

Otto-Friedrich-Universität Bamberg
Lehrstuhl für Soziologie, insbes. Survey-Methodologie

Fortgeschrittene Methoden der Datenerhebung:
Datenerhebung und Fehlerquellen
Sommersemester 2020



Gibt es verdächtige Interviewer?

Eine Analyse zur Identifikation potenziell gefälschter
Interviews im EES 2002

Dozent: Herr Prof. Dr. Mark Trappmann
Prüfling: Kevin Glock
E-Mail: kevin.glock@stud.uni-bamberg.de
Semster: 2. FS, MiSS
Matrikelnr.: 2020296
Abgabe: XX. Februar 2020
Zeichenanzahl (ohne Tabellen):

1. EINLEITUNG	5
2. FORSCHUNGSSTAND	7
2.1. FORMEN VON FÄLSCHUNGEN	7
2.2. VERZERRUNGEN DURCH FÄLSCHUNGEN	8
2.3. PSYCHOLOGISCHE SICHTWEISE AUF INTERVIEWFÄLSCHUNG	9
2.4. ERKENNUNG POTENZIELLER FABRIKATE	12
2.5. FORSCHUNGSFRAGE	15
3. DATEN & METHODIK	17
3.1. VARIABILITÄT UND RELIABILITÄT	19
4. ERGEBNISSE.....	23
5. ZUSAMMENFASSUNG.....	26
6. LITERATUR.....	29
7. EIGENSTÄNDIGKEITSERKLÄRUNG	33

Bias eines Schätzers

Summe der Abweichungsquadrate

Cronbachs Alpha

Summe der Abweichungsquadrate

Bias der Variabilität

Totale Variabilität des Interviewer I_i

Tabellenverzeichnis

TABELLE 1 ANZAHL DER KONTAKTVERSUCHE DES ERSTKONTAKTES (OHNE 10 UND 99=SYS MIS), QUELLE ESS 2002 ROUND 1 HUNGARY	17
TABELLE 2 ANZAHL DER INTERVIEWER PRO INTERVIEWS BASIEREND AUF DEN ERSTELLTEN SUBSET VON KONTAKTVERSUCHEN, QUELLE: EIGENE BERECHNUNGEN	18
TABELLE 3 GLOBALES INTERVIEWER-RANKING ERSTE 29 PERSONEN NACH SCHÄTZER T_i UND PLAUSIBILITÄT FÜR POTENZIELLE FÄLSCHUNG, QUELLE: EIGENE BERECHNUNGEN	24
TABELLE 4 MITTLERE TOTALE VARIANZ INNERHALB EINES CLUSTERS MIT STANDARDABWEICHUNG, QUELLE: EIGENE BERECHNUNGEN	25

Abbildungsverzeichnis

ABBILDUNG 1 THEORETISCHE KOSTEN-NUTZEN-FUNKTIONEN VON BEFRAGENDEN UND AUFTRAGGEBERN ...	11
ABBILDUNG 2 POSTDIKTIVER ERKLÄRUNGSWEG	13
ABBILDUNG 3 THEORETISCHE ITEM-DIMENSIONALITÄT.....	20
ABBILDUNG 4 EMPIRISCHE (SCHWARZ) UND SIMULIERTE (GRAU) DICHTEFUNKTION DER TOTALEN VARIABILITÄT, QUELLE: EIGENE BERECHNUNGEN	23

Formelverzeichnis

(1) BIAS EINES SCHÄTZERS $\hat{\vartheta}$	8
(2) SUMME DER ABWEICHUNGSQUADRATE	8
(3) CRONBACHS ALPHA	19
(4) SUMME DER ABWEICHUNGSQUADRATE	20
(5) THEORETISCHER BIAS DER VARIABILITÄT	21
(6) TOTALE VARIABILITÄT DES INTERVIEWER I_i	21

Abstract

Fake interviews are a central problem of empirical social research and can lead to significant distortions of parameter estimators. The prevention and detection of falsifications are therefore essential to ensure reliable data through interviews and valid estimates. This paper describes and evaluates the procedure of identifying potential originators of counterfeits using the variability method, using the example of the European Social Survey (Round 1) of the partial survey Hungary 2002. Four interviewers could thus be identified who require follow-up investigation. The variability method has the advantage that it can be implemented in a cost and time efficient manner. According to literature it provides equally reliable results compared to advanced statistical methods. To ensure full reproducibility, the analysis is performed in R. The analysis code can be found in the Annex.

Keywords: Fake interviews, Interview fabrication, variability, R, resample, survey research

Zusammenfassung

Interviewfälschungen sind ein zentrales Problem der empirischen Sozialforschung und können zu bedeutenden Verzerrungen von Parameterschätzern führen. Die Vermeidung und Erkennung von Fälschungen sind daher essenziell, um reliable Daten durch Befragungen und valide Schätzungen zu gewährleisten. Die Arbeit beschreibt und evaluiert das Vorgehen der Erkennung von potenziellen Verursachern von Fälschungen anhand der Variabilitätsmethode am Beispiel des European Social Surveys der Teilerhebung Ungarn 2002 (Round 1). 4 Interviewer konnten damit entdeckt werden, die einer Nachuntersuchung bedürfen. Die Variabilitätsmethode hat den Vorteil, dass sie kosten- und zeiteffizient implementierbar ist. Laut Literatur liefert sie gegenüber erweiterten statistischen Verfahren ebenso zuverlässige Ergebnisse. Um eine volle Reproduzierbarkeit zu gewährleisten erfolgt die Analyse in R. Der Analysecode findet sich im Anhang.

Keywords: Interviewfälschung, Interviewfabrikation, Variabilität, R, Resampling, Surveyforschung

1. Einleitung

Die systematische Generierung von Informationen, um eine Gruppe von Individuen mittels zufälliger und komplexer Stichprobenziehungsmethoden zuverlässig quantitativ beschreiben zu können, ist ein viel diskutiertes Forschungsgebiet in der Surveyforschung. Daten zu generieren, die valide wissenschaftliche Aussagen erlauben, sind dabei insbesondere in großangelegten sozialwissenschaftlichen Bevölkerungsbefragungen eine besonders komplexe Herausforderung. Viele sozialwissenschaftliche Fragestellungen sind jedoch nur mittels der systematischen Methode der Befragung beantwortbar. Dabei ist es nötig sich auf die Bereitschaft von befragten Personen und Befragenden zu verlassen, um gültige Messergebnisse zu erlangen.

Eine bedeutende Problematik stellt dabei eine wachsende Unit-Non-Response-Rate dar (Schnell 2013), die dazu führt, dass zunehmend größere Auswahlgesamtheiten definiert werden müssen, um eine hinreichend große Ausschöpfungsquote generieren zu können. Damit wird die Feldarbeit zunehmend belastender für Interviewende und kann zu Frustrationen durch Nichterreichbarkeit und häufige Re-Issuings bei gleicher Bezahlung bei mehr Arbeitsaufwand führen. Vorsätzliche Interviewfälschungen sind daher ein einfaches Mittel für Interviewende, um ihren Lohn, bei Minimierung des Arbeitsaufwandes, zu maximieren. Systematische Verzerrungen durch Fälschungen stellen jedoch ein schwerwiegendes Problem für die Konstruktion erwartungstreuer und konsistenter Schätzer dar, und wären an sich leicht vermeidbar. Obwohl die Vermeidung und Detektion von Fälschungen ein zentrales Problem in der empirischen Sozialforschung darstellen (Blasius und Thiessen 2012, 2013; Blasius und Friedrichs 2012), gibt es bisher nur mäßig viele Publikationen in Bezug auf Interviewfälschungen (Blasius 2014, S. 323).

Die vorliegende Arbeit verfolgt daher die Frage, ob potenzielle Interviewfälschungen adäquat anhand ihrer abweichenden Streuung gefiltert werden können, um damit potenzielle Fälschungen einer Nachuntersuchung zuzuführen. Dies erfolgt anhand des European Social Survey Datensatzes der Teilerhebung Ungarn 2002 (Round1) (European Social Survey ERIC (ESS ERIC) 2003). Potenzielle Interviewfälschungen werden auf Grundlage der geclusterten totalen Variabilität im Antwortverhalten identifiziert, um Interviews zu entdecken, die höchst wahrscheinlich fabriziert wurden und einer weiteren Nachuntersuchung bedürften.

