.
- Graham, J.W. & Donaldson, S.I. (1993). Evaluating interventions with differential attrition: The importance of nonresponse mechanisms and use of follow-up-data. *Journal of Applied Psychology*, 78, 119–128.
- Graham, J.W., Taylor, B.J. & Cumsille, P.E. (2001). Planned missing-data designs in analysis of change. In: Collins, L.M. & Sayer, A.G. (Hrsg.), *New methods for the analysis of change*, S. 335–353. Washington: American Psychological Association.
- Grombach, H. (1976). Consistency and change of personality variables in late life. In: Thomae, H. (Hrsg.), *Patterns of aging*, S. 51–67. Basel, New York: Karger.
- Grombach, H. (1987). Konstanz und Veränderlichkeit der Persönlichkeit als eine Funktion methodischer Grundlagen. In: Lehr, U. & Thomae, H. (Hrsg.), *Formen seelischen Alterns*, S. 18–38. Stuttgart: Enke.
- Hager, W. & Westermann, R. (1983). Planung und Auswertung von Experimenten. In: Bredenkamp, J. & Feger, H. (Hrsg.), *Hypothesenprüfung. Enzyklopädie der Psychologie, Themenbereich B, Serie I, Band 5*, Kap. 2, S. 24–239. Göttingen, Toronto, Zürich: Hogrefe.
- Haisken-De New, J.P. & Frick, J.R. (2000). *Desktop Companion to the German Socio-Economic Panel Study (GSOEP)*, Version 4.0. Berlin: Deutsches Institut für Wirtschaftsforschung.
- Hasher, L. & Zacks, R.T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General*, 108, 356–388.
- Hasher, L. & Zacks, R.T. (1984). Automatic processing of fundamental information: The case of frequency of occurrence. *American Psychologist*, 39, 1372–1388.
- Havighurst, R.J. (1961). Successful aging. *Gerontologist*, 1, 8–13.

- Hayduk, L.A. & Glaser, D.N. (2000). Jiving† the four-step, waltzing around factor analysis, and other serious fun. *Structural Equation Modeling*, 7, 1–35.
- Hayes, N. & Joseph, S. (2003). Big 5 correlates of three measures of subjective well-being. *Personality and Individual Differences*, 34, 723–727.
- Hays, W.L. (1994). *Statistics* (5. Aufl.). Fort Worth: Harcourt Brace.
- Headey, B., Veenhoven, R. & Wearing, A. (1991). Top-down versus bottom-up theories of subjective well-being. *Social Indicators Research*, 24, 81–100.
- Headey, B. & Wearing, A. (1991). Subjective well-being: a stocks and flows framework. In: Strack, F., Argyle, M. & Schwarz, N. (Hrsg.), *Subjective well-being*, S. 49–73. Oxford: Pergamon Press.
- Heckhausen, J. & Brim, O.G. (1997). Perceived problems for self and others: Self-protection by social downgrading throughout adulthood. *Psychology and Aging*, 12, 610–619.
- Heckhausen, J., Dixon, R.A. & Baltes, P.B. (1989). Gains and losses in development throughout adulthood as perceived by different adult age groups. *Developmental Psychology*, 25, 109–121.
- Heise, D.R. (1969). Separating reliability and stability in test-retest correlation. *American Sociological Review*, 34, 93–101.
- Hershberger, S.L., Molenaar, P.C.M. & Corneal, S.E. (1996). A hierarchy of univariate and multivariate structural time series models. In: Marcoulides, G.A. & Schumacker, R.E. (Hrsg.), *Advanced statistical modeling*, S. 159–194. Mahwah: Lawrence Erlbaum.
- Herzog, A.R. & Rodgers, W.L. (1981). Age and satisfaction: Data from several large surveys. *Research on Aging*, 3, 142–165.
- Hoff, A., Tesch-Römer, C., Wurm, S. & Engstler, H. (2003). „Die zweite Lebenshälfte“: Längsschnittliche Konzeption des Alterssurveys. In: Karl, F. (Hrsg.), *Einführung in die sozial- und verhaltenswissenschaftliche Gerontologie*, S. 158–204. Weinheim: Juventa.
