

S/SS

**Schriftenreihe des Instituts für Sozialwissenschaften
der Universität Stuttgart**

No. 1 / 2013

***Panelanalyse mit
Mehrebenenmodellen.
Eine anwendungsorientierte
Einführung.***

Thomas Krause

Dieter Urban



Universität Stuttgart

Abteilung für Soziologie und empirische Sozialforschung

**SISS:
Schriftenreihe
des Instituts für Sozialwissenschaften
der Universität Stuttgart:
No. 1 / 2013**

Panelanalyse mit Mehrebenenmodellen.
Eine anwendungsorientierte Einführung.

Thomas Krause
Dieter Urban

ISSN 0945-9197

**Universität Stuttgart
Institut für Sozialwissenschaften
Abt. f. Soziologie u. empirische Sozialforschung
70550 Stuttgart**

Thomas Krause und Dieter Urban

Panelanalyse mit Mehrebenenmodellen. Eine anwendungsorientierte Einführung.

SISS–Schriftenreihe des Instituts für Sozialwissenschaften der Universität Stuttgart. No. 1/2013.
(<http://www.uni-stuttgart.de/soz/institut/forschung/2013.SISS.1.pdf>)

Z U S A M M E N F A S S U N G: Dieser Beitrag beschreibt, wie Mehrebenenmodelle zur Längsschnittanalyse mit Panel-Daten eingesetzt werden können. Besondere Aufmerksamkeit wird dabei sog. Hybridmodellierungen geschenkt, welche einige Vorteile von Fixed-Effect- und Random-Effects-Panelregressionen miteinander kombinieren. Es werden grundlegende Konzepte von Mehrebenenanalyse (MLA) und Panelanalyse vorgestellt, aber auch Besonderheiten wie z.B. die Variablenzentrierung in Längsschnittuntersuchungen verdeutlicht. Der konkrete Analyseablauf einer solchen Analyse wird schrittweise anhand eines Beispiels mit den Daten des Sozio-oekonomischen Panels und der SPSS-MIXED-Software aufgezeigt. Jede/-r Sozialforscher/-in sollte nach der Lektüre dieses Beitrags dazu befähigt sein, seine/ihre erste eigene Längsschnittanalyse unter Zuhilfenahme einer Mehrebenenmodellierung durchzuführen.

Thomas Krause and Dieter Urban

Panel analysis with multilevel models. An application-oriented introduction.

SISS–Schriftenreihe des Instituts für Sozialwissenschaften der Universität Stuttgart. No. 1/2013.
(<http://www.uni-stuttgart.de/soz/institut/forschung/2013.SISS.1.pdf>)

A B S T R A C T: This paper describes how to use multi-level models for longitudinal studies with panel data. A special focus is set on the specification and estimation of hybrid models combining some advantages of models for fixed-effect and random-effects panel regression. The paper addresses some basic principles of multi-level- and panel analysis but it also discusses some very special problems of longitudinal analysis like centering of variables. It presents an example of a stepwise, applied panel regression analysis within a multi-level-model approach using the data of the SOEP (the national German socio-economic panel study) and the SPSS-MIXED-software. After reading this paper every social research scholar should be able to perform his/her first longitudinal analysis by multi-level modeling.

1 Einführung

Für die sozialwissenschaftliche Untersuchung von Variablenzusammenhängen hat die Analyse von Längsschnittdaten entscheidende Vorteile gegenüber einer Analyse von Querschnittsdaten. Um diese Vorteile jedoch auch nutzen zu können, müssen Sozialforscher/innen das dazu erforderliche, methodische Handwerkzeug überhaupt erst einmal kennen und einsetzen können. Ziel der vorliegenden Schrift ist es, eine Einführung in die Anwendung der Mehrebenenanalyse (engl.: "multi level analysis" MLA) zur statistischen Auswertung von Paneldaten bzw. von "repeated measures data" (RMD) zu geben. Ein besonderes Augenmerk wird dabei auf sogenannte "Hybridmodellierungen"¹ gelegt, welche die Vorteile von Fixed-Effect-(Panel-)Modellen² und Random-Effects-Modellen³ miteinander verbinden.

In den folgenden Ausführungen werden, nach ersten einführenden Bemerkungen zu den grundlegenden Logiken von Mehrebenen- und Längsschnittanalysen, die Besonderheiten und Vorteile der Mehrebenenanalyse (MLA) zur Untersuchung von Paneldaten aufgezeigt. Sodann wird im Hauptteil dieser Schrift die schrittweise Durchführung einer MLA zur Analyse von Paneldaten anhand eines Beispiels zur Untersuchung der Effekte von gesundheitsbezogenen Prädiktoren auf die allgemeine Lebenszufriedenheit in der deutschen Bevölkerung vorgestellt.

Zur praktischen Durchführung der Analyse wird das MIXED-Modul des Statistikprogramm Pakets SPSS eingesetzt. In unseren praxisorientierten Erläuterungen wird sowohl auf das grundlegende Datenmanagement, die stufenweise Modellkonstruktion und Modellschätzung sowie auf die damit verbundene Modellinterpretation eingegangen. Die zur Analyse verwendeten Daten stammen aus dem Sozio-Oekonomischen Panel (SOEP), der umfangreichsten und laufend fortgesetzten deutschen Panelstudie.⁴ Zum Abschluss unserer Einführung wird auf die Problematik der Modelldiagnostik eingegangen und es werden diesbezüglich SPSS-spezifische Einschränkungen diskutiert.

¹ Vgl. dazu Allison 2009: 23ff.

² Bei fixed-effect-Modellierungen wird davon ausgegangen, dass es in einer Population für die Stärke eines Effekts von X auf Y einen wahren Wert β gibt, der sich bei sehr vielen Stichprobenschätzungen als häufigster Wert einer Normalverteilung von b ergäbe. Im Beispiel könnte der β -Effekt der Effekt der Einkommenshöhe auf das Armutsrisiko sein, dessen wahrer Wert in der Population der deutschen Bevölkerung identisch mit seinem wahren Wert in der Population der belgischen Bevölkerung wäre. Der spezifische Vorteil von Fixed-Effect-Panel-Modellen liegt in ihrem Umgang mit (zeitlich konstanter) unbeobachteter Heterogenität (vgl. Abschnitt 2.3).

³ Bei random-effects-Modellierungen wird davon ausgegangen, dass es in einer Population für die Stärke eines Effekts von X auf Y nicht nur einen wahren Wert β , sondern viele wahre Werte β_i gibt. Bei sehr vielen Stichprobenschätzungen könnte sich dementsprechend als häufigster Wert von b ein β_1 oder ein β_2 ergeben, die alle zur Verteilung der wahren Werte von β in der Population gehören. Im Beispiel könnte der β -Effekt der Effekt der Einkommenshöhe auf das Armutsrisiko sein, dessen wahrer Wert in der Population der deutschen Bevölkerung ein anderer sein könnte als sein wahrer Wert in der Population der belgischen Bevölkerung.

⁴ Allgemeine Informationen und einen Überblick bzgl. Inhalt, Aufbau, Verlauf und Umgang mit dem Datensatz des Sozio-oekonomischen Panels lassen sich im "Desktop Companion" des SOEP finden (vgl. Haisken-DeNew/Frick 2005): http://www.diw.de/documents/dokumentenarchiv/17/diw_01.c.38951.de/dtc.409713.pdf (zuletzt zugegriffen am 20.05.2013).

Die SPSS-Syntax-Files für alle hier vorgestellten Modellschätzungen werden im Anhang zu diesem SISS-Beitrag abgedruckt und können vom Leser mit wenigen Modifikationen direkt für eigene RMD-MLA in SPSS eingegeben werden.

Zum Verständnis der folgenden Ausführungen werden grundlegende Kenntnisse mehrebenenanalytischer Verfahren vorausgesetzt. Der/die Leser/in sollte nach Lektüre des vorliegenden Textes in der Lage sein, erste Längsschnittanalysen mit einer abhängigen metrischen (kontinuierlichen) Variablen auf der Basis von Mehrebenenmodellierungen und unter Verwendung der im Anhang abgedruckten SPSS-Syntax-Files durchzuführen.

2 Wozu Längsschnittuntersuchungen, wozu Mehrebenenmodellierung?

Während multivariate Querschnittsanalysen zur Bestimmung des Effekts mehrerer unabhängiger Prädiktoren auf eine abhängige Outcome-Variable allein die Kovarianzen zwischen zeitgleich erhobenen Individualmerkmalen auswerten, basieren Längsschnittanalysen auf Daten zur zeitübergreifenden individuellen Veränderung bzw. zum sozialen Wandel von Individualmerkmalen. Betreffen diese Individualmerkmale die Eigenschaften von einzelnen Personen, so sind zur Längsschnittanalyse nahezu zwangsläufig personengebundene Paneldaten erforderlich, welche bestimmte Merkmale desselben Personenkreises über mehrere Messzeitpunkte hinweg abbilden und somit Wandel und Veränderungen auf der Individualebene dokumentieren können.

Klassische Mehrebenenmodelle beschäftigen sich typischerweise mit der Einbettung von Individuen in organisatorische Einheiten. Beispielsweise sind Schüler in Klassen, Klassen in Schulen und Schulen in Schulbezirke eingebettet. Dementsprechend können in der Mehrebenenanalyse alle Schülermerkmale als empirische Einheiten einer ersten Beobachtungsebene (Level 1) definiert werden und die Kontextmerkmale von Schulklassen als empirische Einheiten einer übergeordneten zweiten Ebene (Level 2) behandelt werden.

Bei der häufigsten Kombination von RMD- und ML-Analysen werden typischerweise Personen als diejenigen kontextuellen Einheiten definiert, die auf der zweiten Ebene der Analyse (Level 2) angesiedelt sind (es können aber auch Vereine, Unternehmen oder Organisationen als Level 2-Einheiten behandelt werden). Dementsprechend liegen die Merkmale oder Eigenschaften von Personen (oder anderen L2-Einheiten), die in Längsschnitterhebungen mehrfach im Zeitablauf als Paneldaten gemessen werden, auf der ersten Analyseebene von RMD-MLA-Modellen.⁵

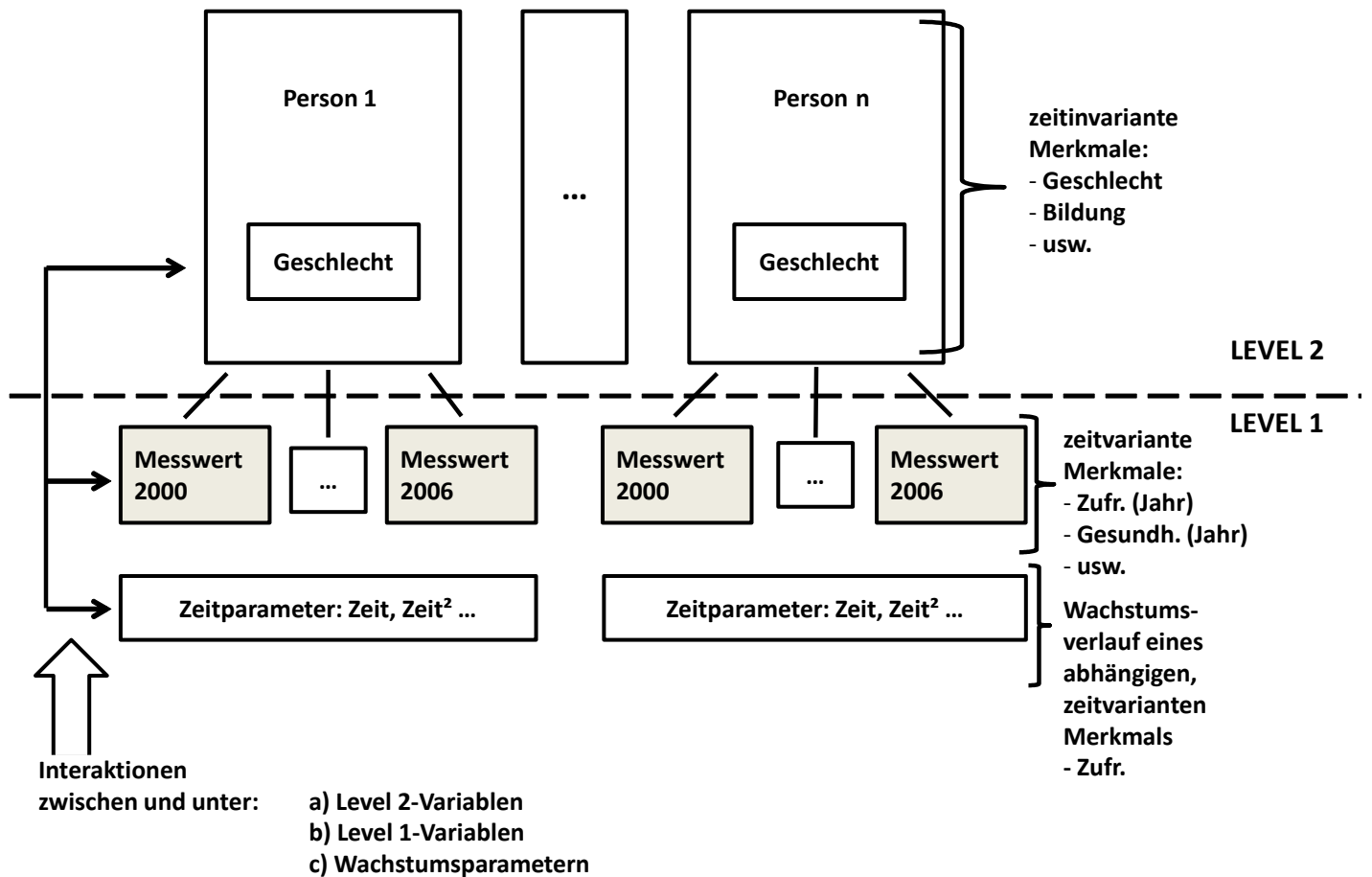
⁵ Prinzipiell besteht in der Mehrebenenanalyse natürlich die Möglichkeit, auch mehr als nur zwei Ebenen zu untersuchen. So können z.B. Personen auch wiederum als Einheiten von Haushalten betrachtet werden, womit sich eine dritte Ebene der Analyse ergäbe. Im vorliegenden Beitrag wird zur Vereinfachung der Darstellung auf die Untersuchung einer solchen dritten Analyseebene verzichtet.

Mithin werden bei dieser Form der Modellierung von Längsschnittdaten die einzelnen, zeitspezifischen Beobachtungswerte als empirische Merkmale betrachtet, die in den Kontext "Person" eingebettet (engl.: nested) sind. Die empirischen Messwerte sind somit als Einheiten der ersten Ebene von MLA-Modellen (Level 1) zu verstehen, wohingegen die zweite Ebene der Modelle (Level 2) die befragten Personen (oder Vereine, Unternehmen usw.) enthält. Auf der ersten Ebene können dann z.B. die Muster "intraindividuelle" Veränderungen (wie z.B. lineares oder kurvilineares Wachstum von Zufriedenheitswerten) über die verschiedenen Erhebungszeitpunkte hinweg beobachtet werden, während auf der zweiten Ebene die Unterschiede dieser Muster d.h. deren Stabilität oder Variation zwischen den untersuchten Personen (oder anderen L2-Einheiten) zu untersuchen sind und somit zusätzlich zu den "intraindividuellen" Veränderungen (s.o.) auch die "interindividuellen" Unterschiede beobachtet werden können.

Die folgende Abbildung 1 verdeutlicht die konstruktive Logik von RMD-MLA-Modellierungen. In der Abbildung werden die Messwerte für zwei Personen (Person 1 und Person 2) auf zwei Analyseebenen (Level 2 und Level 1) dargestellt. Auf Level 2 findet sich als Beispiel für ein (üblicherweise) zeitinvariantes Merkmal die L2-Variable "Geschlecht". Dieses Merkmal wird für jede Person nur ein einziges Mal erhoben. Auf Level 1 befinden sich alle Messwerte für zeitvariante Personenmerkmale (z.B. für das individuelle Ausmaß von Lebenszufriedenheit). Diese zeitvarianten Merkmale werden für jede Person zu jedem Erhebungszeit erneut gemessen (in unserem Beispiel jeweils einmal jährlich in den Jahren 2000, 2001, 2003, 2006). Zusätzlich werden auf Level 2 für jede Person diejenigen Zeitparameter festgelegt, die das Verlaufsmuster des zeitvarianten, abhängigen Merkmals definieren (in unserem Beispiel das Verlaufsmuster der abhängigen Variablen "Lebenszufriedenheit"). Dieses Verlaufsmuster kann beliebige Formen annehmen. Es kann z.B. sowohl strikt linear als auch kurvilinear (etwa in U-Form) sein.

In der Analyse können die partiellen Effekte von L1- und L2-Variablen sowie von Wachstumsparametern auf eine abhängige Variable geschätzt werden. Dabei können diese Effekte als Einzel- oder Interaktionseffekte spezifiziert werden, wobei die Interaktionseffekte durch das Zusammenwirken von Variablen auf nur einer Ebene oder von Variablen auf zwei verschiedenen Ebenen (als "cross level effects") entstehen können.

Abbildung 1: Die konstruktive Logik von RMD-MLA-Modellierungen



Eine Mehrebenenmodellierung von Längsschnittdaten kombiniert die Vorteile von MLA-Modellen mit den Vorteilen der Analyse von RM-Daten. Als wichtigste Vorteile, von denen die meisten auch in den folgenden Beispielsanalysen ausführlich erläutert werden, sind dabei herauszustellen:

1. es können populationsspezifische Verlaufsmuster von Variablenveränderungen getestet werden;
2. es können gruppentypische Variationen von populationsspezifischen Verlaufsmustern ermittelt werden;
3. es können die zeitvarianten Messwerte für jede Person unterschiedlich häufig und mit unterschiedlichen Zeitabständen erhoben werden;
4. es können die Kovarianzen zwischen den wiederholten Messungen verschieden modelliert und getestet werden;
5. es können die Effekte von zeitkonstanten und zeitveränderlichen Variablen geschätzt werden;

6. es können Makro-, Mikro- und kombinierte Makro/Mikro-Effekte geschätzt werden;
7. es können geschachtelte Modelle mit mehreren Makroebenen spezifiziert werden;
8. es können (teilweise) die verzerrenden Effekte unbeobachteter Heterogenität aufgehoben werden;
9. es können Daten analysiert werden, die nicht unabhängig voneinander sind, und es können die Standardfehler der diesbezüglichen Modellschätzungen unverzerrt ermittelt werden.

Die genannten Vorteile bedeuten u.a., dass auch solche Personen in die Analyse einbezogen werden können, für die nicht zu jedem Messzeitpunkt und für jede Individualvariable gültige Messwerte vorliegen. Die Mehrebenenmodellierung benötigt folglich keinen balancierten Datensatz (vgl. Hox 2010: 79), in welchem zu jedem Messzeitpunkt gleich viele Beobachtungswerte für alle Befragten gegeben sind. Deshalb können auch Personen in die Analyse aufgenommen werden, welche (a) aus dem Panel vorzeitig ausgeschieden sind, (b) erst zu einem späteren Zeitpunkt aufgenommen wurden, (c) zu bestimmten Erhebungszeitpunkten zwischen den beobachteten Wellen nicht befragt werden konnten (temporary dropouts) oder (d) zu einem Messzeitpunkt fehlende Werte aufweisen. Es konnte sogar empirisch gezeigt werden, dass eine MLA mit unvollständigen Datensätzen bessere Ergebnisse liefern kann als eine MLA, bei der die fehlenden Werte durch Anwendung von Imputationsmethoden geschätzt werden (vgl. Twisk 2011: 107).