Im Folgenden wird dafür der Stand der Forschung evaluiert und verschiedene Methoden diskutiert. Im Weiteren erfolgt die Darstellung der Ergebnisse der Variabilitätsanalyse zur Identifikation potenzieller Interviewfälschungen. Am Ende der Arbeit erfolgt eine ausführliche Diskussion der Ergebnisse, da sich dieses Verfahren zur Detektion potenzieller nicht reliabler Interviews eignet, dabei kosten- und zeiteffizient und dabei einfach implementierbar ist. Das Verfahren erfolgt mittels deskriptiver Statistik und nutzt einen Resampling-Ansatz zur Überprüfung auf Plausibilität der potenziellen Fälschung, indem empirische Streuungen in jeweiligen Interview-Clustern mit zu erwartenden Varianzen verglichen werden. Dies entspricht einem statistischen linksseitigem Test. Nachteilig ist, dass nur potenzielle Fälschungen betrachtet werden, weshalb eine endgültige Entscheidung einer Hypothese bis zur Nachuntersuchung aussteht.

2. Forschungsstand

Im Folgenden soll auf die wichtigsten Erkenntnisse aus der Forschung zu Interviewfabrikation hinsichtlich der Folgen für Schätzverfahren, der psychologischen Beweggründe von Fälschern und der methodischen Aufdeckung potenzieller Fabrikate eingegangen werden.

2.1. Formen von Fälschungen

Unterschieden werden verschiedene Arten von Fälschungen, die sich nach dem Grad der Fälschung richten. Einerseits werden Duplikate in Datensätzen als Fälschungen angesehen (Blasius und Thiessen 2012, 2015; Kuriakose und Robbins 2016), andererseits aber auch die falsche Umsetzung einer Interviewanweisung oder systematisch fabrizierte Interviews (Blasius 2014, S. 325), die nie eine zu befragende Person erreicht haben. Darüber hinaus gibt es speziellere Formen von Teilfälschungen, wie Nutzung von abweichenden Erhebungsmethoden oder das Eingreifen der befragenden Person, die stellvertretend antwortet oder aus eigenem Antrieb einen passenden Wert für ein Item angibt (Schräpler und Wagner 2005, S. 7).

Blasius (2014) führt aus, dass laut „AAPOR (2003: 2) [...] ein Interview bereits gefälscht [ist], wenn eine der folgenden Aussagen zutrifft: >>a) fabricating all or part of an interview – the recording of data that are not provided by a designated survey respondent and reporting them as answers of that respondent; b) deliberately misreporting disposition codes and falsifying process data (e.g., the recording of a refusal case as ineligible for the sample; reporting a fictitious contact attempt); c) deliberately miscoding the answer to a question in order to avoid follow-up questions; d) deliberately interviewing a non-sampled person in order to reduce effort required to complete an interview; or e) otherwise, intentionally misrepresenting the data collection process to the survey management.<<“ (S. 325).

Schnell (1991) betrachtet hingegen Fälschungen von Interviewern als eine spezielle Form der Imputation, die auf den Annahmen der fälschenden Person beruht, wie einzelne Charakteristiken von Befragten streuen. Sie entsprechen also einem Expertenschätzung (Schnell 1991, S. 26).

In der vorliegenden Arbeit werden Fabrikate als bewusste Fälschungen durch Imputation von Werten durch die befragende Person im Sinne von Schnell (1991) und

Schräpler und Wagner (2005) verstanden. Dabei werden Teilfälschungen und auch vollständig fabrizierte Interviews als gleichwertig schwerwiegend verstanden.

2.2. Verzerrungen durch Fälschungen

Fälschungen von Interviews können zu bedeutenden Verzerrungen von Parameterschätzern führen und machen Daten im schlimmsten Fall unbrauchbar, weshalb ihre Erkennung und Aussortierung notwendig sind. Im Folgenden werden diese Verzerrungen erläutert.

Der Bias eines Schätzers $\hat{\vartheta}$ ist die systematische Abweichung des Schätzers $\hat{\vartheta}$ vom wahren Wert ϑ in der Grundgesamtheit.

$$Bias(\hat{\vartheta}) = \hat{\vartheta} - \vartheta$$

(1)

Er ist bedeutend für die konsistente und erwartungstreue Schätzung mittels einer Stichprobe. Die systematische Imputation von Werten durch Befragende oder auch Befragte verursacht solch einen Bias und sollte möglichst minimal sein. Durch die Testkonstruktion, also Nutzung von Filterfragen, variablen Fragekonstruktionen und Sortierungen der Antwortkategorien kann der Survey so konstruiert werden, dass der Test nicht ermüdend wirkt, so dass es zu Extremwerttendenz, Tendenz zur Mitte und anderen Testeffekten kommt. Ziel der Testkonstruktion sollte daher sein, mögliche Varianzveränderungen adäquat herbeizuführen und damit eine hohe Variabilität des Antwortverhaltens von Befragten abdecken zu können.

Schnell (1991) betrachtet Fälschungen von Interviewern als eine spezielle Form der Imputation, die auf den Annahmen dieser beruht, wie einzelne Charakteristiken von Befragten streuen. Diese Imputation führen in der Folge zu einem Bias des Schätzers $\hat{\vartheta}$ für einzelne Items, Gruppen von Items oder aber ganze Befragungen, je nach Anteil der Fälschungen.

Schlussfolgernd daraus lässt sich damit herleiten, dass die Variabilität über alle Stichprobenstatistiken Y_i mit $i=1, \dots, n$, als Summe der Abweichungsquadrate

$$SQ(Y_i) = \sum_{i=1}^n (y_i - \bar{y})^2$$

(2)

hinreichend groß sein sollte, das heißt alle Kategorien durch Befragte abgedeckt werden und damit eine Tendenz zur Mitte nicht vorliegt. Extremwerttendenzen hingegen würden zu einer Maximierung von $SQ(Y_i)$ führen, da sie in allen Fällen maximal vom globalen Mittelwert abweichen. Anzunehmen wäre hiernach, dass Fälschungen tendenziell eine zu geringe Variabilität aufweisen und Konsistenz zeigen.

Durch analytische Abschätzung schlussfolgert Schnell (1991), dass einfache Statistiken bei größeren Fallzahlen und kleinen Anteilen von Fälschungen robuste Werte abbilden (S. 35). In Anbetracht das durch Fälschungen deutlicher Verzerrungen multivariater Analysen herbeigeführt werden können (Menold et al. 2013; Schräpler und Wagner 2005; Schnell 1991), sind Fabrikate in den Daten schwerwiegend, da Ergebnisse solcher Interviews im weitreichendsten Falle unbrauchbar sind. Schnell (1991) erläutert, dass Imputationen durch Fälscher an sich nicht problematisch für die Abschätzung der Verzerrung wären, sofern zufällig generierte Werte eingesetzt würden (S. 29).

Das Problem ist hierbei die potenzielle Anwendung von Laientheorien über die Streuung der Ausprägungen, die zu einem systematischen Fehler führen (Schnell 1991, S. 29). Es ist zudem wahrscheinlich, dass Interviewfälschungen ähnliche Antwortmuster aufwiesen, innerhalb als auch zwischen Interviews einer befragenden Person. Das heißt Item-Batterien werden ähnlich ausgefüllt, demografische Merkmale wiederholen sich unplausibel und Filterfragen werden übergangen. „[D]ie Konsistenz der Angaben in den gefälschten Interviews ist etwas größer als in echten Interviews“ (Schnell 1991, S. 28). Laut (Schäfer et al. 2004a, S. 11) ist evident, dass Fälschungen daher eine geringere Variabilität innerhalb eines Interviewer-Clusters aufweisen. Somit decken sich die Definitionen. Die Vermeidung und Aufdeckung von Fabrikaten muss daher essenzieller Bestandteil des Qualitätsmanagements in methodologischen Arbeitsbereichen von Umfrageinstituten sein.

2.3. Psychologische Sichtweise auf Interviewfälschung

Im Folgenden soll auf die wichtigsten psychologischen Beweggründe eingegangen werden, die in der Literatur erwähnt werden. Interviewende produzieren aus unterschiedlichen Gründen Fälschungen im Rahmen ihrer Tätigkeit (Blasius 2014; Schäfer et al. 2004b). Ein Grund für Fälschungen sei die Demoralisation von Befragenden (Crespi 1945). Einerseits durch die Institute selbst andererseits durch die Befragenden (Blasius 2014). „So können eine zu lange Dauer des Interviews, zu viele Nachfragen [...], zu komplexe und zu schwierige Fragen den Interviewer demoralisieren. Ein

weiterer möglicher Grund ist der Druck durch den Supervisor [...] eine bestimmte Anzahl von Interviews durchzuführen“ (Nelson und Kiecker 1996, S. 1110; Blasius 2014, S. 324).

Andererseits stimuliert riskantes Verhalten Individuen mit einem psychologischen Reiz der Belohnung, was als möglicher Grund angeführt wird, Interviewfälschung zu betreiben (Harrison und Krauss 2002). Insbesondere, wenn Teilfälschungen von Individuen betrieben werden, die ein hohes Wissen über Populationsparameter (bspw. Studenten) und einen hohen Grad der Motivation aufweisen riskantes Verhalten auszuprägen, ist das Risiko sehr hoch, dass diese Art der Fälschung betrieben wird.