- Horley, J. & Lavery, J.J. (1995). Subjective well-being and age. *Social Indicators Research*, 34, 275–282.
- Hox, J.J. (2002). *Multilevel analysis*. Mahwah: Lawrence Erlbaum.
- Hox, J.J. & Bechger, T.M. (1998). An introduction to structural equation modeling. *Family Science Review*, 11, 354–373.
- Hoyle, R.H. (Hrsg.) (1995). *Structural equation modeling*. Thousand Oaks, London, New Dehli: Sage.
- Hoyle, R.H. & Panter, A.T. (1995). Writing about structural equation models. In: Hoyle, R.H. (Hrsg.), *Structural equation modeling*, S. 158–176. Thousand Oaks, London, New Dehli: Sage.
- Hu, L. & Bentler, P.M. (1995). Evaluating model fit. In: Hoyle, R.H. (Hrsg.), *Structural equation modeling*, S. 76–99. Thousand Oaks, London, New Dehli: Sage.

- Hu, L. & Bentler, P.M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, 3, 424–453.
- Hu, L. & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1–55.
- Hultsch, D.F., Hertzog, C., Dixon, R.A. & Small, B.J. (1998). *Memory changes in the aged*. Cambridge: Cambridge University Press.
- Infratest Sozialforschung (2001). *SOEP 2000. Methodenbericht zum Befragungsjahr 2000 (Welle 17) des Sozio-oekonomischen Panels*. München: Infratest Sozialforschung.
- Infratest Sozialforschung (2002). *Verbesserung der Datengrundlagen für Mortalitäts- und Mobilitätsanalysen: Verbleibstudie bei Panelausfällen im SOEP*. München: Infratest Sozialforschung.
- Inglehart, R. (1990). *Culture shift in advanced industrial society*. Princeton: Princeton University Press.
- Jöreskog, K.G. (1967). Some contributions to maximum likelihood factor analysis. *Psychometrika*, 32, 443–482.
- Jöreskog, K.G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, 34, 183–202.
- Jöreskog, K.G. (1970). Estimation and testing of simplex models. *British Journal of Mathematical and Statistical Psychology*, 23, 121–145.
- Kenward, M.G. (1998). Selection models for repeated measurements with non-random dropout: An illustration of sensitivity. *Statistics in Medicine*, 17, 2723–2732.
- Kozma, A. & Stones, M.J. (1978). Some research issues and findings in the study of psychological well-being in the aged. *Canadian Psychological Review*, 19, 241–249.
- Kozma, A. & Stones, M.J. (1983). Predictors of happiness. *Journal of Gerontology*, 38, 626–628.
- Kraemer, H.C. & Thiemann, S. (1987). *How many subjects?* Newbury Park, London, New Delhi: Sage.
- Kunzmann, U., Little, T.D. & Smith, J. (2000). Is age-related stability of subjective well-being a paradox? Cross-sectional and longitudinal evidence from the Berlin Aging Study. *Psychology and Aging*, 15, 511–526.
- Labouvie-Vief, G. (1998). Cognitive-emotional integration in adulthood. In: Schaie, K.W. & Lawton, M.P. (Hrsg.), *Annual Review of Gerontology and Geriatrics*, S. 206–237. New York: Springer.
- Lachman, M.E. & Weaver, S.L. (1998). The sense of control as a moderator of social class differences in health and well-being. *Journal of Personality and Social Psychology*, 74, 763–773.
- Lance, C.E., Lautenschlager, G.J., Sloan, C.E. & Varca, P.E. (1989). A comparison between bottom-up, top-down, and bidirectional models of relationships between global and life facet satisfaction. *Journal of Personality*, 57, 601–624.

- Lance, C.E., Mallard, A.G. & Michalos, A.C. (1995). Tests of the causal directions of global-life facet satisfaction relationship. *Social Indicators Research*, 34, 69–92.