Auch sei noch einmal darauf hingewiesen, dass bei der längsschnittlichen Mehrebenenmodellierung die Zeit-Intervalle zwischen den verschiedenen Messzeitpunkten unterschiedlich lang sein können (wie dies auch im hier analysierten Untersuchungsbeispiel der Fall sein wird) und die unterschiedlich großen Zeitabstände sogar zwischen den einzelnen Befragten variieren können (vgl. Tabachnick/Fidell 2010: 822f). Mithin kann bei einer RMD-MLA jeder Befragte einen eigenen Erhebungsschedule aufweisen, solange die verschiedenen Erhebungszeitpunkte einer exakt definierten Zeitmetrik unterliegen.

Ein weiterer wichtiger Vorteil von MLA, die zur Längsschnittanalyse eingesetzt werden, ist die flexible und einfache Modellierung von zeitvarianten und zeitinvarianten Variableneffekten (vgl. Hox 2010: 98f). So können sowohl die Effekte von zeitlich veränderlichen Determinanten (z.B. im folgenden Anwendungsbeispiel vom jeweiligen Gesundheitszustand) als auch die Einflüsse von zeitlich konstanten Merkmalen (z.B. vom Geschlecht) sowie die Interaktion zwischen beiden Determinanten (hier: Gesundheitszustand und Geschlecht) in der Analyse untersucht werden. In diesem Falle sind zeitinvariante Merkmale, wie beispielsweise das Geburtsjahr oder das Geburtsland, als Variablen

der zweiten Modellebene zu verstehen. Demgegenüber sind die empirischen Ausprägungen von zeitveränderlichen Variablen wie Einstellungen, Einkommen oder die jeweilige, individuelle Stimmungslage als Messgrößen auf der ersten Modellebene zu betrachten.⁶

2.1 Zentrierungsoptionen

Eine Mehrebenenanalyse von RM-Daten sollte mit zentrierten Beobachtungsdaten durchgeführt werden. Dadurch eröffnen sich verschiedene Vorteile: Die Ergebnisse von Modellschätzungen mit zentrierten Variablen sind leichter zu interpretieren, die Schätzergebnisse werden nicht durch Multikollinearitäten beeinflusst und unerwünschte Korrelationen zwischen Level 1- und Level 2-Variablen werden weitestgehend vermieden.

Zur Zentrierung des Zeitparameters, der in der Längsschnittanalyse die zeitlichen Abstände zwischen den Einzelbeobachtungen repräsentiert, und zur Zentrierung der unabhängigen Prädiktorvariablen gibt es verschiedene Möglichkeiten. Welche davon ausgewählt werden sollte, hängt in hohem Maße von der zugrundeliegenden Forschungsfrage ab. Wir wollen an dieser Stelle allein drei gängige Zentrierungsverfahren vorstellen:

(1) Die Zentrierung um den Nullpunkt (oder um andere, empirisch "sinnvolle" Werte). Dieses Verfahren ist insbesondere für die Zentrierung des Zeitparameters von Bedeutung. Der Parameter mit dem 0-Wert sollte dabei so gewählt werden, dass der in der Modellschätzung ermittelte Wert für das Regressionsintercept (bzw. für die Regressionskonstante) sinnhaft interpretiert werden kann. So sollte der Wert "Null" in der Regel den Beginn einer Panelerhebung (1.Welle) repräsentieren.⁷ Des Weiteren sollte der Zentrierungswert innerhalb des tatsächlich beobachteten Zeit-/Werte-Intervalls und nicht vor Beginn oder nach Abschluss einer Messreihe liegen.

(2) Die Zentrierung um den gruppenübergreifenden Gesamtmittelwert (engl.: grand mean). Dieses Verfahren ist insbesondere zur Analyse von zeitinvarianten Merkmalen von Bedeutung. Bei der Zent-

⁶ In der weiteren Analyse werden wir dann sehen, dass diese Variablen nicht nur auf der ersten sondern auch auf der zweiten Modellebene zu verorten sind, da zeitlich veränderliche Variablen sowohl Variation innerhalb der Befragten (Within-Varianz) als auch zwischen den Befragten (Between-Varianz) transportieren.

⁷ Bei speziellen Fragestellungen kann es allerdings auch sinnvoll sein, um andere Werte zu zentrieren. Dies kann z.B. auch der letzte Erhebungszeitpunkt sein. So kann etwa ein durchschnittlicher Intercept-Wert der abhängigen Variablen "Vermittlungschance auf dem Arbeitsmarkt" interessieren nachdem eine arbeitsmarktbezogene Schulungsmaßnahme durchgeführt wurde. Dann lässt sich mit solch einer Modellierung danach fragen, welches durchschnittliche Niveau die Vermittlungschance nach der schulischen Intervention zum Ende eines längeren Beobachtungszeitraums erreicht und lässt sich auch klären, ob sich dieses Niveau signifikant zwischen den beobachteten Personen unterscheidet.

rierung um den "grand mean" wird um den Gesamtmittelwert einer Variablen in der gesamten Stichprobe zentriert. Modelle mit grand-mean-zentrierten Variablen sind im modelltechnischen Sinne als äquivalent zu Modellschätzungen mit Rohvariablen zu behandeln. Sie unterscheiden sich weder im Fit, in den Residuen noch in den vorausgesagten Werten voneinander und die geschätzten Parameter lassen sich ineinander überführen (vgl. Kreft/De Leeuw 1998: 109). Der dabei geschätzte Interceptwert ist folglich als Voraussagewert für Personen mit durchschnittlicher Merkmalsausprägung (bzw. für Personen mit Prädiktorwerten von 0) zu interpretieren und liegt somit auf jeden Fall nicht außerhalb des Intervalls der beobachteten Messwerte.

(3) Die Zentrierung um gruppenspezifische Mittelwerte (engl.: group means). Bei diesem Verfahren werden Level 1-Variablen um den Mittelwert einer jeden Person über alle Erhebungszeitpunkte hinweg zentriert, da in der Analyse die einzelnen Personen als "Gruppe" bzw. "Kontext" bzw. "L2-Einheit" betrachtet werden. Group-mean-zentrierte Modellschätzungen sind allerdings (in aller Regel)⁸ im Gegensatz zu grand-mean-zentrierten Modellschätzungen nicht mehr äquivalent zu Modellen auf der Grundlage von Rohdaten (s.o.).

Alle drei Zentrierungsoptionen werden im folgenden Anwendungsbeispiel verwendet, wobei für die Wahl einer bestimmten Zentrierungsmethode jeweils unterschiedliche Gründe eine Rolle spielen: Generell wird durch die Zentrierung eine bessere Interpretierbarkeit der Ergebnisse von MLA-Schätzungen erreicht, denn die Schätzwerte für partielle Effekte sind immer als Veränderungen der abhängigen Variablen unter der Voraussetzung, dass alle anderen Prädiktoren das Ausmaß ihrer Mittelwerte aufweisen, zu verstehen (und nicht, wie bei Variablen mit Rohwerten, einen Wert von 0,00 annehmen). Zudem wird dadurch den unerwünschten Effekten der nicht-essentiellen Multikollinearität entgegengewirkt (vgl. dazu Urban/Mayerl 2011: 225ff). So sind beispielsweise group-mean-zentrierte Variablen mit allen Level 2-Variablen unkorreliert (vgl. Kreft/Leeuw 1998: 136). Auch die Korrelationen zwischen Level 2-Variablen und Cross-Level-Interaktionen sind null und somit kein Problem (vgl. ebd.). Einzig die Korrelationen unter Level 1-Variablen und von L1-Variablen mit Cross-Level-Interaktionen stellen ein potentiell Problem bei group-mean-zentrierten Modellen dar (vgl. ebd.). Sind diese Korrelationen von beträchtlichem Ausmaß, so sollten mithin die Variablen grand-mean-zentriert werden.

⁸ Ausgenommen hiervon sind Modelle mit random Intercept und fixed Slopes, in welchen die Gruppen-/Personen-Mittelwerte auf der zweiten Ebene wiedereingeführt werden, diese sind äquivalent zu Modelle mit Rohdaten (vgl. Keft/De Leeuw 1998: 109). Gleiches gilt bei Datensätzen mit dem unwahrscheinlichen Fall, dass die einzelnen Gruppenmittelwerte identisch sind (vgl. ebd.).

2.2 Die Varianz/Kovarianzmatrix

Wie die traditionelle Regressionsanalyse, setzt auch eine erfolgreiche RMD-MLA die Gültigkeit einiger Annahmen über die Verteilung der Schätz-Residuen voraus (vgl. dazu Urban/Mayerl 2011: 177ff).⁹ Bei der RMD-MLA betreffen diese Annahmen insbesondere die Abwesenheit von Residuen-Heteroskedastizität und Residuen-Autokorrelation. Allerdings kann die Gültigkeit dieser Annahmen in der RMD-MLA recht einfach getestet und kontrolliert werden (zumindest für einen bestimmten Typ von Residuen, s.u.). Dazu kann per Vorentscheid die Struktur der Residuen-Kovarianzmatrix zu Beginn der Modellanalyse festgelegt werden. Einige der vielen Möglichkeiten zur Festlegung von verschiedenen Varianz/Kovarianzmatrix-Typen werden wir im Folgenden vorstellen.

Charakteristisch für eine Mehrebenenanalyse ist die Zerlegung des Fehlers in zwei Teile: in den Teil, der für jeden einzelnen Messwert in besonderer Weise entsteht, und den Teil, der für alle Messwerte einer jeden Kontexteinheit (hier: einzelne Personen) entsteht. Da diese Unterscheidung für die folgenden Ausführungen sehr bedeutsam ist, soll sie anhand von Abbildung 2 noch einmal grafisch veranschaulicht werden.

Abbildung 2 zeigt als Datenpunkte im Y/T-Koordinatensystem die Lage von jeweils vier Y-Werten, die für zwei verschiedene Personen (Person 1, Person 2) zu vier aufeinander folgenden Messzeitpunkten (t_1, t_2, t_3, t_4) erhoben wurden. Betrachten wir nun allein die beiden Datenpunkte $Y_{3,1}$ und $Y_{4,2}$. Der Datenpunkt $Y_{3,1}$ markiert die Lage für denjenigen Y-Wert, der zum Zeitpunkt t_3 für die Person 1 gemessen wurde, während der Datenpunkt $Y_{4,2}$ die Lage für denjenigen Y-Wert markiert, der zum Zeitpunkt t_4 für die Person 2 gemessen wurde. Diese beiden Y-Messwerte (und alle anderen Y-Messwerte auch) weisen einen bestimmten Abstand zum Mittelwert aller Y-Werte " \bar{Y} " auf. Dieser Mittelwert aller Y-Werte ist der beste Schätzwert für jeden Y-Wert, wenn weder der Zeitpunkt noch die Person bekannt ist, für den/die ein bestimmter Wert von Y geschätzt werden soll.

Natürlich wird bei solch einer Schätzung auch immer ein Schätzfehler gemacht, der für unsere beiden Datenpunkte vom Ausmaß der Länge der beiden fett eingezeichneten Linien ist: Im Falle einer Schätzung von $Y_{3,1}$ durch \bar{Y} wird der wahre Y-Wert erheblich unterschätzt, und im Falle einer Schätzung von $Y_{4,2}$ durch \bar{Y} wird der wahre Y-Wert erheblich überschätzt. Werden die Abstände aller beobachteten Y-Werte vom Y-Gesamtmittelwert \bar{Y} zu allen Zeiten und von allen Personen addiert, so lässt sich daraus die Gesamtvarianz von Y berechnen.

Man kann aber auch den Abstand eines jeden Datenpunkts zum personenspezifischen Y-Mittelwert \bar{Y}_1 oder \bar{Y}_2 betrachten. Dann schaut man in Abbildung 2 auf die gepunkteten Linien und es ist zu

⁹ Zur Erinnerung: Als Residuen werden hier die Differenzen zwischen den mit der Modellschätzung prognostizierten, zeitabhängigen Werten der abhängigen Y-Variablen (inkl. der Prognose ihrer Verlaufsform, s.o.) und den tatsächlich beobachteten, empirischen Y-Werten verstanden. Die Residuen werden oftmals auch als Fehler oder Fehlerterm (engl.: error / error term) bezeichnet.

erkennen, dass diese kürzer sind als die fett eingezeichneten Linien. Die Fehler der Schätzung werden somit kleiner als sie es bei einer Schätzung mit \bar{Y} sind. Sie können auch als zeitspezifische Residuen einer bestimmten Person $e_{t,i}$ bezeichnet werden (t =spezifischer Zeitpunkt, i =spezifische Person). In der MLA werden diese Residuen auch L1-Residuen genannt. Aus der Summe aller L1-Residuen (bzw. aller gepunkteten Linien über alle Personen) lässt sich die Y-Varianz "innerhalb" aller Personen bzw. die sogenannte "Varianz(within)" von Y berechnen.

In Abbildung 2 sind aber auch noch gestrichelte Linien zu erkennen. Sie markieren die Abstände zwischen den personenspezifischen Y-Mittelwerten \bar{Y}_i und dem Gesamtmittelwert \bar{Y} und variieren nicht mit dem Erhebungszeitpunkt. Sie können als personenspezifische Residuen u_i (i =Person) bezeichnet werden. Diese Residuen werden in der MLA auch als L2-Residuen bezeichnet. Aus der Summe aller dieser Residuen (bzw. aller gestrichelten Linien über alle Personen) lässt sich die Y-Varianz "zwischen" den Personen bzw. die sogenannte "Varianz(between)" von Y berechnen.

Die folgende Abbildung 2 zeigt mithin, dass sich jeder Y-Datenpunkt als Summe berechnen lässt. Die Berechnung dieser Summe verdeutlicht auch noch einmal die folgende Gleichung (1):

$$Y_{t,i} = \text{Schätzwert } (\bar{Y}) + \text{zeitspezifisches personales Residuum } (e_{t,i}) + \text{personales Residuum } (u_i) \quad (1)$$

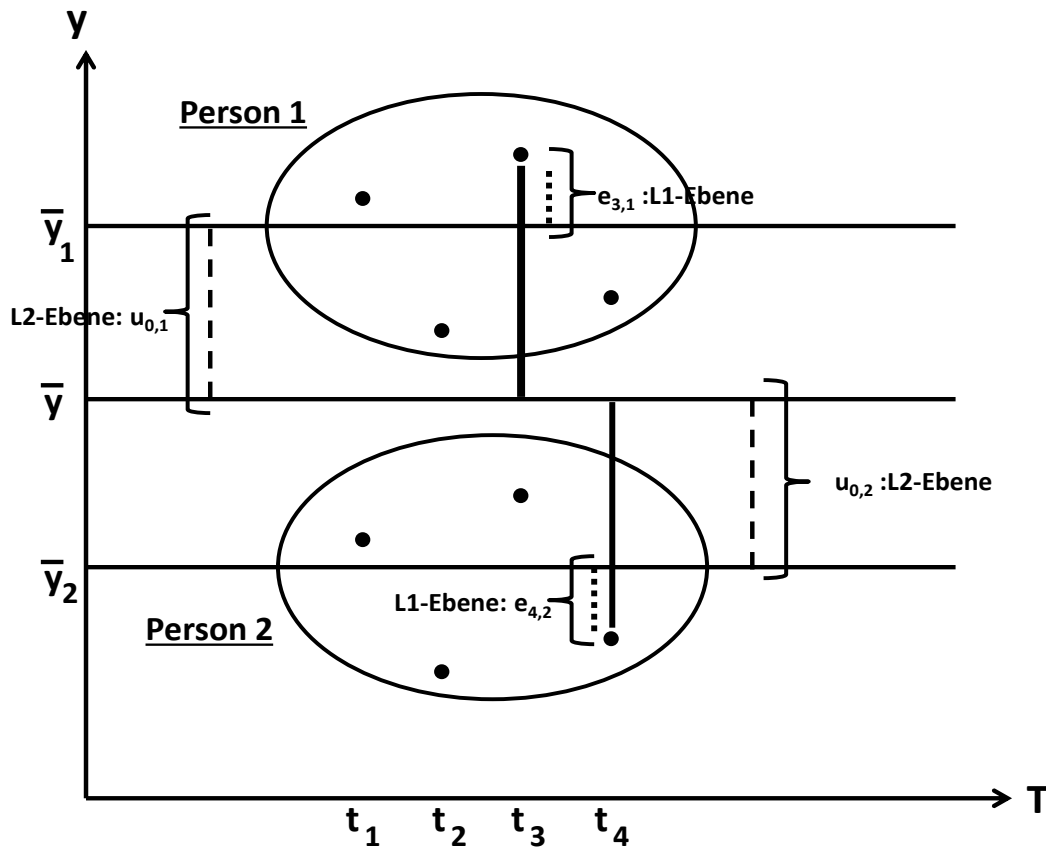
Abbildung 2 verweist auch darauf, dass sich bei einer MLA mit zwei Ebenen (L1 und L2) die Y-Gesamtvarianz in zwei Teilvarianzen zerlegen lässt:

$$\text{Gesamtvarianz (fette Linien)} = \text{Varianz, within (gepunktete Linien)} + \text{Varianz, between (gestrichelte Linien)} \quad (2)$$

Natürlich lässt sich $Y_{j,i}$ nicht nur mit dem Y-Gesamtmittelwert (\bar{Y}) schätzen (wie in Gleichung 1 gezeigt), sondern auch mittels einer linearen Regressionsschätzung mit einem oder mehreren Prädiktoren. Wird dabei der Prädiktor "Zeit" benutzt entsteht aus Gleichung (1) die folgende Gleichung (3):

$$Y_{t,i} = b_0 + b_1 (\text{Zeit } t_i) + e_{t,i} + u_{0,i} \quad (3)$$

Gleichung (3) bezeichnet ein Mehrebenen-Regressionsmodell mit einem "random intercept"-Wert. Die Variation des Intercept-Wertes ergibt sich daraus, dass bei Prognose von $Y_{t,i}$ der personenspezifische u -Wert zum Intercept-Wert b_0 hinzuaddiert werden muss.

Abbildung 2: Die Logik der Residuen- und Varianzzerlegung in Mehrebenenmodellen

Kommen wir nunmehr zurück zu unserem eigentlichen Thema in diesem Abschnitt: "die Definition von Varianz/Kovarianz-Matrizen" in der RMD-MLA:

Wie bereits oben erwähnt, kann in jeder RMD-MLA eine Varianz/Kovarianz-Matrix für die L1-Residuen ($e_{t,i}$) und die L2-Residuen (u_i) festgelegt werden. Die folgende Abbildung 3 zeigt eine L1-Matrix für eine Analyse von RM-Daten über vier Messzeitpunkte. Die einzelnen Matrix-Elemente werden darin als " σ " bezeichnet. In der Diagonalen der Matrix befinden sich die Varianzen der L1-Residuen für jeden Messzeitpunkt. Unterhalb der Diagonalen befinden sich die Kovarianzen zwischen den Residuen von jeweils zwei Messzeitpunkten. Da es sich um eine quadratische Matrix handelt, sind unterhalb und oberhalb der Diagonalen identische Kovarianzen zu finden. In der Analyse reicht es also aus, allein die Elemente der unteren Dreiecksmatrix zu betrachten.