Allerdings sollten dabei weitere motivationale Faktoren, wie Persönlichkeitsmerkmale und Einstellungen (Buehl und Melchers 2017, S. 1), in Betracht gezogen werden, um fälschendes Verhalten zu zeigen. Dabei sind das Risiko entdeckt zu werden und die Konsequenzen tendenziell gering (Menold und Kemper 2014; Blasius 2014) was einen zusätzlichen Anreiz für fälschende Personen darstellt.

Einheitlich wird in der Literatur hingegen auf rationale Beweggründe verwiesen (Blasius 2014; Schäfer et al. 2004b). So bedeutet der Erhebungsprozess im Feld für Interviewende und für Erhebungsinstitute einen erheblichen Zeit- und Kostenaufwand. Einerseits nimmt die Bereitschaft an Befragungen teilzunehmen seit Jahren zunehmend ab (Schnell 2013). Eine zunehmend steigende Unit-Non-Response-Rate, sorgt in Folge für steigenden Druck der Interviewenden, im ursprünglich ersten Versuch ein Interview durchführen zu können. Andererseits führt dies durchschnittlich auch zu einer höheren Anzahl von Versuchen eine beliebige Kontaktperson in einem Wohngebiet zu erreichen und eine hinreichende Anzahl an Interviews in der Stunde durchführen zu können, um eine angemessene Bezahlung zu erreichen, vornehmlich wenn sie per durchgeführten Interview und nicht per Stunde entlohnt werden (Blasius und Friedrichs 2012) oder Re-Issuings nur einen geringeren Lohn bedeuten.

Die rationalistische Vereinfachung bedeutet also, dass sich die individuellen Kosten-Nutzen-Funktionen von Interviewenden und Auftraggeber diametral entgegenstehen (siehe Abbildung 1). Je mehr erneute Re-Issuings erfolgen, desto größer wird dieser Trade-Off für die Interviewenden, was im Falle des ESS damit unterbunden wird, dass Re-Issuings, wenn überhaupt, durch andere Interviewende durchgeführt werden. Betrachtet man Interviewende im Szenario des Erhebungsprozesses als rationale Akteure so lässt sich der Nutzen (Entlohnung per Interview) individuell erhöhen, wenn

auf Ehrlichkeit und Arbeitsaufwand verzichtet wird und Interviews systematisch gefälscht werden, um so den individuell größten Nutzen bei geringsten Kosten (Zeitaufwand, körperlicher Einsatz durch Erreichen von Wohngebiet und Wohneinheiten, psychischer Aufwand durch potenziell harsche Abweisung durch Interview-Verweigerer) zu erreichen.

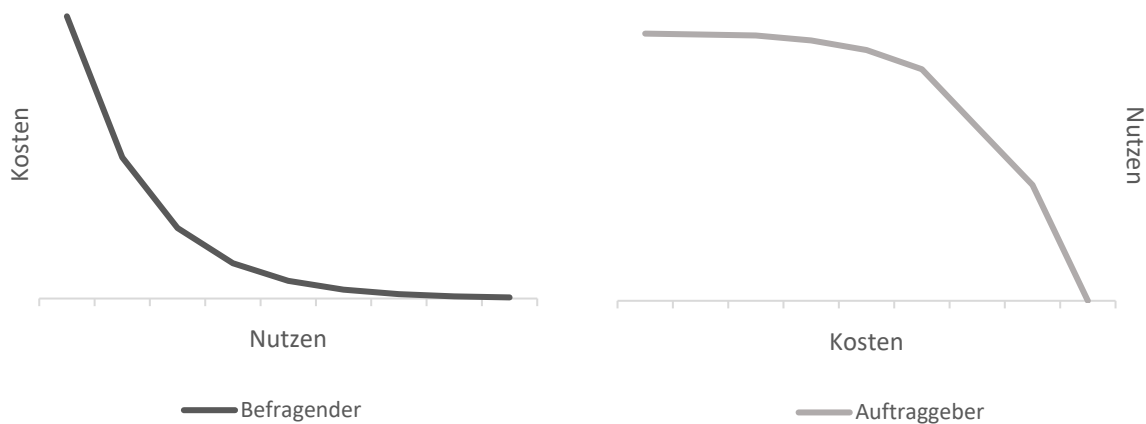


Abbildung 1 Theoretische Kosten-Nutzen-Funktionen von Befragenden und Auftraggebern

Betrachtet man Interviewfabrikation aus persönlichkeitspsychologischer Sicht, kann diese als Lügen oder Unehrllichkeit (Menold und Kemper 2014) verstanden werden. Je nach Reichweite des Fabrikationsbegriffs beabsichtigen Interviewer durch bewusste oder unbewusste Handlungen sich einen persönlichen Vorteil zu ermöglichen, dessen Erlangung ohne Lüge nur mit höherem Aufwand erreichbar wäre beziehungsweise dessen Erfolgswahrscheinlichkeit tendenziell subjektiv geringer ausfällt.

Fabrikation ist folglich eine komplexe Lügenkonstruktionen, weshalb es nötig wird diagnostische Denkweisen auf die Daten anzuwenden. So führen Schräpler und Wagner (2005) und Blasius (2014) an, dass „Interviewer at risk“ (Bushery et al. 1999) häufig den „shortest path through the interview“ (Hood und Bushery 1997, S. 821) nehmen, „[they] avoid choosing the category >>others, please specify<<“ (Menold und Kemper 2014, S. 45) und „producing survey participants who live in one person families and are >>white, non-smoker, [without] health problems and no health insurance<<“ (Hood und Bushery 1997, S. 821; Menold und Kemper 2014, S. 45), also Durchschnittspersonen mit geringer Streuung. Diese Definition deckt sich somit auch mit der theoretischen Annahme von, dass Fälschungen eine geringere Variabilität aufweisen (Schäfer et al. 2004a, S. 11).

Um divergentes Verhalten von Interviewenden zu detektieren ist es nötig kriminologische Fragen an die Absichten des Fälschers zu stellen (Weinauer 2019), um nachzuvollziehen wieso bestimmte Verhaltensweisen gewählt werden, um mit der Fälschung verbundene Risiken zu minimieren. Um Fälscher aufzudecken ist es nötig, verstehen zu können welche Kommunikationsmuster genutzt werden, um auf allgemeine Taten induzieren zu können.

Menold und Kemper (2014) führen aus, dass sich Unterschiede in Verhaltensmerkmalen und Einstellungsmerkmalen der Antwortenden zeigen. Verhaltensmerkmale seien sehr leicht durch Fälscher reproduzierbar. Hingegen seien Einstellungsmerkmale, das heißt persönliche Urteile und individuelles Verhaltenswissen schwieriger zu fälschen. Ferner seien individuell psychologische Aspekte sowie subjektive und kollektive Sichtweisen besonders schwer zu fälschen (Menold und Kemper 2014).

Aus den Erläuterungen Menold und Kempers (2014) ergeben sich also zwei Dimensionen der Intelligenz, die Fälscher vereinen müssen. Einerseits müssen sie ein breites Abstraktionswissen über individuelle und Populationsparameter und -anteile besitzen, insbesondere auch in Subgruppen und andererseits müssen sie eine Vorhersagefähigkeit mitbringen, da sie die Konstruktion des Fragebogens abschätzen können müssen, welche Antwortmuster zu welchen Ergebnissen führen. Diese Überlegung wird dadurch gestützt, dass gefälschte Interviews nach Professionalität des Interviewers variieren. So produzieren langjährige Interviewer weniger Totalfälschungen und modifizieren eher, wohingegen unerfahrene Interviewer tendenziell schlechtere Teil- und mehr Totalfälschungen produzieren (Schnell 1991, S. 28).

2.4. Erkennung potenzieller Fabrikate

Im Folgenden wird dargestellt, was die statistische Erkennung potenzieller Fabrikate bedeutet und mit welchen Methoden sie erfolgen kann. Dabei wird sich auf bereits vorhandene Fabrikate in Datensätzen beschränkt, die im Nachhinein (ex-post) kausal erklärt und aufgedeckt werden sollen (Postdiktives Verfahren). Damit erfolgen die Analysen im Sinne von Pearl und Mackenzie (2018) kontrafaktisch und basieren auf der Argumentation, wie müsste ein Outcome ausgeprägt sein, wenn es keine Fälschung gegeben hätte und schlussfolgern daraus was bedeutet eine Fälschung für ein Outcome.