- Landua, D. (1993). Veränderungen von Zufriedenheitsangaben in Panelbefragungen. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 45, 553–571.
- Larson, R. (1978). Thirty years of research on the subjective well-being of older Americans. *Journal of Gerontology*, 33, 109–125.
- Lawton, M.P., Kleban, M.H., Rajagopal, D. & Dean, J. (1992). Dimensions of affective experience in three age groups. *Psychology and Aging*, 7, 171–184.
- Lee, S.-Y., Poon, W.-Y. & Bentler, P.M. (1992). Structural equation models with continuous and polytomous variables. *Psychometrika*, 57, 89–105.
- Lee, S.-Y., Poon, W.-Y. & Bentler, P.M. (1995). A two-stage estimation of structural equation models with continuous and polytomous variables. *British Journal of Mathematical and Statistical Psychology*, 48, 339–358.
- Lehr, U. (1991). *Psychologie des Alterns* (7 Aufl.). Heidelberg, Wiesbaden: Quelle & Meyer.
- Lewis-Beck, M.S. (Hrsg.) (1994). *Factor analysis and related techniques*. Thousand Oaks, London, New Dehli: Sage.
- Lienert, G.A. (1961). *Testaufbau und Testanalyse*. Weinheim, Berlin, Basel: Beltz.
- Lienert, G.A. & Raatz, U. (1994). *Testaufbau und Testanalyse* (5. Aufl.). Weinheim: Psychologie Verlags Union.
- Little, R.J. & Rubin, D.B. (1987). *Statistical analysis with missing data*. New York: Wiley.
- Little, R.J. & Schenker, N. (1995). Missing data. In: Arminger, G, Clogg, C.C. & Sobel, M.E. (Hrsg.), *Handbook of statistical modeling for the social and behavioral sciences*, S. 39–75. New York: Plenum.
- Little, T.D., Lindenberger, U. & Maier, H. (2000). Selectivity and generalizability in longitudinal research: On the effects of continuers and dropouts. In: Little, T.D., Schnabel, K.U. & Baumert, J. (Hrsg.), *Modeling longitudinal and multilevel data: Practical issues, applied approaches, and specific examples*, S. 187–200. Mahwah: Lawrence Erlbaum.
- Long, J.S. (1983). *Covariance structure models*. Newbury Park, London, New Dehli: Sage.
- Luszcz, M.A. (1999). *Psychological aging: themes and variations*. Adelaide: South Australian Network for Research on Aging (SANRA), Occasional Paper No. 3. Internet-Dokument: www.cas.flinders.edu.au/sanra/occasional/paper3.html (30.05.2003).
- MacCallum, R.C. (1995). Model specification: Procedures, strategies and related issues. In: Hoyle, R.H. (Hrsg.), *Structural equation modeling*, S. 16–37. Thousand Oaks, London, New Dehli: Sage.

- Mallard, A.G., Lance, C.E. & Michalos, A.C. (1997). Culture as a moderator of overall life satisfaction - life facet satisfaction relationship. *Social Indicators Research*, 40, 259–284.
- Mardia, K.V. (1985). Mardia's test of multinormality. In: Kotz, S. & Johnson, N.L. (Hrsg.), *Encyclopedia of statistical sciences*, Bd. 5, S. 217–221. New York: Wiley.
- Marsh, H.W., Balla, J.R. & Hau, K.-T. (1996). An evaluation of incremental fit indices: A clarification of mathematical and empirical properties. In: Marcoulides, G.A. & Schumacker, R.E. (Hrsg.), *Advanced statistical modeling*, S. 315–353. Mahwah: Lawrence Erlbaum.
- Martin, P., Grünendahl, M. & Schmitt, M. (2000). Persönlichkeit, kognitive Leistungsfähigkeit und Gesundheit in Ost und West: Ergebnisse der Interdisziplinären Längsschnittstudie des Erwachsenenalters (ILSE). *Zeitschrift für Gerontologie und Geriatrie*, 33, 111–123.
