Ludwig-Maximilians-Universität
Institut für Statistik

Bachelorarbeit

Reichtumsforschung in Deutschland - Herausforderung der Datengewinnung auf Basis messtheoretischer Grundlagen und Anwendung geeigneter statistischer Verfahren zur Analyse etablierter Datenquellen

Denise Gawron

Betreuung: Professor Dr. Thomas Augustin

29.09.2014

Abstract

Der Versuch, die abstrakte Größe Reichtum zu messen, stellt hohe Anforderungen an Methoden der Datenerhebung und -auswertung. In dieser Arbeit werden zwei der wichtigsten Datenquellen Deutschlands, das Sozio-oekonomische Panel und die Einkommens- und Verbrauchsstichprobe, vorgestellt und unter Berücksichtigung messtheoretischer Grundlagen eingehend betrachtet. Dabei zeigt sich, dass das Sozio-oekonomische Panel durch eine gezielte Überrepräsentation ausreichend Beobachtungen von Hocheinkommenshaushalten liefert, aufgrund seiner Fragebogenkonzeption für Einkommen und Vermögen jedoch möglicherweise Annahmen der Klassischen Testtheorie verletzt werden. Die Einkommens- und Verbrauchsstichprobe verfügt dagegen über sehr exakte Angaben zum Haushaltseinkommen, kann aufgrund der geringen Beobachtungszahl jedoch keine Aussagen über vermögende Haushalte treffen, weshalb diese zensiert werden. Um diesem Problem entgegenzuwirken, werden Methoden für den Umgang mit zensierten Daten betrachtet. Das Tobit-Modell kann wegen seiner restriktiven Normalverteilungsannahme nicht verwendet werden, eine Umwandlung in das verteilungsfreie Rang-Tobit-Modell ist jedoch möglich. Da das Einkommen eine verhältnisskalierte, positive Variable ist, werden zusätzlich Methoden der Lebensdaueranalyse auf ihre Eignung überprüft. Hier zeigt sich, dass Transformationsmodelle geeignet scheinen, zensierte Einkommensdaten zu schätzen.

Inhaltsverzeichnis

1 Einleitung

1.1 Reichtumsforschung in Deutschland

Analysen zu Einkommen und Vermögen sind ein in Politik und Wissenschaft häufig diskutiertes Thema, doch im Allgemeinen liegt der Fokus dieser Diskussionen auf der linken Seite jener Verteilungen: der Armut. Das andere Ende, Haushalte und Personen mit hohen Einkommens- und Vermögenswerten, schienen jenseits von Fragen zur Steuerpolitik kein übermäßiges Interesse zu generieren, da ihre finanzielle Lage keinen Grund für Handlungsbedarf liefert.
Im Zuge von immer weiter angeheizten Debatten zur sozialen Ungleichheit, der (gefühlt) wachsenden Schere zwischen Armut und Reichtum, ändert sich diese einseitige Betrachtung jedoch allmählich. Der vierte Armuts- und Reichtumsbericht der Bundesregierung geht explizit Fragen zur geeigneten Messung von Reichtum nach und das Institut für angewandte Wirtschaftsforschung e. V. hat im Auftrag des Bundesministeriums für Arbeit und Soziales einen Forschungsbericht über "Möglichkeiten und Grenzen der Reichtumsberichterstattung" vorgelegt.
Die Ausweitung des Blicks auf diesen Aspekt der Einkommens- und Vermögensverteilung ist nicht nur politisch wichtig, sondern auch auf Forschungsebene interessant, da sich hier viele Frage- und Problemstellungen ergeben, die in der Armutsforschung nur teilweise beobachtbar sind. Datenquellen, die für Analysen der unteren Einkommensgrenzen (jenseits der sogenannten "bitteren Armut", die noch einmal einen anderen Problembereich darstellt) umfassende und detaillierte Beobachtungen liefern, zeigen bei der Erhebung vermögender Haushalte plötzlich deutliche Schwächen.
Die Schwierigkeiten und Fehlerquellen einer korrekten Erfassung von Reichtum sind vielfältig. Sie beginnen bei der korrekten Operationalisierung, die sich mit der Trennung von Einkommen und Vermögen, dem Einbezug nicht-finanzieller Aspekte und dem Finden einer geeigneten, relativen Grenze auseinandersetzen muss, gehen über die Stichprobenziehung, die mit einer kleinen Grundgesamtheit wirklich reicher Haushalte und dem Problem des systematischen Non-Response zu kämpfen hat, bis hin zum korrekten Aufbau von Messinstrumenten, um Messfehler zu vermeiden.
Diese Herausforderungen schlagen sich nicht nur in der Interpretation der gewonnenen Daten nieder, sondern sind bereits bei deren Auswertung zu beachten. Der Einsatz von Hochrechnungsfaktoren und Imputationen ist sinnvoll und notwendig, doch auch jenseits dieser relativ üblichen Verfahren finden sich Methoden, die geeignet sind, möglichst viele Informationen aus den gegebenen Daten zu ziehen.
Damit ist die Reichtumsforschung vielfältigen, multidisziplinären Problemen unterworfen, für die

teilweise bereits Lösungen existieren, Ansätze aus anderen Forschungsbereichen übernommen werden können, oder gänzlich neue Wege beschritten werden müssen.

1.2 Eingrenzung des Themas und Zielsetzung der Arbeit

Diese Arbeit soll einen Teilbereich der Problemstellungen in der Reichtumsmessung präsentieren, unter statistischen Gesichtspunkten bewerten und Methoden vorstellen, die geeignet sind, Aussagen über den rechten Bereich der Einkommens- und Vermögensverteilungen zu treffen.
Zunächst muss nicht nur der thematische, sondern auch räumliche Schwerpunkt eingegrenzt werden. Eine internationale Betrachtung der Reichtumsforschung ist sinnvoll und in vielen Fällen auch erwünscht, führt jedoch auch zu einer Vergrößerung des Problembereichs, weshalb sich diese Arbeit ausschließlich mit Aspekten der Reichtumsforschung in Deutschland beschäftigen wird.
Dies beginnt mit einer Diskussion der wichtigsten Datenquellen, die häufig für die Armuts- und Reichtumsmessung herangezogen werden. Diese sollen nicht nur oberflächlich vorgestellt, sondern, auf Basis messtheoretischer Grundlagen und der sich daraus ergebenden Anforderungen an die Messung, ausführlich betrachtet werden. Das Hauptaugenmerk liegt dabei auf der Gewinnung von Einkommens- und Vermögensangaben, sowie dem Umgang mit Haushalten auf der rechten Seite der Einkommensverteilung. Hierzu werden zunächst die drei verbreitetsten Messtheorien beschrieben, verglichen und anschließend die für sie relevanten Gütekriterien abgeleitet.
Als wichtigste Datenquellen wurden das Sozio-oekonomische Panel und die Einkommens- und Verbrauchsstichprobe ausgewählt. Diese werden seit Jahren konstant weiterentwickeln, verfügen über eine große Beobachtungszahl, ausführliche Metainformationen und werden regelmäßig für zahlreiche Studien, zur Errechnung wirtschaftlicher Kenngrößen und den Armuts- und Reichtumsbericht der Bundesregierung herangezogen.
Weiterhin verfügen sie über verschiedene Erhebungs- und Stichprobendesigns, wenden unterschiedliche Verfahren zur Erhebung von Einkommensdaten an und verfolgen ein konträres Konzept im Umgang mit Hocheinkommenshaushalten. Das macht ihre Betrachtung, die zunächst mit einem einfachen Überblick über die Datenstruktur beginnt und in die Anwendung der beschrieben Messtheorien mündet, besonders interessant.
Anschließend sollen am Beispiel der Einkommens- und Verbrauchsstichprobe statistische Methoden vorgestellt werden, die die Analyse von zensierten Einkommensdaten ermöglichen. Dies geschieht insbesondere durch die Tobit-Regression und den Versuch, Lebenszeitmodelle auf Einkommensdaten auszuweiten.

2 Messtheoretische Grundlagen

Der nachfolgende Abschnitt stellt eine kurze Einführung in die verbreitetsten Messtheorien und den damit verbundenen Konzepten der Gütekriterien, sowie Aspekte der Klassischen Testtheorie dar. Dies bildet die Grundlage, um die später vorgestellten, häufig zur Einkommensanalyse herangezogenen Datenquellen ausführlich nicht nur auf deren praktische, sondern auch theoretische Konzeption zu diskutieren.
Die drei bekanntesten Messtheorien sind die Klassische Messtheorie, die Operationale Messtheorie und die Repräsentationale Messtheorie. Obwohl jede der drei Theorien mit dem Ziel entwickelt wurde, Regeln zu finden, die die Zuordnung gemessener Werte zu einem Merkmal ermöglichen, unterscheiden sie sich in ihren grundsätzlichen Anforderungen an den Messvorgang.
Jeder Ansatz setzt hierbei verschiedene, teilweise widersprüchliche Bedingungen. Welche Theorie zugrunde gelegt werden kann und sollte, ist abhängig von dem interessierenden Merkmal, der Eigenschaft der Messwerte, der praktischen Durchführung der Messung und der weiteren Verwendung der gewonnenen Werte.

2.1 Messtheorien

2.1.1 Klassische Messtheorie

Die Klassische Messtheorie ist die älteste der drei vorgestellten Theorien. Ihr Name geht auf die Arbeiten von Euklid und Aristoteles zurück, in denen sich bereits Spuren der Anwendung dieser Theorie finden lassen (vgl. Michell 1986, S. 404).
Ihre Anforderungen und daraus folgend ihre möglichen Anwendungsgebiete, unterscheiden sich deutlich von den beiden anderen, insbesondere in den Sozialwissenschaften gebräuchlicheren Theorien.
Die Anwendung der Klassischen Messtheorie setzt voraus, dass das zu messende Attribut des interessierenden Objekts quantitativ ist, sich also sowohl ordnen als auch addieren lässt. Wichtig ist hierbei, dass dies eben genau für das Attribut, jedoch nicht zwingend für das eigentliche Objekt gilt. Diese Voraussetzung ist als Hypothese zu betrachten, die vor der Messung überprüft und gegebenenfalls falsifiziert werden muss.
Dies stellt die eigentliche Herausforderung einer Messung unter den Bedingungen der Klassischen Messtheorie dar. Nicht die Zuordnung von Zahlen zu Objekten, die die empirische Wirklichkeit wiedergeben sollen, wird in den Vordergrund gerückt, sondern vielmehr die Entdeckung bereits

bestehender numerischer Größen (ebd., S. 405).
Aufgrund dieser Bedingungen unterliegt die Messung nach der Klassischen Messtheorie Einschränkungen, die ihre Anwendung in sozialwissenschaftlichen Bereichen erschweren. Beispielsweise ließe sich diskutieren, ob die Merkmale "Einkommen" und "Vermögen" dem Objekt "Mensch" zuordenbare, quantitative Attribute und damit auf Grundlage der Klassischen Messtheorie erfassbar sind.
Wie wir später, insbesondere beim Sozio-oekonomischen Panel, sehen werden, ist die reine Betrachtung von Einkommens- und/oder Vermögensverteilungen für aktuelle Forschungsfragen kaum relevant. Attribute wie "Lebenszufriedenheit", die häufig unter anderem mit Bezug auf das Haushaltseinkommen erklärt werden sollen, können nach den Grundsätzen der Klassischen Messtheorie jedoch nicht erfasst werden. Weiterhin müsste sich das Attribut "Reichtum" auf rein finanzielle Aspekte beziehen und andere, ergänzende Betrachtungsweisen ausklammern.
Da das Ausschließen von für eine Messtheorie ungeeigneten Größen nicht Ziel einer korrekten Messung sein sollte, scheint die Klassische Messtheorie für viele Fragestellungen der Reichtumsforschung ungeeignet. Allerdings darf hierbei nicht vergessen werden, dass eben genau die Aufdeckung schon bestehender numerischer Größen einer der zentralen Punkte dieser Theorie ist. Möglicherweise kann die Annahme, dass diese existieren, jedoch noch nicht gefunden wurden, zukünftige Messungen um neue Blickwinkel bereichern.

2.1.2 Operationale Messtheorie

Die Operationale Messtheorie verfolgt einen gänzlich anderen Ansatz als die Klassische Messtheorie und ist in ihrer Definition wesentlich freier als die später vorgestellte Repräsentationale Messtheorie. Weder sollen durch eine Messung, die ihren Grundsätzen folgt, bereits bestehende numerische Größen aufgedeckt, noch reale empirische Ordnungen durch die Zuordnung von Werten korrekt wiedergegeben werden. Vielmehr sieht diese Theorie vor, dass das interessierende Merkmal eines Objekts erst durch dessen Messung definiert wird (vgl. Hand 1996, S. 453).
Dieser Ansatz bietet sowohl Vor- als auch Nachteile. Durch diese Definition, die jede Operation, die exakt spezifiziert wurde und eine Zahl generiert, als gültige Messung betrachtet, werden Messungen ermöglicht, die im Rahmen der anderen Messtheorien nicht durchführbar wären.
Das oben angesprochene Problem, dass eine über finanzielle Aspekte hinaus führende Definition von Reichtum mit der Klassischen Messtheorie (noch) nicht erfasst werden kann, wird durch den alternativen Ansatz der Operationalen Messtheorie umgangen. Sofern diese durch konkrete Operationen entstanden sind, gelten alle den interessierenden Merkmalen zuordenbare Zahlen als gültige Messungen. Dies gilt sogar unabhängig von der real existierenden Größe "Reichtum".
Eine Messtheorie, die die reale empirische Ordnung ausklammert, mag im ersten Moment sinnlos

erscheinen, kann sich jedoch auch in Bereichen der Armuts- und Reichtumsmessung als nützlich erweisen.
Andreas Klocke hat in seiner im Jahr 2000, in der Zeitschrift für Soziologie erschienenen Studie "Methoden der Armutsmessung. Einkommens-, Unterversorgungs-, Deprivations- und Sozialhilfekonzept im Vergleich" verschiedene Konzepte der Armutsmessung miteinander verglichen. Diese folgten alle der Definition der Europäischen Kommission, die Armut als Zustand beschreibt, in dem „[...]Menschen über nur so geringe materielle, kulturelle und soziale Mittel verfügen, dass sie von der Lebensweise ausgeschlossen sind, die in einer Gesellschaft als unterste Grenze des Akzeptablen annehmbar ist" (Europäische Kommission 1995, in: Klocke 2000, S. 313).
Diese Definition charakterisiert zwar den theoretischen Zustand der Armut, gibt jedoch keine Auskunft darüber, wie dieses Merkmal konkret gemessen werden kann.
Klockes Studie betrachtet insgesamt vier Armutskonzepte, die geeignet erscheinen, den definierten Zustand abzubilden und messbar zu machen. Im direkten Vergleich werden durch diese Konzepte jedoch nicht nur Bevölkerungsanteile, die zwischen 8.8 % und 14.9 %, schwanken, als arm ausgewiesen, sondern es zeigt sich ebenfalls, dass es innerhalb der Gruppen mitunter nur wenige Überschneidungen gibt. Legt man alle vier Konzepte an, gelten nur 2.9 % der Bevölkerung als arm.
Geht man nun davon aus, dass die Messung eine real existierende Größe abbilden soll, weckt ein solches Ergebnis Zweifel an der Korrektheit dieser Messung, da drei, möglicherweise alle vier Konzepte, manche Merkmalsträger irrtümlich als arm ausweisen oder eben gar nicht erst erfassen, obwohl tatsächlich Armut vorliegt. Die Operationalisierung von Reichtum unterscheidet sich zwar von der der Armut, ist jedoch ebenfalls komplex, weshalb davon ausgegangen werden sollte, dass hier ähnliche Probleme auftreten können.
Die Operationale Messtheorie bietet aufgrund ihrer Definition die Möglichkeit, trotz Klockes Ergebnissen, weiterhin an einem oder sogar mehreren der Armutskonzepte festzuhalten, da die korrekte Abbildung der realen Armut keine notwendige Bedingung für die Messung darstellt. Dies kann insbesondere auf politischer Ebene von Bedeutung sein und erlaubt beispielsweise die Zahlung von Sozialhilfe an Personen, die durch die Messung als empfangsberechtigt definiert wurden, unabhängig davon, ob diese tatsächlich arm sind.
Trotz dieser Vorteile, wurde die Operationale Messtheorie inzwischen größtenteils von der Repräsentationalen Messtheorie abgelöst. Dies liegt unter anderem daran, dass die Ausprägungen der Merkmale, die erst durch die Messung definiert wurden, nur innerhalb genau dieser Messung miteinander verglichen werden können, da sie auf keine real existierende Größe zurückzuführen sind. Unterscheiden sich die Spezifikationen zweier Operationen voneinander, muss davon ausgegangen werden, dass diese nicht dasselbe Merkmal definieren.
Dies erschwert den Vergleich der Ergebnisse verschiedener Studien, sofern sie nicht dieselben Da-

tenquellen verwenden. Auch Veränderungen an den Messinstrumenten führen zu einer fehlenden Vergleichbarkeit; ein Problem, mit dem sich insbesondere Längsschnitt- und Panelstudien konfrontiert sehen. Legt man die Operationale Messtheorie zugrunde, führt eine nachträgliche Anpassung der Messinstrumente (z. B. eines Fragebogens) dazu, dass die Ergebnisse der Befragungen vor dieser Veränderung nicht mehr mit den danach folgenden Ergebnissen verglichen werden können. Veränderungen über den Zeitverlauf sind auf diese Weise nur schwer darstellbar.
Weiterhin verhindert dieser Ansatz zwar die Ungültigkeit der Ergebnisse aufgrund einer mangelhaften Operationalisierung, kann jedoch das Auftreten anderer Messfehler nicht verhindern. So können befragte Individuen beispielsweise fehlerhafte Antworten geben, oder systematische Antwortverweigerungen auftreten. Auch die beiden Gütekriterien Objektivität und Reliabilität, auf die später noch genauer eingegangen wird, müssen erfüllt werden.