Abbildung 3: Unstrukturierte Kovarianzmatrix für 4 Messzeitpunkte (mit 10 Parametern)

$$\begin{array}{cccc}
 \sigma_{11} & \sigma & \sigma & \sigma \\
 \sigma_{12} & \sigma_{22} & \sigma & \sigma \\
 \sigma_{13} & \sigma_{23} & \sigma_{33} & \sigma \\
 \sigma_{14} & \sigma_{24} & \sigma_{34} & \sigma_{44}
 \end{array}$$

Wenn in der SPSS-Implementation der RMD-MLA vom Anwender keine spezielle L1-Matrix definiert wird, geht das EDV-Programm von einer L1-Matrix des Typs "scaled identity" aus (SPSS-Abk.: ID).¹⁰ Eine solche Matrix legt fest, dass die Residuen-Varianzen homogen sind und nicht miteinander korrelieren. Somit wird auf diese Weise angenommen, dass die Fehlerterme zu den einzelnen Messzeitpunkten unabhängig voneinander sind.

Eine spezielle Form dieses Matrix-Typs ist eine Matrix vom Typ "compound symmetry" (SPSS-Abk.: CS), nach der auch alle Varianzen gleich sind, nach der aber alle Residuen miteinander kovariieren und alle diese Kovarianzen vom gleichen Ausmaß sind, also einen konstanten Wert aufweisen. Sicherlich werden die Annahmen der Compound Symmetry-Matrix in der Forschungspraxis nur selten erfüllt sein. Auch wenn eine konstante Varianz zu allen Messzeitpunkten noch plausibel sein mag, trifft dies auf eine identische Kovarianz aller Messungen wohl eher nicht zu. So kann in der Regel angenommen werden, dass die Werte von Messzeitpunkten, welche zeitlich näher beieinander liegen, auch stärker miteinander kovariieren als Messwerte, welche zeitlich weiter voneinander entfernt sind. Allerdings gehen Random-Intercept-Modelle implizit vom Muster der "compound symmetry" aus.

Der Gegenpol zum Muster der "compound symmetry" ist das Muster der "unstrukturierten" Kovarianzmatrix (SPSS-Abk.: UN). Bei diesem Matrix-Typ erfolgt keine Festlegung hinsichtlich einer Varianz- und/oder Kovarianzhomogenität. Alle Elemente der Matrix können frei geschätzt werden. Während also die Compound Symmetry-Matrix noch mit der Schätzung von zwei Parametern auskommt (unabhängig davon, wie viele Messzeitpunkte vorhanden sind), müssen bei der unstrukturierten Kovarianzmatrix insgesamt " $p(p+1)/2$ " (mit p =Anzahl der Messzeitpunkte) Parameter geschätzt werden. Die vorausgegangene Abbildung 3 zeigt eine solche Matrix für Werte von vier Messzeitpunkten.

In der Forschungspraxis dominiert die Einschätzung, dass bei Analysen mit vielen Messzeitpunkten (z.B. mit mehr als 5 Zeitpunkten) die Schätzung der unstrukturierten Kovarianzmatrix zu komplex wird und sie deshalb eher bei sparsameren Modellvarianten benutzt werden sollte (vgl. Heck et al. 2010: 164).

¹⁰ Dies gilt allerdings nur bei einer Steuerung der SPSS MIXED-Prozedur über Syntax-Dateien (so wie sie hier im Anhang abgedruckt sind). Wird die Menü-Steuerung von SPSS MIXED benutzt, lautet die Voreinstellung "diagonale Matrix" (SPSS-Abk.: DIAG) (mehr dazu im folgenden Haupttext). Wenn also in der Analyse eine "scaled identity"-Matrix bevorzugt wird, muss in der Syntax-Steuerung überhaupt kein Matrix-Typ angegeben werden, während in der Menü-Steuerung explizit eine ID-Matrix aufgerufen werden muss.

Aber auch bei Datenstrukturen, welche viele personenspezifische Messzeitpunkte und viele fehlende Werte aufweisen, stößt die unstrukturierte Kovarianzmatrix schnell an ihre Grenzen.¹¹ Des Weiteren erlaubt sie keine Varianz in Abhängigkeit von der Variation der "random slopes" der zeitvarianten Parameter¹² (vgl. Raudenbush/Bryk 2010: 197ff).

Alternativ zur "radikalen" Matrixstruktur des komplett unstrukturierten Matrix-Typs sollen hier noch zwei andere Matrix-Typen erwähnt werden:

Die Diagonal-Matrix (SPSS-Abk.: DIAG) ermöglicht eine freie Schätzung der Varianzen aller L1-Residuen, erlaubt also Heteroskedastizität, fixiert aber alle Kovarianzen auf 0.00, womit Korrelationen zwischen den Residuen ausgeschlossen bzw. nicht zugelassen werden. Ob der Ausschluss von Residuen-Korrelationen empirisch sinnvoll ist, kann durch Vergleich des Schätzerfolgs von Modellen mit unterschiedlichen Matrix-Typen ermittelt werden.

Eine (empirisch betrachtet) eher sinnvolle Spezifikation des L1-Matrix-Typs ergänzt die Schätzung heterogener Varianzen um die Schätzung autoregressiver Kovarianzen erster Ordnung (SPSS-Abk.: ARH1). Dabei wird angenommen, dass zeitlich benachbarte Residuen mit einer fixen Stärke (r) untereinander korrelieren und dass zeitlich weiter auseinanderliegende Residuen mit einer deutlich reduzierten, aber auch konstanten Korrelation (r^2) kovariieren. Somit müssen für diesen Matrix-Typ nur $p+1$ Parameter (p =Anzahl der Messzeitpunkte) geschätzt werden.

Analog zur Auswahl eines Matrix-Typs für Level 1-Residuen kann auch ein Matrix-Typ für Level 2-Residuen festgelegt werden. Die Elemente der L2-Matrix repräsentieren die Varianzen und Kovarianzen der Random-Effekte einer MLA-Spezifikation. Wird z.B. im Modell ein "random intercept" und nur der "random slope" eines einzigen Prädiktors (und deren Kovarianz) geschätzt, ergibt sich eine 2×2 -Matrix der L2-Residuen. Die folgende Abbildung 4 zeigt eine solche Matrix. Wiederum markieren die Matrix-Elemente auf der Diagonalen die Varianzen (hier: von Intercept und Slope), während die Matrix-Elemente unterhalb der Diagonalen die entsprechenden Kovarianzen bezeichnen (zur Erinnerung: auch hier entsprechen die Elemente unterhalb der Diagonalen den Elementen oberhalb der Diagonalen, so dass nur die Werte unterhalb der Diagonalen geschätzt und interpretiert werden müssen).

Zur Schätzung der L2-Matrix empfehlen wir, den Matrix-Typ einer unstrukturierten Matrix (SPSS-Abk.: UN) auszuwählen. Ohne diesbezügliche Angabe wählt die SPSS MIXED-Prozedur als

¹¹ Personenspezifische bzw. individuelle Messzeitpunkte würden die unstrukturierte Kovarianzmatrix um eine Vielzahl zusätzlicher Spalten/Zeilen erweitern, während gleichzeitig die dazugehörigen Zellenbesetzungen äußerst gering wären. Eine erfolgreiche (konvergierende) Modellschätzung wäre somit nicht nur aufgrund der zusätzlichen Parameter, sondern auch aufgrund der Unbalanciertheit der Daten äußerst unwahrscheinlich.

¹² Die Modellierung der Kovarianz in Abhängigkeit von der Variation von Random-Slopes (auch zusätzlich zu Zeitparametern) ist im Rahmen von Mehrebenenmodellen allerdings möglich.

Voreinstellung eine Matrix vom Typ "variance components" aus (SPSS-Abk.: VC), womit die Kovarianz der Random-Effekts auf null fixiert wird.

Abbildung 4: Unstrukturierte Kovarianzmatrix für zwei Random-Effekte

$$\begin{array}{cc} \sigma_I & \sigma \\ \sigma_{I,S} & \sigma_S \end{array}$$

Auch der Zeitparameter in RMD-MLA-Modellen (vgl. Gl. 3, weiter oben) kann als Random-Effekt spezifiziert werden. Dadurch lässt sich ein praktikabler Kompromiss zwischen einer Schätzung unter den restriktiven Annahmen einer Compound-Symmetry-Matrix und einer Schätzung mit der hohen (und oftmals zu hohen) Komplexität einer unstrukturierten Matrix erreichen. Dabei wird eine Kovarianzmatrix generiert, deren Varianzen und Kovarianzen in Abhängigkeit von der zwischen den Erhebungszeitpunkten abgelaufenen Zeit variieren kann, ohne hierfür jeweils einen eigenen Parameter schätzen zu müssen. Wie dies praktisch umzusetzen ist, wird in den folgenden Beispielsanalysen gezeigt werden.

Für Experten: SPSS operiert bei der RMD-MLA-Schätzung mit einer integrierten Kovarianzmatrix V . Eine direkte Spezifikation und Ausgabe der Kovarianzmatrix V ist in der SPSS MIXED-Prozedur nicht möglich. Die Kovarianzmatrix V generiert sich allerdings aus der Kombination der Kovarianzmatrix der L2-Residuen (G) und der Kovarianzmatrix der L1-Residuen (R):

$$V = Z \times G \times Z^T + R$$

(vgl. Hoffman 2013: 8ff),¹³ deren Spezifikation und Ausgabe in SPSS MIXED möglich ist.¹⁴ Durch diese Kombination der Matrizen ist es möglich, Kovarianzmatrizen zu konstruieren, deren Elemente in Abhängigkeit von Zeitparametern stärker oder schwächer variieren bzw. kovariieren, indem ein/mehrere "random slope/s" für Zeitparameter spezifiziert wird/werden. Diese Spezifikationsoptionen sind insbesondere dann von Bedeutung, wenn die einzelnen Messzeitpunkte zeitlich unterschiedlich weit voneinander entfernt sind. So gilt auch bei benachbarten Messzeitpunkten, dass eine größere zeitliche Entfernung in der Regel mit geringeren Kovarianzwerten einhergehen wird. Durch Zeitparameter kann dies berücksichtigt werden.

Allerdings kann angemerkt werden, dass Fehlspezifikationen im Random-Teil eines Modells, den Fixed-Teil des Modells wenig beeinflussen (vgl. Heck et al. 2010:164). Somit sind Analysen, deren Fragestellungen vorwiegend den Fixed-Teil des Modells betreffen, nur gering durch Fehlspezifikationen im Random-Teil gefährdet.

¹³ Z beschreibt eine Matrix der Größe (Anzahl-Messzeitpunkte*Anzahl Random-Effekts), welche die Werte zu den jeweiligen Messzeitpunkten enthält. Z^T stellt lediglich die transponierte Matrix dar.

¹⁴ Zur Ausgabe beider Matrizen wird in der MIXED-Prozedur der Unterbefehl "/PRINT= G R" benötigt.

2.3 Hybridmodelle bzw. Besonderheiten bei zeitvarianten Variablen

Sogenannte Hybridmodelle vereinigen die Vorteile von Fixed-Effect- und Random-Effects-Modellen. Mit diesen Modellen kann sowohl der Einfluss von zeitinvarianten Merkmalen (wie z.B. von Geschlecht oder Schulbildung) als auch der Effekt von zeitvarianten Prädiktoren (wie z.B. Alter oder Einkommen) in die Modellspezifikation aufgenommen und geschätzt werden.

Um den reinen Effekt der Veränderung von zeitvarianten Determinanten bestimmen zu können, müssen die entsprechenden (zeitveränderlichen) Variablen adäquat umkodiert werden. Häufig muss aber auch die Spezifikation des Modells auf die Analyse von zeitvarianten Effekten ausgerichtet werden. Dafür gibt es zwei unterschiedliche Strategien:

- 1) die gemeinsame Spezifikation der zeitvarianten Prädiktoren auf der ersten Modellebene unter gleichzeitiger Einbeziehung der personenspezifischen Prädiktoren-Mittelwerte auf der zweiten Modellebene;
- 2) die Zerlegung der zeitvarianten Prädiktoren in ihre jeweiligen Within-Personen- und Between-Personen-Anteile und deren separate Spezifikation im Modell.

Da die erste Option zu Verzerrungen in den Schätzergebnissen führen kann,¹⁵ wird in den folgenden Beispielen zur Analyse der Determinanten von Lebenszufriedenheit nur die zweite Option benutzt. Bevor jedoch die technische Umsetzung dieser Option erläutert wird, soll hier zunächst verdeutlicht werden, warum in aller Regel eine Modellspezifikation unter einfacher Berücksichtigung von zeitvarianten Prädiktoren für eine RMD-MLA nicht ausreichend ist.

Wenn im RMD-MLA-Modell zeitveränderliche Determinanten als reine Einzelvariablen spezifiziert werden, welche zu jeder Panelwelle einen anderen Variablenwert aufweisen können, ergibt sich der Schätzwert der dazugehörigen Effektkoeffizienten aus einer Mischung von zwei verschiedenen Typen von Kovarianzen, ohne dass die entsprechenden Anteile voneinander unterscheidbar sind. Denn zeitveränderliche Determinanten setzen sich im Rahmen von Längsschnittanalysen zumeist aus zwei verschiedenen Formen von Variation zusammen, so dass sie auch als zwei Variablen verstanden werden müssen (vgl. Hoffman/Stawski 2009: 105). Bei diesen zwei Formen von Variablenvariation handelt es sich zum einen um die (zeitliche) Veränderung der Variablenausprägung bei jedem einzelnen Befragten (within-cluster) und zum anderen um die durchschnittliche Variablenausprägung bei jedem Befragten (über die Gesamtzeit) und dessen Variation (dieser durchschnittlichen Variablenausprägung) zwischen den befragten Personen (between-cluster) (vgl. zu dieser Unterscheidung die vorangegangene Abbildung 1 und die diesbezüglichen Erläuterungen).

¹⁵ Entsprechende Verzerrungen treten dann auf, wenn der Effekt der zeitveränderlichen Determinante auf der zweiten Ebene so spezifiziert wird, dass er über die Befragten variieren kann (bzw. als "random" gesetzt wird) (vgl. Hoffman 2009: 117).

Die Unterscheidung von zwei Formen von Variablenvariation ist wichtig, weil nicht davon ausgegangen werden kann, dass die geschätzten Koeffizienten für die Within- und Between-Anteile gleich stark ausfallen. Eher werden die diesbezüglichen Effekte unterschiedlich stark ausgeprägt sein. Werden diese Unterschiede in der Analyse nicht berücksichtigt, ergeben sich verzerrte Schätzwerte als eine Folge unbeobachteter Heterogenität.¹⁶

Denn Verzerrungen der geschätzten Koeffizienten durch unbeobachtete (individuelle) Heterogenität ergeben sich immer dann, wenn Variablen, welche nicht in die Modellschätzung einbezogen werden konnten, mit den anderen erklärenden Variablen und der zu erklärenden Variablen in Zusammenhang stehen (was oft der Fall sein dürfte). In der Analyse kann jedoch durch Verwendung der separierten Within-Werte von zeitvarianten Prädiktoren die Abhängigkeit der dazugehörigen "Teilvariable" von allen zeitlich konstanten Variablen kontrolliert werden,¹⁷ weil durch die Separation des Within-Anteils für die entsprechende Prädiktor-Teilvariable kein Zusammenhang mehr mit zeitlich stabilen Größen besteht, wodurch auch die diesbezügliche Effektschätzung nicht mehr verzerrt sein kann.

Der Vorteil dieses Verfahrens wird allerdings mit einer damit verbundenen, geringeren statistischen Teststärke (power) erkaufte, da die bei der Effektschätzung berücksichtigte Fallzahl durch die alleinige Betrachtung der Within-Varianz reduziert wird. So fallen beispielsweise alle Befragten aus der Analyse heraus, welche im Zeitverlauf der Untersuchung keine Veränderungen bei den zeitvarianten Prädiktoren aufweisen. Deshalb werden auch die Effektschätzungen, die sich auf die Within-Variation beziehen, mit dem Akronym ATET (Average Treatment Effect on the Treated) bezeichnet, da eben nicht alle Personen zur Schätzung des Within-Effekts beitragen, sondern nur solche, welche tatsächlich eine Veränderung (Treatment) während des Untersuchungszeitraums aufweisen (vgl. Brüderl 2010: 971).^{18,19}

¹⁶ Für eine umfangreiche Darstellung des Problems der unbeobachteten Heterogenität und dessen Bearbeitung siehe Brüderl 2010.

¹⁷ Allerdings gilt trotz automatischer Kontrolle der zeitkonstanten unbeobachteten/individuellen Heterogenität, dass die geschätzten Koeffizienten der Within-Variablen immer noch durch zeitveränderliche Einflussgrößen, nach welchen nicht kontrolliert wird, verzerrt werden können.

¹⁸ Diese Einschränkung reduziert in den folgenden Beispielsanalysen die Anzahl der Einzelbeobachtungen, die zur Schätzung der L1-Effekte benutzt werden können, von 83.313 auf 56.426 (68%) Beobachtungswerte.

¹⁹ Probleme bzgl. der Generalisierbarkeit der Ergebnisse ergeben sich hierbei nur, wenn angenommen wird, dass die Gruppe, welche Veränderungen aufweist, sich in einer (relevanten) systematischen Weise von den Personen ohne Veränderung unterscheidet (vgl. Brüderl 2010: 971).

Fassen wir also noch einmal zusammen: Da ein separat geschätzter Within-Personeneffekt um die Einflüsse aller zeitlich konstanten Merkmale bereinigt ist, ermöglicht dieser eine genauere Annäherung der Schätzung an den "wahren" Effekt in der effektrelevanten Population. Um diesen Vorteil in der RMD-MLA nutzen zu können, sollten die Within- und Between-Anteile eines jeden zeitvarianten Prädiktors nicht vermengt, sondern getrennt werden. Was bedeutet das für die Praxis der Mehrebenenanalyse?

Um die Within- und Between-Komponenten zu separieren, müssen die zeitveränderlichen Determinanten in zwei (Teil-)Variablen zerlegt werden. Dabei wird der Between-Anteil der Determinante (gemessen als seine pro Person beobachtete, zeitlich durchschnittliche Ausprägung: \bar{x}_i) als Level-2-Variable operationalisiert. Und der Within-Anteil der Determinante wird als Level-1-Variable, welche über die Abweichung vom individuellen Mittelwert gebildet wird ($\Delta x_{ti} = x_{ti} - \bar{x}_i$), berechnet, was rein technisch betrachtet der zuvor vorgestellten Group-Mean-Zentrierung entspricht. Beide (Teil-)Prädiktoren werden sodann gemeinsam in die Modellschätzung aufgenommen.²⁰ Folglich wird der Within-Personeneffekt einer Determinanten über die Spezifikation eines group-mean-zentrierten Prädiktors geschätzt, wohingegen der Between-Personeneffekt über die Spezifikation eines grand-mean-zentrierten Prädiktors ermittelt wird. Um beide Prädiktoren in unseren Analysen voneinander unterscheiden zu können, werden sie mit speziellen Bezeichnungen gekennzeichnet. Die within-ausgerichteten Prädiktoren werden mit dem Namenspräfix " Δ " und die between-ausgerichteten Prädiktoren mit dem Namenssuffix "_mean" gekennzeichnet.