- Mayer, K.U. & Baltes, M.M. (Hrsg.) (1996). *Die Berliner Altersstudie*. Berlin: Akademie Verlag.
- Mayring, P. (1987). Subjektives Wohlbefinden im Alter. Stand der Forschung und theoretische Weiterentwicklung. *Zeitschrift für Gerontologie*, 20, 367–376.
- Mayring, P. (1990). Glück und Wohlbefinden im Alter. In: Mayring, P. & Saup, W. (Hrsg.), *Entwicklungsprozesse im Alter*, S. 167–183. Stuttgart, Berlin, Köln: Kohlhammer.
- Mayring, P. (1991). Die Erfassung subjektiven Wohlbefindens. In: Abele, A. & Becker, P. (Hrsg.), *Wohlbefinden*, S. 167–183. Weinheim, München: Juventa.
- McArdle, J.J. (1986). Latent variable growth within behavior genetic models. *Behavior Genetics*, 16, 163–200.
- McArdle, J.J. (1998). Modeling longitudinal data by latent growth curve methods. In: Marcoulides, G.A. (Hrsg.), *Modern methods for business research*, S. 359–406. Mahwah: Lawrence Erlbaum.
- McArdle, J.J. & Aber, M.S. (1990). Patterns of change within latent variable structural equation models. In: von Eye, A. (Hrsg.), *Statistical models in longitudinal research*, Bd. 1, S. 151–224. San Diego, London: Academic Press.
- McArdle, J.J. & Epstein, D. (1987). Latent growth curves within developmental structural equation models. *Child Development*, 58, 110–133.
- McArdle, J.J. & Hamagami, F. (2001). Latent difference score structural models for linear dynamic analyses with incomplete longitudinal data. In: Collins, L.M. & Sayer, A.G. (Hrsg.), *New methods for the analysis of change*, S. 139–175. Washington: American Psychological Association.
- McArdle, J.J. & Nesselroade, J.R. (1994). Using multivariate data to structure developmental change. In: Cohen, S.H. & Reese, H.W. (Hrsg.), *New methods for the analysis of change*, S. 223–267. Hillsdale: Lawrence Erlbaum.
- McNeil, J.K., Stones, M.J. & Kozma, A. (1986). Longitudinal variation in domain indicators of happiness. *Social Indicators Research*, 18, 119–124.
- Meredith, W. & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, 55, 107–122.

- Michalos, A.C. (1985). Multiple discrepancies theory (MDT). *Social Indicators Research*, 16, 347–413.
- Michalos, A.C. (1986). An application of multiple discrepancies theory (MDT) to seniors. *Social Indicators Research*, 18, 349–373.
- Mirowski, J. & Ross, C.E. (1999). Well-being across the life course. In: Horwitz, A.V. & Scheid, T.L. (Hrsg.), *A handbook for the mental health: Social contexts, theories, and systems*, S. 328–347. New York: Cambridge University Press.
- Mroczek, D.K. & Spiro, A. (2003). Modeling intraindividual change in personality traits: Findings from the Normative Aging Study. *Journal of Gerontology: Psychological Sciences*, 58B, P153–P165.
- Mulaik, S.A. & James, L.R. (1995). Objectivity and reasoning in science and structural equation modeling. In: Hoyle, R.H. (Hrsg.), *Structural equation modeling*, S. 118–137. Thousand Oaks, London, New Dehli: Sage.
- Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*, 49, 115–132.
- Muthén, B., Kaplan, D. & Hollis, M. (1987). On structural equation modeling with data that are not missing at random. *Psychometrika*, 52, 431–462.
- Okun, M.A. (1998). Subjective well-being. In: Maddox, G.L. (Hrsg.), *The encyclopedia of aging* (2 Aufl.), S. 909–912. New York: Springer.
- Palmore, E. & Kivett, V. (1977). Change in life satisfaction: A longitudinal study of persons aged 46-70. *Journal of Gerontology*, 32, 311–316.