2.1.3 Repräsentationale Messtheorie

Die Repräsentationale Messtheorie ist inzwischen die in der Literatur und Praxis vorherrschende Messtheorie. Sie geht davon aus, dass durch die korrekt spezifizierte Zuordnung von Zahlen eine reale, empirische Ordnung ausgedrückt wird (vgl. Hand 1996, S. 449).
Ihr Grundgedanke ist damit mit dem der Klassischen Messtheorie vergleichbar, die ebenfalls versucht, bereits bestehende Größen zu finden. Allerdings fällt bei der Repräsentationalen Messtheorie die Beschränkung auf ausschließlich quantitativ erfassbare Attribute weg, weshalb - ebenso wie in der Operationalen Messtheorie - auch Messwerte, die sich auf einer Nominal- oder Ordinalskala bewegen, erfasst werden können.
Messungen, die nach den Grundsätzen der Repräsentationalen Messtheorie durchgeführt werden, sollen zudem strukturtreu sein. Hierfür wird zunächst eine Beziehung R der zu messenden Objekte o zueinander definiert, die als empirisches Relativ α bezeichnet wird. Das numerische Relativ β beschreibt wiederum eine Menge von Zahlen, über die eine Relation S definiert wurde.
Ziel ist es, eine Zuordnungsregel zu finden, mit der das numerische Relativ der Ordnung des empirischen Relativs entspricht. Diese strukturtreuen Abbildungen werden auch als „Morphismen“ bezeichnet, wobei man zwischen umkehrbar eindeutigen, isomorphen, und nicht umkehrbar eindeutigen, homomorphen, Abbildungen unterscheidet.
Ist eine Abbildung φ isomorph, so gilt

$$o_1 R_i o_2 \Leftrightarrow \varphi(o_1) S_i \varphi(o_2)$$

Das bedeutet, von der dem Objekt zugeordneten Zahl kann direkt auf das eigentliche Objekt rück-

geschlossen werden. Ist das nicht möglich, da dieselbe Zahl mehreren Objekten zugeordnet wird, spricht man von einem Homomorphismus.
Dann gilt (vgl. Schnell et al. 2011, S. 131f)

$$o_1 R_i o_2 \Rightarrow \varphi(o_1) S_i \varphi(o_2)$$

Isomorphismen sind letztlich bijektive Homomorphismen und führen zu einer Umbenennung der Elemente, ohne deren Struktur zu verändern (vgl. Karpfinger 2013, S. 13).
Ein einfaches Beispiel, das diesen Gedanken verdeutlicht, ist das der Längenmessung. Legt man mehrere Bretter nebeneinander, so dass ihre unteren Kanten auf einer Linie sind, lässt sich vergleichen, ob dies ebenfalls für die oberen Kanten gilt.
Damit sind die Bretter das zu messende Objekt o und ihr Größenunterschied die interessierende Beziehung R. Nun soll eine Abbildung φ gefunden werden, die diese Beziehung wiedergibt. Ordnet man die Bretter nach ihrer Größe, so dass die obere Kante des rechten Bretts mindestens so hoch ist wie die des linken, kann man das empirische Relativ α in ein numerisches Relativ β umsetzen, indem man den Brettern Zahlen zuordnet, die die reale Beziehung der Bretter zueinander darstellen.
Durch das geordnete Aneinanderreihen der Bretter, haben wir eine Relation zwischen ihnen aufgezeigt, die sich mit $o_i \leq o_j$ beschreiben lässt. Ordnen wir nun jedem der Bretter, beginnend mit dem ganz links, aufsteigende Zahlen zu, haben wir eine Zuordnungsregel gefunden, die das empirische Relativ wiedergibt, denn wir wissen, dass eine größere Zahl auch ein größeres Brett bedeutet.
Wird jede der Zahlen nur einmal vergeben, da die Länge der Bretter exakt (bspw. in Nanometern) erfasst wird, handelt es sich zudem um einen Isomorphismus, da sie einen direkten Rückschluss auf das entsprechende Brett zulassen. Erfolgt hingegen eine gröbere Zahlenzuordnung, die Längenunterschiede, die eine bestimmte Größe unterschreiten, durch Rundungen ignoriert, liegt lediglich ein Homomorphismus vor.
Die Bedingung, dass der Messwert die reale, zugrundeliegende Größe abbilden muss, lässt sich auch durch die Klassische Testtheorie ausdrücken:

$$X = T + \epsilon$$

Der wahre Wert T kann nicht direkt beobachtet werden und soll durch das Messinstrument X dargestellt werden. Da dieses Messinstrument T jedoch nur ungenau erfasst, wird mit einem Fehler ϵ gerechnet. Dieser Fehler unterliegt folgenden Annahmen:
(1) $\mu(\epsilon) = 0$
(2) $\rho_{T\epsilon} = 0$

(3) $\rho_{\epsilon_1 \epsilon_2} = 0$
(4) $\rho_{\epsilon_1 T_1} = 0$
(1) Der Erwartungswert des Fehlers ist gleich Null, (2) es gibt keine Korrelation zwischen dem wahren Wert und dem Messfehler, (3) die Messfehler zweier wiederholter Messungen sind unkorreliert und (4) der Messfehler korreliert nicht mit dem wahren Wert einer anderen Messung (vgl. Diekmann 2008, S. 262f).

Während die Gleichung $X = T + \epsilon$ trivial ist, da eine große Differenz zwischen dem wahren und dem gemessenen Wert durch einen entsprechend großen Fehler ϵ erklärt wird, sind die Annahmen für diesen Fehler in der Praxis nicht immer gegeben.
Die eben angesprochene Erklärung einer großen Abweichung zwischen X und T, die durch den Messfehler ausgeglichen wird, widerspricht Annahme (1), nach der der Erwartungswert des Messfehlers gleich Null sein soll.
In Bezug auf die Reichtumsforschung ist es denkbar, dass exakte Angaben über hohe Einkommens- und Vermögenswerte - auch von den Befragten selbst - schwerer einzuschätzen sind, als bei geringeren Summen, so dass der Messfehler mit steigendem Vermögen respektive Einkommen ebenfalls zunimmt. Dies wäre eine Verletzung der Annahme (2). Weiterhin könnte beispielsweise der Bildungsabschluss Einfluss auf den Überblick haben, den der Befragte über sein eigenes Vermögen hat. Kommt es deshalb zu Messfehlern, wäre Annahme (4) verletzt.
Aufgrund der Bedingung, dass die Messung die empirische Wirklichkeit abbilden soll, ist die Anwendung der Repräsentationalen Messtheorie komplexer als die der Operationalen Messtheorie und bedarf einer noch gründlicheren Prüfung. Dafür sorgt diese Bedingung gleichzeitig für eine bessere Vergleichbarkeit der Ergebnisse, auch, wenn die Daten nicht durch dasselbe Messinstrument generiert worden sind, da sich diese dennoch auf dasselbe empirische Relativ beziehen. Weiterhin ermöglicht sie - im Gegensatz zu der Klassischen Messtheorie - Messungen von nicht rein quantitativ erfassbaren Attributen.
Ein weiterer Vorteil dieser Theorie liegt darin, dass sie intuitiv einfacher zu erfassen ist, insbesondere als die Operationale Messtheorie. Die Vorstellung, dass die durch diese Messung generierten Zahlen der direkte Ausdruck einer real existierenden Größe sind, erleichtert die Interpretation der gewonnenen Werte.
Dennoch kann es auch hier zu Verzerrungen aufgrund von fehlerhaften Messwerten oder Antwortverweigerungen kommen. Möchte man Messungen, die auf Basis der Repräsentationalen Messtheorie entstanden sind überprüfen, müssen nicht nur die Gütekriterien der Objektivität und Reliabilität erfüllt sein, zusätzlich muss das Messinstrument auch valide sein.

2.2 Gütekriterien

2.2.1 Objektivität

Die Objektivität ist eines von insgesamt drei Gütekriterien, mit denen sich die Qualität einer Messung überprüfen lassen. Dieses Kriterium besagt, dass Messergebnisse nicht von externen Einflüssen abhängig sein dürfen (vgl. Wolf 2010, S. 240).
Wiederholte Messungen sollen also immer dieselben Ergebnisse liefern, unabhängig davon, wer sie durchführt, auswertetet oder interpretiert. Damit ist dieses Kriterium für alle drei Messtheorien relevant und eine Grundvoraussetzung um das Kriterium der Reliabilität zu erfüllen.

2.2.2 Reliabilität

Die Reliabilität führt das Kriterium der Objektivität einen Schritt weiter. Sie fordert, dass Messergebnisse nicht nur unabhängig von den Umständen der Messung sein sollen, sondern allgemein reproduzierbar sind. Das heißt, wiederholtes Anlegen desselben Messinstruments muss zu denselben Ergebnissen führen (vgl. Häder 2010, S. 109).
Die Reliabilität ist ein Maß für den Anteil der Varianz des wahren Werts T an der Varianz des gemessenen Werts X. Formal lässt sich dies folgendermaßen ausdrücken:

$$r = \frac{Var(T)}{Var(X)}$$

Dieser Reliabilitätskoeffizient bewegt sich zwischen 0 und 1, wobei ein Wert von 1 auf ein perfekt reliables Messinstrument schließen lässt, da auch bei wiederholten Messungen dasselbe Ergebnis erzielt wird (vgl. Moosbrugger 2007, S. 110).
Ein Problem dieses Koeffizienten, neben der fehlenden Beobachtbarkeit des wahren Werts, ist, dass für dessen korrekte Bestimmung die Annahmen der Testtheorie erfüllt sein müssen. Der Reliabilitätskoeffizient drückt die Testgenauigkeit der gesamten Population aus und trifft daher keine Aussage über einzelne Beobachtungen. Nimmt man nun an, dass der Fehler ϵ mit dem wahren Wert einer Messung korreliert ist, womit Annahme (2) beziehungsweise (4) der Testtheorie verletzt wären, ist das Messinstrument trotz eines möglicherweise hohen Koeffizienten r für Teilbereiche der beobachteten Population nicht geeignet (ebd., S. 138). Da dies insbesondere niedrige respektive hohe Werte des Messbereichs betrifft, muss diese Einschränkung bei der Reichtumsforschung unbedingt bedacht werden.
Ein weiteres Problem mit der Bestimmung des Reliabilitätskoeffizienten würde entstehen, wenn

das Messinstrument niedrige Werte über- und hohe Werte unterschätzt, also auch hier der Messfehler mit dem wahren Wert korreliert wäre. Dieser Fall würde dazu führen, dass die beobachtete Streuung geringer ist als die wahre Streuung, womit der Koeffizient nicht mehr zwischen 0 und 1 liegen, sondern theoretisch beliebig groß werden könnte.
In der Praxis lässt sich der wahre Wert nicht bestimmen, weshalb verschiedene Verfahren entwickelt wurden, die zur Überprüfung der Reliabilität herangezogen werden können.
In der Reichtumsforschung bietet sich die Test-Retest-Methode an. Hier wird den Untersuchungsobjekten nach Ablauf einer bestimmten Frist derselbe Test ein weiteres Mal vorgelegt. Ist das Messinstrument reliabel, müssen die beiden Ergebnisse eine hohe Korrelation aufweisen. Allerdings ist bei der Wahl dieser Methode und der Festlegung der Frist zu beachten, dass zwischenzeitlich weder eine Veränderung des interessierenden Attributs stattgefunden haben darf, noch Gedächtniseffekte das Ergebnis verfälschen (vgl. Häder 2010, S. 110)
Wird die Frist zu kurzfristig gewählt, ist es möglich, dass sich die Befragten an ihre zuletzt gegebene Antwort erinnern und diese einfach wiederholen. Eine zu lange Frist kann dazu führen, dass sich die Einkommens- und Vermögensverhältnisse zwischenzeitlich verändert haben, beispielsweise durch eine berufliche Veränderung, Gehaltsanpassungen oder, wenn nach dem letzten Monatseinkommen gefragt wird, Sonderzahlungen, wie Weihnachts- oder Urlaubsgeld. Dies kann verhindert werden, indem der Retest diese Veränderungen ebenfalls abfragt. Beim Einsatz des Test-Retest-Designs ist insbesondere Annahme (3) der Klassischen Testtheorie, die Unkorreliertheit zweier wiederholter Messungen, zu beachten.
Methoden, wie der Parallel- oder Split-Half-Test sind bei Einkommens- und Vermögensfragen weniger geeignet, obwohl man das Vorgehen in der Einkommens- und Verbrauchsstichprobe, die einen direkten Vergleich zwischen Einnahmen und Ausgaben eines Haushalts ermöglicht, als Form des Paralleltests verstehen kann.
Interessant ist auch die Frage, ob das Kriterium der Reliabilität in der Operationalen Messtheorie erfüllt sein muss. Der wahre Wert T der Klassischen Testtheorie wird üblicherweise im Zusammenhang mit der Repräsentationalen Messtheorie angenommen, die ein empirisches Relativ ausdrücken möchte. Da diese Einschränkung bei der Operationalen Messtheorie nicht gegeben ist, jedoch davon auszugehen ist, dass auch Messungen die dieser Theorie folgen, an der Reproduzierbarkeit ihrer Messwerte interessiert sind, muss T in diesem Zusammenhang nicht als empirische Wirklichkeit, sondern als durch die Messung definiertes, erwünschtes Ergebnis betrachtet werden. Auch in diesem Fall muss überprüft werden, ob das gewählte Messinstrument zuverlässig, also reliabel ist.
Da die Annahmen der Klassischen Testtheorie für eine sinnvolle Prüfung der Reliabilität erfüllt sein müssen, dies in der Praxis jedoch nicht immer gewährleistet werden kann, werden gelegentlich andere Theorien zugrunde gelegt, die die Reliabilität eines Messinstruments nicht mehr über

die Gesamtpopulation bestimmen, sondern für jede einzelne Beobachtung, in Abhängigkeit von der Schwierigkeit des Messinstruments, gemessen an der Merkmalsausprägung der Untersuchungseinheit (Moosbrugger 2007, S. 138f). Ein bekanntes Beispiel für ein Verfahren mit diesem Aufbau ist das sogenannte Rasch-Modell.