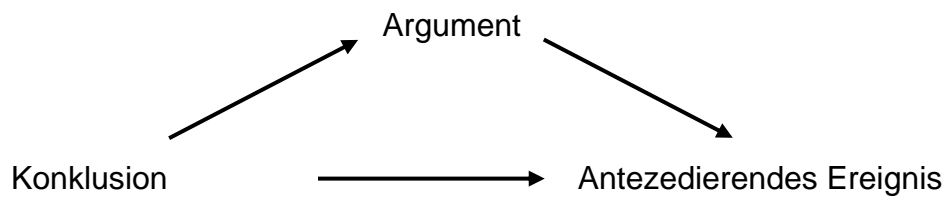


Abbildung 2 Postdiktiver Erklärungsweg

Für die Vermeidung von Interviewfälschungen (Präventive Verfahren) sei auf Gwartney (2013) und Schnell (2019) verwiesen.

Interpretiert man das Aufdecken von Fabrikaten aus kommunikationswissenschaftlicher Signalverarbeitung ist die Erkennung solcher Fabrikate der Zugriff auf Informationen ohne spezifische Zusammenarbeit mit dem Absender (die fälschende Person). Um potenzielle Fabrikate zu erkennen ist es nötig das Vorhandensein oder Fehlen von Kommunikationssignalen zu detektieren, um Verzerrungen durch Fabrikate zu extrahieren. Die Erkennung basiert auf der Unterscheidbarmachung von Rauschen (zufällige Antwortabweichungen) und der Filterung von offensichtlichen (systematische Antwortabweichungen) und verborgenen Signalen (spezielles abweichendes Antwortverhalten), um die Kommunikation des Absenders zu rekonstruieren. Die Erkennung versteckter Signale (Verzerrungen) in verdächtigem Material (Interviews), basiert häufig nur auf der Wahrscheinlichkeit, mit der ein möglicherweise verborgenes Signal enthalten ist, da durch statistische Verfahren nicht vollständig geklärt werden kann, ob es sich um eine Fälschung handelt oder nicht. Dies ist ein Nachteil der Aufdeckung. Deshalb immer anschließende Nachuntersuchungen der Interviews sowie Rekontaktierungen der Befragten (Schnell 2019, Kap 8.7.1) und die Erhebung weitere Paradata (Felder et al. 2014) erfolgen sollte, um diesen Nachteil zu minimieren. Dieses Problem besteht auch mit dem genutzten Verfahren der Variabilitätsprüfung, die lediglich potenzielle Fälscher aufgrund zu geringer geclusterter Totalvarianz filtert, die einer Nachbefragung zugeführt werden sollten.

Die Forschung zur Detektion von Fabrikaten nutzt das gesamte methodische Repertoire der deskriptiven und inferenzstatistischen Statistik sowie datengetriebene und algorithmische Ansätze (Bredl et al. 2012; Murphy et al. 2004; Kemper und Menold 2014; Haas und Winker 2016; Birnbaum 2012; Kuchen 2018; Schäfer et al. 2004b; Bredl et al. 2011; Bergmann et al. 2019). Datengetriebene und algorithmische Methoden haben das größte Potenzial für die zukünftige Evaluation von Surveydaten, um Fälschungen zu detektieren, bieten in vorliegenden Fall jedoch den Nachteil, dass sie vorhandene

Trainingsdaten benötigen, das heißt bereits vorhandene Fälschungen bekannt sein müssen, um auf potenzielle weitere Fälschungen zu schließen. Zudem ist ein hoher Aufwand in der Implementierung nötig, um adäquate Ergebnisse zu liefern.

Einfache deskriptive Verfahren, die sich auf Korrelationswerte und Varianzen fokussieren, können hingegen nur Tendenzen abbilden, ob einzelne Interviews oder mehrere Interviews einzelner oder mehrerer befragender Personen auffällig vom Gesamt abweichen. Diese Verfahren beruhen auf der Annahme der internen Konsistenz von Item-Skalen. Einzelne Items einer Skala sollten eine gewisse Korrelation aufweisen, die bei Fälschungen abweicht, auf Grundlage, dass keine Eindimensionalität besteht (siehe Kap. 3), kontrafaktisch müsste sich jedoch eine höhere Korrelation für Item-Batterien ergeben, da Fälscher ein konsistenteres Antwortverhalten zeigen.

Auffällig ist, dass gefälschte Interviews einen höheren Anteil an Middle Responses zeigen und einen geringeren Anteil an Extremwertaussagen (Bredl et al. 2012; Schäfer et al. 2004b). Zudem zeigen gefälschte Interviews deutlich häufiger geringere Anteile an „anderes“-Kategorien, da fälschende Personen den kürzesten Weg durch ein Interview nehmen (Menold et al. 2013, S. 30), basierend auf der rationalen Annahme, den Aufwand zu minimieren und ihren Nutzen zu maximieren. Daraus ergibt sich, dass die Streuung in gefälschten Daten deutlich geringer ausfällt als es in echten Daten der Fall wäre, das gilt nicht nur für Befragungsdaten, sondern auch für andere gefälschte experimentelle Daten (Al-Marzouki et al. 2005). Die fälschende Person geht dabei davon aus, dass insbesondere quasi-experimentelle Daten, wie Bevölkerungsumfragen in ihren Response-Mustern nicht so leicht reproduzierbar wären und damit eher als abweichende Einzelfälle betrachtet würden denn als vorsätzlich produzierte Daten. Um solche bereits vorhandenen Daten zu reproduzieren und auf eine Population hochzurechnen, macht sich das Re-Sampling von Daten nützlich, dass N Stichproben mit zurücklegen aus der bereits vorhandenen Stichprobe zieht. Durch die Simulation der Stichprobe ergeben sich N Stichproben, die eine Quasi-Grundgesamtheit abbilden, die erwartungstreue und konsistente Schätzungen für die echte Grundgesamtheit erlauben.

Da keine Daten, bei denen gefälschte Interviews bekannt sind vorliegen, so dass Trainingsdaten genutzt werden könnten fokussiert die Arbeit die Detektion von potenziell gefälschten Interviews. Dies wird anhand deskriptiver Werte evaluiert, die mittels Simulationsverfahren geschätzt und bewertet werden. Die evaluierten Interviews zeigen signifikante Abweichungen vom vorhandenen Datensatz und sollten basierend auf der

Analyse einer Nachuntersuchung zugeführt werden, um Fälschungen valide ausschließen zu können. Bestätigte Fälschungen könnten dann für weitere Untersuchungen genutzt werden.

2.5. Forschungsfrage

Aus den bestehenden Erkenntnissen in der Forschungsliteratur zu Interviewfälschungen lässt sich sagen, dass Fälscher ein Kommunikationsverhalten zeigen, dass vom üblichen Antwortverhalten abweicht. Die Erkennung solcher Fabrikate stellt damit den Zugriff auf Informationen ohne spezifische Zusammenarbeit mit dem Absender dar. Um zu erkennen welches Interview eine Fälschung ist, ist es nötig enthaltene Signale zu filtern, die solchen von echten Befragten unterscheidet. Somit kann angenommen werden, dass Fälscher Spuren in den Daten hinterlassen. *Je deutlicher die fälschende Person von echten Kommunikationssignalen abweicht, desto wahrscheinlicher ist es, dass sie entdeckt wird.*

Aus der Literatur ergibt sich, dass Fälschungen ein klares Muster in den Antwortkategorien zeigen sollten. Fälscher vermeiden Filterfragen, da sie dem Kalkül der Nutzenmaximierung widersprechen, die Kosten zu minimieren. *Nach Interviewer gruppierte Fälschungen weisen folglich eine kürzere Interviewlänge auf.*

Aus dem Nutzenkalkül ergibt sich, dass Interviewer möglichst wenige Kontaktversuche anstreben, da Re-Issuings mehr Zeitaufwand bedeuten und in der Regel nur realisierte Interviews entlohnt werden. *Die Wahrscheinlichkeit einer Fälschung steigt mit der Anzahl der Re-Issuings durch die gleiche Person.*

Um nicht aufzufallen, werden tendenziell Extremwerte vermieden und mittlere Kategorien bevorzugt. Interviewer zeichnet dabei eine unterschiedliche Wahrscheinlichkeit aus, dass sie fälschen. Sodass sich Fälschungen bei einzelnen Interviewern häufen. Folglich ergibt sich eine höhere Konsistenz. Das divergente Verhalten führt zu einem systematischen Fehler des Schätzers. *Gruppierte Fälschungen weisen damit durchschnittlich eine geringere Streuung der Antwortmuster innerhalb des Clusters auf, und eine höhere zwischen Clustern.*

Die zweigeteilte Forschungsfrage basierend auf den abgeleiteten Arbeitshypothesen lautet folglich: *Weisen einzelne Interviews signifikant geringere Streuungen auf?*

Sind somit potenzielle Fälschungen auf Grundlage ihrer abweichenden Streuung innerhalb von Interviewer-Clustern zu identifizieren?