In unseren Analysen ergibt sich somit der personale Wert eines Between-Prädiktors als Mittel über alle seine zeitbedingten Einzelwerte (bei einer bestimmten Person), wohingegen der personale Wert eines Within-Prädiktors die Variation um dieses Mittel widerspiegelt (vgl. Brüderl 2010: 983). So betrifft etwa der Prädiktor " Δ Gesundheit_{ti}" die Variation der Determinante "subjektive Gesundheitseinschätzung" innerhalb von Personen und spiegelt somit den individuellen Wandel von Bewertungen wieder, welcher für das folgende Beispiel von besonderem Interesse ist. Im Unterschied dazu betrifft der Prädiktor "Gesundheit_mean" den individuellen Mittelwert (Between-Anteil) der zeitveränderlichen Determinante "subjektive Gesundheitseinschätzung" und bezieht sich allein auf die Variation der Determinante zwischen den befragten Personen.

In der folgenden Abbildung 5 wird die Logik der Konstruktion von group-mean-zentrierten Prädiktoren verdeutlicht:

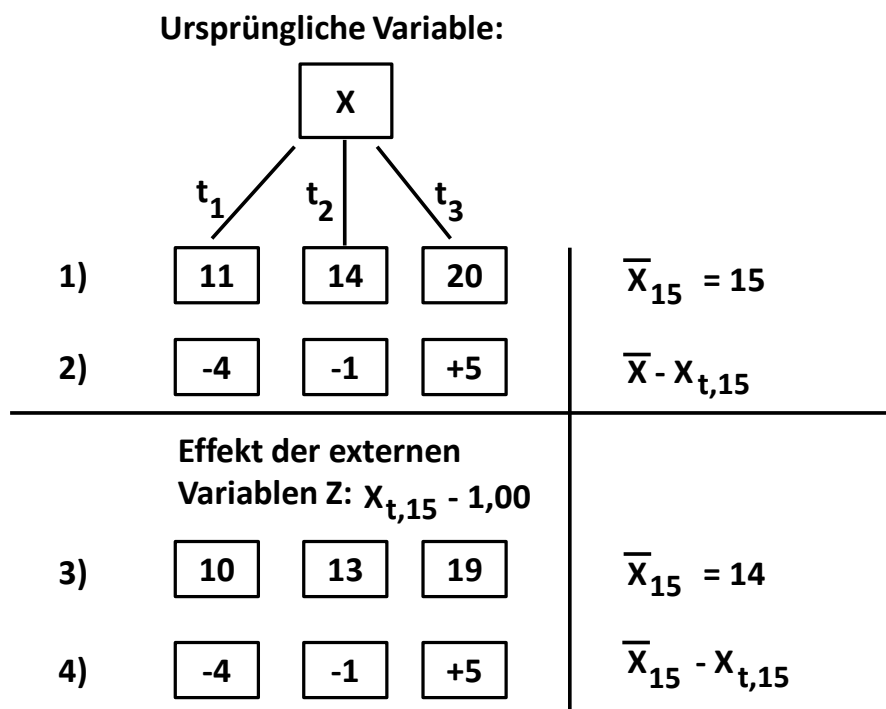
Abbildung 5 zeigt eine zeitvariante Variable "X", die über drei Zeitpunkte hinweg gemessen wird (t_1 , t_2 , t_3) und für Person No. 15 die drei verschiedenen Ausprägungen: 11, 14 und 20 annimmt (vgl. Abb. 5, Zeile 1). Nach der personenspezifischen (= group-mean) Zentrierung hat die X-Variable die drei

²⁰ Vgl. dazu auch Hedeker/Gibbons 2006: 73.

Ausprägungen: -4, -1 und +5 (vgl. Zeile 2). Nunmehr interveniere eine zeitkonstante, externe Variable "Z", die alle X-Werte um -1 verkleinert, so dass die X-Variable neue Werte annehme: 10, 13 und 19 (Zeile 3). Wenn nun diese neuen Werte wiederum zentriert werden, ergeben sich in Zeile 4 die gleichen Werte wie in Zeile 2. Somit wird durch die personenspezifische Zentrierung der Effekt der Variablen Z aufgehoben.

Im Unterschied zum Within-Prädiktor ist der geschätzte Effekt des Between-Prädiktors in aller Regel nicht von inhaltlichem Interesse. Der Between-Prädiktor kann aber in der Modellschätzung als Kontrollvariable für weitere zeitkonstante Einflussgrößen verwendet werden. Zudem kann sein Schätzwert als Vergleichs- bzw. Referenzwert für die Effektstärke des Within-Effekts dienen, da ein deutlicher (signifikanter) Unterschied zwischen den geschätzten Koeffizienten des Within- und Between-Prädiktors ein Hinweis darauf ist, dass in den Modelldaten eine zeitkonstante unbeobachtete Heterogenität besteht, welche ohne die hier vorgestellte Kontrolle das ermittelte Schätzergebnis verzerrt hätte (vgl. dazu Allison 2009: 25).

Abbildung 5: Logik der Konstruktion von group-mean-zentrierten Prädiktoren



3 Analysebeispiel: Die Bedeutung der subjektiven Gesundheitseinschätzung für die Bewertung der individuellen Lebenszufriedenheit

Mit den folgenden statistischen Analysen wird kein substantiell begründetes Forschungsinteresse verfolgt. Die hier vorgestellte Modellierung dient lediglich als illustratives Beispiel für die Durchführung einer Längsschnittanalyse mittels Mehrebenenmodellierung (RMD-MLA). Deshalb wird auch auf die Berücksichtigung der Effekte von Kontrollvariablen in der Modellschätzung verzichtet. Allerdings wird, wie zuvor erläutert, durch die Art der von uns verwendeten Längsschnittmethodik ein guter Teil von Drittvariablen (bzw. von zeitkonstanten Merkmalen) ganz automatisch kontrolliert.

Im untersuchten Analysebeispiel soll der Einfluss der subjektiven Gesundheitseinschätzung auf die Bewertung der individuellen Lebenszufriedenheit untersucht werden. Dabei soll auch der Frage nachgegangen werden, ob dieser Einfluss über alle befragten Personen im Zeitverlauf signifikant variiert und ob eine evtl. dabei festzustellende Varianz mit zeitinvarianten Personenmerkmalen statistisch zu erklären ist.

Zur Untersuchung dieser beiden Fragen werden im Folgenden verschiedene RMD-Mehrebenenschätzungen vorgestellt. Wir beginnen mit einfachen "random intercept" Mehrebenenmodellen und erweitern diese schrittweise zu Prädiktor-Modellen mit "random slopes" und "cross-level-interactions".

Bevor diese Analysebeispiele besprochen werden, wird in Abschnitt 3.1 zunächst das vor der statistischen Schätzung erforderliche, grundlegende Datenmanagement vorgestellt. Im Anschluss daran werden in Abschnitt 3.2 der Grundgedanke des Analyseablaufs sowie die damit verbundenen Tests für den Modellvergleich erläutert.

Unter der subjektiven Einschätzung einer allgemeinen Lebenszufriedenheit²¹ wird eine kognitiv geprägte, generalisierte Evaluation des eigenen Lebens anhand individueller Standards verstanden. Zur Messung der Bewertung des individuellen Gesundheitszustandes wurde in den Erhebungen eine 5-Punkte-Ratingskala verwendet.²²

²¹ Die subjektive Bewertung der allgemeinen Lebenszufriedenheit wurde durch das Frage-Item erhoben: "Wie zufrieden sind Sie gegenwärtig, alles in allem, mit Ihrem Leben?". Für die Bewertung stand eine 11-stufige Ratingskala zur Verfügung, wobei die jeweiligen Endkategorien 0 und 10 mit, "ganz und gar unzufrieden" und "ganz und gar zufrieden" verbalisiert wurden. Die Lebenszufriedenheit wird für die weiteren Analysen als metrisch skaliert aufgefasst.

²² Die subjektive Selbsteinschätzung wurde auf einer 5-stufigen Ratingskala mit den verbalisierten Kategorien "sehr gut", "gut", "zufriedenstellend", "weniger gut" und "schlecht" erhoben. In den weiteren Analysen wird die Variable als metrisch skaliert aufgefasst.

Der Lebensbereich "Gesundheit" hat nach der Selbsteinschätzung vieler Befragter eine überaus hohe Bedeutsamkeit für ihr subjektives Wohlbefinden. Im Wohlfahrtssurvey 1978 nannten beispielsweise die westdeutschen Befragten den Gesundheitsbereich als denjenigen Bereich, der für ihr Leben am wichtigsten ist (vgl. Glatzer 1984a: 198f). Betrachtet man allerdings nicht ausschließlich die Bedeutsamkeit des Gesundheitsbereichs sondern auch den Zusammenhang zwischen Gesundheitseinschätzung und einer globalen Lebenszufriedenheit, so hat die Gesundheitseinschätzung zwar einen bedeutsamen Einfluss auf die Lebenszufriedenheit, dieser ist aber bei weitem nicht mehr so dominant wie der Gesundheitsbereich bei der subjektiven Wichtigkeitseinschätzung (vgl. Glatzer 1984b: 236f; Diener 2009: 35). Der Zusammenhang zwischen Gesundheitszustand und subjektivem Wohlbefinden ist eher als moderat stark (und positiv gerichtet) einzustufen.

Allerdings bleibt dieser Zusammenhang auch unter Kontrolle von anderen Variablen erhalten (vgl. Diener 2009: 35). Dies gilt insbesondere für die subjektive Selbsteinschätzung des Gesundheitszustandes, wohingegen "objektivere" Maße, wie die Anzahl der Arztbesuche oder die Bewertung von Medizinern, eher schwächere Zusammenhangsstärken ergeben (vgl. Lucas/Diener 2009: 85). Es ist also immer noch das eigene gesundheitliche "Empfinden", das für die Bewertung des eigenen Lebens am wichtigsten ist.

So sind Personen, welche der Meinung sind, gesundheitliche Probleme zu haben, relativ unabhängig von ihrem objektiven Gesundheitszustand weniger zufrieden (vgl. Dette 2005: 57), wobei natürlich die Einflusstärke von subjektiv wahrgenommenen Gesundheitsveränderungen auf die Lebenszufriedenheit von Person zu Person variiert und davon abhängig ist, welche individuelle Wichtigkeit diesem Lebensbereich zugeordnet wird.

3.1 Datenmanagement

Für die folgenden Analysen wurden Variablen des SOEP-Datensatzes (s.o.) benutzt (Lebenszufriedenheit, Gesundheitseinschätzung und Geschlecht). Diese wurden jeweils viermal (2000, 2001, 2003 und 2006) gemessen. Dabei ist es für unsere Analysen wichtig (wichtig zu illustrativen Zwecken), dass diese Messzeitpunkte unterschiedlich weit voneinander entfernt sind.

Unsere Analysen beschränken sich auf Befragte, welche zumindest an zwei der vier Erhebungswellen erfolgreich teilgenommen haben, denn, wie oben ausgeführt, ist es für Längsschnittanalysen mit Mehrebenenmodellen nicht erforderlich, nur Fälle mit kompletten individuellen Zeitreihen in die Schätzung einzubeziehen (wichtiger Vorteil dieser Modellierung!). Jedoch wurde eine zweimalige Teilnahme als Mindesthäufigkeit definiert, damit für jeden Befragten auch genügend Informationen zur Berechnung der Within-Varianz (s.o.) von zeitvarianten Merkmalen vorliegen und somit auch tatsächlich stattgefundenen individueller Wandel untersucht werden kann.

Auf diese Weise ergibt sich für unsere Untersuchungen eine Fallzahl von 24.869 Befragten mit 83.313 Einzelbeobachtungen, was auch bedeutet, dass jede Person in der untersuchten Stichprobe durchschnittlich an 3,35 Wellen teilgenommen hat. Von den 83.313 Einzelbeobachtungen konnten für die Schätzung der L1-Effekte von Δ Zeit und Δ Gesundheit nur 56.426 Einzelbeobachtungen genutzt werden, da dazu nur die Daten solcher Fälle ausgewertet werden konnten, die über alle Wellen hinweg mindestens zwei verschiedene Gesundheitseinschätzungen geliefert haben (s.o.). Die Fallzahl von 24.869 Befragten wird dadurch nicht reduziert.

Bevor in einer RMD-MLA die ersten Modelle geschätzt werden können, müssen zunächst die Rohdaten in ein geeignetes Format gebracht werden. Diese Datenaufbereitung gestaltet sich für Paneldaten schwieriger und aufwendiger als für Querschnittsdaten:

Üblicherweise liegen Datensätze im sogenannten Wide-Format vor, in welchem für jeden Befragten eine eigene Zeile vorgesehen ist. Handelt es sich bei diesen Wide-Datensätzen um Paneldaten, finden sich in jede Zeile bzw. bei jedem Befragten mehrere Spalten, welche sowohl die Werte für einzelne zeitkonstante Variablen (z.B. Geschlecht) als auch mehrere Werte für zeitvariante Variablen enthalten, die pro Erhebungszeitpunkt einen spezifischen Wert aufweisen (z.B. Lebenszufriedenheit_2000, Lebenszufriedenheit_2001 usw.). Dementsprechend beinhaltet der Wide-Datensatz insgesamt k Spalten pro Befragten ($k = \text{Zahl der Variablen} \times \text{Anzahl der Erhebungswellen}$). Die Struktur eines solchen Datensatzes verdeutlicht Abbildung 6.²³

Abbildung 6: SPSS-Daten im Wide-Format²⁴

	PERSNR	gender	Lebenszufriedenheit_2000	Lebenszufriedenheit_2001	Lebenszufriedenheit_2003	Lebenszufriedenheit_2006	gesund_2000	gesund_2001	gesund_2003	gesund_2006
1	201	,00	9,00	9,00	7,00	6,00	3,00	2,00	3,00	4,00
2	203	1,00	8,00	7,00	8,00	5,00	2,00	1,00	2,00	2,00
3	601	1,00	8,00	8,00	2,00	.	2,00	2,00	2,00	.
4	602	,00	5,00	4,00	5,00	5,00	3,00	2,00	3,00	3,00

Bevor dann aber Mehrebenenmodelle geschätzt werden können, müssen die Daten vom Wide-Format (vgl. Abb. 6) in das Long-Format (vgl. die folgende Abbildung 7) transformiert werden. Das Long-Format enthält pro Befragten nicht mehr nur eine Zeile, sondern bei Teilnahme von allen Befragten

²³ Dies ist prinzipiell unabhängig davon, ob im Datensatz auch wirklich gültige Werte für jeden Befragten in jeder Panelwelle enthalten sind.

²⁴ Die Matrix zeigt nur die Rohwerte der Variablen vor der Zentrierung.

an allen Erhebungswellen insgesamt z Datenzeilen ($z = \text{Anzahl der Befragten} \times \text{Anzahl der Erhebungswellen}$). Deshalb ist auch zur eindeutigen Identifikation bzw. Zuordnung der Daten für jeden Befragten eine eindeutige Identifikationsnummer nötig, die sich dann in mehreren Zeilen wiederfindet (hier die Variable "PERSNR", vgl. Abb. 7).

Neben der Personenkennziffer wird im Long-Format-Datensatz zusätzlich eine Zählvariable für die Wellenzuordnung benötigt (hier: Index1), welche durch SPSS im Befehl zur Umformatierung von Daten in das Long-Format (nahezu) automatisch generiert wird.²⁵

Wie bereits zuvor mehrfach erörtert, sollten die empirisch gemessenen Variablen im Datensatz zentriert werden (u.a. um Multikollinearitätseffekte zu vermeiden und die Interpretation der Modellschätzung zu vereinfachen).

Bei Fällen mit fehlenden Werten sollte überlegt werden, ob diese Werte über ein Imputationsverfahren geschätzt und ergänzt werden können. Sollte dies nicht möglich sein, sollten aus dem Long-Format-Datensatz alle Fälle ausgeschlossen werden, welche

- a) fehlende Werte bei zeitkonstanten Prädiktoren aufweisen;
- b) fehlende Werte bei allen Wellen von zeitvariablen Prädiktoren aufweisen;
- c) an weniger als j Wellen teilgenommen haben (hier: $j=2$).

Beim Ausschluss einzelner Fälle müssen evtl. nicht nur alle Zeilen einer Person sondern auch einzelne Messzeitpunkt-Zeilen gelöscht werden. Falls dieser Ausschluss nicht im Vorhinein vorgenommen wird, werden die häufig mehrfach durchgeführten Modellschätzungen einer RMD-MLA auf unterschiedlichen Fallzahlen/Fällen basieren, wodurch ein Modellvergleich extrem erschwert bzw. verunmöglicht wird.

²⁵ Die für die Umformatierung (vom Wide- in das Long-Format) benutzte SPSS-Syntax findet sich im Anhang dieses Textes (vgl. Anhang A1).

Abbildung 7: SPSS-Daten im Long-Format²⁶

	PERSNR	gender	Index1	Lebenszufriedenheit	Gesundheit	zeit_delta
1	201	,00	1	9,00	2,00	,00
2	201	,00	2	9,00	3,00	1,00
3	201	,00	3	7,00	2,00	3,00
4	201	,00	4	6,00	1,00	6,00
5	203	1,00	1	8,00	3,00	,00
6	203	1,00	2	7,00	4,00	1,00
7	203	1,00	3	8,00	3,00	3,00
8	203	1,00	4	5,00	3,00	6,00
9	601	1,00	1	8,00	3,00	,00
10	601	1,00	2	8,00	3,00	1,00
11	601	1,00	3	2,00	3,00	3,00
12	602	,00	1	5,00	2,00	,00
13	602	,00	2	4,00	3,00	1,00
14	602	,00	3	5,00	2,00	3,00
15	602	,00	4	5,00	2,00	6,00

3.2 Analyseablauf

Die im Weiteren vorgestellte Analysestrategie folgt der "bottom-up" Logik nach Hox (ders. 2010). Danach werden in MLA zunächst einfache Modelle konstruiert und geschätzt, welche dann Schritt für Schritt um zusätzliche Parameter erweitert werden, bis die Analyse beim jeweiligen Endmodell angelangt ist (vgl. Hox 2010: 56). Nach jedem Schritt ist dabei zu untersuchen, ob die Schätzung zusätzlich eingeführter Parameter eine Verbesserung der Modellanpassung gegenüber der vorausgegangenen Modellschätzung erbringt.²⁷

Zum Modellvergleich wird bei der hier benutzten Analysestrategie jeweils ein Likelihood-Ratio-Test (LRT) verwendet. Eine Bedingung zur korrekten Anwendung des Likelihood-Ratio-Tests ist die

²⁶ Die Matrix zeigt bis auf die Variable "zeit_delta", bei der die Jahreszahlen (2000, 2001, 2003, 2006) auf die erste Jahreszahl zentriert wurden (0, 1, 3, 6), nur die Rohwerte vor der Zentrierung. Da Person 601 an der 2006er Befragung nicht teilnahm, gibt es für sie im Datensatz nur vier und nicht fünf Zeilen.