2.2.3 Validität

Das Gütekriterium der Validität ist für Messungen, die auf der Repräsentationalen Messtheorie basieren entscheidend, da es Auskunft darüber gibt, ob ein Messinstrument das empirische Relativ korrekt abbildet. Während Messungen, denen die Operationale Messtheorie zugrunde gelegt wurde, lediglich objektiv und reliabel sein müssen, ist eine nicht valide Messung unter den Gesichtspunkten der Repräsentationalen Messtheorie wertlos, weil damit die Grundforderung dieser Theorie verletzt ist.
Ähnlich wie bei der Reliabilität ist die Prüfung, ob diese Forderung erfüllt wurde, in der Praxis sehr schwierig. Im Groben wird zwischen drei Typen der Validität unterschieden, die einzeln oder ergänzend überprüft werden können (vgl. Häder 2010, S. 113).
Die *Inhaltsvalidität* beschreibt die inhaltliche Analyse des Messverfahrens (vgl. Wolf 2010, S. 250). Ist ein Test inhaltsvalide, wird das zu messende Attribut vollständig durch die Messung abgebildet. Um dies bestimmen zu können, muss der gesamte Aufbau des Messverfahrens überprüft werden; ein Verfahren, das in den Sozialwissenschaften nur selten angewandt werden kann und deshalb durch Expertenratings - die eigentlich nur einen Teil der Prüfung ausmachen würden - ersetzt wird.
Im Gegensatz zur Inhaltsvalidität ist die *Kriteriumsvalidität* einfacher zu erfassen und objektiv besser prüfbar, jedoch nicht immer anwendbar. Sie setzt die Existenz eines Vergleichkriteriums voraus, das ebenso wie das zu prüfende Messinstrument die interessierende Größe abbildet.
Die Schwierigkeit dieser Form der Validitätsprüfung liegt darin, dass ein solches Kriterium nicht nur existieren, sondern bereits erfolgreich auf Validität überprüft worden sein muss, da ein Vergleich andernfalls nicht sinnvoll ist. Weiterhin muss überprüft werden, ob bei Existenz eines solchen Kriteriums die Einführung eines neuen Messinstruments überhaupt notwendig ist.
Sind diese Bedingungen erfüllt, kann die Kriteriumsvalidität eines Messinstruments wie folgt geschätzt werden (ebd., S. 252)

$$r_{TiTj} = \frac{r_{XiXj}}{\sqrt{r_i}\sqrt{r_j}}$$

r_{TiTj} steht für die Korrelation der wahren Werte des zu überprüfenden Messinstruments i und des

Vergleichkriteriums j, r_{XiXj} für die Korrelation der gemessenen Werte i und j und r_i respektive r_j sind die Reliabilitätskoeffizienten der Testwerte i und j.
Dabei handelt es sich um eine korrigierte Formel, die die Reliabilität r der beiden Messinstrumente i und j miteinbezieht. Würde man diese ignorieren, würde die errechnete Validität der wahren Werte zu gering ausfallen, da die Messungenauigkeiten des Instruments, die durch die Reliabilität ausgedrückt werden, nicht beachtet würden.
Der *Konstruktvalidität* liegt die Annahme zugrunde, dass ein gemeinsamer Ursprung, beispielsweise die politische Einstellung des Probanden, die Beantwortung eines Variablensets beeinflusst. Mithilfe eines Fragebogens, der die verschiedenen Konstrukte erfasst, kann anschließend empirisch geprüft werden, ob die Antworten die theoretischen Überlegungen bestätigen.
Eine Methode, um die zuvor getroffenen Annahmen über Dimensionen eines Variablensets zu prüfen, ist die Faktorenanalyse. Mit deren Hilfe kann beispielsweise kontrolliert werden, ob ein Messinstrument, das die sogenannten Big Five der Persönlichkeit messen soll, tatsächlich eine fünfdimensionale Struktur aufweist (ebd., S. 253).
Einen anderen Ansatz stellten Campbell und Fiske 1959 vor; die Multitrait-Multimethod-Matrix (MMM-Matrix). Diese ermöglicht es, mindestens zwei Konstrukte zu vergleichen, sofern für beide mehr als ein Erhebungsverfahren existiert (vgl. Häder, S. 115).
Sie berechnet die diskriminante Validität, das heißt, die Korrelation zwischen zwei verschiedenen Konstrukten, die durch dieselbe Erhebungsmethode (z. B. einen Fragebogen) gemessen wurden und die konvergente Validität, das heißt, die Korrelation zwischen verschiedenen Erhebungsmethoden, die dasselbe Konstrukt messen. Idealerweise ist dabei die konvergente Validität größer als die diskriminante Validität. Dies würde bedeuten, dass die Messergebnisse methodenunabhängig sind und sich die verschiedenen Konstrukte nachweisbar voneinander unterscheiden.
Die Validität einer Messung ist eine zentrale Bedingung der Repräsentationalen Messtheorie und gleichzeitig nur schwer nachzuprüfen. Bezogen auf die Reichtumsmessung, stellen sich gleich mehrere Schwierigkeiten.
Zum einen ist noch ungeklärt, was “Reichtum” überhaupt bedeutet, zum anderen muss überprüft werden, ob und wie zuverlässig sich die verschiedenen Aspekte des Reichtums erfassen lassen. Die folgend vorgestellten Datensätze, liefern aufgrund ihres Erhebungsdesigns und der Fragebogengestaltung unterschiedliche Ansätze, um Einkommens- und Vermögenswerte zu messen, auf deren Basis Fragen zur Reichtumsforschung analysiert werden können.

3 Datengewinnung

Die Frage nach dem Einkommen wird häufig zu den sogenannten soziodemographischen Strukturdaten gezählt und daher in einer Vielzahl von Erhebungen abgefragt. Dies geschieht jedoch meist in Kategorien und gilt als eine der Fragen, mit den meisten Antwortverweigerungen.
Für eine tiefgreifende Einkommensanalyse ist das nicht ausreichend; dennoch stehen in Deutschland mehrere Datenquellen zur Verfügung, die einen detaillierten Blick auf diese Größe zulassen. Zu nennen sind unter anderem die Allgemeine Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS), die ein Vorhaben des GESIS Leibnitz-Institut für Sozialwissenschaften ist, oder auch die SAVE-Panelstudie des Max-Planck-Instituts für Sozialrecht und Sozialpolitik, die sich speziell mit dem Sparverhalten deutscher Haushalte auseinandersetzt.
Als besonders wichtig gelten jedoch das Sozio-oekonomische Panel und die Einkommens- und Verbrauchsstichprobe. Diese erheben Fragen zum Einkommen und Vermögen deutscher Haushalte detailliert und in regelmäßigen Abständen, verfügen über einen großen Stichprobenumfang und werden als Grundlage für viele einkommensbasierte Fragestellungen verwendet.
Dennoch stehen auch sie vor Herausforderungen: Die Zahl wirklich reicher Haushalte ist so gering, dass sie auch im Fall einer perfekten Zufallsstichprobe nicht oder nur in sehr geringer Zahl ausgewählt werden würden, es kann zu systematischen Antwortverweigerungen kommen, oder die Angaben zu Einkommen und Vermögen werden versehentlich oder mutwillig verfälscht.
Weiterhin müssten für eine umfassende Beurteilung der Größe “Reichtum” Daten zu nicht-finanziellen Aspekten vorliegen und das Einkommen eines Haushalts in Zusammenhang mit dessen Ausgaben, die durch regionale Preisunterschiede beeinflusst sein können, gesetzt werden.

3.1 Das Sozio-oekonomische Panel (SOEP)

Das Sozio-oekonomische Panel ist eine der wichtigsten Datenquellen Deutschlands. Es bildet nicht nur die Grundlage für multidisziplinäre Forschungsfragen, sondern wird auch für internationale Vergleiche und wichtige, politische Publikationen wie den Armuts- und Reichtumsbericht herangezogen (vgl. Wagner et al. 2008, S. 303).
Dabei handelt es sich um eine Panelstudie, die seit 1984 im jährlichen Rhythmus erhoben wird. Im Startjahr wurde eine Zufallsstichprobe mit rund 4.500 westdeutschen Haushalten gezogen, sowie eine spezielle Stichprobe mit 1.300 Haushalten, deren Haushaltsvorstand aus einem der damals typischen Gastarbeiterländer (Türkei, Italien, Spanien, Griechenland und das ehemalige Jugoslawien) stammte.
Diese Haushalte werden seit ihrer ersten Ziehung wiederholt im jährlichen Rhythmus befragt. Eine

der Besonderheiten des SOEP ist hierbei, dass es nicht nur einen Fragebogen gibt, der von dem Haushaltsvorstand ausgefüllt wird, sondern zusätzliche Personenfragebögen, die von jeder in dem ausgewählten Haushalt lebenden Person ab siebzehn Jahren beantwortet werden (ebd., S. 308). Auch für jüngere Haushaltsmitglieder existieren Fragebögen, die im Allgemeinen von einem ihrer Elternteile beantwortet werden, bis sie das Mindestalter für eine eigenständige Befragung erreicht haben.

Dieses Befragungsdesign erlaubt es Forschern, das Leben von an dem Panel teilnehmenden Personen nicht nur konstant weiterzuverfolgen und Veränderungen zu beobachten, sondern weiterhin wird durch die Einzelbefragung ermöglicht, über die zeitliche Eingrenzung "Von der Wiege bis zur Bahre" hinaus zu gehen. So können werdende Mütter bereits während ihrer Schwangerschaft befragt werden, nicht erst nach der Geburt des Kindes.

Für die Reichtumsforschung relevanter ist jedoch der Teil "bis zur Bahre". Durch das Studiendesign, das eine Befragung aller im Haushalt lebenden Personen vorsieht, können durch andere Haushaltsmitglieder nachträglich noch Fragen zu finanziellen Aspekten, wie der Hinterbliebenenrente und Erbschaften beantwortet werden, die eine bessere Einschätzung der Vermögensverhältnisse erlauben.

In der Praxis wird diese Möglichkeit allerdings nicht immer vollständig umgesetzt. Der Fragebogen 2013 (Welle 30) des SOEP, der sich an die Hinterbliebenen des Verstorbenen richtet, enthält beispielsweise keinerlei Fragen zu finanziellen Verhältnissen. Dies mag damit zusammenhängen, dass finanzbezogene Fragen, sowie die Befragung von Hinterbliebenen noch immer als sensible Themen gelten und zu Antwortverweigerungen oder schlimmstenfalls dem vollständigen Ausfall aus der Stichprobe führen könnten. Fragen zu einmaligen Zahlungen, die auch Erbschaften mit einschließen und regelmäßigen Einnahmen, wie etwa eine Hinterbliebenenrente, werden jedoch mittels der allgemeinen Haushalts- und Personenfragebögen erhoben.

Um eine über die Jahre stattfindende Ausdünnung der Stichprobe zu verhindern, wurde unter anderem ein Schneeballsystem eingeführt. Dies bedeutet, dass jede Person, die in einen der in der Stichprobe enthaltenen Haushalte zieht, ebenfalls aufgenommen wird. Gleichzeitig werden ehemalige Haushaltsmitglieder auch nach ihrem Umzug in einen anderen Haushalt weiterverfolgt (ebd., S. 308).

Um diese Änderungen der Wohnverhältnisse auch im Datensatz sichtbar zu machen, werden für jede befragte Person mehrere Identifikationsnummern vergeben. Diese setzen sich aus der Nummer des ursprünglichen Haushalts, der Nummer des aktuellen Haushalts (die im ersten Befragungsjahr identisch sind) und der unveränderlichen Personennummer zusammen. Diese Personennummer wird für jede im Befragungshaushalt lebende Person vergeben, unabhängig davon, ob diese an der Befragung teilnimmt (ebd., S. 318f).

Wie in Abbildung 2.1, die die Entwicklung der Teilstichproben von 1984 bis 2010 abbildet zu

sehen ist, kann dieses Schneeballsystem die Stichprobenausfälle nicht vollständig kompensieren, weshalb die Zahl der teilnehmenden Haushalte stetig sinkt. Aus diesem Grund werden regelmäßig Refreshment-Samples gezogen, mit denen die Zahl der teilnehmenden Haushalte nicht nur konstant gehalten werden, sondern die Stichprobengröße in den vergangenen Jahren deutlich erhöht werden konnte.

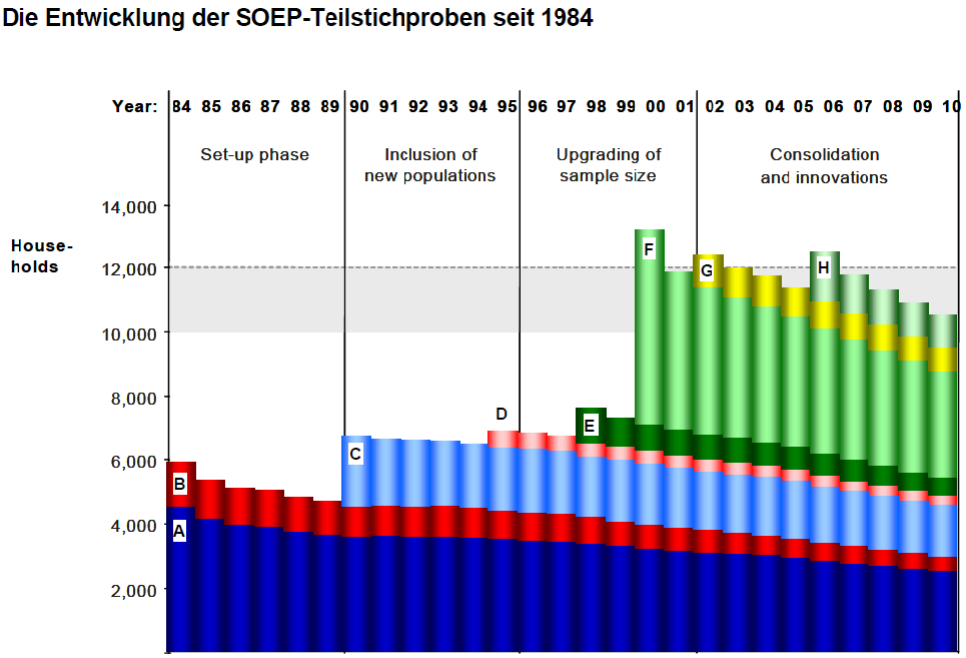

ABBILDUNG 3.1: Quelle: High Incomes in Household Surveys: Sample G of the German Socio-Economic Panel Study (SOEP)