3. Daten & Methodik

Im Folgenden werden die in dieser Arbeit verwendeten Daten beschrieben und die genutzte Methodik nachvollziehbar vorgestellt, um eine einfache Implementation für weitere Arbeiten zu gewährleisten. Der genutzte Algorithmus wird in Form von Code im Anhang aufgeführt, um eine volle Reproduzierbarkeit zu gewährleisten.

In der vorliegenden Arbeit wurde auf den European Social Survey 2002 der ersten Runde in Ungarn (ESS Round 1 Hungary) zurückgegriffen. Der European Social Survey (ESS Round 1 Hungary) enthält Daten von Befragten der ungarischen Wohnbevölkerung. Erhoben wurden die Daten im Wohnumfeld der Befragten mittels *computer-assisted personal interviews* (CAPI). Insgesamt sollten 2484 Zielpersonen erreicht werden. 1685 verwertbare Interviews wurden realisiert. Das entspricht einer Unit-Non-Responderate von 32,17 Prozent. Und einer Ausschöpfung von 67,83%.

Für den Erstkontakt waren 198 Personen eingeteilt, 2461 Interviews sollten durch diese Personen als Erstkontakt gestartet werden, 23 sind als NA vermerkt. 689 Befragte wurden nicht durch den Erstkontakt aufgesucht. 436 Befragte wurden einmal durch den Erstkontakt aufgesucht. 224 Befragte wurden zweimal aufgesucht. 124 Personen wurden dreimal durch den Erstkontakt angesteuert.

Kontaktversuche durch Erstkontakt	Anzahl der Befragten	Kontaktversuche durch Erstkontakt	Anzahl der Befragten
0	689	5	16
1	436	6	8
2	224	7	2
3	124	8	1
4	45	9	2

Tabelle 1 Anzahl der Kontaktversuche des Erstkontaktes (ohne 10 und 99=sys mis), Quelle ESS 2002 Round 1 Hungary

Nach dem vierten Kontaktversuch durch den Erstkontakt geht die Anzahl der Kontaktversuche gegen Eins, wobei zwei Personen einen neunten und einer einen zehnten Kontaktversuch gestartet haben. 19 Personen waren für ein Re-Issuing eingeteilt, das nicht von der Person für den Erstkontakt durchgeführt wurde. Wiederum 2 Personen waren als Drittinterviewer eingeteilt. 838 der 1685 realisierten Interviews wurden innerhalb der ersten vier Kontaktversuche durch einen der drei Interviewer realisiert, die meisten durch den Erstkontakt.

Die Angaben, wie viele Kontaktversuche insgesamt unternommen wurden, Unit-Non-Responses eingerechnet, fehlen in den meisten Fällen: für die Person des Erstkontaktes fehlen 937, für die Zweitinterviewer 2449 und für den Dritrinterviewer 2479 Angaben. Hier ist nicht klar, wie viele Kontaktversuche unternommen wurden, auch wenn kein Interview realisiert werden konnte.

Aus den theoretischen Ergebnissen folgernd wurde für die Analyse ein Subset gebildet, das Interviews bis zum neunten Kontaktversuch beinhaltet. Dabei wurden sowohl die Erstkontaktpersonen sowie die Ersatzinterviewer miteinbezogen. So ergeben sich 166 Interviewer mit unterschiedlicher Anzahl an Interviews. Insgesamt wurden 1095 Interviews miteinbezogen.

Anzahl der Interviewer pro Anzahl der Interviews

1	22	12	4
2	8	13	2
3	4	14	3
4	24	15	1
5	17	17	2
6	18	18	2
7	11	20	1
8	19	21	1
9	17	22	1
10	5	27	1
11	3		
Summe	1095 Interviews bei 166 Interviewern		

Tabelle 2 Anzahl der Interviewer pro Interviews basierend auf den erstellten Subset von Kontaktversuchen, Quelle: eigene Berechnungen

Der Datensatz umfasst diverse Themenbereiche, die über insgesamt 560 Variablen abgedeckt werden, wobei nicht alle Variablen im Datensatz abgefragt wurden. Da es sich um einen integrierten Datensatz handelt, sind auch Variablen enthalten, die nicht im Fragebogen für Befragte in Ungarn enthalten sind. Solche Variablen, die nicht für die Erhebung relevant sind, wurden vor der Analyse aus dem Datensatz entfernt bzw. für die Analyse als Nullwert umkodiert, da eine Berechnung mit *logischen NAs* selbst im Ergebnis ein *NA* zurückgibt. Die Analyse erfolgte mittels der Programmiersprache R und diverser Datenpakete zur Datenmanipulation vornehmlich aus dem *tidyverse*.

Als Grundlage der Analyse dienen die Interviewer-Nummern, wobei ausschließlich solche der ursprünglich ersten Interviews genutzt werden. Da im ESS ein Re-Issuing nicht durch dieselben Personen durchgeführt wird, werden diese in der vorliegenden Analyse nicht betrachtet, sollten bei weiteren Analysen aber separat überprüft, oder in die Analyse mit aufgenommen werden, da ein Restrisiko der Fälschung für Re-Issuings verbleibt.

3.1. Variabilität und Reliabilität

Laut Schnell (2008) stellt Messen die regelgeleitete Zuordnung von Zahlen und Objekten dar. Der Vorgang der Messung ist somit eine homomorphe Abbildung eines empirischen Relativs in ein numerisches (Schnell 2008). Im Rahmen der Überprüfung der Güte dieser Zuordnung ist das Ziel eine korrekte Messung zu gewährleisten. Um insbesondere die Zuverlässigkeit einer Messung eines Konstruktes zu gewährleisten, wird in der Literatur häufig Cronbachs Alpha herangezogen, um die Item-Reliabilität zu überprüfen.

“When items are used to form a scale, they need to have internal consistency. The items should all measure the same thing, so they should be correlated with one another. [...] The formula is:

$$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum s_i^2}{s_T^2} \right)$$

(3)

where k is the number of items, s_i^2 is the variance of the ith item and s_T^2 is the variance of the total score formed by summing all the items (Bland und Altman 1997).

Dieses Konzept der Eindimensionalität einer Skala bei der Indexkonstruktion zu gewährleisten einer zuverlässigen Messung wird meist auch im Falle der Identifikation von Falsifikationen genutzt, da theoretisch anzunehmen ist, dass Interviewer nicht immer im Sinne des Instrumentes antworten und somit Mehrdimensionalität im Instrument verursachen, somit Korrelationen, für einzelne Interviewer niedriger sind.

Durch systematische Konstruktion von verschiedenen Itemtypen ist es möglich Interviewfälschungen effizienter zu vermeiden und aufzuspüren. Die Variabilitätsmethode macht sich ebendieses Phänomen der Mehrdimensionalität auf einfache Weise zu nutze. Durch abweichendes Antwortverhalten ergeben sich Unterschiede in der Streuung der Ausprägungen einzelner Instrumente.

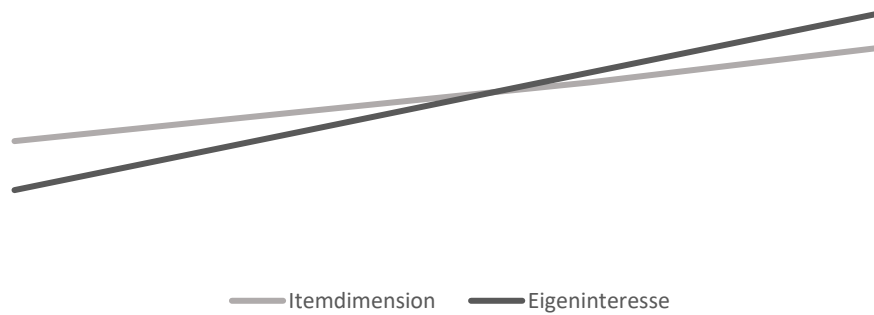


Abbildung 3 Theoretische Item-Dimensionalität

So weisen Fälschungen eine geringere Variabilität in den Ausprägungen einzelner Items auf, da nicht eine wahre Antwort gegeben wird, sondern eine willkürliche Imputation der fälschenden Person erfolgt. Diese Imputation entspricht damit nicht der Itemdimension, sondern dem Eigeninteresse der fälschenden Person, wodurch es zu einer Abweichung zur erwarteten Streuung des Items für eine einzelne Person bzw. ein Cluster von Interviews kommt. Kontrafaktisch wollen Personen jedoch sich innerhalb einer Lüge nicht widersprechen, weshalb fälschende Person eine höhere interne Konsistenz innerhalb einer Item-Batterie aufweisen müssten. Aus dem Nutzenkalkül ergibt sich ebenso ein konsistentes Fälschungsverhalten über mehrere Fragebögen.

