

methoden daten analysen

ZEITSCHRIFT FÜR EMPIRISCHE SOZIALFORSCHUNG

mda

2010, Jahrgang 4, Heft 1



*Leander Steinkopf, Gerrit Bauer
und Henning Best*

Nonresponse und Interviewer-Erfolg im
Telefoninterview.

Empirische Untersuchungen zum Einfluss
stimmlicher Eigenschaften der Interviewer

Gerhard Krug

Fehlende Daten bei der Verknüpfung von
Prozess- und Befragungsdaten.

Ein empirischer Vergleich ausgewählter
Missing Data Verfahren

Herausgegeben von

*Christof Wolf
Marek Fuchs
Bärbel Knäuper
Petra Stein*

Methoden – Daten – Analysen. Zeitschrift für Empirische Sozialforschung

Die Zeitschrift wird herausgegeben von GESIS – Leibniz-Institut für Sozialwissenschaften.

Herausgeber: Christof **Wolf** (Mannheim, geschäftsführend), Marek **Fuchs** (Kassel), Bärbel **Knäuper** (Montreal), Petra **Stein** (Duisburg-Essen)

Wissenschaftlicher

Beirat: Hans-Jürgen **Andreß** (Köln), Andreas **Diekmann** (Zürich), Sabine **Häder** (Mannheim), Udo **Kelle** (Marburg), Dagmar **Krebs** (Gießen), Frauke **Kreuter** (College Park, Maryland), Edith **de Leeuw** (Utrecht), Norbert **Schwarz** (Ann Arbor)

Redaktion: Paul **Lüttinger**

GESIS – Leibniz-Institut für Sozialwissenschaften

Postfach 12 21 55

68072 Mannheim

Tel.: 0621 – 1246-268

E-Mail: mda@gesis.org

Internet: www.gesis.org/MDA/

Die MDA deckt alle Fragestellungen aus dem Bereich der Empirischen Sozialforschung ab, insbesondere aus dem Bereich der Umfragemethodik. Im Vordergrund stehen Artikel, die die methodischen und/oder statistischen Kenntnisse der Profession erweitern, sowie Beiträge, die sich mit der Anwendung der Methoden der Empirischen Sozialforschung in der Forschungspraxis beschäftigen, oder solche, in denen ein statistisches Verfahren exemplarisch angewandt wird. Obwohl der Schwerpunkt auf Umfragemethoden liegt, sind Beiträge zu anderen methodischen Bereichen willkommen.

Alle Beiträge, die zur Veröffentlichung in der MDA eingereicht werden, werden von mindestens zwei unabhängigen Gutachtern blind begutachtet.

Der Nachdruck von Beiträgen ist nach Absprache möglich. Die MDA erscheint zweimal im Jahr und steht als Printversion und online zur Verfügung. Die Registrierung für den Bezug der MDA erfolgt über die Web-Seiten von GESIS:

<http://www.gesis.org/forschung-lehre/gesis-publikationen/zeitschriften/mda/bestellung/>

Druck: Concordia-Druckerei König oHG, Mannheim-Sandhofen

Gedruckt auf chlorfrei gebleichtem Papier.

ISSN 1864-6956

4. Jahrgang 2010 © GESIS, Mannheim, Juni 2010

Inhalt

FORSCHUNGSBERICHTE

- 3 Nonresponse und Interviewer-Erfolg im Telefoninterview.
Empirische Untersuchungen zum Einfluss stimmlicher
Eigenschaften der Interviewer
Leander Steinkopf, Gerrit Bauer und Henning Best
- 27 Fehlende Daten bei der Verknüpfung von Prozess- und
Befragungsdaten. Ein empirischer Vergleich ausgewählter
Missing Data Verfahren
Gerhard Krug
-

REZENSIONEN

- 58 Confirmatory Factor Analysis for Applied Research
Timothy A. Brown, 2006
Peter Schmidt
- 62 Sozialforschung. Methoden und Anwendungen.
Ein Überblick für die BA-Studiengänge
Uwe Flick, 2009
Jürgen Schiener
- 64 Clusteranalyse mit SPSS. Mit Faktorenanalyse
Christian FG Schendera, 2010
Michael Wiedenbeck
-

ANKÜNDIGUNGEN

- 67 Frühjahrstagung 2011 der Sektion Jugendsoziologie
„Methoden der Jugendforschung – angemessene
Antworten auf neue Herausforderungen“
TU Darmstadt, 23. - 24. März 2011 – Call for Papers
- 70 European Labour Force Survey (EU-LFS) and
European Union Statistics on Income and Living
Conditions (EU-SILC): 2nd European User Conference
Mannheim, March 31 - April 1, 2011 – Call for Papers
- 72 Hinweise für unsere Autorinnen und Autoren

Nonresponse und Interviewer-Erfolg im Telefoninterview

Empirische Untersuchungen zum Einfluss stimmlicher Eigenschaften der Interviewer

Nonresponse in CATI-Surveys

An Empirical Study on the Effects of Interviewers' Vocal Characteristics

Leander Steinkopf, Gerrit Bauer und Henning Best

Zusammenfassung

Basierend auf der Messung von objektiven Eigenschaften und subjektiven Bewertungen der Stimmen von 56 weiblichen Telefoninterviewern wird der Zusammenhang zwischen Stimmeigenschaften und dem Interviewerfolg empirisch untersucht. Theoretisch lässt sich argumentieren, dass die Interviewer-Stimmen einerseits Nutzenerwartungen der Respondenten beeinflussen können, indem sie z. B. belohnend wirken oder Seriosität vermitteln. Ausgehend von dual-process-Theorien ist außerdem zu erwarten, dass die Stimme einen situativen Hinweisreiz darstellt, der – ohne rationale Abwägung – Kooperationsnormen aktivieren kann und hierdurch die Wahrscheinlichkeit der Teilnahme am Interview erhöht. Empirisch zeigt sich, dass subjektiv eingeschätzte Stimmerkmale in keinem Zusammenhang mit der Erfolgsquote der Interviewer stehen. Objektive Merkmale der Stimme hingegen, insbesondere die Stimmhöhe, haben substanziellen Einfluss auf die Erfolgsquote. Es ist jedoch wichtig zu bemerken, dass der Zusammenhang zwischen der Stimmhöhe oder der Sprechgeschwindigkeit und dem Interviewerfolg nicht linear ist, sondern einen umgekehrt U-förmigen Verlauf hat. Wir finden empirisch, dass Interviewerinnen mit durchschnittlicher Stimmhöhe und durchschnittlicher Sprechgeschwindigkeit die höchsten Erfolgsquoten aufweisen.

Abstract

This study examines nonresponse in telephone surveys. Our analysis relates response rates to interviewer voice characteristics. Drawing on theory we argue that interviewer voices can affect the respondents' utility expectations, e.g. by indicating integrity, or by providing an intrinsic reward. Based on dual-process theories, one can additionally expect the voice to be a situative cue. We use contact data generated in a CATI survey and supplement these data with information on interviewers' voices. The survey was conducted during 2007 and 2008 at the University of Mannheim, Germany. To obtain metadata, we recorded the interviewers' voices and generated phonetic measures of vocal characteristics. This enabled us to study the determinants of interviewer effectiveness and nonresponse with a special focus on objective voice characteristics (pitch, speech rate, etc.). Additionally, we accounted for a variety of subjective voice attributes (friendliness, trustworthiness, etc.). The results show that objective vocal characteristics have greater explanatory power than subjective indicators. Furthermore, the objective voice characteristics are related to the interviewers' productivity in a nonlinear way.

1 Einleitung

Telefonische Befragungen haben sich in den Sozialwissenschaften und insbesondere der Markt- und Meinungsforschung längst zu einem Standardinstrument entwickelt. CATI-Befragungen bieten insgesamt eine Reihe von Vorteilen, etwa die im Vergleich zu persönlich-mündlichen Interviews deutlich geringeren Kosten und die einfachere Feldsteuerung. Allerdings bleiben die Ausschöpfungsquoten telefonischer Befragungen nach wie vor hinter Face-to-Face durchgeführten Studien zurück, wenngleich sich durch Telefonsurveys bessere Response-Raten erzielen lassen als durch Websurveys (vgl. z. B. Fricker 2005; Roster et al. 2004). Die Ausschöpfungsquote einer Befragung kann grundsätzlich als aggregierte Folge von Teilnahmeentscheidungen der potentiellen Befragten verstanden werden. Aus Untersuchungen zum Befragtenverhalten hat sich unterdessen ein umfangreiches Forschungsfeld entwickelt, denn Nonresponse ist – gleich welche Befragungsmethode genutzt wird – stets ein wesentliches Problem sozialwissenschaftlicher Umfrageforschung. Weil Ausfälle häufig selektiv erfolgen, können die Ausfallgründe erstens mit den die Forscher interessierenden Merkmalen korreliert sein, zweitens verringert sich die Repräsentativität einer Befragung. Somit treten mit großer Wahrscheinlichkeit Verzerrungen auf (Nonresponse bias, siehe z. B. Groves 1989; Schnell 1997). In der Praxis der Umfrageforschung besteht daher das Ziel, den Ursachen des Ausfalls ganzer Erhebungseinheiten nachzugehen und Nonresponse zu reduzieren.

Unit Nonresponse, also der vollständige Ausfall einer Erhebungseinheit (Groves et al. 2001; Schnell 1997), kann seine Ursache(n) auf verschiedenen Ebenen haben: der soziale und kulturelle Kontext, die Erhebungsorganisation und deren Auftraggeber, das Erhebungsdesign, der Interviewer und die Erhebungseinheit selbst (de Leeuw/de Heer 2001). Trotz der umfangreichen Forschungsarbeiten zum Nonresponse bleiben dennoch bestimmte, die Befragungssituation kennzeichnende Merkmale, bisher wenig beachtet: Merkmale der Stimme des Interviewers. Im deutschsprachigen Raum liegt hierzu lediglich eine Studie vor (Hüfken/Schäfer 2003), und auch international ist der Forschungsstand unbefriedigend. Dies ist überraschend, da in der Kommunikationsforschung unumstritten ist, dass im Gespräch sowohl verbale Inhalte als auch nonverbale Ausdrücke – zu denen stimmliche Merkmale üblicherweise gezählt werden – übertragen werden. Schon Goffman unterscheidet zwischen diesen beiden Informationskanälen, dem „Ausdruck, den er [der Akteur] sich selbst gibt, und der Ausdruck, den er ausstrahlt“ (Goffman 1969: 6). Unser Beitrag setzt an letzterem an und analysiert systematisch den Einfluss sowohl objektiv gemessener als auch subjektiv bewerteter Stimmmerkmale von Interviewern auf die Teilnahme an telefonischen Befragungen. Während in Face-to-Face-Surveys weitere

physische Interviewereigenschaften, etwa körperliche Attraktivität, ein gepflegtes Erscheinungsbild oder eine ansprechende Mimik und Gestik das Antwortverhalten und vor allem die Kooperationsbereitschaft des Befragten beeinflussen können, sind diese potentiellen Faktoren in Telefoninterviews ausgeschaltet. Der Befragte nimmt lediglich die Stimme des Interviewers wahr.¹

Wir werden im Folgenden zunächst die theoretischen Mechanismen beschreiben, die einen Einfluss der Stimme von Interviewern auf die Teilnahme des Respondenten nahelegen. Sodann geben wir einen Überblick über den bisherigen Forschungsstand und stellen das Design unserer Studie vor. Es folgt eine Darstellung der empirischen Ergebnisse, die wir schließlich vor dem Hintergrund der Theorien zum Befragtenverhalten diskutieren.

2 Theorie und Forschungsstand

2.1 Theorie

Eine klassische Erklärung der Teilnahme an Befragungen findet sich bei Esser (1986), der die Entscheidung als eine rational geleitete Wahlhandlung modelliert. Ein Befragter wird dann an einer Untersuchung teilnehmen, wenn sich daraus ein positiver Erwartungsnutzen ergibt. Dieser Nutzen wird von den direkten Kosten einer Befragung, den Opportunitätskosten und eventuellen Belohnungen mitbestimmt. In diesem Sinne ist etwa die Wahrscheinlichkeit, dass eine Person mit wenig Zeit an einer Befragung teilnimmt, sicherlich eher gering, da sie hohe Opportunitätskosten einkalkulieren muss (vgl. Esser 1986: 39). Für eine Person hingegen, die Sozialkontakte vermisst und Gesprächsbedarf hat, ist ein Interview demnach mit sozialem Nutzen verbunden.²

1 In telefonischen Befragungen sind selbstverständlich auch Überzeugungsstrategien der Interviewer von Bedeutung. Diese sind jedoch nicht Gegenstand des vorliegenden Beitrags.

2 Bezogen auf die Teilnahme an Befragungen hat die Rational-Choice-Theorie durch Groves und Couper (1998) sowie Groves et al. (2000) eine Erweiterung erfahren. Unter der Bezeichnung „Leverage-Saliency“-Theorie wird der Bedeutung von Überzeugungsstrategien der Interviewer Rechnung getragen. Die Theorie geht also davon aus, dass bestimmte Attribute die Kosten-Nutzen-Entscheidung umso stärker beeinflussen, desto salienter sie sind. Interviewer versuchen gleich zu Beginn der Kontaktphase, die Saliency bestimmter Merkmale der Umfrage zu erhöhen. Sie betonen dabei solche Merkmale der Umfrage, von denen sie denken, dass sie vom jeweiligen Respondenten positiv bewertet werden (etwa Thema, Incentive, Auftraggeber etc.). Wird eine möglichst vollständige Erklärung von Nonresponse angestrebt, so sind die Überlegungen von Groves et al. von großer Bedeutung. Da in der vorliegenden Studie gezielt Hypothesen zum Einfluss stimmlicher Merkmale abgeleitet und überprüft werden sollen, sind Überzeugungsstrategien hier nicht Untersuchungsgegenstand.

Doch in welchem Zusammenhang steht die rationale Entscheidung zur Interviewteilnahme mit Eigenschaften der Stimme des Interviewers? Bereits Esser (1986) merkt an, dass „die Entscheidung zur Teilnahme meist aus einer Art Indifferenz heraus erfolgt“. Speziell in einer solchen Situation seien dann Einflüsse „relativ periphere[r] und zufällige[r] Faktoren“ (Esser 1986: 39) zu erwarten. Sowohl die direkten Kosten als auch der Nutzen eines Interviews sind mit stimmlichen Eigenschaften des Interviewers verknüpft. Wir halten zwei Kosten/Nutzen-Aspekte für besonders diskussionswürdig: Erstens kann ein Gespräch mit einer Person mit sympathischer, attraktiver oder freundlicher Stimme durchaus als angenehm empfunden werden und entsprechend für den Respondenten mit Nutzen verbunden sein. Im Gegensatz hierzu ist ein Gespräch mit einer unsympathischen, schrillen Stimme mitunter unangenehm und somit kostenträchtig (vgl. Feldstein et al. 2001; Scherer 1978).

Zweitens sind subjektive Risiken der Interviewteilnahme zu beachten – insbesondere Befürchtungen hinsichtlich einer Entanonymisierung der Befragungsergebnisse. Da telefonische Befragungen in aller Regel als „cold calls“, also ohne Vorankündigung, durchgeführt werden, hat der Befragte keine oder nur geringe Informationen über den Auftraggeber einer Studie, seine Vertrauenswürdigkeit und die Einhaltung von Datenschutzbestimmungen. Die Stimme des Interviewers kann in dieser Situation Seriosität vermitteln und Kompetenz ausstrahlen, Vertrauen wecken, oder eben auch unprofessionell wirken und verunsichern. Beispielsweise berichten Apple et al. (1979), dass höhere Stimmen eher als nervös empfunden werden, langsam sprechende Personen hingegen als tendenziell weniger glaubwürdig und überzeugend. Neuere Studien deuten darauf hin, dass stimmliche Merkmale des Empfängers bei der Beurteilung des Senders als Referenz herangezogen werden und so Einschätzungen von Kompetenz und Vertrauenswürdigkeit entstehen (Feldstein et al. 2001). Folglich werden durch nonverbal vermittelte Persönlichkeitsmerkmale die erwarteten Kosten einer Interviewteilnahme beeinflusst, da sich auf Seiten des Respondenten die subjektiv eingeschätzte Wahrscheinlichkeit eines Datenmissbrauchs verändert. Insgesamt kann argumentiert werden, dass die rationale Entscheidung zur Interviewteilnahme von subjektiven Eigenschaften, die der Stimme zugeschrieben werden, mit bestimmt werden kann.

In neueren Entwicklungen der Entscheidungstheorie wird jedoch argumentiert, dass viele Entscheidungen nicht kalkuliert gefällt werden, sondern vielmehr einer Routine folgen oder spontan getroffen werden – insbesondere, wenn Kosten und Nutzen einer Handlung eher gering sind. Wichtige Anstöße für die Analyse solcher Entscheidungen geben sozialpsychologische dual-process-Theorien (siehe z. B. Chaiken/Trope 1999) und das Modell der Frame Selektion von Esser (1990) und Kroneberg (2007).

Komplexe Erwägungen von Opportunitätskosten oder Seriositätseinschätzungen sind in diesen theoretischen Modellen nur bedingt zu erwarten. Das „Elaborationswahrscheinlichkeitsmodell der Einstellungsänderung“ (ELM) von Petty und Cacioppo (1984) sagt einen starken Einfluss peripherer Merkmale – so genannter Hinweisreize oder „situational cues“ – voraus, wenn die Akteure nur ein geringes Interesse am jeweiligen Handlungsgegenstand haben. Effekte stimmlicher Merkmale wurden vor diesem theoretischen Hintergrund beispielsweise im Bereich der Werbewirkungsforschung untersucht (Gelinac-Chebat/Chebat 2001). Angewandt auf den Gegenstand des Telefoninterviews ist auch in einer solchen low-involvement Situation von einem Effekt der Sprechstimme auszugehen. Im Gegensatz zum Rational-Choice-Modell des Nonresponse wird jedoch weder eine explizite Nutzenkalkulation angenommen, noch wird erwartet, dass der Akteur die situativen Reize einer aufmerksamen Prozessierung unterzieht. In telefonischen Befragungen bedeutet dies konkret, dass sich potentielle Teilnehmer in den ersten Sekunden der Kontaktphase häufig nicht intensiv mit dem Inhalt des Einleitungstextes befassen werden. Stattdessen oder zusätzlich orientieren sie sich an Hinweisreizen. Im Telefongespräch können solche Stimuli insbesondere von Stimme und Sprache des Interviewers ausgehen (Groves et al. 1992). Diese Hinweisreize können einerseits durch subjektiv zugeschriebene Eigenschaften der Stimme (kompetent, sympathisch, etc.) gesetzt werden, wie sie bereits im Abschnitt zur Rationalität der Interviewteilnahme diskutiert wurden. Darüber hinaus ist von einem Einfluss objektiver Stimmeigenschaften auszugehen (Tonhöhe, Lautstärke, etc.). Es ist anzunehmen, dass auch das Bewerten einer Stimme nach den genannten subjektiven Zuschreibungen kognitiven Aufwand erfordert. In einer peripheren Situation wird der Akteur jedoch versuchen, seinen kognitiven Aufwand zu minimieren. Der Kooperationsstimulus würde in diesem Sinne nicht von einer explizit „sympathischen“ oder „angenehmen“ Stimme ausgehen, sondern vielmehr würde die Stimme als solche einen Hinweisreiz setzen, ohne hierbei explizit bewertet zu werden. In der sozialpsychologischen Forschung wird argumentiert, dass Stereotype über Stimmatraktivität bestehen und diese in unbewussten Prozessen als Heuristik für eine schnelle Beurteilung herangezogen werden (Bartsch 2008). Da Stereotype nicht notwendigerweise eine akkurate Beschreibung einer Person darstellen, prägen sie besonders den ersten Eindruck (Zuckerman et al. 1990). Dies bedeutet, dass auch objektive stimmliche Eigenschaften ohne eine subjektive Evaluation bzw. ohne eine bewusste Zuschreibung von Eigenschaften verhaltenswirksam werden können. In diesem Sinne wären gerade in Telefoninterviews direkte Effekte z. B. der Stimmlage, der Tonhöhe oder der Sprechgeschwindigkeit zu erwarten, da die Stimme in dieser Situation der einzige Hinweisreiz ist und zunächst keine weiteren Informationen über den Interviewer vorliegen.

2.2 Forschungsstand

Obwohl Einflussfaktoren auf die Teilnahme an Interviews vielfach untersucht wurden (z. B. kürzlich in dieser Zeitschrift von Schnauber/Daschmann 2008 oder Weidmann et al. 2008), sind zur Wirkung von Interviewerstimmen auf das Verhalten der Befragten nur wenige Studien erschienen (Groves et al. 2008; Hüfken/Schäfer 2003; Oksenberg/Cannell 1988; Oksenberg et al. 1986; Sharf/Lehman 1984; van der Vaart et al. 2005).

Die grundlegende Arbeit stammt von Oksenberg et al. (1986). Die Autoren messen objektive wie subjektive stimmliche Eigenschaften von insgesamt 12 weiblichen Interviewern mit besonders hohen bzw. niedrigen Verweigerungsraten und vergleichen diese Extremgruppen miteinander. Oksenberg et al. finden, dass sich erfolgreiche und erfolglose Interviewer stärker hinsichtlich objektiver stimmlicher Eigenschaften (insbes. Tonhöhe und Lautstärke) als bezüglich subjektiver Zuschreibungen (Attraktivität und positive Einstellung) unterscheiden. Für die Eigenschaften Sprachkompetenz, Sprechgeschwindigkeit, Betonung, Einschätzung als kompetent, das Alter und die soziale Schicht zeigten sich keine oder nur sehr geringe Effekte. Die Ergebnisse von Oksenberg et al. (1986) sollten jedoch aufgrund der sehr geringen Fallzahl und des Extremgruppenvergleichs mit Vorsicht betrachtet werden. In einer Follow-Up-Studie überprüfen Oksenberg und Cannell (1988) die Ergebnisse an einer etwas größeren Stichprobe und ergänzen den Extremgruppenvergleich um eine Korrelationsanalyse. Die Ergebnisse sind jedoch recht uneinheitlich und schwer zu interpretieren. So zeigt sich in der Korrelationsanalyse (n=25) ein negativer Einfluss der Stimmhöhe auf die Erfolgsquote sowie ein positiver Effekt von Sprechgeschwindigkeit, Lautstärke und der Einschätzung als vertrauenswürdig. Auch im Extremgruppenvergleich (getrennt durchgeführt an drei Samples von je 2*6 Interviewerinnen) führen lauterer und schnelleres Sprechen sowie eine Einschätzung der Interviewerin als „kompetent“ oder „vertrauenswürdig“ zu geringeren Verweigerungsraten. Im Gegensatz zur Korrelationsanalyse zeigt sich hier jedoch teilweise ein positiver Einfluss der Stimmhöhe. Freundlichkeit, Interessiertheit, Natürlichkeit, die Betonung³ oder die Einschätzung der Stimme als angenehm haben keinen Effekt.

Mit Effekten objektiv messbarer stimmlicher Eigenschaften, vor allem der Fundamentalfrequenz, der Intonation und der Länge von Sprachpausen befassten sich bereits Sharf und Lehman (1984). Ihre Analyse bezieht sich auf sechs weibliche

3 Die Betonung wurde in dieser Studie einerseits als subjektive Einschätzung erhoben, andererseits als Variation in der Fundamentalfrequenz gemessen. Die Ergebnisse waren unsystematisch (mal positiv, mal negativ, mal Null).

Interviewer, wobei positive Effekte einer höheren Stimme und einer größeren Variation der Stimmhöhe gefunden werden. Auch Interviewer mit einer flüssigeren Sprechweise mit kürzeren Sprachpausen hatten höhere Erfolgsquoten, und eine fallende Sprachmelodie erwies sich in diesem Kontext förderlich. Diese älteren Studien leiden unter der geringen Stichprobengröße, dem Extremgruppenvergleich (selection on the dependent variable) und damit einhergehenden Verzerrungen sowie der Nichtverwendung von multivariaten Verfahren.

Neuere Studien (Groves et al. 2008; Hüfken/Schäfer 2003; van der Vaart et al. 2005) versuchen, die Defizite teilweise auszugleichen. Hüfken und Schäfer (2003) verwenden eine Variante des Extremgruppenvergleichs. Aus ihrem Sample von 51 Interviewern wählen sie 10 mit hohen Erfolgsquoten, 10 mit niedrigen und 4 mit mittleren Quoten aus. Für die Stimmen dieser 24 Interviewer erheben sie subjektive Ratings zu verschiedenen stimmlichen und persönlichen Merkmalen für jeden einzelnen Interviewversuch, den die Interviewer durchführen. Hierdurch ergibt sich ein geclustertes Sample von $n=219$ (leider ignorieren die Autoren die Clustering bei ihren Analysen, behandeln die 219 Interviewversuche also als unabhängige Beobachtungen, was zu einer Unterschätzung der Standardfehler führt). Empirisch zeigen sich in den multivariaten Logit-Modellen ein signifikant negativer Einfluss der Sprechgeschwindigkeit sowie ein positiver Effekt „enthusiastischen Auftretens“. Der Koeffizient der Variation der Stimmhöhe ist negativ und in einzelnen Modellen signifikant. Die anderen Eigenschaften (Stimmhöhe, Lautstärke, Betonung, Wortfluss, angenehme Stimme, Interesse, Höflichkeit u. a.) haben keinen signifikanten Einfluss. Es muss jedoch beachtet werden, dass Hüfken und Schäfer Interviewer beiderlei Geschlechts in ihrer Stichprobe haben, dies in den Modellen jedoch nicht adäquat berücksichtigen (z. B. bei der Stimmhöhe). Zudem verwenden sie ausschließlich subjektive Ratings für Stimmeigenschaften, die technisch auch objektiv messbar wären.

Van der Vaart et al. (2005) untersuchen die Effekte objektiver und subjektiver Stimmeigenschaften von 31 weiblichen Interviewern. Es ist positiv anzumerken, dass die Autoren im Gegensatz zu den vorgenannten Studien ihr Sample nicht nach der Erfolgsquote selektieren, also keinen Extremgruppenvergleich durchführen. In ihren bivariaten Analysen haben die objektiven Messdaten (Stimmhöhe, Variation der Höhe, Sprechgeschwindigkeit und Pausen) keinen Einfluss auf den Interviewerfolg. Es zeigen sich jedoch positive bivariate Effekte der Lautstärke, des subjektiv eingeschätzten Redeflusses sowie der Autorität. Die Einschätzung des Interviewers als zuverlässig oder liebenswürdig hat keinen signifikanten Effekt.

In einer aktuellen Studie von Groves et al. (2008) mit ca. 60 Interviewern⁴ beiderlei Geschlechts bestätigt sich der positive Einfluss einer höheren Tonlage und einer lauterer Stimme, wenn auch die Variation im Geschlecht der Interviewer nicht adäquat berücksichtigt wird. Negativ auf die Interviewteilnahme wirken sich darüber hinaus maskuline Stimmen sowie solche mit rauchigem Klang aus. Wirkt der Einleitungstext nicht frei gesprochen sondern abgelesen, ist dies der Kooperation des Respondenten ebenfalls nicht zuträglich.

Insgesamt stellt sich der Forschungsstand eher uneinheitlich dar, und die Interpretierbarkeit der Ergebnisse einiger Studien wird durch methodische Probleme eingeschränkt. Neben der geringen Fallzahl halten wir insbesondere die systematische Selektion der Probanden für problematisch, da hierbei Effekte leicht überschätzt und durch Ausreißer verzerrt werden können. Multivariate Methoden werden tendenziell nur selten eingesetzt (sicherlich auch aufgrund der geringen Fallzahlen), so dass keine zuverlässige Aussage über die Effekte einzelner Variablen gemacht werden kann. Inhaltlich zeichnet sich dennoch ab, dass insbesondere höhere und lautere Stimmen zu verbesserten Kooperationsraten führen. Zu beachten ist, dass gerade der Einfluss der Stimmhöhe (Fundamentalfrequenz) erstens nicht unabhängig vom Geschlecht sein muss und zweitens ein nicht-linearer Verlauf des Zusammenhangs angenommen werden sollte. Beidem wurde in bisherigen Modellierungen aber nicht adäquat Rechnung getragen.

3 Datengrundlage und Methode

Die vorliegende Studie basiert auf Informationen zu 56 Interviewerinnen im Alter von 20–25 Jahren, die an der Universität Mannheim im Bachelorstudiengang Soziologie an einem Lehrforschungsprojekt teilnahmen. Wir beschränken uns hier aus praktischen Gründen auf weibliche Interviewer: Stimmliche Eigenschaften sollten unbedingt getrennt nach Geschlechtern untersucht werden, da eine Analyse der gepoolten Daten zur Verzerrungen führt. Ergebnisse der phonetischen als auch der sozialpsychologischen Forschung zeigen, dass Stimmen von Frauen und Männern nach anderen Kriterien beurteilt werden, während das Geschlecht der Rater (Empfänger) nur eine untergeordnete Rolle spielt (Addington 1968; Aronovitch 2001). Zwar wäre es wünschenswert, männliche und weibliche Interviewer miteinander vergleichen zu können. Leider war in unserem Sample die Zahl der männlichen Interviewer jedoch zu gering (s. u.), um zuverlässige Ergebnisse für Männer zu erhalten.

4 Leider ist es in dem Text von Groves et al. (2008) ausgesprochen schwierig Angaben zu finden, wie hoch die Fallzahlen der einzelnen Analysen waren.

Um den Einfluss stimmlicher Merkmale des Interviewers auf die Befragungsteilnahme zu überprüfen, müssen verschiedene Daten getrennt voneinander gemessen und miteinander verknüpft werden. So werden Angaben zur *Erfolgsquote* der Interviewer und Daten zu *objektiven Eigenschaften* und *subjektiven Bewertungen* der Interviewerstimmen benötigt.

3.1 Messung der Erfolgsquote

Die Erfolgsquoten berechnen sich aus Prozessdaten, die in zwei CATI-Befragungen anfielen. Die Befragungen wurden in den Jahren 2007 und 2008 unter kontrollierten Bedingungen im CATI-Labor der Universität Mannheim durchgeführt. Im Jahr 2007 führten 32 Interviewerinnen und 13 Interviewer eine bundesweite Telefonumfrage zum Thema „Soziale Sicherheit“ durch. Hierfür wurde eine Gabler-Häder-Stichprobe von N=4.045 Telefonnummern verwendet (siehe Gabler/Häder 1999). Abzüglich nicht vergebener Telefonnummern, Geschäftsanschlüssen und Faxnummern verblieben 1.341 zu befragende Personen, von denen 250 tatsächlich befragt werden konnten (Ausschöpfungsquote von 19 %). Die Befragung im Jahr 2008 war auf die Rhein-Neckar-Region (Vorwahl 0621) beschränkt und widmete sich dem Thema „Familie und Beruf“. Insgesamt waren 40 studentische Interviewer beteiligt (29 weiblich, 11 männlich), keiner von diesen war bereits bei der Studie im Jahr 2007 als Interviewer tätig gewesen. Die Interviewerstäbe der beiden Studien sind also vollständig disjunkte Gruppen. Auch im zweiten Erhebungsjahr wurde eine Gabler-Häder-Stichprobe verwendet (N=4.500), die zu 1.046 zu befragenden Personen führte, von denen 359 befragt werden konnten (Ausschöpfungsquote von 34 %). Diese vergleichsweise hohe Ausschöpfungsquote könnte sich zumindest teilweise aus Regionalitätseffekten und einer verbesserten Öffentlichkeitsarbeit (Artikel in der Lokalzeitung „Mannheimer Morgen“) erklären.

Die Erfolgsquote der Interviewer wurde als Quotient der individuell realisierten Interviews und der Anzahl der Telefonate, bei denen eine Kontaktperson erreicht wurde, berechnet.⁵ In der Befragung 2007 lag die mittlere Erfolgsquote

5 Eine Schwierigkeit bei der Berechnung von Erfolgsquoten stellen Termine dar. So kann man argumentieren, es sei bereits ein Erfolg, einen Termin für ein Interview abzusprechen, und andererseits könne ein Interview, das aufgrund eines Termins realisiert wird, der von einer anderen Person vereinbart wurde, nicht als Erfolg gewertet werden. Im Laufe der Befragung zeigte sich jedoch, dass die Absprache von Terminen auf eine weiche Variante der Teilnahmeverweigerung hindeuten kann. Dies gilt umso mehr, da die Interviewer geschult waren, im Falle einer „keine Zeit“-Reaktion des Befragten unbedingt zu versuchen, einen Termin abzusprechen. Da außerdem die Anrufe nach Terminabsprache den Interviewern von der CATI-Software nicht systematisch zugewiesen wurden, ist keine Verzerrung der Ergebnisse zu befürchten.

der Interviewerinnen bei 14,3 % ($s=5,4$), im Jahr 2008 bei 16,0 % ($s=7,6$). Diese Erfolgsquote ist wohlgerne nicht identisch mit der Ausschöpfungsquote, da letztere auf der Zahl der zu befragenden Personen beruht. Die individuelle Erfolgsquote basiert hingegen auf der Zahl der getätigten Anrufe mit Personenkontakt, und pro Respondent können mehrere Anrufe notwendig sein. Eine Erfolgsquote von 16 % (wie im Jahr 2008) bedeutet, dass eine durchschnittlich produktive Interviewerin 6,25 Anrufe mit Personenkontakt benötigte, um ein Interview zu realisieren.