Da weder das Schneeballsystem, noch die allgemeinen Refreshment-Samples eine ausreichende Repräsentation bestimmter Haushaltstypen gewährleisten können, werden diese beiden Methoden durch eine weitere Variante ergänzt, der Ziehung spezieller Sub-Samples.
Dies geschah beispielsweise bereits in der ersten Befragungswelle, indem durch das Sub-Sample B speziell Haushalte mit einem aus den damals üblichen Gastarbeiterländern stammenden Haushaltsvorstand in das SOEP aufgenommen wurden. Im Jahr 1990 wurde das SOEP durch das Sub-Sample C ergänzt, das ostdeutsche Haushalte beinhaltete, im Jahr 1995 wurde das Sample D, das Zuwandererhaushalte enthielt und im Jahr 2002 schließlich die Hocheinkommensstichprobe mit Haushalten, deren Nettoäquivalenzeinkommen 4.500 EUR überstieg, gezogen.
Diese Sub-Samples sind auch im Hinblick auf Einkommensanalysen entscheidend, da nur durch sie für eine umfassende Repräsentation der deutschen Gesamtbevölkerung in der Stichprobe gesorgt wird. Würden Haushalte mit sehr geringen, oder sehr hohen Einkommen respektive Vermögen sys-

tematisch unterrepräsentiert, oder schlimmstenfalls gar nicht beachtet, wären Analysen zur Armut und zum Reichtum, die in Deutschland im Allgemeinen als relative Größen betrachtet werden, verfälscht.
Wegen des Schneeballsystems, des Ausfalls einiger Haushalte aufgrund eines systematisch anderen Antwortverhaltens, sowie der durch die speziellen Sub-Samples existierenden Überrepräsentation bestimmter Haushaltstypen, müssen dennoch Hochrechnungsfaktoren beachtet werden.
Die Schätzung dieser Faktoren folgt dem Ansatz von Horwitz und Thompson, über den Kehrwert der Auswahlwahrscheinlichkeit der jeweiligen Stichprobeneinheit. Da sich dieser Ansatz jedoch auf Querschnitterhebungen bezieht, werden für das SOEP weiterentwickelte Verfahren angewandt, die zusätzlich die Antwort- beziehungsweise Verbleibwahrscheinlichkeiten in den weiteren Wellen miteinbeziehen (vgl. Wagner 2008, S. 313).
Die Hocheinkommensstichprobe G, die gerade für Fragen der Reichtumsforschung besonders relevant ist, führte zu einer Veränderung der Hochrechnungsfaktoren der Haushalte, deren monatliches Nettoäquivalenzeinkommen 4.500 EUR übersteigt. Da das SOEP zwar auch Personenbefragungen beinhaltet, sich das Nettoeinkommen jedoch auf den gesamten Haushalt bezieht, ist die Bestimmung des Äquivalenzeinkommens von besonderer Wichtigkeit, da nicht jeder Haushalt mit hohem Nettoeinkommen als Hocheinkommenshaushalt gesehen werden kann. Teilt sich das Einkommen auf viele Personen auf, sind diese nicht mehr als einkommensstark zu betrachten; dies gilt insbesondere, da das SOEP neben Personenhaushalten auch Wohnheime und staatliche Einrichtungen mit in die Stichprobe aufnimmt. Bei der Berechnung des Nettoäquivalenzeinkommens richtet sich das SOEP nach den Vorgaben der OECD-Skala, unter anderem, um eine internationale Vergleichbarkeit zu gewährleisten.
Vor dem Sub-Sample G lagen 97,5 % aller in der Stichprobe vertretenden Haushalte unter der Hocheinkommensschwelle. Durch die Hinzunahme von rund 1.200 Haushalten, die diese Grenze zum Zeitpunkt der ersten Erhebung überschritten und die daraus folgende Überrepräsentation dieser Haushalte, mussten die Hochrechnungsfaktoren angepasst werden, um die Gesamtverteilung der Einkommens- und Vermögenswerte nicht zu verzerren. Dennoch ist eine solche Überrepräsentation hinsichtlich der Reichtumsforschung sinnvoll, da diese Haushalte aufgrund ihrer geringen Grundgesamtheit, selbst bei einer perfekten Zufallsstichprobe, zu selten gezogen werden, um zuverlässige Aussagen über sie treffen zu können. Diesem Umstand wurde durch das Sub-Sample G Rechnung getragen.
Erfreulicherweise zeigte sich im Zuge dieser speziellen Stichprobe außerdem, dass die Vermutung, Hocheinkommenshalte würden sich seltener an Umfragen beteiligen für das SOEP nicht bestätigt werden kann. Allerdings sprechen Wagner et al. in ihrem Bericht später von der Relevanz der Imputation von Vermögenswerten im SOEP, da hier ein Strukturunterschied im Antwortverhalten beobachtet werden konnte und ein Weglassen der fehlenden Werte zu Verzerrungen führen würde

(vgl. Wagner et al. 2007b, S. 14).
Zudem wurde nach der ersten Vermögensbilanz, die im Jahr 1988 erhoben wurde, eine erhöhte Drop-Out-Rate beobachtet. Dies war der Grund, weshalb bis zum Jahr 2002, dem Jahr, in dem auch die Hocheinkommensstichprobe gezogen wurde, auf Befragungen zum Vermögen verzichtet wurde. Erst, nachdem die Stichprobengröße konsequent erhöht worden war, wurde ein erneuter Versuch gestartet und in den Jahren 2007 und 2012 wiederholt. Bedacht werden sollte hierbei, dass ein großer Stichprobenumfang das Ausscheiden von Befragten zwar abfangen kann, es jedoch zu Schwierigkeiten führen könnte, sollte dieses Ausscheiden struktureller Natur sein, also beispielsweise hauptsächlich vermögende Haushalte betreffen.
Weiterhin muss bedacht werden, dass gleich hohe Antwortwahrscheinlichkeiten zwar wünschenswert sind, da sie die Gefahr von Verzerrungen verringern, Messfehler jedoch dennoch nicht ausgeschlossen werden können, sofern nicht angenommen werden kann, dass diese unabhängig vom Einkommen respektive Vermögen auftreten.
So ist es beispielsweise denkbar, dass hohe Einkommens- und Vermögenswerte schwerer eingeschätzt werden können und deshalb stärker über- oder unterschätzt werden als dies bei geringen Einkommen und Vermögen der Fall ist. Dies würde wiederum zu einer Verletzung der Annahme (2) der Testtheorie führen, nach der der Messfehler nicht mit dem wahren Wert der Messung korreliert sein darf. Durch das Fragebogendesign des SOEP, das eine retrospektive Beantwortung dieser Fragen vorsieht, wird diese Gefahr zusätzlich erhöht.
Weiterhin werden Steuerabgaben und Sozialversicherungsbeiträge im SOEP nicht direkt durch die Befragten angegeben, sondern anschließend simuliert. Im Fall der Sozialversicherungsbeiträge ist dies eine relativ zuverlässige Methode, doch das komplizierte Steuersystem kann schnell zu fehlerhaften Werten führen. Hier sind insbesondere Hocheinkommenshaushalte betroffen, da Steuerabgaben für diese Art der Haushalte schnell überschätzt werden (vgl. Becker et al. 2003, S. 71f). Dies führt zu einer systematischen Verzerrung der errechneten Werte und verletzt Annahme (4) der Testtheorie, die besagt, dass der Messfehler nicht mit dem wahren Wert einer anderen Messung korrelieren darf.
Ein Vorteil der bereits erwähnten Hochrechnungsfaktoren ist, dass durch diese bekannte Messfehler ausgeglichen werden können. So ist beispielsweise bekannt, dass die Messfehler bei Einkommensangaben in der ersten Befragungswelle besonders hoch sind. Daher wird der Hochrechnungsfaktor für diese Angaben auf 0 gesetzt, sodass diese Werte nicht die Datenanalyse einfließen (vgl. Wagner et al. 2007b, S. 8). Dieses Vorgehen wird durch das Panel-Design ermöglicht.
Weiterhin ist es denkbar, dass Messfehler, die die Annahmen der Klassischen Testtheorie verletzen, durch eine geeignete Datentransformation reduziert werden können. So könnte in dem Fall, dass - wie oben beschrieben - der Messfehler der Einkommensmessung direkt mit dessen wahrem Wert korreliert ist, ein steigendes Einkommen also zu einem steigenden Messfehler führt, eine Lo-

garithmierung der Werte diesen Effekt verringern. Später werden wir noch sehen, dass eine solche Transformation auch im Hinblick auf die Anwendung statistischer Methoden zur Datenanalyse sinnvoll sein kann.
Ein weiteres Problem aller Haushaltsumfragen, das die Reichtumsforschung indirekt betrifft, ist die sogenannte bittere Armut. Personen, die von dieser betroffen sind, tauchen im Allgemeinen nicht in diesen Umfragen auf, da sie häufig von Obdachlosigkeit betroffen sind und daher für Haushaltsumfragen nicht zur Grundgesamtheit zählen und nicht in einer Stichprobe gezogen werden können. Dies führt zu einer weiteren Verzerrung der Einkommensverteilung, die wiederum grundlegend für die Reichtumsforschung ist, sofern Reichtum als relative Größe betrachtet wird.
Aufgrund der Multidisziplinarität des SOEP beinhaltet der Fragebogen nicht nur Fragen zu soziodemographischen Angaben, sondern bietet die Möglichkeit, eine Reihe weiterer Werte zu ermitteln. So lag von Beginn an ein wesentlicher Teil des Interesses auf Fragen zur Lebenszufriedenheit, doch inzwischen wurden sogar medizinische Daten wie der in Welle 29 vorgenommene Greifkrafttest erhoben.
Dies schafft nicht nur die Voraussetzung für die Verwendung des SOEP in unterschiedlichen, wissenschaftlichen Disziplinen, sondern ermöglicht es Forschern ebenfalls nicht-finanzielle Aspekte von Reichtum zu betrachten. Das Einbeziehen solcher Daten führt, wie bei den messtheoretischen Grundlagen bereits ausgeführt, zum Ausschluss der Klassischen Messtheorie als Grundlage der Messung. Der Aufbau der Erhebungsinstrumente des SOEP legt daher nahe, dass entweder die Operationale oder die Repräsentationale Messtheorie zugrunde gelegt wurden.
Da letztere derzeit bestimmend ist und besonders häufig als Grundlage verwendet wird, kann vermutet werden, dass auch die Messinstrumente des SOEP darauf basieren. Dies hätte den Vorteil, dass die gewonnenen Daten einfach zu interpretieren und mit anderen Studien vergleichbar wären. Allerdings ist zu beachten, dass Forscher, die die Daten des SOEP verwenden, diese frei wählen und daher ihre in ihrer Arbeit verwendeten Variablen auf jene beschränken können, die auch nach den Maßstäben der Klassischen Messtheorie hätten erhoben werden können. Die Frage, welche Messtheorie dem gesamten SOEP zugrunde liegt, ist daher kaum zu beantworten, sondern muss im Einzelfall, mit Hinblick auf die Forschungsfrage, die Analysemethoden und die weitere Verwendung der gewonnenen Daten betrachtet werden.
Die Bereiche des SOEP, die sich mit dem Einkommen und Vermögen der Teilnehmer befassen, wurden in Anlehnung an die Canberra Group entwickelt, einer internationalen Expertengruppe, die 1996 auf Initiative des Australian Bureau of Statistics gegründet wurde und sich mit Problemstellungen der Einkommensmessung auseinandersetzt (vgl. Canberra Group 2001, Preface xi). Dieser Ansatz wurde gewählt, um eine bessere internationale Vergleichbarkeit der erhobenen Daten zu gewährleisten (vgl. Wagner et al. 2007a, S. 161).
Ein exakt festgelegtes Messinstrument, das dazu dient, durch verschiedene Studien erhobene Da-

ten zu vergleichen, ist ein Hinweis auf die Operationale Messtheorie, da nur so garantiert werden kann, dass dieselben Messwerte generiert werden. Die Annahme, dass zwei verschiedene Messinstrumente dennoch denselben Wert abbilden, kann in diesem Fall nicht getroffen werden, da diese Theorie ohne die Annahme eines zugrundeliegenden wahren Werts auskommt.
Andererseits kann argumentiert werden, dass das SOEP der Repräsentationalen Messtheorie folgt und die Canberra Group als Expertengruppe dient, deren Arbeit als Form der Inhaltsvalidität für Einkommens- und Vermögensmessungen betrachtet werden kann. In diesem Fall ist das Ziel weniger die Konstruktion eines einheitlichen Messinstruments, sondern die Abbildung des gesuchten empirischen Relativs.
Ergänzend könnte ein Vergleich mit der Volkswirtschaftlichen Gesamtrechnung als Überprüfung der Kriteriumsvalidität gesehen werden. 2002 kam das SOEP, unter Einbezug der Hocheinkommensstichprobe auf eine Nachweisquote von 70 % des VGR-Aggregats (vgl. Wagner et al. 2007b, S. 14). Damit liegt das SOEP nicht nur im internationalen Vergleich sehr hoch, sondern übertrifft sogar die Einkommens- und Verbrauchsstichprobe.
Die Differenz wird im Allgemeinen damit erklärt, dass einerseits Haushalte ihr Vermögen nicht korrekt angeben und andererseits besonders vermögende Haushalte nicht in den jeweiligen Stichproben enthalten sind. Dieses Problem tritt auch im SOEP, trotz der speziellen Hocheinkommensstichprobe auf, da die Grundgesamtheit wirklich vermögender Haushalte so klein ist, dass diese zwar theoretisch in der Stichprobe enthalten sein könnten, faktisch jedoch nicht gezogen werden (ebd., S. 12).
Es kann vermutet werden, dass das SOEP zum Ziel hat, Messinstrumente zu entwickeln, die wahre Werte abbilden sollen und damit der Repräsentationalen Messtheorie folgt. Dies ist insbesondere wegen des Panel-Designs sinnvoll, da der Operationalen Messtheorie folgend jede Veränderung der Messinstrumente eine Vergleichbarkeit mit den zuvor gewonnenen Werten unmöglich macht. Allerdings wird auch der Tatsache, dass eine korrekte Abbildung des empirischen Relativs nicht immer gewährleistet werden kann, Rechnung getragen und Messinstrumente so aufgebaut, dass diese denen anderer relevanter Datenquellen, auch international betrachtet, ähneln.
Neben Fragen zur Validität der erhobenen Daten, müssen diese, unabhängig von der zugrunde gelegten Messtheorie, objektiv und reliabel sein.
Die Fragebögen des SOEP wurden früher per Paper And Pencil Interview (PAPI) erhoben, eine Befragungsmethode, die später durch Computer Assisted Personal Interviews (CAPI) ergänzt wurde (vgl. Wagner et al. 2007a, S. 152). Ende 2006 wurden spezielle Interviewerbefragungen durchgeführt, um Interviewereffekte aufzudecken und gegebenenfalls in Zukunft in das weitere Vorgehen miteinbeziehen zu können (ebd., S. 157). Durch diese Maßnahmen kann von einer hohen Objektivität ausgegangen werden.
Das Panel-Design des SOEP kann als eine Form des Test-Retest-Designs zur Prüfung der Re-

liabilität betrachtet werden, da den Befragten mit gewissem zeitlichen Abstand dieselben Fragen vorgelegt werden. Da die letzte Befragung zu diesem Zeitpunkt mindestens ein Jahr zurück liegt, können Gedächtniseffekte weitestgehend ausgeschlossen werden.
Allerdings ist diese Methode aus demselben Grund nur bei Variablen sinnvoll, die kaum Veränderungen unterworfen sind. Andernfalls können Abweichungen nur schwer auf mangelnde Reliabilität des Messinstruments zurückgeführt werden. Dies gilt insbesondere bei den Befragungsschwerpunkten, die nur etwa alle fünf Jahre erhoben werden, zu denen auch die Vermögensbilanz zählt. Das Einkommen wird zwar im jährlichen Abstand erfasst, kann sich aber innerhalb eines Jahres schnell verändern. Da im SOEP nicht nur nach dem gesamten Nettoeinkommen gefragt, sondern dessen genaue Zusammensetzung erhoben wird, können Veränderungen jedoch relativ gut nachvollzogen werden.
Insgesamt zeigt sich, dass das Sozio-oekonomische Panel ein Format ist, das aufgrund seines Stichproben-, Studien- und Fragebogendesigns wertvolle Daten liefert, die insbesondere auch geeignet sind, um Fragen zur Reichtumsforschung zu untersuchen. Es wird beständig weiterentwickelt, um Fehlerquellen aufzudecken oder zu beheben und Daten für verschiedenste wissenschaftliche Disziplinen bereitzustellen.