Es ist daher anzunehmen, dass für einzelne Befragte die Summe der Abweichungsquadrate

$$SQ_i = \sum_{i=1}^{n_i} (x_i - \bar{x}_i)^2$$

(4)

des Antwortverhaltens charakteristisch ist. Wobei SQ_i die „wahre“ Streuung der i-ten ursprünglich befragten Person ist, n_i die Anzahl der Items der i-ten ursprünglich befragten Person, x_i das i-te Item ist und \bar{x}_i der Mittelwert aller Items, die die ursprünglich befragte Person beantwortet hätte.

Wenn nun Interviewer systematisch Fälschungen produzieren, führt das dazu, dass diese eine systematisch abweichende Variabilität aufweisen, die entweder höher oder niedriger ist als diese von durchschnittlichen Befragten. Was zu einem Bias der Variabilität

$$BIAS_{li}(SQ) = \left(\sum_{j=1}^{m_i} (x_{ij} - \bar{x}_i)^2 \right) - \left(\sum_{l=1}^{n_i} (x_{il} - \bar{x}_l)^2 \right)$$

(5)

innerhalb des Clusters von j Interviews des i-ten Interviewers I führt. Somit wird für die Umsetzung der Methode zuerst die jeweilige Variabilität innerhalb des li-ten Interviewer-Fragebogenclusters berechnet, wobei m_i die Anzahl aller Interviews j ist, die die befragende Person geführt hat. In einem zweiten Schritt wird die Statistik mit der erwarteten Varianz für einen Fragebogen-Cluster der gegebenen Größe auf der gesamten Umfrage verglichen (Schäfer et al. 2004).

„Somit ergeben sich l_i , mit $i = 1, \dots, n$, Interviewern l_i und n , die Anzahl der Interviewer, die die Umfrage durchgeführt haben. Die Anzahl der Fragebögen Q_j wird mit $j = 1, \dots, m$ angegeben. Wobei $m = m_1 + \dots + m_i$, mit m_i die Anzahl der vom Interviewer l_i gelieferten Fragebögen bezeichnet. Ohne Rücksicht auf die Bedeutung der Antworten – ob 5 für '5 Jahre' oder für 'Ich bin anderer Meinung' kodiert – wird die Summe der Abweichungsquadrate für jede Frage $Q_j(k)$ mit $k = 1, \dots, l$ für alle Fragebögen Q_j eines Interviewers l_i berechnet und über alle Fragen summiert:

$$T_{li} = \sum_{k=1}^l \sum_{j=1}^{m_i} (Q_j(k) - \overline{Q(k)})^2$$

(6)

Hier bezeichnet $\overline{Q(k)}$ den Mittelwert für Frage $Q(k)$ und der Index j erfasst alle Fragebögen Q_j mit $j = m_{i1}, \dots, m_i$ des Interviewers l_i . Die Verteilung der Teststatistik T_{li} wird mit einem Resampling-Ansatz für die gesamte Umfrage geschätzt. Aus dieser Verteilung kann eine Wahrscheinlichkeit für die beobachteten Wert abgeleitet werden. Im Folgenden werden diese Wahrscheinlichkeiten mit Plausibilität bezeichnet. Durch Sortierung der Interviewer in Bezug auf die erreichte Plausibilität erhält man ein Interviewer Ranking. Die Interviewer mit der geringsten Plausibilität stehen an der Spitze der Rangfolge. Sie werden als potenzielle Fälscher betrachtet. Der Wert von T_i , der dem Interviewer l_i zugeordnet ist, wird mit der entsprechenden Verteilung von T , die mit einem Resampling-Ansatz geschätzt wird, verglichen. Die Fläche unter der Dichtekurve auf der linken Seite der Realisierung T_i definiert die Plausibilität. Wenn die Plausibilität zu gering ist, wird der Interviewer als potenzieller Fälscher angesehen. Das

Verfahren entspricht einem einseitigen statistischen Test“ (Schäfer et al. 2004a, S. 11–12, eigene Übersetzung).

4. Ergebnisse

Im Folgenden Abschnitt werden die Ergebnisse der Variabilitätsanalyse der Interviewer-Cluster vorgestellt.

Abbildung 4 zeigt das Ergebnis des Resamplings. Insgesamt wurden 166 Interviewer mit 1095 Interviews in die Analyse miteingeschlossen. Die meisten Interviewer zeigen eine totale Variabilität zwischen XXX und XXXX. XX Interviewer mit auffällig niedriger Variabilität ergeben sich aus dem linksseitigem Testverfahren, dies entspricht XXX% der Interviewer, die als potenzielle Fälscher aufgrund des Testes eingestuft werden. XX Interviews von 1095 insgesamt entsprechen damit XXX% als potenzielle Interviewfälschungen.

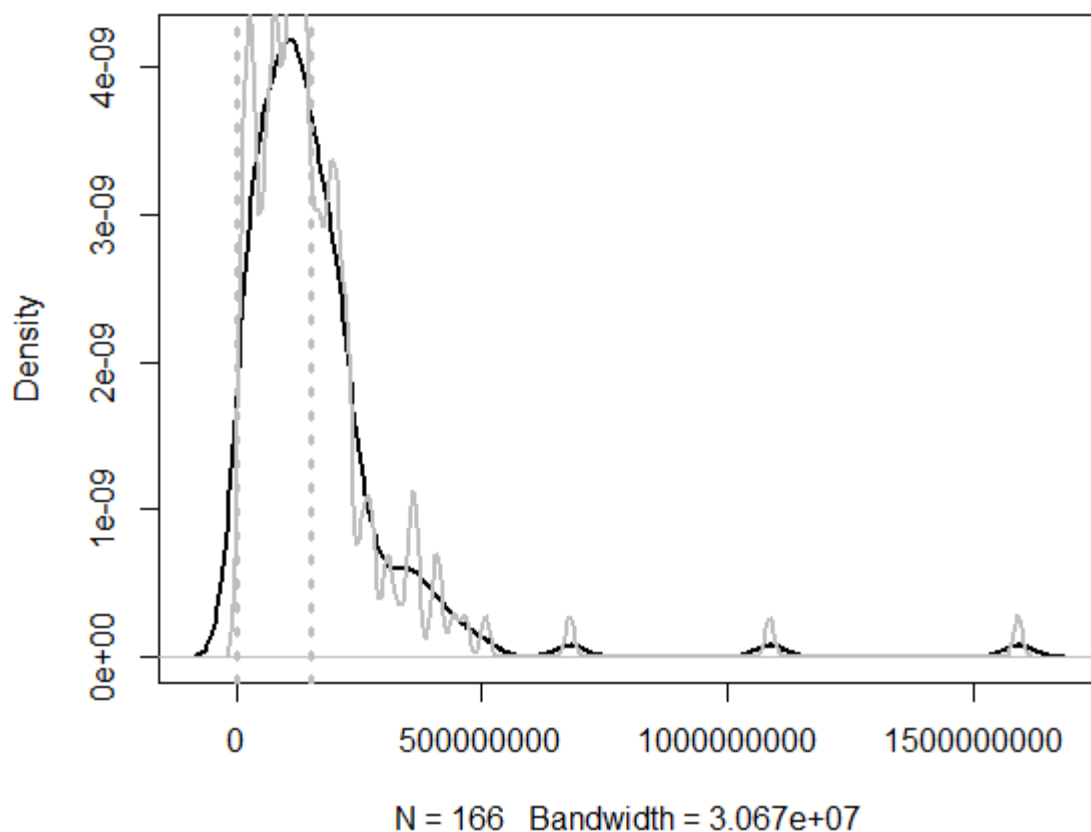


Abbildung 4 Empirische (schwarz) und simulierte (grau) Dichtefunktion der totalen Variabilität, Quelle: eigene Berechnungen

Die Tabelle 3 liefert das Ranking für die ersten 29 Interviews. Es zeigt sich, dass insbesondere Interviewer mit nur einem Interview sehr niedrige Plausibilitätswerte aufweisen, was nachvollziehbar erscheint. Die summierte Totalstreuung eines Interviewers mit nur einem Interview ist geringer als, die eines Interviewers mit mehreren Interviews. Dies muss bei der Evaluation berücksichtigt werden. Weshalb für die Entscheidung ebenfalls der Mittelwert der Totalstreuung eines Clusters nach Anzahl der

ID	ROWSUM	PLAUSIBILITY	CUM PLAUSIBILITY	TOTINT
2086	5913534	0.000220	0.000220	1
26015	8541190	0.000318	0.000538	1
24097	9511784	0.000354	0.000893	1
19021	10524707	0.000392	0.001285	1
3131	10870728	0.000405	0.001690	1
39057	11236628	0.000419	0.002108	1
25106	13220952	0.000492	0.002601	1
3023	13230701	0.000493	0.003094	1
8234	15396418	0.000573	0.003667	1
3156	17335184	0.000646	0.004313	1
25004	23675976	0.000882	0.005195	1
9022	25150264	0.000937	0.006131	3
15047	25237306	0.000940	0.007072	1
8118	25816866	0.000962	0.008033	1
10152	25962941	0.000967	0.009000	1
9061	27201328	0.001013	0.010013	2
9118	29452818	0.001097	0.011111	1
34048	30482082	0.001135	0.012246	1
2102	30529847	0.001137	0.013383	1
3038	31387870	0.001169	0.014552	1
23038	31594649	0.001177	0.015729	2
8787	32827805	0.001223	0.016952	1
4134	32957652	0.001228	0.018180	1
30027	36727607	0.001368	0.019548	2
1054	37991876	0.001415	0.020963	3
9009	42751960	0.001592	0.022555	4
3135	43281082	0.001612	0.024168	1
25053	45629888	0.001700	0.025867	4
15055	48071596	0.001791	0.027658	5
12046	51416390	0.001915	0.029573	4
...