- Palmore, E. & Kivett, V. (1985). Change in life satisfaction. In: Palmore, E., Busse, E.W., Maddox, G.L., Nowlin, J.B. & Siegler, I.C. (Hrsg.), *Normal aging III*, S. 373–380. Durham: Duke University Press.
- Pedersen, N.L., McClearn, G.E., Plomin, R., Nesselroade, J.R, Berg, S. & DeFaire, U. (1991). The Swedish Adoption/Twin Study of Aging: An update. *Acta Geneticae Medice et Gemellologiae*, 40, 7–20.
- Perrig-Chiello, P., Perrig, W.J., Stähelin, H.B., Krebs-Roubicek, E. & Ehram, R. (1996). Wohlbefinden, Gesundheit und Autonomie im Alter: Die Basler IDA Studie (Interdisziplinäre Altersstudie). *Zeitschrift für Gerontologie und Geriatrie*, 29, 95–109.
- Raykov, T. & Widaman, K.F. (1995). Issues in applied structural equation modeling research. *Structural Equation Modeling*, 2, 289–318.
- Rietz, C., Rudinger, G. & Andres, J. (1996). Lineare Strukturgleichungsmodelle. In: Erdfelder, E., Mausfeld, R., Meiser, T. & Rudinger, G. (Hrsg.), *Handbuch Quantitative Methoden*, S. 253–268. Weinheim: Psychologie Verlags Union.

- Rindskopf, D. (1983). Parameterizing inequality constraints on unique variances in linear structural equation models. *Psychometrika*, 48, 73–83.
- Rindskopf, D. (1984). Using phantom and imaginary latent variables to parameterize constraints in linear structural equation models. *Psychometrika*, 49, 37–47.
- Rovine, M.J. & von Eye, A. (1991). *Applied computational statistics in longitudinal research*. San Diego, London: Academic Press.
- Rubin, D.B. (1976). Inference and missing data. *Biometrika*, 63, 581–592.
- Rubin, D.B. (1987). *Multiple imputation for nonresponse in surveys*. New York: Wiley.
- Rudinger, G. & Rietz, C. (1993). Und mal wieder: Reliabilität und Stabilität in Panelmodellen. *ZUMA-Nachrichten* 32, 17, 60–75.
- Rudinger, G. & Rietz, C. (1998). The neglected time dimension? Introducing a longitudinal model testing latent growth curves, stability, and reliability as time bound processes. *Methods of Psychological Research Online*, 3, 109–130. Internet: <http://www.mpr-online.de/> (15.09.2003).
- Rudinger, G. & Rietz, C. (2001). Structural equation modeling in longitudinal research on aging. In: Birren, J.E. & Schaie, K.W. (Hrsg.), *Handbook of the psychology of aging* (5 Aufl.), S. 29–52. San Diego (et al.): Academic Press.
- Rule, S.J. (1976). A general experimentwise error rate for multiple significance tests. *Perceptual and Motor Skills*, 43, 1263–1277.
- SAS Institute Inc. (1990). *SAS procedures guide, version 6*. Cary, NC: SAS Institute Inc.
- Schafer, J.L. & Graham, J.W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, 7, 147–177.
- Schafer, J.L. & Olsen, M.K. (1998). Multiple imputation for multivariate missing-data problems: A data analyst's perspective. *Multivariate Behavioral Research*, 33, 545–571.
- Schaie, K.W. (1964). A general model for the study of developmental problems. *Psychological Bulletin*, 64, 92–107.
- Schaie, K.W. (1977). Quasi experimental designs in the psychology of aging. In: Birren, J.E. & Schaie, K.W. (Hrsg.), *Handbook of the psychology of aging*, S. 39–58. New York: Van Nostrand Reinhold.
- Schaie, K.W. (1983). The Seattle Longitudinal Study: A 21-year exploration of psychometric intelligence in adulthood. In: Schaie, K.W. (Hrsg.), *Longitudinal studies of adult psychological development*, S. 64–135. New York, London: Guilford Press.