3.2 Messung von stimmlichen Eigenschaften

Insgesamt erklärten sich 56 der 61 weiblichen Interviewer bereit, eine Stimmprobe abzugeben. Die Aufnahme der tatsächlichen Interviews war uns technisch nicht möglich⁶ und ist datenschutzrechtlich bedenklich. Daher baten wir die Interviewer im Anschluss an die Feldphase um eine Aufnahme ihrer Stimme. Die Interviewer verlasen unter kontrollierten Bedingungen den Einleitungstext der Befragung. Wie in der realen Interviewsituation sprachen sie hierfür in ein Computer-Headset folgenden Text:

Guten Tag, hier ist <NAME INTERVIEWER/IN> von der Universität Mannheim. Wir führen eine <deutschlandweite / regionale> wissenschaftliche Umfrage zum Thema <„Soziale Sicherheit“ / „Familie und Beruf“> durch. Dazu würde ich gern mit derjenigen Person in Ihrem Haushalt sprechen, die bereits 18 Jahre alt ist und als letzte Geburtstag hatte. Können Sie mir bitte sagen, auf wen das in Ihrem Haushalt zutrifft?

Die Aufnahmen wurden jeweils kurz nach der Feldphase der jeweiligen Umfrage unter gleichen technischen Bedingungen durchgeführt. Sie entstammen also nicht der tatsächlichen Interviewsituation, sondern wurden nachträglich angefertigt. So erhielten wir valide Messungen der Stimmhöhe⁷ und ähnlicher Eigenschaften, nicht aber der Lautstärke. Diese wird in der realen Interviewsituation durch eine Reihe von teilweise nicht kontrollierbaren Faktoren beeinflusst. Hierzu zählen einerseits Einstellungen des Mikrofons bzw. Headsets und des CATI-Computers (diese sind kontrollierbar), andererseits Einstellungen am Telefon des Befragten. Letztere können vom Interviewer weder wahrgenommen noch beeinflusst werden. In unseren Analysen verzichteten wir folglich auf eine Betrachtung der Lautstärke.

6 Eine unmittelbare Aufnahme der Stimme des Interviewers ist in der eingesetzten Telefonsoftware (WinCATI) nicht vorgesehen. Auch gelang es uns mit den zur Verfügung stehenden Mitteln nicht, einen Mitschnitt der Einleitungssequenz über die Siemens-Telefonsoftware „Opti Client“ zu realisieren.

7 Die individuelle Stimmhöhe eines Menschen wird vor allem durch die Länge und Dicke der Stimmbänder, durch die Größe des Kehlkopfs bestimmt (Fährmann 1982). Im Gegensatz zu Akzentuierung, Betonung und Sprechgeschwindigkeit wird die Fundamentalfrequenz daher als ein nicht (dauerhaft) verstellbares Stimmmerkmal eingestuft (Bartsch 2008). Es ist jedoch zu beachten, dass auch habituelle Spannungszustände die Stimmhöhe mit beeinflussen können (Fährmann 1982).

Die Stimmproben wurden um die Stille am Anfang und Ende der Aufnahme gekürzt und mit dem phonetischen Analyseprogramm *praat* analysiert (Boersma/Weenink 2009). Es wurden die durchschnittliche Fundamentalfrequenz der Stimme F0, ihre Variation sowie – als Indikator für die Sprechgeschwindigkeit – die Dauer der Aufnahme bestimmt (siehe Tabelle 1). Die Frequenzberechnung erfolgte unter Verwendung der Autokorrelationsmethode in einem Frequenzrahmen von 75–600 Hz, mit 100 frames/sec (vgl. Hewlett/Beck 2006; Reetz/Jongman 2009 für eine Einführung in die Phonetik). Wenn im weiteren Verlauf von „Stimmhöhe“ gesprochen wird, ist immer die hier beschriebene durchschnittliche Fundamentalfrequenz gemeint.

Tabelle 1 Gemessene stimmliche Eigenschaften
(Mittelwerte, sd in Klammern)

	2007	2008
Stimmhöhe [Hz]	230,0 (18,5)	232,1 (16,1)
Variation [sd(F0)]	74,4 (12,8)	67,7 (15,5)
Dauer [sec]	17,8 (1,5)	17,1 (1,4)
N	29	27

Die durchschnittliche Frequenz der Stimmproben unterscheidet sich nicht substantiell zwischen den Befragungen, und auch die Sprechgeschwindigkeit variiert nicht systematisch. Deutliche Unterschiede sind jedoch hinsichtlich der Variation der Stimmhöhe zu erkennen. Die Gründe für Differenzen in der Variation der Stimmhöhe sind unklar, es ist jedoch davon auszugehen, dass sie artifiziell sind. Wir gehen also davon aus, dass es zwischen den Erhebungsjahren zu einer unbeobachteten Veränderung (u. U. am Aufnahmegerät) gekommen ist. Dies beeinträchtigt die Reliabilität der Messungen je Erhebungsjahr aber nicht. Um Verzerrungen durch diese Unterschiede zu vermeiden, werden in den statistischen Analysen stets gruppenzentrierte Maße der objektiv gemessenen Stimmeigenschaften verwendet (d. h. von den Messwerten der 2007er-Interviewer wurde der 2007er-Durchschnittswert abgezogen; bei den Interviewern aus 2008 der Mittelwert dieser Befragung).

Vor der *subjektiven Bewertung* der Stimmen wurden die Aufnahmen anonymisiert, indem der Interviewername herausgeschnitten und durch einen 45 Hz Sinuston von 2,3 Sekunden Dauer ersetzt wurde. Um Einflüsse der Aufnahme-Lautstärke auf die subjektiven Stimmratings auszuschließen, wurden die Stimmproben zudem in ihrer Lautstärke angeglichen.

Für die Ratings wurde auf eine anfallende Stichprobe von 159 Nutzern der Mannheimer Universitätsbibliothek zurückgegriffen (Dezember 2008). Das Durch-

schnittsalter der Rater lag bei 25,9 Jahren ($sd=8,9$), 98 waren weiblich, 61 männlich. Die Bibliotheksnutzer wurden beim Betreten des Ausleihbereichs der Zentralbibliothek gebeten, an einem Experiment zur Einschätzung von Stimmen teilzunehmen. Als Incentive wurde eine saisonal übliche Schokoladenfigur angeboten. Die Messung der Stimmeigenschaften erfolgte an Laptops mit der Software MediaLab, wobei jeder Rater zehn zufällig ausgewählte Stimmen in zufälliger Reihenfolge bewerten sollte.⁸ Bewertet wurde anhand von 7-stufigen Differenzialskalen, bei denen nur die Endpunkte benannt waren. In Anlehnung an die Literatur (siehe den Literaturüberblick weiter oben) wurden subjektive Angaben erhoben, inwieweit die Stimme bzw. der Sprecher kompetent, vertrauenswürdig, angenehm, freundlich oder selbstsicher erscheint. Zudem wurde abgefragt, wie deutlich ein regionaler oder ausländischer Akzent erkennbar ist, wie stark die Stimme männlich oder weiblich klingt, ob die Stimme dem Rater insgesamt gefällt und für wie gutaussehend er den Sprecher einschätzt. Auch die Einschätzung des Aussehens erfolgte ausschließlich auf Basis der Stimmprobe.⁹ Zum Abschluss wurde der Rater gefragt, wie gerne er von der Person interviewt werden würde. Wir bezeichnen letzteres im Folgenden als subjektive Bereitschaft der Teilnahme.

Um Ausstrahlungseffekte der vorhergehenden Stimmsamples zu verringern, wurden zwischen den Aufnahmen einige Takte eines amerikanischen Weihnachtslieds eingeblendet („Who Comes this Night?“). Im Durchschnitt liegen 19,8 Ratings pro Stimme vor (12*19, 44*20 Ratings), die zur Bildung des Skalenwertes gemittelt wurden. Die Werte variieren zwischen 1 und 7, wobei hohe Werte eine starke Ausprägung des jeweiligen Merkmals bedeuten (siehe Tabelle 2 für Kennwerte der Messungen). Die subjektiven Eigenschaften (kompetent, vertrauenswürdig, angenehm, freundlich, selbstsicher, gut aussehend) sind empirisch hoch miteinander korreliert und laden in einer Hauptkomponentenanalyse auf einem Faktor.

8 Die Ratings liegen für männliche und weibliche Stimmen vor (1.590 Ratings). In diesem Artikel werden aufgrund der Fallzahl nur weibliche Stimmen analysiert. Folglich werden nur die 1.108 Ratings der weiblichen Stimmen genutzt.

9 Ein Foto des Interviewers lag den Ratern nicht vor, es geht hier also um die Vorstellung des Raters. Mehrere Studien haben gezeigt, dass es einen Zusammenhang zwischen stimmlichen und körperlichen Merkmalen gibt. So berichten Zuckerman und Driver (1989) Ergebnisse in Bezug auf Merkmale des Gesichts. Huges, Dispenza und Gallup (2004) demonstrieren, dass Körpermaße mit stimmlichen Eigenschaften korrelieren (Verhältnisse von Schulter- zu Hüftumfang und Verhältnis von Taillen- zu Hüftumfang). Als mögliche Ursache für diese Zusammenhänge diskutieren sie hormonelle Faktoren. Ferner werden über stimmliche Merkmale auch Alters-Schätzungen des Senders ermöglicht (Mulac/Giles 1996). Alter ist zwar sicherlich eine wichtige, durch die Stimme vermittelte Eigenschaft, den Ratern war in unserer Studie allerdings klar, dass es sich bei den Stimmproben um in etwa gleichaltrige Kommilitonen handelt.

Tabelle 2 Messergebnisse und Reliabilität

	Gesamt	Durchschnittswerte		Inter-Rater-Reliabilität	
		2007	2008	ICC	Reliabilität ^a
kompetent	5,01	5,00	5,03	0,19	0,82
vertrauenswürdig	5,12	5,11	5,14	0,15	0,78
angenehm	4,98	5,00	4,96	0,19	0,82
freundlich	5,39	5,36	5,42	0,16	0,78
selbstsicher	4,80	4,68	4,92	0,20	0,83
maskulin	1,92	1,90	1,94	0,09	0,65
gefällt insgesamt	4,69	4,68	4,71	0,20	0,83
gut aussehend	4,71	4,78	4,65	0,17	0,80
gerne interviewt	4,52	4,53	4,52	0,17	0,80
ausl. Akzent	1,82	1,73	1,91	0,72	0,98
regionaler Akzent	2,10	2,17	2,04	0,14	0,77

a Reliabilität berechnet als $\rho_k = (k \cdot ICC) / (1 + (k-1) \cdot ICC)$, siehe z. B. Winer (1971).

Die tabellierten Reliabilitätskoeffizienten der Messergebnisse liegen größtenteils im zufriedenstellenden Bereich. Betrachtet man statt des Reliabilitätskoeffizienten die Intraklassenkorrelationen, wird jedoch deutlich, dass die Durchschnittswerte zwar insgesamt als annähernd reliable Messung angesehen werden können, es aber zwischen den einzelnen Ratern durchaus unterschiedliche Einschätzungen der gleichen Stimme gibt. Entsprechend liegen die ICCs für „weiche“ Eigenschaften wie „angenehm“ oder „kompetent“ nur um 0,19, bei einem klar definierten und einfach einzuschätzenden Merkmal wie „ausländischer Akzent“ hingegen bei 0,72. Die Reliabilitätskoeffizienten und ICCs unserer Messungen decken sich mit den Ergebnissen anderer Studien (z. B. Oksenberg et al. 1986; Groves et al. 2008).¹⁰

4 Ergebnisse

Im Folgenden diskutieren wir zunächst bivariate Befunde, ehe wir multivariate Modelle zum Einfluss stimmlicher Eigenschaften auf die Erfolgsquote weiblicher Interviewerinnen vorstellen.

10 Das bedeutet letztlich, dass eine ausreichend große Anzahl von Ratings (10-20) benötigt wird, wenn subjektive Eigenschaften von Stimmen mittels Rating-Skalen reliabel gemessen werden sollen. Die Ergebnisse von Van der Vaart et al. 2005 deuten jedoch darauf hin, dass eine Magnitude-Messung zu höheren Übereinstimmungen zwischen den Ratern führt. Für eine Diskussion messtheoretischer und praktischer Probleme beim Raten von Stimmen vgl. Shrivastav und Sapienza (2005), Kreiman et al. (2006) sowie Kreiman und Gerratt (1993).

4.1 Ergebnisse bivariater Analysen

Bivariate Ergebnisse lassen zwar nur bedingt Aussagen über tatsächliche Zusammenhänge zu, allerdings sind diese Befunde am ehesten mit denen früherer Forschungsarbeiten vergleichbar. Tabelle 3 fasst erstens Korrelationen zwischen den durch (subjektive) Ratings erhobenen Stimmerkmalen und der Erfolgsquote der Interviewerinnen zusammen. Zweitens werden auch Korrelationen zwischen diesen Merkmalen und der subjektiven Bereitschaft des Raters berichtet, an einem Interview teilzunehmen.

Tabelle 3 Bivariate lineare Korrelationen der subjektiven Stimmerkmale mit den Erfolgsmaßen

	Erfolgsquote	subjektive Bereitschaft
subjektive Bereitschaft	0,14	1,00
kompetent	0,19	0,86***
vertrauenswürdig	0,15	0,91***
angenehm	0,13	0,95***
maskulin	-0,19	-0,49***
freundlich	0,13	0,78***
selbstsicher	0,18	0,80***
gefällt insgesamt	0,16	0,97***
gut aussehend	0,12	0,81***
ausländischer Akzent	-0,28**	-0,44***
regionaler Akzent	0,20	-0,23*

*N=56. * $p < 0,1$; ** $p < 0,05$; *** $p < 0,01$*

Es zeigt sich zunächst, dass die Korrelation zwischen der Erfolgsquote und der subjektiven Bereitschaft, sich von einer spezifischen Interviewerin befragen zu lassen, recht gering ist ($r=0,14$). Dies weist darauf hin, dass es sich bei dem objektiven Interviewerfolg und der subjektiven Bereitschaft offensichtlich um voneinander unabhängige Merkmale handelt, eine subjektive „Erfolgseinschätzung“ der Interviewerin also nicht zur Vorhersage ihrer tatsächlichen Performanz geeignet ist.

Sieht man von den Merkmalen eines regionalen und ausländischen Akzentes ab, so sind die (linearen) bivariaten Befunde überraschend einhellig. Alle subjektiven Indikatoren korrelieren stark mit der subjektiven Bereitschaft, wohingegen sich keinerlei signifikante Zusammenhänge zwischen den subjektiv eingeschätzten stimmlichen Merkmalen und der Erfolgsquote berichten lassen.

Als einflussreiches Merkmal, auch hinsichtlich der tatsächlichen Erfolgsquote, erweisen sich hingegen Akzente des Interviewers, und zwar insbesondere

eine ausländische Färbung. Ein ausländischer Akzent hat einen negativen Einfluss auf die Erfolgsquote ($r=-0,28$) und reduziert auch die Bereitschaft des Raters, sich von einer Frau mit einem solchen Stimmerkmal befragen zu lassen ($r=-0,44$). Ein regionaler Akzent erweist sich hingegen als der Erfolgsquote tendenziell förderlich ($r=0,20$), wenngleich regionale Sprachcharakteristika von den studentischen Ratern offenbar eher negativ bewertet werden ($r=-0,23$).

Betrachten wir nun noch die bivariaten linearen Korrelationen zwischen objektiven Stimmeigenschaften und der Erfolgsquote sowie der potentiellen Teilnahmebereitschaft des Raters. Die in Tabelle 4 dargestellten Zusammenhänge weisen insgesamt auf schwache und zumeist insignifikante lineare Effekte stimmlicher Eigenschaften hin. Lediglich die positive Korrelation zwischen der Varianz der Stimmhöhe, also der Modulation der Tonhöhe, und der Erfolgsquote ist auf dem 5-Prozent-Niveau signifikant.

Tabelle 4 Bivariate lineare Korrelationen der objektiven Stimmerkmale mit den Erfolgsmaßen

	Erfolgsquote	subjektive Bereitschaft
Stimmhöhe (gruppenzentriert)	-0,11	-0,08
Standardabweichung der Stimmhöhe (gruppenzentriert)	0,26**	0,13
Dauer (gruppenzentriert)	-0,07	-0,23

*N=56. * $p<0,1$, ** $p<0,05$, *** $p<0,01$*

Zusammenfassend lässt sich sagen, dass die subjektiven Angaben der Rater vergleichsweise stark miteinander korrelieren. Die Bereitschaft aber, von einer bestimmten Person gerne befragt werden zu wollen, ist weitgehend unabhängig von der tatsächlichen Erfolgsquote der Interviewerin. Subjektive Ratings spezifischer Stimmerkmale stehen mit der Erfolgsquote in keinem nennenswerten linearen Zusammenhang, und auch die bivariaten Ergebnisse zu objektiven Stimmeigenschaften sind unbefriedigend.

4.2 Ergebnisse multivariater Analysen

Bisherige Forschungsarbeiten zu Effekten der Stimme des Interviewers auf die Erfolgsquote von Interviewern nutzten nur selten multivariate Analysemethoden und unterstellten jeweils lineare Zusammenhänge. Diese Modellierung lässt sich jedoch nicht unmittelbar aus den oben skizzierten theoretischen Überlegungen ableiten.

Vielmehr ist es durchaus vorstellbar, dass gerade durchschnittliche Ausprägungen von Tonhöhe, deren Variation oder der Sprechgeschwindigkeit in beiden Modi der Informationsverarbeitung die Kooperationswahrscheinlichkeit des Respondenten fördern, wohingegen extreme Werte auffällig und u. U. sogar unangenehm erscheinen könnten. In den multivariaten Analysen prüfen wir daher systematisch auch nichtlineare Zusammenhänge; hierfür nehmen wir stets quadratische Terme der Prädiktoren mit in die Regressionsmodelle auf. Zudem wird in sämtlichen Modellen das Erhebungsjahr kontrolliert. Dies erscheint geboten, da die Umfragen der Jahre 2007 und 2008 in Erhebungsgebiet und Ausschöpfungsquote differieren. Unterschieden in den Ausprägungen der Prädiktoren wurde zuvor bereits durch eine gruppenspezifische Zentrierung auf die Jahres-Mittelwerte entgegengewirkt.

In Bezug auf *subjektiv* bewertete Merkmale der Stimme finden wir auch in den multivariaten Analysen keinen Zusammenhang mit der Erfolgsquote. Auch mit linearem und quadratischen Term der bewerteten Eigenschaften können wir keine Zusammenhänge zwischen der Erfolgsquote und den gerateten Charakteristika erkennen. Die Zusammenhänge sind nur von geringer Stärke und nicht einmal auf dem 10-Prozent Niveau statistisch signifikant. Die Ergebnisse der multivariaten, nichtlinearen Analysen bestätigen diesbezüglich unsere bivariaten Resultate, so dass wir die Ergebnisse zu subjektiven Merkmalen nicht weiter behandeln müssen. Der folgende Abschnitt konzentriert sich folglich auf eine multivariate Untersuchung *objektiv* gemessener stimmlicher Aspekte.¹¹

Tabelle 5 zeigt sukzessiv aufgebaute Regressionsmodelle zum Einfluss der Stimmhöhe, ihrer Variation sowie der Sprechgeschwindigkeit (Dauer in Sekunden, die für das Lesen des Einleitungstextes benötigt wurde) auf die Erfolgsquote – stets unter Kontrolle des Erhebungsjahres. Signifikante Unterschiede zwischen den Erfolgsquoten der beiden Erhebungsjahre bestehen in den multivariaten Modellen nicht. Die Koeffizienten des Erhebungsjahres verweisen aber auf die tendenziell leicht höhere Erfolgsquote im Jahr der regional durchgeführten Erhebung (2008).

Die Tonhöhe, die in den bivariaten Analysen praktisch keinen Effekt auf die Erfolgsquote hatte, erweist sich nach der Aufnahme eines quadratischen Terms als statistisch hoch signifikant und substanziell bedeutsam (Modell 1). Mit steigender Stimmhöhe steigt zunächst die Erfolgsquote an. Bei einer Tonhöhe um 225 Hertz, und damit nahe am Stichprobenmittelwert von 231 Hertz, erzielen weibliche Interviewerinnen demnach die höchste Erfolgsquote, die bei einer weiteren Zunahme der Grundfrequenz aber wieder sinkt. Die von Modell 1 vorhergesagten Erfolgs-

11 Bei der vergleichsweise geringen Fallzahl ist eine simultane Modellierung subjektiv gerateter und objektiv gemessener Einflussgrößen leider kaum möglich.

quoten sind in Abbildung 1 veranschaulicht.¹² Der nichtlineare Zusammenhang ist hier klar zu erkennen und fittet die Datenpunkte vergleichsweise gut.¹³ Der Anteil der erklärten Varianz ($R^2=0,15$) ist vollständig auf die unterschiedlichen Stimmhöhen der Interviewerinnen zurückführbar.

Tabelle 5 Determinanten der Erfolgsquote in telefonischen Interviews (Lineare Regression, unstandardisierte Koeffizienten, z-Werte in Klammern)

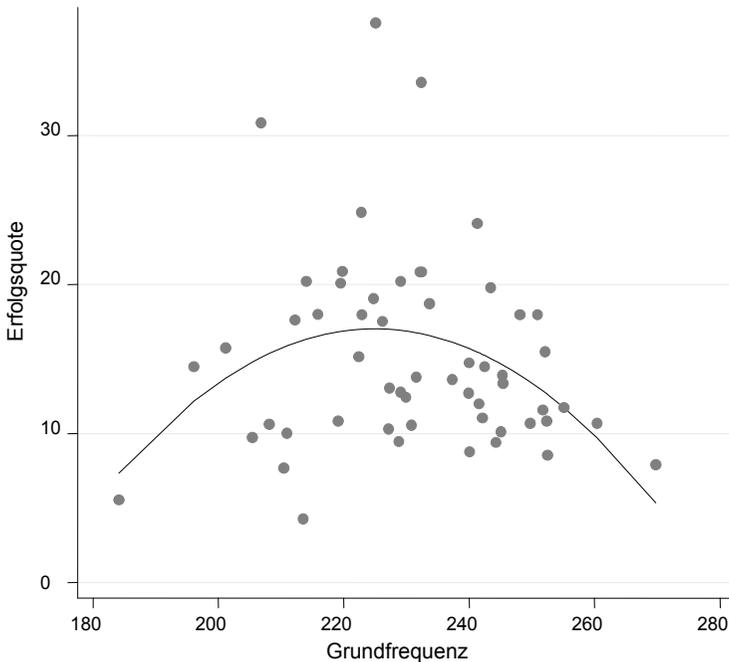
	Modell 1	Modell 2	Modell 3	Modell 4	Modell 5
<i>Stimmhöhe</i>					
linear	-0,71 (-1,43)			-0,55 (-1,11)	-0,49 (-0,99)
quadriert	-0,59*** (-2,76)			-0,56** (-2,64)	-0,45** (-2,02)
<i>Variation der Stimmhöhe</i>					
linear		1,44** (2,07)		1,26* (1,86)	1,04 (1,51)
quadriert		-0,22 (-0,64)		-0,21 (-0,64)	-0,11 (-0,32)
<i>Dauer</i>					
linear			0,14 (0,22)		-0,02 (-0,04)
quadriert			-0,73** (-2,48)		-0,43 (-1,36)
Jahr (Ref: 2007)	1,32 (0,79)	1,94 (1,12)	1,62 (0,97)	1,50 (0,91)	1,41 (0,85)
Konstante	16,19*** (12,04)	14,61*** (11,19)	15,77*** (11,99)	16,42*** (11,50)	16,81*** (11,58)
R^2	0,15	0,10	0,13	0,21	0,24
adjustiertes R^2	0,11	0,04	0,08	0,13	0,13
N	56	56	56	56	56

* $p<0,1$, ** $p<0,05$, *** $p<0,01$

Erfolgsquote in Prozent. Sowohl Stimmhöhe als auch deren Variation wurden aus Gründen der besseren Lesbarkeit vor der Berechnung der Schätzer durch 10 dividiert.

- 12 Die drei im Scatterplot auffälligen Beobachtungen mit den höchsten Erfolgsquoten verfügen zwar alle über eine in etwa mittlere Stimmlage. Tests auf Anfälligkeit des Modells gegenüber Extremwerten haben aber ergeben, dass diese und einzelne andere Datenpunkte die Regressionskoeffizienten und damit den Verlauf der Kurve nicht verzerren.
- 13 Eine nichtparametrische Lowess-Modellierung führt im Wesentlichen zu dem gleichen Kurvenverlauf.

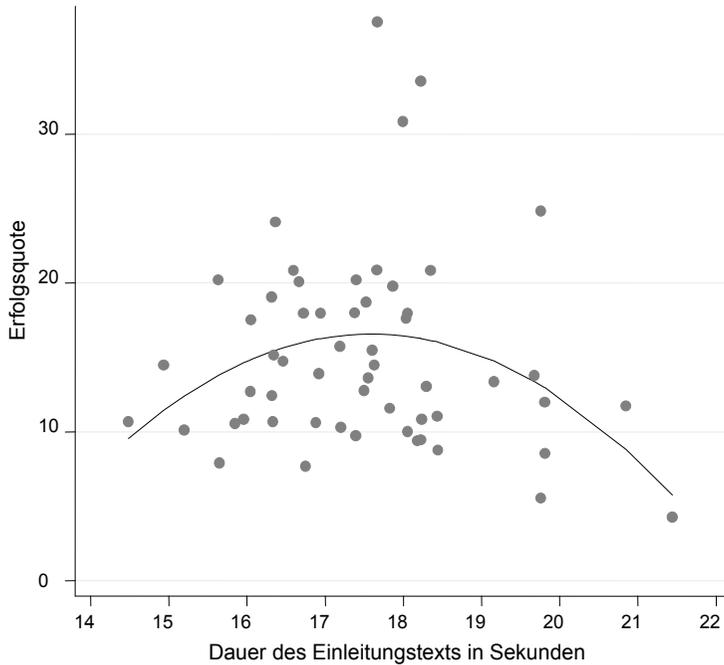
Abbildung 1 Effekte der Stimmhöhe auf die Erfolgsquote



Die Variation der Stimmhöhe (Modell 2) hatte sich bereits in den bivariaten Analysen als einflussreich erwiesen, was sich in den multivariaten Analysen zunächst auch bestätigt. Grundsätzlich wird hier ein positiver Effekt der Variation der Stimmhöhe (Modulation) errechnet, der allerdings ab einem Wert um 90 Hertz leicht abflacht und darüber hinaus keine weitere Erhöhung der Erfolgsquote mehr bedingt. Eine zu eintönige, schlecht betonte Sprechweise geht daher mit einer niedrigen Erfolgsquote einher. Die Erfolgsquote verbessert sich zunächst mit der Betonung, steigt aber bei Überbetonung nicht weiter an.

Der Effekt der Sprechgeschwindigkeit (Modell 3 und Abbildung 2) ähnelt dagegen wieder dem der Stimmhöhe: Sowohl zu langsames als auch zu schnelles Ableesen des Einleitungstextes führt zu niedrigen Erfolgsquoten. Das Maximum mit über 16,5 % wird bei einer für das Lesen benötigten Dauer von 17,5 Sekunden erreicht, also bei durchschnittlicher Sprechgeschwindigkeit. Bei einer um zwei Standardabweichungen erhöhten beziehungsweise reduzierten Dauer von 20,3 und 14,7 Sekunden liegt die vorhergesagte Erfolgsquote dahingegen nur noch bei etwa 10 %.

Abbildung 2 Effekte der Sprechgeschwindigkeit auf die Erfolgsquote



In Modell 4 wird nun simultan der Effekt der durchschnittlichen Tonhöhe während des Sprechens des Einleitungstextes und die dabei gezeigte Variation der Stimmhöhe untersucht. Es ist in diesem wie auch dem folgenden Modell jedoch zu beachten, dass – in Relation zur Zahl der verwendeten Parameter – die Zahl der Datenpunkte recht gering ist, so dass die Power der Modelle zurück geht. Entsprechend schwächen sich lineare und quadratische Effekte im Vergleich zu den Effekten in den Modellen 1 und 2 in Stärke und Signifikanz leicht ab, bleiben aber substantziell unverändert. Wird zusätzlich die Sprechgeschwindigkeit berücksichtigt (Modell 5), so bleiben signifikante Effekte lediglich der Stimmhöhe bestehen; auch die Effekte der Variation derselben sowie der Sprechgeschwindigkeit werden zumindest in ihrer Richtung nicht substantziell beeinflusst. Insgesamt erklärt dieses Modell nun ca. 24 % der Varianz der Erfolgsquote der Interviewerinnen. Dieser Anteil ist, wie Analysen ohne Berücksichtigung des Erhebungsjahres gezeigt haben, zu einem großen Teil auf objektiv gemessene stimmliche Charakteristika der Interviewerinnen zurückzuführen.

5 Diskussion und Fazit

Basierend auf Daten von 56 weiblichen Interviewern wurde in diesem Beitrag der Einfluss von stimmlichen Eigenschaften auf die Erfolgsquote in telefonischen Interviews untersucht. Hierdurch lassen sich Erkenntnisse gewinnen, die dazu dienen können, Unit-Nonresponse zu verringern und in der Folge die Datenqualität in telefonischen Befragungen zu verbessern.

Theoretisch gingen wir von einem erweiterten entscheidungstheoretischen Modell aus. Wir haben argumentiert, dass sich die Wahrscheinlichkeit eines Respondenten, am Telefoninterview teilzunehmen, erhöhen sollte, wenn die Interviewer eine Stimme haben, die als angenehm empfunden wird. In einer kalkulierten Entscheidung, wie sie in der Rational-Choice-Theorie angenommen wird, ergibt sich dies durch eine Veränderung der Kosten-Nutzen-Relation (Esser 1986; Schnell 1997). Eine angenehme Stimme, die sich durch bestimmte Merkmale auszeichnet (Apple et al. 1979; Feldstein et al. 2001; Scherer 1978), kann entweder den subjektiv empfundenen Nutzen an der Befragungsteilnahme erhöhen oder die erwarteten Kosten reduzieren, was letztlich gleichermaßen zu einer Erhöhung des Nettonutzens führt. In einem peripheren Informationsverarbeitungsmodus hingegen, wie er in dual-process-Theorien konzeptionalisiert wird (Chaiken/Trope 1999; Esser 1990; Kroneberg 2007; Petty/Cacioppo 1984), bleibt diese kognitiv anspruchsvolle Evaluation zwar aus, aber stimmliche Eigenschaften der Interviewerin können als Hinweisreize (cues) Wirksamkeit erlangen. Der entscheidende Unterschied zu einem reinen RC-Modus ergibt sich daraus, dass es sich um unbewusste Prozesse handelt, die keine aktive Evaluation und Bewertung stimmlicher Merkmale erfordern (Bartsch 2008; Zuckerman et al. 1990).

Empirisch zeigt sich, dass stimmliche Eigenschaften der Interviewer insgesamt einen nicht unerheblichen Einfluss auf ihre Erfolgsquote haben. Die Ergebnisse müssen jedoch differenziert betrachtet werden. So finden wir, dass *subjektive Ratings* der Stimmen (z. B. die Bewertung der Stimme als freundlich, angenehm o. ä.) nicht relevant für den Interviewerfolg sind. Vielmehr sind es die *objektiv gemessenen* Merkmale der Stimme, die den Erfolg als Interviewerin wirkungsvoll prognostizieren. Allerdings, und das ist ein wichtiges Ergebnis dieser Studie, ist der Zusammenhang zwischen z. B. der Tonhöhe oder der Sprechgeschwindigkeit mit den Erfolgsquoten nicht linear, sondern umgekehrt u-förmig. Solch umgekehrt u-förmige Kurvenverläufe haben sich, auch unter Kontrolle anderer Variablen, sowohl für die Stimmhöhe als auch für die Sprechgeschwindigkeit beobachten lassen. Die Maxima der vorhergesagten Erfolgsquote liegen in beiden Fällen sehr nah am Stichprobenmittelwert der beiden erklärenden Faktoren. Inhaltlich bedeutet dies, dass gerade

die *Durchschnittlichkeit* einer Stimme zu hohen Erfolgsquoten führt. Je normaler eine (weibliche) Stimme ist, desto positiver wirkt sie sich im telefonischen Interview aus, und weder schnelles noch langsames Sprechen führt zu irgendwelchen Vorteilen. Im Gegenteil ist es auch hier der mittlere Bereich, der die höchsten Erfolgsquoten aufweist. Offen und damit weiterhin erklärungsbedürftig bleibt, warum gerade die Durchschnittlichkeit kooperatives Verhalten stimuliert. Denkbar wäre, dass eine solche Stimme mit einem Attraktivitäts-Stereotyp besetzt ist. Alternativ wird in der phonetischen Forschung diskutiert, ob durch soziale Vergleichsprozesse gerade solche Stimmen überzeugen, die der des Empfängers ähnlich sind (Feldstein et al. 2001; Festinger 1954). Gegenüber der Gesamtheit der Respondenten ergäbe sich daraus ein Vorteil für Interviewer mit durchschnittlichen Stimmeigenschaften. Die Tatsache schließlich, dass wir keinen Zusammenhang zwischen den subjektiven Bewertungen der Stimme und der beobachteten Erfolgsquote, wohl aber zwischen der Erfolgsquote und objektiven Stimmeigenschaften finden, deutet darauf hin, dass der Einfluss der Stimme auf die Interviewteilnahme kein bewusster Prozess ist. Für den Fall einer Entscheidung, die auf rational-kalkulierten Nutzenerwartungen basiert, hätten wir einen stärkeren Effekt der Bewertungen erwartet. Vielmehr spricht unser Befund dafür, dass stimmliche Merkmale ihre Wirkung mitunter unbewusst über periphere Informationsverarbeitungsprozesse entfalten.¹⁴

Unsere Ergebnisse erklären auch, zumindest teilweise, den recht inkonsistenten Forschungsstand (z. B. Oksenberg/Cannell 1988; Sharf/Lehman 1984; van der Vaart et al. 2005). Offenbar ist eine rein lineare Spezifikation des Modells, wie sie in den bivariaten Korrelationsstudien unterstellt wird, den tatsächlichen Zusammenhängen nicht angemessen. Vielmehr ist die Möglichkeit nicht-linearer Beziehungen unbedingt zu beachten, was durch Befunde zu Zusammenhängen zwischen Stimme und Persönlichkeitsmerkmalen bereits nahegelegt wurde (Apple et al. 1979; Feldstein et al. 2001). Ausgehend von unseren Ergebnissen erscheint *weitere Forschung* bezüglich des Zusammenhangs von stimmlichen Eigenschaften und Interviewerfolg vielversprechend. Eine wichtige Erweiterung wäre es, die Basis der Untersuchung auszuweiten, also größere Fallzahlen zu verwenden und es dadurch zu ermöglichen, auch männliche Interviewer zu untersuchen. Unsere Arbeit zeigt zudem, dass es wichtig ist, die tatsächliche Erfolgsquote der Interviewer –

14 In Ergänzung zu dieser inhaltlichen Schlussfolgerung sollte jedoch auch die Möglichkeit eines Artefakts durch unzureichende Messung der *subjektiven* stimmlichen Eigenschaften durch Ratingverfahren in Betracht gezogen werden. Es lässt sich nicht mit Sicherheit ausschließen, dass mittels einer Messung mit Magnitude-Skalen (anstatt der in dieser Untersuchung genutzten 7-stufigen Differenzialskalen), andere Ergebnisse erzielt werden könnten. Diese Einschränkung gilt nicht für *objektive* stimmliche Eigenschaften.

und nicht die subjektive Teilnahmebereitschaft – zu untersuchen, da letztere keinen Zusammenhang mit der tatsächlichen Erfolgsquote zeigt, also nicht valide ist. Unerlässlich ist schließlich die getrennte Analyse männlicher und weiblicher Interviewer, die Verwendung multivariater Verfahren und vor allen die Berücksichtigung nichtlinearer Zusammenhänge.