Tabelle 3 Globales Interviewer-Ranking erste 29 Personen nach Schätzer T_i und Plausibilität für potenzielle Fälschung, Quelle: eigene Berechnungen

Interviews und dessen Standardabweichung berechnet wurde, um Interviewer innerhalb ihres Clusters zu bewerten. Tabelle 4 zeigt die Mittelwerte der totalen Variabilität gruppiert nach Anzahl der Interviews und die untere Schwelle von minus einer Standardabweichung. Interviews von Interviewern die unter diesem Wert liegen, werden als potenzielle Fälschungen eingestuft.

Anzahl der Interviews pro Interviewer	Mittlere totale Varianz (MTV) (Std.Abw.)	Untere Schwelle (MTV minus 1 Std.Abw.)
1	21700019,1 (13496430)	8203589,06
2 KORREKTE WERTE EINFÜGEN	42221381,8 (24060654,3)	18160727,6
3	62141494,1 (26824466,6)	35317027,5
4	73574743 (23557844,9)	50016898,1
5	114019537 (40894429,1)	73125107,9
6	156412898,6 (45103393,82)	111309504,8
7	176281794,5 (39490123,96)	136791670,5
8	- (-)	-

Tabelle 4 Mittlere totale Varianz innerhalb eines Clusters mit Standardabweichung, Quelle: eigene Berechnungen

5. Zusammenfassung

Die Arbeit hat ausgehend von Forschungsergebnissen zur Interviewfälschung die Variabilitätsmethode zur Aufdeckung potenziell gefälschter Interviews praktisch anhand des Datensatzes des European Social Surveys 2002 in Ungarn vorgestellt.

Es erfolgte eine Literaturrecherche zum Thema Interviewfälschung. Mittels dieser wurde Literatur zu den unterschiedlichen Aspekten der Forschung im Feld der Survey-Methodik gesichtet und diskutiert. Es zeigte sich, dass ein breites Spektrum an Forschungssichtweisen auf Fälschungen besteht, wobei psychologische Aspekte der Fälschung und der Detektion den größten Teil ausmachen.

Ergebnisse, Ergebnisse, Ergebnisse.

Ausgehend von den methodischen Sichtweisen der Detektion wurde in der vorliegenden Arbeit die Variabilitätsmethode zur Aufdeckung potenziell gefälschter Interviews mit der statistischen Programmiersprache R angewandt und mittels Resampling einem statistischen Test unterzogen. Basierend auf den Ergebnissen ergaben sich **XXX** potenziell gefälschte Interviews. Diese konnten durch eine Abweichung der quadratischen Abweichungssumme innerhalb des Interviewer-Fragebogenclusters ermittelt werden. Die Plausibilität für eine Fälschung wurde mittels eines Resampling-Ansatzes ermittelt und stellt die linksseitige Fläche unter der Verteilung dar. Die identifizierten Interviewenden sollten in einer Nachuntersuchung einer weiteren Prüfung unterstellt werden, um eine Fabrikation ausschließen oder bestätigen zu können. Da nicht eindeutig geklärt werden kann, ob es sich um Fälschungen handelt, kann keine Hypothese darüber bestätigt werden.

Die Methodik weist daher offensichtlich die Schwäche auf, dass nicht eindeutig geklärt werden kann, ob es sich bei den identifizierten Interviewenden tatsächlich um Fälscher handelt. Diese Schwäche haben allerdings alle ex-post Methodiken inne, die in der Literatur zur Aufdeckung von potenziellen Fabrikaten beschrieben werden. Eine Identifikation von potenziellen Fälschern ist mit vollständiger Sicherheit nur möglich, wenn der oder die Interviewende von selbst eine Fälschung aufdeckt bzw. in Folge der Nachuntersuchung aufgedeckt werden. Eine weitaus bedeutendere Schwäche der Methodik ist, die geringe Reliabilität. Die Methodik beruht auf der Annahme, dass Fälschungen eine höhere Konsistenz des Antwortverhaltens abbilden. Dies ist nicht trennscharf auf eine Fälschung zurückzuführen. Zudem muss eine hinreichend große Anzahl an Interviews pro Interviewer vorliegen, um überhaupt zu sagen, dass die Variabilität geringer ist. Professionelle Fälscher weisen ein höheres und breiteres Erfahrungswissen über Parameter und den Fragebogen auf, sodass sie von dieser Methode unentdeckt blieben.

Ein großer Vorteil der vorgestellten Methodik ist die einfache Implementierung und Umsetzung. Aufwendigere Verfahren liefern im Vergleich tendenziell nicht immer reliablere Ergebnisse bei der Aufdeckung von potenziellen Fälschern und bedeuten einen wesentlich höheren Ressourcenaufwand und/oder sind schwer implementierbar.

Eine weitere wichtige Maßnahme im Erhebungsmanagement ist daher die erfolgreiche Verhinderung der Erstellung von Fabrikaten durch geeignete Maßnahmen, wie Supervision, angemessenen Lohn und Vorbildungsmaßnahmen (Schnell 2019, 8.5; 8.6; 8.7.2).

Abschließend kann festgehalten werden, dass ebenso wichtig wie die Aufdeckung von Fabrikaten, die Verhinderung dieser zu erachten ist. Da eine Aufdeckung von Betrug schwer methodisch zu erreichen ist und keine letztendliche Garantie besteht, dass einzelne Fälschungen insbesondere Teilfälschungen unentdeckt bleiben. Weitere Forschung in diesem Bereich ist daher notwendig, um bestehende Methodiken zu Prävention und Detektion von Interviewfälschungen zu verbessern, um robuste Schätzungen in Bevölkerungsumfragen vornehmen zu können und Erkenntnis daraus abzuleiten.

6. Literatur

- Al-Marzouki, Sanaa; Evans, Stephen; Marshall, Tom; Roberts, Ian (2005): Are these data real? Statistical methods for the detection of data fabrication in clinical trials. In: *BMJ* 331 (7511), S. 267–270. DOI: 10.1136/bmj.331.7511.267.
- Bergmann, Michael; Schuller, Karin; Malter, Frederic (2019): Preventing interview falsifications during fieldwork in the Survey of Health, Ageing and Retirement in Europe (SHARE). In: *Longitudinal and Life Course Studies* 10 (4), S. 513–530. Online verfügbar unter <https://doi.org/10.1332/175795919X15694136530293>.
- Birnbaum, Benjamin (2012): Algorithmic Approaches to Detecting Interviewer Fabrication in Surveys.
- Bland, J. Martin; Altman, Douglas G. (1997): Statistics Notes: Cronbach's Alpha. In: *BMJ* 314 (7080), S. 572. Online verfügbar unter <http://www.jstor.org/stable/25173851>.
- Blasius, Jörg (2014): Fälschungen von Interviews. In: Nina Baur und Jörg Blasius (Hg.): *Handbuch Methoden der empirischen Sozialforschung*. Wiesbaden: Springer Fachmedien Wiesbaden, S. 323–329.
- Blasius, Jörg; Friedrichs, Jürgen (2012): Faked Interviews. In: Samuel Salzborn, Eldad Davidov und Jost Reinecke (Hg.): *Methods, Theories, and Empirical Applications in the Social Sciences*. Festschrift für Peter Schmidt. Wiesbaden: Springer VS, S. 49–56.
- Blasius, Jörg; Thiessen, Victor (2012): *Assessing the Quality of Survey Data*. London: Sage.
- Blasius, Jörg; Thiessen, Victor (2013): Detecting Poorly Conducted Interviews. In: Peter Winkler, Natalja Menold und Rolf Porst (Hg.): *Interviewers' Deviations in Surveys. Impact, Reasons, Detection and Prevention*. Frankfurt: Peter Lang, S. 67–88.
- Blasius, Jörg; Thiessen, Victor (2015): Should we Trust Survey Data? Response Simplification and Data Fabrication. In: *Social Science Research*.