²⁷ Neben der Modellverbesserung sind auch eher praktische Kriterien für die Modellbewertung zu beachten, wie die Beibehaltung von nicht signifikanten Haupteffekten bei der Schätzung von Interaktionseffekten. So sollten in der Regel auch nicht signifikante Haupt(-Fixed-)Effekte im Modell verbleiben, wenn sie in späteren Modellen als "random" spezifiziert werden (dazu mehr im Folgenden).

Verwendung des Maximum-Likelihood-Schätzverfahrens (MLE) zur Modellschätzung (vgl. Singer/Willett 2003: 118).²⁸ In diesem Test wird die Differenz zwischen den Devianzwerten zweier aufeinander aufbauender (geschachtelter) Modelle²⁹ berechnet und der Differenzwert mittels Verwendung der Chi-Quadrat-Verteilung auf Signifikanz getestet. Der Likelihood-Ratio-Test ermöglicht die simultane Testung mehrerer Parameterveränderungen und vermeidet in der Tendenz die fälschliche Verwerfung der Nullhypothese bzw. den Typ 1-Fehler.³⁰

Zusätzlich zum Likelihood-Ratio-Test wird in unseren Analysen danach geschaut, wie hoch der zusätzliche Anteil der geschätzten Varianzen (auf den verschiedenen Modellebenen) ist, der sich durch die Hinzunahme zusätzlicher Modellparameter zusätzlich statistisch erklären lässt. Dabei lässt sich die Reduktion der jeweiligen Varianz auf jeder Modellebene über die Formel $(\sigma^2_{\text{Vormodell}} - \sigma^2_{\text{Folgemodell}}) / \sigma^2_{\text{Vormodell}}$ (für geschachtelte Modelle) abschätzen.

Hox (ders. 2010: 56ff) schlägt für den Ablauf von Mehrebenenanalysen fünf prototypische Schritte vor:

1. die Schätzung des Null/Intercept-Only-Modells und die Berechnung des Intraklassenkorrelationskoeffizienten (ICC);
2. die Hinzufügung von unabhängigen L1-Prädiktoren zur Schätzung von zusätzlichen "fixed effects";
3. die Hinzufügung von unabhängigen L2-Prädiktoren zur Schätzung weiterer "fixed effects";
4. die Definition von Level 1-Variablen als "random-effects" und eine dementsprechend modifizierte Modellschätzung;
5. die Hinzufügung von Cross-Level-Interaktionen zur Erklärung von evtl. vorhandenen Variabilitäten in den "random slopes" und eine dementsprechend modifizierte Modellschätzung.

Dieser Analyseablauf wird im Weiteren aufgegriffen, wobei jedoch einige Modifikationen aufgrund der Besonderheiten einer Längsschnittanalyse vorgenommen werden. So wird u.a. der Effekt der jeweils vergangenen Prozesszeit zunächst als Fixed- und später als Random-Effekt geschätzt, um die durchschnittliche zeitliche Veränderung von Variablen abzubilden und später die strikten Annahmen

²⁸ Auch weitere Modellgütekriterien wie z.B. Devianz, Akaike information criterion (AIC) und Bayesian information criterion (BIC) werden mittels ML-Informationen errechnet.

²⁹ Für diesen Test müssen die zu vergleichenden Modelle ineinander geschachtelt (nested) sein, was bedeutet dass sich das weniger komplexe Modell aus dem komplexen Modell ergeben muss, wenn im komplexeren Modell ein oder mehrere Modellparameter auf null gesetzt werden bzw. aus der Analyse ausgeschlossen werden. Auch muss dafür bei jeder Modellschätzung der gleiche Datensatz mit der gleichen Anzahl von Fallzeilen (s.o.) benutzt werden.

³⁰ Anders betrachtet, kann der Likelihood-Ratio-Test auch als zu konservativ bezüglich der Testung von Varianz- und Kovarianzparametern betrachtet werden, wodurch die Nullhypothese nicht oft genug verworfen wird (vgl. Hedeker/Gibbons 2006: 52).

der "compound symmetry" (s.o.) zu entspannen. Des Weiteren wird Modellierungsschritt 2 mit Teilen von Schritt 3 verbunden, indem die Within- und Between-Anteile der zeitveränderlichen Determinanten (s.o.) jeweils parallel in die Modelle aufgenommen werden.

In den im Anschluss gezeigten Analysen werden folgende Variablen benutzt:

- (-) Lebenszufriedenheit, zu erklärende Variable, zeitvariant, nicht-zentriert;
- (-) Δ Zeit, Prädiktor, zeitvariant, L1-Prädiktor, nullwert-zentriert;
- (-) Δ Gesundheit, zeitvariant, L1-Prädiktor, personen-zentriert;
- (-) Gesundheit_mean, zeitinvariant, L2-Prädiktor, grand-mean-zentriert;
- (-) Geschlecht, zeitinvariant, L2-Prädiktor (kodiert: m=1/w=0).

3.3 Schritt 1: Intercept-only-Modell (M1.0)

Der erste Schritt der Analyse beginnt mit der Spezifizierung des Intercept-Only-Modells (M1.0). Das auch als Nullmodell bezeichnete Modell "M1.0" enthält noch keine erklärenden Variablen und dient zunächst nur für eine erste Schätzung der Varianzanteile auf den verschiedenen Ebenen. Hierbei wird angenommen, dass die Fehlerterme auf Ebene 1 " ε_{it} " und auf Ebene 2 " u_{0i} " jeweils:³¹

- 1.) unabhängig voneinander sind (z.B. keine Autokorrelationen bestehen);
- 2.) einen Erwartungswert von Null besitzen;
- 3.) normalverteilt sind;
- 4.) keine Varianzungleichheiten (Heteroskedastizität) aufweisen.

Im Folgenden werden die beiden Modellgleichungen für Level 1 und Level 2 des Nullmodells vorgestellt:

Auf der ersten Ebene ergibt sich für jeden Messzeitpunkt (t) ein spezieller Y-Wert (der Wert der abhängigen Variablen) für jeden i-ten Befragten (Y_{it}). Dieser setzt sich zusammen aus einem Intercept (β_{0i}), das bei jedem Befragten für alle Messzeitpunkte gleich ist und nur zwischen den Befragten variiert, und einem Zufallswert (ε_{it}) der zwischen den Messzeitpunkten und zwischen den Befragten variiert.

Auf der zweiten Ebene wird das Intercept (β_{0i}) aus der ersten Ebene als abhängige Variable betrachtet. Es wird in seiner Ausprägung bestimmt von einer Konstanten (γ_{00}), die unabhängig vom Messzeitpunkt und unabhängig vom Befragten stets den gleichen Wert aufweist, und von einem Zufallswert (u_{0i}), der von Befragtem zu Befragtem unterschiedlich sein kann. Somit wird in diesem

³¹ Vgl. dazu Snijders/Boskers 2012: 249.

Nullmodell das Intercept als "random" behandelt, d.h. in diesem Modell variiert das Intercept zwischen den einzelnen Befragten in Abhängigkeit von einer Zufallsgröße (u) auf der Ebene 2 (=Personenebene).

$$\text{M1.0} \quad \left| \begin{array}{l} \text{Level 1:} \\ Y_{ti} = \beta_{0i} + \varepsilon_{ti} \end{array} \right. \quad \left| \begin{array}{l} \text{Level 2:} \\ \beta_{0i} = \gamma_{00} + u_{0i} \end{array} \right.$$

t=Messzeitpunkt; i=Befragte/r

In Längsschnittmodellen, in welchen lediglich das Intercept als "random" spezifiziert ist (wie hier im Nullmodell), wird üblicherweise von einer "compound symmetry" (s.o)³² ausgegangen (vgl. Hox 2010: 99).

Dies bedeutet, dass die Kovarianzmatrix der zu schätzenden Parameter eine Form besitzt, in welcher alle L2-Parameter-Varianzen den gleichen Wert annehmen und alle Kovarianzen ebenfalls einen identischen Wert aufweisen. In späteren Modellen können diese Annahmen aufgegeben werden, indem dort u.a. auch der Zeitparameter als "random" spezifiziert wird und somit die Kovarianzmatrix in Abhängigkeit von der Zeit variieren kann (vgl. Snijders/Boskers 2012: 253; Hox 2010: 102).

Mittels der im Nullmodell geschätzten Varianzen $\text{Var}(\varepsilon_{ti})$ und $\text{Var}(u_{0i})$ (vgl. die folgende Tabelle 1.0) lässt sich der Interkorrelationskoeffizient (ICC) nach der folgenden Formel berechnen:

$$\text{ICC} = \frac{\text{Var}(u_{0i})}{\text{Var}(\varepsilon_{ti}) + \text{Var}(u_{0i})}$$

Für das geschätzte Nullmodell M1.0 ergibt sich dabei ein Wert von 0,50, wonach in einer ersten Annäherung ca. 50% der Varianz der Lebenszufriedenheit auf die Unterschiede zwischen den Befragten (=Personen, =Gruppen, =Kontexte) zurückgeführt werden kann. Somit kann auch ca. 50% der Varianz der Lebenszufriedenheit auf Unterschiede zwischen den Messzeitpunkten zurückgeführt werden. Dabei ist jedoch zu beachten, dass dieser Wert nicht nur durch den intraindividuellen Wandel der Lebenszufriedenheit sondern auch durch zufällige Schwankungen bzw. zufällige Messfehler entsteht (vgl. Klein/Pötschke 2004: 444). Bei einem so hohen ICC-Wert wie 0,50 sollte die abhängige Variable nicht mehr mittels linearer Ein-Ebenen-Regression und gepoolten Daten sondern nur noch mittels RMD-MLA durchgeführt werden.

³² Genauer gesagt, gehen Random-Intercept-Modelle nur dann von einer "compound symmetry" aus, wenn die Fehler als unabhängig angenommen werden (vgl. Hedeker/Gibbons 2006: 52).

Tabelle 1.0: Schätzung des Nullmodells M1.0

Modellbezeichnung	M1.0
Fixed Effects	Koeffizient (S.E.)
Konstante (γ_{00})	7,020*** (,0090)
Random Effects	Varianz (S.E.)
Residuum (Varianz ϵ_{it})	1,571*** (,0092)
Konstante (Varianz u_{0i})	1,559*** (,0187)
Modellmaße	
Devianz	309959,296
AIC	309965,296
BIC	309993,287
ICC	0,498
Anzahl geschätzter Parameter	3

Die Koeffizienten- und Varianzschätzung erfolgte mit REML-Verfahren. Zur Berechnung der Modellgüte wurden die ML-Schätzwerte benutzt. Fallzahl: 24.869 Befragte mit 83.313 Einzelbeobachtungen. Signifikanz: n.s.= $p \geq 0.05$, * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$.

Die Ergebnisse dieses ersten Modellierungsschritts zeigen mit den jeweiligen einseitigen Wald-Tests³³ für die Varianzwerte (für $\text{Var}(\epsilon_{it})$: Wald $Z=170,95$; $p < 0.001$ und für $\text{Var}(u_{0i})$: Wald $Z=83,50$; $p < 0.001$), dass sowohl ausreichend Varianz zwischen den Befragten als auch zwischen den Messzeitpunkten der jeweiligen Befragten besteht, welche durch Hinzunahme von zeitvarianten und zeitinvarianten Determinanten erklärt werden kann.

3.4 Schritt 2/3 – Fixed Effects M2.3.1-M3.0

Im nächsten Schritt wird zunächst das Modell M2.3.1 spezifiziert. In Modell M2.3.1 wird die Zeit³⁴ als linearer Parameter spezifiziert, dessen Koeffizient zunächst für alle Befragten als identisch angenommen wird (vgl. Tabelle 2.1). Die folgende Auflistung zeigt die dementsprechenden Modellgleichungen der RMD-MLA. Die Ergebnisse der Modellschätzung zeigt die folgende Tabelle 2.1.

$$\begin{array}{l|l}
 \text{M2.3.1} & \text{Level 1:} \\
 & Y_{ti} = \beta_{0i} + \beta_{1i} (\Delta\text{Zeit})_{ti} + \epsilon_{ti} \\
 & t=\text{Messzeitpunkt}; i=\text{Befragte}/r \\
 & \text{Level 2:} \\
 & \beta_{0i} = \gamma_{00} + u_{0i}; \quad \beta_{1i} = \gamma_{10}
 \end{array}$$

³³ Die Testgröße für den Wald-Test ergibt sich aus dem Quotienten: (geschätzte Varianz)/(Standardfehler der geschätzten Varianz).

³⁴ Der Zeitparameter nimmt die Werte 0, 1, 3 und 6 an und bildet somit die vergangene Prozesszeit ab, die sich aus den Messzeitpunkten 2000, 2001, 2003, 2006 ergibt.

Um den Schätzerfolg dieses Modells mit demjenigen des Nullmodells zu vergleichen, wird ein Likelihood-Ratio-Test (s.o.) durchgeführt. Der LRT zeigt mit einem Wert von 1042,70 ($309959,296 - 308916,595 = 1042,70$; $df=1$) eine signifikant bessere Modellpassung gegenüber dem Nullmodell an ($p < 0.001$).

Tabelle 2.1: Modelle M2.3.1 und M2.3.2

Modellbezeichnung	M2.3.1	M2.3.2
Fixed Effects	Koeffizient (S.E.)	
Konstante (γ_{00})	7,158*** (.0101)	7,117*** (.0089)
Δ Zeit (γ_{10})	-0,067*** (.0021)	-0,051*** (.0020)
Δ Gesundheit (γ_{20})	-	0,505*** (.0082)
Gesundheit mean (γ_{01})	-	0,879*** (.0097)
Random Effects	Varianz (S.E.)	
Residuum (Varianz ϵ_{it})	1,538*** (.0090)	1,445*** (.0085)
Konstante (Varianz u_{0i})	1,586*** (.0188)	1,096*** (.0142)
Modellmaße		
Devianz	308916,595	298109,800
AIC	308924,595	298121,800
BIC	308961,916	298177,782
Anzahl Parameter	4	6

Die Koeffizienten- und Varianzschätzung erfolgte mit REML-Verfahren. Zur Berechnung der Modellgüte wurden die ML-Schätzwerte benutzt. Fallzahl: 24.869 Befragte mit 83.313 Einzelbeobachtungen. Signifikanz: $n.s. = p \geq 0.05$, * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$.

Der einfache t-test des geschätzten Δ Zeit-Parameters bestätigt die Angemessenheit der Inklusion des Zeitparameters ($t = \gamma_{10}/SE = -32,51$; $p < 0.001$). Nach Modell M2.3.1 sinkt für jedes vergangene Jahr die Lebenszufriedenheit im Schnitt um ca. 0,07 Einheiten.

Das Intercept von 7,16 bezieht sich nun auf den durchschnittlichen Wert zum Messzeitpunkt 2000, um welchen die Zeitvariable zentriert wurde (s.o.). Vergleicht man die L1-Varianzen zwischen den Modellen M1.0 und M2.3.1, so zeigt sich, dass durch die Aufnahme der Prozesszeit ca. 2,1% der Varianz der ersten Ebene statistisch erklärt werden kann.

Vergleicht man hingegen die L2-Varianzen zwischen den Modellen M1.0 und M2.3.1, so gelangt man zu einem negativen Wert.³⁵ Diese scheinbare Abnormität ist bei Mehrebenenanalysen mit Messwiederholungen nicht ungewöhnlich und ergibt sich aus dem Samplingprozess des Verfahrens (vgl. Hox 2010: 88). Zur Lösung dieses Problems wird vorgeschlagen, für die weitere Modellanalyse nicht mehr das Nullmodell als Referenzmodell sondern das Modell M2.3.1 zu benutzen. Denn durch die Aufnahme der Prozesszeit in Modell M2.3.1 wird die Variabilität zwischen den Messzeitpunkten und zwischen den Befragten realistischer als im Nullmodell abgebildet, da das Intercept-Only-Modell die Varianz der Messzeitpunkte überschätzt und die Varianz zwischen den Befragten unterschätzt (vgl.

³⁵ Berechnet nach der Formel " $(\sigma^2_{\text{Vormodell}} - \sigma^2_{\text{Folgemodell}}) / \sigma^2_{\text{Vormodell}}$ " ergibt sich ein Wert von: $(1,559 - 1,586) / 1,559 = -0.0173$.

ebd.: 88f). Somit dient M2.3.1 als geeigneteres Referenzmodell für die weitere Modellkonstruktion. Im folgenden Modell 2.3.2 werden die beiden Prädiktoren des Between- und Within-Person-Anteils der subjektiven Gesundheitseinschätzung als zusätzliche Parameter in die Modellschätzung aufgenommen. Die folgende Auflistung zeigt die diesbezüglichen Modellgleichungen:

M2.3.2	Level 1: $Y_{ti} = \beta_{0i}$ $+ \beta_{1i} (\Delta\text{Zeit})_{ti} + \beta_{2i} (\Delta\text{Gesundheit})_{ti} + \varepsilon_{ti}$	Level 2: $\beta_{0i} = \gamma_{00} + \gamma_{01}(\text{Gesundheit_mean})_i + u_{0i}$ $\beta_{1i} = \gamma_{10}; \quad \beta_{2i} = \gamma_{20}$
	t=Messzeitpunkt; i=Befragte/r	

Der LRT der neuen Modellschätzung M2.3.2 indiziert mit einem Wert von 10806,80 (df=2) eine signifikant bessere Modellpassung als sie mit M2.3.1 erreicht werden konnte ($p < 0.001$).

Auch die t-Tests bestätigen einen signifikanten Zusammenhang zwischen dem Wandel der subjektiven Gesundheitseinschätzung sowie der durchschnittlichen Gesundheitseinschätzung und der Lebenszufriedenheit (γ_{20} : $t=61,40$, $p < 0.00$; γ_{01} : $t=90,83$, $p < 0.001$). Durch den Within-Anteil der subjektiven Gesundheitseinschätzung kann ca. 6,0% ($(1,538-1,445)/1,538=0,06$) der Varianz zwischen den Befragungszeitpunkten statistisch erklärt werden. Zusätzlich kann die Varianz zwischen den Befragten durch den Between-Anteil um ca. 31% reduziert werden ($((1,586-1,096)/1,586=0,31)$).

Darüber hinaus zeigt sich, dass der Wert des Zeitkoeffizienten durch die Kontrolle des Wandels der Gesundheitseinschätzung abnimmt. Es liegt nahe, dass dieser Rückgang auf den mit der Alterung einhergehenden, durchschnittlich schlechteren Gesundheitszustand zurückzuführen ist, welcher unter gleichzeitiger Berücksichtigung des Gesundheitswandels als selbständiger Effekt auspartialisiert wird.