Zwar sind die hier präsentierten Analysen sicherlich nicht abschließend – allein schon, da sie sich ausschließlich auf weibliche Interviewer beziehen –, dennoch lassen sich auch Ansatzpunkte für die Praxis der Umfrageforschung finden. Unsere Ergebnisse legen zunächst einmal nahe, dass in Befragungssituationen Eigenschaften des Interviewers, die über seine fachliche Kompetenz hinausgehen, eine wichtige Rolle spielen. Allein die stimmlichen Merkmale erklärten in unseren Regressionsmodellen mehr als 20 % der Varianz der Erfolgsquote. Allerdings sollte es bei der Rekrutierung von Interviewerinnen nicht darum gehen, Personen mit ganz besonderen stimmlichen Merkmalen zu selektieren. Im Gegenteil, gerade solche mit normalen, durchschnittlichen Stimmen scheinen die geringsten Verweigerungsraten zu erzielen und sind für telefonische Befragungen daher besonders geeignet.

Literatur

- Addington, D. W., 1968: The Relationship of Selected Vocal Characteristics to Personality Perception. *Speech Monographs* 35: 492-508.
- Apple, W., L. A. Streeter und R. M. Krauss, 1979: Effects of Pitch and Speech Rate on Personal Attributions. *Journal of Personality and Social Psychology* 37: 715-727.
- Aronovitch, C. D., 2001: The Voice of Personality: Stereotyped Judgments and their Relation to Voice Quality and Sex of Speaker. *The Journal of Social Psychology* 99: 207-220.
- Bartsch, S., 2008: „What Sounds Beautiful Is Good?“ How Employee Vocal Attractiveness Affects Customer's Evaluation of the Voice-to-Voice Service Encounter. S. 45-68 in: B. Strauss (Hg.): *Aktuelle Forschungsfragen im Dienstleistungsmarketing*. Wiesbaden: Gabler.
- Boersma, P. und D. Weenink, 2009: Praat, Version 5.0.46, Software zur Sprachanalyse. www.praat.org.
- Chaiken, S. und Y. Trope (Hg.), 1999: *Dual-Process Theories in Social Psychology*. New York: Guilford Press.
- de Leeuw, E. D. und W. de Heer, 2001: Trends in Household Survey Nonresponse: A Longitudinal and International Comparison. S. 41-54 in: R. M. Groves, D. A. Dillman, J. L. Eltinge und R. J. A. Little (Hg.): *Survey Nonresponse*. New York: Wiley.
- Esser, H., 1986: Über die Teilnahme an Befragungen. *ZUMA-Nachrichten* 18: 38-47. www.gesis.org/fileadmin/upload/forschung/publikationen/zeitschriften/zuma_nachrichten/zn_18.pdf (3.3.2010).
- Esser, H., 1990: „Habits“, „Frames“ und „Rational Choice“: Die Reichweite von Theorien der rationalen Wahl (am Beispiel der Erklärung des Befragtenverhaltens). *Zeitschrift für Soziologie* 19: 231-247.
- Fährmann, R., 1982: Elemente der Stimme und Sprechweise. S. 138-164 in: K. R. Scherer (Hg.): *Vokale Kommunikation*. Weinheim, Basel: Belz.

- Feldstein, S., F.-A. Dohm und C. L. Crown, 2001: Gender and Speech Rate in the Perception of Competence and Social Attractiveness. *The Journal of Social Psychology* 141: 785-806.
- Festinger, L., 1950: A Theory of Social Comparison Processes. *Human Relations* 7: 117-140.
- Fricker, S., M. Galesic, R. Tourangeau und T. Yan, 2005: An Experimental Comparison of Web and Telephone Surveys. *Public Opinion Quarterly* 69: 370-392.
- Gabler, S. und S. Häder, 1999: Erfahrungen beim Aufbau eines Auswahlrahmens für Telefonstichproben in Deutschland. *ZUMA-Nachrichten* 23: 45-61. www.gesis.org/fileadmin/upload/forschung/publikationen/zeitschriften/zuma_nachrichten/zn_44.pdf (3.3.2010).
- Gelinas-Chebat, C. und J.-C. Chebat, 2001: Effects of Two Voice Characteristics on the Attitudes Towards Advertising Messages. *The Journal of Social Psychology* 132: 447-459.
- Goffman, E., 1969 [1959]: *Wir alle spielen Theater*. München: Piper.
- Groves, R. M., 1989: *Survey Errors and Survey Costs*. New York: Wiley.
- Groves, R. M., R. B. Cialdini und M. C. Couper, 1992: Understanding the Decision to Participate in a Survey. *Public Opinion Quarterly* 56: 475-495.
- Groves, R. M. und M. P. Couper, 1998: *Nonresponse in Household Interview Surveys*. New York: Wiley.
- Groves, R. M., D. A. Dillman, J. L. Eltinge und R. J. A. Little, 2001: *Survey Nonresponse*. New York: Wiley.
- Groves, R. M., B. O'Hare, D. Gould-Smith, J. Benki und P. Maher, 2008: Telephone Interviewer Voice Characteristics and the Survey Participation Decision. S. 385-400 in: J. M. Lepkowski, C. Tucker, J. M. Brick, E. D. de Leeuw, L. Japac, P. J. Lavrakas, M. W. Link und R. L. Sangster (Hg.): *Advances in Telephone Survey Methodology*. New York: John Wiley & Sons, Inc.
- Groves, R. M., E. Singer und A. Corning, 2000: Leverage-Salience Theory of Survey Participation. Description and an Illustration. *Public Opinion Quarterly* 64: 299-308.
- Hewlett, N. und J. M. Beck, 2006: *An Introduction to the Science of Phonetics*. Philadelphia: Larence Erlbaum Associates.
- Hüfken, V. und A. Schäfer, 2003: Zum Einfluss stimmlicher Merkmale und Überzeugungsstrategien der Interviewer auf die Teilnahme in Telefonumfragen. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 55: 321-339.
- Huges, S. M., F. Dispenza und G. G. Gallup Jr., 2004: Ratings of Voice Attractiveness Predict Sexual Behavior and Body Configuration. *Evolution and Human Behavior* 25: 295-304.
- Kreiman, J. und B. R. Gerratt, 1993: Perceptual evaluation of voice quality: Review, tutorial, and a framework for future research. *Journal of Speech and Hearing Research* 36: 21-40.
- Kreiman, J., B. R. Gerratt und M. Ito, 2006: When and Why Listeners Disagree in Voice Quality Assessment Tasks. *Journal of the Acoustical Society of America* 122: 2354-2364.
- Kroneberg, C., 2007: Wertrationalität und das Modell der Frame-Selektion. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 59: 215-239.
- Mulac, A. und H. Giles, 1996: „You're Only As Old As You Sound“: Perceived Vocal Age and Social Meanings. *Health Communications* 8: 199-215.
- Oksenberg, L. und C. F. Cannell, 1988: Effects of Interviewer Vocal Characteristics on Nonresponse. S. 257-269 in: R. M. Groves, P. P. Biemer, L. E. Lyberg, J. T. Massey, W. L. Nicholls und J. Waksberg (Hg.): *Telephone Survey Methodology*. New York: Wiley.
- Oksenberg, L., Coleman, L. und C. F. Cannell, 1986: Interviewers' Voices and Refusal Rates in Telephone Surveys. *Public Opinion Quarterly* 50: 97-111.
- Petty, R. E. und J. T. Cacioppo, 1984: The Effects of Involvement on Responses to Argument Quantity and Quality – Central and Peripheral Routes to Persuasion. *Journal of Personality and Social Psychology* 46: 69-81.
- Reetz, H. und A. Jongman, 2009: *Phonetics: Transcription, Production, Acoustics and Perception*. Chichester: Wiley-Blackwell.
- Roster, C. A., R. D. Rogers, G. Albaum und D. Klein, 2004: A Comparison of Response Characteristics from Web and Telephone Surveys. *International Journal of Market Research* 46: 359-373.

- Scherer, K. R., 1978: Personality Inference From Voice Quality: The Loud Voice of Extroversion. *European Journal of Social Psychology* 8: 467-487.
- Schnauber, A. und G. Daschmann, 2008: States oder Traits? Was beeinflusst die Teilnahmebereitschaft an telefonischen Befragungen? *Methoden – Daten – Analysen. Zeitschrift für Empirische Sozialforschung* 2: 97-123. www.gesis.org/fileadmin/upload/forschung/publikationen/zeitschriften/mda/Vol.2_Heft_2/03_Schnauber.pdf (3.3.2010).
- Schnell, R., 1997: Nonresponse in Bevölkerungsumfragen. Opladen: Leske und Budrich.
- Sharf, D. J. und M. E. Lehman, 1984: Relationship Between the Speech Characteristics and Effectiveness of Telephone Interviewers. *Journal of Phonetics* 12: 219-228.
- Shrivastav, R., C. M. Sapienza und V. Nandur, 2005: Application of Psychometric Theory to the Measurement of Voice Quality Using Rating Scales. *Journal of Speech, Language, and Hearing Research* 48: 323-335.
- van der Vaart, W., Y. Ongena, A. Hoogendoorn und W. Dijkstra, 2005: Do Interviewers' Voice Characteristics Influence Cooperation Rates in Telephone Surveys? *International Journal of Public Opinion Research* 18: 488-499.
- Weidmann, C., P. Schmich und S. Schiller-Born, 2008: Der Einfluss von Kontrollüberzeugungen der Interviewer auf die Teilnahme an telefonischen Befragungen, *Methoden – Daten – Analysen. Zeitschrift für Empirische Sozialforschung* 2: 125-147. www.gesis.org/fileadmin/upload/forschung/publikationen/zeitschriften/mda/Vol.2_Heft_2/04_Schmich.pdf (3.3.2010).
- Winer, B. J., 1971: *Statistical Principles in Experimental Design*. New York: McGraw-Hill.
- M. Zuckerman und R. E. Driver, 1989: What Sounds Beautiful is Good: The Vocal Attractiveness Stereotype. *Journal of Nonverbal Behavior* 13: 67-82.
- Zuckerman, M., H. Hodgins und K. Miyake, 1990: The Vocal Attractiveness Stereotype: Replication and Elaboration. *Journal of Nonverbal Behavior* 14: 97-12.

Anschrift der Autoren

Leander Steinkopf, B. A.
Ostender Straße 49a
13353 Berlin

Gerrit Bauer, Dipl. Soz.
Universität Mannheim
Mannheimer Zentrum für Europäische
Sozialforschung (MZES)
68131 Mannheim
gerrit.bauer@mzes-uni-mannheim.de

PD Dr. Henning Best
Universität Mannheim
Fakultät für Sozialwissenschaften
68131 Mannheim
best@uni-mannheim.de

Fehlende Daten bei der Verknüpfung von Prozess- und Befragungsdaten

Ein empirischer Vergleich ausgewählter Missing Data Verfahren

Missing Data due to Record Linkage of Register and Survey Information

An Empirical Comparison of Selected Missing Data Techniques

Gerhard Krug

Zusammenfassung

Die Verknüpfung von Prozess- und Befragungsdaten gewinnt in der empirischen Sozialforschung zunehmend an Bedeutung. Aus Datenschutzgründen können Befragte die Verknüpfung aber ablehnen, weshalb die verbleibende Stichprobe selektiv sein kann. Hier können Missing Data Techniken helfen, eventuelle Selektionsverzerrungen in empirischen Analysen zu korrigieren. Dieses Papier nutzt eine Befragung, in der unter anderem die Zustimmung zur Verknüpfung erbeten wurde, um den Erfolg ausgewählter Missing Data Techniken bei der Ausfallkorrektur im Rahmen einer Fallstudie zu vergleichen. Bei nicht zustimmenden Befragten werden ihre faktisch gegebenen Antworten auf „fehlend“ gesetzt, um so pseudo-fehlende Werte auf Basis eines empirischen (im Vergleich zu einem statistisch simulierten) Ausfallmechanismus zu erzeugen. Eine KQ-Regressionsanalyse wird durchgeführt und eventuelle Verzerrungen durch den Datenausfall werden jeweils alternativ durch fallweisen Ausschluss von Beobachtungen mit fehlenden Werten, eine Multiple Imputation

Abstract

Linking register to survey data is becoming more and more important for empirical social science. Due to reasons of data protection the respondents have been asked for their permission to link their data. The resulting sample can therefore be selective. Missing data techniques can be used to correct for any record linkage bias. In this paper I use a survey where participants were asked permission for combining the survey with administrative data (record linkage). Based upon this survey the performance of different missing data techniques is compared. For those who refuse their permission I set their survey answers to missing, creating pseudo-missing data following an empirical relevant but unknown mechanism (rather than a statistical simulation of a missing data process). OLS Regression is performed using casewise deletion, multiple imputation and two versions of Heckman's sample selection model, respectively, to correct for the pseudo-missing data. The results are compared to a regression that is based on the complete data set and that gives us the

(Ergänzung) der fehlenden Werte und durch Selektionskorrektur nach Heckman korrigiert. Die Ergebnisse der Korrekturverfahren werden mit Regressionsanalysen auf Basis der vollständigen Daten verglichen, welche die „wahren“ Regressionskoeffizienten liefern. Es zeigt sich in einer Beispielanalyse mit *geringer* Selektivität des Datenausfalles, dass hier alle Korrekturverfahren ähnlich gut abschneiden. In einer zweiten Analyse mit *starker* Selektivität lieferte ausschließlich die Multiple Imputation gute Ergebnisse, jedoch nur, wenn die abhängige Variable keine fehlenden Werte aufwies.

“true” regression parameters. In an empirical example analysis characterized by weak selectivity of the missing data, all missing data techniques performed quite well. In a second example analysis with strong selectivity, it was only multiple imputation that was able to correct for the record linkage bias, given that missing values were present only in one or more independent variables. In the case of strong selectivity and missing values in the dependent variable, none of the missing data techniques eliminated the bias.

1 Einleitung¹

Standardisierte Befragungen sind ein zentrales Element der empirischen Sozialforschung. Durch die Gründung von Forschungsdatenzentren und die Aufbereitung administrativer Daten zu Scientific Use Files rücken aber auch sogenannte prozessproduzierte Daten in den Blick der Forschung (Wirth/Müller 2004; Allmendinger/Kohlmann 2005). Angesichts der Tatsache, dass die Vorteile beider Datenquellen zum Teil auf unterschiedlichen Gebieten liegen (vgl. Hartmann/Krug 2009), bietet es sich an, die Aussagekraft empirischer Analysen durch ihre Kombination zu erweitern. Eine Möglichkeit besteht in der Datenverknüpfung (Record Linkage), bei der auf der individuellen Ebene Befragungsdaten mit prozessproduzierten Informationen zum selben Individuum angereichert werden.² In vielen Fällen gilt allerdings, dass vor einer solchen Verknüpfung die Erlaubnis der betroffenen Personen einzuholen ist. Obwohl erfahrungsgemäß die Zustimmungsbereitschaft der Befragten relativ hoch ist (z. B. Hartmann et al. 2008: 57), wird diese natürlich auch von einem Teil der Befragten verweigert. Diese Personen weisen dann in dem angereicherten Datensatz bei den entsprechenden Variablen fehlende Werte auf.

1 Für wertvolle Hinweise und Verbesserungsvorschläge zu früheren Versionen des Textes danke ich Jörg Drechsler, Hans Kiesl, André Pahnke, Martin Spieß, Gesine Stephan sowie den Herausgebern und zwei anonymen Gutachtern der MDA. Für das Korrekturlesen des Textes danke ich Katrin Drasch und Christiane Spies. Verbliebene Fehler liegen in meiner Verantwortung.

2 Vom Record Linkage ist das statistische Matching (Rässler 2002) zu unterscheiden, bei dem Informationen von aus statistischer Sicht möglichst ähnlichen Individuen miteinander verknüpft werden.

Für die empirische Forschung mit solchen Daten stellt sich damit die Frage des Umgangs mit den fehlenden Werten. Die einfachste Lösung besteht darin, für Analysen alle Fälle mit fehlenden Werten auszuschließen (fallweiser Ausschluss). Dies setzt jedoch einen zufälligen Ausfall der nicht verwendeten Beobachtungen voraus. Ist dies nicht der Fall, werden Schätzungen etwa von Mittelwerten oder Regressionskoeffizienten verzerrt sein. Statistische Verfahren, wie die Multiple Imputation (Ergänzung) fehlender Werte oder eine Selektionskorrektur nach Heckman, versprechen hier Abhilfe. Dabei gehen sie von bestimmten Annahmen über den Datenausfallprozess aus, so dass bei Nichterfüllung dieser Annahmen das Ziel unverzerrter Schätzungen aber eventuell verfehlt wird.

Da diese Annahmen im konkreten Anwendungsfall meist nicht testbar sind, ist die Entscheidung für das eine oder andere Verfahren zum Umgang mit fehlenden Werten (sogenannte Missing Data Verfahren) oft schwierig. Die vorliegende Arbeit prüft im Rahmen einer Fallstudie einige ausgewählte Verfahren und zeigt auf, inwiefern sie bei der Korrektur von Stichprobenausfällen zu unterschiedlichen Ergebnissen führen. Aufgrund des gewählten Analysedesigns kann an ausgewählten Beispielen nicht nur nachvollzogen werden, ob die Verfahren in der Forschungspraxis zum selben, sondern auch zum *richtigen* Ergebnis gelangen. Es wird hierzu eine Befragung genutzt, bei der die Zustimmung zur Verknüpfung mit administrativen Daten abgefragt wurde. Solche empirischen Vergleiche sind etwa geeignet, um die Robustheit empirischer Forschungsergebnisse im Hinblick auf das gewählte Verfahren unter realistischen Anwendungsbedingungen zu analysieren (vgl. z. B. Ridder 1992). Im Unterschied zu vollständig simulierten Daten ermöglicht der echte Datensatz in Kombination mit dem gewählten Analysedesign, das Fehlen von Prozessdaten aufgrund tatsächlicher empirischer Teilnahmeentscheidungen von Befragten nachzuahmen. So liefern die folgenden Analysen im Gegensatz zu Simulationen einen Hinweis, ob die Missing Data Verfahren geeignet sind, den *empirisch* aufgrund bestehender Datenschutzregelungen auftretenden Datenausfall bei der Verknüpfung von Prozess- und Befragungsdaten auszugleichen.

Die vorliegende Arbeit ist hierzu wie folgt aufgebaut. Zunächst werden in Abschnitt 2 Annahmen zu verschiedenen Ausfallmechanismen und mit diesen korrespondierende Verfahren zum Umgang mit fehlenden Werten vorgestellt. Im Anschluss (Abschnitt 3) wird der empirische Vergleich durchgeführt, wobei zunächst die Datenbasis vorgestellt sowie die empirischen Determinanten des Datenausfalles bestimmt werden (3.1). Danach wird das Design des Vergleichs vorgestellt (3.2), die Implementation der Missing Data Verfahren besprochen (3.3) und schließlich werden die Ergebnisse präsentiert (3.4). In Abschnitt 4 erfolgt eine Zusammenfassung und Diskussion der Ergebnisse und Abschnitt 5 schließt mit Schlussfolgerungen aus dem empirischen Vergleich der Verfahren.

2 Ausfallmechanismen und Missing Data Verfahren

Die Zusammenspielung von Prozess- und Befragungsdaten kann unterschiedlichen Zwecken dienen. Im Folgenden wird davon ausgegangen, dass mit den verknüpften Prozess- und Befragungsdaten eine Regressionsanalyse durchgeführt werden soll, wobei mit Y_i die abhängige Variable und mit \mathbf{X}_i der Vektor der $j = 1, 2, \dots, J$ unabhängigen Variablen bezeichnet wird. Von Interesse seien die Koeffizienten der Regression $Y_i = \mathbf{X}_i \boldsymbol{\beta} + \varepsilon_i$ (zur übersichtlicheren Darstellung wird im Folgenden auf den Personenindex i verzichtet), welche im Folgenden als *Analysegleichung* oder *Analysemodell* bezeichnet wird.

Grundsätzlich lassen sich hinsichtlich des Datenausfalls im Allgemeinen und so auch im Fall der Datenverknüpfung drei unterschiedliche Situationen unterscheiden (Rubin 1987; Little/Rubin 1987; Collins/Schafer/Kam 2001): missing at random (MAR), missing completely at random (MCAR) und missing not at random (MNAR). Diese Situationen unterscheiden sich danach, welche Annahmen über die Beziehung zwischen den in der konkreten inhaltlichen Analyse relevanten Variablen und den Determinanten des Ausfallprozesses gerechtfertigt sind (vgl. zum Folgenden Horton/Lipsitz 2001; Horton/Kleinman 2007): Für jede befragte Person kann \mathbf{X} in zwei Komponenten unterteilt werden: \mathbf{X}^{obs} bezeichnet die Variablen ohne fehlende Werte (die beobachtete Komponente) und \mathbf{X}^{mis} die Variablen mit fehlenden Werten (die fehlende Komponente), entsprechendes gilt für Y^{obs} und Y^{mis} . Sei \mathbf{R} ein Vektor von $j = 1, 2, \dots, J$ Indikatorvariablen, die für jede x-Variable angeben, ob der entsprechende Wert fehlt ($R_j = 1$, falls das j-te Element von \mathbf{X} fehlt, $R_j = 0$ sonst) und sei ϕ der Parametervektor, der den Ausfallprozess (ein Element wird beobachtet oder nicht) kennzeichnet. $P(\mathbf{R} | Y, \mathbf{X})$ sei die Wahrscheinlichkeit für das Auftreten eines bestimmten Ausfallmusters.

Nach Little und Rubin (1987) ist missing completely at random (MCAR) definiert als

$$P(\mathbf{R} | Y, \mathbf{X}) = P(\mathbf{R} | Y^{obs}, Y^{mis}, \mathbf{X}^{obs}, \mathbf{X}^{mis}) = P(\mathbf{R} | \phi) \quad (1)$$

wobei ϕ und β (zu schätzende Parameter) als distinkt angenommen werden. Weniger technisch ausgedrückt bedeutet MCAR: „the process generating missing values bears no statistical relationship (e.g. correlations) with our variables of interest“ (Collins/Schafer/Kam 2001: 333). Diese Annahme erscheint jedoch gerade dann, wenn der Ausfall auf Entscheidungen der befragten Individuen beruht, als problematisch. Die missing at random (MAR) Annahme ist dagegen weniger restriktiv und lautet:

$$P(\mathbf{R} | Y, \mathbf{X}) = P(\mathbf{R} | Y^{obs}, \mathbf{X}^{obs}, \phi) \quad (2)$$

Die MAR-Annahme besagt damit, dass der Ausfallprozess zwar mit den Variablen der interessierenden Analyse zusammenhängt, diese Beziehung aber vollständig von den beobachteten Daten erfasst wird. Graham (2009: 553) spricht daher auch von „conditionally missing at random“. Im Fall MCAR wie auch MAR wird der Ausfallprozess (missing data mechanism) als ignorierbar (ignorable) bezeichnet. Dagegen besagt die MNAR-Annahme, dass der Ausfallprozess $P(\mathbf{R} | Y, \mathbf{X})$ nicht weiter vereinfacht werden kann, da er auch auf unbeobachteten Daten beruht; er ist nicht ignorierbar (nonignorable) (Little/Rubin 1987):

$$P(\mathbf{R} | Y, \mathbf{X}) \neq P(\mathbf{R} | Y^{obs}, \mathbf{X}^{obs}, \phi) \quad (3)$$

Es gibt eine Reihe von Verfahren, mit fehlenden Daten umzugehen (Missing Data Verfahren). Diese lassen sich danach unterscheiden, welche der drei genannten Annahmen zum Datenausfall mindestens erfüllt sein muss, damit die Verfahren anwendbar sind.

2.1 Fallweiser Ausschluss: Missing completely at random (MCAR)

Der fallweise Ausschluss ist meist eine sehr naheliegende und unkomplizierte Möglichkeit, mit fehlenden Daten umzugehen. Hier werden für die Analyse nur diejenigen Fälle verwendet, für die alle Variablen beobachtete Werte aufweisen. Im vorliegenden Fall wären das also ausschließlich diejenigen Personen, welche der Verknüpfung ihrer Daten zugestimmt haben. Dabei wird allerdings – meist implizit – davon ausgegangen, dass die Teilstichprobe der Zustimmenden eine einfache Zufallsstichprobe aus allen Befragten darstellt und damit beim Datenausfall die missing completely at random – Annahme gerechtfertigt ist. Ist diese Annahme tatsächlich erfüllt, können auf Basis der verfügbaren Fälle unverzerrte Schätzungen vorgenommen werden, wenn auch die Schätzung wegen geringerer Fallzahl an Effizienz verliert. Ist dies nicht der Fall, führt etwa eine Regressionsanalyse zu verzerrten Parameterschätzungen.

2.2 Multiple Imputation: Missing at random (MAR)

Unter Multipler Imputation (MI; vgl. Rubin 1976, 1987; Weins 2006) versteht man ein Verfahren, bei dem die fehlenden Werte in den Daten mit $m > 1$ plausiblen Werten ersetzt werden, wodurch ebenso viele Datensätze entstehen. Dabei ist es

grundsätzlich irrelevant, ob zu den ausfallbehafteten Variablen auch die abhängige Variable Y gehört. Daher können im Folgenden Y und \mathbf{X} zu \mathbf{Z} zusammengefasst werden, wobei $\mathbf{Z}^{mis} = (Y^{mis}, \mathbf{X}^{mis})$ und $\mathbf{Z}^{obs} = (Y^{obs}, \mathbf{X}^{obs})$.

Das Verfahren der Multiplen Imputation setzt voraus, dass die Beziehung zwischen Datenausfall und der ausfallbehafteten Variable bzw. den ausfallbehafteten Variablen vollständig von beobachteten Daten abhängt (MAR). Das Vorgehen bei der Multiplen Imputation kann in drei Teilschritte zerlegt werden: Imputation, Datenanalyse und Kombination der Ergebnisse.

Im *Imputationsschritt* werden zunächst mit $m > 1$ mehrere plausible Werte für die fehlenden Werte erzeugt. Die MAR-Annahme garantiert dabei, dass $(\mathbf{z}^{\{1\}}, \mathbf{z}^{\{2\}}, \dots, \mathbf{z}^{\{m\}})$ ergänzte Datensätze aus der Verteilung $f(\mathbf{Z}^{mis} | \mathbf{Z}^{obs})$ erzeugt werden können, da nach der Konditionierung auf \mathbf{Z}^{obs} der Datenausfall – im Bezug auf die betrachteten Variablen – zufällig erfolgt.³ Unabhängig davon wie die Imputationen konkret erzeugt werden, erfolgt im nächsten Schritt die *Datenanalyse* stets in den m generierten ergänzten Datensätzen, wobei Standardverfahren der statistischen Analyse verwendet werden können, z. B. Regressionsanalysen. Im *Kombinationsschritt* werden dann die Ergebnisse der m separaten Analysen aus den verschiedenen imputierten Datensätzen gemäß einfacher Kombinationsregeln (Rubin 1987) miteinander verknüpft. Die MI-Schätzung des Regressionsparameters β erfolgt etwa als einfacher Mittelwert aus den Regressionsparametern β_m , die man aus den $m = 1, \dots, M$ imputierten Datensätzen erhält: Es sei $m = 1, \dots, M$ die Zahl der Imputationen, dann ist der MI-Schätzer für den Regressionskoeffizienten

β der einfache Durchschnitt über alle M Imputationen:
$$\bar{\beta}_M = \frac{1}{M} \sum_{m=1}^M \hat{\beta}_m.$$

Zur Bestimmung der Standardfehler wird zunächst die Varianz innerhalb der

imputierten Datensätze (within imputation) $\bar{W}_M = \frac{1}{M} \sum_{m=1}^M W_m$ berechnet, mit

$W_m = \text{Var}(\hat{\beta}_m)$ in der m -ten Imputation, sowie die Varianz zwischen den Datensätzen (between imputations) $B_M = \frac{1}{M-1} \sum_{m=1}^M (\hat{\beta}_m - \bar{\beta}_M)^2$. Der Schätzer für die Ge-

3 Es existiert eine Vielzahl von Varianten zur Erzeugung der Imputationen: z. B. Propensity Score Methoden, Predictive Mean Matching, Diskriminanzanalysen oder logistische Regressionen (vgl. Horton/Kleinman 2007). Bei komplexeren Ausfallmustern bieten sich meist Markov Chain Monte Carlo (MCMC) Methoden an. Hier wird eine Markov-Kette erzeugt, um Ziehungen aus der sogenannten Posteriorverteilung $f(\mathbf{Z}^{mis} | \mathbf{Z}^{obs})$ zu simulieren. Eine Markov-Kette ist eine Sequenz von Zufallsvariablen, in der die Verteilung eines jeden Elementes vom Wert des vorherigen abhängt. Bei der MCMC-Methode wird eine Kette erzeugt, die lang genug ist, damit die Elemente zu einer stabilen (stationären) Verteilung konvergieren, hier $f(\mathbf{Z}^{mis} | \mathbf{Z}^{obs})$. Die Implementation der MCMC-Methode kann etwa über den IP-Algorithmus erfolgen (Schafer 1999a).

samtvarianz kombiniert beide Werte auf folgende Weise: $V_M = \overline{W}_M + \frac{M+1}{M} B_M$, wobei $\sqrt{V_M}$ schließlich der Schätzer für die Standardfehler der Regressionskoeffizienten ist.

2.3 Heckman-Korrektur: Missing not at random (MNAR)

Kann man nicht davon ausgehen, dass alle relevanten Einflüsse auf die Zustimmung in den beobachteten Daten erfasst sind, ist eine Selektionskorrektur nach Heckman (zum Folgenden Heckman 1979; Engelhardt 1999) eine mögliche Alternative zur Multiplen Imputation. Der typische Fall ist, dass lediglich in der abhängigen Variable Y fehlende Werte auftreten, während die Kontrollvariablen vollständig beobachtet werden. Das Verfahren ist aber grundsätzlich auch bei Ausfällen in der abhängigen und/oder in mehreren unabhängigen Variablen anwendbar.

Aus der MNAR-Annahme ergibt sich, dass eine Kleinste-Quadrate-Schätzung der Analysegleichung (wiederum wird im Folgenden der Personenindex i zur besseren Lesbarkeit weggelassen)

$$Y = \mathbf{X}\boldsymbol{\beta} + \varepsilon \quad (4)$$

zu verzerrten Parameterschätzungen führt. Um dies zu vermeiden, wird eine zweite Gleichung formuliert, die den Selektionsprozess beschreibt, soweit dieser durch die vorhandenen Daten abzubilden ist.

$$D^* = \mathbf{C}\boldsymbol{\alpha} + \nu \quad (5)$$

Dabei ist D^* eine latente Variable, etwa die latente Bereitschaft, der Datenverknüpfung zuzustimmen. Übersteigt die latente Variable einen bestimmten Wert (z. B. 0), dann stimmt eine Person dem Zusammenspielen zu und sonst nicht:

$$D = \begin{cases} 1 & \text{falls } D^* > 0, \\ 0 & \text{sonst} \end{cases}$$

Demnach wird Y nur für solche Personen beobachtet, für die $\nu > -\mathbf{C}\boldsymbol{\alpha}$ ist, weshalb der Erwartungswert von Y in der Teilpopulation nicht $\mathbf{X}\boldsymbol{\beta}$ ist, sondern $E(Y | D = 1, \mathbf{X}) = \mathbf{X}\boldsymbol{\beta} + E(\varepsilon | D = 1) = \mathbf{X}\boldsymbol{\beta} + E(\varepsilon | \nu > -\mathbf{C}\boldsymbol{\alpha})$.