3.2 Die Einkommens- und Verbrauchsstichprobe

Die Einkommens- und Verbrauchsstichprobe ist eine der, wenn nicht sogar die wichtigste Datenquelle Deutschlands bezüglich Einkommens- und Vermögenserhebungen. Sie bildet die Grundlage für die Bemessung der Regelsätze des Arbeitslosengelds II, der Ermittlung des Verbraucherpreisindex und wird für Schätzungen wichtiger volkswirtschaftlicher Gesamtgrößen benötigt. Weiterhin werden ihre Daten für den Armuts- und Reichtumsbericht der Bundesregierung genutzt (vgl. Qualitätsbericht 2012, S. 4).
Die erste Erhebung fand 1962 und 1963 statt und wurde von da an im fünfjährigen Rhythmus durchgeführt. 1993 wurde auch Ostdeutschland in die Stichprobe miteinbezogen und seit 1998 werden die Daten der EVS deutschlandweit erhoben. Gleichzeitig wurden erst in diesem Jahr auch Haushalte mit ausländischem Haushaltsvorstand in die Grundgesamtheit aufgenommen.
Die EVS ist, ebenso wie das Sozio-oekonomische Panel eine Haushaltsstichprobe, definiert jedoch anders als das SOEP Gemeinschaftsunterkünfte wie beispielsweise Wohnheime nicht als Haushalt und kann daher keine Aussagen über diese Form der Wohnverhältnisse treffen (ebd., S. 3).
Die Teilnahme an der EVS ist freiwillig, die Auswahl der Haushalte basiert auf einer Quotenstichprobe (vgl. Becker et al. 2003, S. 61). Mit rund 60.000 teilnehmenden Haushalten ist ihr Stichpro-

benumfang nicht nur innerhalb Deutschlands, sondern EU-weit einzigartig und richtet sich nach den gesetzlich maximal möglichen 0,2 % der Erhebungsgesamtheit des Mikrozensus. Dieser wird auch als Grundlage für die Quotenauswahl herangezogen (vgl. Qualitätsbericht 2012, S. 6).
Dieses Quotendesign kann in der statistischen Auswertung zu Problemen führen, da es möglicherweise zu einer Verzerrung der Varianzschätzung führt. Dies ist insbesondere der Fall, wenn es sich um eine Randquotenstichprobe handelt, die Quotenmerkmale also voneinander getrennt betrachtet werden. Um dennoch einen unverzerrten Schätzer zu liefern, müssten die verwendeten Randmerkmale unabhängig sein und wenigstens ein Teil der Merkmale eine Gleichverteilung aufweisen (vgl. Quatember 1997, S. 22).
Eine Auswahl nach kombinierten Quoten kann hingegen als geschichtete Zufallsstichprobe behandelt werden und liefert daher, zumindest unter Urnenbedingungen, einen unverzerrten Schätzer (ebd., S. 21). Da die Stichprobe der EVS einer Auswahl nach kombinierten Quoten entspricht, können daher Methoden, die eine Varianzschätzung beinhalten, angewandt werden, sollten jedoch dennoch mit Berücksichtigung des Auswahlverfahrens interpretiert werden. Dieser Argumentation folgt auch der Qualitätsbericht des Statistischen Bundesamts zur Einkommens- und Verbrauchsstichprobe 2008.
Da die Teilnahme an der EVS freiwillig ist, kann auch die quotierte Auswahl der teilnehmenden Haushalte keine Garantie für eine ausreichende Repräsentation der Grundgesamtheit liefern, da es durch Teilnahmeverweigerungen weiterhin zu Verzerrungen kommen kann (vgl. Becker et al. 2003, S. 66). Im Jahr 2008 lag die Ausfallquote aller angeschriebenen Haushalte vor Beantwortung des ersten Fragebogens bei 24 %. Ausgehend von den Haushalten, die den ersten Fragebogen eingereicht hatten, kam es innerhalb des Berichtszeitraums zu einem Ausfall von 6,5 %.
Wie das Sozio-oekonomische Panel, kann auch die EVS keine Aussagen über bittere Armut treffen, da Obdachlose per Definition nicht zur Grundgesamtheit gehören. Ein in Bezug auf Reichtumsmessung größeres Problem ist jedoch die Unterrepräsentation von Haushalten mit sehr hohem Einkommen. Da diese Haushalte aufgrund ihrer kleinen Grundgesamtheit nur selten ausgewählt werden und möglicherweise zusätzlich eine höhere Wahrscheinlichkeit der Antwortverweigerung aufweisen, kann durch die Daten der EVS keine Aussage über sie getroffen werden, weshalb Haushalte mit einem Einkommen von mehr als 18.000 EUR aus der Betrachtung ausgeschlossen werden. Ebenso wie Personen ohne festen Wohnsitz und Personen, die in Gemeinschaftsunterkünften oder Anstalten leben, zählen sie nicht zu der für die EVS relevante Grundgesamtheit (vgl. Statistisches Bundesamt 2012, S. 3).
Ein Schwerpunkt der Erhebung der EVS ist die Erfassung sämtlicher Einnahmen und Ausgaben von Privathaushalten. Hierzu erfolgt die Erhebung in mehreren Schritten. Zunächst wird ein Allgemeiner Fragebogen beantwortet, der soziodemographische Angaben enthält, unter anderem die Einkommenskategorie. Dieser dient später für die Hochrechnung der Daten, die Einkommenskate-

gorie wird für keine weiteren Analysen herangezogen. Weiterhin beschäftigt sich ein Fragebogen mit dem Geld- und Sachvermögen der teilnehmenden Haushalte.
Anschließend muss jeder Haushalt ein Haushaltsbuch führen. Dieses wurde bis einschließlich der Befragung 1993 über das ganze Jahr erhoben, seit der Erhebung 1998 werden die teilnehmenden Haushalte in vier Teile gesplittet, sodass jeder Haushalt die Aufzeichnungen nur noch über ein Quartal führen muss.
Dieses Vorgehen verringert den Arbeitsaufwand der Haushalte und kann zu einer höheren Teilnahmebereitschaft führen, hat jedoch den Nachteil, dass Sondereinnahmen wie Urlaubs-, oder Weihnachtsgeld und Dividendenausschüttungen nicht mehr für alle Haushalte erfasst werden können. Dies führt zu einer eingeschränkten Vergleichbarkeit der älteren Erhebungen mit denen aus jüngeren Jahren (vgl. Becker et al. 2003, S. 69).
Das Führen eines Haushaltsbuchs bietet Vorteile gegenüber einfachen, wenn auch ausführlichen retrospektiven Befragungen, wie sie beim SOEP angewandt werden, da die Angaben kontinuierlich eingetragen und nicht aus dem Gedächtnis beantwortet werden. Dies kann zu einer Verringerung von Messfehlern führen, die, wie oben beschrieben, im Fall des SOEP insbesondere bei Hocheinkommenshaushalten kritisch zu betrachten sind.
Da es sich bei der EVS jedoch um eine Längsschnittstudie statt eines Panel-Designs handelt, fallen hier die Ergebnisse des SOEP, dass Fragen zum Einkommen in der ersten Erhebungswelle besonders hohe Messfehler aufweisen, deutlich stärker ins Gewicht, da mit jeder Erhebung andere Haushalte an der Befragung teilnehmen. Ein Ausschluss der Ergebnisse durch Anpassung des Hochrechnungsfaktors ist also nicht möglich. Durch die exaktere Erhebung mittels des Haushaltsbuchs ist es jedoch auch vorstellbar, dass die Messfehler insgesamt geringer ausfallen und dieser Nachteil aufgehoben wird.
Weiterhin werden in der EVS, im Gegensatz zum SOEP, Sozialversicherungsbeiträge und Steuerabgaben nicht simuliert, sondern direkt abgefragt, eine weitere Möglichkeit, um Fehlerquellen zu verringern. Allerdings werden eventuelle Steuernachzahlungen, respektive -erstattungen aufgrund der Jahresgrenze nicht berücksichtigt (ebd., S. 71).
Die Frage, welche Messtheorie bei der Gestaltung der EVS zugrunde gelegt wurde, ist wie beim SOEP nur schwer und kaum eindeutig zu beantworten. Da sie sich hauptsächlich auf finanzielle Aspekte konzentriert, ist ein Aufbau nach den Grundsätzen der Klassischen Messtheorie nicht auszuschließen, schränkt die Erfassung interessierender Merkmale jedoch dennoch so weit ein, dass diese Annahme unwahrscheinlich ist.
Vielmehr können auch hier sowohl Aspekte der Operationalen, wie auch der Repräsentationalen Messtheorie gefunden werden. Da die EVS die Grundlage für viele politische Entscheidungen bildet, kann argumentiert werden, dass ihre Erhebungen wahre, in der Bevölkerung vorkommende Werte abbilden sollten. Andererseits wurde bereits bei der Vorstellung der Operationalen Mess-

theorie angeführt, dass die Abbildung eines wahren Werts keine Voraussetzung für politische Entscheidungen darstellen muss, sofern die durch die Messung erzeugten Werte dennoch einer geeigneten Definition folgen.
Die internationale Vergleichbarkeit kann, zumindest EU-weit, auch durch gesetzlich vorgeschriebene Regelungen zur Haushaltsbudgeterhebung erreicht werden. Eine solche Regelung existiert derzeit zwar noch nicht, doch es gibt bereits ein sogenanntes "gentlemen's agreement", das zu eben diesem Zweck getroffen wurde (vgl. Qualitätsbericht 2012, S. 9).
Die vier Erhebungsinstrumente der EVS werden, mit Ausnahme des ersten Fragebogens, der auch Online beantwortet werden kann, in Papierform erhoben. Der Hauptteil der Befragung, das Haushaltsbuch, wird von den Haushalten selbstständig ausgefüllt, wodurch Interviewereffekte verhindert werden. Die differenzierte Aufstellung zu Einnahmen und Ausgaben führen ebenfalls zu der Annahme, dass die Erhebung das Kriterium der Objektivität erfüllt.
Durch das Längsschnitt-Design der EVS ist eine Überprüfung der Reliabilität mittels Test-Retest-Designs, wie sie beim SOEP möglich ist, nicht gegeben, zumal der lange Erhebungszeitraum und der damit verbundene hohe Aufwand einer erneuten Befragung von Haushalten, die bereits einmal teilgenommen haben, nicht sinnvoll scheint.
Da die EVS jedoch nicht nur Einnahmen, sondern gleichermaßen Ausgaben aller Haushalte erfasst und diese einander entsprechen sollten, ist dennoch eine Plausibilitätsüberprüfung der Angaben möglich. Dies könnte man als Parallel-Test-Design betrachten, da beide Varianten dieselbe Komponente, nämlich das zur Verfügung stehende Vermögen, innerhalb derselben Erhebung abfragen. Auf Basis dieser Angaben kontrollieren die Statistischen Landesämter die eingegangenen Unterlagen, bevor sie sie an das Statistische Bundesamt weiterleiten. Dieses Vorgehen ermöglicht eine Rücksprache mit den Haushalten, deren Angaben unplausibel waren und bietet die Möglichkeit der zeitnahen Korrektur (ebd., S. 7). Dadurch kann die Ausfallquote weiter gesenkt werden.
Durch den großen Stichprobenumfang, die umfassenden Angaben zu Einnahmen und Ausgaben der teilnehmenden Haushalte und die dadurch ermöglichte Plausibilitätsüberprüfung, stellt die EVS eine der wichtigsten Datenquellen zur Einkommens- und Vermögensforschung Deutschlands dar. Da sie nicht nur die Einnahmen der Haushalte betrachtet, sondern auch deren Ausgaben und damit regionale Preisunterschiede beobachten kann, die wichtig für die differenzierte Festlegung von Armuts- und Reichtumsgrenzen sind, ist sie dem SOEP in der rein finanziellen Betrachtung dieser Größen überlegen.
Jedoch ignoriert sie nicht-finanzielle Aspekte des Vermögens und steht in der Kritik, aus Vermögen generiertes Einkommen, sowie Einkommen aus selbstständiger Tätigkeit zu unterschätzen. Die Nichterfassung von Hocheinkommenshaushalten stellt insbesondere in der Reichtumsforschung ein Defizit der EVS dar, das nur schwer ausgeglichen werden kann.

4 Methoden für zensierte Daten

Die Tatsache, dass Haushalte, deren monatliches Nettoeinkommen über 18.000 EUR liegt, in der Auswertung der Einkommens- und Verbrauchsstichprobe keine Beachtung finden, erschwert deren Einsatz für die Reichtumsforschung. Dies ist bedauerlich, da mit der EVS eine umfangreiche Stichprobe vorliegt, die die Einnahmen und Ausgaben deutscher Haushalte mit einer Genauigkeit erfasst, die bisher keine andere Datenquelle erreichen konnte.
Die Problematik der zensierten Daten tritt in verschiedenen wissenschaftlichen Disziplinen auf und über die Jahre wurden Verfahren entwickelt, die diese weder ignorieren, noch aus der Analyse ausschließen, sondern sich vielmehr die darin enthaltenen Informationen zunutze machen.
Daher ist es sinnvoll, nach Wegen zu suchen, die Daten der EVS dennoch auch für Aussagen über Hocheinkommenshaushalte nutzbar zu machen. Hierzu ist es notwendig, zunächst den genauen Zensierungsmechanismus zu betrachten, der dieser Datenquelle zugrunde liegt.
Wie im obigen Abschnitt über die EVS bereits beschrieben, umfasst deren Stichprobe rund 60.000 Haushalte, die nach einem Quotenverfahren ausgewählt werden. In dieser Stichprobe können auch Haushalte mit einem monatlichen Nettoeinkommen enthalten sein, das die gesetzte Grenze von 18.000 EUR überschreitet. Da dies jedoch in geringem Umfang geschieht und, anders als im Sozio-oekonomischen Panel, davon ausgegangen wird, dass hier die Antwortbereitschaft deutlich geringer ist als bei Haushalten mit mittleren Einkommen, werden diese nicht zur Analyse herangezogen und gelten damit auch nicht als Teil der Grundgesamtheit, über die die EVS Aussagen zu treffen vermag (vgl. Statistisches Bundesamt 2012, S. 3).
An dieser Stelle sollen zunächst drei Zensierungsmodelle vorgestellt werden, die in der Realität häufig auftreten und das Vorgehen in der Datenanalyse beeinflussen. Hierfür wird folgende Notation verwendet:
$n =$ Untersuchungsobjekt, mit $n = 1, ..., N$
$c_n =$ Beobachtungszeitraum
$T_n =$ Verweildauer
$d_n =$ Zensierungsindikator mit $d = 0$, wenn T_n zensiert und $d = 1$, wenn T_n unzensiert ist
Modell 1 geht von einem zuvor festgelegten Beobachtungszeitraum c_n aus, der für jedes Individuum n verschieden, oder eine Konstante c sein kann. Beobachtbar ist in diesem Fall lediglich $min(T_n, c_n)$.
In *Modell 2* ist c eine für alle Individuen einheitliche Konstante, die jedoch zunächst unbekannt ist. Stattdessen tritt die Zensierung nach einer zuvor festgelegten Anzahl von Ereignissen auf. Damit wird c zu einer Zufallsvariable.
Modell 3 betrachtet c_n ebenfalls als Zufallsvariable, die nun jedoch für jedes Individuum verschie-

den und von T_n unabhängig ist (vgl. Fahrmeir 1996, S. 303).
Alle drei Modelle haben gemein, dass die Zensierung erst nach Beginn des Beobachtungszeitraums c_n eintritt. Dies bedeutet, dass die Untersuchungsobjekte bis zu diesem Zeitpunkt beobachtet werden konnten und daher ihre minimale Verweildauer dem Beobachtungszeitraum entspricht. Diese Form der Zensierung ist allgemein als Rechtszensierung bekannt.
Im Fall der EVS werden Hocheinkommenshaushalte zwar in die Stichprobe mit aufgenommen, in der späteren Auswertung jedoch nicht mehr beachtet. Weiterhin ist fraglich, ob die genauen Einkommensdaten zugänglich sind, oder aus Datenschutzgründen anonymisiert werden, um keine Rückschlüsse auf die Befragten zuzulassen.
Dennoch liegen für diese Haushalte Beobachtungen vor. Geht man nun davon aus, dass deren monatliches Einkommen im Datensatz nicht mehr exakt ausgewiesen, sondern mit > 18.000 EUR angegeben wird, kann man dies als festen Zensierungszeitpunkt c betrachten, womit der Zensierungsvorgang Modell 1 entspräche.
Da nun bekannt ist, welcher Vorgang den Daten der EVS zugrunde liegt, können Modelle geprüft werden, die auf diesen eingehen und, trotz deren Beschränkung, einen möglichst hohen Informationsgewinn aus den Daten ermöglichen.