Die Interpretation der zeitvarianten Determinante "Gesundheitseinschätzung" muss die Effekte von zwei Gesundheitsprädiktoren berücksichtigen, da sowohl die Within- als auch die Between-Anteile der Determinante als erklärende Variablen im Modell spezifiziert wurden. Für die "Between"-Interpretation gilt auf Basis der Schätzung von Modell M2.3.2, dass Befragte, welche sich um eine Einheit bei der durchschnittlichen gesundheitlichen Selbsteinschätzung unterscheiden, auch einen Unterschied von ca. 0,88 Skaleneinheiten bzgl. der Lebenszufriedenheit aufweisen. Die "Within"-Interpretation konstatiert, dass eine Verbesserung der Gesundheit um einen Skaleneinheit (unter Kontrolle der Prozesszeit) mit einem durchschnittlichen Anstieg der Lebenszufriedenheit um 0,5 Skaleneinheiten verbunden ist. Beide Effekte sind hochgradig signifikant.

Es muss dazu jedoch grundsätzlich angemerkt werden, dass bei den sequenziellen Modellschätzungen einer RMD-MLA auch nicht-signifikante Level 1-Effekte weiterhin in den Modellen verbleiben können. Selbst bei einem durchschnittlich nicht von null unterscheidbaren (Fixed-)Effekt kann

eine signifikante Varianz des Effekts (signifikanter Random-Effekt) zwischen den Befragten bestehen (vgl. Singer/Willet 2003: 222), welche wiederum über personenspezifische Merkmale erklärt werden könnte. Dies bedeutet konkret, dass auch dann, wenn der Level 1-Effekt von Δ Gesundheit nicht signifikant gewesen wäre, dennoch der Effekt Varianz auf der zweiten Modellebene (zwischen den Befragten) eine bedeutsame Varianz aufweisen könnte.

Um den dritten Schritt der Modellkonstruktion abzuschließen muss noch eine Modellschätzung vorgenommen werden, bei der alle zeitinvarianten Merkmale in das Modell aufgenommen werden. So muss in der Schätzung zusätzlich zum Prädiktor "Gesundheit_mean" auch noch der Prädiktor "Geschlecht"³⁶ als weitere zeitinvariante Größe auf der zweiten Modellebene berücksichtigt werden. Die nachfolgende Auflistung zeigt die dementsprechenden Modellgleichungen und die folgende Tabelle 2.2 zeigt die Ergebnisse der Modellschätzungen.

M3.0	Level 1: $Y_{ti} = \beta_{0i}$ $+ \beta_{1i} (\Delta\text{Zeit})_{ti} + \beta_{2i} (\Delta\text{Gesundheit})_{ti} + \varepsilon_{ti}$	Level 2: $\beta_{0i} = \gamma_{00} + \gamma_{01}(\text{Gesundheit_mean})_i + \gamma_{02}(\text{Geschlecht})_i + u_{0i}$ $\beta_{1i} = \gamma_{10}; \quad \beta_{2i} = \gamma_{20}$
t=Messzeitpunkt; i=Befragte/r		

Tabelle 2.2: Modelle M3.0 und M4.0

Modellbezeichnung	M3.0		M4.0	
Fixed Effects	Koeffizient (S.E.)			
Konstante (γ_{00})	7,165***	(,0117)	7,164***	(,0118)
Δ Zeit (γ_{10})	-0,051***	(,0020)	-0,051***	(,0021)
Δ Gesundheit (γ_{20})	0,505***	(,0082)	0,501***	(,0082)
Gesundheit_mean (γ_{01})	0,882***	(,0097)	0,883***	(,0097)
Geschlecht (γ_{02})	-0,100***	(,0158)	-0,100***	(,0158)
Random Effects	Varianz (S.E.)			
Residuum (Varianz ε_{ti})	1,445***	(,0085)	1,347***	(,0095)
Konstante (Varianz u_{0i})	1,094***	(,0141)	1,214***	(,0188)
Konstante-Zeit (Cov u_{0i}, u_{1i})	-		-0,041***	(,0035)
Δ Zeit (Varianz u_{1i})	-		0,016***	(,0010)
Modellmaße				
Devianz	298069,933		297782,227	
AIC	298083,933		297800,227	
BIC	298149,246		297884,201	
Anzahl Parameter	7		9	

Die Koeffizienten- und Varianzschätzung erfolgte mit REML-Verfahren. Zur Berechnung der Modellgüte wurden die ML-Schätzwerte benutzt. Fallzahl: 24.869 Befragte mit 83.313 Einzelbeobachtungen. Signifikanz: n.s.= $p \geq 0.05$, * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$.

³⁶ Die Variable "Geschlecht" wurde mit 1=männlich und 0=weiblich kodiert.

Als Level 2-Variable sollte das Geschlecht die Unterschiede zwischen den Befragten (Between-Varianz) (statistisch) erklären können. Der LR-Test zeigt denn auch, dass Modell M3.0 eine signifikant bessere Anpassung besitzt als Modell M.2.3.2 (Devianz-Unterschied (df=1)=39.87; $p<0.001$).

Der Haupteffekt von Geschlecht wird signifikant negativ ($p<0.001$), was bedeutet, dass unter Kontrolle der restlichen Parameter die befragten Männer eine durchschnittlich um 0,10 Skaleneinheiten geringere Lebenszufriedenheit aufweisen als Frauen.

Allerdings kann durch die zusätzliche Berücksichtigung des Geschlechts lediglich 0,2% der Level 2-Varianz gebunden werden ($((1,096-1,094)/1,096=0,002$), woraus der Schluss gezogen werden kann, dass dem Haupteffekt von Geschlecht keine substantielle Bedeutung zukommt. Wie erwartet, bleibt die Level 1-Varianz konstant, da lediglich der Level 2-Parameter "Geschlecht" dem Modell hinzugefügt wurde.

3.5 Schritt 4 – Random Slopes M4.0-M4.3

Im vorletzten Modellierungsschritt werden die Level 1-Variablen als "random" spezifiziert und auf Signifikanz der jeweiligen Varianz untersucht. Zunächst wird hierzu der Effekt der Prozesszeit zwischen den Befragten freigegeben (vgl. Tabelle 2.2: M4.0), wodurch nicht mehr von einem einheitlichen zeitlichen Wandel für alle Befragten ausgegangen wird. Dadurch wird zudem, wie zuvor angesprochen, die Annahme einer "compound symmetry" (s.o.) aufgegeben. Durch Freisetzung der Prozesszeit können nunmehr die Varianzen und die Kovarianzen der Messzeitpunkte unterschiedliche Werte annehmen³⁷ (vgl. Hox 2010: 102 f.; Snijders/Boskers 2012: 253). Die nachfolgende Auflistung zeigt die dementsprechenden Modellgleichungen, und die oben abgedruckte Tabelle 2.2 zeigt die Ergebnisse der Modellschätzung.

M4.0	Level 1: $Y_{it} = \beta_{0i}$ $+ \beta_{1i} (\Delta\text{Zeit})_{it} + \beta_{2i} (\Delta\text{Gesundheit})_{it} + \varepsilon_{it}$	Level 2: $\beta_{0i} = \gamma_{00} + \gamma_{01}(\text{Gesundheit_mean})_i + \gamma_{02}(\text{Geschlecht})_i + u_{0i}$ $\beta_{1i} = \gamma_{10} + u_{1i}; \quad \beta_{2i} = \gamma_{20}$
	t=Messzeitpunkte; i=Befragte	

Der LRT zeigt bei einem Devianz-Unterschied von 287,71 (df=2, da eine Kovarianz und eine Varianz) eine deutlich verbesserte Modellpassung der Schätzung M4.0 im Vergleich zur Schätzung M3.0 an ($p<0.001$). Der Test unterstützt damit die Vermutung, dass die Annahme einer "compound symmetry" mit den Daten nicht völlig in Einklang zu bringen ist. Der einseitige Wald-Test ($\text{Var}(u_{1i})$): Wald-Z=15,83; $p<0.001$) bestätigt ebenfalls, dass zwischen den Befragten eine signifikante Varianz

³⁷ Die Varianz für einen Messzeitpunkt t errechnet sich hier (Modell M4.0) über die Formel: $\text{Var}(\varepsilon_{it}) + \text{Var}(u_{0i}) + 2 \times \text{Cov}(u_{0i}, u_{1i}) \times t + \text{Var}(u_{1i}) \times t^2$. Die Kovarianz zwischen zwei Messzeitpunkten t_x und t_y berechnet sich hingegen über: $\text{Var}(\varepsilon_{it}) + \text{Cov}(u_{0i}, u_{1i}) \times (t_x + t_y) + \text{Var}(u_{1i}) \times t_x \times t_y$ (nach Singer/Willet 2003: 252, 254).

bzgl. des Zeiteinflusses besteht.

Die Level 1-Varianz innerhalb der Befragten reduziert sich durch die Freisetzung des Zeiteffekts um ca. 7,0% von 1,45 auf 1,35, was jedoch keine substantielle Bedeutung hat, sondern lediglich durch Verschiebung der Varianzanteile auf die höhere Ebene entsteht.

Des Weiteren besteht eine signifikante Kovarianz zwischen u_{0i} und u_{1i} , welche anzeigt, dass Personen mit hoher Lebenszufriedenheit zum ersten Messzeitpunkt und unter Kontrolle der restlichen Variablen eine negativere zeitliche Steigung³⁸ aufweisen als Personen mit einem geringeren Startniveau an Lebenszufriedenheit.³⁹

Im nächsten Schritt gilt es zu testen, ob der Einfluss des Gesundheitswandels auf die Lebenszufriedenheit eine signifikante Varianz zwischen den Befragten aufweist und, wenn ja, ob diese über zeitinvariante Personenmerkmale (statistisch) erklärt werden kann.⁴⁰

Modell M4.1 enthält somit drei weitere Parameter: die Varianz des Gesundheitswandels ($\text{Var}(u_{2i})$) sowie deren Kovarianzen mit u_{0i} und u_{1i} . Die nachfolgende Auflistung zeigt die dementsprechenden Modellgleichungen und die folgende Tabelle 2.3 zeigt die Ergebnisse der Modellschätzung.

M4.1	Level 1: $Y_{ti} = \beta_{0i}$ $+ \beta_{1i} (\Delta\text{Zeit})_{ti} + \beta_{2i} (\Delta\text{Gesundheit})_{ti} + \varepsilon_{ti}$	Level 2: $\beta_{0i} = \gamma_{00} + \gamma_{01}(\text{Gesundheit_mean})_i + \gamma_{02}(\text{Geschlecht})_i + u_{0i}$ $\beta_{1i} = \gamma_{10} + u_{1i}; \quad \beta_{2i} = \gamma_{20} + u_{2i}$
	t=Messzeitpunkt; i=Befragte/r	

Der Vergleich von Modell M4.0 mit M4.1 mittels LRT zeigt eine signifikant bessere Passung von Modell M4.1 (Devianz-Unterschied (df=3)=985,18; $p < 0.001$). Auch der Wald-Test bestätigt, dass der Einfluss des Gesundheitswandels über die Befragten variiert ($\text{Var}(u_{2i})$: Wald-Z=21,76; $p < 0.001$).⁴¹

³⁸ Da sich nach der Normalverteilungsannahme der Random-Slopes (vgl. Hox 2010: 89f) ca. 95% der Befragten einen Zeit-Koeffizienten im Intervall [-0.2989, 0.1969] aufweisen, muss an dieser Stelle der zugegebenermaßen ungeschickte Begriff "negativere Steigung" verwendet werden.

³⁹ Für Kovarianzparameter ist der zweiseitige Wald-Test angebracht, da Kovarianzen sowohl negative als auch positive Werte annehmen können.

⁴⁰ Durch die Spezifikation eines weiteren Random-Slopes verändern sich die Kovarianzmatrizen der Folgemodelle nochmals gegenüber M4.0, da die Kovarianzmatrix nun auch in Abhängigkeit der Variation des Gesundheitszustands geformt wird.

⁴¹ Mangels substanziellen Interesses wird hier auf die Interpretation und Darstellung der Kovarianzen verzichtet.

Somit wird die zusätzliche Freisetzung des Effekts des Gesundheitswandels durch die Daten gestützt, sodass es sich im nächsten Schritt anbietet, diese Varianz über Personenmerkmale zu erklären. Zunächst bleibt allerdings festzuhalten, dass der Einfluss des Gesundheitswandels auf die Lebenszufriedenheit zwischen den Befragten signifikant variiert.

Tabelle 2.3: Modell M4.1 der Lebenszufriedenheit

Modellbezeichnung	M4.1	
Fixed Effects	Koeffizient (S.E.)	
Konstante (γ_{00})	7,163***	(,0119)
Δ Zeit (γ_{10})	-0,051***	(,0021)
Δ Gesundheit (γ_{20})	0,481***	(,0095)
Gesundheit mean (γ_{01})	0,875***	(,0097)
Geschlecht (γ_{02})	-0,099***	(,0158)
Random Effects	Varianz (S.E.)	
Residuum (Varianz ε_{ii})	1,232***	(,0097)
Konstante (Varianz u_{0i})	1,264***	(,0190)
Konstante-Zeit (Cov u_{0i}, u_{1i})	-0,043***	(,0036)
Δ Zeit (Varianz u_{1i})	0,015***	(,0010)
Konstante- Δ Gesundheit (Cov u_{0i}, u_{2i})	-0,153***	(,0128)
Δ Zeit- Δ Gesundheit (Cov u_{1i}, u_{2i})	0,010***	(,0026)
Δ Gesundheit (Varianz u_{2i})	0,332***	(,0152)
Modellmaße		
Devianz	296797,044	
AIC	296821,044	
BIC	296933,008	
Anzahl Parameter	12	

Die Koeffizienten- und Varianzschätzung erfolgte mit REML-Verfahren. Zur Berechnung der Modellgüte wurden die ML-Schätzwerte benutzt. Fallzahl: 24.869 Befragte mit 83.313 Einzelbeobachtungen. Signifikanz: n.s.= $p \geq 0.05$, * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$.

3.6 Schritt 5 – Cross-Level-Interaktionen M5.0a-M5.0b

Nachdem gezeigt werden konnte, dass der Einfluss des Wandels der Gesundheitseinschätzung auf die Lebenszufriedenheit eine signifikante Varianz aufweist, kann untersucht werden, inwieweit sich diese Varianz über Personenmerkmale erklären lässt. Deshalb soll im letzten Modellierungsschritt die Varianz des Within-Anteils der zeitveränderlichen Determinanten "Gesundheitseinschätzung" über zeitinvariante Merkmale erklärt werden. Die folgenden Modelle können somit auch als "slope as outcome"- Modelle bezeichnet werden (vgl. Snijders/Boskers 2012: 80), da vorrangig versucht wird, die Varianz der Random-Slopes über Personenmerkmale (statistisch) zu erklären. In einem ersten Modell soll zunächst die diesbezügliche Bedeutung des Prädiktors "Geschlecht" getestet werden. Es wird dazu angenommen, dass die individuelle Wichtigkeit von Gesundheitszustandsveränderungen für die Lebenszufriedenheit über die beiden Geschlechter systematisch variiert. Die nachfolgende Auflistung zeigt die dementsprechenden Modellgleichungen und die folgende Tabelle 2.4 zeigt die Ergebnisse der Modellschätzung.

M5.0a	Level 1: $Y_{ti} = \beta_{0i}$ $+ \beta_{1i} (\Delta\text{Zeit})_{ti} + \beta_{2i} (\Delta\text{Gesundheit})_{ti}$ $+ \varepsilon_{ti}$	Level 2: $\beta_{0i} = \gamma_{00} + \gamma_{01}(\text{Gesundheit_mean})_i + \gamma_{02}(\text{Geschlecht})_i + u_{0i}$ $\beta_{1i} = \gamma_{10} + u_{1i}$ $\beta_{2i} = \gamma_{20} + \gamma_{21}(\text{Geschlecht})_i + u_{2i}$
	t=Messzeitpunkt; i=Befragte/r	

Anhand der Schätzung von Modell M5.0a lässt sich erkennen, dass die Annahme über eine geschlechtsspezifische Variation des Effekts der Gesundheitsveränderung nicht zutrifft. Der t-test zeigt an, dass der Interaktionsterm zwischen $\Delta\text{Gesundheit}$ und Geschlecht nicht signifikant wird (vgl. Tabelle 2.4: M5.0a).

Auch der LRT zwischen Modell M4.1 und M5.0a bestätigt, dass durch die Berücksichtigung des Interaktionsterms keine signifikante Verbesserung der Modellgüte stattgefunden hat (Devianz-Unterschied ($df=1$) = 0,02; $p > 0.05$). Dementsprechend kann auch kein Varianzanteil des Random-Slopes von $\Delta\text{Gesundheit}$ (statistisch) erklärt werden. Im Modellvergleich bleiben sämtliche Varianz- u. Kovarianzschätzungen durchweg stabil. Somit gehen wir zunächst davon aus, dass sich die Stärke des Einflusses von Gesundheitsveränderungen auf die Lebenszufriedenheit nicht anhand des Geschlechts differenziert.

Ein weiteres "slope as outcome"-Modell kann durch Schätzung eines Interaktionseffekts zwischen den Variablen "Gesundheit_mean" und " $\Delta\text{Gesundheit}$ " getestet werden. Das neue Modell M5.0b soll untersuchen, ob die Stärke des Einflusses von Gesundheitsveränderungen vom durchschnittlichen Gesundheitszustand einer Person abhängt.⁴² So scheint es plausibel anzunehmen, dass Gesundheitsveränderungen für Personen mit einem allgemein guten Gesundheitszustand weniger zentral für die Lebenszufriedenheit sind als für Personen, welche im Schnitt ohnehin schon gesundheitlich angeschlagen sind. Die nachfolgende Auflistung zeigt die dementsprechenden Modellgleichungen und die folgende Tabelle 2.4 zeigt die Ergebnisse der Modellschätzung.

⁴² Derlei ad hoc getroffene Annahmen, so wie sie hier für die Interaktion zwischen durchschnittlichem Gesundheitszustand (oder Geschlecht) und Gesundheitsveränderungen gemacht wurden, sollten natürlich niemals Teil einer sauber durchgeführten Forschungsarbeit sein. Im vorliegenden Text werden diese Interaktionseffekte allein zu illustrativen Zwecken ohne ausreichende analytische oder theoretische Argumentation eingeführt.

M5.0b	Level 1: $Y_{ti} = \beta_{0i}$ $+ \beta_{1i} (\Delta\text{Zeit})_{ti} + \beta_{2i} (\Delta\text{Gesundheit})_{ti}$ $+ \varepsilon_{ti}$	Level 2: $\beta_{0i} = \gamma_{00} + \gamma_{01}(\text{Gesundheit_mean})_i + \gamma_{02}(\text{Geschlecht})_i + u_{0i}$ $\beta_{1i} = \gamma_{10} + u_{1i}$ $\beta_{2i} = \gamma_{20} + \gamma_{22}(\text{Gesundheit_mean})_i + u_{2i}$
	t=Messzeitpunkt; i=Befragte/r	

Die Schätzergebnisse von Modell M5.0b untermauern die angestellten Vermutungen. Der LRT zwischen Modell M4.1 und M5.0b zeigt, dass eine signifikante Modellverbesserung durch die Berücksichtigung der Interaktion zwischen der durchschnittlichen Gesundheitseinschätzung und Gesundheitszustandsveränderungen erreicht werden konnte (Devianz-Unterschied (df=1) = 200,91; $p < 0.001$). Die Schätzung von Modell M5.0b ergibt, dass sich für jeden zusätzlichen Skalenpunkt bei der durchschnittlichen Gesundheitseinschätzung der Einfluss von Gesundheitszustandsveränderungen auf die Lebenszufriedenheit um 0,18 Punkte verringert.