- Bredl, Sebastian; Storfinger, Nina; Menold, Natalja (2011): A literature review of methods to detect fabricated survey data. Discussion Papers. Justus Liebig University Giessen, Center for international Development and Environmental Research (ZEU) (56). Online verfügbar unter <https://EconPapers.repec.org/RePEc:zbw:zeudps:56>.
- Bredl, Sebastian; Winker, Peter; Koetschau, Kerstin (2012): A statistical approach to detect interviewer falsification of survey data. In: *Survey Methodology* 38, S. 1–10.
- Buehl, Anne-Kathrin; Melchers, Klaus G. (2017): Individual Difference Variables and the Occurrence and Effectiveness of Faking Behavior in Interviews. In: *Frontiers in psychology* 8, S. 1–15. DOI: 10.3389/fpsyg.2017.00686.
- Bushery, John M.; Reichert, Jennifer W.; Albright, Keith A.; Rossiter, John C. (1999): Using Date and Time Stamps to Detect Interviewer Falsification. In: *Proceedings of the Survey Research Methods Section, ASA* (9), S. 316–320. Online verfügbar unter http://www.asasrms.org/Proceedings/papers/1999_053.pdf, zuletzt geprüft am 27.01.2021.
- Crespi, Leo P. (1945): The Cheater Problem in Polling. In: *Opinion Quarterly* (9), S. 431–445.
- European Social Survey ERIC (ESS ERIC) (2003): European Social Survey (ESS), Round 1 - 2002.
- Felder, Barabara; Birg, Alexandra; Kreuter, Frauke (2014): Paradata. In: Nina Baur und Jörg Blasius (Hg.): *Handbuch Methoden der empirischen Sozialforschung*. Wiesbaden: Springer Fachmedien Wiesbaden, S. 357–366.
- Gwartney, Patricia (2013): Mischief Versus Mistakes: Motivating Interviewers to not Deviate. In: Peter Winker, Natalja Menold und Rolf Porst (Hg.): *Interviewers' Deviations in Surveys. Impact, Reasons, Detection and Prevention*. Frankfurt am Main: Peter Lang, S. 195–215.
- Haas, Samuel de; Winker, Peter (2016): Detecting Fraudulent Interviewers by Improved Clustering Methods – The Case of Falsifications of Answers to Parts of a Questionnaire. In: *Journal of Official Statistics* 32 (3), S. 643–660. DOI: 10.1515/jos-2016-0033.

- Harrison, David E.; Krauss, Stefanie L. (2002): Interviewer Cheating: Implications for Research on Entrepreneurship in Africa. In: *Journal of Developmental Entrepreneurship* (7), S. 319–330.
- Hood, C. C.; Bushery, John M. (1997): Getting More Bang from the Reinterview Buck: Identifying 'At Risk' Interviewers". In: *Proceedings from Section on Survey Research Methods, ASA*, S. 820–824, zuletzt geprüft am 27.01.2021.
- Kemper, Christoph J.; Menold, Natalja (2014): Nuisance or Remedy? The Utility of Stylistic Responding as an Indicator of Data Fabrication in Surveys. In: *Methodology* 10 (3), S. 92–99. DOI: 10.1027/1614-2241/a000078.
- Kuchen, Robert (2018): Detecting Fabricated Interview Data in a Two-Level Approach Using the Mahalanobis Distance. Johannes Gutenberg-Universität Mainz, Universitätsmedizin.
- Kuriakose, Noble; Robbins, Michael (2016): Don't get duped: Fraud through duplication in public opinion surveys. In: *SJI* 32 (3), S. 283–291. DOI: 10.3233/SJI-160978.
- Menold, Natalja; Winker, Peter; Storfinger, Nina; Kemper, Christoph J. (2013): A method for ex-post identification of falsifications in survey data. In: Peter Winkler, Natalja Menold und Rolf Porst (Hg.): Interviewers' deviations in surveys. Frankfurt am Main: Peter Lang, S. 25–48. Online verfügbar unter <https://search.gesis.org/publication/gesis-bib-141178>.
- Menold, Natalja.; Kemper, Christoph. J. (2014): How Do Real and Falsified Data Differ? Psychology of Survey Response as a Source of Falsification Indicators in Face-to-Face Surveys. In: *International Journal of Public Opinion Research* 26 (1), S. 41–65. DOI: 10.1093/ijpor/edt017.
- Murphy, Joe; Baxter, Rodney; Eyerman, Joe; Cunningham, David; Kennet; Joel (2004): A System for Detecting Interviewer Falsification. [American Association for Public Opinion Research 59th Annual Conference]. Hg. v. American Association for Public Opinion Research Section on Survey Research Methods. Phoenix, Arizona. Online verfügbar unter <https://www.amstat.org/Sections/Srms/Proceedings/y2004/Files/Jsm2004-000517.pdf>, zuletzt geprüft am 27.01.2021.

- Nelson, James E.; Kiecker, Pamela L. (1996): Marketing Research Interviewers and Their Perceived Necessity of Moral Compromise. In: *Journal of Business Ethics* (15), S. 1107–1117.
- Pearl, Judea; Mackenzie, Dana (2018): The Book of Why. The New Science of Cause and Effect. New York, NY: Basic Books.
- Schäfer, Christin; Schräpler, Jörg-Peter; Müller, Klaus-Robert (2004a): Identification, Characteristics and Impact of Faked and Fraudulent Interviews in Surveys. Deutsches Institut für Wirtschaftsforschung (DIW). Berlin. Online verfügbar unter https://www.diw.de/documents/dokumentenarchiv/17/diw_01.c.41963.de/paper2004_schaeferetal.pdf, zuletzt geprüft am 27.01.2021.
- Schäfer, Christin; Schräpler, Jörg-Peter; Müller, Klaus-Robert; Wagner, Gert G. (2004b): Automatic Identification of Faked and Fraudulent Interviews in Surveys by Two Different Methods. Deutsches Institut für Wirtschaftsforschung (DIW). Berlin (DIW Discussion Papers, 441). Online verfügbar unter <http://hdl.handle.net/10419/18293>, zuletzt geprüft am 02.05.2020.
- Schnell, Rainer (1991): Der Einfluss gefälschter Interviews auf Survey-Ergebnisse. In: *Zeitschrift für Soziologie* (20), S. 25–35.
- Schnell, Rainer (2013): Nonresponse in Bevölkerungsumfragen: Ausmaß, Entwicklung und Ursachen: Springer-Verlag.
- Schnell, Rainer (2019): Survey-Interviews. Methoden standardisierter Befragungen, 2. Auflage. Wiesbaden: Springer Fachmedien Wiesbaden, zuletzt geprüft am 27.01.2021.
- Schräpler, Jörg-Peter; Wagner, Gert G. (2005): Characteristics and impact of faked interviews in surveys. An analysis of genuine fakes in the raw data of SOEP. In: *Allgemeines Statistisches Archiv* 89, S. 7–20, zuletzt geprüft am 02.05.2020.
- Weinauer, Marlene (2019): Be a detective for a day: How to detect falsified interviews with statistics. In: *SJI* 35, S. 569–575. DOI: 10.3233/SJI-190524.

7. Eigenständigkeitserklärung

Ich, Kevin Glock, An der Spinnerei 11-17, 96047 Bamberg,

Matrikelnr.: 2020296 versichere an Eides Statt durch meine Unterschrift, dass ich die vorstehende Arbeit selbständig und ohne fremde Hilfe angefertigt und alle Stellen, die ich wörtlich oder dem Sinne nach aus Veröffentlichungen entnommen habe, als solche kenntlich gemacht habe, mich auch keiner anderen als der angegebenen Literatur oder sonstiger Hilfsmittel bedient habe.

Ich versichere an Eides Statt, dass ich die vorgenannten Angaben nach bestem Wissen und Gewissen gemacht habe und dass die Angaben der Wahrheit entsprechen und ich nichts verschwiegen habe.

Die Strafbarkeit einer falschen eidesstattlichen Versicherung ist mir bekannt, namentlich die Strafandrohung gemäß § 156 StGB bis zu drei Jahren Freiheitsstrafe oder Geldstrafe bei vorsätzlicher Begehung der Tat bzw. gemäß § 163 Abs.1 StGB bis zu einem Jahr Freiheitsstrafe oder Geldstrafe bei fahrlässiger Begehung.

Bamberg, Xx. Februar 2020

Unterschrift