4.1 Tobit-Regression

Hier sei zunächst das sogenannte Tobit-Modell erwähnt, das zum ersten Mal 1958 in einer Untersuchung zum Konsumverhalten von Haushalten Anwendung fand. James Tobin stellte darin fest, dass die Nichtbeachtung zensierter Variablen zu einer Verzerrung der geschätzten Regressionsparameter β_i führt und etablierte mit dem Tobit-Modell eine Möglichkeit, den Zensierungsvorgang in die Schätzung miteinzubeziehen (vgl. Windzio 2013, S. 261).
Für die Anwendung des Tobit-Modells spielt es keine Rolle, ob die Daten links-, rechts- oder beidseitig zensiert sind. Da Tobins Modell ursprünglich linkszensierte Daten enthielt, werden diese häufig zur Illustration des Modells herangezogen; es lässt sich jedoch analog für rechtszensierte Daten anwenden. Entscheidend ist, dass diese an einer festen Stelle c zensiert wurden, eine Bedingung, die im Fall der EVS erfüllt wird. Da bei dieser eine Rechtszensierung der Daten vorliegt, wird das Modell im Folgenden an diesem Beispiel vorgestellt.
Das Modell geht davon aus, dass eine latente Variable y^* nur bis zum Erreichen eines Schwellenwertes τ beobachtbar ist. Das heißt, y^* ist beobachtbar, wenn $y^* < \tau$ gilt. Daraus folgt

$$y_i = \begin{cases} y_i^*, & \text{wenn } y_i^* < \tau \\ \tau, & \text{sonst} \end{cases}$$

Beziehungsweise im Fall der EVS, unter der Annahme, dass zensierte Daten als solche ausgewiesen und nicht einfach als fehlende Werte behandelt werden

$$y_i = \begin{cases} y_i^*, & \text{wenn } y_i^* < 18.000 \\ 18.000, & \text{sonst} \end{cases}$$

Wird y^* durch eine Regression geschätzt, gilt $y^* = x'\beta + \epsilon$ und damit $x'\beta + \epsilon < \tau$ beziehungsweise $\epsilon < -x'\beta$.
Daraus lässt sich ableiten, dass die Wahrscheinlichkeit, dass y^* beobachtbar ist, der Wahrscheinlichkeit entspricht, dass der Fehler ϵ kleiner als $-x'\beta$ ist.
Mit diesen Überlegungen lässt sich nun der Inverse Mills-Ratio aufstellen, der die Verteilung der latenten Variable auf den beobachteten Bereich konditioniert (vgl. Windzio 2013, S. 261):

$$\frac{\phi_i(\frac{\tau - x'\beta}{\sigma})}{\Phi_i(\frac{\tau - x'\beta}{\sigma})} = \lambda_i$$

Durch die Abhängigkeit von $x'\beta$ wird der Inverse Mills Ratio für jede Beobachtung spezifisch berechnet.
Daraus ergibt sich der bedingte Erwartungswert von ϵ_i

$$E(\epsilon_i | \epsilon_i < -x'\beta) = \sigma \frac{\phi_i(\frac{\tau - x'\beta}{\sigma})}{\Phi_i(\frac{\tau - x'\beta}{\sigma})} = \sigma \frac{\phi_i(\delta_i)}{\Phi_i(\delta_i)} = \sigma \lambda_i(\delta_i)$$

und damit letztlich die Tobit-Regression

$$E(y^* | y^* < \tau) = x'\beta + \sigma \lambda_i(\delta_i)$$

Mit σ als Regressionskoeffizient, der den Einfluss der unabhängigen Variable λ modelliert. λ wird kleiner, je geringer die Zensierungswahrscheinlichkeit für y^* ausfällt (vgl. Windzio 2013, S. 265), wobei $x'\beta$ von δ_i durch ein Probit-Modell geschätzt wird.
Das bedeutet, je höher die Wahrscheinlichkeit einer Zensierung ist, desto stärker wird die Schätzung der latenten Variable y^* durch λ korrigiert. Der Vorteil dieses Modells liegt in der einfachen Anwendung und der Möglichkeit, es sowohl für links-, als auch rechts-, oder sogar intervallzensierte Daten zu verwenden.
Allerdings unterliegt das Tobit-Modell als einfaches Regressionsmodell der Normalverteilungsannahme und es hat sich gezeigt, dass die Koeffizienten des Modells sehr sensitiv auf Abweichungen von dieser Verteilung reagieren (Breen in: Windzio 2013, S. 273). Sollte eine Normalverteilung

der Fehler nicht plausibel sein, muss das Tobit-Modell angepasst werden.
Im Fall der EVS werden die Daten zensiert, da für jene Haushalte mit einem monatlichen Nettoeinkommen von mehr als 18.000 EUR nicht ausreichend Beobachtungen vorliegen, um Aussagen über sie treffen zu können. Da diese Zensierung einseitig stattfindet, also nicht auch auf den Niedrigeinkommensbereich angewandt wird, liegt die Vermutung nahe, dass hier mehr Beobachtungen vorliegen (wieder unter dem Ausschluss der bitteren Armut, die wegen der Haushaltsdefinition im Allgemeinen ebenfalls nicht beobachtet werden kann). Dies spricht gegen eine Normalverteilung der abhängigen Variable und kann zu Verzerrungen führen, wenn das Einkommen über eine einfache Tobit-Regression geschätzt werden soll.
Zunächst sollte überprüft werden, ob eine Transformation der Daten eine annähernde Normalverteilung bewirkt. Dies kann zum Beispiel durch Logarithmierung geschehen, sofern eine Lognormalverteilung vorliegt, die zudem den Vorteil hätte, dass mit dem Einkommen korrelierte Messfehler, deren Risiko unter 3.1 angesprochen wurde, reduziert werden könnten. Allerdings ist zu beachten, dass dadurch, dass die Tobit-Regression empfindlicher auf Abweichungen von der Normalverteilungsannahme reagiert als ein gewöhnlicher KQ-Schätzer, unter Umständen auch eine annähernde Normalverteilung noch immer problematisch ist.
Alternativ kann versucht werden, das Tobit-Modell in ein nicht-parametrisches Rang-Modell umzuwandeln. Diesen Ansatz stellten Conover und Iman 1981 für ein einfaches Regressionsmodell vor, indem sie jeder Beobachtung innerhalb einer Variablengruppe einen Rang zuordneten und damit am Ende den Rang der abhängigen Variable schätzten (vgl. Conover 1984, S. 127).
Diesem Ansatz folgend, testeten Ballenberger et al. 2012 verschiedene Methoden, die sich im Umgang mit zensierten Daten bewährt haben, mittels einer Simulation, die zensierte, nicht normalverteilte Daten generierte und verglichen dabei unter anderem auch das einfache Tobit-Modell mit einem Rang-Tobit-Modell. Dabei wurde das Rang-Tobit-Modell folgendermaßen aufgebaut

$$R(y^*) = \frac{n+1}{2} + \beta(R(x) - \frac{n+1}{2})$$

$$R(y_i) = \begin{cases} R(y_i^*), & \text{wenn } R(y_i^*) > R(\tau) \\ R(\tau), & \text{sonst} \end{cases}$$

mit R als Rang für die entsprechende Variable (vgl. Ballenberger et al. 2012, Supplement S1).
Im Fall der EVS, die ja von einer oberen Beobachtungsgrenze ausgeht, müssten die Vorzeichen umgedreht werden.

$$R(y_i) = \begin{cases} R(y_i^*), & \text{wenn } R(y_i^*) < R(\tau) \\ R(\tau), & \text{sonst} \end{cases}$$

Der Vergleich der beiden Modelle ergab, dass sie bei erfüllter Normalverteilungsannahme gleichwertig sind, während das Rang-Tobit-Modell bei Verletzung dieser Annahme deutlich besser abschneidet (vgl. Abbildung 4.1, sowie Ballenberger et al. 2012, S. 5).

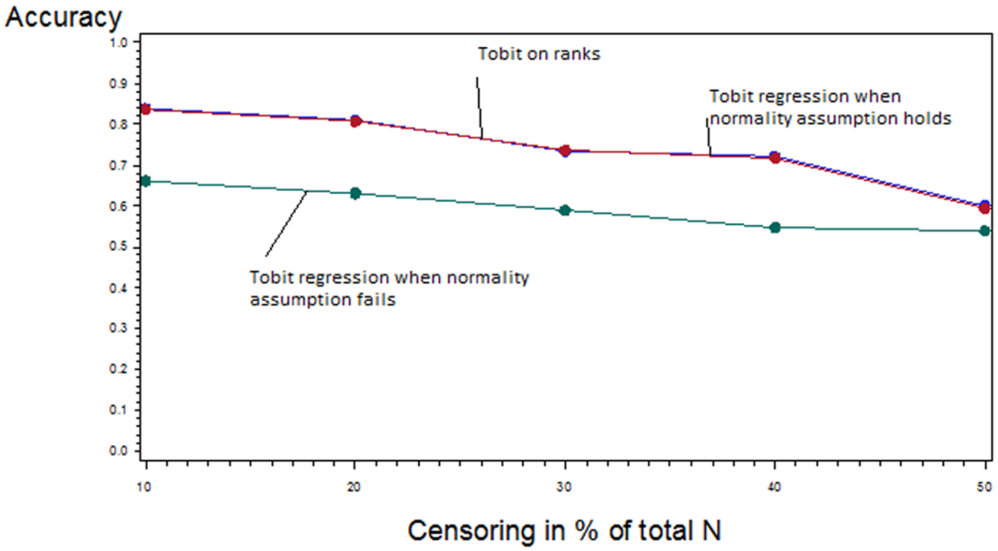

ABBILDUNG 4.1: Quelle: Ballenberger et al. 2012, Novel Statistical Approaches for Non-Normal Censored Immunological Data: Analysis of Cytokine and Gene Expression Data

Unter der Voraussetzung, dass diese Ergebnisse auf rechtszensierte Daten übertragen werden können, was aufgrund des Aufbaus des Tobit-Modells jedoch möglich sein sollte, stellt die Konstruktion eines Rang-Tobit-Modells eine Möglichkeit dar, auf Fälle angewandt zu werden, in denen eine Normalverteilung nicht angenommen werden kann. Damit wäre dies eine mögliche Methode für die nähere Betrachtung der Einkommensdaten der EVS.
Neben dem Einsatz des Rang-Tobit-Modells und der Transformation der Zielvariablen, existiert noch die Möglichkeit, das Tobit-Modell für andere Verteilungen abzuwandeln, sofern Annahmen über diese getroffen werden können. Dies wäre ein Ansatz für Vermögensangaben, die, falls Schulden als negatives Vermögen betrachtet werden, auch nicht-positive Werte annehmen können. Die Zensierung findet bei der EVS jedoch für das Haushaltseinkommen, nicht dessen Vermögen statt, weshalb von einer genaueren Betrachtung dieses Vorgehens an dieser Stelle abgesehen wird.
Möchte man das Einkommen als abhängige Variable schätzen, bietet, wie gerade gezeigt wurde, das Tobit-Modell Möglichkeiten dazu, selbst, wenn keine Normalverteilung angenommen werden

kann. Allerdings können in diesem Fall, wie bei einer einfachen Regression, Schätzungen im negativen Bereich auftreten, die inhaltlich nicht gewollt sind. Diese Kombination, zensierte Daten, deren Schätzungen nicht-negativ sein sollen, führt direkt zu den Lebenszeitmodellen, die in anderen Bereichen bereits vielseitig und erfolgreich angewandt werden und nachfolgend vorgestellt werden sollen.

4.2 Lebensdaueranalyse

Das Problem der zensierten Daten tritt in vielen Bereichen auf, unter anderem in der Industrie, wenn beispielsweise die Lebensdauer von technischen Geräten modelliert werden soll, oder auch in klinischen Studien, die die Wirksamkeit neuer Behandlungsmethoden untersuchen.
Insbesondere Letztere haben häufig das Problem, dass Probanden noch während der Laufzeit aus der Studie aussteigen und nicht weiter beobachtet werden können, oder der Untersuchungszeitraum endet, bevor alle interessierenden Ereignisse eingetreten sind.
Lebenszeitmodelle wurden entwickelt, um trotz dieser Problematik Schätzungen für die interessierende Variable zu liefern und dabei auch den Informationsgehalt der zensierten Daten zu nutzen.
Die oben vorgestellten Zensierungsarten wurden bei der Entwicklung dieser Modelle berücksichtigt und es wurde bereits argumentiert, dass sich die Situation der EVS darauf übertragen lässt, wobei Zensierungsmodell 1, mit $c_n = c = 18.000$, gilt.
Allerdings werden Lebenszeitmodelle üblicherweise eben genau dafür, nämlich Zeiträume als interessierende Variable, verwendet. Theoretisch ließen sie sich jedoch auch auf Einkommensdaten anwenden. Das Einkommen kann, ebenso wie Verweildauern, nicht negativ sein und ist verhältnisskaliert. Die Bedingung, dass die Variable > 0 sein muss, schließt lediglich eine Übertragung der Lebenszeitmodelle auf Vermögensangaben aus, da diese durch die mögliche Existenz von Schulden auch negative Werte annehmen können.
Ein weiteres wichtiges Kriterium für die Anwendung von Lebenszeitmodellen ist die Existenz eines genau definierten Startzeitpunkts und des Ereigniseintritts. Auch diese sind bei der Variable "Einkommen" gegeben. Beobachtet wird hier nur nicht mehr "die Zeit, bis das Ereignis eintritt", sondern "das Einkommen, bis die Einnahmen enden".
Zunächst sollen einige grundsätzliche Begriffe geklärt werden, bevor der Versuch gestartet wird, eine geeignete Anwendung von Lebenszeitmodellen auf Einkommensdaten zu finden.
Zwei der wichtigsten Größen der Lebenszeitanalyse sind die Survivorfunktion $S(t)$ und die Hazardrate $h(t)$.
Für die Survivorfunktion gilt

$$S(t) = P(T \geq t)$$

Sie beschreibt also die Wahrscheinlichkeit, dass das Untersuchungsobjekt einen bestimmten Zeitpunkt t überlebt. Die Funktion ist monoton fallend und bewegt sich als Wahrscheinlichkeit zwischen 1 und 0, wobei zum Startzeitpunkt gilt $S(t = 0) = 1$.
In der Realität ist $S(t)$ wegen der beschränkten Beobachtungszahl im allgemeinen keine glatte Funktion, sondern stufenförmig und nicht immer bis zum Ende, also bis $S(t) = 0$ beobachtbar (vgl. Kleinbaum 2012, S. 12).
Dies ist beispielsweise auch bei der EVS der Fall, da die obere Grenze der Einnahmen als Studienende nach einem zuvor festgelegten Zeitraum $c = 18.000$ interpretiert werden kann. Danach sind keine weiteren Beobachtungen mehr möglich, anhand der vorliegenden Daten ist lediglich ersichtlich, dass es Untersuchungsobjekte gibt, die bis zu dieser Grenze "überlebt" haben. Wann das letzte Ereignis eingetreten ist und $S(t) = 0$ gilt, lässt sich nicht ermitteln.
Die Hazardrate $h(t)$ bildet sozusagen das Gegenstück zur Survivorfunktion. Mit