Heckman (1979) zeigt, dass die Schätzung des Erwartungswertes von Y für die gesamte Population mit dem Problem beim Fehlen einer Variablen vergleichbar ist und ähnliche Konsequenzen für die unverzerrte Schätzung der Regressionspara-

meter β hat. Diese „fehlende Variable“ ist hier $E(\varepsilon | v > -\mathbf{C}\alpha)$. Unter der Annahme, dass die Störgrößen in den Gleichungen (4) und (5) einerseits bivariat normalverteilt sind $(\varepsilon, v) \sim N(0, 0, \sigma_\varepsilon^2, \sigma_v^2, \rho_{\varepsilon v})$, mit σ für die Varianz der jeweiligen Störgrößen und $\rho_{\varepsilon v}$ als ihr Korrelationskoeffizient, und andererseits unabhängig von \mathbf{X} und \mathbf{C} , lässt sich diese „fehlende Variable“ approximieren durch $E(\varepsilon | v > -\mathbf{C}\alpha) = \rho_{\varepsilon v} \sigma_\varepsilon \lambda$ (vgl. Heckman 1979). Die Variable $\lambda = \frac{\phi(\mathbf{C}\hat{\alpha})}{\Phi(\mathbf{C}\hat{\alpha})}$ wird meist als inverse Mills Ratio bezeichnet.

Im sogenannten Two-Step-Verfahren, bei dem die Selektionskorrektur in eine Kleinste-Quadrate-Schätzung einsetzt wird, wird λ in einem ersten Schritt aus der vorhandenen Stichprobe mit $\frac{\phi(\mathbf{C}\hat{\alpha})}{\Phi(\mathbf{C}\hat{\alpha})}$ geschätzt. Die Werte $\mathbf{C}\hat{\alpha}$ sind aus einer Probitregression der Selektionsgleichung (5) zu schätzen und Φ bzw. ϕ bezeichnen die (kumulative) Standardnormalverteilung.

Im zweiten Schritt wird λ in die Analysegleichung eingesetzt. Eine KQ-Schätzung der resultierenden Regressionsgleichung $Y = \mathbf{X}\beta + \rho_{\varepsilon v} \sigma_\varepsilon \hat{\lambda} + \varepsilon$ liefert dann eine unverzerrte Schätzung der Regressionsparameter β , wobei $\rho_{\varepsilon v} \sigma_\varepsilon$ der Koeffizient der Variable $\hat{\lambda}$ ist.

Um Kollinearitätsprobleme zu vermeiden sollte dabei \mathbf{C} mindestens eine Instrumentvariable enthalten, also ein Element das zwar in der Selektionsgleichung signifikant ist, nicht jedoch in der Analysegleichung für Y und daher nicht auch in \mathbf{X} enthalten ist (exclusion restriction, vgl. Puhani 2000).

Anstatt des Einsetzens in die KQ-Schätzung kann die Selektionskorrektur allerdings auch durch eine simultane Schätzung beider Gleichungen als Maximum Likelihood (ML)-Schätzung erfolgen, im Folgenden auch als Selektionskorrektur in der ML-Variante bezeichnet. Diese gilt jedoch als noch weniger robust gegenüber Verletzungen der Verfahrensannahmen als die KQ-Variante, wenn auch gesicherte Erkenntnisse hierzu kaum vorliegen (vgl. Winship/Mare 1992). Allerdings ist eine ML-Schätzung für den Fall, dass die abhängige Variable der Analysegleichung eine binäre Variable ist, die einzig möglich Variante der Selektionskorrektur nach Heckman.

3 Empirischer Vergleich

Im Folgenden wird zunächst die Datenbasis, der Ausfallprozess und das Untersuchungsdesign zum Vergleich der Missing Data Verfahren vorgestellt. Dabei wird auf konkrete Forschungsfragen Bezug genommen, die jedoch selbst inhaltlich nicht

von Interesse sind, sondern nur dem Verfahrensvergleich dienen. Danach werden kurz die konkreten Varianten der verwendeten Missing Data Verfahren vorgestellt und schließlich die Ergebnisse des Vergleichs präsentiert.

3.1 Datenbasis und empirischer Ausfallprozess

Die Datenbasis für den empirischen Vergleich bildet eine Befragung zur Kombilohnförderung „Mainzer Modell“. Im Rahmen der Evaluation dieser zunächst regional begrenzten, später bundesweit eingesetzten Kombilohnförderung, wurden von TNS Infratest Sozialforschung im Auftrag des Instituts für Arbeitsmarkt- und Berufsforschung (IAB) der Bundesagentur für Arbeit (BA) Geförderte und eine Gruppe ungeförderter Vergleichspersonen befragt. Die Stichprobe umfasste Personen, die im Zeitraum Januar 2001 bis März 2003 von Arbeitslosigkeit in Beschäftigung übergingen (für weitere Details siehe Hartman 2004). Alle Befragten wurden um die Erlaubnis gebeten, ihre Befragungsdaten mit Prozessdaten verknüpfen zu dürfen. Insgesamt stimmten 74,4 % diesem Anliegen zu. Damit ist die Zustimmungquote, verglichen mit anderen Erhebungen im Bereich der Arbeitsmarktforschung, relativ niedrig.

Grundsätzlich stand der Zugang zur Förderung nach dem Mainzer Modell allen Personen offen, die im Inland eine Beschäftigung aufnehmen dürfen. Daher war die Befragung nicht auf eine bestimmte Teilgruppe beschränkt. Allerdings handelt es sich bei den Geförderten zum großen Teil um Personen im Niedriglohnbereich. Im Anschluss an vorherige Analysen (Kaltenborn et al. 2005; Krug 2009) erfolgt hier zusätzlich eine Beschränkung auf die zwei Gruppen: die Kombilohnbezieher, welche keine abgeschlossene Berufsausbildung aufweisen oder vorher langzeitarbeitslos (mindestens seit einem Jahr ununterbrochen arbeitslos gemeldet) waren und die Vergleichspersonen, für die dies ebenfalls zutrifft.

Für die Umsetzung der Korrekturverfahren (siehe Abschnitt 3.3) ist es relevant, welche Faktoren die Zustimmung zur Zusammenspielung und damit den potentiellen Ausfall von Prozessdaten beeinflussen. Da eine Auseinandersetzung mit dem hier behandelten Ausfallmechanismus bereits an anderer Stelle erfolgte, soll dies hier nur kurz geschehen.⁴ Hartmann und Krug (2009) unterschieden in Anlehnung an theoretische Überlegungen zur Teilnahmeverweigerung bei Befragungen zwischen zwei potentiell wirksamen Einflussfaktoren: zum einen allgemeine Einflüsse und zum andern untersuchungsspezifische Einflüsse auf das Zustim-

4 Für die ausführliche Diskussion der Determinanten des Datenausfalls siehe Hartmann/Krug (2009: 125ff.).

mungsverhalten. Als allgemeine, also unabhängig von Thema und Auftraggebern der Untersuchung wirksamen Einflüsse, werden neben den soziodemographischen Merkmalen Alter, Bildung, Region (Ost-/Westdeutschland), Geschlecht und Nationalität auch die subjektive Wertschätzung der Freizeit, die Arbeitszeit, der Haushaltskontext (ist ein Partner vorhanden, leben minderjährige Kinder im Haushalt) sowie das Erwerbseinkommen bzw. fehlende Angaben beim Erwerbseinkommen identifiziert, letzteres als Indikator für ein geringes Vertrauen der Befragten gegenüber dem Interviewer oder der Befragung an sich. Im Gegensatz dazu sind untersuchungsspezifische Einflüsse solche, deren Effekt auf das Zustimmungsverhalten sich speziell aus Gegenstand und Auftraggeber der Befragung ergibt. Im Fall der vorliegenden Befragung ergeben sich aus theoretischer Sicht mögliche Einflüsse einer Tätigkeit im öffentlichen Dienst, der Arbeitslosigkeit vor Beschäftigungsantritt und der Ausbildung. Ebenfalls relevant könnte sein, ob die angetretene Stelle vom (damaligen) Arbeitsamt vermittelt wurde, ob eine aufgenommene Stelle aktuell noch andauert, ob es sich um eine kombilohngeförderte Stelle handelte und ob vor dem Antritt der Beschäftigung Sozialhilfe bezogen wurde.⁵

Tabelle 1, vollständiges Modell, zeigt den empirischen Einfluss der aufgeführten potentiellen Determinanten.⁶ Während sich einige der Variablen und Indikatoren als signifikant erweisen, ist eine Vielzahl der Determinanten des allgemeinen Befragungsteilnahmeverhaltens für die Zustimmung zum Datenzusammenspielen scheinbar wenig relevant. Das kann, wie auch die recht hohe Zustimmungquote, daran liegen, dass sich die Befragten bereits für die Teilnahme an der Befragung entschieden haben und damit bereits eine positive Einstellung vorliegt. Entsprechend liegt der Wert des Pseudo R^2 lediglich bei ca. 0,03, das heißt die aufgeführten Variablen liefern einen nur sehr geringen Erklärungsbeitrag für den Datenausfall. Trotzdem gibt es auch signifikante Unterschiede zwischen zustimmenden und nicht zustimmenden Befragten.

Die nicht signifikanten Variablen können ohne wesentlichen Informationsverlust aus dem Modell entfernt werden (die Informationskriterien AIC bzw. BIC sinken dementsprechend) und es verbleiben nur die empirisch relevanten Faktoren des Ausfallmechanismus (restringiertes Modell).

- 5 Im Unterschied zu den Analysen in Hartmann/Krug 2009 werden hier nur erwerbstätige Personen untersucht, weshalb sich die verwendeten Variablen sowie die Koeffizienten leicht unterscheiden. Zudem sind die Interviewermerkmale in den verwendeten Daten nicht enthalten.
- 6 Dazu ist zu beachten, dass die Analyse mit den vollständigen Daten durchgeführt wird; je nach Ausfallszenario wäre der Einfluss mancher Variablen auf das Zustimmungsverhalten für den Anwender also nicht testbar. So kann unter Szenario 4 (vgl. weiter unten) etwa der Einfluss der Region Ostdeutschland auf das Zustimmungsverhalten nicht mehr überprüft werden, da die Variable nur für die Zustimmungsvorlieger vorliegt.

Es zeigt sich, dass ein verknüpfungsbedingter Datenausfall bei Frauen signifikant häufiger vorkommt, ebenso bei Personen mit ausländischer Staatsbürgerschaft. Dagegen kommt er bei ostdeutschen Befragten seltener vor. Tendenziell stimmen Personen mit hohem Erwerbseinkommen seltener der Verknüpfung zu als Andere. Bei Personen mit Kindern im Haushalt kommt ein verknüpfungsbedingter Datenausfall seltener vor, der vorherige Sozialhilfebezug wirkt in eine ähnliche Richtung. Schließlich stimmen Personen, die ihre Stelle durch Eigeninitiative statt über die Arbeitsvermittlung gefunden haben, ebenfalls seltener der Verknüpfung zu und sind damit häufiger vom Datenausfall betroffen. Demnach erweist sich der Ausfallprozess zwar als systematisch, da hinsichtlich einiger Variablen Unterschiede zwischen Personen mit und ohne Zustimmung bestehen. Die Systematik ist jedoch sehr gering, was am niedrigen R^2 abzulesen ist.

Wie verhält es sich nun zu den in Abschnitt 2 aufgeführten Annahmen über den Ausfallmechanismus? Gegeben, es handelt sich (im restringierten Modell) um eine vollständige Erfassung aller Determinanten des Datenausfalls, hängt die Gültigkeit der Annahmen bezüglich des Ausfallmechanismus *für ein konkretes Analysemodell* nun davon ab, ob und wie stark die Variablen in der Analysegleichung mit den hier aufgeführten Determinanten zusammenhängen. Besteht kein Zusammenhang der Analysevariablen zu den Variablen Geschlecht, Staatsbürgerschaft, etc., so spricht dies für MCAR. Besteht ein Zusammenhang, wird dieser aber mutmaßlich vollständig von den beobachteten Variablen abgedeckt, so spricht dies eher für MAR. Besteht ein Zusammenhang, man vermutet aber, dass es sich nicht um eine vollständige Auflistung der zentralen Determinanten handelt und dass die beobachteten Variablen nur einen Teil des Zusammenhanges erfassen, wäre von MNAR auszugehen. In einem gegebenen Analysefall ist empirisch nur die MCAR-Annahme zu widerlegen, sowohl MAR als auch MNAR sind empirisch nicht testbar (Schafer/Graham 2002: 151).⁷

7 Auch geringe Abweichungen von der MAR haben nicht unbedingt ein Versagen von Verfahren zur Folge, welche diese Annahme treffen. Daher ist selbst mit dem vorliegenden Untersuchungsdesign, bei dem das korrekte Ergebnis bei vollständigen Daten bekannt ist, das Zutreffen der MAR Annahme nicht eindeutig zu belegen (vgl. Schafer/Graham 2002: 151ff.), z. B. durch den Umkehrschluss „wenn die Korrektur durch Multiple Imputation erfolgreich war, dann lag MAR vor“.

Tabelle 1 Empirische Determinanten des Datenausfalls
(Probitregression mit vollständigen Daten)

Abhängige: Befragter stimmt Zusammen- spielen nicht zu (=fehlende Prozessdaten)	vollständiges Modell Koeff. (Std.Fehler)	restringiertes Modell Koeff. (Std.Fehler)
Kombilohnförderung: ja	-0,068 (0,074)	
Alter (Ref.: Bis unter 25 Jahre)		
25 bis 34 Jahre	0,070 (0,104)	
35 bis 44 Jahre	-0,034 (0,103)	
45 bis 54 Jahre	0,018 (0,109)	
55 Jahre oder mehr	-0,191 (0,180)	
Bildung (Ref.: kein Abschluss)		
Volks-/Hauptschulabschluss	-0,149 (0,156)	
Volks-, Hauptschule	-0,084 (0,164)	
Mittlere Reife, POS 10.Klasse	-0,041 (0,154)	
Fachhochschulreife, Abitur	0,078 (0,162)	
Keine Angabe	0,037 (0,226)	
Geschlecht: weiblich	0,125* (0,069)	0,111* (0,063)
Keine deutsche Staatsbürgerschaft	0,221** (0,105)	0,238** (0,102)
Region: Ost	-0,182*** (0,070)	-0,160*** (0,058)
Freizeit sehr wichtig	-0,052 (0,064)	
Teilzeit (< 30 Std.): ja	-0,029 (0,066)	
Bruttoeinkommen: (Ref.: 1.024 Euro oder mehr)		
1 bis 325 Euro	-0,462*** (0,175)	-0,439** (0,171)
326 bis 511 Euro	-0,088 (0,111)	-0,113 (0,109)
512 bis 1.023 Euro	-0,245*** (0,069)	-0,258*** (0,068)
Keine (valide) Einkommensangabe	-0,393*** (0,080)	-0,368*** (0,078)

Abhängige: Befragter stimmt Zusammen- spielen nicht zu (=fehlende Prozessdaten)	vollständiges Modell Koeff. (Std.Fehler)	restringiertes Modell Koeff. (Std.Fehler)
Ursprüngliche Beschäftigung zum Befragungszeitpunkt beendet	0,057 (0,059)	
Lebensform: mit Partner	-0,043 (0,061)	
Kinder im Haushalt: ja	-0,204*** (0,068)	-0,216*** (0,057)
Beschäftigt im öffentlichen Dienst: ja	-0,011 (0,094)	
Kumulierte Dauer der Arbeitslosigkeit (Ref.: unter einem Monat)		
Ein bis unter sechs Monate	0,060 (0,080)	
Sechs bis unter zwölf Monate	0,005 (0,090)	
Zwölf bis unter vierundzwanzig Monate	0,146 (0,092)	
Vierundzwanzig Monate und mehr	0,059 (0,112)	
Vor Beschäftigung Sozialhilfe: ja	-0,184** (0,080)	-0,199** (0,079)
Wie wurde die angetretene Stelle gefunden (Ref.: Arbeitsamt)		
Bekannte/Freunde	-0,017 (0,079)	-0,020 (0,078)
Eigene Initiative	0,142* (0,075)	0,142* (0,074)
Sonstige Suchwege	0,059 (0,081)	0,059 (0,080)
Konstante	-0,306 (0,191)	-0,366*** (0,092)
Pseudo R ²	0,029	0,024
p-Wert (chi ² -Test)	0,000	0,000
N	2604	2604
AIC	2934,02	2912,468
BIC	3121,694	2988,711
Likelihood-Ratio-Test auf gemeinsame Insignifikanz der im restringierten Modell ausgelassenen Variablen	chi ² = 16,45; df = 18; p = 0,63	

*p<0,1; **p<0,05; ***p<0,01

3.2 Untersuchungsdesign

Für den nachfolgenden empirischen Vergleich wird zunächst auf Basis der vollständigen, unbearbeiteten Befragungsdaten eine Analysegleichung zur Bestimmung des Einflusses einiger unabhängiger Variablen $\mathbf{X} = (T, U, V_j, W_k)$ auf die abhängige Variable Y in Form einer Regression aufgestellt:

$$Y = \alpha_0 + \beta T + \gamma U + \sum_{j=1}^J \delta_j V_j + \sum_{k=1}^K \eta_k W_k + \varepsilon \quad (6)$$

Dabei soll T die Variable sein, deren Effekt β auf Y hauptsächlich interessiert, während die anderen Elemente von \mathbf{X} lediglich als Kontrollvariablen dienen. Die Ergebnisse aus der Regression mit den vollständigen Befragungsdaten stellen die „wahren Werte“⁸ der Koeffizienten dar. Sie dienen somit als Maßstab für den empirischen Vergleich der Missing Data Verfahren, an dem deren Erfolg bei der Korrektur gemessen wird.

Um die Situation nach einer tatsächlichen Datenverknüpfung zu simulieren (die faktisch jedoch nicht stattfand) werden nun einzelne Variablen zu Prozessdaten „umdeklariert“⁹, indem bei den Personen ohne Zustimmung die entsprechenden Werte gelöscht werden. Im Anschluss an die Löschung werden die Missing Data Verfahren auf die Daten mit diesen pseudo-fehlenden Werten angewendet und es wird geprüft, ob die erzielten Ergebnisse den „wahren Werten“ entsprechen. Dabei werden verschiedene Szenarien durchgespielt (Tabelle 2). Diese unterscheiden sich zunächst darin, ob die Prozessdaten als Kontrollvariablen oder als abhängige Variable in der Analyse dienen. Dienen sie als Kontrollvariablen, so werden schrittweise immer mehr Kontrollvariablen des Analysemodells als Prozessdaten behandelt und mit zustimmungsbedingt fehlenden Werten simuliert. Je mehr Variablen fehlende Werte aufweisen, desto schwieriger sollte die Situation für die Korrekturverfahren werden: Bei der Heckman-Korrektur muss argumentiert werden, dass die Störgröße bzw. die „fehlende(n) Variable(n)“ nicht mit den im Modell enthaltenen Variablen \mathbf{X} in Gleichung 4 bzw. \mathbf{C} in Gleichung 5 korreliert sind. Dies wird mit der zunehmenden Zahl ausfallbehafteter Variablen schwieriger. Bei der Multiplen Imputation muss entsprechend für jede einzelne ausfallbehaftete Variable plausibel sein, dass

8 „Wahre Werte“ steht hier in Anführungszeichen, da es sich ja auch bei den Ergebnissen mit vollständigen Daten um eine Schätzung handelt.

9 Unabhängig von der Zustimmung lagen bei allen Befragten einige wenige Informationen aus den administrativen Daten der Bundesagentur für Arbeit bereits vor (vgl. Hartmann et al. 2002: 173ff.). Dazu gehört neben der Variablen *Arbeitslosigkeitsdauer* auch die Variable *Stellung im Beruf*. Alle anderen Variablen stammen aus der Befragung, werden aber zum Zweck des Vergleichs je nach Szenario wie Prozessdaten mit pseudo-fehlenden Werten versehen.

die MAR-Annahme gegeben ist und auch hier gilt, je mehr solche Variablen, desto eher können Abweichungen von der MAR-Annahme auftreten.

In Szenario 1 wird angenommen, dass nur eine Variable eine ausfallbehaftete Prozessdatenvariable darstellt, nämlich die abhängige Variable Y des Analysemodells (vgl. Gleichung 6). Hierzu werden im Datensatz die Werte der Variable Y für diejenigen Befragten auf „fehlend“ gesetzt, welche ihre Zustimmung zu einer Datenverknüpfung verweigert hatten. Auf Basis dieses Datensatzes mit pseudo-fehlenden Prozessdaten bei den Personen, die einer Zusammenspielung nicht zugestimmt hatten, werden schließlich Schätzungen der Analysegleichung (6) unter Verwendung der in Abschnitt 2 vorgestellten Möglichkeiten des Umgangs mit den fehlenden Werten durchgeführt. Die Ergebnisse der korrigierten Schätzungen können dann mit den tatsächlich auf Basis der vollständigen Daten durchgeführten Schätzungen verglichen werden.

In Szenario 2 wird demgegenüber angenommen, dass lediglich die unabhängige Variable U eine Prozessdatenvariable darstellt und damit nur dort fehlende Werte aufgrund der Verknüpfung entstehen. In Szenario 3 werden Ausfälle in der Variable U und einer Reihe weiterer Kontrollvariablen (V_j) simuliert und in Szenario 4 werden schließlich alle oben aufgeführten Kontrollvariablen (U, V_j, W_k) auf „fehlend“ gesetzt, wenn die Zustimmung zur Datenverknüpfung nicht vorliegt.¹⁰ Die Variable T enthält in keinem der Szenarien fehlende Werte, ist also immer eine Variable aus den Befragungsdaten. Dies soll einerseits garantieren, dass eventuelle Abweichungen zwischen vollständigen und ausfallkorrigierten Daten auf die Qualität der Korrektur der *anderen* Variablen zurückgehen. Andererseits bleiben die Koeffizienten dann über die verschiedenen Szenarien vergleichbar, da sie – z. B. im Fall der Multiplen Imputation – nie imputierte Werte aufweisen.¹¹

Tabelle 2 Szenarien des Datenausfalls durch zustimmungspflichtige Datenverknüpfung

Szenario 1	Fehlende Werte in der abhängigen Variable Y
Szenario 2	Fehlende Werte in U , einer unabhängigen Variable
Szenario 3	Fehlende Werte in mehreren unabhängigen Variablen (U, V_j)
Szenario 4	Fehlende Werte in allen unabhängigen Variablen (U, V_j, W_k)

10 Natürlich wären noch weitere Szenarien bzw. Abstufungen zwischen Szenario 2 und 4 möglich, die vorliegende Analyse beschränkt sich auf diese eine, zumal der Informationsgehalt zusätzlicher Abstufungen wohl eher gering ist.

11 Für Anhaltspunkte, wie die Korrekturverfahren abschneiden, wenn eine *Prozessdatenvariable* im inhaltlichen Interesse der Analyse steht, siehe die Tabellen zum Aufsatz (<http://www.gesis.org/forschung-lehre/gesis-publikationen/zeitschriften/mda/jg-4-2010-heft-1/>), vor allem die Variable „Arbeitslosigkeitsdauer“.

Alle Szenarien werden für zwei unterschiedliche Analysemodelle mit je unterschiedlich starker Beziehung zu den Ausfalldeterminanten durchgespielt.¹² Im Analysemodell 1 wird eine empirische Analyse durchgeführt, die nur gering durch die Selektivität des Datenausfalls betroffen ist. Eine solche Analyse stellt die folgende Untersuchung des Einflusses der Kombilohnförderung „Mainzer Modell“ auf die Arbeitszeit dar (vgl. Hartmann/Krug 2009)¹³:

Analysemodell 1:

$$AZEIT = \alpha_0 + \beta MZM + \sum_{l=1}^4 \gamma_l ALO_l + \sum_{j=1}^J \delta_j KONT_j^1 + \sum_{k=1}^K \eta_k KONT_k^2 + \varepsilon$$

Dabei steht *AZEIT* für die Arbeitszeit der Person, *MZM* ist ein binärer Indikator für die Kombilohnförderung im Mainzer Modell (1, falls Mainzer Modell) und *ALO* steht für Dummyvariablen (1 bis unter 6 Monate, 6 bis unter 12 Monate, 12 bis unter 24 Monate und ab 24 Monate), welche die klassierte „kumulierte Dauer der Arbeitslosigkeit im Erwerbsleben“ abbilden. Schließlich stehen *KONT¹* und *KONT²* für zwei Vektoren diskreter sowie kontinuierlicher Kontrollvariablen. Da sich der Datenausfall nur gering auswirkt, sollte Analysemodell 1 eine eher geringe Herausforderung für die Korrekturverfahren darstellen und sie sollten zu Ergebnissen führen, die sehr nah an den „wahren“ Werten liegen. Allerdings weisen alle Verfahren noch weitere Anwendungsbedingungen auf, so dass nicht automatisch von einer erfolgreichen Korrektur des Datenausfalls auszugehen ist. So können eventuelle Abweichungen von den wahren Daten auf eine Verletzung der verfahrensspezifischen Anwendungsbedingungen schließen lassen bzw. darauf, dass die Verfahren nicht korrekt durchgeführt wurden.

Analysemodell 2 ist analog zu Analysemodell 1 formuliert, wird hingegen aber so konstruiert, dass das Analysebeispiel stark von der verknüpfungsbedingten Selektivität betroffen ist. Ein solcher Fall ist die folgende Regression, bei der eine der zentralen Determinanten des Zustimmungsverhaltens als abhängige Variable der Regression gewählt wird:

12 Um diese beiden unterschiedlichen Analysemodelle zu identifizieren, musste auf das Ergebnis bei einem fallweisen Ausschluss vorgegriffen werden, weshalb bereits vor der Durchführung streng genommen vorhersehbar war, dass der fallweise Ausschluss in Analysemodell 2 nicht funktioniert. Der Erfolg der anderen beiden Verfahren war natürlich dennoch vollkommen offen.

13 Explizites Ziel im Mainzer Modell war neben der Aktivierung Arbeitsloser zur Aufnahme niedrig entlohnter Beschäftigung vor allem auch die Förderung von Teilzeitbeschäftigung (vgl. Kaltenborn et al. 2005).

Analysemodell 2:

$$EKJA = \alpha_0 + \beta ENDE + \sum_{l=1}^4 \gamma_l ALO_l + \sum_{j=1}^J \delta_j KONT_j^1 + \sum_{k=1}^K \eta_k KONT_k^2 + \varepsilon$$

Hier ist die abhängige Variable *EKJA* eine binäre Variable dafür, ob eine Person in der Befragung (eine valide) Einkommensangabe gemacht hat (1, falls ja), *ENDE* ein binärer Indikator, ob das Beschäftigungsverhältnis, auf das sich die Einkommensangabe bezieht, zum Zeitpunkt der Befragung noch andauert (1, falls nein). *KONT¹* und *KONT²* stehen wiederum für zwei Vektoren diskreter sowie kontinuierlicher Kontrollvariablen, die sich allerdings von denen in beiden Analysemodellen unterscheiden.

Die beiden Analysemodelle sollen zwar möglichst realitätsnah sein, es wird aber nicht der Anspruch erhoben, dass diese Analysen auch genau so durchgeführt würden, wenn es um eine inhaltliche Analyse und nicht um einen Methodenvergleich ginge. Vielmehr wurden sie mit Blick auf die methodischen Schwierigkeiten gewählt. So sind in beiden Analysen sowohl signifikante als auch insignifikante Kontrollvariablen enthalten, was für die Korrekturverfahren bedeutet, dass die Möglichkeit besteht, dass bei fehlerhafter Korrektur signifikante Variablen insignifikant werden und umgekehrt.

3.3 Implementation der Missing Data Verfahren

Die technisch am wenigsten aufwändige Variante ist der fallweise Ausschluss. Hier bedarf es keiner gesonderten Schätzverfahren und/oder Software. Es werden lediglich alle Fälle mit fehlenden Werten aus der Analyse entfernt, und auf die verbliebenen Fälle werden die üblichen statistischen Schätzverfahren angewendet.

Zur Multiplen Imputation der pseudo-fehlenden Prozessdaten wurde in der folgenden Analyse die Methode der verketteten Regressionen („Sequential Regression Multivariate Imputation“ SRMI) verwendet, die in der Software IVEWare implementiert ist (Raghuathan et al. 2001; Raghuathan/Solenberger/van Hoeweyk 2002). IVEWare bietet im vorliegenden Anwendungsfall den Vorteil, dass hier komplexe Datenstrukturen berücksichtigt werden können, wie sie in Befragungen häufig auftreten. Neben Imputationsroutinen für kontinuierliche Merkmale bietet IVEWare auch solche für Zähldaten, dichotome und kategoriale Variablen, und es können u. a. Filterbedingungen berücksichtigt werden, z. B. dass nur für Personen mit Kindern die Anzahl oder das Alter dieser Kinder imputiert wird.

Wird bei der Selektionskorrektur zwischen Analyse- und Selektionsgleichung unterschieden, so ist bei der Imputation zwischen dem Analysemodell (die

Variablen in der Regression der Arbeitszeit bzw. der Regression der validen Einkommensangabe) und dem Imputationsmodell (für die Imputation verwendete Variablen) zu unterscheiden. Das Imputationsmodell sollte neben den Variablen des Analysemodells zusätzlich die Variablen enthalten, die a) mit der (den) ausfallbelasteten Variable(n) und b) mit dem Ausfall selbst zusammenhängen (Schafer 1999a: 143). Variablen der Kategorie b) sind vor allem diejenigen Variablen aus Tabelle 1, restringiertes Modell, welche nicht bereits im Analysemodell enthalten sind. Variablen der Kategorie a) sind zwar nicht mit dem Datenausfall korreliert, aber mit der ausfallbelasteten Variable. Neben den Variablen des Analysemodells werden daher im Imputationsmodell noch eine Reihe zusätzlicher Variablen aufgenommen (vgl. Tabelle 5 im Anhang). Für eine ausführliche Diskussion, welche Variablen sinnvollerweise in das Imputationsmodell aufgenommen werden sollen und der Vor- und Nachteile verschiedener Strategien siehe Collins/Schafer/Kam (2001).

Zur Imputation im Rahmen der SRMI Methode wird wie folgt vorgegangen (Raghunathan et al. 2001). Seien P_1, P_2, \dots, P_q die q Prozessdatenvariablen mit fehlenden Werten für Personen, die ihre Zustimmung zur Datenverknüpfung verweigert haben und sei \mathbf{B} der Vektor der Befragungsvariablen, die also vollständige Beobachtungen enthalten. Im ersten Schritt des ersten Durchgangs wird die Variable P_1 auf \mathbf{B} regressiert und auf dieser Basis die fehlenden Werte in P_1 imputiert. Die Art der Regression ist vom Skalenniveau des Merkmals abhängig.¹⁴ Im zweiten Schritt wird die Variable P_2 auf P_1, \mathbf{B} regressiert – wobei P_1 nun sowohl die beobachteten als auch imputierten Werte enthält, im dritten Schritt P_3 auf P_1, P_2, \mathbf{B} , etc. Im zweiten Durchgang wird dann die jeweilige abhängige Variable, zum Beispiel P_1 , auf alle anderen Variablen, also $P_2, \dots, P_q, \mathbf{B}$ regressiert, und die im ersten Durchgang imputierten Werte in P_1 werden durch die neuen Werte des zweiten Durchgangs ersetzt. Raghunathan/Solenberger/van Hoeweyk (2002:16) geben an, dass nach zehn Durchgängen die Imputation beendet und ein Datensatz mit vollständigen Daten erzeugt werden kann.

Eine Regressionsanalyse der Gleichung 6 auf Basis dieser imputierten Daten würde nicht die Unsicherheit berücksichtigen, die dadurch entsteht, dass die imputierten Prozessdatenwerte Schätzungen darstellen. Dies wird gewährleistet, indem nicht nur eine (Single Imputation), sondern mehrere Ergänzungen (Multiple Imputation) durchgeführt werden. Zudem enthält die Imputation eine stochastische Komponente (vgl. Fn. 14). Im Allgemeinen gelten bei bis zu 50 % fehlender

14 Implementiert sind lineare, logistische und multinomiale logistische Regressionen sowie Poissonregressionen. Dabei wird allerdings zur Prognose von P_i den jeweiligen Regressionskoeffizienten aus der Regression von P_i auf \mathbf{B} ein normalverteilter Zufallsterm hinzuaddiert.

Informationen fünf Imputationen und somit auch fünf verschiedene imputierte Datensätze als ausreichend (Schafer 1999b: 7; kritisch dazu Bodner 2008; Graham/Olchowski/Gilreath 2007).

Zur Schätzung der Regressionsparameter durch Heckmans Selektionskorrektur wurde für das Analysemodell 1 der Stata-Befehl *heckman* verwendet, wobei sowohl die Kleinste-Quadrate als auch die Maximum-Likelihood-Variante implementiert ist. Bei der KQ-Variante wird zunächst die Probitregression der Zustimmung zur Datenverknüpfung durchgeführt $P(D = 1 | C) = \phi(C\alpha)$ (vgl. Gleichung 5) und daraus die inverse Mills Ratio berechnet. Hierzu können nur die Variablen ohne fehlende Werte, also die Befragungsvariablen, als Regressoren verwendet werden. Die inverse Mills Ratio wird dann in eine KQ-Regression der Analysegleichung als zusätzlicher Regressor aufgenommen, wodurch die Schätzung der Regressionskoeffizienten auf Basis der Personen, die der Datenverknüpfung zugestimmt haben, unverzerrt erfolgen kann. In der ML-Variante werden die Probitregression der Zustimmung und die lineare Regression der Arbeitszeit simultan geschätzt. Im Analysemodell 2, das sich u. a. durch eine binäre abhängige Variable auszeichnet, wurde der Stata-Befehl *heckprob* verwendet, wobei ausschließlich eine ML-Schätzung möglich ist und dabei zwei Probitregressionen simultan geschätzt werden.