Ohne eine theoretisch oder zumindest analytisch abgeleitete Begründung des Wirkmechanismus, der hinter diesem Zusammenhang steht, fällt eine inhaltliche Interpretation des geschätzten Effekts schwer. Möglicherweise wird der Effekt durch einen im Hintergrund wirkenden, hier nicht spezifizierten Alterseffekt erzeugt. Elaboriertere Modelle mit weiteren Cross-Level-Interaktionen könnten darüber Aufschluss geben.

Tabelle 2.4: Modelle M5.0a und M5.0b der Lebenszufriedenheit

Modellbezeichnung	M5.0a	M5.0b	M5.0b (stand. Koeff.)
Fixed Effects			
	Koeffizient (S.E.)		
Konstante (γ_{00})	7,163*** (.0119)	7,165*** (.0119)	-
Δ Zeit (γ_{10})	-0,051*** (.0021)	-0,052*** (.0021)	-0,064
Δ Gesundheit (γ_{20})	0,482*** (.0130)	0,464*** (.0095)	0,135
Gesundheit_mean (γ_{01})	0,875*** (.0097)	0,884*** (.0097)	0,404
Geschlecht (γ_{02})	-0,099*** (.0158)	-0,099*** (.0158)	-0,056
Geschlecht* Δ Gesundheit (γ_{21})	-0,002 n.s. (.0188)	-	-
Gesundheit_mean* Δ Gesundheit (γ_{22})	-	-0,175*** (.0123)	-0,038
Random Effects			
	Varianz (S.E.)		
Residuum (Varianz ϵ_i)	1,232*** (.0097)	1,234*** (.0097)	
Konstante (Varianz u_{0i})	1,264*** (.0190)	1,263*** (.0190)	
Konstante-Zeit (Cov u_{0i}, u_{1i})	-0,044*** (.0036)	-0,043*** (.0035)	
Δ Zeit (Varianz u_{1i})	0,015*** (.0010)	0,015*** (.0010)	
Konstante- Δ Gesundheit (Cov u_{0i}, u_{2i})	-0,154*** (.0128)	-0,156*** (.0127)	
Δ Zeit- Δ Gesundheit (Cov u_{1i}, u_{2i})	0,010*** (.0027)	0,012*** (.0027)	
Δ Gesundheit (Varianz u_{2i})	0,332*** (.0152)	0,314*** (.0150)	
Modellmaße			
Devianz	296797,029	296596,130	
AIC	296823,029	296622,130	
BIC	296944,324	296743,424	
Anzahl Parameter	13	13	

Die Koeffizienten- und Varianzschätzung erfolgte mit REML-Verfahren. Zur Berechnung der Modellgüte wurden die ML-Schätzwerte benutzt. Fallzahl: 24.869 Befragte mit 83.313 Einzelbeobachtungen. Geschlecht (γ_{02}) in M5.0b (stand. Koeff.) nicht standardisiert; Signifikanz: n.s.= $p \geq 0,05$, * = $p < 0,05$, ** = $p < 0,01$, *** = $p < 0,001$.

Der Haupteffekt von Δ Gesundheit (γ_{20}) stellt unter Anwesenheit der Interaktionsterme den Einfluss des Wandels dar, wenn die durchschnittliche Gesundheitseinschätzung den Wert Null annimmt, was bei Grand-Mean-Zentrierung einer durchschnittlichen Ausprägung entspricht.

Zum Vergleich der einzelnen Effektstärken können auch standardisierte Koeffizienten berechnet werden (vgl. Tabelle 2.4, letzte Spalte). Dazu wird die RMD-MLA-Schätzung von Modell M5.0b nicht mit den zentrierten Rohwerten, sondern mit den z-transformierten Werten der unabhängigen und abhängigen Modellvariablen durchgeführt (vgl. Hox 2010: 23).⁴³ Aus den standardisierten Werten lässt sich dann beispielsweise ablesen, dass der Haupteffekt der Gesundheitsveränderungen mehr als doppelt so stark wie der Effekt des zeitlichen Wandels bzw. der Alterung ist.⁴⁴ Des Weiteren wird deutlich, dass die getrennte Spezifikation des Within- und Between-Anteils der Gesundheitsvariablen durchaus sinnvoll ist, da sich die entsprechenden Effekte deutlich voneinander unterscheiden.

⁴³ Allerdings sollten alle Dummy-Variablen (wie hier "Geschlecht") nicht standardisiert bzw. z-transformiert werden, sondern auch für eine standardisierte Modellschätzung in ihrer Rohform belassen werden.

⁴⁴ Zu einem ähnlichen Ergebnis gelangt man, wenn die (statistisch) erklärten Varianzanteile der ersten Modellebene betrachtet werden.

3.7 Modelldiagnostik

Alle hier vorgestellten Modellschätzungen beruhen auf mehreren Modellannahmen. Insbesondere wird, wie bereits weiter oben erwähnt, angenommen, dass die Fehlerterme auf Modellebene 1 " ε_{ti} " und auf Modellebene 2 " u_{0i} " jeweils:

- 1.) unabhängig voneinander sind (dass z.B. keine Autokorrelationen bestehen);
- 2.) einen Erwartungswert von Null besitzen;
- 3.) normalverteilt sind;
- 4.) keine Varianzungleichheiten (Heteroskedastizität) aufweisen.

Dabei ist zu berücksichtigen, dass diese, für Schätzungen von linearen Modellen generell geltenden Annahmen, bei der Schätzung von MLA-Modellen noch vervielfacht werden. Denn dort betreffen sie nicht nur die Residuen einer einzigen Analyseebene, sondern müssen für die Residuen von mindestens zwei Modellebenen zutreffen. Darüber hinaus müssen in der MLA auch Annahmen bzgl. des Verhältnisses zwischen den Variablen der einzelnen Modellebenen, des Verhältnisses der Random-Effekte zueinander und des Verhältnisses zwischen den Variablen der Modellebenen und den Random-Effekten getroffen werden (vgl. Raudenbush/Bryk 2010: 255).

Leider liegt eine große Schwachstelle der SPSS MIXED-Prozedur in den beschränkten Möglichkeiten der damit ermöglichten Modelldiagnostik. Denn Annahmen, die die Verteilung der L2-Residuen betreffen, können mit dieser Prozedur nicht getestet werden. Die SPSS MIXED-Prozedur erlaubt es nicht, in Modellen mit mehr als einem Random-Effekt die getrennten Residuen der zweiten Ebene auszulesen (vgl. auch West et al. 2007: 213f). Die Prozedur ermöglicht allein, die Level 1-Residuen, die vorausgesagten Werte des Gesamtmodells sowie die vorausgesagten Werte des Fixed-Modellteils auszulesen.⁴⁵ Somit können zumindest die Annahmen für die erste Modellebene überprüft werden.⁴⁶

4 Resümee

Mehrebenenmodelle sind vor allem als Hybridmodellierungen ein sehr flexibles Werkzeug zur Analyse von Paneldaten (auch bezeichnet als "repeated measure data - multiple level analysis": RMD-MLA). Mit diesen Modellen können sowohl reine Veränderungseffekte als auch Einflüsse von zeitkonstanten Merkmalen analysiert werden. Besonders vorteilhaft sind die damit verbundenen Mög-

⁴⁵ Der entsprechende Unterbefehl zum Auslesen besagter Werte in der MIXED-Prozedur lautet: /SAVE=FIXPRED PRED RESID

⁴⁶ Zu Techniken der Residuendiagnostik vgl. Urban/Mayerl 2011: 177-274.

lichkeiten zur Einbeziehung von Datenreihen mit fehlenden Messwerten und mit individuell unterschiedlich distanzierten Erhebungszeitpunkten. Auch die Möglichkeiten zur Modellschätzung mit unterschiedlich spezifizierten Residuen-Kovarianzmatrizen sind vielfältig und reichen von einfachen und sparsamen Modellspezifikationen bis hin zu äußerst komplexen Strukturen.

Mögliche Beschränkungen für die Analyse können aber durch ungenügende Software-Implementierungen entstehen. So sind die überaus umfangreichen Modellvoraussetzungen von RMD-MLA bei Verwendung der SPSS MIXED-Prozedur nur bedingt zu überprüfen. Auch muss der Benutzer von MLA-Software bei komplizierten Modellspezifikationen oder Datenlagen stets mit einer potentiellen Nicht-Konvergenz seiner Modellschätzungen rechnen.

Der vorliegende Beitrag hatte das Ziel, den praktischen Einstieg in die Verwendung von Mehrebenenanalysen zur Untersuchung von Längsschnittdaten (RMD-MLA) zu erleichtern und einen möglichen Analyseablauf unter Einsatz der SPSS MIXED-Prozedur anhand eines konkreten, empirischen Beispiels zu verdeutlichen.

Literaturverzeichnis

- Allison, P.D., 2009: Fixed effects regression models. Los Angeles: Sage.
- Andreß, H.-J./Golsch, K./Schmidt, A., 2013: Applied panel data analysis for economic and social surveys. Heidelberg, Berlin: Springer.
- Brüderl, J. 2010: Kausalanalyse mit Paneldaten. In: Wolf, C./Best, H. (Hg.): Handbuch der sozialwissenschaftlichen Datenanalyse. Wiesbaden: VS Verlag für Sozialwissenschaften, S. 963–994.
- Detle, D.E., 2005: Berufserfolg und Lebenszufriedenheit. Eine längsschnittliche Analyse der Zusammenhänge. Erlangen-Nürnberg: Dissertation.
- Diener, E. (Hg.), 2009: The science of well-being. Dordrecht: Springer.
- Glatzer, W., 1984a: Zufriedenheitsunterschiede zwischen Lebensbereichen. In: Glatzer, W./Zapf, W. (Hg.): Lebensqualität in der Bundesrepublik. Objektive Lebensbedingungen und subjektives Wohlbefinden. Frankfurt a.M.: Campus Verlag, S. 192–205.
- Glatzer, W., 1984b: Determinanten der Zufriedenheit. In: Glatzer, W./Zapf, W. (Hg.): Lebensqualität in der Bundesrepublik. Objektive Lebensbedingungen und subjektives Wohlbefinden. Frankfurt a.M.: Campus Verlag, S. 234–245.
- Haisken-DeNew, J.P./Frick, J.R., 2005: DTC. Desktop Companion to the German Socio-Economic Panel (SOEP). In: http://www.diw.de/documents/dokumentenarchiv/17/diw_01.c.38951.de/dtc.409713.pdf (zuletzt zugegriffen am 14.01.2013).
- Heck, R.H./Thomas, S.L./Tabata, L.N., 2010: Multilevel and longitudinal modeling with IBM SPSS. New York: Routledge.
- Hedeker, D./Gibbons, R.D., 2006: Longitudinal data analysis. Hoboken, N.J.: Wiley.
- Hoffman, L., 2013: Introduction to Random Effects of Time and Model Estimation, In: http://psych.unl.edu/psycrs/944/944_Lecture05_Random_Effects.pdf (zuletzt zugegriffen am 20.05.2013).
- Hoffman, L./Stawski, R.S., 2009: Persons as Contexts: Evaluating Between-Person and Within-Person Effects in Longitudinal Analysis. In: Research in Human Development, 6, S. 97–120.
- Hox, J. J., 2010: Multilevel analysis. Techniques and applications (2. Aufl.). New York, NY: Routledge.
- Klein, M./Pötschke, M., 2004: Die intra-individuelle Stabilität gesellschaftlicher Wertorientierungen. Eine Mehrebenenanalyse auf der Grundlage des Sozio-Oekonomischen Panels (SOEP). In: Kölner Zeitschrift für Soziologie und Sozialpsychologie, 3, S. 432–456.
- Raudenbush, S.W./Bryk, A.S., 2010: Hierarchical linear models. Applications and data analysis methods. Thousand Oaks, Calif.: Sage.

- Singer, J.D./Willett, J.B., 2003: Applied longitudinal data analysis. Modeling change and event occurrence. Oxford: Oxford Univ. Press.
- Snijders, T.A.B./Bosker, R.J., 2012: Multilevel analysis. An introduction to basic and advanced multilevel modeling (2. Aufl.). Los Angeles: Sage.
- Tabachnick, B.G./Fidell, L.S., 2010: Using multivariate statistics (5. Aufl.). Boston: Pearson.
- Twisk, J.W.R., 2011: Applied multilevel analysis. A practical guide (5. Aufl.). Cambridge: Cambridge University Press.
- Urban, D./Mayerl, J., 2011: Regressionsanalyse: Theorie, Technik und Anwendung (4. Aufl.). Wiesbaden: VS-Verlag.
- West, B./Welch, K.B./Galecki, A.T., 2007: Linear mixed models. A practical guide using statistical software. Boca Raton: Chapman & Hall/CRC.

Anhang

A0 SPSS Syntax zur Zentrierung von Variablen

```

* Syntax modifiziert nach
* http://www.ats.ucla.edu/stat/spss/faq/centering\_variables.htm
* (zuletzt zugegriffen am 03.04.2013)

#####Funktion Grand Mean Zentrierung.
define grand_cvars( vlist = !charend('/')
                   /suffix = !cmdend )

compute one_temp = 1.
exe.

!do !vname !in (!vlist)
!!let !nname = !concat(!vname, !suffix)
aggregate
/outfile=* mode=addvariables overwrite = yes
/break =one_temp
/y_temp=mean(!vname).

compute !nname = !vname - y_temp.
exe.

!doend
delete variables y_temp one_temp.
!enddefine.

#####Funktion Group Mean Zentrierung.

define group_cvars( group = !charend('/')
                   /vlist = !charend('/')
                   /suffix = !charend('/')
                   /suffix2 = !cmdend )

!do !vname !in (!vlist)
!!let !nname = !concat(!vname, !suffix)
!!let !nname2 = !concat(!vname, !suffix2)

aggregate
/outfile=* mode=addvariables overwrite = yes
/break =!group
/y_temp=mean(!vname).

compute !nname = !vname - y_temp.
compute !nname2 = y_temp.
exe.
!doend

delete variables y_temp.
!enddefine.

*Funktion group_cvars aufrufen - hier eingeben person-mean Zentrierung fuer within.
*Ausgabe: group mean zentrierte Variable + Variable mit Groupmean-Wert.
group_cvars group = PERSNR
/vlist = Gesundheit
/suffix = _gc
/suffix2= _mean.

*Funktion grand_cvars aufrufen - hier angeben between Werte.
*Ausgabe grand mean zentrierte Variable.
grand_cvars vlist = Gesundheit_mean
/suffix = _c.

```

A1 SPSS Syntax zur Transformation vom Wide- ins Long-Datenformat

```

VARSTOCASES
/MAKE Lebenszufriedenheit
  FROM Lebenszufriedenheit_2000 Lebenszufriedenheit_2001 Lebenszufriedenheit_2003 Lebenszufriedenheit_2006
/MAKE Gesundheit
  FROM gesundheit_2000 gesundheit_2001 gesundheit_2003 gesundheit_2006
/MAKE Flag_missing_long
  FROM flag_missing_Q flag_missing_R flag_missing_T flag_missing_W
/INDEX=Index1(4)
/KEEP=HHNR PERSNR gender GEBJAHR PSAMPLE immer_teilnahme Missing_irgendwo inStichprobe
/NULL=KEEP.

```

A2 SPSS-Syntax für Modelle M1.0 – M5.0b

*Modell M1.0 Nullmodell.

```

MIXED lebenszufriedenheit
/CRITERIA=CIN(95) MXITER(100) MXSTEP(10) SCORING(1) SINGULAR(0.00000000001) HCONVERGE(0,
  ABSOLUTE) LCONVERGE(0, ABSOLUTE) PCONVERGE(0.000001, ABSOLUTE)
/FIXED= | SSTYPE(3)
/METHOD=REML
/PRINT=G R SOLUTION TESTCOV
/RANDOM=INTERCEPT | SUBJECT(PERSNR) COVTYPE(VC).

```

MIXED lebenszufriedenheit

```

/CRITERIA=CIN(95) MXITER(100) MXSTEP(10) SCORING(1) SINGULAR(0.00000000001) HCONVERGE(0,
  ABSOLUTE) LCONVERGE(0, ABSOLUTE) PCONVERGE(0.000001, ABSOLUTE)
/FIXED= | SSTYPE(3)
/METHOD=ML
/PRINT=G R SOLUTION TESTCOV
/RANDOM=INTERCEPT | SUBJECT(PERSNR) COVTYPE(VC).

```

*Modell M2.3.1 mit prozesszeit.

```

MIXED lebenszufriedenheit with zeit_delta
/CRITERIA=CIN(95) MXITER(100) MXSTEP(10) SCORING(1) SINGULAR(0.00000000001) HCONVERGE(0,
  ABSOLUTE) LCONVERGE(0, ABSOLUTE) PCONVERGE(0.000001, ABSOLUTE)
/FIXED= zeit_delta | SSTYPE(3)
/METHOD=REML
/PRINT=G R SOLUTION TESTCOV
/RANDOM=INTERCEPT | SUBJECT(PERSNR) COVTYPE(VC).

```

MIXED lebenszufriedenheit with zeit_delta

```

/CRITERIA=CIN(95) MXITER(100) MXSTEP(10) SCORING(1) SINGULAR(0.00000000001) HCONVERGE(0,
  ABSOLUTE) LCONVERGE(0, ABSOLUTE) PCONVERGE(0.000001, ABSOLUTE)
/FIXED= zeit_delta | SSTYPE(3)
/METHOD=ML
/PRINT=G R SOLUTION TESTCOV
/RANDOM=INTERCEPT | SUBJECT(PERSNR) COVTYPE(VC).

```

*Modell M2.3.2 Gesundheitseinschätzung aufgenommen.

```

MIXED lebenszufriedenheit with zeit_delta gesundheit_gc gesundheit_mean_c
/CRITERIA=CIN(95) MXITER(100) MXSTEP(10) SCORING(1) SINGULAR(0.00000000001) HCONVERGE(0,
  ABSOLUTE) LCONVERGE(0, ABSOLUTE) PCONVERGE(0.000001, ABSOLUTE)
/FIXED=zeit_delta gesundheit_gc gesundheit_mean_c | SSTYPE(3)
/METHOD=REML
/PRINT= G R SOLUTION TESTCOV
/RANDOM=INTERCEPT | SUBJECT(PERSNR) COVTYPE(VC).

```

MIXED lebenszufriedenheit with zeit_delta gesundheit_gc gesundheit_mean_c

```

/CRITERIA=CIN(95) MXITER(100) MXSTEP(10) SCORING(1) SINGULAR(0.00000000001) HCONVERGE(0,
  ABSOLUTE) LCONVERGE(0, ABSOLUTE) PCONVERGE(0.000001, ABSOLUTE)
/FIXED=zeit_delta gesundheit_gc gesundheit_mean_c | SSTYPE(3)
/METHOD=ML
/PRINT= G R SOLUTION TESTCOV
/RANDOM=INTERCEPT | SUBJECT(PERSNR) COVTYPE(VC).