$$h(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} P(t \leq T < t + \Delta t \mid T \geq t)$$

beschreibt sie das Potenzial pro Zeiteinheit für das Eintreten des Ereignisses, unter der Bedingung, dass die Untersuchungseinheit bis zum Zeitpunkt t überlebt hat (ebd., S. 11).
Sie bewegt sich zwischen 0 und ∞ und ist damit offensichtlich keine Wahrscheinlichkeit. Zudem ist sie im Gegensatz zur Survivorfunktion nicht zwingend monoton fallend, sondern kann auch steigen, oder sogar eine unimodale Form annehmen. Sie kann genutzt werden, um eine spezifische Modellform der Lebensdauern zu identifizieren.
Eine weitere wichtige Größe in der Lebensdaueranalyse ist die Dichtefunktion $f(t)$.
Diese drei Größen $S(t), h(t)$ und $f(t)$ stehen in Beziehung zueinander. So ergibt sich aus der Definition von $h(t)$ beispielsweise:

$$h(t) = \tfrac{f(t)}{S(t)} \tag{4.1}$$

und

$$h(t) = -[\tfrac{dS(t)/dt}{S(t)}] \tag{4.2}$$

respektive

$$S(t) = \exp[-\textstyle\int_0^t h(u)du] \tag{4.3}$$

und

$$S(t) = \textstyle\int_t^\infty f(u)du \tag{4.4}$$

sowie

$$f(t) = h(t)S(t) \tag{4.5}$$

Das bedeutet, dass es ausreichend ist, eine der Größen zu bestimmen, um die anderen daraus abzuleiten; eine Eigenschaft, die sich die Lebenszeitanalyse zu Nutzen macht, wie wir später noch sehen werden.
Nun, da die wichtigsten Begriffe geklärt wurden, wenden wir uns der Frage zu, wie die Schätzung der Verweildauern unter Einbeziehung zensierter Daten funktioniert. Die Lebensdaueranalyse liefert hierfür verschiedene Ansätze, die abhängig von der genauen Datensituation und der interessierenden Fragestellung angewandt werden können.
Eine der ältesten und einfachsten Methoden für die Schätzung von Verweildauern, ist die sogenannte Sterbetafelmethode, die, ihrem Namen entsprechend, ursprünglich verwendet wurde, um die Überlebenszeiten einer interessierenden Population zu beobachten, jedoch auch auf viele andere Situationen angewandt werden kann (vgl. Fahrmeir 1996, S. 317).
Sie wird verwendet, wenn Zeitintervalle und die Zahl der innerhalb eines Intervalls eingetretenen Ereignisse gegeben sind. So wird die Hazardrate im k-ten Intervall geschätzt durch

$$\hat{h}_k = \frac{m_k}{n_k - w_k/2}$$

mit
m_k als Anzahl der Ereignisse, die im k-ten Intervall eingetreten sind,
n_k als Zahl der Untersuchungseinheiten, die im k-ten Intervall noch beobachtet werden (auch als Risikomenge bezeichnet),
w_k als Zahl der zensierten Fälle im k-ten Intervall.
Dabei ist $w_k/2$ ein Korrekturterm, um die aufgetretenen Zensierungen in die Risikomenge einfließen zu lassen. Man geht also davon aus, dass bei der Hälfte der zensierten Beobachtungseinheiten ein Ereignis im k-ten Intervall aufgetreten ist, weshalb diese aus der Risikomenge herausgerechnet werden. Diese Annahme ist jedoch willkürlich gewählt (vgl. Fahrmeir 1996, S. 319).
Der Einfluss von Kovariablen wird bei dieser Methode nicht explizit modelliert, ergibt sich jedoch aus der Möglichkeit, geschätzte Verweildauern von Gruppen, die sich durch ein Merkmal unterscheiden, miteinander zu vergleichen.
Im Fall der EVS ist diese Methode trotz ihrer Einfachheit kaum geeignet, um Aussagen über das Einkommen zu treffen. Zum Einen ist die Intervallkonstruktion unnötig, da die Einkommensangaben der EVS sehr exakt und nicht innerhalb von Kategorien erfasst werden und zum anderen

berücksichtigt die Schätzung der Sterbetafelmethode zwar zensierte Daten, jedoch nur, wenn diese innerhalb eines Intervalls auftreten. Bei der EVS sind die für uns interessanten Daten aber alle am rechten Rand zensiert, über den mit der Sterbetafelmethode keine Aussagen getroffen werden können.
Eine Weiterentwicklung der Sterbetafelmethode stellt der Kaplan-Meier-Schätzer dar, der nicht länger von festen Intervallgrenzen ausgeht. Stattdessen werden die Intervallgrenzen über die geordneten Beobachtungszeitpunkte der eingetretenen Ereignisse konstruiert. Dabei gilt

$$\hat{h}_k = \frac{m_k}{|R_k|}$$

und

$$\hat{S}(t) = \prod_{k|t_(k)<t} (1 - \frac{m_k}{|R_k|})$$

mit m_k als Zahl der aufgetretenen Ereignisse zum Zeitpunkt k und $|R_k|$ als Risikomenge zum Zeitpunkt k. Auch hier werden keine Kovariablen mit einbezogen, so dass die Untersuchung auf Einflussfaktoren nur durch den Vergleich der geschätzten Verweildauern verschiedener Gruppen, zum Beispiel durch den Log-Rank-Test, möglich ist (vgl. Kleinbaum 2012, S. 67).
Wegen der im Fall der EVS verwendeten Zensierungsform, ist der KM-Schätzer ebenso ungeeignet wie die Sterbetafelmethode, da $\hat{S(t)}$ nur für die Zeitspanne bis zum größten Ereigniszeitpunkt als definiert betrachtet wird (vgl. Fahrmeir 1996, S. 321).
Ein für Verweildauern besonders beliebtes Modell ist das Proportional-Hazards- oder auch Cox-Modell. Dabei handelt es sich um einen semi-parametrischen Ansatz, der den großen Vorteil hat, dass die Verteilung der Verweildauer vor der Modellierung nicht bekannt sein muss. Weiterhin wird in diesem Modell der Einfluss von Kovariablen auf die Lebensdauer modelliert.
Für dieses Modell wird für die Hazardfunktion folgende Form angenommen:

$$h(t|x) = h_0(t) \exp(x'\beta)$$

Dabei stellt $h_0(t)$ die sogenannte Baseline Hazardrate dar, die nicht näher spezifiziert werden muss und lediglich von der Zeit abhängt, während $\exp(x'\beta)$ den Einfluss der Kovariablen auf die Hazardfunktion modelliert und dabei zeitunabhängig ist (unter der Bedingung, dass die Kovariablen zeitunabhängig sind, andernfalls muss das Modell angepasst werden)(vgl. Fahrmeir 1996, S. 314f). β wird durch die sogenannte Maximum-Partial-Likelihood geschätzt

$$PL(\beta, x_1, ..., x_N) = \prod_{n=1}^{k} \frac{exp(x'_{(n)}\beta)}{\sum_{l \in R(t_{(n)})} exp(x'_{(l)}\beta)}$$

Der Begriff der *Partial*-Likelihood entstand, da diese Schätzung nur die Wahrscheinlichkeiten der Untersuchungsobjekte beachtet, bei denen ein Ereignis eingetroffen ist. Dennoch gehen auch die Informationen, die die zensierten Daten liefern, in die Schätzung mit ein, da sie wie bei dem KM-Schätzer bis zum Zeitpunkt ihrer Zensierung in die Risikomenge einfließen (vgl. Kleinbaum 2012, S. 113).
Ein Problem des Cox-Modelles ist, dass die Güte der Schätzung unter kleinen Stichproben abnehmen kann (vgl. Fahrmeir 1996, S. 329). Nun verfügt die EVS zwar mit rund 60.000 Haushalten über eine beachtliche Stichprobengröße, doch gerade die uns interessierenden Beobachtungen am rechten Rand der Verteilung sind eher spärlich gesät. Daher wollen wir, trotz der Beliebtheit des Cox-Modells, einen alternativen Ansatz, den der Transformationsmodelle, näher betrachten.
Bei dieser Form der Regressionsmodelle handelt es sich um parametrische Modelle, bei denen von einer bekannten Fehlerverteilung für ϵ ausgegangen wird und die die Verweildauer unter Einfluss von Kovariablen modellieren.
Damit ähneln sie den herkömmlichen Regressionsmodellen, die auch außerhalb der Lebensdaueranalyse verwendet werden, unterliegen jedoch der Einschränkung, dass Lebensdauern nur nichtnegative Werte annehmen können. Um komplizierte Restriktionen für die Regressionsparameter zu vermeiden, wird die Verweildauer daher einer linearen Transformation

$$g(T) = x'\beta + \sigma\varepsilon$$

unterworfen (im Fall eines linearen Modells).
In der Praxis ist dies meist

$$g(T) = \ln T$$

Mit dieser Transformation erhält man ein lineares Modell mit logarithmierten Verweildauern
$y = lnT = x'\beta + \sigma\epsilon$
$\Rightarrow T = exp(x'\beta + \sigma\epsilon) = exp(x'\beta)exp(\sigma\epsilon) = exp(\beta_0)exp(x_i'\beta_i)exp(\sigma\epsilon)$ mit $i = 1, ..., n$
Das bedeutet, dass sich die Kovariablen multiplikativ auf die Verweildauer auswirken, weshalb diese Modellklasse auch als Accelerated-Failure-Time-Modell bezeichnet wird. Das exakte Modell ist abhängig von der angenommenen Fehlerverteilung (vgl. Fahrmeir 1996, S. 310).
In der Lebensdaueranalyse häufig verwendete Verteilungen sind die Exponentialverteilung, die einen Spezialfall der Weibullverteilung darstellt, das Log-logistische Modell, das Lognormale Modell und die Generalisierte Gammaverteilung.
Anhand der Transformationsmodelle lässt sich noch einmal der Zusammenhang zwischen den einzelnen Größen der Lebensdaueranalyse $S(t), h(t)$ und $f(t)$ darstellen. Dies soll exemplarisch an der Weibullverteilung geschehen.

Die Dichtefunktion $f(t)$ der Weibullverteilung lautet

$$f(t) = \lambda p t^{p-1} exp - (\lambda t^p)$$

aus (4.4) $S(t) = \int_t^\infty f(u)du$ ergibt sich

$$S(t) = 1 - F(t) = 1 - (1 - exp(-\lambda t^p)) = exp(-\lambda t^p)$$

und somit aus (4.2) $h(t) = -[\frac{dS(t)/dt}{S(t)}]$

$$h(t) = -\frac{d\ exp(-\lambda t^p)/dt}{exp(-\lambda t^p)} = -\frac{exp(-\lambda t^p)p(-\lambda t^{p-1})(-\lambda)}{exp(-\lambda t^p)} = \lambda p t^{p-1}$$

mit $\lambda = exp(x'\beta)$ und p als Shape-Parameter.
Der Shape-Parameter p bestimmt die Hazardfunktion
$p > 1$ führt zu zunehmender Hazardfunktion
$p < 1$ führt zu abnehmender Hazardfunktion
$p = 1$ bedeutet konstante Hazardfunktion und entspricht dem Spezialfall der Exponentialverteilung
Im log-logistischen Modell führt der Parameter p zu einer unimodalen Hazardfunktion, die zunächst zunimmt und erst nach Erreichen ihres Hochpunktes abnimmt, sofern $p > 1$, andernfalls nimmt die Hazardfunktion monoton ab. Dies gilt ebenfalls für das log-normale Modell, das, aufgrund der komplizierteren Berechnung, häufig durch das sehr ähnliche log-logistische Modell ersetzt wird (vgl. Fahrmeir 1996, S. 311f).
Der Ansatz der Transformationsmodelle hat große Ähnlichkeit mit dem bereits vorgestellten Tobit-Modell; tatsächlich können die Hazardrate und der Inverse Mills Ratio als analog zueinander betrachtet werden (vgl. Windzio 2013, S. 260); und ermöglicht es uns nicht nur, andere Verteilungen als die Normalverteilung zu betrachten, sondern konditioniert die geschätzte Verweildauer (bzw. in unserem Fall das Einkommen) auf einen nicht-negativen Bereich. Damit sind Transformationsmodelle dem Tobit-Modell und letztlich auch dem Rang-Tobit-Modell vorzuziehen, insbesondere, wenn Annahmen über die Fehlerverteilung getroffen werden können.
Die Ähnlichkeit zwischen dem Tobit-Modell und den Transformationsmodellen zeigt sich insbesondere, wenn wir das Lognormal-Modell betrachten, das ebenfalls von einer Normalverteilung der Fehler ausgeht.
In diesem Fall ergibt sich für die Hazardfunktion der transformierten Variablen (vgl. Fahrmeier

1996, S. 312)

$$\lambda(y) = \frac{\sigma^{-1}\phi(\frac{y-x'\beta}{\sigma})}{1-\Phi(\frac{y-x'\beta}{\sigma})}$$

Diese Form erinnert stark an den Inversen Mills Ratio, den wir zuvor für das Tobit-Modell betrachtet haben. Allerdings gibt es in diesem Fall keine Zensierungskonstante τ, stattdessen wird der transformierte Zeitpunkt t der Verweildauer verwendet.
Da das Einkommen eine Größe ist, deren Messung, wie im zweiten Teil der Arbeit gezeigt, zwar ihre Tücken hat, jedoch häufig durchgeführt wird, gibt es Datenquellen abseits der EVS, die herangezogen werden können, um Annahmen über die Fehlerverteilung zu treffen. Hier ist beispielsweise das SOEP zu nennen, das durch die spezielle Hocheinkommensstichprobe über ausreichend Werte am äußeren Rand der Verteilung verfügt.
Die Parameter der Transformationsmodelle werden in der Regel durch einen ML-Schätzer ermittelt. Dazu muss zunächst die Likelihood-Funktion aufgestellt werden, die wiederum abhängig von dem zugrundegelegten Zensierungsmodell ist.
Im Fall der EVS, die mit $c_n = c = 18.000$ eine konstante, nicht-zufällige Beobachtungszeit hat und damit Modell 1 entspricht, ergibt sich die Likelihood aus dem Produkt der Beiträge der einzelnen Untersuchungseinheiten n

$$L_n = f_n(t_n)^{d_n} S_n(c_n)^{1-d_n}$$

und lautet somit

$$\prod_{n=1}^{N} f_n(t_n)^{d_n} S_n(c_n)^{1-d_n}$$

beziehungsweise, wegen der Konstanten c

$$\prod_{n=1}^{N} f_n(t_n)^{d_n} S_n(c)^{1-d_n}$$

Durch den Zensierungsindikator d als Exponent, geht $f_n(t_n)$ in die Likelihood ein, wenn das Untersuchungsobjekt nicht zensiert wurde, oder im Fall einer Zensierung $S_n(c)$, was der Wahrscheinlichkeit des Überlebens am Ende des Beobachtungszeitraums entspricht.
Durch die Abhängigkeit von $f_n(t)$ und $S_n(t)$ von n kommt die Möglichkeit der Einbeziehung von Kovariablen zum Ausdruck (vgl. Fahrmeier 1996, S. 323).
Nachdem die Likelihood aufgestellt wurde, können die Parameter des Transformationsmodells durch die partielle Ableitung der logarithmierten Likelihood geschätzt werden (vgl. Kleinbaum 2012, S. 321).