- Schaie, K.W. & Hofer, S.M. (2001). Longitudinal studies in aging research. In: Birren, J.E. & Schaie, K.W. (Hrsg.), *Handbook of the psychology of aging* (5 Aufl.), S. 53–77. San Diego (et al.): Academic Press.

- Scheier, M.F. & Carver, C.S. (1985). Optimism, coping, and health: Assessment and implications of generalized outcome expectancies. *Health Psychology*, 4, 219–247.
- Scherpenzeel, A. (1995). *A question of quality: Evaluating survey questions by multitrait-multimethod studies*. Amsterdam: Nimmo.
- Scherpenzeel, A. & Saris, W. (1996). Causal direction in a model of life satisfaction: The top-down/bottom-up controversy. *Social Indicators Research*, 38, 161–180.
- Schilling, O. (1998). *Grundkurs: Statistik für Psychologen*. München: Fink.
- Schimmack, U., Radhakrishnan, P., Oishi, S., Dzokoto, V. & Ahadi, S. (2002). Culture, personality, and subjective well-being: Integrating process models of life satisfaction. *Journal of Personality and Social Psychology*, 82, 582–593.
- Schmitz-Scherzer, R. & Thomae, H. (1983). Constancy and change of behavior in old age: Findings from the Bonn Longitudinal Study on Aging. In: Schaie, K.W. (Hrsg.), *Longitudinal studies of adult psychological development*, S. 191–221. New York, London: Guilford Press.
- Schönemann, P.H. & Borg, I. (1969). Von der Faktorenanalyse zu den Strukturgleichungsmodellen. In: Erdfelder, E., Mausfeld, R., Meiser, T. & Rudinger, G. (Hrsg.), *Handbuch Quantitative Methoden*, S. 241–252. Weinheim: Psychologie Verlags Union.
- Schräpler, J.P. (1995). *Die Schätzung von Reliabilität und Stabilität der Zufriedenheitsangaben im Sozio-oekonomischen Panel*. Bochum: Ruhr-Universität, Fakultät für Sozialwissenschaft, Diskussionspapier Nr. 95-13.
- Schwarz, N. & Clore, G.L. (1983). Mood, misattribution and judgements of well-being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology*, 45, 513–523.
- Schwarz, N. & Strack, F. (1991). Evaluating one's life: a judgement model of subjective well-being. In: Strack, F., Argyle, M. & Schwarz, N. (Hrsg.), *Subjective well-being*, S. 27–47. Oxford: Pergamon Press.
- Sedlmeier, P. (1996). Jenseits des Signifikanztest-Rituals: Ergänzungen und Alternativen. *Methods of Psychological Research Online*, 1, 41–63. Internet: <http://www.mpr-online.de/> (15.09.2003).
- Shapiro, S.S. & Wilk, M.B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52, 591–611.
- Shmotkin, D. (1991). The role of time orientation in life satisfaction across the life span. *Journal of Gerontology: Psychological Sciences*, 46, P243–P250.
- Singer, J.D. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Sciences*, 24, 323–355.
- Small, B.J., Hertzog, C., Hulstsch, D.F. & Dixon, R.A. (2003). Stability and change in adult personality over 6 years: Findings from the Victoria Longitudinal Study. *Journal of Gerontology: Psychological Sciences*, 58B, P166–P176.

- Smith, J., Fleeson, W., Geiselmann, B., Settersten, R. & Kunzmann, U. (1996). Wohlbefinden im hohen Alter: Vorhersagen aufgrund objektiver Lebensbedingungen und subjektiver Bewertung. In: Mayer, K.U. & Baltes, P.B. (Hrsg.), *Die Berliner Altersstudie*, S. 497–523. Berlin: Akademie Verlag.
- Smith, J. & Freund, A.M. (2002). The dynamics of possible selves in old age. *Journal of Gerontology: Psychological Sciences*, 57B, P492–P500.
- Smith, J., Maas, I., Mayer, K.U., Helmchen, H., Steinhagen-Thiessen, E. & Baltes, P.B. (2002). Two-wave longitudinal findings from the Berlin Aging Study: Introduction to a collection of articles. *Journal of Gerontology: Psychological Sciences*, 57B, P471–P473.