```

*Modell M3.0 geschlecht mit aufgenommen.

```

MIXED lebenszufriedenheit with zeit_delta gesundheit_gc gesundheit_mean_c gender
/CRITERIA=CIN(95) MXITER(100) MXSTEP(10) SCORING(1) SINGULAR(0.00000000001) HCONVERGE(0,
  ABSOLUTE) LCONVERGE(0, ABSOLUTE) PCONVERGE(0.000001, ABSOLUTE)
/FIXED=zeit_delta gesundheit_gc gesundheit_mean_c gender | SSTYPE(3)
/METHOD=REML
/PRINT= G R SOLUTION TESTCOV
/RANDOM=INTERCEPT | SUBJECT(PERSNR) COVTYPE(VC).

```

MIXED lebenszufriedenheit with zeit_delta gesundheit_gc gesundheit_mean_c gender
 /CRITERIA=CIN(95) MXITER(100) MXSTEP(10) SCORING(1) SINGULAR(0.00000000001) HCONVERGE(0,
 ABSOLUTE) LCONVERGE(0, ABSOLUTE) PCONVERGE(0.000001, ABSOLUTE)
 /FIXED=zeit_delta gesundheit_gc gesundheit_mean_c gender | SSTYPE(3)
 /METHOD=ML
 /PRINT= G R SOLUTION TESTCOV
 /RANDOM=INTERCEPT | SUBJECT(PERSNR) COVTYPE(VC).

*Modell M4.0 zeit random gesetzt.

MIXED lebenszufriedenheit with zeit_delta gesundheit_gc gesundheit_mean_c gender
 /CRITERIA=CIN(95) MXITER(100) MXSTEP(10) SCORING(1) SINGULAR(0.00000000001) HCONVERGE(0,
 ABSOLUTE) LCONVERGE(0, ABSOLUTE) PCONVERGE(0.000001, ABSOLUTE)
 /FIXED=zeit_delta gesundheit_gc gesundheit_mean_c gender | SSTYPE(3)
 /METHOD=REML
 /PRINT= G R SOLUTION TESTCOV
 /RANDOM=INTERCEPT zeit_delta | SUBJECT(PERSNR) COVTYPE(UN).

MIXED lebenszufriedenheit with zeit_delta gesundheit_gc gesundheit_mean_c gender
 /CRITERIA=CIN(95) MXITER(100) MXSTEP(10) SCORING(1) SINGULAR(0.00000000001) HCONVERGE(0,
 ABSOLUTE) LCONVERGE(0, ABSOLUTE) PCONVERGE(0.000001, ABSOLUTE)
 /FIXED=zeit_delta gesundheit_gc gesundheit_mean_c gender | SSTYPE(3)
 /METHOD=ML
 /PRINT= G R SOLUTION TESTCOV
 /RANDOM=INTERCEPT zeit_delta | SUBJECT(PERSNR) COVTYPE(UN).

*Modell M4.1 gesundheit random gesetzt.

MIXED lebenszufriedenheit with zeit_delta gesundheit_gc gesundheit_mean_c gender
 /CRITERIA=CIN(95) MXITER(100) MXSTEP(10) SCORING(1) SINGULAR(0.00000000001) HCONVERGE(0,
 ABSOLUTE) LCONVERGE(0, ABSOLUTE) PCONVERGE(0.000001, ABSOLUTE)
 /FIXED=zeit_delta gesundheit_gc gesundheit_mean_c gender | SSTYPE(3)
 /METHOD=REML
 /PRINT= G R SOLUTION TESTCOV
 /RANDOM=INTERCEPT zeit_delta gesundheit_gc | SUBJECT(PERSNR) COVTYPE(UN).

MIXED lebenszufriedenheit with zeit_delta gesundheit_gc gesundheit_mean_c gender
 /CRITERIA=CIN(95) MXITER(100) MXSTEP(10) SCORING(1) SINGULAR(0.00000000001) HCONVERGE(0,
 ABSOLUTE) LCONVERGE(0, ABSOLUTE) PCONVERGE(0.000001, ABSOLUTE)
 /FIXED=zeit_delta gesundheit_gc gesundheit_mean_c gender | SSTYPE(3)
 /METHOD=ML
 /PRINT= G R SOLUTION TESTCOV
 /RANDOM=INTERCEPT zeit_delta gesundheit_gc | SUBJECT(PERSNR) COVTYPE(UN).

*Modell M5.0a cross-level-interaktion gender mit gesundheit_gc.

MIXED lebenszufriedenheit with zeit_delta gesundheit_gc gesundheit_mean_c gender inter_G_G
 /CRITERIA=CIN(95) MXITER(100) MXSTEP(10) SCORING(1) SINGULAR(0.00000000001) HCONVERGE(0,
 ABSOLUTE) LCONVERGE(0, ABSOLUTE) PCONVERGE(0.000001, ABSOLUTE)
 /FIXED=zeit_delta gesundheit_gc gesundheit_mean_c gender inter_G_G | SSTYPE(3)
 /METHOD=REML
 /PRINT= G R SOLUTION TESTCOV
 /RANDOM=INTERCEPT zeit_delta gesundheit_gc | SUBJECT(PERSNR) COVTYPE(UN).

MIXED lebenszufriedenheit with zeit_delta gesundheit_gc gesundheit_mean_c gender inter_G_G
 /CRITERIA=CIN(95) MXITER(100) MXSTEP(10) SCORING(1) SINGULAR(0.00000000001) HCONVERGE(0,
 ABSOLUTE) LCONVERGE(0, ABSOLUTE) PCONVERGE(0.000001, ABSOLUTE)
 /FIXED=zeit_delta gesundheit_gc gesundheit_mean_c gender inter_G_G | SSTYPE(3)
 /METHOD=ML
 /PRINT= G R SOLUTION TESTCOV
 /RANDOM=INTERCEPT zeit_delta gesundheit_gc | SUBJECT(PERSNR) COVTYPE(UN).

*Modell M5.0b cross-level-interaktion durchschnittliche Gesundheit_mean mit gesundheit_gc.

MIXED lebenszufriedenheit with zeit_delta gesundheit_gc gesundheit_mean_c gender inter_G_G_mean
 /CRITERIA=CIN(95) MXITER(100) MXSTEP(10) SCORING(1) SINGULAR(0.00000000001) HCONVERGE(0,
 ABSOLUTE) LCONVERGE(0, ABSOLUTE) PCONVERGE(0.000001, ABSOLUTE)
 /FIXED=zeit_delta gesundheit_gc gesundheit_mean_c gender inter_G_G_mean | SSTYPE(3)
 /METHOD=REML
 /PRINT= G R SOLUTION TESTCOV
 /RANDOM=INTERCEPT zeit_delta gesundheit_gc | SUBJECT(PERSNR) COVTYPE(UN).

MIXED lebenszufriedenheit with zeit_delta gesundheit_gc gesundheit_mean_c gender inter_G_G_mean
 /CRITERIA=CIN(95) MXITER(100) MXSTEP(10) SCORING(1) SINGULAR(0.00000000001) HCONVERGE(0,
 ABSOLUTE) LCONVERGE(0, ABSOLUTE) PCONVERGE(0.000001, ABSOLUTE)
 /FIXED=zeit_delta gesundheit_gc gesundheit_mean_c gender inter_G_G_mean | SSTYPE(3)
 /METHOD=ML
 /PRINT= G R SOLUTION TESTCOV
 /RANDOM=INTERCEPT zeit_delta gesundheit_gc | SUBJECT(PERSNR) COVTYPE(UN).

*Modell M5.0b mit standardisierten Werten für standardisierte Koeffizienten, cross-level-interaktion.

DESCRIPTIVES VARIABLES=zeit_delta

gesundheit_gc gesundheit_mean_c Lebenszufriedenheit gender inter_G_G inter_G_G_mean

/Save.

MIXED Zlebenszufriedenheit with Zzeit_delta Zgesundheit_gc Zgesundheit_mean_c Zgender Zinter_G_G_mean

/CRITERIA=CIN(95) MXITER(100) MXSTEP(10) SCORING(1) SINGULAR(0.000000000001) HCONVERGE(0, ABSOLUTE) LCONVERGE(0, ABSOLUTE) PCONVERGE(0.000001, ABSOLUTE)

/FIXED=Zzeit_delta Zgesundheit_gc Zgesundheit_mean_c Zgender Zinter_G_G_mean | SSTYPE(3)

/METHOD=REML

/PRINT= G R SOLUTION TESTCOV

/RANDOM=INTERCEPT Zzeit_delta Zgesundheit_gc | SUBJECT(PERSNR) COVTYPE(UN).

MIXED Zlebenszufriedenheit with Zzeit_delta gesundheit_gc Zgesundheit_mean_c Zgender Zinter_G_G_mean

/CRITERIA=CIN(95) MXITER(100) MXSTEP(10) SCORING(1) SINGULAR(0.000000000001) HCONVERGE(0, ABSOLUTE) LCONVERGE(0, ABSOLUTE) PCONVERGE(0.000001, ABSOLUTE)

/FIXED=Zzeit_delta Zgesundheit_gc Zgesundheit_mean_c Zgender Zinter_G_G__mean | SSTYPE(3)

/METHOD=ML

/PRINT= G R SOLUTION TESTCOV

/RANDOM=INTERCEPT Zzeit_delta Zgesundheit_gc | SUBJECT(PERSNR) COVTYPE(UN).

SISS: Schriftenreihe des Instituts für Sozialwissenschaften der Universität Stuttgart
<http://www.uni-stuttgart.de/soz/institut/forschung/siss.html>

bisher sind erschienen:

- No. 1/1994 "Vertrauen" - soziologisch betrachtet. Ein Beitrag zur Analyse binärer Interaktionssysteme.
Peter Antfang, Dieter Urban
- No. 2/1994 Report on the German Machine Tool Industry.
Frank C. Englmann, Christian Heyd, Daniel Köstler, Peter Paustian
with the assistance of Susanne Baur and Peter Bergmann
- No. 3/1994 Neue württembergische Rechtstatsachen zum Unternehmens- und Gesellschaftsrecht.
Udo Kornblum
- No. 4/1994 Rechtstatsachen zum Unternehmens- und Gesellschaftsrecht aus den neuen Bundesländern.
Udo Kornblum
- No. 1/1995 Die Bedeutung Neuronaler Netze in der Ökonomie.
Hermann Schnabl
- No. 2/1995 Regionale Strukturprobleme.
Sammelband der Beiträge zum Symposium vom 13. und 14. Oktober 1994.
Frank C. Englmann (Hrsg.)
- No. 3/1995 Latent Attitude Structures Directing the Perception of New Technologies.
An Application of SEM-Methodology to the Construction of Attitude
Measurement Models Related to Technologies of Prenatal Genetic
Engineering and Testing.
Dieter Urban
- No. 4/1995 Handbuch zur empirischen Erhebung von Einstellungen/Kognitionen zur Bio- und Gentechnologie (inklusive Diskette)
(zweite, überarbeitete und erweiterte Auflage)
Uwe Pfenning, Dieter Urban, Volker Weiss
- No. 5/1995 Social Indicators in a Nonmetropolitan County: Testing the Representativeness of a Regional Nonrandom Survey in Eastern Germany.
Dieter Urban, Joachim Singelmann
- No. 1/1996 Jugend und Politik im Transformationsprozeß. Eine Fallstudie zur Stabilität und Veränderung von politischen Einstellungen bei ostdeutschen Jugendlichen zwischen 1992 und 1995.
Dieter Urban, Joachim Singelmann, Helmut Schröder
- No. 2/1996 Einstellungsmessung oder Einstellungsgenerierung? Die Bedeutung der informationellen Basis bei Befragten für die empirische Rekonstruktion von Einstellungen zu gentechnischen Anwendungen.
Martin Slaby
- No. 1/1997 Gentechnik: „Fluch oder Segen“ versus „Fluch und Segen“.
Bilanzierende und differenzierende Bewertungen der Gentechnik in der öffentlichen Meinung.
Dieter Urban und Uwe Pfenning

(Fortsetzung ...)

SISS: Schriftenreihe des Instituts für Sozialwissenschaften der Universität Stuttgart
<http://www.uni-stuttgart.de/soz/institut/forschung/siss.html>

bisher sind erschienen (Fortsetzung):

- No.2/1997 Die soziale Vererbung von Ausländer"feindlichkeit". Eine empirische Längsschnittanalyse der intra- und intergenerativen Transmission von sozialen Einstellungen.
Dieter Urban und Joachim Singelmann
- No. 3/1997 Politische Sozialisation im Transformationsprozeß: Die Entwicklung demokratiebezogener Einstellungen von ostdeutschen Jugendlichen und deren Eltern zwischen 1992 und 1996.
Barbara Schmidt, Dieter Urban, Joachim Singelmann
- No.1/1998 Bewertende Einstellungen zur Gentechnik: ihre Form, ihre Inhalte und ihre Dynamik. Kurzbericht zu Ergebnissen des Forschungsprojektes „Einstellungen zur Gentechnik“.
Dieter Urban, Uwe Pfenning, Joachim Allhoff
- No.2/1998 Technikeinstellungen: gibt es die überhaupt? Ergebnisse einer Längsschnittanalyse von Bewertungen der Gentechnik.
Dieter Urban
- No.3/1998 Zur Interaktion zwischen Befragten und Erhebungsinstrument. Eine Untersuchung zur Konstanz des Meinungsurteils von Befragten im Interviewverlauf.
Martin Slaby
- No.1/1999 Role Models and Trust in Socio-Political Institutions: A Case Study in Eastern Germany, 1992-96.
Joachim Singelmann, Toby A. Ten Ayck, Dieter Urban
- No.1/2000 Die Zufriedenheit von Stuttgarter Studierenden mit ihrer Lebens- und Wohnsituation. Erste deskriptive Ergebnisse einer sozialwissenschaftlichen Studie zu allgemeinen und bereichsspezifischen Zufriedenheiten der Studierenden des Campus Vaihingen und des Campus Hohenheim.
Projektgruppe Campus: Slaby, M.; Grund, R.; Mayerl, J.; Noak, T.; Payk, B.; Sellke, P.; Urban, D.; Zudrell, I.
- No.2/2000 Längsschnittanalysen mit latenten Wachstumskurvenmodellen in der politischen Sozialisationsforschung.
Dieter Urban
- No.1/2001 Unser „wir“ - ein systemtheoretisches Modell von Gruppenidentitäten.
Jan A. Fuhse
- No.2/2001 Differentielle Technikakzeptanz, oder: Nicht immer führt die Ablehnung einer Technik auch zur Ablehnung ihrer Anwendungen.
Eine nutzentheoretische und modell-statistische Analyse.
Martin Slaby, Dieter Urban
- No.3/2001 Religiosität und Profession. Longitudinale Analysen zur Entwicklung des religiösen Selbstbildes bei Erzieherinnen.
Heiko Lindhorst

(Fortsetzung ...)

bisher sind erschienen (Fortsetzung):

- No.4/2001 Ist Glück ein affektiver Sozialindikator subjektiven Wohlbefindens?
Dimensionen des subjektiven Wohlbefindens und die Differenz
zwischen Glück und Zufriedenheit.
Jochen Mayerl
- No.1/2002 Risikoakzeptanz als individuelle Entscheidung.
Zur Integration der Risikoanalyse in die nutzentheoretische
Entscheidungs- und Einstellungsforschung.
Martin Slaby, Dieter Urban
- No.2/2002 Vertrauen und Risikoakzeptanz. Zur Relevanz von Vertrauen
bei der Bewertung neuer Technologien.
Martin Slaby, Dieter Urban
- No.3/2002 Probleme bei der Messung individueller Veränderungsraten.
13 empirisch und methodisch induzierte Effekte, die es schwierig machen,
Veränderungen von generalisierten Bewertungen zu ermitteln.
Dieter Urban
- No.1/2003 Systeme, Netzwerke, Identitäten. Die Konstitution sozialer Grenzziehungen
am Beispiel amerikanischer Straßengangs.
Jan A. Fuhse
- No.2/2003 Können Nonattitudes durch die Messung von Antwortreaktionszeiten ermittelt werden?
Eine empirische Analyse computergestützter Telefoninterviews.
Jochen Mayerl
- No.1/2004 Erhöht ein Opfer-Täter-Zyklus das Risiko, Sexualstraftaten als pädosexuelle Straftaten zu
begehen? Ergebnisse einer ereignisanalytischen Pilotstudie
Dieter Urban, Heiko Lindhorst
- No.1/2005 Persönliche Netzwerke in der Systemtheorie
Jan A. Fuhse
- No.2/2005 Analyzing cognitive processes in CATI-Surveys with response latencies:
An empirical evaluation of the consequences of using different
baseline speed measures.
Jochen Mayerl, Piet Sellke, Dieter Urban
- No.1/2006 Ist Bildung gleich Bildung? Der Einfluss von Schulbildung auf ausländerablehnende
Einstellungen in verschiedenen Alterskohorten.
Klaus Hadwiger
- No.2/2006 Zur soziologischen Erklärung individuellen Geldspendens.
Eine Anwendung und Erweiterung der Theory of Reasoned Action unter Verwendung von
Antwortlatenzzeiten in einem Mediator-Moderator-Design.
Jochen Mayerl
- No.1/2007 Antwortlatenzzeiten in TRA-Modellen. Zur statistischen Erklärung von (Geld)-
Spendenverhalten.
Dieter Urban, Jochen Mayerl

SISS: Schriftenreihe des Instituts für Sozialwissenschaften der Universität Stuttgart
<http://www.uni-stuttgart.de/soz/institut/forschung/siss.html>

bisher sind erschienen (Fortsetzung):

- No.1/2008 Berufseintritt und Berufssituation von Soziologieabsolventen der Universität Stuttgart.
Deskriptive Ergebnisse einer Absolventenbefragung aus dem Jahr 2007.
Jochen Mayerl, Dieter Urban
- No.1/2010 Der Bystander-Effekt in alltäglichen Hilfsituationen:
Ein nicht-reaktives Feldexperiment.
Katrin Alle, Jochen Mayerl
- No.2/2010 Das Working-Poor-Problem in Deutschland.
Empirische Analysen zu den Ursachen von Armut trotz Erwerbstätigkeit.
Leonie Hellmuth, Dieter Urban
- No.3/2010 Binär-logistische Regressionsanalyse.
Grundlagen und Anwendung für Sozialwissenschaftler.
Jochen Mayerl, Dieter Urban
- No.1/2011 Pädosexuelle Viktimisierung und pädosexuelle Straffälligkeit. Ein „harter“ empirischer
Hypothesentest.
Dieter Urban, Joachim Fiebig
- No.1/2012 Meta-Analyse und "publication bias". Eine empirisch-statistische Fallstudie.
Joachim Fiebig, Gerrit Rüdibusch, Dieter Urban
- No.1/2013 Panelanalyse mit Mehrebenenmodellen. Eine anwendungsorientierte Einführung.
Thomas Krause, Dieter Urban