Typischerweise fällt es schwer, geeignete Instrumentvariablen zur Durchführung der Heckman-Korrektur zu finden (Puhani 2000), das heißt in den konkreten Anwendungsfällen eine oder mehrere Variablen, die in der Selektionsregression signifikant sind, bei Aufnahme in die Analysegleichung jedoch nicht. Grundsätzlich kommen genau die Variablen aus dem restringierten Modell in Tabelle 1 in Frage, welche nicht bereits im jeweiligen Analysemodell vorkommen. Für die Analyse der Arbeitszeit (Analysemodell 1) kommen somit lediglich die Variablen „ausländische Staatsbürgerschaft“, „vorheriger Sozialhilfebezug“ sowie das „Finden der Stelle über ...“ in Frage, da die Variable „Region Ostdeutschland“ explizit und die Variablen „Geschlecht“ und „Kinder vorhanden“ implizit in der Variable „Haushaltskontext“ enthalten sind. Für das Bruttomonatseinkommen gilt, dass dieses mit der Arbeitszeit korreliert ist, so dass es ebenfalls als Instrument ausscheidet.¹⁵ Für die Regression der vorhandenen Einkommensangabe der Arbeitszeit (Analysemodell 2) sind die verfügbaren Instrumente entsprechend „ausländische Staatsbürgerschaft“, „Region Ostdeutschland“ und „Finden der Stelle über ...“.

15 Die Einkommensvariable wurde nicht als erklärende Variable in Analysemodell 1 aufgenommen, da die Arbeitszeit das Bruttomonatseinkommen beeinflusst und nicht umgekehrt.

3.4 Ergebnisse des empirischen Vergleiches

In diesem Abschnitt soll die Qualität der Ausfallkorrektur durch die verschiedenen Verfahren miteinander verglichen werden. Ein Ergebnis des Abschnitts 3.1 sowie früherer Analysen mit dem vorliegenden Datensatz (vgl. Hartmann/Krug 2009) war, dass es nur hinsichtlich weniger Variablen systematische Unterschiede zwischen Personen mit und ohne Erlaubnis zur Datenzusammenspielung gibt. Dennoch kann die Analyse bestimmter Fragestellungen *im Einzelfall* stärker oder weniger stark mit diesen wenigen Variablen zusammenhängen. Daher erfolgt der Test der Korrekturverfahren zunächst für eine Fragestellung, welche eher gering von der Selektivität des Datenausfalls betroffen ist (Analysemodell 1), wobei die Ergebnisse unter allen vier Szenarien des Ausfalls den Ergebnissen aus vollständigen Daten gegenüber gestellt werden. Danach wird entsprechend mit dem Analysemodell 2 vorgegangen, das wesentlich stärker von der Ausfallselektivität betroffen ist. Wichtig ist hierbei allerdings, dass ein Anwender im Normalfall statistisch nur bedingt testen kann, ob er bei seinen Analysen von starker oder schwacher Selektivität ausgehen muss, so dass die Entscheidung auf Basis fundierter Überlegungen über den Datenausfall erfolgen muss.

3.4.1 Analysemodell 1: schwache Selektivität

Zunächst wird der Vergleich der verschiedenen Missing Data Verfahren für das Analysemodell 1 durchgeführt, der durch eher schwache Selektivität gekennzeichnet ist. Berücksichtigt man nur die Fälle, in denen die Befragungsdaten keine fehlenden Werte aufweisen, so beträgt die Ausfallquote ca. 25 %. Tabelle 3 zeigt die Ergebnisse der verschiedenen Missing Data Verfahren für die verschiedenen Szenarien im Analysemodell 1 und vergleicht sie mit den Ergebnissen der echten, vollständigen Daten. Der Fokus des Vergleichs liegt auf dem Koeffizienten des Dummies für die Kombilohnförderung (β in Gleichung 6). Darüber hinaus liefert grundsätzlich auch der Vergleich der geschätzten Koeffizienten der pseudo-ausfallbehafteten unabhängigen Variable(n) mit den „wahren“ Koeffizienten bei vollständigen Daten Anhaltspunkte zur Bewertung der Missing Data Verfahren. Während die vorliegende Analyse lediglich die Korrektur des Koeffizienten einer nicht ausfallbehafteten Variable durch die ausgewählten Verfahren untersucht, kann aus den Tabellen auch abgelesen werden, wie es sich bei der Korrektur ausfallbehafteter Variablen verhält. Aufgrund ihres Umfangs werden die entsprechenden vollständigen Tabellen nur auf der Webseite der mda (<http://www.gesis.org/forschung-lehre/gesis-publikationen/zeitschriften/mda/jg-4-2010-heft-1/>) veröffentlicht.

Tabelle 3 Analysemodell 1 – Koeffizienten und Standardfehler der Dummyvariable „Kombilohnförderung (MZM)“ nach Szenarien

OLS ^a - Regressionen der wöchentlichen Arbeitszeit in Stunden	vollständige Daten	fallweiser Ausschluss	Multiple Imputation	Heckman- Korrektur (OLS)	Heckman- Korrektur (ML)
	Koeff. (Std.Fehler)	Koeff. (Std.Fehler)	Koeff. (Std.Fehler)	Koeff. (Std.Fehler)	Koeff. (Std.Fehler)
Szenario 1	-4,769*** (0,550)	-4,462*** (0,632)	-4,135*** (0,660)	-4,402*** (0,641)	-4,270*** (0,682)
Szenario 2	-4,769*** (0,550)	-4,462*** (0,632)	-4,752*** (0,551)	-4,370*** (0,650)	-4,180*** (0,680)
Szenario 3	-4,769*** (0,550)	-4,462*** (0,632)	-4,349*** (0,674)	-4,342*** (0,663)	-4,131*** (0,664)
Szenario 4	-4,769*** (0,550)	-4,462*** (0,632)	-4,455*** (0,577)	-4,296*** (0,685)	-4,041*** (0,665)

* $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$

^a Ausnahme SSM(ML)

In der zweiten Spalte der Tabelle 3 findet sich die Schätzung auf Basis der vollständigen Daten. Die Ergebnisse zeigen, dass die Kombilohnförderung einen signifikant negativen Einfluss auf die Arbeitszeit hatte. Im Mittel arbeiten Geförderte 4,8 Stunden weniger als ungefördert Beschäftigte, dies gilt (natürlich) für alle Szenarien gleichermaßen.

Fallweiser Ausschluss

Die Analyse mit fallweisem Ausschluss kommt, da sie nur schwach von Selektivität betroffen ist, zu sehr ähnlichen Ergebnissen wie eine Analyse unter Vollständigkeit der Daten. Allerdings liegt bei fallweisem Ausschluss eine leichte Unterschätzung des negativen Einflusses der Förderung vor (d. h. der negative Effekt wird als leicht geringer eingeschätzt). Bedingt durch die geringere Fallzahl zeigt sich auch ein größerer Standardfehler. Dennoch bleibt der Effekt auf dem 0,1 %-Niveau signifikant, so dass sich die aus der Analyse ergebende Schlussfolgerung nicht verändert. Auch bei den anderen Variablen (Tabelle 6, siehe <http://www.gesis.org/forschung-lehre/gesis-publikationen/zeitschriften/mda/jg-4-2010-heft-1/>) halten sich die Unterschiede zu den Ergebnissen bei vollständigen Daten in Grenzen. Lediglich beim Dummy für die kumulierte Arbeitslosigkeitsdauer „Sechs bis unter 12 Monaten“ wird aus einem

insignifikanten ein stärkerer und nun signifikanter Effekt.¹⁶ Stets bleiben (nicht) signifikante Einflüsse (nicht) signifikant, wenn es auch zu Unterschieden in der Größe der Koeffizienten kommt, etwa den Dummies für „Angestellte“ bzw. „Reinigungsgewerbe“. Beim Dummy „Freizeit sehr wichtig“ dreht sich zwar das Vorzeichen, der entsprechende Koeffizient ist und bleibt jedoch insignifikant. Aufgrund des Datenausfallmusters bei der Datenverknüpfung (die ausfallbehafteten Beobachtungen sind stets dieselben, unabhängig davon, welche Variablen betrachtet werden) verändert sich die Schätzung über die verschiedenen Szenarien nicht.

Multiple Imputation

Im Gegensatz zum fallweisen Ausschluss unterscheiden sich bei den anderen statistischen Korrekturverfahren die Ergebnisse je nach Szenario. Zunächst zur Multiplen Imputation: In Szenario 1 bestehen lediglich in der abhängigen Variable des Analysemodells fehlende Werte. Hier steigt zum einen der Standardfehler im Vergleich zur Analyse mit vollständigen Daten, zum anderen schneidet die Imputation schlechter ab als der fallweise Ausschluss (oder auch die Heckman-Korrektur in der KQ-Variante, vgl. unten). Kaum Unterschiede ergeben sich zwischen Multipler Imputation und der Selektionskorrektur nach Heckman in der Maximum-Likelihood-Variante. In Szenario 2 mit nur einer pseudo-ausfallbelasteten Kontrollvariable zeigt sich hingegen, dass die Schätzung für den Effekt der Förderung sehr nahe an den „wahren“ Ergebnissen mit vollständigen Daten liegt. Enthalten mit Szenario 3 und 4 ein großer Teil bzw. alle unabhängigen Variablen fehlende Werte, so entfernt sich die Schätzung vom Referenzwert aus den vollständigen Daten. Sie ist aber immer noch nahe am Ergebnis mit vollständigen Daten, wenn sie auch erhöhte Standardfehler aufweist.

Heckman-Korrektur

Die Selektionskorrektur in der KQ-Variante weist über alle Szenarien eine ähnliche Qualität der Ergebnisse auf. Dabei sind die Schätzergebnisse stets etwas weiter von den Ergebnissen mit vollständigen Daten entfernt als beim fallweisen Ausschluss. Auch die Standardfehler sind vergleichsweise groß. Im Übergang von Szenario 1 zu Szenario 2 zeigt sich im Gegensatz zur Imputation allerdings eine nur geringe Veränderung der Koeffizienten. Sie entsteht dadurch, dass die Variable *ALO* in Sze-

16 Im Gegensatz dazu führt beim Dummy „Arbeit sehr wichtig“ die leichte Erhöhung des Standardfehlers zur Insignifikanz des Effektes, der Koeffizient war aber bereits bei vollständigen Daten an der Schwelle zur Insignifikanz.

nario 2 nun ausfallbehaftet ist und daher nicht zur Berechnung von λ durch die Selektionsgleichung verwendet werden kann. Im Vergleich zur Multiplen Imputation liefert die KQ-Variante der Heckman-Korrektur also bei fehlenden Werten in der unabhängigen Variable schlechtere, in einer abhängigen Variable bessere Ergebnisse und bei mehreren unabhängigen Variablen in etwa ähnliche (Szenario 3) bis leicht schlechtere Ergebnisse (Szenario 4). Demgegenüber liefert die ML-Variante der Selektionskorrektur nach Heckman die Koeffizientenschätzungen mit der größten Abweichung zum vollständigen Datensatz, am stärksten ist die Abweichung in Szenario 4. Doch selbst hier hält sich die Abweichung in Grenzen. Der Koeffizient ist auch in Szenario 4 signifikant von Null verschieden und mit 4,041 im Vergleich zum „wahren Wert“ 4,769 um nur 15 % unterschätzt.

3.4.2 Analysemodell 2: starke Selektivität

In Tabelle 4 sind die Ergebnisse der Korrekturverfahren unter der Bedingung einer starken Betroffenheit des Analysemodells von der Selektivität des Datenausfalls abgebildet. Berücksichtigt man wiederum nur die Fälle, in denen die Befragungsdaten keine fehlenden Werte aufweisen, so ist die Ausfallquote mit ca. 22 % leicht geringer als im Analysemodell 1. Erneut liegt der Schwerpunkt des Vergleichs auf dem Koeffizienten β , das heißt in Analysemodell 2 dem Koeffizienten der Variable „Beschäftigung zum Interviewzeitpunkt bereits beendet“. Für die vollständigen Tabellen sei auf Anhang B (siehe <http://www.gesis.org/forschung-lehre/gesis-publikationen/zeitschriften/mda/jg-4-2010-heft-1/>) verwiesen. In den vollständigen Daten zeigt sich hier, dass die Wahrscheinlichkeit sinkt, dass Befragte eine valide Einkommensangabe machen, wenn die angesprochene Beschäftigung zum Zeitpunkt der Befragung bereits beendet ist.

Fallweiser Ausschluss

Der starke Einfluss der Selektivität zeigt sich deutlich, wenn man die Ergebnisse der Regression mit vollständigen Daten und bei fallweisem Ausschluss vergleicht. Ist der Effekt ohne verknüpfungsbedingten Datenausfall mit -0,174 signifikant negativ, so zeigt sich in der Regression mit fallweisem Ausschluss ein weitaus kleinerer und insignifikanter Effekt von -0,089. Damit kommt es durch den Datenausfall im Vergleich zum Fall mit schwacher Selektivität zu inhaltlich deutlich anderen Aussagen. In ähnlicher Weise sind auch andere Variablen des Analysemodells 2 betroffen. So ist etwa der Effekt des Geschlechts nach fallweisem Ausschluss nicht mehr signifikant, während umgekehrt der Effekt des Alters nun als signifikant für die valide Einkommensangabe erscheint.

Tabelle 4 Analysemodell 2 – Koeffizienten und Standardfehler der Dummyvariable „Beschäftigung zum Befragungszeitpunkt bereits beendet (*ENDE*)“ nach Szenarien

Probitregressionen „Einkommensangabe vorhanden“	vollständige Daten Koeff. (Std.Fehler)	fallweiser Ausschluss Koeff. (Std.Fehler)	Multiple Imputation Koeff. (Std.Fehler)	Heckman- Korrektur (ML) Koeff. (Std.Fehler)
Szenario 1	-0,174** (0,065)	-0,089 (0,075)	-0,085 (0,065)	-0,066 (0,096)
Szenario 2	-0,174** (0,065)	-0,089 (0,075)	-0,168** (0,065)	-0,066 (0,086)
Szenario 3	-0,174** (0,065)	-0,089 (0,075)	-0,176** (0,065)	-0,061 (0,069)
Szenario 4	-0,174** (0,065)	-0,089 (0,075)	-0,171* (0,067)	-0,079 (0,082)

* $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$

Multiple Imputation

Geht man davon aus, dass die Selektivität durch beobachtete Daten erfasst werden kann, so sollte die Multiple Imputation die bei fallweisem Ausschluss vorliegenden Verzerrungen beheben können. In Szenario 1 mit fehlenden Werten in der abhängigen Variable ist dies nicht der Fall. Der Koeffizient der Variable „Beschäftigung beendet“ unterscheidet sich kaum vom verfälschten Wert bei fallweisem Ausschluss.¹⁷ Fehlen die Werte allerdings lediglich in einer der unabhängigen Variablen (Szenario 2), so gelingt es der Multiplen Imputation die ausfallbedingten Verzerrungen weitgehend auszugleichen. Der Koeffizient ist mit -0,168 sehr nahe am Ergebnis mit vollständigen Daten und wird ebenfalls als signifikant ausgewiesen. In Szenario 2 muss lediglich *MAR* im Bezug auf die Variable *ALO* vorliegen, in Szenario 3 muss diese Annahme für *ALO* und zusätzlich alle Variablen *KONT*¹ erfüllt sein. Aber auch hier liegt das Ergebnis der Korrektur mit -0,176 sehr nahe am „wahren“ Koeffizienten und ist signifikant von Null verschieden. Zeigen mit Szenario 4 alle Kontrollvariablen fehlende Werte, wird der Koeffizient mit -0,171 ebenfalls als signifikant ausgewiesen, wenn auch der Standardfehler im Vergleich zu den

17 Um zu testen, ob sich das Ergebnis der Multiplen Imputation mit mehr als 5 Imputationen verbessert, wurde die Analyse auch mit 10 und 20 Imputationen durchgeführt. Die Koeffizienten und Standardfehler (in Klammern) unterschieden sich mit -0,087 (0,064) nach 10 und -0,084 (0,066) nach 20 Imputationen nicht wesentlich von den Ergebnissen mit 5 Imputationen.

vollständigen Daten etwas ansteigt. Damit führt die Multiple Imputation auch im Fall starker Selektivität zu einer guten Korrektur des Datenausfalls. Dies gilt jedoch nicht für fehlende Werte in der abhängigen Variable.

Heckman-Korrektur

Im Gegensatz zur Multiplen Imputation geht das Heckman-Verfahren auch von unbeobachteten Einflüssen auf den Datenausfall aus und berücksichtigt diese bei der Ausfallkorrektur. Ist die abhängige Y-Variable binär, so kommt ausschließlich die ML-Variante der Heckman-Korrektur in Frage. Trotz der weniger restriktiven Annahmen über den Ausfallprozess zeigt sich das Heckman-Verfahren allerdings als wenig erfolgreich bei der Korrektur des Datenausfalls. In allen vier Szenarien wird das Ausmaß des negativen Effektes der Variable „Beschäftigung beendet“ mit Werten zwischen $-0,06$ und $-0,08$ im Vergleich zum „wahren Wert“ $-0,174$ deutlich unterschätzt. Zum andern wird der Effekt im Gegensatz zu den Ergebnissen aus vollständigen Daten als insignifikant ausgewiesen. Damit liegt die Heckman-Korrektur sogar noch unter den Ergebnissen des fallweisen Ausschlusses.

4 Zusammenfassung und Diskussion der Ergebnisse

Beschränkt man sich beim Umgang mit fehlenden Prozessdaten auf den fallweisen Ausschluss, so setzt man voraus, dass der Datenausfall ohne Zusammenhang mit den Analysevariablen ist. In der durchgeführten Fallstudie führte eine schwache Selektivität des Datenausfalls im Bezug auf das entsprechende Analysemodell zu nur geringen Unterschieden im Vergleich zu Analysen mit vollständigen Daten. Hat ein Anwender also Grund für die Annahme, dass der Datenausfall bei der Zusammenspielung zufällig oder zumindest nur schwach systematisch erfolgt, so erscheint der fallweise Ausschluss als einfache und sinnvolle Alternative zu komplexen Korrekturverfahren. Lag der Fall vor, dass die Analysevariablen stark mit den Determinanten des Ausfallprozesses zusammenhängen, so verwundert es aufgrund der MCAR-Annahme des fallweisen Ausschlusses kaum, dass gravierende Abweichungen zu den „wahren“ Ergebnissen mit vollständigen Daten entstanden.

Die Multiple Imputation setzt als Verfahren die Annahme voraus, dass der Datenausfall zwar systematisch ist, diese Systematik aber durch die beobachteten Daten erfasst wird. Beginnend mit der Situation schwacher Selektivität, lieferte die Multiple Imputation bei fehlenden Werten in einer oder mehreren unabhängigen Variablen entweder leicht bessere oder ähnlich gute Ergebnisse wie der fallweise

Ausschluss. War die abhängige Variable eine Prozessdatenvariable, führte die Verwendung der Multiplen Imputation allerdings zu einer leichten Verschlechterung der Schätzung. Insgesamt waren in allen Szenarien die Abweichungen von Ergebnissen mit vollständigen Daten aber eher gering, so dass sich das Verfahren im Großen und Ganzen ähnlich gut für die Ausfallkorrektur geeignet zeigte, wie der fallweise Ausschluss. Bei der Analyse mit starker Selektivität erwies sich die Imputation als das einzige unter den getesteten Korrekturverfahren, das für den interessierenden Koeffizienten eine Schätzung lieferte, welche das Ergebnis aus den vollständigen Daten reproduzieren konnte. Dies gelang bei fehlenden Werten in einer, mehreren und sogar allen Kontrollvariablen, allerdings aber gerade nicht für die Ausfälle in der Y-Variable, wo sich das Ergebnis bei Imputation kaum von dem beim fallweisen Ausschluss unterschied.

Für die Heckman-Korrektur ist trotz der MNAR-Annahme zu beobachten, dass sie sowohl beim Vorliegen schwacher als auch starker Selektivität keine wesentliche Verbesserung im Vergleich zum fallweisen Ausschluss erzielte. Letzteres bedeutet wiederum, dass dort aufgetretene Verzerrungen nicht ausgeglichen wurden. Dass die Heckman-Korrektur die Schätzung im Vergleich zum fallweisen Ausschluss nicht verbessert, mag darin begründet sein, dass sie zwar die geringsten Annahmen über den Ausfallprozess macht, aber sehr strenge Anforderungen an die Verteilung der Störgrößen in Selektions- und Analysegleichung (bivariate Normalverteilung) stellt sowie auf Instrumente angewiesen ist. Womöglich sind die verwendeten Instrumente im vorliegenden Fall nicht ausreichend und/oder die Verteilungsannahmen sind nicht gerechtfertigt.

Da es sich um eine Fallstudie handelt, können diese Ergebnisse keine uneingeschränkte Verallgemeinerbarkeit beanspruchen. Zum einen geht es hier nur um den Ausfallmechanismus „keine Zustimmung zur Datenverknüpfung“ und nicht um einen generellen Vergleich der Leistungsfähigkeit der Missing Data Verfahren. Zum andern kann in anderen Studien, die Prozess- und Befragungsdaten verknüpfen, der Ausfallprozess eventuell besser oder auch schlechter mit den verfügbaren Informationen erfasst werden, als mit den hier verwendeten Daten. Dies kann natürlich die Genauigkeit beeinflussen, mit der es den Missing Data Verfahren gelingt, den verknüpfungsbedingten Datenausfall zu korrigieren. So kann sich die Leistungsfähigkeit der hier untersuchten Verfahren in anderen Forschungszusammenhängen von der hier Berichteten natürlich unterscheiden. Allerdings ist anzunehmen, dass die allgemeinen empirischen Determinanten des Zustimmungsverhaltens zur Datenverknüpfung – und damit des Datenausfalls – auch in anderen Befragungen ähnlich wirksam sind, und damit die Selektivität der Stichprobe nach Zusammenspielung ähnlich sein wird (unabhängig davon, ob Determinanten nun beobachtete

oder unbeobachtete Einflüsse darstellen). Daher wird diese Fallstudie einige wichtige Hinweise für die Anwender liefern können, die in einer konkreten Forschungsarbeit beide Datenquellen nutzen wollen und vor der Wahl eines entsprechenden Korrekturverfahrens stehen.

Schließlich ist noch zu diskutieren, ob die Anwendung der Korrekturverfahren *in dieser Fallstudie* nicht noch in der Hinsicht verbessert werden könnte, so dass sie bessere Ergebnisse liefert. Eindeutig (und negativ) ist die Antwort lediglich beim fallweisen Ausschluss, da hier keine weiteren Varianten möglich sind. Für die Multiple Imputation stellt sich diese Frage lediglich bei fehlenden Werten in der abhängigen Variable. Hier fällt auf, dass das Ergebnis der Korrektur sowohl bei schwacher als auch starker Selektivität zu wünschen übrig lässt. Ob der Grund hierfür in der Verletzung der MAR-Annahme lag, oder ob es sich um ein generelles Problem bei der Imputation fehlender Werte in abhängigen Variablen einer Regression handelt (z. B. Hippel 2007)¹⁸, kann hier nicht abschließend geklärt werden. Für ein besseres Abschneiden der Heckman-Korrektur könnten eventuell mehr oder bessere Instrumentvariablen helfen. Dies setzt allerdings voraus, dass es möglich ist, trotz der vorangegangenen intensiven Auseinandersetzung mit Determinanten des Datenausfalles (Hartmann/Krug 2009) noch weitere Determinanten zu finden, die zudem noch die Instrumenteneigenschaft aufweisen. Zumindest in den vorliegenden Daten ist dies wohl wenig wahrscheinlich. Hinzu kommt, dass laut Wilde (2000) die bivariate Variante des Heckmanverfahrens (Analysemodell 2) in vielen Fällen auch ohne Instrumentvariablen funktionieren sollte. In dem Fall bleibt nur der Schluss, dass in der vorliegenden Analyse die verfahrensbedingten Annahmen über die unbeobachtete Heterogenität (bivariate Normalverteilung, Unkorreliertheit mit den enthaltenen Variablen) nicht zutreffen. Dass diese sehr restriktiv sind und ihre Verletzung problematisch für die Qualität der Heckman-Korrektur ist, dazu vergleiche etwa Winship/Mare (1992), zum Umgang mit Annahmeverletzungen siehe z. B. Newey (2009).

18 Hippel zeigt, dass im Rahmen einer Multiplen Imputation solche Beobachtungen, welche ausschließlich in der Y-Variable fehlende Werte aufweisen, zunächst keinerlei Information zur Ausfallkorrektur beitragen, sondern die Schätzung der Koeffizienten lediglich mit statistischem Rauschen versehen. Zusätzliche Information enthalten die imputierten Werte allerdings – wie in der vorliegenden Analyse – durch die a)- und b)-Variablen (Tabelle 5; vgl. Hippel 2007: 108ff.), so dass das schlechte Abschneiden der Multiplen Imputation in Szenario 1 nur teilweise durch die Ergebnisse von Hippel erklärbar ist.

5 Schluss

Die Verknüpfung von Informationen über ein Individuum aus verschiedenen Quellen (Record Linkage) gewinnt in der empirischen Sozialforschung zunehmend an Bedeutung (z. B. Schnell/Bachteler/Reiher 2009), bedingt unter anderem auch durch die verstärkte Aufbereitung von Prozessdaten für wissenschaftliche Zwecke. Aus Datenschutzgründen sind solche Verknüpfungen, etwa die von Befragungsdaten mit Prozessdaten, oft zustimmungspflichtig, so dass es hier zu Datenausfällen kommen kann. In der vorliegenden Fallstudie wurden die Resultate mehrerer Verfahren beim Umgang mit fehlenden Werten im Rahmen einer zustimmungspflichtigen Datenverknüpfung empirisch miteinander verglichen.

Fasst man die Ergebnisse des Verfahrensvergleiches zusammen, so sprechen diese – unter Berücksichtigung der stets eingeschränkten Verallgemeinerbarkeit solcher Fallstudien – dafür, bei Bedarf eine Anreicherung von Befragungsdaten mit Prozessdaten vorzunehmen, auch wenn es durch die Zustimmungspflichtigkeit zu Datenausfällen kommt. Ein Grund ist, dass nach bisherigen Erkenntnissen der Fall starker Korrelationen von Analysevariablen mit Zustimmungsdeterminanten wohl eher selten ist, da es nur wenige solcher Determinanten zu geben scheint. Daher werden viele Analysen von nur schwacher Selektivität betroffen sein und in diesem Fall sind – bei sachgerechter Anwendung – alle hier behandelten Korrekturverfahren ähnlich gut geeignet. Geht ein Anwender trotz schwacher Selektivität von starkem Ausfall und entweder beobachteten oder zusätzlichen unbeobachteten Ausfalldeterminanten aus (und verwendet entsprechend komplexere Korrekturverfahren), so begeht er laut dieser Fallstudie wohl keinen Fehler.

Anders ist es allerdings, wenn der wohl eher seltenere Fall der starken Selektivität vorliegt. Dies ist etwa der Fall, wenn Aspekte der Datenqualität im Zentrum der Untersuchung stehen, wie etwa im Analysemodell 2. In diesem Fall riskieren Forscher mit der Verwendung des fallweisen Ausschlusses einerseits und der Heckman-Korrektur andererseits – zumindest unter zu der vorliegenden Untersuchung ähnlichen Bedingungen – falsche inhaltliche Schlussfolgerungen aus ihren Analysen. Dies gilt bei der Verwendung von ausfallbelasteten Prozessdatenvariablen als abhängige Regressionsvariablen auch für die Multiple Imputation. Nutzt ein Anwender eine oder mehrere Prozessdatenvariablen als Kontrollvariablen, so scheint die Multiple Imputation ein Verfahren, welches – wiederum unter zu der vorliegenden Untersuchung ähnlichen Bedingungen –, helfen kann, falsche Schlussfolgerungen zu vermeiden. Je weiter natürlich die Bedingungen einer Analyse mit verknüpften Daten von denen in dieser Fallstudie abweichen, etwa hinsichtlich des Analysemodells oder des Themas der Befragung, desto weniger aussagekräftig sind die hier angeführten Ergebnisse.

Literatur

- Allmendinger, J. und A. Kohlmann, 2005: Datenverfügbarkeit und Datenzugang am Forschungsdatenzentrum der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung. *Allgemeines Statistisches Archiv* 88: 159-182.
- Bodner, T. E., 2008: What Improves With Increased Missing Data Imputations? *Structural Equation Modeling* 15(4): 651-675.
- Collins, L. M., J. L. Schafer und C. M. Kam, 2001: A Comparison of Inclusive and Restrictive Missing-Data Strategies in Modern Missing-Data Procedures. *Psychological Methods* 6: 330-351.
- Engelhardt, H., 1999: Lineare Regression mit Selektion: Möglichkeiten und Grenzen der Heckman-Korrektur. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 51: 706-723.
- Graham, J. W., A. E. Olchowski und T. D. Gilreath, 2007: How Many Imputations Are Really Needed? – Some Practical Clarifications of Multiple Imputation Theory. *Prevention Science* 8(3): 206-213.
- Graham, J. W., 2009: Missing Data Analysis: Making It Work in the Real World. *Annual Review of Psychology* 60: 549-576.
- Hartmann, J., 2004: Repräsentative Erhebung zur Evaluation des Mainzer Modells. S. 49-139 in: T. Gewiese, J. Hartmann, G. Krug, und H. Rudolph (Hg.): *Das Mainzer Modell aus Sicht der Arbeitnehmer und Betriebe. Befunde aus der Begleitforschung*. Nürnberg: IAB. <http://doku.iab.de/externe/2004/k040823w09.pdf> (04.05.2009).
- Hartmann, J., K. Brink, R. Jäckle und N. Tschersich, 2008: IAB-Haushaltspanel im Niedrigeinkommensbereich – Methoden- und Feldbericht. *FDZ Methodenreport 7/2008* Nürnberg: IAB. http://doku.iab.de/fdz/reporte/2008/MR_07-08.pdf (04.05.2009).
- Hartmann, J., A. Holleder, B. Kaltenborn, H. Rudolph, A. Vanselow, C. Weinkopf, und E. Wiedemann, 2002: Vom arbeitsmarktpolitischen Sonderprogramm CAST zur bundesweiten Erprobung des Mainzer Modells. 2. Zwischenbericht des Forschungsverbunds „Evaluierung CAST“. *BMWA-Dokumentation* Nr. 516.
- Hartmann, J. und G. Krug, 2009: Verknüpfung von personenbezogenen Prozess- und Befragungsdaten. Selektivität durch fehlende Zustimmung der Befragten? *Zeitschrift für ArbeitsmarktForschung* 42(2): 121-139.
- Heckman, J. J., 1979: Sample Selection Bias as a Specification Error. *Econometrica* 47(1): 153-162.
- Hippel, P. T., 2007: Regression With Missing Ys: An Improved Strategy for Analyzing Multiply-Imputed Data. *Sociological Methodology* 37(1): 83-117.
- Horton N. J. und K. P. Kleinman, 2007: Much Ado About Nothing: A Comparison of Missing Data Methods and Software to Fit Incomplete Data Regression Models. *The American Statistician* 61: 79-90.
- Horton, N. J. und S. R. Lipsitz, 2001: Multiple Imputation In Practice: Comparison of Software Packages for Regression Models With Missing Variables. *American Statistician* 55: 244-254.
- Kaltenborn, B., G. Krug, H. Rudolph, C. Weinkopf und E. Wiedemann, 2005: Evaluierung der arbeitsmarktpolitischen Sonderprogramme CAST und Mainzer Modell. *Bundesministerium für Wirtschaft und Arbeit. Forschungsbericht* Nr. 552, Berlin.
- Krug, G., 2009: In-Work Benefits For Low-Wage Jobs. Can Additional Income Reduce Employment Stability? *European Sociological Review* 25(4): 459-474.
- Little, R. J. A. und D. B. Rubin, 1987: *Statistical Analysis with Missing Data*. New York: Wiley.
- Newey, W. K., 2009: Two-Step Series Estimation of Sample Selection Models. *Econometrics Journal* 12: 217-229.
- Puhani, P. A., 2000: The Heckman Correction for Sample Selection and Its Critique. *Journal of Economic Surveys* 14(1): 53-68.

- Raghunathan, T. E., J. M. Lopkowski, J. van Hoeweyk und P. Solenberger, 2001: A Multivariate Technique for Multiply Imputing Missing Values Using a Series of Regression Models. *Survey Methodology* 27: 85-96.
- Raghunathan, T. E., P. Solenberger und J. van Hoeweyk, 2002: IVEware: Imputation and Variance Estimation Software. User Guide. <http://www.isr.umich.edu/src/smp/ive> (04.05.2009).
- Schnell, R., T. Bachteler und J. Reiher, 2009: Entwicklung einer neuen fehlertoleranten Methode bei der Verknüpfung von personenbezogenen Datenbanken unter Gewährleistung des Datenschutzes. *MDA – Methoden, Daten, Analysen. Zeitschrift für Empirische Sozialforschung* 3(2): 203-217. http://www.gesis.org/fileadmin/upload/forschung/publikationen/zeitschriften/mda/Vol.3_Heft_2/06_Schnell_et_al.pdf (14.4.2010).
- Rässler, S., 2002: *Statistical Matching. A Frequentist Theory, Practical Applications and Alternative Bayesian Approaches*. New York, Berlin: Springer.
- Ridder, G., 1992: An Empirical Evaluation of Some Models For Non-Random Attrition In Panel Data. *Structural Change and Economic Dynamics* 3(2): 337-355.
- Rubin, D. B., 1976: Inference With Missing Data (With Discussion). *Biometrika* 63: 581-592.
- Rubin, D. B., 1987: *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley.
- Schafer, J. L., 1999a: *Analysis of Incomplete Multivariate Data*. London, u. a.: Chapman & Hall/CRC.
- Schafer, J. L., 1999b: Multiple Imputation: A Primer. *Statistical Methods in Medical Research* 8(1): 3-15.
- Schafer, J. L. und J. W. Graham, 2002: Missing Data: Our View of the State of the Art. *Psychological Methods* 7(2): 147-177.
- Weins, C., 2006: Multiple Imputation. S. 205-216 in: J. Behnke, V. Gschwend, D. Schindler, und K.-U. Schnapp (Hg.): *Methoden der Politikwissenschaft. Neuere quantitative Methoden Analyseverfahren*. Baden-Baden: Nomos.
- Wilde, J., 2000: Identification of Multiple Equation Probit Models With Endogenous Dummy Regressors. *Economics Letters* 69: 309-312.
- Winship, C. und R. D. Mare, 1992: Models for Sample Selection Bias. *Annual Review of Sociology* 18: 327-50.
- Wirth, H. und W. Müller, 2004: Mikrodaten der amtlichen Statistik als eine Datengrundlage der empirischen Sozialforschung. S. 93-127 in: A. Diekmann (Hg.): *Methoden der Sozialforschung (Sonderheft 44 der Kölner Zeitschrift für Soziologie und Sozialpsychologie)*. Wiesbaden: VS Verlag für Sozialwissenschaften.