- Snijders, T.A.B. & Bosker, R.J. (1999). *Multilevel analysis*. London, Thousand Oaks, New Dehli: Sage.
- SOEP Group (2001). The German Socio-Economic Panel (GSOEP) after more than 15 years - Overview. *Vierteljahrshefte zur Wirtschaftsforschung*, 70(1), 7–14.
- Staudinger, U. (2000). Viele Gründe sprechen dagegen, und trotzdem geht es vielen Menschen gut: Das Paradox des subjektiven Wohlbefindens. *Psychologische Rundschau*, 51, 185–197.
- Staudinger, U., Marsiske, M. & Baltes, P.B. (1995). Resilience and reserve capacity in later adulthood: Potentials and limits of development across the life span. In: Cichetti, D. & Cohen, D. (Hrsg.), *Developmental Psychopathology*, Bd. 2, S. 801–847. New York: Wiley.
- Steel, P. & Ones, D.S. (2002). Personality and happiness: A national level analysis. *Journal of Personality and Social Psychology*, 83, 767–781.
- Steiger, J.H. (2000). Point estimation, hypothesis testing, and intervall estimation using the RMSEA: some comments and a reply to Hayduk and Glaser. *Structural Equation Modeling*, 7, 149–162.
- Stevens, J. (1996). *Applied multivariate statistics for the social sciences* (3. Aufl.). Lawrence Erlbaum.
- Steyer, R. & Eid, M. (1993). *Messen und Testen*. Berlin, Heidelberg: Springer.
- Steyer, R., Eid, M. & Schwenkmezger, P. (1997). Modeling true intraindividual change: True change as a latent variable. *Methods of Psychological Research Online*, 2, 21–33. Internet: <http://www.mpr-online.de/> (15.09.2003).
- Steyer, R., Partchev, I. & Shanahan, M.J. (2000). Modeling true intraindividual change in structural equation models: The case of poverty and children's psychosocial adjustment. In: Little, T.D., Schnabel, K.U. & Baumert, J. (Hrsg.), *Modeling longitudinal and multilevel data: Practical issues, applied approaches, and specific examples*, S. 109–126. Mahwah: Lawrence Erlbaum.
- Stock, W.A., Okun, M.A. & Benin, M. (1986). The structure of subjective well-being among the elderly. *Psychology and Aging*, 1, 91–102.

- Strack, F., Schwarz, N. & Gschneidinger, E. (1985). Happiness and reminiscing: The role of time perspective, mood, and mode of thinking. *Journal of Personality and Social Psychology*, 49, 1460–1469.
- Stumpf, H. (1996). Klassische Testtheorie. In: Erdfelder, E., Mausfeld, R., Meiser, T. & Rudinger, G. (Hrsg.), *Handbuch Quantitative Methoden*, S. 411–430. Weinheim: Psychologie Verlags Union.
- Tanaka, J.S. (1993). Multifaceted conceptions of fit in structural equation models. In: Bollen, K.A. & Long, J.S. (Hrsg.), *Testing structural equation models*, S. 10–39. Newbury Park, London, New Dehli: Sage.
- Tesch-Römer, C.T., Salewski, C. & Schwarz, G. (Hrsg.) (1997). *Psychologie der Bewältigung*. Weinheim: Psychologie Verlags Union.
- Thomae, H. (1987). Alternsformen – Wege zu ihrer methodischen und begrifflichen Erfassung. In: Lehr, U. & Thomae, H. (Hrsg.), *Formen seelischen Alterns*, S. 172–195. Stuttgart: Enke.
- Thomae, H. (1993). The Bonn Longitudinal Study on Aging (BOLSA). *Zeitschrift für Gerontologie*, 26, 142–150.
- Thomas, D. & Diener, E. (1990). Memory accuracy in the recall of emotions. *Journal of Personality and Social Psychology*, 59, 291–297.