Mit dem Tobit-Modell und der Lebensdaueranalyse existieren Methoden, die dazu geeignet sind, auch dann noch Informationen aus den vorliegenden Daten zu ziehen, wenn diese, wie im Fall der EVS, ab einem bestimmten Grenzwert zensiert sind.
Das Tobit-Modell ist wegen seiner relativ restriktiven Annahme der Normalverteilung zunächst nur bedingt geeignet, kann jedoch durch Transformation in ein verteilungsfreies Rang-Modell, oder Spezifikation einer alternativen Fehlerverteilung angepasst werden. Allerdings ist sein Einsatz durch die mangelnde Beschränkung auf positive Werte für Einkommensanalysen nicht ideal.
Hier sollte den Transformationsmodellen der Lebensdaueranalyse der Vorzug gegeben werden, die eben für genau diese Situationen entwickelt worden sind. Das ändert sich, wenn die interessierende Variable nicht mehr das Einkommen, sondern das Vermögen ist. In diesem Fall sind Transformationsmodelle nicht mehr sinnvoll einsetzbar. Die Anwendung auf nicht-finanzielle Aspekte dürfte sich bei beiden Ansätzen schwierig gestalten, da diese mindestens eine Intervallskalierung der abhängigen Variable annehmen.
Welches Modell letztlich verwendet werden sollte, ist abhängig von dem Zensierungsmodell, dem Wertebereich der abhängigen Variable und der (angenommenen) Fehlerverteilung.

5 Fazit und Ausblick

Nachdem lange Zeit das andere Ende der Verteilung von großem Interesse war, findet die Reichtumsforschung zunehmend mehr Beachtung. Diese kann jedoch nicht einfach nur als Kehrseite der Armutsmessung betrachtet werden, sondern führt teilweise zu gänzlich anderen Problemstellungen.
Die Fehlerquellen innerhalb dieses Forschungsbereichs sind vielfältig und nicht nur auf eine wissenschaftliche Disziplin beschränkt. Selbst häufig genutzte Datenquellen, deren Messungen konstant überprüft und weiterentwickelt werden, sehen sich im Bezug zur Reichtumsmessung mit neuen Herausforderungen konfrontiert, die über die bloße Operationalisierung des Begriffs hinausgehen.
Eine Messung, die den Grundsätzen der Repräsentationalen Messtheorie folgt, kann weder für das Sozio-oekonomische Panel, noch für die Einkommens- und Verbrauchsstichprobe angenommen werden, wobei Ersteres immerhin die Erfassung nicht-finanzieller Aspekte ermöglicht. Dennoch ist der Reichtumsbegriff derzeit noch so schwammig, dass eine Messung, die von einem zugrunde liegenden wahren Wert ausgeht, angezweifelt werden muss.
Allerdings ist die zugrunde gelegte Theorie weniger von der Datenquelle selbst, als den Nutzern dieser abhängig, die darüber entscheiden, welche Variablen sie zur Konstruktion und Überprüfung ihrer Forschungsfrage verwenden möchten. Beide Datenquellen orientieren sich zudem an

internationalen Standards, um eine bessere Vergleichbarkeit unterschiedlicher Erhebungen zu gewährleisten.
Während die Validität eines Messwerts daher auch von den Nutzern abhängt, ist es die Aufgabe des SOEP und der EVS objektive und reliable Daten bereitzustellen. Hier muss jedoch ebenfalls beachtet werden, dass die Überprüfung dieser Gütekriterien schwierig ist und deren Ergebnisse immer hinterfragt werden sollten.
Sowohl das SOEP als auch die EVS sind bemüht, Objektivität herzustellen. Dies geschieht bei beiden durch größtenteils schriftliche Befragungen, die enge Antwortmöglichkeiten vorgeben. Die exakte Erfassung aller Einnahmen und Ausgaben durch ein Haushaltsbuch macht dies bei der EVS besonders deutlich. Dafür hat das SOEP eine spezielle Interviewerbefragung entworfen und durchgeführt, die es ermöglichen sollte, Verzerrungen aufzudecken und adäquat darauf einzugehen.
Auch bei der Bestimmung der Reliabilität unterscheiden sich beide Datenquellen. So wird durch das Paneldesign des SOEP ein Test-Retest-Design ermöglicht, dessen Ergebnisse aufgrund der langen Abstände der Schwerpunktbefragungen jedoch kritisch zu betrachten sind. Bei der EVS ermöglicht dagegen die exakte Erfassung der Einnahmen und Ausgaben eines Haushalts eine Plausibilitätsprüfung, die als Paralleltest-Design betrachtet werden kann.
Ein weiteres Problem der Messungen ist die Anwendung der Klassischen Testtheorie, deren Annahmen unter Umständen verletzt werden. Dies ist insbesondere beim SOEP der Fall, das zum einen eine rückwirkende Schätzung der Einkommens- und Vermögenswerte des letzten Jahres misst, was die Gefahr von mit diesen Werten korrelierten Messfehlern erhöht, und zum anderen Steuerabgaben simuliert. Ein Vorgehen, das zur Überschätzung der Abgaben von Hocheinkommenshaushalten und damit zu einer Korrelation des Messfehlers mit dem wahren Wert einer anderen Variable führt. Ist dies der Fall, muss auch der Reliabilitätskoeffizient infrage gestellt werden.
Durch das wesentlich aufwändigere Haushaltsbuch der EVS wird die Wahrscheinlichkeit dieser Annahmeverletzungen reduziert. Allerdings werden bei dieser Hocheinkommenshaushalte mit einem monatlichen Nettoeinkommen von 18.000 EUR systematisch ausgeschlossen, da zu wenige Beobachtungen vorliegen. Ein Problem, das das SOEP durch das Ziehen eines speziellen Subsamples gelöst hat.
Durch den Ausschluss der Hocheinkommenshaushalte liegen bei der EVS, unter der Voraussetzung, dass diese Daten nicht vollständig aus der Stichprobe entfernt oder als fehlende Werte behandelt werden, zensierte Daten vor. Für diese Zensierungsmodelle wurden Methoden entwickelt, um den Informationsgehalt, der diesen innewohnt, dennoch auszunutzen.
Das Tobit-Modell ist ein Regressionsmodell, das seine Schätzung anhand der Zensierungswahrscheinlichkeit korrigiert und damit dem gewöhnlichen KQ-Schätzer im Fall zensierter Daten überlegen ist. Für die Anwendung des Modells ist es unerheblich, ob die vorliegenden Daten rechts-, links-, oder intervallzensiert sind, sofern sie über eine feste Grenze verfügen.

Allerdings ist dieses Modell mit der Annahme einer Normalverteilung der Fehler recht restriktiv und für Einkommensverteilungen vermutlich nicht anwendbar. Möglicherweise hilft hier bereits eine Datentransformation, andernfalls kann ein verteilungsfreies Rang-Tobit-Modell angewendet werden. Dieses löst jedoch auch nicht das Problem, dass die abhängige Variable Einkommen nicht negativ werden kann, da es sich auch hierbei lediglich um ein einfaches Regressionsmodell handelt.
Eine Alternative sind die Lebenszeitmodelle, die speziell für stetige, nicht-negative Variablen entwickelt wurden und sich daher auch auf das Einkommen übertragen lassen.
Aufgrund der Zensierung an einer festen, einheitlichen Grenze, sind die Methoden, die ohne die Modellierung von Einflussvariablen auskommen, im Fall der EVS nicht geeignet. Auch das semiparametrische Cox-Modell kann aufgrund der geringen Beobachtungszahl am rechten Rand zu Problemen führen.
Transformationsmodelle, die dem Tobit-Modell ähneln, sind hingegen parametrische Modelle, für die die Fehlerverteilung der Variable bekannt sein muss. Da diese durch andere Datenquellen, wie dem SOEP geschätzt werden kann, stellen Transformationsmodelle eine geeignete Alternative zum Cox-Modell dar und haben im Gegensatz zur Tobit-Regression den Vorteil, dass sie aufgrund ihrer Transformation ausschließlich positive Werte schätzen.
Mit den Transformationsmodellen und der Tobit-Regression liegen damit zwei ähnliche Methoden vor, um mit zensierten Daten umzugehen. Während die Transformationsmodelle gut geeignet sind um das Einkommen zu schätzen, könnte ein Rang-Tobit-Modell für die Schätzung von Vermögen verwendet werden.
Diese Arbeit hat sich größtenteils mit dem Einkommen befasst. Die Betrachtung des Vermögens, das positive und negative Werte annehmen kann und eine wesentlich komplexere Größe ist, da hierzu beispielsweise auch angesammelte Sozialleistungsansprüche gezählt werden können, kann noch einmal zu vollkommen anderen Problemstellungen führen. Zudem muss bedacht werden, dass bei der Bestimmung von Reichtum immer beide Größen beachtet werden müssen, da diese, anders als das im Allgemeinen bei der Armut der Fall ist, auch unabhängig voneinander auftreten können. Zählt man schließlich noch nicht-finanzielle Aspekte zu der Reichtumsdefinition, die häufig nicht intervallskaliert sind, werden wieder vollkommen neue Herausforderungen an die Datengewinnung und -analyse gestellt.
Weiterhin muss bedacht werden, dass die Zensierung von Hocheinkommenshaushalten bei der EVS nicht zufällig war, sondern aufgrund der geringen Stichprobengröße in diesen Einkommensbereichen zustande kam. Das Problem der Einkommensmessung ist daher grundsätzlich weniger ein Problem der Zensierung, sondern das einer rechtsschiefen Verteilung mit stark ausgeprägtem Randbereich. Auch hierfür existieren Methoden, oder deren Ansätze, die für eine korrekte Reichtumsmessung angewendet und weiterentwickelt werden sollten.

Literatur

[1] *Final report and recommendations*. The Canberra Group, Ottawa, 2001.

[2] *Einkommens- und Verbrauchsstichprobe 2008: Qualitätsbericht*. Wiesbaden, 2012.

[3] *Lebenslagen in Deutschland. Armuts und Reichtumsberichterstattung der Bundesregierung: Der vierte Armuts- und Reichtumsbericht der Bundesregierung*. Bonn, März 2013.

[4] ARNDT, CHRISTIAN, ROLF KLEIMANN, MARTIN ROSEMANN, JOCHEN SPÄTH und JÜRGEN VOLKERT: *Lebenslagen in Deutschland: Armuts- und Reichtumsbericht der Bundesregierung: Forschungsprojekt: Möglichkeiten und Grenzen der Reichtumsberichterstattung: Schlussbericht*. Bonn, April 2010.

[5] BALLENBERGER, NIKOLAUS, ANNA LLUIS, ERIKA VON MUTIUS, SABINA ILLI, BIANCA SCHAUB und ANDREW ROWLAND DALBY: *Novel Statistical Approaches for Non-Normal Censored Immunological Data: Analysis of Cytokine and Gene Expression Data*. PLoS ONE, 7(10):e46423, 2012.

[6] BECKER, IRENE, JOACHIM R. FRICK, MARKUS M. GRABKA, RICHARD HAUSER, PETER KRAUSE und GERT G. WAGNER: *A Comparison of the Main Household Income Surveys for Germany: EVS and SOEP*. In: HAUSER, RICHARD und IRENE BECKER (Herausgeber): *Reporting on Income Distribution and Poverty*, Seiten 55–90. Springer Berlin Heidelberg, Berlin, Heidelberg, 2003.

[7] CONOVER, W. J. und RONALD L. IMAN: *Rank Transformations as a Bridge Between Parametric and Nonparametric Statistics*. The American Statistician, 35(3):124, 1981.

[8] DIEKMANN, ANDREAS: *Empirische Sozialforschung: Grundlagen, Methoden, Anwendungen*, Band 55678 der Reihe *Rororo Rowohlts Enzyklopädie*. Rowohlt-Taschenbuch-Verl., Reinbek bei Hamburg, 19 Auflage, 2008.

[9] FAHRMEIR, L., ALFRED HAMERLE und GERHARD TUTZ: *Multivariate statistische Verfahren*. Walter de Gruyter, Berlin, 2., überarbeitete Aufl Auflage, 1996.

[10] HÄDER, MICHAEL: *Empirische Sozialforschung: Eine Einführung*. Lehrbuch. VS Verlag für Sozialwissenschaften, Wiesbaden, 2 Auflage, 2010.

[11] HAND, DAVID J.: *Statistics and the Theory of Measurement*. Journal of the Royal Statistical Society. Series A (Statistics in Society), 159(3):445–492, 1996.

[12] HAUSER, RICHARD und IRENE BECKER (Herausgeber): *Reporting on Income Distribution and Poverty*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2003.

[13] KARPFINGER, CHRISTIAN und KURT MEYBERG: *Algebra*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.

[14] KLEINBAUM, DAVID G. und MITCHEL KLEIN: *Survival analysis: A self-learning text*. Statistics for biology and health. Springer, New York, NY, 3 Auflage, 2012.

[15] KLOCKE, ANDREAS: *Methoden der Armutsmessung. Einkommens-, Unterversorgungs-, Deprivations- und Sozialhilfekonzept im Vergleich*. Zeitschrift für Soziologie, 29(4):313–329, 2000.

[16] MICHELL, JOEL: *Measurement scales and statistics: A clash of paradigms*. Psychological Bulletin, 100(3):398–407, 1986.

[17] MOOSBRUGGER, HELFRIED und AUGUSTIN KELAVA: *Testtheorie und Fragebogenkonstruktion*. Springer-Lehrbuch. Springer Medizin Verlag Heidelberg, Berlin, Heidelberg, 2007.

[18] QUATEMBER, ANDREAS: *Die Tauglichkeit des Quotenverfahrensschemas unter Urnenmodellbedingungen*. Austrian Journal of Statistics, 26(2):7–25, 1997.

[19] SCHNELL, RAINER, PAUL B. HILL und ELKE ESSER: *Methoden der empirischen Sozialforschung*. Oldenbourg, R, München, 9 Auflage, 2011.

[20] WAGNER, GERT G., JOACHIM R. FRICK, JAN GOEBEL, GRABKA, MARKUS, M., RAINER PISCHNER und JÜRGEN SCHUPP: *High Incomes in Household Surveys: Sample G of the German Socio-Economic Panel Study (SOEP)*. Berlin, 2007.

[21] WAGNER, GERT G., JOACHIM R. FRICK und JÜRGEN SCHUPP: *The German Socio-Economic Panel Study (SOEP) – Scope, Evolution and Enhancements*. Schmollers Jahrbuch, 127(1):139–169, 2007.

[22] WAGNER, GERT G., JAN GÖBEL, PETER KRAUSE, RAINER PISCHNER und INGO SIEBER: *Das Sozio-oekonomische Panel (SOEP): Multidisziplinäres Haushaltspanel und Kohortenstudie für Deutschland – Eine Einführung (für neue Datennutzer) mit einem Ausblick (für erfahrene Anwender)*. AStA Wirtschafts- und Sozialstatistisches Archiv, 2(4):301–328, 2008.

[23] WINDZIO, MICHAEL: *Regressionsmodelle für Zustände und Ereignisse: Eine Einführung*. Studienskripten zur Soziologie. Springer VS, Wiesbaden, 2013.

[24] WOLF, CHRISTOF: *Handbuch der sozialwissenschaftlichen Datenanalyse*. VS, Verl. für Sozialwiss., Wiesbaden, 1 Auflage, 2010.

Eigenständigkeitserklärung

Ich versichere, dass ich die vorgelegte Bachelorarbeit “Reichtumsforschung in Deutschland - Herausforderung der Datengewinnung auf Basis messtheoretischer Grundlagen und Anwendung geeigneter statistischer Verfahren zur Analyse etablierter Datenquellen” eigenständig und ohne fremde Hilfe verfasst, keine anderen als die angegebenen Quellen verwendet und die den benutzten Quellen entnommenen Passagen als solche kenntlich gemacht haben. Diese Bachelorarbeit ist in dieser oder einer ähnlichen Form in keinem anderen Kurs vorgelegt worden.

München, 29. September 2014 Denise Gawron