Anschrift des Autors

Dr. Gerhard Krug
Institut für Arbeitsmarkt- und Berufsforschung (IAB)
der Bundesagentur für Arbeit (BA)
Weddigenstr. 20-22
90478 Nürnberg
Gerhard.Krug@iab.de

Anhang

Tabelle 5 Zusätzliche Variablen im Imputationsmodell

	Analysefall 1	Analysefall 2
Suchanstrengungen (Zahl der Suchwege)	a)-Variable	Analysevariable
Bereits einmal Stelle wegen zu niedrigem Einkommen abgelehnt (Dummy)	a)-Variable	Analysevariable
Wie wurde die Stelle gefunden (Arbeitsvermittlung; Bekannte/Freunde; Eigene Initiative; Sonstiges)	b)-Variable	b)-Variable
Sozialhilfebezug vor Beschäftigungsaufnahme (Dummy)	b)-Variable	Analysevariable
Von ABM in Beschäftigung (Dummy)	a)-Variable	Analysevariable
Von and. Maßnahme in Beschäftigung (Dummy)	a)-Variable	a)-Variable
Vorher Weiterbildung (Dummy)	a)-Variable	Analysevariable
Vorher Lohnersatzleistungen vom Arbeitsamt (Arbeitslosengeld; Arbeitslosenhilfe; Keine)	a)-Variable	Analysevariable
Beschäftigung befristet (Dummy)	a)-Variable	a)-Variable
Bruttolohn in Euro (falls Angabe)	b)-Variable	a)-Variable
Bruttolohn: Angabe nicht verweigert (Dummy)	b)-Variable	Analysevariable
Geschlecht weiblich (Dummy)	b)-Variable	Analysevariable
Zufrieden mit Lohn (Dummy)	a)-Variable	Analysevariable
Zeit in (geförderter) Beschäftigung bis Interviewzeitpunkt	a)-Variable	a)-Variable
Beschäftigung zum Interviewzeitpunkt beendet?	a)-Variable	Analysevariable
Nationalität deutsch (Dummy)	b)-Variable	b)-Variable
Berufserfahrung in Jahren	a)-Variable	a)-Variable
Nettohaushaltseinkommen	a)-Variable	Analysevariable

a)-Variable: Variable hängt potentiell mit der ausfallbelasteten Variable zusammen;

b)-Variable: Variable hängt potentiell mit dem Ausfall zusammen;

Analysevariable: Variable bereits im Analysemodell enthalten (vgl. Schafer 1999a: 143).

Anmerkung: Um das Vorgehen übersichtlich und verständlich zu halten, wird in der Untersuchung ein einziges Imputationsmodell für alle Szenarien verwendet. Dieses Modell wurde ursprünglich lediglich für die Imputation der Variable „Arbeitslosigkeitsdauer“ aufgestellt (Szenario 2). Grundsätzlich müssten mit jeder neuen ausfallbelasteten Variable aber weitere a)-Variablen hinzugefügt werden, um die MAR-Annahme zu stützen. Das Imputationsmodell scheint sich aber – so zeigen die positiven Ergebnisse – auch ohne Erweiterung der a)-Variablen für die Szenarien 3 und 4 zu eignen. Für Szenario 1 gibt es hingegen definitionsgemäß keine a)-Variablen, da diese bereits im Analysemodell enthalten sein sollten.

Anhang A Analysemodell 1 (schwache Selektivität)

Tabelle 6 Ausfallkorrektur durch fallweisen Ausschluss (alle Szenarien) - KQ-Regressionen der wöchentlichen Arbeitszeit in Stunden

	vollständige Daten	Szenario 1-4
	Koeff. (Std.Fehler)	Koeff. (Std.Fehler)
Kombilohnförderung: ja	-4,769*** (0,550)	-4,462*** (0,632)
Kumulierte Dauer der Arbeitslosigkeit (Ref.: unter einem Monat)		
Ein bis unter sechs Monate	-0,730 (0,591)	-1,156 (0,667)
Sechs bis unter zwölf Monate	-1,178 (0,656)	-1,475* (0,737)
Zwölf bis unter vierundzwanzig Monate	-2,165** (0,682)	-2,030** (0,777)
Über vierundzwanzig Monate	-1,685* (0,818)	-2,222* (0,923)
Alter	0,282 (0,173)	0,110 (0,197)
Alter (quadr.)	-0,004 (0,002)	-0,002 (0,003)
Stellung im Beruf (Ref.: Arbeiter)		
Angestellter	-3,212*** (0,456)	-3,874*** (0,520)
Sonstiges	-2,476* (1,152)	-3,456** (1,339)
Branche (Ref.: Leih-/Zeitarbeitsfirma)		
Reinigungsgewerbe	-9,196*** (0,940)	-7,989*** (1,086)
Hotel- und Gaststättengewerbe	-2,694** (0,961)	-2,741* (1,095)
Callcenter	-4,573*** (1,202)	-3,898** (1,367)
Energie- und Wasser, Bergbau, Verarbeitendes Gewerbe	0,102 (0,774)	-0,104 (0,867)
Handel	-4,319*** (0,739)	-4,455*** (0,827)
Verkehr, Nachrichtenübermittlung, Banken, Versicherungen, Öffentlicher Dienst, Sozialversicherung	-4,230*** (0,870)	-3,883*** (0,981)
Bereich andere Dienstleistungen, k. A.	-2,996*** (0,674)	-2,677*** (0,757)
Haushaltskontext (Ref.: Alleinstehend)		
Alleinerziehend	-2,211** (0,823)	-2,388* (0,950)
Nicht erwerbstätiger Partner, ohne Kind	1,424 (0,879)	1,205 (1,031)
Erwerbstätiger Partner, ohne Kind	2,351** (0,892)	2,135* (1,020)
Nicht erwerbstätiger Partner, Kind	3,204*** (0,881)	3,461*** (1,010)
Erwerbstätiger Partner, Kind	-1,260 (0,905)	-1,639 (1,038)

	vollständige Daten	Szenario 1-4
	Koeff. (Std.Fehler)	Koeff. (Std.Fehler)
Zahl der Kinder (unter 18 Jahren im Haushalt)	0,101 (0,293)	0,174 (0,332)
Nebentätigkeit vorhanden	-3,452*** (1,009)	-3,324** (1,156)
Einkommen sehr wichtig	0,333 (0,432)	0,375 (0,491)
Arbeit sehr wichtig	0,954* (0,437)	0,958 (0,499)
Freizeit sehr wichtig	-0,242 (0,483)	0,264 (0,545)
Familie sehr wichtig	-1,080* (0,532)	-1,486* (0,608)
Region: Ost	4,718*** (0,444)	4,598*** (0,505)
Art der Kündigung letzte Beschäftigung (Ref.: Sonstiges)		
Arbeitnehmerkündigung	1,558 (0,825)	1,545 (0,909)
Arbeitgeberkündigung	1,423** (0,453)	1,645** (0,519)
Konstante	33,678*** (3,027)	36,712*** (3,479)
Adjusted R ²	0,2125	0,2305
N	2605	1940

* $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$

Grau unterlegt: Koeffizienten ausfallbelasteter Variablen.

Tabelle 7 Ausfallkorrektur durch Multiple Imputation nach Szenarien – KQ-Regressionen der wöchentlichen Arbeitszeit in Stunden

	Vollständige Daten	Szenario 1	Szenario 2	Szenario 3	Szenario 4
	Koeff. (Std.Fehler)	Koeff. (Std.Fehler)	Koeff. (Std.Fehler)	Koeff. (Std.Fehler)	Koeff. (Std.Fehler)
Kombilohnförderung: ja	-4,769*** (0,550)	-4,135*** (0,660)	-4,752*** (0,551)	-4,349*** (0,674)	-4,455*** (0,577)
Kumulierte Dauer der Arbeitslosigkeit (Ref.: unter einem Monat)					
Ein bis unter sechs Monate	-0,730 (0,591)	-0,929 (0,598)	-0,916 (0,671)	-0,880 (0,592)	-0,873 (0,596)
Sechs bis unter zwölf Monate	-1,178 (0,656)	-1,199 (0,692)	-1,504* (0,725)	-1,252 (0,658)	-1,163 (0,682)
Zwölf bis unter vierundzwanzig Monate	-2,165** (0,682)	-1,708* (0,700)	-2,351** (0,801)	-2,244** (0,688)	-2,274** (0,700)
Über vierundzwanzig Monate	-1,685* (0,818)	-1,737* (0,826)	-2,309* (0,913)	-1,926* (0,820)	-1,863* (0,831)
Alter	0,282 (0,173)	0,206 (0,197)	0,272 (0,173)	0,054 (0,181)	0,127 (0,192)
Alter (quadr.)	-0,004 (0,002)	-0,003 (0,003)	-0,004 (0,002)	-0,001 (0,002)	-0,002 (0,002)
Stellung im Beruf (Ref.: Arbeiter)					
Angestellter	-3,212*** (0,456)	-3,867*** (0,523)	-3,213*** (0,456)	-4,050*** (0,509)	-3,897*** (0,472)
Sonstiges	-2,476* (1,152)	-2,964 (1,526)	-2,581* (1,152)	-1,922 (1,888)	-0,632 (1,128)
Branche (Ref.: Leih-/Zeitarbeitsfirma)					
Reinigungsgewerbe	-9,196*** (0,940)	-8,087*** (1,194)	-9,121*** (0,942)	-7,891*** (1,050)	-7,757*** (1,127)
Hotel- und Gaststättengewerbe	-2,694** (0,961)	-2,853** (1,087)	-2,650** (0,962)	-2,859* (1,206)	-3,033** (1,092)
Callcenter	-4,573*** (1,202)	-3,343** (1,265)	-4,543*** (1,203)	-4,136** (1,329)	-4,013*** (1,163)
Energie- und Wasser, Bergbau, Verarbeitendes Gewerbe	0,102 (0,774)	-0,183 (0,809)	0,134 (0,774)	0,345 (1,142)	0,083 (0,879)
Handel	-4,319*** (0,739)	-4,298*** (0,789)	-4,251*** (0,740)	-4,313*** (0,796)	-4,500*** (0,812)
Verkehr, Nachrichtenübermittlung, Banken, Versicherungen, Öffentli- cher Dienst, Sozialversicherung	-4,230*** (0,870)	-3,734*** (0,914)	-4,189*** (0,871)	-3,597** (1,153)	-4,120*** (1,175)
Bereich andere Dienstleistungen, k. A.	-2,996*** (0,674)	-2,543*** (0,705)	-2,962*** (0,674)	-2,609*** (0,743)	-2,714*** (0,797)
Haushaltskontext (Ref.: Alleinstehend)					
Alleinerziehend	-2,211** (0,823)	-2,366** (0,870)	-2,195** (0,826)	-2,117* (0,906)	-2,199* (0,860)
Nicht erwerbstätiger Partner, ohne Kind	1,424 (0,879)	1,353 (0,931)	1,417 (0,880)	0,906 (1,040)	1,329 (1,013)
Erwerbstätiger Partner, ohne Kind	2,351** (0,892)	2,322* (0,969)	2,332** (0,892)	2,250 (1,169)	2,401* (1,117)
Nicht erwerbstätiger Partner, Kind	3,204*** (0,881)	3,414*** (0,979)	3,224*** (0,883)	3,603*** (1,065)	3,174*** (0,856)
Erwerbstätiger Partner, Kind	-1,260 (0,905)	-1,405 (0,936)	-1,254 (0,907)	-1,430 (0,951)	-1,517 (0,981)

	Vollständige Daten	Szenario 1	Szenario 2	Szenario 3	Szenario 4
	Koeff. (Std.Fehler)	Koeff. (Std.Fehler)	Koeff. (Std.Fehler)	Koeff. (Std.Fehler)	Koeff. (Std.Fehler)
Zahl der Kinder (unter 18 Jahren im Haushalt)	0,101 (0,293)	0,007 (0,310)	0,100 (0,293)	0,030 (0,398)	0,144 (0,266)
Nebentätigkeit vorhanden	-3,452*** (1,009)	-2,462* (1,009)	-3,461*** (1,012)	-3,066** (1,176)	-4,093** (1,475)
Einkommen sehr wichtig	0,333 (0,432)	0,361 (0,428)	0,365 (0,432)	0,305 (0,440)	0,122 (0,491)
Arbeit sehr wichtig	0,954* (0,437)	1,016* (0,442)	0,941* (0,438)	1,089* (0,447)	1,096* (0,468)
Freizeit sehr wichtig	-0,242 (0,483)	0,010 (0,574)	-0,256 (0,483)	-0,180 (0,485)	-0,119 (0,476)
Familie sehr wichtig	-1,080* (0,532)	-1,314* (0,611)	-1,073* (0,532)	-1,216* (0,545)	-1,241 (0,648)
Region: Ost	4,718*** (0,444)	4,439*** (0,471)	4,772*** (0,446)	4,729*** (0,447)	4,758*** (0,495)
Art der Kündigung letzte Beschäftigung (Ref.: Sonstiges)					
Arbeitnehmerkündigung	1,558 (0,825)	1,766* (0,847)	1,485 (0,826)	1,453 (0,826)	1,099 (0,985)
Arbeitgeberkündigung	1,423** (0,453)	1,865*** (0,454)	1,353** (0,455)	1,400** (0,460)	1,752*** (0,486)
Konstante	33,678*** (3,027)	34,674*** (3,553)	33,988*** (3,034)	37,564*** (3,134)	36,350*** (3,511)
Adjusted R ²	0,2125	n.v.	n.v.	n.v.	n.v.
N	2605	2605	2605	2605	2605

* $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$; n.v.: nicht verfügbar

Standardfehler basieren auf Rubin 1987; Variablen des Imputationsmodells siehe Tabelle 5;

Grau unterlegt: Koeffizienten ausfallbelasteter Variablen.

Tabelle 8a Ausfallkorrektur durch Heckman-Korrektur (KQ) nach Szenarien – KQ-Regressionen der wöchentlichen Arbeitszeit in Stunden

	Vollständige Daten	Szenario 1	Szenario 2	Szenario 3	Szenario 4
	Koeff. (Std.Fehler)	Koeff. (Std.Fehler)	Koeff. (Std.Fehler)	Koeff. (Std.Fehler)	Koeff. (Std.Fehler)
Kombilohnförderung: ja	-4,769*** (0,550)	-4,402*** (0,641)	-4,370*** (0,650)	-4,342*** (0,663)	-4,296*** (0,685)
Kumulierte Dauer der Arbeitslosigkeit (Ref.: unter einem Monat)					
Ein bis unter sechs Monate	-0,730 (0,591)	-1,213 (0,675)	-1,147 (0,661)	-1,152 (0,661)	-1,149 (0,661)
Sechs bis unter zwölf Monate	-1,178 (0,656)	-1,475* (0,740)	-1,470* (0,732)	-1,474* (0,731)	-1,471* (0,732)
Zwölf bis unter vierundzwanzig Monate	-2,165** (0,682)	-2,197** (0,825)	-2,016** (0,771)	-2,022** (0,771)	-2,015** (0,771)
Über vierundzwanzig Monate	-1,685* (0,818)	-2,267* (0,928)	-2,221* (0,915)	-2,224* (0,915)	-2,220* (0,925)
Alter	0,282 (0,173)	0,097 (0,199)	0,093 (0,199)	0,106 (0,195)	0,106 (0,195)
Alter (quadr.)	-0,004 (0,002)	-0,001 (0,003)	-0,001 (0,003)	-0,002 (0,003)	-0,002 (0,003)
Stellung im Beruf (Ref.: Arbeiter)					
Angestellter	-3,212*** (0,456)	-3,926*** (0,529)	-3,937*** (0,531)	-3,877*** (0,516)	-3,877*** (0,516)
Sonstiges	-2,476* (1,152)	-3,430* (1,343)	-3,435* (1,344)	-3,479** (1,328)	-3,491** (1,329)
Branche (Ref.: Leih-/Zeitarbeitsfirma)					
Reinigungsgewerbe	-9,196*** (0,940)	-8,303*** (1,202)	-8,330*** (1,213)	-7,990*** (1,078)	-7,987*** (1,078)
Hotel- und Gaststättengewerbe	-2,694** (0,961)	-2,907* (1,131)	-2,925* (1,136)	-2,730* (1,087)	-2,733* (1,087)
Callcenter	-4,573*** (1,202)	-4,140** (1,426)	-4,162** (1,433)	-3,912** (1,356)	-3,908** (1,356)
Energie- und Wasser, Bergbau, Verarbeitendes Gewerbe	0,102 (0,774)	-0,163 (0,875)	-0,169 (0,877)	-0,091 (0,861)	-0,082 (0,861)
Handel	-4,319*** (0,739)	-4,550*** (0,844)	-4,560*** (0,847)	-4,460*** (0,820)	-4,454*** (0,820)
Verkehr, Nachrichtenübermittlung, Banken, Versicherungen, Öffentli- cher Dienst, Sozialversicherung	-4,230*** (0,870)	-4,024*** (1,010)	-4,037*** (1,014)	-3,877*** (0,973)	-3,870*** (0,973)
Bereich andere Dienstleistungen, k. A.	-2,996*** (0,674)	-2,877*** (0,826)	-2,893*** (0,832)	-2,672*** (0,751)	-2,664*** (0,751)
Haushaltskontext (Ref.: Alleinstehend)					
Alleinerziehend	-2,211** (0,823)	-2,014 (1,129)	-1,991 (1,138)	-2,349* (0,944)	-2,353* (0,944)
Nicht erwerbstätiger Partner, ohne Kind	1,424 (0,879)	1,229 (1,033)	1,239 (1,035)	1,203 (1,022)	1,194 (1,022)
Erwerbstätiger Partner, ohne Kind	2,351** (0,892)	2,398* (1,108)	2,411* (1,111)	2,127* (1,012)	2,123* (1,012)
Nicht erwerbstätiger Partner, Kind	3,204*** (0,881)	3,873** (1,212)	3,895** (1,220)	3,469*** (1,001)	3,460*** (1,001)
Erwerbstätiger Partner, Kind	-1,260 (0,905)	-1,309 (1,170)	-1,295 (1,172)	-1,665 (1,030)	-1,669 (1,030)

	Vollständige Daten	Szenario 1	Szenario 2	Szenario 3	Szenario 4
	Koeff. (Std.Fehler)	Koeff. (Std.Fehler)	Koeff. (Std.Fehler)	Koeff. (Std.Fehler)	Koeff. (Std.Fehler)
Zahl der Kinder (unter 18 Jahren im Haushalt)	0,101 (0,293)	0,174 (0,333)	0,173 (0,334)	0,180 (0,330)	0,180 (0,330)
Nebentätigkeit vorhanden	-3,452*** (1,009)	-3,351** (1,159)	-3,351** (1,160)	-3,330** (1,147)	-3,322** (1,146)
Einkommen sehr wichtig	0,333 (0,432)	0,302 (0,507)	0,304 (0,505)	0,314 (0,503)	0,373 (0,487)
Arbeit sehr wichtig	0,954* (0,437)	1,118* (0,564)	1,126* (0,566)	1,089* (0,547)	0,950 (0,495)
Freizeit sehr wichtig	-0,242 (0,483)	0,345 (0,562)	0,343 (0,561)	0,329 (0,557)	0,275 (0,541)
Familie sehr wichtig	-1,080* (0,532)	-1,480* (0,609)	-1,478* (0,610)	-1,413* (0,621)	-1,486* (0,603)
Region: Ost	4,718*** (0,444)	4,775*** (0,583)	4,763*** (0,569)	4,744*** (0,562)	4,610*** (0,501)
Art der Kündigung letzte Beschäftigung (Ref.: Sonstiges)					
Arbeitnehmerkündigung	1,558 (0,825)	1,776 (0,987)	1,793 (0,993)	1,774 (0,991)	1,507 (0,904)
Arbeitgeberkündigung	1,423** (0,453)	1,624** (0,521)	1,625** (0,521)	1,637** (0,520)	1,621** (0,516)
Konstante	33,678*** (3,027)	35,446*** (4,048)	35,395*** (4,055)	35,341*** (4,155)	35,379*** (4,037)
$\rho_{\varepsilon v}$		0,269	0,281	0,262	0,289
$\rho_{\varepsilon} \sigma_{\varepsilon}$		2,697 (4,379)	2,819 (4,419)	2,630 (4,436)	2,911 (4,573)
Adjusted R ²	0,2125	(n.v.)	(n.v.)	(n.v.)	(n.v.)
N	2605	1940	1940	1940	1940

* $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$; n.v.: nicht verfügbar; robuste Standardfehler
 Grau unterlegt: Koeffizienten ausfallbelasteter Variablen.

Tabelle 8b Ausfallkorrektur durch Heckman-Korrektur (KQ) nach Szenarien - Selektionsmodell (Probitregression des Datenausfalls)

	Szenario 1	Szenario 2	Szenario 3	Szenario 4
	Koeff. (Std.Fehler)	Koeff. (Std.Fehler)	Koeff. (Std.Fehler)	Koeff. (Std.Fehler)
Kombilohnförderung: ja	0,029 (0,074)	0,052 (0,073)	0,078 (0,063)	0,104 (0,062)
Kumulierte Dauer der Arbeitslosigkeit (Ref.: unter einem Monat)				
Ein bis unter sechs Monate	-0,054 (0,080)			
Sechs bis unter zwölf Monate	-0,004 (0,090)			
Zwölf bis unter vierundzwanzig Monate	-0,146 (0,092)			
Über vierundzwanzig Monate	-0,038 (0,111)			
Alter	-0,008 (0,023)	-0,010 (0,023)		
Alter (quadr.)	0,000 (0,000)	0,000 (0,000)		
Stellung im Beruf (Ref.: Arbeiter)				
Angestellter	-0,040 (0,062)	-0,046 (0,061)		
Sonstiges	0,051 (0,154)	0,044 (0,154)		
Branche (Ref.: Leih-/Zeitarbeitsfirma)				
Reinigungsgewerbe	-0,256* (0,126)	-0,266* (0,125)		
Hotel- und Gaststättengewerbe	-0,148 (0,130)	-0,155 (0,130)		
Callcenter	-0,194 (0,162)	-0,201 (0,162)		
Energie- und Wasser, Bergbau, Verarbeitendes Gewerbe	-0,069 (0,107)	-0,070 (0,107)		
Handel	-0,083 (0,102)	-0,086 (0,102)		
Verkehr, Nachrichtenübermittlung, Banken, Versicherungen, Öffentli- cher Dienst, Sozialversicherung	-0,128 (0,119)	-0,132 (0,119)		
Bereich andere Dienstleistungen, k. A.	-0,177 (0,093)	-0,181 (0,093)		
Haushaltskontext (Ref.: Alleinstehend)				
Alleinerziehend	0,260* (0,110)	0,263* (0,110)		
Nicht erwerbstätiger Partner, ohne Kind	0,021 (0,116)	0,026 (0,116)		
Erwerbstätiger Partner, ohne Kind	0,208 (0,119)	0,208 (0,119)		
Nicht erwerbstätiger Partner, Kind	0,325** (0,120)	0,325** (0,120)		
Erwerbstätiger Partner, Kind	0,274* (0,122)	0,273* (0,122)		
Zahl der Kinder (unter 18 Jahren im Haushalt)	-0,003 (0,040)	-0,004 (0,040)		
Nebentätigkeit vorhanden	-0,026 (0,135)	-0,023 (0,135)		

	Szenario 1	Szenario 2	Szenario 3	Szenario 4
	Koeff. (Std.Fehler)	Koeff. (Std.Fehler)	Koeff. (Std.Fehler)	Koeff. (Std.Fehler)
Einkommen sehr wichtig	-0,060 (0,058)	-0,055 (0,058)	-0,049 (0,057)	
Arbeit sehr wichtig	0,136* (0,059)	0,136* (0,059)	0,112 (0,058)	
Freizeit sehr wichtig	0,058 (0,066)	0,055 (0,065)	0,046 (0,065)	
Familie sehr wichtig	0,004 (0,071)	0,006 (0,071)	0,060 (0,068)	
Region: Ost	0,136* (0,062)	0,122* (0,061)	0,120* (0,058)	
Art der Kündigung letzte Beschäftigung (Ref.: Sonstiges)				
Arbeitnehmerkündigung	0,222 (0,116)	0,228* (0,116)	0,227* (0,115)	
Arbeitgeberkündigung	0,000 (0,061)	0,002 (0,061)	0,014 (0,060)	
<i>Instrumentvariablen</i>				
Keine deutsche Nationalität	-0,263* (0,104)	-0,249* (0,104)	-0,214* (0,100)	-0,243* (0,098)
Vor Beschäftigung Sozialhilfe: ja	0,196* (0,081)	0,202* (0,081)	0,220** (0,079)	0,185* (0,076)
Wie wurde die angetretene Stelle gefunden (Ref.: Arbeitsamt)				
Bekannte/Freunde	0,022 (0,079)	0,024 (0,079)	0,006 (0,078)	0,008 (0,078)
Eigene Initiative	-0,140 (0,074)	-0,137 (0,074)	-0,144 (0,074)	-0,147* (0,073)
Sonstige Suchwege	-0,062 (0,081)	-0,059 (0,081)	-0,055 (0,080)	-0,056 (0,080)
Konstante	0,616 (0,406)	0,615 (0,401)	0,476*** (0,092)	0,627*** (0,073)
N	2605	2605	2605	2605

* $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$; robuste Standardfehler
Grau unterlegt: Koeffizienten ausfallbelasteter Variablen.

Tabelle 9a Ausfallkorrektur durch Heckman-Korrektur (ML) nach Szenarien – ML-Schätzung der wöchentlichen Arbeitszeit in Stunden

	Vollständige Daten	Szenario 1	Szenario 2	Szenario 3	Szenario 4
	Koeff. (Std.Fehler)	Koeff. (Std.Fehler)	Koeff. (Std.Fehler)	Koeff. (Std.Fehler)	Koeff. (Std.Fehler)
Kombilohnförderung: ja	-4,769*** (0,550)	-4,270*** (0,682)	-4,180*** (0,680)	-4,131*** (0,664)	-4,041*** (0,665)
Kumulierte Dauer der Arbeitslosigkeit (Ref.: unter einem Monat)					
Ein bis unter sechs Monate	-0,730 (0,591)	-1,328 (0,721)	-1,103 (0,655)	-1,134 (0,656)	-1,119 (0,656)
Sechs bis unter zwölf Monate	-1,178 (0,656)	-1,504 (0,799)	-1,504* (0,726)	-1,527* (0,726)	-1,529* (0,727)
Zwölf bis unter vierundzwanzig Monate	-2,165** (0,682)	-2,537** (0,840)	-1,981** (0,764)	-2,011** (0,764)	-1,989** (0,764)
Über vierundzwanzig Monate	-1,685* (0,818)	-2,342* (0,998)	-2,206* (0,908)	-2,199* (0,908)	-2,194* (0,908)
Alter	0,282 (0,173)	0,065 (0,213)	0,057 (0,212)	0,083 (0,194)	0,079 (0,194)
Alter (quadr.)	-0,004 (0,002)	-0,001 (0,003)	-0,001 (0,003)	-0,001 (0,003)	-0,001 (0,003)
Stellung im Beruf (Ref.: Arbeiter)					
Angestellter	-3,212*** (0,456)	-4,016*** (0,562)	-4,037*** (0,561)	-3,865*** (0,513)	-3,862*** (0,513)
Sonstiges	-2,476* (1,152)	-3,629* (1,441)	-3,636* (1,440)	-4,067** (1,330)	-4,107** (1,332)
Branche (Ref.: Leih-/Zeitarbeitsfirma)					
Reinigungsgewerbe	-9,196*** (0,940)	-8,930*** (1,172)	-8,965*** (1,171)	-8,015*** (1,072)	-8,000*** (1,073)
Hotel- und Gaststättengewerbe	-2,694** (0,961)	-3,364** (1,185)	-3,389** (1,184)	-2,955** (1,081)	-2,973** (1,080)
Callcenter	-4,573*** (1,202)	-4,601** (1,477)	-4,625** (1,476)	-3,857** (1,351)	-3,869** (1,347)
Energie- und Wasser; Bergbau, Verarb. Gewerbe	0,102 (0,774)	-0,221 (0,940)	-0,234 (0,939)	0,127 (0,858)	0,137 (0,858)
Handel	-4,319*** (0,739)	-4,765*** (0,897)	-4,783*** (0,896)	-4,537*** (0,817)	-4,549*** (0,816)
Verkehr/Nachrichtenüberm., Ban- ken, Versicherungen, Öffentlicher Dienst/Sozialversicherung	-4,230*** (0,870)	-4,406*** (1,063)	-4,417*** (1,062)	-4,060*** (0,967)	-4,068*** (0,966)
Bereich andere Dienstleistungen, k. A.	-2,996*** (0,674)	-3,354*** (0,824)	-3,374*** (0,824)	-2,837*** (0,750)	-2,821*** (0,750)
Haushaltskontext (Ref.: Alleinstehend)					
Alleinerziehend	-2,211** (0,823)	-1,159 (1,032)	-1,153 (1,032)	-2,033* (0,941)	-2,026* (0,941)
Nicht erwerbstätiger Partner, ohne Kind	1,424 (0,879)	1,215 (1,105)	1,240 (1,104)	1,057 (1,017)	1,021 (1,017)
Erwerbstätiger Partner, ohne Kind	2,351** (0,892)	2,838** (1,101)	2,836** (1,100)	1,912 (0,995)	1,874 (0,995)
Nicht erwerbstätiger Partner, Kind	3,204*** (0,881)	4,679*** (1,095)	4,680*** (1,094)	3,381*** (0,995)	3,371*** (0,995)
Erwerbstätiger Partner, Kind	-1,260 (0,905)	-0,652 (1,123)	-0,661 (1,123)	-1,696 (1,027)	-1,682 (1,027)
Zahl der Kinder (unter 18 Jahren im Haushalt)	0,101 (0,293)	0,153 (0,360)	0,151 (0,359)	0,155 (0,330)	0,140 (0,330)
Nebentätigkeit vorhanden	-3,452*** (1,009)	-3,412** (1,246)	-3,404** (1,245)	-3,388** (1,139)	-3,363** (1,138)

	Vollständige Daten	Szenario 1	Szenario 2	Szenario 3	Szenario 4
	Koeff. (Std.Fehler)	Koeff. (Std.Fehler)	Koeff. (Std.Fehler)	Koeff. (Std.Fehler)	Koeff. (Std.Fehler)
Einkommen sehr wichtig	0,333 (0,432)	0,152 (0,531)	0,170 (0,530)	0,191 (0,528)	0,354 (0,485)
Arbeit sehr wichtig	0,954* (0,437)	1,429** (0,541)	1,423** (0,540)	1,329* (0,537)	0,936 (0,492)
Freizeit sehr wichtig	-0,242 (0,483)	0,537 (0,590)	0,519 (0,590)	0,477 (0,587)	0,366 (0,537)
Familie sehr wichtig	-1,080* (0,532)	-1,471* (0,655)	-1,467* (0,655)	-1,278* (0,649)	-1,523* (0,599)
Region: Ost	4,718*** (0,444)	5,160*** (0,548)	5,098*** (0,545)	5,061*** (0,540)	4,693*** (0,500)
Art der Kündigung letzte Beschäftigung (Ref.: Sonstiges)					
Arbeitnehmerkündigung	1,558 (0,825)	2,260* (0,993)	2,274* (0,992)	2,247* (0,988)	1,507 (0,893)
Arbeitgeberkündigung	1,423** (0,453)	1,600** (0,559)	1,604** (0,558)	1,638** (0,555)	1,668** (0,512)
Konstante	33,678*** (3,027)	33,134*** (3,763)	33,134*** (3,753)	33,181*** (3,456)	33,674*** (3,446)
Ath ($\rho_{\varepsilon v}$)		0,883*** (0,106)	0,876*** (0,107)	0,844*** (0,107)	0,860*** (0,104)
Ln (σ_{ε})		2,418*** (0,029)	2,416*** (0,029)	2,411*** (0,030)	2,415*** (0,029)
Adjusted R ²	0,2125	n.v.	n.v.	n.v.	n.v.
N	2605	2605	2605	2605	2605

* $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$; n.v.: nicht verfügbar; robuste Standardfehler
 Grau unterlegt: Koeffizienten ausfallbelasteter Variablen.