- Tisak, J. & Meredith, W. (1990). Descriptive and associative developmental models. In: von Eye, A. (Hrsg.), *Statistical methods in longitudinal research*, Bd. II, S. 387–406. San Diego: Academic Press.
- Tisak, J. & Tisak, M.S. (1996). Longitudinal models of reliability and validity: A latent curve approach. *Applied Psychological Measurement*, 20, 275–288.
- Tobin, S.S. & Neugarten, B.L. (1961). Life satisfaction and social interaction in the aging. *Journal of Gerontology*, 16, 344–346.
- van der Kamp, L. & Bijleveld, C. (1998). Methodological issues in longitudinal research. In: Bijleveld, C. & van der Kamp, L. (Hrsg.), *Longitudinal data analysis*, S. 1–45. London, Thousand Oaks, New Dehli: Sage.
- van der Leeden, R. (1998). Multilevel analysis of longitudinal data. In: Bijleveld, C. & van der Kamp, L. (Hrsg.), *Longitudinal data analysis*, S. 269–317. London, Thousand Oaks, New Dehli: Sage.
- Veenhoven, R. (1984). *Conditions of happiness*. Dordrecht: D. Reidel.
- Veenhoven, R. (1991). Is happiness relative? *Social Indicators Research*, 24, 1–34.
- Veenhoven, R. (1994). Is happiness a trait? *Social Indicators Research*, 32, 101–160.
- Veenhoven, R. (1996). Developments in satisfaction research. *Social Indicators Research*, 37, 1–46.

- Veenhoven, R. (2003). *World database of happiness, correlational findings*. Internet-Dokument: www.eur.nl/fws/research/happiness (01.08.2003).
- Verbeke, G. & Molenberghs, G. (2000). *Linear mixed models for longitudinal data*. New York: Springer.
- Wahl, H.-W. (2003). Verhaltens- und sozialwissenschaftliche Gerontologie in ihrer geschichtlichen Entwicklung. In: Karl, F. (Hrsg.), *Einführung in die Sozial- und verhaltenswissenschaftliche Gerontologie*, S. 87–110. Weinheim: Juventa.
- Wahl, H.-W. & Heyl, V. (2003). *Gerontologie: Einführung und Geschichte*. Stuttgart: Kohlhammer.
- Werner, J. (1997). *Lineare Statistik*. Weinheim: Psychologie Verlags Union.
- Werts, C.E., Jöreskog, K.G. & Linn, R.L. (1971). Comment on "The estimation of measurement error in panel data". *American Sociological Review*, 36, 110–113.
- West, S.G., Finch, J.F. & Curran, P. J. (1995). Structural equation models with nonnormal variables. In: Hoyle, R.H. (Hrsg.), *Structural equation modeling*, S. 56–75. Thousand Oaks, London, New Dehli: Sage.
- Wiley, D.E. & Wiley, J.A. (1970). The estimation of measurement error in panel data. *American Sociological Review*, 35, 112–117.
- Wilson, W. (1967). Correlates of avowed happiness. *Psychological Bulletin*, 67, 294–306.
- Wothke, W. (1993). Nonpositive definite matrices in structural modeling. In: Bollen, K.A. & Long, J.S. (Hrsg.), *Testing structural equation models*, S. 256–293. Newbury Park, London, New Dehli: Sage.
- Wothke, W. (2000). Longitudinal and multigroup modeling with missing data. In: Little, T.D., Schnabel, K.U. & Baumert, J. (Hrsg.), *Modeling longitudinal and multilevel data: Practical issues, applied approaches, and specific examples*, S. 219–240. Mahwah: Lawrence Erlbaum.
- Zapf, W. & Habich, R. (Hrsg.) (1996). *Wohlfahrtsentwicklung im vereinten Deutschland*. Berlin: edition sigma.
- Zimprich, D. (2002). *Kognitive Entwicklung im Alter*. Dissertation, Fakultät für Sozial- und Verhaltenswissenschaften der Ruprecht-Karls-Universität Heidelberg, Heidelberg.