Tabelle 9b Ausfallkorrektur durch Heckman-Korrektur (ML) nach Szenarien – Selektionsmodell (Probitregression des Datenausfalls)

	Szenario 1 Koeff. (Std.Fehler)	Szenario 2 Koeff. (Std.Fehler)	Szenario 3 Koeff. (Std.Fehler)	Szenario 4 Koeff. (Std.Fehler)
Kombilohnförderung: ja	0,024 (0,073)	0,046 (0,072)	0,043 (0,063)	0,066 (0,062)
Kumulierte Dauer der Arbeitslosigkeit (Ref.: unter einem Monat)				
Ein bis unter sechs Monate	-0,058 (0,079)			
Sechs bis unter zwölf Monate	0,002 (0,088)			
Zwölf bis unter vierundzwanzig Monate	-0,144 (0,090)			
Über vierundzwanzig Monate	-0,035 (0,109)			
Alter	-0,005 (0,023)	-0,008 (0,023)		
Alter (quadr.)	0,000 (0,000)	0,000 (0,000)		
Stellung im Beruf (Ref.: Arbeiter)				
Angestellter	-0,049 (0,061)	-0,056 (0,061)		
Sonstiges	0,134 (0,154)	0,125 (0,154)		
Branche (Ref.: Leih-/Zeitarbeitsfirma)				
Reinigungsgewerbe	-0,251* (0,124)	-0,260* (0,124)		
Hotel- und Gaststättengewerbe	-0,099 (0,129)	-0,109 (0,129)		
Callcenter	-0,201 (0,159)	-0,209 (0,159)		
Energie- und Wasser; Bergbau, Verarb. Gewerbe	-0,101 (0,106)	-0,102 (0,105)		
Handel	-0,068 (0,101)	-0,072 (0,100)		
Verkehr/Nachrichtenüberm., Ban- ken, Versicherungen, Öffentlicher Dienst/Sozialversicherung	-0,085 (0,117)	-0,090 (0,116)		
Bereich andere Dienstleistungen, k. A.	-0,136 (0,092)	-0,140 (0,092)		
Haushaltskontext (Ref.: Alleinstehend)				
Alleinerziehend	0,196 (0,110)	0,202 (0,110)		
Nicht erwerbstätiger Partner, ohne Kind	0,040 (0,115)	0,045 (0,115)		
Erwerbstätiger Partner, ohne Kind	0,237* (0,118)	0,239* (0,118)		
Nicht erwerbstätiger Partner, Kind	0,334** (0,118)	0,336** (0,118)		
Erwerbstätiger Partner, Kind	0,265* (0,122)	0,265* (0,122)		
Zahl der Kinder (unter 18 Jahren im Haushalt)	0,006 (0,040)	0,004 (0,040)		
Nebentätigkeit vorhanden	-0,024 (0,133)	-0,022 (0,133)		

	Szenario 1 Koeff. (Std.Fehler)	Szenario 2 Koeff. (Std.Fehler)	Szenario 3 Koeff. (Std.Fehler)	Szenario 4 Koeff. (Std.Fehler)
Einkommen sehr wichtig	-0,056 (0,058)	-0,052 (0,058)	-0,048 (0,057)	
Arbeit sehr wichtig	0,132* (0,058)	0,132* (0,058)	0,108 (0,058)	
Freizeit sehr wichtig	0,045 (0,065)	0,042 (0,065)	0,033 (0,064)	
Familie sehr wichtig	0,012 (0,070)	0,013 (0,070)	0,066 (0,067)	
Region: Ost	0,111 (0,061)	0,099 (0,060)	0,107 (0,057)	
Art der Kündigung letzte Beschäftigung (Ref.: Sonstiges)				
Arbeitnehmerkündigung	0,204 (0,114)	0,211 (0,114)	0,212 (0,113)	
Arbeitgeberkündigung	-0,015 (0,060)	-0,014 (0,060)	-0,003 (0,059)	
<i>Instrumentvariablen</i>				
Keine deutsche Nationalität	-0,295** (0,093)	-0,283** (0,093)	-0,236** (0,091)	-0,261** (0,089)
Vor Beschäftigung Sozialhilfe: ja	0,180* (0,073)	0,184* (0,073)	0,200** (0,072)	0,174* (0,070)
Wie wurde die angetretene Stelle gefunden (Ref.: Arbeitsamt)				
Bekannte/Freunde	0,066 (0,072)	0,066 (0,072)	0,046 (0,071)	0,044 (0,071)
Eigene Initiative	-0,106 (0,067)	-0,105 (0,067)	-0,108 (0,067)	-0,112 (0,066)
Sonstige Suchwege	-0,034 (0,073)	-0,032 (0,073)	-0,021 (0,073)	-0,023 (0,072)
Konstante	0,545 (0,403)	0,546 (0,398)	0,486*** (0,090)	0,628*** (0,070)
N	2605	2605	2605	2605

* $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$; robuste Standardfehler
Grau unterlegt: Koeffizienten ausfallbelasteter Variablen.

Anhang B Analysemodell 2 (starke Selektivität)

Tabelle 10 Ausfallkorrektur durch fallweisen Ausschluss (alle Szenarien) – Probitregressionen der validen Einkommensangabe

	vollständige Daten	Szenario 1-4
	Koeff. (Std.Fehler)	Koeff. (Std.Fehler)
Beschäftigung zum Befragungszeitpunkt bereits beendet: ja	-0,174** (0,065)	-0,089 (0,075)
Kumulierte Dauer der Arbeitslosigkeit (Ref.: unter einem Monat)		
Ein bis unter sechs Monate	-0,144 (0,090)	-0,130 (0,102)
Sechs bis unter zwölf Monate	-0,079 (0,099)	-0,044 (0,113)
Zwölf bis unter vierundzwanzig Monate	-0,140 (0,102)	-0,118 (0,117)
Über vierundzwanzig Monate	-0,233* (0,119)	-0,144 (0,135)
Geschlecht: weiblich	0,178* (0,072)	0,142 (0,082)
Alter	-0,024 (0,026)	-0,072* (0,031)
Alter (quadr.)	0,000 (0,000)	0,001* (0,000)
Zahl der Kinder unter 18 Jahren	0,041 (0,034)	0,064 (0,040)
Nettohaushaltseinkommen	0,000*** (0,000)	0,000*** (0,000)
Zufrieden mit dem Lohn	-0,019 (0,061)	-0,043 (0,070)
Arbeitszeit	0,006* (0,003)	0,004 (0,003)
Art der Kündigung letzte Beschäftigung (Ref.: Sonstiges)		
Arbeitnehmerkündigung	-0,037 (0,120)	-0,110 (0,132)
Arbeitgeberkündigung	0,105 (0,069)	0,135 (0,079)
Zahl der genutzten Suchwege: (Ref.: keine)		
Ein bis drei Suchwege	0,163 (0,200)	0,351 (0,232)
Vier bis sechs Suchwege	0,209 (0,136)	0,293 (0,156)
Sieben bis neun Suchwege	0,240 (0,138)	0,381* (0,158)
Bereits ein Stellenangebot vom AA wegen zu niedrigem Einkommen abgelehnt? ja	0,120 (0,103)	0,060 (0,117)
Von ABM in Beschäftigung: ja	-0,172 (0,119)	-0,225 (0,133)
Vor Beschäftigung Weiterbildung: ja	-0,053 (0,146)	0,032 (0,170)
Vor Beschäftigung Sozialhilfe: ja	0,137 (0,084)	0,178 (0,097)
Konstante	0,395 (0,501)	1,247* (0,589)
Pseudo R ²	0,0291	0,0294
N	2270	1775

* $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$; robuste Standardfehler
 Grau unterlegt: Koeffizienten ausfallbelasteter Variablen

Tabelle 11 Ausfallkorrektur durch Multiple Imputation nach Szenarien – Probitregressionen der validen Einkommensangabe

	Vollständige Daten	Szenario 1	Szenario 2	Szenario 3	Szenario 4
	Koeff. (Std.Fehler)	Koeff. (Std.Fehler)	Koeff. (Std.Fehler)	Koeff. (Std.Fehler)	Koeff. (Std.Fehler)
Beschäftigung zum Befragungszeitpunkt bereits beendet: ja	-0,174** (0,065)	-0,086 (0,065)	-0,168** (0,065)	-0,176** (0,065)	-0,171* (0,067)
Kumulierte Dauer der Arbeitslosigkeit (Ref.: unter einem Monat)					
Ein bis unter sechs Monate	-0,144 (0,090)	-0,130 (0,090)	-0,157 (0,102)	-0,174 (0,103)	-0,187 (0,101)
Sechs bis unter zwölf Monate	-0,079 (0,099)	-0,042 (0,096)	-0,051 (0,121)	-0,078 (0,115)	-0,064 (0,113)
Zwölf bis unter vierundzwanzig Monate	-0,140 (0,102)	-0,119 (0,099)	-0,115 (0,122)	-0,141 (0,117)	-0,141 (0,118)
Über vierundzwanzig Monate	-0,233* (0,119)	-0,093 (0,124)	-0,157 (0,130)	-0,207 (0,155)	-0,158 (0,130)
Geschlecht: weiblich	0,178* (0,072)	0,089 (0,073)	0,171* (0,072)	0,151 (0,078)	0,147* (0,074)
Alter	-0,024 (0,026)	-0,041 (0,058)	-0,027 (0,026)	-0,042 (0,027)	-0,031 (0,031)
Alter (quadr.)	0,000 (0,000)	0,001 (0,001)	0,000 (0,000)	0,001 (0,000)	0,000 (0,000)
Zahl der Kinder unter 18 Jahren	0,041 (0,034)	0,056 (0,039)	0,042 (0,034)	0,059 (0,037)	0,048 (0,052)
Nettohaushaltseinkommen	0,000*** (0,000)	0,000 (0,000)	0,000*** (0,000)	0,000*** (0,000)	0,000*** (0,000)
Zufrieden mit dem Lohn	-0,019 (0,061)	-0,031 (0,067)	-0,018 (0,061)	-0,045 (0,067)	-0,066 (0,068)
Arbeitszeit	0,006* (0,003)	0,003 (0,003)	0,006* (0,003)	0,005 (0,003)	0,004 (0,004)
Art der Kündigung letzte Beschäftigung (Ref.: Sonstiges)					
Arbeitnehmerkündigung	-0,037 (0,120)	-0,053 (0,143)	-0,024 (0,120)	-0,025 (0,121)	-0,139 (0,169)
Arbeitgeberkündigung	0,105 (0,069)	0,079 (0,074)	0,113 (0,069)	0,112 (0,070)	0,117 (0,095)
Zahl der genutzten Suchwege: (Ref.: keine)					
Ein bis drei Suchwege	0,163 (0,200)	0,241 (0,217)	0,165 (0,200)	0,161 (0,201)	0,344 (0,254)
Vier bis sechs Suchwege	0,209 (0,136)	0,246 (0,141)	0,198 (0,136)	0,184 (0,137)	0,293 (0,170)
Sieben bis neun Suchwege	0,240 (0,138)	0,324* (0,148)	0,228 (0,138)	0,211 (0,139)	0,373* (0,172)
Bereits ein Stellenangebot vom AA wegen zu niedrigem Einkommen abgelehnt? ja	0,120 (0,103)	0,022 (0,118)	0,118 (0,103)	0,125 (0,104)	0,046 (0,154)
Von ABM in Beschäftigung: ja	-0,172 (0,119)	-0,170 (0,124)	-0,177 (0,119)	-0,165 (0,120)	-0,209 (0,144)
Vor Beschäftigung Weiterbildung: ja	-0,053 (0,146)	0,018 (0,157)	-0,049 (0,147)	-0,057 (0,147)	0,027 (0,185)
Vor Beschäftigung Sozialhilfe: ja	0,137 (0,084)	0,155 (0,088)	0,141 (0,084)	0,125 (0,086)	0,167 (0,102)
Konstante	0,395 (0,501)	0,611 (0,910)	0,470 (0,504)	0,840 (0,514)	0,555 (0,600)
Pseudo R ²	0,0291	n.v.	n.v.	n.v.	n.v.
N	2270	2270	2270	2270	2270

* $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$; n.v.: nicht verfügbar;
Standardfehler basieren auf Rubin 1987; Variablen des Imputationsmodells siehe Tabelle 5;
Grau unterlegt: Koeffizienten ausfallbelasteter Variablen.

Tabelle 12a Ausfallkorrektur durch Heckman-Korrektur (ML) nach Szenarien – Probitregressionen der validen Einkommensangabe

	Vollständige Daten	Szenario 1	Szenario 2	Szenario 3	Szenario 4
	Koeff. (Std.Fehler)	Koeff. (Std.Fehler)	Koeff. (Std.Fehler)	Koeff. (Std.Fehler)	Koeff. (Std.Fehler)
Beschäftigung zum Befragungszeitpunkt bereits beendet: ja	-0,174** (0,065)	-0,066 (0,096)	-0,066 (0,086)	-0,061 (0,069)	-0,079 (0,082)
Kumulierte Dauer der Arbeitslosigkeit (Ref.: unter einem Monat)					
Ein bis unter sechs Monate	-0,144 (0,090)	-0,107 (0,118)	-0,127 (0,098)	-0,125 (0,092)	-0,131 (0,100)
Sechs bis unter zwölf Monate	-0,079 (0,099)	-0,038 (0,109)	-0,055 (0,105)	-0,053 (0,102)	-0,049 (0,111)
Zwölf bis unter vierundzwanzig Monate	-0,140 (0,102)	-0,083 (0,142)	-0,129 (0,109)	-0,125 (0,106)	-0,124 (0,115)
Über vierundzwanzig Monate	-0,233* (0,119)	-0,145 (0,132)	-0,149 (0,129)	-0,142 (0,122)	-0,148 (0,133)
Geschlecht: weiblich	0,178* (0,072)	0,146 (0,080)	0,145 (0,079)	0,145 (0,074)	0,115 (0,089)
Alter	-0,024 (0,026)	-0,068* (0,033)	-0,068* (0,032)	-0,066* (0,028)	-0,070* (0,031)
Alter (quadr.)	0,000 (0,000)	0,001 (0,000)	0,001* (0,000)	0,001* (0,000)	0,001* (0,000)
Zahl der Kinder unter 18 Jahren	0,041 (0,034)	0,052 (0,053)	0,051 (0,049)	0,059 (0,036)	0,064 (0,039)
Nettohaushaltseinkommen	0,000*** (0,000)	0,000 (0,000)	0,000* (0,000)	0,000*** (0,000)	0,000*** (0,000)
Zufrieden mit dem Lohn	-0,019 (0,061)	-0,044 (0,067)	-0,044 (0,067)	-0,037 (0,064)	-0,042 (0,069)
Arbeitszeit	0,006* (0,003)	0,003 (0,004)	0,003 (0,004)	0,003 (0,003)	0,004 (0,004)
Art der Kündigung letzte Beschäftigung (Ref.: Sonstiges)					
Arbeitnehmerkündigung	-0,037 (0,120)	-0,136 (0,127)	-0,138 (0,126)	-0,132 (0,123)	-0,103 (0,133)
Arbeitgeberkündigung	0,105 (0,069)	0,126 (0,082)	0,125 (0,080)	0,119 (0,074)	0,133 (0,079)
Zahl der genutzten Suchwege: (Ref.: keine)					
Ein bis drei Suchwege	0,163 (0,200)	0,297 (0,302)	0,294 (0,275)	0,283 (0,214)	0,337 (0,238)
Vier bis sechs Suchwege	0,209 (0,136)	0,232 (0,242)	0,235 (0,206)	0,228 (0,142)	0,281 (0,163)
Sieben bis neun Suchwege	0,240 (0,138)	0,305 (0,273)	0,307 (0,229)	0,297* (0,144)	0,367* (0,168)
Bereits ein Stellenangebot vom AA wegen zu niedrigem Einkommen abgelehnt? ja	0,120 (0,103)	0,076 (0,113)	0,076 (0,112)	0,080 (0,109)	0,060 (0,115)
Von ABM in Beschäftigung: ja	-0,172 (0,119)	-0,210 (0,153)	-0,209 (0,144)	-0,208 (0,123)	-0,222 (0,132)
Vor Beschäftigung Weiterbildung: ja	-0,053 (0,146)	0,014 (0,168)	0,009 (0,167)	0,010 (0,159)	0,033 (0,167)
Vor Beschäftigung Sozialhilfe: ja	0,137 (0,084)	0,156 (0,134)	0,140 (0,122)	0,135 (0,091)	0,134 (0,095)
Konstante	0,395 (0,501)	1,521* (0,690)	1,529* (0,643)	1,470** (0,534)	1,357* (0,645)
Ath (ρ_{ev})		-0,733 (2,914)	-0,775 (2,481)	-2,473 (83,978)	-0,317 (1,150)
ρ_{ev}		-0,650 (1,433)	-0,625 (1,776)	-0,986 (2,355)	-0,307 (1,041)
Pseudo R ²	0,0291	n.v.	n.v.	n.v.	n.v.
N	2270	1775	1775	1775	1775

* $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$; robuste Standardfehler
 Grau unterlegt: Koeffizienten ausfallbelasteter Variablen

Tabelle 12b Ausfallkorrektur durch Heckman-Korrektur (ML) nach Szenarien – Selektionsmodell (Probitregression des Datenausfalls)

	Szenario 1 Koeff. (Std.Fehler)	Szenario 2 Koeff. (Std.Fehler)	Szenario 3 Koeff. (Std.Fehler)	Szenario 4 Koeff. (Std.Fehler)
Beschäftigung zum Befragungszeitpunkt bereits beendet: ja	-0,080 (0,065)	-0,079 (0,065)	-0,074 (0,063)	-0,071 (0,063)
Kumulierte Dauer der Arbeitslosigkeit (Ref.: unter einem Monat)				
Ein bis unter sechs Monate	-0,096 (0,089)			
Sechs bis unter zwölf Monate	-0,080 (0,099)			
Zwölf bis unter vierundzwanzig Monate	-0,214* (0,103)			
Über vierundzwanzig Monate	-0,024 (0,129)			
Geschlecht: weiblich	-0,135 (0,083)	-0,133 (0,080)		
Alter	0,003 (0,026)	0,000 (0,025)		
Alter (quadr.)	0,000 (0,000)	0,000 (0,000)		
Zahl der Kinder unter 18 Jahren	0,047 (0,036)	0,048 (0,035)		
Nettohaushaltseinkommen	0,000** (0,000)	0,000** (0,000)		
Zufrieden mit dem Lohn	0,024 (0,061)	0,023 (0,061)		
Arbeitszeit	0,000 (0,003)	0,000 (0,003)		
Art der Kündigung letzte Beschäftigung (Ref.: Sonstiges)				
Arbeitnehmerkündigung	0,234 (0,134)	0,238 (0,131)	0,209 (0,125)	
Arbeitgeberkündigung	0,018 (0,069)	0,019 (0,067)	0,029 (0,066)	
Zahl der genutzten Suchwege: (Ref.: keine)				
Ein bis drei Suchwege	0,013 (0,203)	0,019 (0,200)	-0,002 (0,198)	
Vier bis sechs Suchwege	0,100 (0,137)	0,079 (0,136)	0,048 (0,135)	
Sieben bis neun Suchwege	0,156 (0,139)	0,135 (0,139)	0,106 (0,137)	
Bereits ein Stellenangebot vom AA wegen zu niedrigem Einkommen abgelehnt? ja	-0,086 (0,101)	-0,090 (0,099)	-0,103 (0,097)	
Von ABM in Beschäftigung: ja	-0,003 (0,136)	-0,001 (0,131)	0,018 (0,124)	
Vor Beschäftigung Weiterbildung: ja	0,093 (0,153)	0,106 (0,152)	0,079 (0,148)	
Vor Beschäftigung Sozialhilfe: ja	0,198* (0,087)	0,211* (0,087)	0,176* (0,083)	
<i>Instrumentvariablen</i>				
Keine deutsche Nationalität	-0,337** (0,117)	-0,322** (0,116)	-0,282** (0,104)	-0,273* (0,112)
Region: Ost	0,222* (0,087)	0,206** (0,077)	0,173** (0,063)	0,150* (0,065)

	Szenario 1 Koeff. (Std.Fehler)	Szenario 2 Koeff. (Std.Fehler)	Szenario 3 Koeff. (Std.Fehler)	Szenario 4 Koeff. (Std.Fehler)
Wie wurde die angetretene Stelle gefunden (Ref.: Arbeitsamt)				
Bekannte/Freunde	-0,003 (0,087)	-0,004 (0,086)	-0,010 (0,084)	-0,015 (0,085)
Eigene Initiative	-0,088 (0,081)	-0,086 (0,080)	-0,095 (0,079)	-0,105 (0,081)
Sonstige Suchwege	-0,072 (0,088)	-0,073 (0,088)	-0,059 (0,086)	-0,077 (0,090)
Konstante	0,402 (0,497)	0,394 (0,493)	0,689*** (0,148)	0,816*** (0,071)
Pseudo R ²	n.v.	n.v.	n.v.	n.v.
N	2270	2270	2270	2270

* $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$; robuste Standardfehler
 Grau unterlegt: Koeffizienten ausfallbelasteter Variablen.

Rezensionen



TIMOTHY A. BROWN, 2006: *Confirmatory Factor Analysis for Applied Research*. Guilford Press, New York, London. ISBN 978-1-59385-274-0, 475 Pages, EUR 36,99.

At the beginning the author writes in his introduction "This book was written for the simple reason, that no other book of its kind had been published before" and he is right. To achieve the goal of a user friendly text he uses four strategies:

1. Every key concept is accompanied by an applied data set and the syntax and output from the mostly used structural equation software packages AMOS, CALIS, EQS, LISREL and MPLUS.
2. Tables are included that recap the procedures or steps of the methods being presented (e. g. how to write up the results of a CFA study for reports or publications).
3. Numerous figures are provided that graphically illustrate some of the more complicated concepts or procedures (e. g. forms of measurement invariance, types of nonpositive definite matrices, identification of formative indicator models).
4. Many chapters contain appendices with user-friendly illustrations of seemingly complex quantitative operations (e. g. data generation in Monte Carlo Simulation, calculation of matrix determinants and their role in model fit and improper solutions, the effect of model identification on standard errors).

The book is structured in two parts. The first five chapters present the fundamental concepts and procedures of CFA including an introduction to Exploratory Factor Analysis. The second half of the book deals with more specialized issues like analysis of multitrait-multimethod (MTMM) data and issues of validation (chapter 6), different types of constraints and analysis of multiple groups including intercepts and latent means (MGCFA) and in addition Multiple Indicator Multiple Causes Models (MIMIC) which combine formative and reflective indicators in one model (chapter 7). Chapters 8 and 9 contain Higher Order Confirmatory Factor Analysis, CFA approaches to scale reliability estimation and CFA with formative indicators, missing values and non-normally distributed continuous indicators and a detailed overview of confirmatory factor analysis with categorical indicators. The final chapter deals with the problem of sample size and power.

Timothy Brown uses only computer syntax examples out of two plausible reasons. Graphical input does not lend itself easily to the written page and he argues that using syntax may be often faster than graphical input and gives more transparency to the underlying model. However to understand all the possible model specifications in CFA and structural equation modelling graphical input facilitates the learning for beginners significantly. Those, who want to get a good introduction to graphical input as a complement are well served by the books of Barbara Byrne (Byrne 1998, 2006, 2010). In the following section I want to review the chapters in more detail.

Chapter 1 contains a gentle introduction to the topic and summarizes the four main goals of Confirmatory Factor Analysis that is the Psychometric Evaluation of Test Instruments, Construct Validation, Analysis of

Method Effects and Evaluation of Measurement invariance.

The common Factor Model and Exploratory Factor Analysis (EFA) are discussed in detail. The concepts, the terminology and the basic equations are well explained and it is outlined very clearly, that both exploratory and confirmatory factor analysis are based on the common factor model. Here as in the other chapters he uses the usual notation by using Greek symbols. For the illustration of the results the author uses here and in all other chapters path diagrams. As main goal for EFA he defines the determination of the dimensionality of a set of multiple indicators by uncovering the smallest number of interpretable factors needed to explain the correlations between them. In more detail he explains the selection of a specific method to estimate the factor model, the selection of the appropriate number of factors, selection of a technique to rotate the initial factor matrix to foster the interpretability of the solution and the selection of a method to compute factor scores. In the end (p. 38) he gives a very informative table summarizing all necessary steps. From my point of view I miss a short discussion of the shortcoming of EFA statistically and methodologically and especially in the past and present research practice in the sense of inductive methodology and ad-hoc interpretation. A fine example of such a criticism was given by Duncan's (1984) critique of factor analysis or Borsboom's (2006) criticism of classical test theory.

The purposes, parameters and fundamental equations of confirmatory factor analysis and the comparison of EFA and CFA are dealt with in chapter three. Brown discusses standardized and unstandardized solutions, indicator cross-loadings, unique variances, correlated indicator errors and their possible explanation by method factors. These method factors can be for example question wording, social desirability, acquiescence and others and cannot be estimated in EFA but only in CFA. Furthermore he explains

as central advantage of confirmatory factor analysis to specify a priori the number of factors, the relationship between indicators and latent variables and the relations between factors and those between errors. All the available information can be transformed in corresponding constraints which are tested. In addition he shows by giving an example that in contrast to the correlation of sum or composite scores or single items the correlation of factors is corrected from measurement error and therefore not biased. Very instructive is also how he explains the reproduction of an input correlation matrix from the parameter estimates of a two factor measurement model as this procedure is often difficult for beginners to understand. CFA model identification and estimation and descriptive Goodness of Fit Indices are explained in detail technically but also in a very clear and non-technical way and illustrated with a substantive example. The appendix to the chapter shows nicely how to obtain a solution for a just-identified factor model and the hand calculation of the Maximum Likelihood Fitting Function for one model treated before. Chapter four gives a very good account of the specification and interpretation of CFA models. Using the SEM programs LISREL, MPLUS, EQS and CALIS a two factor model of personality is evaluated. In detail overall goodness of fit, interpretability of the parameters including standardized and unstandardized solutions are discussed. Very informative for users of CFA is Table 4.6 which summarizes based on different articles on this topic what to report in a CFA study for an article or report. Maybe in the whole chapter Brown should have stressed one of the major criticisms of the use of SEM that users should more explicitly document the underlying substantive theory for their model specifications (McCallum/Austin 2000). This refers also to the issue of the deciding whether indicators are formative or reflective, which has been an issue of increasing importance (e. g. Bollen 2007, Howell et al. 2007). Two useful appendices

dealing with the topics "Goodness of Model Fit does not ensure meaningful parameter estimates" and an example report of the "Two-factor model of neuroticism and extraversion" ends the chapter.

"CFA Model Revision and Comparison" deals with the extremely important issue of model modification. In a systematic way Brown deals with the sources of poor-fitting CFA solutions (number of factors, misspecified relationships between indicators and constructs, correlated errors) and the problem of improper solutions and nonpositive definite matrices. All the topics are illustrated by different data-sets. In addition he discusses the use of EFA for model modification in a CFA framework, which has been recently intensively dealt with by Aspourov and Muthen (2009). The discussion of the often neglected problem of equivalent models is treated by him especially in deciding between a second-order factor model versus first-order factor models with multiple factors. In chapter six he shows how the application of CFA to multitrait-multimethod-matrices allows the estimation of convergent, discriminant and method effects in the evaluation of the construct validity of latent variables. The different possible parametrizations (like Correlated Methods Models vs. Correlated Uniqueness Models) are discussed in detail and the syntax for the correlated method models for the five different programs is given. In addition he discusses shortly other parametrizations like the direct product model by Browne, which allows interactions between method and trait factors. However, I miss a discussion of the true score MTMM models (Scherpenzeel/Saris 1997) and the application to the modelling of autocorrelated errors in panel data, which could be especially useful for a sociological and political science audience (Finkel 1995). Confirmatory Factor Analysis with Equality Constraints, Multiple Groups and Mean Structures are dealt with in chapter 7. After a discussion of congeneric, Tau-equivalent and parallel indicators and their

implementation in the five programs he introduces the topic of invariance of measurements over time and groups. Furthermore the estimation of latent means and intercepts of indicators is demonstrated. This is done both for multiple time points and for multiple cross-sectional groups and syntax for both cases is given. In addition the relationship between MIMIC models and Multiple Group Confirmatory Factor Analysis is discussed and the test of differential item functioning using MIMIC models. Special emphasis is given to the issue of partial invariance and the practical guidelines discussed in the literature. Higher order factor analysis, scale reliability evaluation and formative indicators are discussed in chapter 8. For the second order factor analysis he gives an example syntax, discusses the necessary steps for testing and describes in general terms and for a specific example the Schmid-Leiman Transformation to estimate the relationship of the observed measures to the higher-order factors. In the next section he discusses procedures based on CFA, which overcome limitations of classical reliability estimates like Cronbach's Alpha, by no longer assuming equal loadings, equal measurement error, unidimensionality and no error correlations. Finally he deals with the issue of models with formative indicators and addresses the problems of identification, estimation and interpretation. Examples are the model specification for the latent variables Life Stress and Socioeconomic Status (SES). Missing, non-normal and categorical data - often neglected problems in practical applications - are treated in chapter 9. In detail the two recommended procedures for treating missing values that is model based full information maximum likelihood (FIML) and multiple imputation are treated and syntax examples are given. Furthermore alternative estimation procedures for non-normal data (robust ML and weighted least squares) are discussed. Finally he describes the estimators for categorical data like weighted least squares and robust weighted least squares (WLSMV) and their strengths

and weaknesses. Very informative is also the section on the comparison of CFA with item response theory (IRT) models normally perceived as two different worlds. He starts with the well known point that factor analysis with binary outcomes is equivalent to a two-parameter normal ogive item response theory model and then discusses other IRT models. Using MPLUS he demonstrate the procedure to estimate a one factor CFA with binary indicators and factor loadings constrained to equality. In the final chapter another neglected topic that is statistical power and sample size is treated. In his discussion he clearly argues that simple general rules like the relation of subjects to number of indicators are not sufficient. One needs to determine the sample size and as a consequence the power of models in the context of a particular model and data-set. He shows how the Satorra-Saris method in LISREL and the Monte Carlo method in MPLUS can be used to reach the goal of the specific determination of power for a given model and data-set. The only important missing information in this chapter but also relevant for chapter 9 is the meta-analysis of the robustness of SEM models in general by Hoagland and Boomsma (1998). Finally an overview of newly developed procedures like multilevel factor models and factor mixture models is given. The only missing point is the treatment of bayesian estimation procedures, which are at least available in AMOS.

Summarizing the review I think that this is presently the best book for applied researchers on confirmatory factor analysis and can be used as a textbook in M.A. courses or Ph.D. courses for confirmatory factor analysis in the social sciences and psychology as the basic textbook. But it also can serve for researchers in the social sciences and psychology as a very useful reference book in modelling measurement models and building scales with confirmatory factor analysis.

References

- Asparouhov, T. and K. A. Bollen, 2009: Exploratory Structural Equation Modeling. *Structural Equation Modeling*, 16: 397-438.
- Bollen, K. A., 2007: Interpretational Confounding is Due to Misspecification, Not to Type of Indicator: Comment on Howell, Breivik and Wilcox. *Psychological Methods*, 12: 219-228.
- Boorsboom, D., 2006: The Attack of the Psychometricians. *Psychometrika*, 71: 425-440.
- Byrne, B., 1998: *Structural Equation Modeling with LISREL, PRELIS and SIMPLIS*. Maywah: Erlbaum.
- Byrne, B., 2006: *Structural Equation Modeling with EQS*. London: Sage.
- Byrne, B., 2010: *Structural Equation Modeling with AMOS*. New York: Taylor and Francis.
- Duncan, O. D., 1984: *Notes on Social Measurement*. New York: Russell Sage.
- Finkel, S. E., 1995: *Causal Analysis with Panel Data*. London: Sage.
- Hoagland, J. J. and A. Boomsma, 1998: Robustness Studies in Covariance Structure Modeling. An Overview and a Metaanalysis. *Sociological Methods and Research*, 26: 329-367.
- Howell, R. D., Breivik, E. and J. B. Wilcox, 2007: Reconsidering Formative Measurement. *Psychological Methods*, 12: 205-218.
- MacCallum, R. C. and J. T. Austin, 2000: Structural Equation Modeling in Psychological Research. *Annual Review of Psychology*, 51: 201-226.
- Scherpenzeel, A. and W. Saris, 1997: The Validity and Reliability of Survey Questions: A Meta-Analysis of MTMM Studies. *Sociological Methods and Research*, 25: 341-383.

PETER SCHMIDT, MARBURG



UWE FLICK, 2009:
Sozialforschung.
Methoden und
Anwendungen. Ein
Überblick für die
BA-Studiengänge.
Reinbek bei
Hamburg: Rowohlt
Taschenbuch Verlag.
ISBN: 978-3-499-
55702-6, 336
Seiten, EUR 12,95.

In Folge der Einführung von sozialwissenschaftlichen Bachelorstudiengängen an den Hochschulen erscheinen eine Reihe von neuen oder grundlegend überarbeiteten Lehrbüchern zu empirischen und statistischen Forschungsmethoden, die den veränderten Anforderungen und Rahmenbedingungen im Studium gerecht werden sollen. Uwe Flicks Lehrbuch zur „Sozialforschung“ verfolgt diesen Anspruch explizit. Demnach ist – so Flick (S. 10) – von der Einführung der BA/MA-Studiengänge im Vergleich zu den Diplomstudiengängen eine „Verkürzung der grundständigen Ausbildung“ und im BA zumeist auch eine „Reduzierung der angebotenen Inhalte“ zu erwarten, obwohl „zumindest grundlegende Kenntnisse der Sozialforschung eine der zu vermittelnden Kernkompetenzen“ bleiben dürften. Ein Vergleich mit den Empfehlungen der Deutschen Gesellschaft für Soziologie zur Ausgestaltung soziologischer Bachelor- und Masterstudiengänge sowie zur Methodenausbildung unterstreicht diese Erwartungen. Der dort empfohlene Anteil der Methodenausbildung an den BA-Studiengängen liegt bei 20 bis 25 Prozent, wobei neben der Statistik sowohl die standardisierte („quantitative“) Sozialforschung als auch die nichtstandardisierte („qualitative“) gleich gewichtet sind (Stichwort: Methodenpluralismus). Deshalb möchte Flick in seinem Lehrbuch einen einflussreichen Überblick über das gesamte Feld

der empirischen Sozialforschung geben, der qualitative und quantitative Ansätze gleichermaßen berücksichtigt, der aber dennoch im engen Zeitkorsett der BA-Studiengänge handhabbar bleibt, weil er weiter reichende Vertiefungen der jeweiligen Spezialliteratur überlässt (S. 11). Ob und in welcher Weise das Buch diesen – auf den ersten Blick widerstrebenden – Ansprüchen gerecht wird, ist nach der Sichtung von Inhalt und Form zu diskutieren.

Das Buch ist in zwölf Kapitel gegliedert, deren Reihenfolge sich in bewährter Weise am sequenziellen Phasenmodell der empirischen Forschungspraxis orientiert. Auf zwei einführende Kapitel zu den Zielsetzungen und Grundformen empirischer Sozialforschung (Kap. 1) sowie zur Entwicklung von forschungsleitenden Fragestellungen (Kap. 2) folgt ein grundlegender Teil zu den erkenntnistheoretischen Hintergründen (Kap. 3) sowie zu Prozess und Gestaltungsmöglichkeiten empirischer Forschung (Kap. 4). Konkrete Verfahren der Datenerhebung bzw. -auswertung werden am Beispiel von Befragung und Beobachtung (Kap. 5) sowie von Techniken der statistischen und interpretativen Datenanalyse (Kap. 6) erläutert. An die verfahrenstechnischen Grundlagen anknüpfend werden die im Forschungsprozess üblicherweise erforderlichen Entscheidungen rekapituliert (Kap. 7), die Kombinations- und Integrationsmöglichkeiten qualitativer und quantitativer Verfahren diskutiert (Kap. 8) und die gängigen Darstellungsformen und Präsentationsmöglichkeiten aufgezeigt (Kap. 9). Im Anschluss geht es um Qualitätskriterien (Kap. 10) und ethische Leitlinien (Kap. 11) für empirische Forschung. Abschließend wird ein Leitfaden zur Umsetzung eigener Forschungsprojekte auch im Hinblick auf empirische Studien- und Abschlussarbeiten (Kap. 12) präsentiert.

Bezüglich der didaktischen Umsetzung der umrissenen Inhalte sind die vielfältigen Beigaben zum Lehrbuch hervorzuheben. Neben den Standardverzeichnissen sind

ein Sach- und ein Personenregister enthalten und darüber hinaus ein umfangreiches Glossar mit den einschlägigen Fachbegriffen. Jedes Kapitel wird mit einem kleinen Inhaltsverzeichnis sowie einer Übersicht der Lernziele eingeleitet. Abgeschlossen werden die Kapitel jeweils mit einer Wiederholung inhaltlicher Kernpunkte und weiterführenden Literaturhinweisen, wobei Letztere oft unspezifisch auf weiterführende Lehrbücher der quantitativen oder qualitativen Sozialforschung verweisen. Differenziertere Hinweise finden sich jedoch im Fließtext. Die zahlreichen Abbildungen und Tabellen im Text dienen zwar z. T. auch der beispielhaften Präsentation von Daten, wie man es von einem Empirielehrbuch erwarten kann, noch häufiger aber der zusammenfassenden Übersicht wesentlicher methodischer Sachverhalte. Dazu kommen die Kästen, die ergänzend zum Fließtext Definitionen, Beispiele und besonders häufig Leitfäden zur Umsetzung von forschungspraktischen Entscheidungen und Abläufen enthalten. Für die praktische Arbeit mit dem Buch ist m. E. lediglich anzumerken, dass die Gliederung des Buches besser bis in die dritte Ebene hinein zu nummerieren und auch die dritte Gliederungsebene noch in das Inhaltsverzeichnis aufzunehmen wäre.

Eine inhaltliche Besonderheit des vorliegenden Lehrbuchs ist die Integration qualitativer und quantitativer Perspektiven und Verfahren. Dabei stellt sich die Frage, ob eine solche integrierte Darstellung der beiden Ansätze überhaupt sinnvoll ist, und falls ja, inwieweit und in welcher Weise? Ohne Zweifel haben die beiden Ansätze breite gemeinsame Grundlagen, doch es wird keinem der beiden gerecht, die Darstellung auf die Gemeinsamkeiten zu beschränken. Hinzukommen muss eine Darstellung der jeweiligen Besonderheiten, ggf. auch der Widersprüche und darüber hinaus der internen Vielfalt und Bandbreite der einzelnen Paradigmen. Diese Ansprüche erfüllt das Lehrbuch praktisch durchgängig. Besonders gewinnbringend für die Einsteiger/innen im Feld sind

z. B. die vergleichenden Gegenüberstellungen zum Forschungsprozess, zur Datenerhebung und -analyse, zur Kombination unterschiedlicher Ansätze und zur Darstellung von Untersuchungsergebnissen. Im Kapitel zu den wissenschaftstheoretischen Grundlagen scheint mir dagegen die Darstellung einzelner Ansätze zu knapp geraten, so z. B. bei den Kernannahmen des kritischen Rationalismus (S. 47-48). Insgesamt aber ist die oben gestellte Frage nach dem Sinn einer integrierten Darstellung qualitativer und quantitativer Forschungsmethoden auch bei kritischer Durchsicht des Werkes positiv zu beantworten.

„Sozialforschung“ von Uwe Flick ist ohne Zweifel ein gutes und empfehlenswertes Lehrbuch für Anfänger/innen im Bereich der empirischen Forschungsmethoden. Es ist inhaltlich plausibel und übersichtlich gegliedert, klar und kompetent geschrieben in einer leicht nachvollziehbaren Sprache. Es ist mit vielen praktischen Beispielen und Handreichungen versehen und nicht zuletzt sorgfältig redigiert. Die Integration qualitativer und quantitativer Verfahren ist innovativ und – von Einzelheiten abgesehen – ebenfalls gelungen. Das Buch kann gut dazu beitragen, die curriculare Trennung von qualitativer und quantitativer Sozialforschung zu überbrücken, die im Bereich der Lehre vor allem dort existiert, wo die Methodenausbildung einen breiten Raum im Studiengang einnimmt und auf mehrere Schultern verteilt ist. Damit hat Uwe Flick seine Zielsetzung, einen einführenden, breiten Überblick über das gesamte Feld der empirischen Sozialforschung zu geben, voll und ganz erreicht. Sorgfältig zu bedenken sind dennoch die Einsatzgebiete des Lehrbuchs. Zum Selbststudium ist es bestens geeignet und auch zur begleitenden Lektüre eines Studienmoduls empirischer Sozialforschung. Dagegen kann es weder alleinige Grundlage einer einzelnen Lehrveranstaltung noch eines gesamten Moduls sein, weil es einerseits zu breit angelegt ist und andererseits die einzelnen Erhebungs- und Auswertungsme-

thoden nicht tief genug erläutert werden. Doch das sei hier nicht als Kritik sondern vielmehr als Gebrauchsanweisung im Sinne des Autors selbst gelesen (S. 11).

JÜRGEN SCHIENER, MAINZ

* * * * *



CHRISTIAN FG SCHENDERA, 2010: Clusteranalyse mit SPSS. Mit Faktorenanalyse. Oldenburg Wissenschaftsverlag. München. ISBN 978-3-486-58691-6, 435 Seiten, EUR 49,80.

Die Monographie „Clusteranalyse mit SPSS“ von Schendera reiht sich ein in die Sammlung von Begleitbüchern zur Anwendung von SPSS, für die offenbar mangels SPSS-eigener Publikationen Nachfrage besteht.

Das im Haupttitel und Kapitel 1 behandelte Thema, die Clusteranalyse, ist nicht der einzige Gegenstand des Werks. Im Untertitel „Mit Faktorenanalyse“ (Kapitel 2) wird dem Leser noch ein weiteres Verfahren offeriert und als unangekündigter Überraschungsgast taucht im dritten Kapitel noch die „Diskriminanzanalyse“ auf. Der Autor nimmt in seinem Vorwort die zu erwartende Frage des Lesers vorweg: „Warum die augenscheinlich grundverschiedenen „Verfahren“ Cluster- und Faktorenanalyse in einem Buch?“. Seine „einfache Antwort“ ist: „Beiden Verfahrensgruppen liegt dasselbe Prinzip zugrunde: Klassifikation mit dem Ziel maximaler Homogenität.“ Da auf der Plattform dieses Prinzips noch eine ganze Reihe weiterer statistischer Methoden

untergebracht werden könnte, ergänzt der Autor den prinzipiellen Grund um eine Liste weiterer – m. E. nicht sehr triftiger – Spezialgründe, die im Einzelnen zu behandeln hier nicht der Ort ist.

Weniger ausführlich wird die Einbeziehung der Diskriminanzanalyse begründet – sie wird mehr oder weniger als ein zur Clusteranalyse komplementäres Verfahren vorgestellt: „Die Diskriminanzanalyse kehrt dabei die Vorgehensweise der Clusteranalyse um und versucht aus den clusteranalytisierten Variablen auf die ermittelte Clusterzugehörigkeit zu schließen. Aus dieser Perspektive heraus kann man die Diskriminanzanalyse *auch* als einen statistischen Plausibilitätstest der erzielten Clusterlösung verstehen.“ (S. 299)

Im Schlusskapitel 4 sind dann den Verfahren zugrundeliegende Formeln aufgelistet.

Wer das Buch in die Hand nimmt, um etwas über die thematisch benannten Verfahren zu lernen, der wird enttäuscht sein – es handelt sich im ersten Teil eher um ein „Kochbuch“. Es hat demzufolge auch das Problem, die Darstellung des Verfahrens und seiner Umsetzung in eine SPSS-Syntax mit anschließender Erläuterung der Ausgabe in einen einheitlichen Textfluss zu bringen. Dies gelingt, kurz gesagt, für den Bereich der Clusteranalyse nahezu nie. Bereits das Vorwort ist in Vorwegnahme später dargestellter Inhalte befrachtet mit Fachterminologie und Meinungen zur richtigen Anwendungsweise und Interpretation, die für den „Einsteiger“ kaum verständlich sind. Das Buch erreicht in der Darstellung der Clusteranalyse stellenweise bestenfalls das Niveau einschlägiger Lehrbücher, die aber den unvergleichlichen Vorteil haben, ausschließlich der Sachlogik folgen zu dürfen und einen Leser auch befähigen, SPSS anzuwenden, eventuell mit kleinen Unterstützungen aus dem Hilfe-Menu.

Zunächst beginnt der Teil der Clusteranalyse noch gut mit einem heuristischen Beispiel aus dem Ferienalltag – Gruppierung von

Muscheln am Strand. Leider ist dieses Beispiel nicht konsequent durchgeführt. Der Leser erfährt nicht, für welche Clusterlösung er sich – etwa aufgrund des Dendrogramms – entscheiden könnte. Außerdem sind bei der Durchführung des Beispiels in der Berechnung der Distanzmatrix nach dem zweiten Fusionsschritt hässliche Schönheitsfehler passiert. Das in diesem Schritt neu gebildete Aggregat aus Muschel 2 und 5 erhält nicht die neu zu berechnenden Aggregatdistanzen, es werden vielmehr alte Maßzahlen aus der Matrix nach dem ersten Fusionsschritt übernommen – verwirrend für Einsteiger.

Aber dann schlägt der Text eine Tonlage an, die spekulativ ist und zu dem Verfahren keinen stringenten Bezug besitzt.

Richtig ist der Hinweis (S. 6 unten, S. 7 oben) darauf, dass die Ergebnisse einer Analyse von der Wahl des Distanzmaßes und des „Fusionsalgorithmus“, also der Definition der Distanz von Aggregaten und der Skalierung der Variablen abhängen. So wenig aber wie am Beispiel überhaupt ein „Ergebnis“ festgehalten wird, so wenig werden diese Abhängigkeiten an dem gewählten Beispiel demonstriert (es ist allerdings auch nicht dafür geeignet). Dass das Ergebnis auch noch von „Merkmalen der zu gruppierenden Elemente (u. a. beobachtbare vs. latente Merkmale)“ abhängen soll, ist kaum zu verstehen. Der nachfolgende Text zeigt jedoch, dass es sich dabei um ein Problem der Interpretation handeln soll.

Angeblich liegt der Clusteranalyse „eine unausgesprochene, aber absolut zentrale, und zwar gleich zweifache Annahme zugrunde: a) Dass die noch nicht geclusterten Daten überhaupt Cluster enthalten, und b) dass diese Cluster Sinn machen. Zusammen münden diese beiden Aspekte wieder in die zu prüfende Voraussetzung, ob die Daten überhaupt Strukturen i. S. v. sinnhaltigen Gruppierungen enthalten (können).“ (S. 7) Das ist schlicht irreführend: mit der Anwendung eines explorativen Verfahrens setzt

man nicht die Struktur schon voraus, die man zu finden hofft.

Die anschließenden Ausführungen über manifeste und latente Variable und ihre jeweilige Bedeutung für die Interpretation von Clustern sollen den Leser zur Vorsicht gemahnen, helfen ihm aber im Verständnis der Clusteranalyse nicht weiter.

Das Problem einer Rezension des vorliegenden Buchs ist, dass sie zu viele einzelne Kritikpunkte erfordern würde, wollte man alle Details des Texts angemessen würdigen. Ich greife daher nur einige besonders typische Passagen des Buches in der Rezension auf.

Der Kochbuchcharakter wird sehr deutlich im weiteren Verlauf der Darstellung der hierarchisch-agglomerativen Clusteranalyse in Abschnitt 1.2. Die vielfältigen Möglichkeiten von SPSS für die verschiedenen Typen von Variablen werden hier tabellarisch quasi plakatiert. Der Begleittext kommentiert und gibt Empfehlungen zur Anwendung, teilweise allerdings nur mit minimalen Informationen. Immerhin ist es vorstellbar, dass sich auch ein „Einsteiger“ nach einer solchen Lektüre vor den PC setzen kann und eine Clusteranalyse rechnet. Aber der Text ist eher für ein „Durchwursteln“ geeignet als für den systematischen Erwerb von Kenntnissen.

Die Anwendungen zeigen die Abhängigkeit der Clusterlösungen von den gewählten Verfahren, dann folgt die Eingabemöglichkeit von verschiedenen Datentypen (Fall vs. Variable – Matrix oder Distanzmatrix) und konkrete Beispiele mit Variablen unterschiedlichen Skalenniveaus. Leider werden bei den Anwendungen die didaktischen Möglichkeiten verschenkt, die beispielhafte Analysen bieten würden. Hier müsste ausführlicher darauf eingegangen werden, wie der Anwender die Entscheidungen über die Verfahrensparameter treffen sollte. Der Leser erfährt hier nur etwas über die korrekte SPSS-Syntax, aber nichts über eine auch inhaltlich befriedigende Datenanalyse.

Besonders problematisch ist außerdem der Vorschlag, der in Abschnitt 1.2.9 „Analyse von gemischten Daten II“ entwickelt wird. Bei einem Satz von Clustervariablen, der aus einem Teil binärer und einem Teil metrischer Variablen besteht, soll für jeden Teil eine eigene Clusteranalyse gerechnet werden. Nach Auswahl je einer Clusterlösung sollen dann die Beobachtungen entsprechend den beiden Clusterzugehörigkeitsvariablen kreuztabelliert werden. Jede Zelle der Kreuztabelle sei dann ein eigenes Cluster. Dieses Vorgehen mag in gewissen Spezialfällen zu sinnvoll interpretierbaren Clustern führen, ist aber generell nicht begründbar. Es wird dann scheitern, wenn die Cluster sich durch bestimmte Interaktionen zwischen den metrischen und den binären Variablen auszeichnen.

Das Two-Step-Clusterverfahren ist unter den hierarchisch-agglomerativen Verfahren die Spezialität von SPSS. Bis auf ca. drei Seiten, auf denen das Verfahren heuristisch beschrieben wird, und die sich stark an Bacher et al. und den von SPSS zitierten Originalarbeiten orientieren und auf diese verweisen, ist dieser Teil im Wesentlichen der Gang durch eine Beispielanalyse mit Erläuterungen.

Auch der Abschnitt 1.4 „Partitionierendes Verfahren: Clusterzentrenanalyse (k-means)“ gibt im Wesentlichen eine korrekte Beschreibung. Allerdings fehlt eine genauere Beschreibung des Sortieralgorithmus, der auf der Minimal-Distanzeigenschaft einer optimalen Partition beruht. Da es mehrere solcher Partitionen geben kann, besitzt das k-means Verfahren mitunter keine eindeutige, sondern mehrere startpunktabhängige Lösungen.

Die beiden „Nebenthemen“ des vorliegenden Buchs, Faktorenanalyse und Diskriminanzanalyse, erscheinen mir dagegen insgesamt gut gelungen. Das liegt sicherlich daran, dass beiden Methoden eine geschlossene mathematische Theorie zugrunde liegt, zum anderen aber auch daran, dass in der Dar-

stellung der Theorie durch Schendera das „Kochbuchprinzip“ aufgegeben worden ist. Insbesondere wird nun die Theorie der jeweiligen Methode in eigenständigen Texten beschrieben. Sie sind zwar alles andere als für „Einsteiger“ geeignet, da sie in relativ dichter Form einen Überblick über das jeweilige Gebiet geben und die Theorie nicht im Stil eines Lehrbuchs entwickeln. Für den Leser, der sich hier schon einmal umgetan hat, geben sie jedoch einen guten Überblick.

Das gilt besonders für die Darstellung der Faktorenanalyse, bei der die Darstellung der Grundformen und Varianten sowie der Grundzüge der Anwendung und der Schätzungen sehr systematisch entwickelt sind. In fünf sorgfältig durchgeführten Beispielanalysen werden Varianten der Faktorenanalyse vorgeführt und die Ergebnisse interpretiert. Abschließend werden noch die „Faktorenanalyse von Fällen“, ein an Clusteranalyse erinnernder Ansatz, und die „Faktorenanalyse eingeleiteter Fälle“ dargestellt. Letzteres Verfahren ist etwa bei ordinalen Variablen anzuwenden, bei denen anstelle der Korrelationskoeffizienten Spearman- oder Kendallkoeffizienten analysiert werden.

Analoges ist von der Darstellung der Diskriminanzanalyse zu sagen. Einer Darstellung der Theorie, die ebenfalls nicht so einfach für „Einsteiger“ geeignet sein dürfte, folgen zwei ebenfalls sorgfältig präsentierte Beispielanalysen.

Man möchte dem Buch wünschen, dass in künftigen Auflagen das Thema Clusteranalyse eine sorgfältige Überarbeitung erfährt, die es auf das Niveau der beiden letzten Teile bringt.

Literatur

Bacher, J., K. Wenzig und M. Vogler, 2004: SPSS Two-Step Cluster – A First Evaluation. Friedrich-Alexander-Universität Erlangen-Nürnberg. Sozialwissenschaftliches Institut. Lehrstuhl für Soziologie. Arbeits- und Diskussionspapiere 2004-2 (2. Auflage).

Ankündigungen

Call for Papers

Frühjahrstagung 2011 der Sektion Jugendsoziologie

„Methoden der Jugendforschung –
angemessene Antworten auf
neue Herausforderungen“

TU Darmstadt

23. - 24. März 2011

Frühjahrstagung 2011 – Sektion Jugendsoziologie

Die Frühjahrstagung 2011 der Sektion Jugendsoziologie beschäftigt sich mit methodischen Problemen der Jugendforschung und fragt nach Perspektiven für die Weiterentwicklung der Methodologie der Jugendforschung. Ziel der Tagung ist es, sowohl einen Überblick über die aktuellen methodischen Herausforderungen der empirisch arbeitenden Jugendsoziologie als auch über die methodischen Innovationen, Designs, Methoden und Techniken zu geben, mit denen die Jugendforschung auf diese Herausforderungen reagiert.

In der empirischen Jugendforschung haben sich für verschiedene Gegenstände und Themen Traditionen von üblicherweise genutzten qualitativen und quantitativen Forschungsmethoden herausgebildet. Das Anliegen dieser Tagung ist es, die Angemessenheit dieser methodischen Forschungstraditionen zu reflektieren, sowohl aus thematischer und theoretisch-inhaltlicher Sicht als auch mit Blick auf die sozialen, kognitiven und rechtlichen Besonderheiten der jungen Beforschten, und Anstöße für die Weiterentwicklung des methodischen Instrumentariums der Jugendforschung zu geben.

Mit dem Wandel von Jugend und Jugendphase verändern sich die inhaltlichen Fragestellungen der Jugendforschung, z. B. durch die vermehrte Nutzung mobiler Kommunikation (Handys) und des virtuellen Raumes (SchülerVZ, Twitter, Facebook, ICQ, Onlinespiele). Daraus ergeben sich jedoch auch neue methodische Fragestellungen und Herausforderungen (qualitative Verfahren mit Messenger-Systemen, Online-Gruppendiskussionen, Mobiltelefonstichproben, Online-Befragungen).

Die erkenntnisleitende Frage der Tagung wird sein, wie sich die empirische Jugendsoziologie in den letzten zwei Jahrzehnten methodisch entwickelt hat, welche Vorgehensweisen und methodischen Konzepte für welche bekannten und neu entdeckten Themenbereiche Anwendung finden, wie sie methodisch und methodologisch zu bewerten sind und welche zukunftsgerichteten Überlegungen vorliegen.

Daraus leiten wir folgende potenzielle thematische Schwerpunkte als Gegenstand der Tagung ab:

- Güte der gegenwärtigen Forschung? Wie methodisch angemessen sind die qualitativen, quantitativen oder integrativen Designs, die Methoden und Techniken der Jugendsoziologie zur Erfassung ihrer Gegenstände? Wo liegen Stärken, wo bestehen Probleme (z. B. bei der Validität der Selbstauskünfte Jugendlicher)? In welchem Verhältnis stehen Kompetenzmessungen und Assessments zu begleitenden Befragungen? Wie verändert die Erfahrung der Jugendlichen mit Kompetenzmessungen ihren Umgang mit sozialwissenschaftlichen Datenerhebungen? Welche Erfahrungen bestehen mit dem Feldzugang und dem Aufenthalt im Feld?
- Allgemeine Methode – spezielle Population? Lassen sich die für Erwachsene etablierten qualitativen und quantitativen Methoden der empirischen Sozialforschung ohne Weiteres auf Jugendliche übertragen? Welche typischen oder spezifischen Einschränkungen (z. B. bei Feldzugang, Einwilligung, Datenschutz, durch geänderte politische Rahmenbedingungen) gibt es? Welche speziellen Methoden sind verfügbar bzw. erforderlich? Welche Besonderheiten bei der Stichprobengenerierung sind zu beachten (komplexe Stichproben, school-based studies, Mobiltelefon- und Onlinestichproben)?
- Historische Entwicklung? Wie haben sich die Methoden und Techniken der empirischen Jugendforschung über die Zeit verändert? In welchem Zusammenhang stand dies mit den Veränderungen der Methodologie und den methodischen Standards in der empirischen Sozialforschung einerseits und der theoretischen Ansätze in der Jugendforschung andererseits? Und umgekehrt: Welcher Einfluss der methodischen Konzepte der Jugendforschung bzw. Jugendsoziologie besteht auf die allgemeine Entwicklung der empirischen Sozialforschung?
- Absehbare und notwendige zukünftige Entwicklungen? Welche neuen Designs und Methoden bieten sich an, um anstehende oder bislang noch nicht behandelte inhaltliche Felder der Jugendsoziologie methodisch angemessen zu bearbeiten?

Angesprochen und eingeladen, sich zu beteiligen, sind Sozialwissenschaftlerinnen und Sozialwissenschaftler, die sich mit der qualitativen und/oder quantitativen empirischen Erhebung und Analyse zu Fragestellungen der Jugendforschung beschäftigen. Gegenstand der Vortragsangebote sollen methodische Fragestellungen sein, nicht dagegen inhaltliche Beiträge, in denen u. a. die verwendeten Methoden diskutiert werden. Abstract (ca. 1.500 Zeichen) sind bis zum **1. Dezember 2010** zu richten an:

Prof. Dr. Marek Fuchs
Professur für Empirische Sozialforschung
Institut für Soziologie
Technische Universität Darmstadt
64283 Darmstadt
fuchs@ifs.tu-darmstadt.de

apl. Prof. Dr. Jens Luedtke
Institut für Gesellschafts- und Politikanalyse
Goethe-Universität Frankfurt a. M.
60054 Frankfurt a. M.
jens.luedtke@soz.uni-frankfurt.de

Die Tagung soll am 23./24. März 2011 an der TU Darmstadt stattfinden; es wird eine Tagungsgebühr von voraussichtlich 40€ erhoben.

Call for Papers

**European Labour Force Survey
(EU-LFS) and
European Union Statistics on Income
and Living Conditions (EU-SILC):
2nd European User Conference**

Organized by
German Microdata Lab, GESIS,
in cooperation with Eurostat

Mannheim

March 31 - April 1, 2011

*European User
Conference
EU-LFS and
EU-SILC*

EU-LFS and EU-SILC are the most important official micro data for comparative social research in Europe. An increasing number of researchers use these data for a wide range of economic and social analyses. Topics addressed include e.g. labour migration, integration of immigrants, monetary poverty, income mobility, income inequality, material deprivation, working poor, gender gaps on the labour market. In the context of the rapidly growing use of EU-LFS and EU-SILC data there is a need to share experience between researchers, to provide feedback to producers, but also to learn more about the way these European statistics are developed, compiled and disseminated.

The 2nd European User Conference for EU-LFS and EU-SILC, organized by the German Microdata Lab, GESIS, in cooperation with Eurostat, will provide researchers with the opportunity to present and discuss their work and share their experience. In addition to fostering the discussion within the research community on both substantive and methodological issues, the conference offers researchers the opportunity to give feedback to the European Statistical System. Given that the legal bases for EU-LFS and EU-SILC will be revised in the near future the conference offers the unique possibility to discuss needs and wants of the research community with Eurostat.

Researchers of all disciplines (e.g. economics, demography, geography, political science, public health and sociology) who use either EU-LFS or EU-SILC micro data are encouraged to participate and to submit an abstract. Substantive topics may include, among others, all aspects of the European labour market, living conditions, migration, income

inequalities, poverty and social exclusion. Methodological topics may include e.g. questions of data quality, cross-national and inter-temporal comparability, and statistical modeling. All presentations must be comparative and include data from at least two countries.

SUBMISSIONS:

The deadline for submissions of abstracts is **October 31, 2010**. Please send your submission to the local organizers of the conference:

Christof Wolf (Christof.Wolf@gesis.org) and Heike Wirth (Heike.Wirth@gesis.org).

Abstracts should not be longer than 1000 words; the abstract should be informative, indicate the dataset(s) used and the countries analyzed. The submitters will be notified by **January 15, 2011**, whether their paper was accepted for presentation or not.

For any further questions please contact the local organizers and visit the conference webpage at

<http://www.gesis.org/forschung-lehre/veranstaltungen/european-user-conference-2/>

Hinweise für unsere Autorinnen und Autoren

Methoden – Daten – Analysen (MDA) veröffentlicht Beiträge aus dem Bereich der Empirischen Sozialforschung, insbesondere aus dem Bereich der Umfragemethodik. Im Vordergrund stehen Artikel, welche die methodischen und/oder statistischen Kenntnisse der Profession erweitern, sowie Beiträge, die sich mit der Anwendung der Methoden der Empirischen Sozialforschung in der Forschungspraxis beschäftigen, oder solche, in denen ein statistisches Verfahren exemplarisch angewandt wird. Obwohl der Schwerpunkt auf Umfragemethoden liegt, sind Beiträge zu anderen methodischen Bereichen willkommen. Die Artikel sollen für eine breite Leserschaft von Wissenschaftlern und Praktikern im Bereich der Empirischen Sozialforschung verständlich sein.

Manuskripte, die bereits an anderer Stelle veröffentlicht sind oder gleichzeitig anderen Publikationsorganen zur Veröffentlichung angeboten worden sind, werden grundsätzlich nicht berücksichtigt. Eine spätere Veröffentlichung eines in der MDA erschienenen Beitrages ist möglich, sofern an exponierter Stelle auf die Ersterscheinung des Beitrages in der MDA hingewiesen wird.

Jeder Beitrag, der zur Veröffentlichung in MDA eingereicht wird, wird zunächst von den Herausgebern danach bewertet, ob er für eine Veröffentlichung grundsätzlich in Frage kommt.

Falls die Herausgeber einer Veröffentlichung grundsätzlich ablehnend gegenüber stehen, werden die Autoren unter Angabe von Gründen für diese Entscheidung informiert.

Falls die Herausgeber zur Ansicht gelangen, dass der Beitrag grundsätzlich zur Veröffentlichung in Frage kommt, wird er anonymisiert an mindestens zwei unabhängige Gutachter verschickt, die um eine Stellungnahme gebeten werden. Im Zweifelsfalle wird ein drittes Gutachten eingeholt.

Wird ein Beitrag nach Beschluss der Herausgeber in das Begutachtungsverfahren gegeben, erfolgt die abschließende Entscheidung über ein Manuskript auf der Basis der Gutachten durch die Herausgeber. Im Falle einer Ablehnung erhalten die Autoren eine ausführliche Begründung für die Ablehnung. Wird eine Überarbeitung eines Beitrages für erforderlich gehalten, erhalten die Autoren detaillierte Überarbeitungshinweise.

Unabhängig vom Ergebnis des Begutachtungsverfahrens werden die Autoren von der Entscheidung durch die Redaktion per E-Mail informiert.

Die folgenden Regeln sind bei der Abfassung von Manuskripten zu beachten:

Manuskripte müssen per E-Mail (mda@gesis.org) eingereicht werden. Der Umfang der Manuskripte soll inklusive Leerzeichen alles in allem nicht mehr als 70.000 Zeichen betragen.

Den Beiträgen sind Abstracts in Deutsch und Englisch (jeweils ca. 15 Zeilen) voranzustellen. Auch der Titel des Beitrages ist in Deutsch und Englisch einzureichen.

Um die Anonymität der Beiträge zu wahren, darf in einem Manuskript nur der Titel des Beitrages enthalten sein, nicht aber Namen oder Anschriften der Autoren; Name und Anschrift der Autoren müssen, gemeinsam mit dem Titel des Beitrages, auf einer separaten Seite eingereicht werden.

Beiträge sind mit dem Dezimalklassifikationssystem zu untergliedern (1 - 2 - 2.1 - 2.2 - 3 usw.). Die Gliederungstiefe geht dabei höchstens auf *eine* Stelle nach dem Punkt.

Tabellen enthalten Tabellenummer und Titel im Tabellenkopf, Abbildungen werden analog behandelt.

Grafiken sind mittels gängiger Grafiksoftware zu erstellen. Ist eine spezielle Grafiksoftware erforderlich, übernimmt der Autor/die Autorin die endgültige Formatierung der Grafiken in eigener Regie.

Bei der Erstellung von Tabellen und Grafiken ist zu berücksichtigen, dass der Satzspiegel 11,5 cm (Breite) x 18,5 cm (Höhe) beträgt. Die Grafiken sind als jpeg- oder tif-Dateien in *Graustufen (CMYK)* mit einer Auflösung von mindestens 300 dpi zu liefern.

Die Beiträge sind unter Wahrung der gültigen Rechtschreiberegungen (neue Rechtschreibung) zu erstellen.

Werden in einem Beitrag empirische Daten verwandt, muss die Möglichkeit der Replikation bestehen. Im Falle einer Veröffentlichung in der MDA erklären sich die Autoren daher schriftlich bereit, Dritten auf deren Anfrage hin die Daten und Programmroutinen zur Verfügung zu stellen.

Anmerkungen und Fußnoten sind mit der Fußnotenfunktion des Schreibprogrammes (im Normalfalle Word) zu erstellen; bitte nicht gesondert formatieren. Fußnoten sind nur für inhaltliche Kommentare vorzusehen, nicht für bibliographische Hinweise.

Literaturhinweise im Text sind nach den folgenden Mustern aufzuführen: Müller (2002) – Schulze und Mayer (2003) – Müller, Mayer und Schulze (2004) – Müller et al. (2005) – Müller (2006: 75) – (vgl. Müller 2007: 75) – (Müller 2008; Mayer/Müller/Schulze 2009).

Das Literaturverzeichnis ist wie folgt zu gestalten:

Buchveröffentlichungen:

Strobl, R. und W. Kühnel, 2000: Dazugehörig und ausgegrenzt. Analysen zu Integrationschancen junger Aussiedler. Weinheim/München: Juventa.

Zeitschriftenbeiträge:

Becker, R., R. Imhof und G. Mehlkop, 2007: Die Wirkung monetärer Anreize auf den Rücklauf bei einer postalischen Befragung und die Antworten auf Fragen zur Delinquenz. Empirische Befunde eines Methodenexperiments. Methoden – Daten – Analysen. Zeitschrift für Empirische Sozialforschung 1 (2): 131-159.

Beiträge in Büchern:

Braun, M. und I. Borg, 2004: Berufswerte im zeitlichen und im Ost-West-Vergleich. S. 179-199 in: R. Schmitt-Beck, M. Wasmer und A. Koch (Hg.): Sozialer und politischer Wandel in Deutschland. Analysen mit ALLBUS-Daten aus zwei Jahrzehnten. Wiesbaden: VS-Verlag für Sozialwissenschaften.

Internetquellen:

Stadtmüller, S. und R. Porst, 2005: Zum Einsatz von Incentives bei postalischen Befragungen. GESIS How-to-Reihe, Nr. 14. Mannheim: GESIS. http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/howto/how-to14rp.pdf (1.12.2008).

Datenfile:

Forschungsgruppe Wahlen, Mannheim: Zur politischen Lage in Niedersachsen im Januar 2008. GESIS Köln, Deutschland ZA4863 Datenfile.